

**Algorithms for Intelligent Systems**

*Series Editors:* Jagdish Chand Bansal · Kusum Deep · Atulya K. Nagar

Anand J. Kulkarni

Suresh Chandra Satapathy *Editors*

---

# Optimization in Machine Learning and Applications



Springer

# **Algorithms for Intelligent Systems**

## **Series Editors**

Jagdish Chand Bansal, Department of Mathematics, South Asian University,  
New Delhi, Delhi, India

Kusum Deep, Department of Mathematics, Indian Institute of Technology Roorkee,  
Roorkee, Uttarakhand, India

Atulya K. Nagar, Department of Mathematics and Computer Science,  
Liverpool Hope University, Liverpool, UK

This book series publishes research on the analysis and development of algorithms for intelligent systems with their applications to various real world problems. It covers research related to autonomous agents, multi-agent systems, behavioral modeling, reinforcement learning, game theory, mechanism design, machine learning, meta-heuristic search, optimization, planning and scheduling, artificial neural networks, evolutionary computation, swarm intelligence and other algorithms for intelligent systems.

The book series includes recent advancements, modification and applications of the artificial neural networks, evolutionary computation, swarm intelligence, artificial immune systems, fuzzy system, autonomous and multi agent systems, machine learning and other intelligent systems related areas. The material will be beneficial for the graduate students, post-graduate students as well as the researchers who want a broader view of advances in algorithms for intelligent systems. The contents will also be useful to the researchers from other fields who have no knowledge of the power of intelligent systems, e.g. the researchers in the field of bioinformatics, biochemists, mechanical and chemical engineers, economists, musicians and medical practitioners.

The series publishes monographs, edited volumes, advanced textbooks and selected proceedings.

More information about this series at <http://www.springer.com/series/16171>

Anand J. Kulkarni · Suresh Chandra Satapathy  
Editors

# Optimization in Machine Learning and Applications

*Editors*

Anand J. Kulkarni  
Department of Mechanical Engineering  
Symbiosis Institute of Technology  
Pune, Maharashtra, India

Suresh Chandra Satapathy  
School of Computer Engineering  
Kalinga Institute of Industrial  
Technology (KIIT)  
Bhubaneswar, Odisha, India

ISSN 2524-7565

ISSN 2524-7573 (electronic)

Algorithms for Intelligent Systems

ISBN 978-981-15-0993-3

ISBN 978-981-15-0994-0 (eBook)

<https://doi.org/10.1007/978-981-15-0994-0>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved 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

# Preface

The book aims to discuss multifaceted and state-of-the-art literature survey of machine learning (ML) techniques along with mathematical formulations of the underlying heuristics and metaheuristics. The contributions may further help to explore new research avenues leading towards multidisciplinary research discussions. It also discusses associated original mathematical problem formulations and solutions, along with the comparative analysis.

More specifically, Chap. 1 elaborates the use of artificial neural network (ANN) for abnormality detection in medical images as one of the major contributions of the book. It exhibited analysis procedures to improvise the classification of X-ray images, which may enhance the detection probability at a very early stage of the disease like cancer. Criminal activities affect our societies in many adverse ways. Predicting the locations where future crimes are more likely to happen would immensely help the police forces all over the world in preventing likely crimes. Chapter 2 presents a deep learning-based approach for predicting the future crime hot spots from past crime data. The fundamentals associated with the ML literature along with a critical survey of associated optimization techniques have been discussed in Chap. 3. It also presents a general description of decision tree and outlines ensembles methods, the different ways to generate them along with their selection processes and criteria. Similarly, with a goal of developing a new algorithm in order to classify types of tumours, Chap. 4 proposes a framework for the stage-wise classification of several magnetic resonance images (MRI). Chapter 5 offers an application of an evolutionary algorithm such as genetic programming for predictive analysis of water quality of reservoirs and lakes. In addition, it also throws light on a functional relationship between features in the associated real hydrological data. Importantly, authors have developed cause-and-effect models and used spectral analysis to eliminate the issues in handling complex issues such as time lag, cross-validation and overfitting. The work related to two well-known nature inspired optimization algorithms, i.e. genetic algorithm (GA) and particle swarm optimization (PSO), in ML domain is critically reviewed in Chap. 6. The comprehensive yet critical study of both the approaches may help researchers to apply similar techniques to identify optimized solution to several types of problems of interest. Chapter 7 discusses an attempt to hybridize the fuzzy C-means

with a robust optimization metaheuristic belonging to the class of socio-inspired optimizers referred to as Cohort Intelligence (CI) algorithm. The method is validated by testing it on the Breast Cancer Wisconsin Diagnostic Data set. Land suitability assessment is an important activity to evaluate the land performance for alternative kind of agriculture based on a variegated parameters. The authors of Chap. 8 proposed a method for climate data process referred to as Day wise Spatial Climate Data Generation Process which attempts to automate the process of generating spatial representation of climate data. This process offers the agricultural experts an easy technique to study the spatial variation of climate parameters and may be of help in contingency planning for the area under consideration. Chapter 9 provides a review of modern advancement made towards video-based group activity recognition technique. The chapter also provides a comprehensive review on the latest progress in deep learning and recent developments in group activity recognition performance. This review may serve as a rich reference discussing diverse applications and the models described in different applications associated with surveillance, sport analytics and video summary, etc. The field of journalism is no exception to artificial intelligence (AI) and ML. Chapter 10 provides a detailed assessment of the implications of AI in journalism worldwide. The chapter highlights the immense impact of AI on the ecosystem of media market across the globe. Also, it is underscored that the AI techniques can have enough scope to create social good to assist the human to navigate the required out of a huge pool of data by personalized recommendations. A review of how the face of public relations has changed with the interventions from AI and ML is highlighted in Chap. 11. The chapter emphasized on man-machine relationship which needs to be judiciously dealt for the greater interest of the society. Chapter 12 focuses on designing an efficient transmission policy for energy harvesting sensors by estimating its state through channel gain estimation using different computational intelligence techniques. The chapter proposes a novel computational technique which exploits roulette wheel selection approach for estimation. Its performance is compared with ANN, extreme learning machine under the same simulation environment. In addition, the chapter contributes a new collaborative transmission policy amongst the nodes of a wireless sensor network for performance improvement.

