
Artificial Intelligence and Machine Learning in Healthcare

Ankur Saxena • Shivani Chandra
Editors

Artificial Intelligence and Machine Learning in Healthcare

 Springer

Editors

Ankur Saxena
Amity University
Noida, Uttar Pradesh, India

Shivani Chandra
Amity University
Noida, Uttar Pradesh, India

ISBN 978-981-16-0810-0

ISBN 978-981-16-0811-7 (eBook)

<https://doi.org/10.1007/978-981-16-0811-7>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2021, corrected publication 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

1	Practical Applications of Artificial Intelligence for Disease Prognosis and Management	1
	Ankur Chaurasia, Sakshi Vats, Abhishek Sengupta, Abhay Bansal, and Priyanka Narad	
2	Automated Diagnosis of Diabetes Mellitus Based on Machine Learning	37
	Manish Edida, N. Jaya Lakshmi, and Neetu Jabalia	
3	Artificial Intelligence in Personalized Medicine	57
	Khammampalli Srija, P. P. R. Prithvi, Ankur Saxena, Alka Grover, Shivani Chandra, and Shalini Jauhari Jain	
4	Artificial Intelligence in Precision Medicine: A Perspective in Biomarker and Drug Discovery	71
	Seneka Santoshi and Dipankar Sengupta	
5	Transfer Learning in Biological and Health Care	89
	Robin Sinha, Ankur Saxena, and Smitha Mony Sreedharan	
6	Visualization and Prediction of COVID-19 Using AI and ML	99
	Dilip Kumar J. Saini, Dharendra Siddharth, and Ajay Kumar	
7	Machine Learning Approaches in Detection and Diagnosis of COVID-19	113
	S. Suba and Nita Parekh	
8	Applications of Machine Learning Algorithms in Cancer Diagnosis	147
	Amita Sharma, Shalini Jain, and Sreemoyee Chatterjee	
9	Use of Artificial Intelligence in Research and Clinical Decision Making for Combating Mycobacterial Diseases	183
	Mukul Sharma and Pushpendra Singh	

10 Bias in Medical Big Data and Machine Learning Algorithms	217
Ankur Saxena, Mohit Saxena, and Alejandra Rodriguez Ilerena	
Correction to: Artificial Intelligence and Machine Learning in Healthcare	C1
Ankur Saxena and Shivani Chandra	

The original version of the book was revised: reposted as an edited volume, chapter authors’ details have been included on the table of contents and each chapter’s opening pages. The erratum to the book is available at https://doi.org/10.1007/978-981-16-0811-7_11

About the Editors

Dr. Ankur Saxena is currently working as an Assistant Professor at Amity University, Noida, Uttar Pradesh. He has been teaching graduate and post-graduate students for more than 15 years and has 3 years of industrial experience in software development. He has published 5 books and more than 40 research articles in reputed journals and is an editorial board member and reviewer for several journals. His research interests include cloud computing, big data, machine learning, evolutionary algorithms, software frameworks, design and analysis of algorithms, and biometric identification.

Dr. Shivani Chandra is an Assistant Professor at Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Noida. She has more than 20 years of experience in biotechnology and molecular biology. Her research interests include genomics analysis, computational biology, and bioinformatics data analysis. She has submitted more than 4000 clones to the NCBI GenBank and was one of the key players in the Rice Genome Sequencing Project. She has published several research articles in genome sequencing, comparative genomics, and genome analysis in reputed journals. She has more than 15 years of teaching experience in computational biology, molecular biology, genetics, recombinant DNA technology, and bioinformatics.