Every chapter submitted to the book was critically evaluated by at least two expert reviewers. The critical suggestions by the reviewers certainly helped and influenced the authors of individual chapter to enrich the quality in terms of experimentation, performance evaluation, representation, etc. The book may serve as a valuable reference for the metaheuristic optimization methods and application associated with machine learning domain.

Pune, India  
Bhubaneswar, India

Anand J. Kulkarni  
Suresh Chandra Satapathy

# Contents

<b>1</b>	<b>Use of Artificial Neural Network for Abnormality Detection in Medical Images . . . . .</b>	<b>1</b>
	Prachi R. Rajarapollu, Debasish Adhikari and Nutan V. Bansode	
<b>2</b>	<b>Deep Learning Techniques for Crime Hotspot Detection . . . . .</b>	<b>13</b>
	Sankar N. Nair and E. S. Gopi	
<b>3</b>	<b>Optimization Techniques for Machine Learning . . . . .</b>	<b>31</b>
	Souad Taleb Zouggar and Abdelkader Adla	
<b>4</b>	<b>A Package Including Pre-processing, Feature Extraction, Feature Reduction, and Classification for MRI Classification . . . . .</b>	<b>51</b>
	Alireza Balavand and Ali Husseinzadeh Kashan	
<b>5</b>	<b>Predictive Analysis of Lake Water Quality Using an Evolutionary Algorithm . . . . .</b>	<b>69</b>
	Mrunalini Jadhav, Kanchan Khare, Sayali Apte and Rushikesh Kulkarni	
<b>6</b>	<b>A Survey on the Latest Development of Machine Learning in Genetic Algorithm and Particle Swarm Optimization . . . . .</b>	<b>91</b>
	Dipti Kapoor Sarmah	
<b>7</b>	<b>A Hybridized Data Clustering for Breast Cancer Prognosis and Risk Exposure Using Fuzzy C-means and Cohort Intelligence . . . . .</b>	<b>113</b>
	Meeta Kumar, Anand J. Kulkarni and Suresh Chandra Satapathy	
<b>8</b>	<b>Development of Algorithm for Spatial Modelling of Climate Data for Agriculture Management for the Semi-arid Area of Maharashtra in India . . . . .</b>	<b>127</b>
	Vidya Kumbhar and T. P. Singh	

<b>9 A Survey on Human Group Activity Recognition by Analysing Person Action from Video Sequences Using Machine Learning Techniques . . . . .</b>	<b>141</b>
Smita Kulkarni, Sangeeta Jadhav and Debashis Adhikari	
<b>10 Artificial Intelligence in Journalism: A Boon or Bane? . . . . .</b>	<b>155</b>
Santosh Kumar Biswal and Nikhil Kumar Gouda	
<b>11 The Space of Artificial Intelligence in Public Relations: The Way Forward . . . . .</b>	<b>169</b>
Santosh Kumar Biswal	
<b>12 Roulette Wheel Selection-Based Computational Intelligence Technique to Design an Efficient Transmission Policy for Energy Harvesting Sensors . . . . .</b>	<b>177</b>
Shaik Mohammad, E. S. Gopi and Vineetha Yogesh	
<b>Author Index . . . . .</b>	<b>197</b>

# About the Editors

**Anand J. Kulkarni** holds a Ph.D. in Distributed Optimization from Nanyang Technological University, Singapore; an M.S. in AI from the University of Regina, Canada; and Bachelor of Engineering from Shivaji University, India. He worked as a Research Fellow on a cross-border supply-chain disruption project at Odette School of Business, University of Windsor, Canada. Currently, he is the Head and an Associate Professor at the Symbiosis Institute of Technology, Pune, India. His research interests include optimization algorithms, multiobjective optimization, multiagent systems, complex systems, swarm optimization, game theory, and self-organizing systems. He is the founder and Chairman of the OAT Research Lab. Anand has published over 40 research papers in peer-reviewed journals and conferences as well as two books.

**Suresh Chandra Satapathy** is a Professor at the School of Computer Engineering, KIIT, Odisha, India. Previously, he was a Professor and the Head of the Department of CSE at ANITS, AP, India. He received his Ph.D. in CSE from JNTU, Hyderabad, and M.Tech. in CSE from the NIT, Odisha. He has more than 27 years of teaching and research experience. His research interests include machine learning, data mining, swarm intelligence and applications. He has published more than 98 papers in respected journals and conferences and has edited numerous volumes for Springer AISC and LNCS. In addition to serving on the editorial board of several journals, he is a senior member of the IEEE and a life member of the Computer Society of India, where he is the National Chairman of Division-V (Education and Research).

# Chapter 1

# Use of Artificial Neural Network for Abnormality Detection in Medical Images



Prachi R. Rajarapollu, Debashis Adhikari and Nutan V. Bansode

## 1 Introduction

This chapter implements classification analysis procedures to improvise the image classification of X-ray images of the image samples which will enhance the detection probability of abnormalities at a very initial stage. Thirty percent of radiologist fails to detect the malignancy at an early stage resulting in a possibility of reducing false-positive results. These false-positive (FP) results can be due to inter-observatory analysis errors due to different faults in rib vessel and its structuring [1, 2]. Thus, reduction in FP images and increased true-positive (TP) images is important for an accurate analysis of the X-ray. Several ways have been identified by researchers for the reduction in the FP results.

There are two prominent methods of reduction in the FP in image processing; feature-based analysis and morphological-based analysis. Analysis is done in two phases like extraction of features and classification of features. Morphological analysis can be done on the basis of circularity, size, contrast and local curvature, etc. The traditional analysis algorithm used is glass lens medical digital-cams in real time. Numerous difficulties may have been faced by conventional auto-focus algorithms. The main prominent difficulty is the repeated interpolation of data and increasing calculations due to the same.

The important tasks the radiologist needs to carry out are detection and identification of cancerous cells. The FP results arise due to factors like rib crossing, vessel

---

P. R. Rajarapollu (✉) · D. Adhikari · N. V. Bansode  
School of Electrical Engineering, MIT AOE, Pune, India  
e-mail: [prrajarapollu@entc.maepune.ac.in](mailto:prrajarapollu@entc.maepune.ac.in)

D. Adhikari  
e-mail: [dadhikari@entc.mitaoe.ac.in](mailto:dadhikari@entc.mitaoe.ac.in)

N. V. Bansode  
e-mail: [nvbansode@entc.maepune.ac.in](mailto:nvbansode@entc.maepune.ac.in)

crossings, rib vessel crossings which can be mistaken as malignant tumors. Thus, the precise detection of the cancer will highly depend on reduction in such factors that tend to be FP images and a considerable increase in the true-positive (TP) results. There are a number of ways to reduce faulty results. These methods mostly work based on extraction of features and classification of features. Some of the authors tried feature extraction and implemented with artificial neural networks (ANN) [3, 4]. The critical task is to identify faulty cells. Even for experienced radiologists, it is very difficult and risky job to distinguish normal and abnormal cells. The tool specifically used by pathologist is WSI, but it is lacking in automation and classification of features which is an important parameter for early diagnosis of disease [5].

## 2 Literature Survey

In case of breast cancer with mammographic images, identification and interpretation of the signs of breast cancer by screening algorithm are a challenging task. In [1], a new technique of pattern recognition and detection of breast cancer has been elaborated. Author worked on bilateral asymmetry identification, detection and classification of regions of interest. Here, only the conclusion drawn was whether benign lesions or malignant tumors are present. With the help of artificial intelligence and various other algorithms, analysis has been done [3]. Malignant pleural mesothelioma from thoracic CT scans method has been used for texture-based segmentation. Automatic sampling and a manual sampling are used to extract statistical features from the MPM texture [6]. Texture analysis of gradient images has been analyzed for the categorization of mammographic masses as benign or malignant. Local feature extraction has been done by wavelet transform and Fisher linear discriminant analysis to obtain better results [7]. Malignant skin tissues have been detected with the help of a comprehensive dielectric spectroscopy study [8]. A statistical analysis has been carried out in [9] to ascertain when classifiers are to be used to differentiate between benign and malignant thyroid nodules. For classification of the benign and malignant state, an algorithm based on fuzzy inference system has been used in [10]. Here, a new medical expert system has been developed which is helping in the diagnosis of pulmonary diseases. Experimental results have been received by using feed forward artificial neural network [11]. The focus of the chapter is on breast cancer detection by using millimeter (mm) waves modulated Gaussian pulse radar system. The measured radar signals are feed to ANN for further calculations and computations [12]. As lungs are covered or placed under the rib, it is difficult to diagnosis correctly with rib bones. Research work focused on suppressing or reducing the contrast of rib and clavicles of chest area to make the lung portion more visible and help in accurate diagnosis. Image processing system had been developed with multi-resolution massive training artificial neural network [2]. The developed system is going to help computer-aided diagnosis system for better results. Programming has been done to differentiate benign and malignant tissues [5]. Receiver operating characteristics analysis is used for finding out the probability of benign and malignant. Linear and

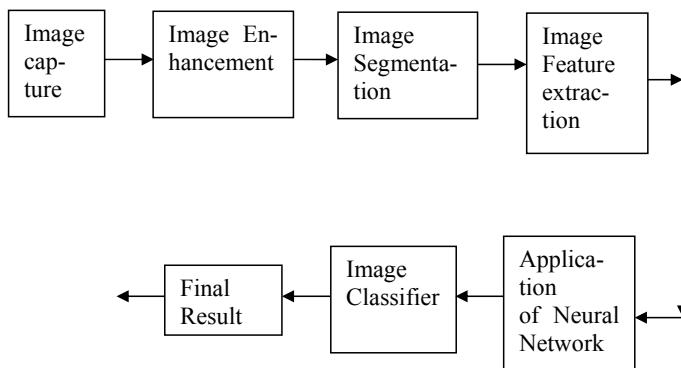
nonlinear classifiers are used for analysis [4]. The aim of the research is to carry out the experiment to improve the image quality. In various medical images, improvement has been done by using the adaptive fractional order derivative. Improved image quality will help in accurate diagnosis [13]. A system is developed for self diagnosis in preliminary level. As per the claim by author, users have to give the details about the symptoms and based on the database available the diagnosis will be done [14]. Author had considered various images collected from MRI, CT and PET for analysis purpose. Image segmentation based on deep convolution neural network has been used. Different fusion schemes like feature learning level, fusing at classifier level and fusing at decision-making level have been performed [15].

### 3 Proposed Method

There are various techniques that have been used for classification of cancer cells. Cancer affected regions have been captured by X-ray images, endoscopy, microscopically zoomed biopsy, etc. In all these methods, automation in diagnosis is lacking for better results. To overcome the drawback of existing system, a new enhanced and effective methodology is implemented.

System block diagram has been shown in Fig. 1. Through this chapter, we propose classification algorithms using ANN to build a prototype for the automatic classification of tumor being malignant or benign.

For this research chapter, we need to have authenticated unaltered and unprocessed [13] images for algorithms to be applied on the images for the enhancement of the precise classification. The experimental database has been collected from private hospitals, Japanese Society of Radiological Technology (JSRT) which is the only open-browsing database for medical image processing research work.