List of Figures

Fig. 1.1	Artificial intelligence, machine learning, and deep learning	2
Fig. 1.2	AI in disease management	5
Fig. 1.3	Overall process of the application of AI in disease prognosis and diagnosis	7
Fig. 1.4	AI/ML techniques help to identify the biomarker of a disease from multidimensional data	9
Fig. 1.5	AI in drug development	11
Fig. 1.6	Kaggle homepage. (www.kaggle.com)	12
Fig. 1.7	An example to show the preview of the file's contents is visible in the data explorer by clicking on the data tab of dataset on Kaggle. (www.kaggle.com)	13
Fig. 1.8	Kaggle search box. (www.kaggle.com)	14
Fig. 1.9	UCI machine learning repository home page (including search box). (archive.ics.uci.edu)	15
Fig. 1.10	Preview of "View ALL Datasets" tab. (archive.ics.uci.edu)	15
Fig. 1.11	List of datasets present in UCI Machine Learning Repository after clicking "View ALL Datasets" tab. (archive.ics.uci.edu) ...	16
Fig. 1.12	Example of dataset window opened in UCI Machine Learning Repository. (archive.ics.uci.edu)	17
Fig. 1.13	HealthData.gov home page. (catalog.data.gov)	17
Fig. 1.14	Artificial neural network model	21
Fig. 1.15	Code for importing the Parkinson's disease data	25
Fig. 1.16	Parkinson's disease dataset imported in MATLAB	25
Fig. 1.17	Code for checking missing value in dataset	25
Fig. 1.18	Output of missing value code	26
Fig. 1.19	Code for outlier detection	26
Fig. 1.20	Feature scaling code	27
Fig. 1.21	Feature selection code	28
Fig. 1.22	Output of explained variance percentage along with graphical representation	29
Fig. 1.23	New table of dataset (after PCA)	30
Fig. 1.24	Building classifier (SVM, KNN and Naive Bayes) code	30

Fig. 1.25	Output of different classifiers	31
Fig. 1.26	Code for dividing the dataset into training and testing set	31
Fig. 1.27	Output of train and test size of dataset	31
Fig. 1.28	Code to train a model	32
Fig. 1.29	Confusion matrix of SVM model	32
Fig. 1.30	Confusion matrix of KNN model	32
Fig. 1.31	Confusion matrix of Naive Bayes model	33
Fig. 2.1	Global prevalence of diabetes mellitus (Source: American Diabetes Association)	39
Fig. 2.2	Basic flow chart of a disease diagnostic AI model	42
Fig. 2.3	Reinforcement learning architecture	44
Fig. 2.4	Machine learning applications in diabetes management	45
Fig. 2.5	Flow chart of methodology	48
Fig. 2.6	Confusion matrix of k-means clustering	49
Fig. 2.7	Performance chart	52
Fig. 2.8	F1 scores of the classification models	52
Fig. 3.1	Most commonly used models of artificial intelligence in healthcare	62
Fig. 3.2	Categories of machine learning used in personalized medicine. The data is obtained by the search of algorithms in PubMed	63
Fig. 3.3	Supervised and unsupervised learning models mostly used in personalized medicine. The data is obtained by the search of algorithms in PubMed	64
Fig. 3.4	Decision-making by classification in SVM	66
Fig. 3.5	Process of the ANN	67
Fig. 4.1	Artificial intelligence can help in gaining insights from the heterogeneous datasets (clinical, omics, environmental, and lifestyle data), mapping genotype-phenotype relationships, and identifying novel biomarkers for patient diagnostics and prognosis against a specific disease	75
Fig. 4.2	Application of artificial intelligence in various steps of drug discovery process (Paul et al. 2020)	80
Fig. 4.3	Some examples of pharmaceutical companies collaborating with artificial intelligence (AI) organization for healthcare improvements in the field of oncology, cardiovascular diseases, and central nervous system disorders (Paul et al. 2020)	84
Fig. 5.1	Modified VGG-16 model	93
Fig. 5.2	Modified EfficientNetB4 model	94
Fig. 5.3	Modified Inception-ResNet-V2 model	95
Fig. 5.4	Modified Inception-V3 model	96
Fig. 5.5	Comparison between accuracies on testing dataset generated by retrained transfer learning models	96
Fig. 5.6	Comparison between various evaluation parameters such as accuracy, sensitivity, specificity, and area under the curve on testing dataset generated by retrained transfer learning models . .	97