**Fig. 1** System block diagram

The image database used in experiment has 24 images with minor growth of cells, 24 images with major growth and 24 images with tuberculosis, and such total 72 images have been used. Each of the images is of  $512 \times 512$  pixels in size. The images are obtained by browsing the public database of JSRT [11]. For preprocessing of data, MATLAB software has been used. The scanned image is saved with the size of  $512 \times 512$  pixels. After scanning, the image quality gets affected by artifacts like non-uniform intensity, speeds and shift. Thus, the preprocessing removes the noise present in scanned images by keeping essential details of the image [9]. Hence, image filtering helps in preprocessing. Image filtering can be done by median filtering given as  $f(x, y) = \text{median}\{g(s, t)\}$ , where  $(x, y)$  are the target pixel coordinates to be replaced by the value of median pixel value at  $(s, t)$ . The first step is segmentation which helps in separating background and tumor cells [10]. With the help of segmentation, it is possible to remove the bony structure which makes analysis easier. In prescribed method by defining the peripheral coordinates, we can define the affected area. The image is made binary by indicating field area by logical value 1 and remaining area by logical value 0. Masks can be used to find out edges in the images by checking for discontinuity in the pixel values. Threshold value is considered as a valley point of two peaks on histogram. This valley will give an approximate value of the threshold to be set for segmenting it as a nodule from the lung region. It has been observed that unwanted gray pixels also get segmented in this method. Morphological erosion and dilation can be done to remove the artifacts. Morphological operations help in maintaining the details necessary for further processing. In image processing, feature extraction stage is an important stage that uses algorithms and techniques by which it is possible to sight and isolate numerous desired parts or shapes (features) of a given image [12]. Different parameters like affected area and size have been obtained from image. Pixels with value 1 denote segmented tumor. Algorithm is implemented and analyzed by 2 by 2 pixel patterns. Here, the peripheral pixels will decide the perimeter of the tumor in image. In morphology of tumors, generally the shape of the tumor is circular. Change in shape has been identified by index measured by equation,

$$I = 4\pi A/p^2$$

where  $P$  is the perimeter of the tumor, and  $A$  is area of the tumor in pixels. Based on contrast and texture, cell classification is possible.

For the classification of malignant or benign, the nodes of the neural network are average gray level, standard deviation, smoothness, third moment, uniformity, entropy, contrast and energy. In the training process of malignant and benign images, the classifier sets threshold for these nodes and upon testing the process classifies it based on the values obtained. It classifies it to the most strongly matched area, i.e., malignant or benign. For training of malignant and benign images, the classifier sets threshold for these nodes. Classification has been done based on the values obtained by calculation using given formulas in equations [1, 3, 6–8]. Based on strongly matched parameters, classification is done.  $P(Z_i)$  is an estimate of the probability of occurrence of gray level  $Z_i$ . The  $L$ th moment of  $Z$  about its mean is defined as follows.

Mean has been calculated by using expression 1, shown as follows:

$$M = \sum_{i=0}^{L-1} (Z_i - m)^n P(Z_i) \quad (1)$$

Standard deviation can be calculated by using following expression:

$$R = 1 - \left( \frac{1}{1 + \sigma^2(Z)} \right) \quad (2)$$

For finding third moment  $n = 3$  has been considered, as follows in eq. 3:

$$\text{Third Movement} = \sum_{i=0}^{L-1} (Z_i - m)^3 P(Z_i) \quad (3)$$

Uniformity is calculated by using Eq. 4

$$\text{Uniformity, } U = \sum_{i=0}^{L-1} (P)^2(Z_i) \quad (4)$$

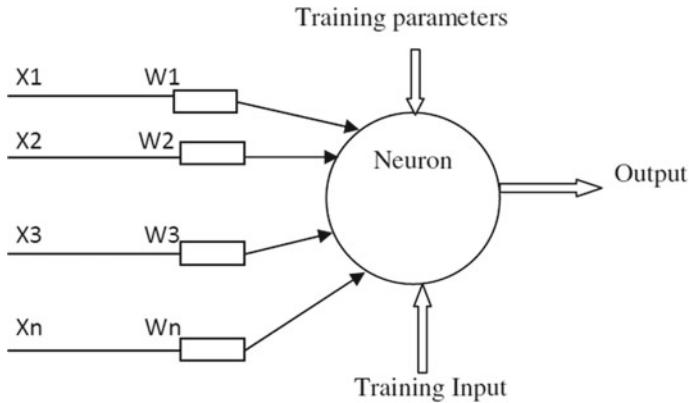
Entropy is calculated by using expression 5

$$\text{Entropy, } e = - \sum_{i=0}^{L-1} P(Z_i) \log_2 P(Z_i) \quad (5)$$

An artificial neural network (ANN) is one of the data processing techniques inspired by the biological nervous systems. It works in a same way as that of human brain's information processing mechanism. It connects a number of neuron or nodes which work in unison to solve the problem. It is application configured for pattern recognition and classification through a learning process. It has two modes, namely learning and using mode. As the below diagram depicts,  $X_i$ s are the node inputs, and  $W_i$ s are the weights which is summed at an activation element to work as a classifier as shown in Fig. 2. The network has been trained by the following steps like:

- Feed forward propagation network has been created.
- Train neural network with the training samples and the group defined for it.
- Train the neural network in such way that it will be possible to identify the particular selected input sample is having any issue or not.
- From the outcomes of network and the samples trained in network, classification rate is calculated.

The chapter deals on the application of automation in the classification of lung cancer. On further applications, we can include deep learning or machine learning



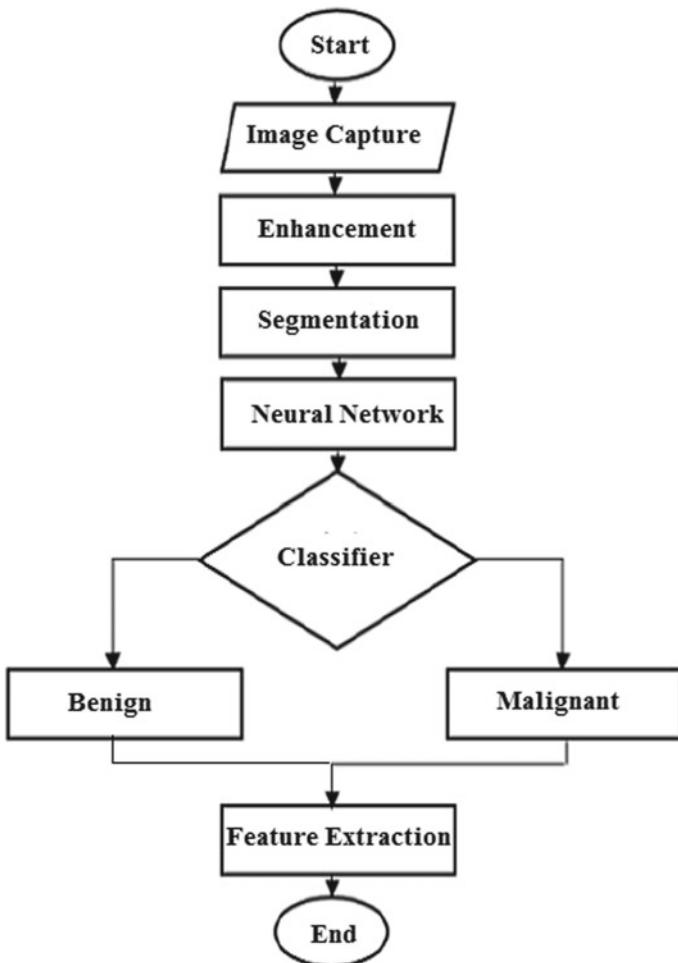
**Fig. 2** Artificial neural network training

which will not only classify images in malignant or benign but also detect and diagnose malignancy. With the culmination of deep learning and data mining techniques, a fully automated, high precision system can be developed for detection, classification and even diagnoses of cancer. Figure 3 shows the flowchart of complete system functioning.

## 4 Results and Conclusion

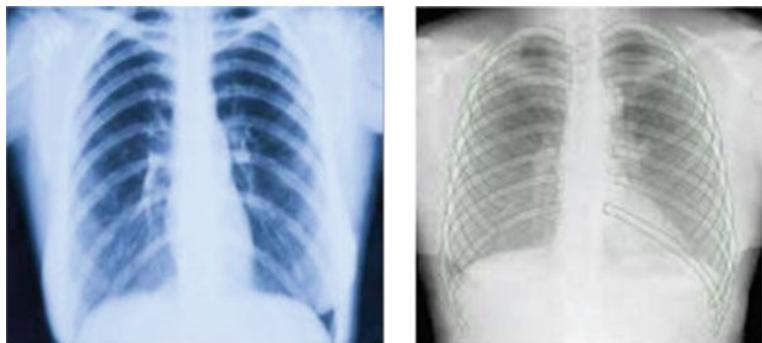
Received results are shown in Figs. 4, 5, 6, 7, 8, 9, 10, 11 and Tables 1 and 2. From received results, it is concluded that we can classify between benign and malignant X-ray images using artificial neural network classifier more accurately. Table 1 gives comparison for benign images, and Table 2 gives the details about malignant images. Comparison has been done by considering various parameters like, avg., gray level, std. deviation, smoothness, third moment, uniformity and entropy. As per the first step of algorithm, the image has been captured by scanning one of the preprocessing step. After preprocessing, there are chances of noise inclusion in the image; hence, image is get filtered out and gets converted to gray form as shown in Fig. 4. Image segmentation is one of the most important steps to separate out the lungs area from rib structure. For better analysis and diagnosis, the shoulder bones, rib, etc., must get filtered out. Segmentation is helping in separation of interested portion from complete image as shown in Fig. 5. This segmentation has been carried out to find benign and malignant as shown in Figs. 6 and 7. Experiment purpose benign image has been taken as shown in Fig. 8. Segmentation process has been done on the image, and results are shown in Fig. 9. Figures 10 and 11 give the details about GUI developed for user interfacing.

We can add multiple nodes or neurons to this network for further precision in the analysis and classification. The merging of data mining and deep learning together



**Fig. 3** Flowchart for functioning of a system

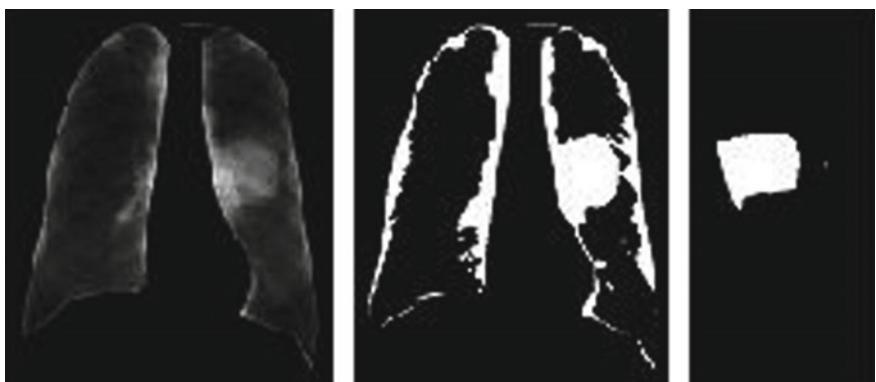
will make this prototype a full-fledged automated cancer classifier and detector. The key feature of algorithm implemented is information processing using artificial intelligence. In artificial neural network, large number of interconnected elements called as neurons are present, which can get trained to solve specific problems.