Fig. 6.1	Daily COVID-19 confirmed, death, and recovered cases	105
Fig. 6.2	Highly affected regions for COVID-19 confirmed, active, recovered, and tested cases in India	106
Fig. 6.3	COVID-19 confirmed, active, recovered, and tested cases in India	108
Fig. 7.1	Flowchart of the study by Fang et al. to assess the performance of CT scans for the detection of COVID-19 comparison to RT-PCR (reproduced from (Fang et al. 2020))	116
Fig. 7.2	Chest X-ray image on day 3 of a COVID-19 patient (left) clearly indicates right mid and lower zone consolidation; on day 9 (right) is seen worsening oxygenation with diffuse patchy airspace consolidation in the mid and lower zones. (Case courtesy of Dr. Derek Smith, Radiopaedia.org, rID: 75249)	116
Fig. 7.3	CT scan image performed to assess the degree of lung injury of the patient in Fig. 7.2 on day 13 (left coronal lung window, right axial lung window). Multifocal regions of consolidation and ground-glass opacifications with peripheral and basal predominance. (Case courtesy of Dr. Derek Smith, Radiopaedia.org, rID: 75249)	117
Fig. 7.4	Typical convolutional network framework for classifying COVID-19 cases, which takes as input CXR images and passes through a series of convolution, pooling, and dense layers and uses a softmax function to classify an image as COVID-19 infected with probabilistic values between 0 and 1	119
Fig. 7.5	ResNet block where the input F_l^k is added to the transformed signal $g_c(F_{l \rightarrow m}^k, k_{l \rightarrow m})$ to enable cross-layer connectivity. (Reproduced from (Khan et al. 2020a))	123
Fig. 7.6	COVID-Net architecture. (Reproduced from (Wang and Wong 2020))	123
Fig. 7.7	CoroNet architecture. AE_H and AE_P are the two autoencoders trained independently on healthy and non-COVID pneumonia subjects, respectively. TFEN is a Feature Pyramid-based Autoencoder (FPAE) network, with seven layers of convolutional encoder blocks and decoder blocks, while CIN is a pre-trained ResNet-18 network. (Reproduced from (Khobahi et al. 2020))	125
Fig. 7.8	COVNet architecture. Features are extracted from each CT scan slice which are combined using max-pooling operation and submitted to a dense layer, which generates scores for the three classes. (Reproduced from (Li et al. 2020))	126
Fig. 7.9	Block diagram of the subsystem (a) performs a 3D analysis of CT scans, for identifying lung abnormalities, and subsystem (b) that performs a 2D analysis of each slice of CT scans, for	

	detecting and marking large-sized ground-glass opacities using proposed method (reproduced from (Gozes et al. 2020))	127
Fig. 7.10	(a) Workflow of the AI system data divided into four nonoverlapping cohorts for training, internal validation, external testing, and expert reader validation. (b) Usage of the AI system—performs lung segmentation on CT images and diagnosis of COVID-19 and locates abnormal slices (reproduced from (Jin et al. 2020))	129
Fig. 7.11	Inception V3 architecture has a deeper architecture compared to ResNet (source https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#d27e)	130
Fig. 7.12	Xception architecture introduced depth-wise separable convolutions (source https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#d27e)	131
Fig. 7.13	DenseNet architecture connects feature maps of all previous layers to subsequent layers (source https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803)	132
Fig. 7.14	VGG architecture has a narrow topology (source https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d#d27e)	132
Fig. 7.15	LSTM architecture employs gates to regulate flow of information across layers (source http://colah.github.io/posts/2015-08-Understanding-LSTMs/)	132
Fig. 7.16	@Original Inception Net Architecture (above), truncated Inception Net architecture (below). (Reproduced from (Das et al. 2020))	133
Fig. 7.17	Dataflow in the DL model using data augmentation (reproduced from (Sedik et al. 2020))	134
Fig. 7.18	Architecture used in the study by (reproduced from (Brunese et al. 2020))	134
Fig. 7.19	Illustration of the COVID-19Net model (reproduced from (Wang et al. 2020))	136
Fig. 7.20	Abnormal lung regions identified by GSInquire leveraged from the update parameters generated by the Inquisitor of the generator-inquisitor pair after probing the response signals from the generated network with respect to the input signal and target label. (Reproduced from (Wang and Wong 2020))	138
Fig. 7.21	Attribution maps for five random patients for the three classifications considered. Yellow regions represent most salient and blue regions the least salient regions as indicated by the color bar (reproduced from (Khobahi et al. 2020))	139