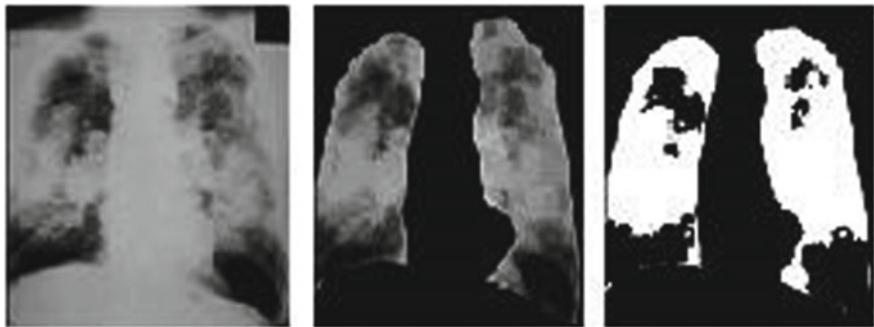
**Fig. 4** Original gray images



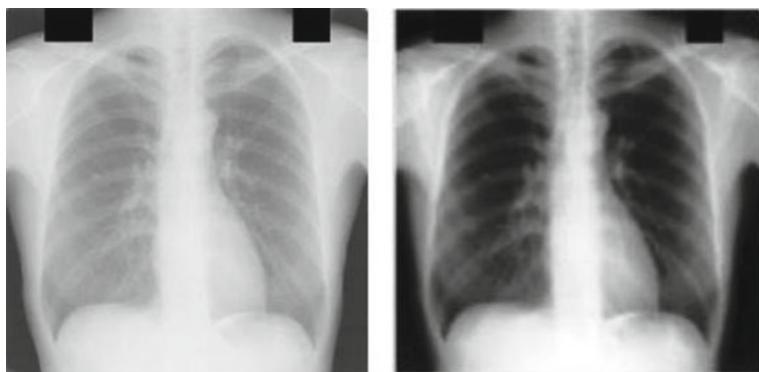
**Fig. 5** Image segmentation



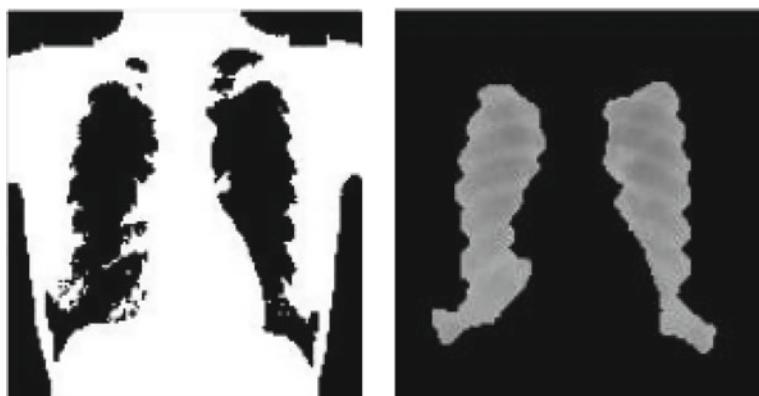
**Fig. 6** Nodule segmentation from the redundant lung region (benign)



**Fig. 7** Nodule segmentation from the redundant lung region (malignant)



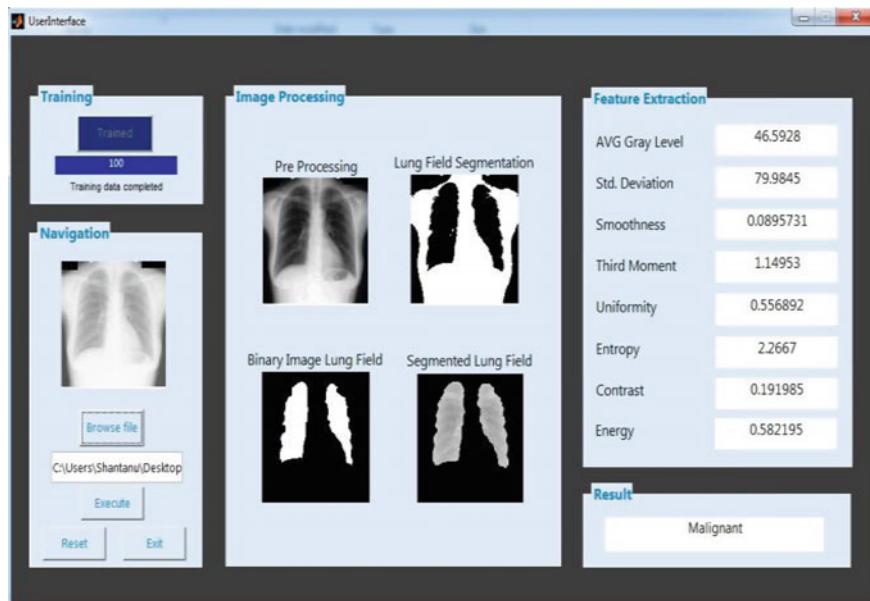
**Fig. 8** Benign image processing and results of preprocessing



**Fig. 9** Benign X-ray and processing by segmentation



**Fig. 10** Result of user interface showing—benign figure



**Fig. 11** Result of user interface showing for malignant

**Table 1** Comparison for benign

Sr. No.	Statistic features	Value
1	Avg. gray level	32.8894
2	Std. deviation	64.4891
3	Smoothness	0.0601132
4	Third moment	1.46329
5	Uniformity	0.629625
6	Entropy	1.83653

**Table 2** Comparison for malignant

Sr. No.	Statistic features	Value
1	Avg. gray level	46.5928
2	Std. deviation	79.9845
3	Smoothness	0.0895731
4	Third moment	1.14953
5	Uniformity	0.556892
6	Entropy	2.2667

## References

- Casti P, Mencattini A, Salmeri M, Rangayyan RM, Enderle JD (2017) Computerized analysis of mammographic images for detection and characterization of breast cancer. In: Computerized analysis of mammographic images for detection and characterization of breast cancer, vol 1. Morgan & Claypool, p 186
- Suzuki K, Abe H, MacMahon H, Doi K (2006) Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). IEEE Trans Med Imaging 25(4):406–416
- Arasi MA, El-Horbaty ESM, Salem AM, El-Dahshan ESA (2017) Computational intelligence approaches for malignant melanoma detection and diagnosis. In: 2017 8th international conference on information technology (ICIT), Amman, Jordan, pp 55–61
- Fogel DB, Wasson EC, Boughton EM, Porto VW, Angeline PJ (1998) Linear and neural models for classifying breast masses. IEEE Trans Med Imaging 17(3):485–488
- Joo S, Moon WK, Kim HC (2004) Computer-aided diagnosis of solid breast nodules on ultrasound with digital image processing and artificial neural network. In: The 26th annual international conference of the IEEE engineering in medicine and biology society. San Francisco, CA, pp 1397–1400
- Brahim W, Mestiri M, Betrouni N, Hamrouni K (2017) Malignant pleural mesothelioma segmentation from thoracic CT scans. In: 2017 international conference on advanced technologies for signal and image processing (ATSIP). Fez, Morocco, pp 1–5
- Rabidas R, Midya A, Chakraborty J, Arif W (2017) Texture analysis of gradient images for benign-malignant mass classification. In: 2017 4th international conference on signal processing and integrated networks (SPIN). Noida, pp 201–205
- Mirbeik-Sabzevari A, Ashinoff R, Tavassolian N (2017) Ultra-wideband millimeter-wave dielectric characteristics of freshly-excised normal and malignant human skin tissues. IEEE Trans Biomed Eng PP(99):1–1
- Patrício M, Oliveira C, Caseiro-Alves F (2017) Differentiating malignant thyroid nodule with statistical classifiers based on demographic and ultrasound features. In: 2017 IEEE 5th Portuguese meeting on bioengineering (ENBENG). Coimbra, pp 1–4

10. Johra FT, Shuvo MMH (2016) Detection of breast cancer from histopathology image and classifying benign and malignant state using fuzzy logic. In: 2016 3rd international conference on electrical engineering and information communication technology (ICEEICT). Dhaka, pp 1–5
11. Economou GK et al (1994) Medical diagnosis and artificial neural networks: a medical expert system applied to pulmonary diseases. In: proceedings of IEEE workshop on neural networks for signal processing. Ermioni, Greece, pp 482–489
12. Lenzi C, Pasian M, Bozzi M, Perregini L, Caorsi S (2017) MM-waves modulated Gaussian pulse radar breast cancer imaging approach based on artificial neural network: preliminary assessment study. In: 2017 Mediterranean microwave symposium (MMS). Marseille, pp 1–4
13. Krouma H, Ferdi Y, Taleb-Ahmedx A (2018) Neural Adaptive fractional order differential based algorithm for medical image enhancement. In: 2018 international conference on signal, image, vision and their applications (SIVA). Guelma, Algeria, pp 1–6
14. Aljurayfani M, Alghernas S, Shargabi A (2019) Medical self-diagnostic system using artificial neural networks. In: 2019 international conference on computer and information sciences (ICCIS). Sakaka, Saudi Arabia, pp 1–5
15. Guo Z, Li X, Huang H, Guo N, Li Q (2018) Medical image segmentation based on multi-modal convolutional neural network: study on image fusion schemes. In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), Washington, DC, pp 903–907

# Chapter 2

## Deep Learning Techniques for Crime Hotspot Detection



Sankar N. Nair and E. S. Gopi

### 1 Introduction

The prevention of crimes is more profitable to a society than to solve a crime after its occurrence. It is essential for police forces across the world to have prior knowledge about the probable locations of future crimes, for more efficient utilization of police resources. Hotspot analysis is a major part of crime mapping studies. A crime hotspot is defined as an area that has a greater than the average number of criminal or disorder events, or an area where people have a higher than average risk of victimization. Accuracy and time complexity are the two major constraints associated with this problem, since real-time results are necessary for a quick response to changing conditions. Also, accuracy is hard to achieve since the problem depends on many dynamic parameters, like gang behavior. Statistical approaches to the problem are more time consuming and are not able to provide the real-time results that are needed. A deep learning-based approach would provide much faster results for a better response from police force. Also, accuracy would be improved due to the ability of the deep neural network to find complex relations that are hidden in the raw data.

Deep learning techniques have been successfully used in similar applications due to its ability to find complex relationships between inputs of large dimensions and corresponding outputs. Deep learning techniques have been proven to work well with huge datasets with large number of parameters. The input datasets are made to be correlated with introducing some overlap in time interval between consecutive datasets, and this correlation would help the deep neural network to identify the

---

S. N. Nair · E. S. Gopi (✉)

Department of Electronics and Communication Engineering,  
National Institute of Technology Trichy, Tiruchirapalli, Tamil Nadu 620015, India  
e-mail: [esgopi@nitt.edu](mailto:esgopi@nitt.edu)

S. N. Nair  
e-mail: [sankarnnair@gmail.com](mailto:sankarnnair@gmail.com)

features of the dataset much better. The raw dataset was converted into heat maps so that two-dimensional convolutional filters could be applied to find local and more global features of the dataset.

## 1.1 Literature Survey

Crime analysis involves exploiting data about crimes to enable law enforcement to better apprehend criminals and prevent crimes. Data used by crime analysts includes the time and locations of crimes and a variety of characteristics, such as methods of entry and items stolen, that vary with the type of crime. Crime analysts use these data with methodologies like aggregate crime rate analysis, hotspots, and space-time point process modeling to analyze and predict the spatial patterns of crimes. Ring-shaped hotspot detection is important for a variety of application domains where finding a ring-shaped hotspot may help focus domain users' efforts to a specific region. For example, finding a ring-shaped hotspot may focus public security officials' efforts to the inner circle of a ring when searching for a possible crime source [1, 2]. Various criminal profiling and criminal behavioral studies [3, 4] have been used as the basis for modeling the probability of a site being selected by criminals as their preferred locations for a crime.

Spatial scan statistics are used to determine hotspots in spatial data and are widely used in epidemiology and bio-surveillance. In [5], experiments regarding the computational study of spatial scan statistics were performed. An algorithm was proposed to find the largest discrepancy region in a domain. Approximation algorithms were developed using these discrepancy functions, which could be used for spatial scan analysis of crime locations.

Aggregate crime rate analysis uses sample units, such as neighborhoods, cities, and schools, to explain the variation in crime rates across those units. The statistical basis of this analytic approach is well established, and it is modeled as a regression problem. Many regression methods, especially Poisson regressions, are well studied and broadly used [6–8].