Fig. 7.22	Attention heatmaps generated by GRAD-CAM. The red regions indicate the activation regions associated with a sample. (Reproduced from (Li et al. 2020))	140
Fig. 7.23	DL discovered suspicious lung areas learned by COVID-19Net. (Reproduced from (Wang et al. 2020))	140
Fig. 8.1	Categorization of machine learning algorithms	151
Fig. 8.2	Machine learning algorithms	152
Fig. 8.3	Tasks and metrics	153
Fig. 8.4	Applications of ML in cancer prediction/prognosis	153
Fig. 8.5	Knowledge discovery process	157
Fig. 8.6	Flowchart for cancer prediction using ML	157
Fig. 8.7	SVM with different classifiers. Source: https://miro.medium.com/max/2560/1*dh0lzq0QNCOyRIX1Ot4Vow.jpeg	160
Fig. 8.8	An example of artificial neural networks	160
Fig. 8.9	The flow diagram of Naive Bayes in machine learning (Source: https://i.stack.imgur.com)	161
Fig. 8.10	ROC curve	162
Fig. 8.11	Flowchart in Orange tool	163
Fig. 8.12	Performance comparison of machine learning models	164
Fig. 8.13a	Confusion matrix for liver cancer dataset using SVM	164
Fig. 8.13b	Confusion matrix for liver cancer dataset using NN	165
Fig. 8.13c	Confusion matrix for liver cancer dataset using Naive Bayes	165
Fig. 8.14a	ROC curve for class 1	166
Fig. 8.14b	ROC curve for class 2	167
Fig. 8.15	Neural networks model using RStudio	168
Fig. 8.16	Predictive model using the Orange tool on prostate cancer dataset	169
Fig. 8.17a	Confusion matrix for prostate cancer dataset using SVM	170
Fig. 8.17b	Confusion matrix for prostate cancer dataset using Naive Bayes	170
Fig. 8.17c	Confusion matrix for prostate cancer dataset using neural networks	171
Fig. 8.18	Curve of receiver operating characteristics for prostate cancer dataset	172
Fig. 8.19	Neural networks model by RStudio	173
Fig. 8.20	Classification matrix of neural networks model by RStudio	173
Fig. 8.21	Performance comparison of machine learning models for breast cancer dataset	174
Fig. 8.22a	Confusion matrix for breast cancer dataset using SVM	175
Fig. 8.22b	Confusion matrix for breast cancer dataset using NN	175
Fig. 8.22c	Confusion matrix for breast cancer dataset using Naive Bayes	176
Fig. 8.23	ROC curve for breast cancer dataset	177

Fig. 8.24	NN model for breast cancer dataset using RStudio	178
Fig. 8.25	Classification matrix of neural networks model by RStudio	178
Fig. 9.1	The picture displays the interconnected gene expression domains, from genome to metabolite. Using microarrays, sequencing, and Mass spectrometry at each stage reveals to get multi-level gene and protein expression, these techniques delivered a multidimensional view of both natural and pathological processes	185
Fig. 9.2	Schematic representation of the steps involved in traditional drug discovery process vs. AI based drug repurposing with the salient features of both the processes	192
Fig. 9.3	Data accumulation at EMBL-EBI by data resource over time. The y-axis shows total bytes for a single copy of the data resource over time. Resources shown are the BioImage Archive, Proteomics IDentifications (PRIDE), European Genome-Phenome Archive (EGA), ArrayExpress, European Nucleotide Archive (ENA), Protein Data Bank in Europe and MetaboLights. The y-axis for both charts is logarithmic, so not only are most data types growing, but the rate of growth is also increasing. For all data resources shown here, growth rates are predicted to continue increasing. From Cook et al., NAR, 2020	194
Fig. 9.4	Schematic representation of the steps involved in AI-based prediction models for genomic applications	197
Fig. 9.5	The image depicts diverse applications of artificial intelligence in healthcare. The ability of AI to learn and rewrite its own rules, through Machine Learning and Deep Learning, offers not only benefits for today but also yet unseen capabilities for tomorrow	201
Fig. 10.1	Overview of Bias	221

List of Tables

Table 1.1	Flow chart of ANN process	20
Table 2.1	List of pathological investigation for diabetes mellitus	40
Table 2.2	Attributes in Pima Indians dataset	48
Table 2.3	Evaluation parameters of different predictive models	51
Table 5.1	<i>Description of dataset</i> : we have in total 253 brain MRI images out of which 155 are having tumor and 98 are normal	92
Table 5.2	<i>Description of dataset type</i> : we have in total 253 brain MRI images. We split our whole dataset into three different parts: training, validation, and testing dataset	92
Table 5.3	<i>Evaluation parameter results of various models</i> : we evaluated our transfer learning models using parameters such as accuracy, sensitivity, specificity, and area under the curve	97
Table 7.1	List of popular architectures reviewed in this chapter	122
Table 8.1	Liver cancer dataset	156
Table 8.2	Prostate cancer dataset	156
Table 8.3	Breast cancer dataset	156
Table 8.4	Confusion matrix generated by ANN for liver cancer dataset in RStudio	168