Hotspot models use past crime data to identify unusual clusters of criminal incidents within a well-defined region. These clusters are commonly referred to as hotspots representing areas that contain unusual amounts of crimes. The term 'hotspots' has become part of the lexicon of crime analysts. In [9], the computational study of spatial scan statistics is extensively studied. First, an exact algorithm for finding the largest discrepancy region is described. Then, a new approximation algorithm is proposed for a large class of discrepancy functions to improve approximation. A survey of the existing techniques for the identification of geographic hotspots for crimes and other application has been studied in some papers [10, 11]. In [12], spatiotemporal hotspots are selected through density estimation techniques using both kernel methods and mixture models.

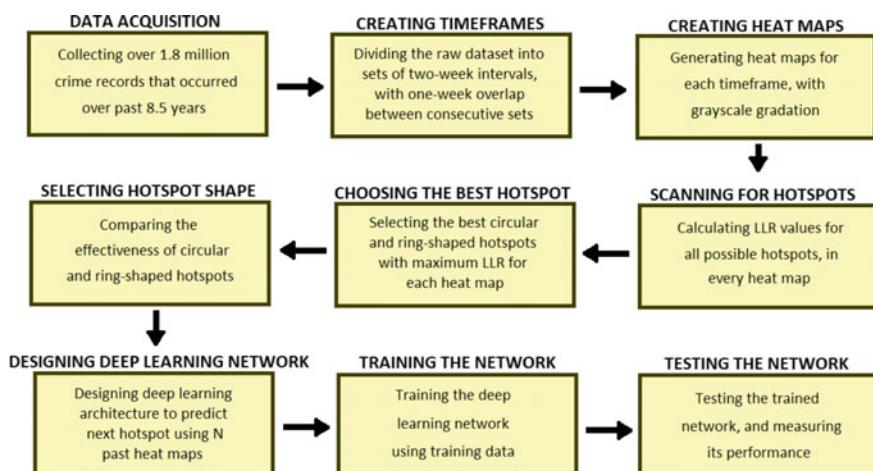
Various studies have also been conducted on the effects of sociological and environmental parameters on the crime rates. Cohen and Felson [13] postulate, based

on human ecological theory, that higher crime rates happen when the dispersion of activities away from households is more. In [14], large data samples of crimes have been collected, and geographic location-based profiling has been done for various types of crimes. In [15, 16], the authors investigate various hidden aspects like the physical and social characteristics of crime sites and people's perceptions of crime locations and the policies that create or maintain these locations.

In [17], the author defines various types of hotspots and elaborates on the accepted theories about the probable root causes for the occurrence of hotspots. Chainey and Dando [18] discuss the different statistical tests and other techniques to effectively identify the crime hotspots. Four different datasets were used to compare the different proposed statistical methods. Some spatial analysis tools to closely study the spatial patterns and locational contexts of crime are examined in [19]. These tools are being used by police forces in various parts of the world with different levels of success.

## 2 Proposed Methodology

The block diagram in Fig. 1 describes the various stages by which the experiments proposed in this paper have been implemented. The raw data consists of date when each crime has occurred and the corresponding geographical location of each crime, as defined by the latitude and longitude where the crime has occurred. The preprocessing of the dataset is done by dividing the crimes into many sets based on the date of occurrence of the crimes. Each set includes crimes that have occurred in a two-week period. Also, two sets that have consecutive two-week intervals have an overlapping period of one week.



**Fig. 1** Block diagram describing various steps involved in the proposed idea

The next step is the generation of heat maps for each of these two-week intervals. The intensity gradation of each pixel is a grayscale value proportional to the number of crimes that happened in the area of geographic location corresponding to the pixel. The best possible circular and ring-shaped hotspots have been identified for each heat map, using the metric log likelihood ratio (LLR). Once it was identified that the circular hotspot had better performance, it was chosen to train a deep neural network which takes the heat maps of past timeframes as inputs and gives the best circular hotspot of the immediate next timeframe as the output.

## **2.1 Heat Maps**

A heat map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per capita income. Heat maps provide an easy way to visualize how a measurement varies across a geographic area or show the level of variability within a region.

Heat maps are made use of to represent the crimes in a 2-D map projection based on the location of crime. The data is pre-processed before converting into heat maps. The dataset contains all crimes that occurred between January 1, 2010, and July 31, 2018, a period of more than  $8\frac{1}{2}$  years. The parameters for each crime include date and time of occurrence, latitude and longitude of the location. The preprocessing involves splitting of the dataset into various sets based on the date of occurrence of the crimes. Each set consists of crimes happened within a time interval of 14 days, and consecutive sets have an overlap of 7 days. For example, the first set contains crimes between January 1, 2010, and January 14, 2010. The second set contains crimes between January 8, 2010, and January 21, 2010, and so on.

The total area in which the crimes in a set occurred is called an activity set. The area is divided into a grid of size  $P \times P$ . Now, each crime is mapped onto a pixel on the  $P \times P$  activity area, corresponding to its actual location on the map. One pixel could have more than one crime happening there. So, the grayscale activity area is graded according to the number of crimes happening there. Lighter shades of gray indicate lesser number of crimes, and darker shades of gray indicate more crimes. White-colored pixels indicate locations without any crimes, and black pixels are locations where maximum crimes have occurred.

## **2.2 Hotspots**

Areas of concentrated crime are often referred to as hotspots. Hotspots can be of various shapes—like circular, ring-shaped, and irregular-shaped. Hotspots are to be computed for heat maps generated for every two-week interval. The shapes of interest in this experiment are ring-shaped hotspots and circular hotspots. The best hotspots of both shapes are to be computed and compared using statistical tests. A ring-shaped

hotspot  $R$  is uniquely defined by its inner radius (denoted by  $r_0$ ), outer radius ( $r_1$ ), and coordinates of the center of the concentric circles ( $x_r$ ,  $y_r$ ). Similarly, a circular hotspot  $C$  is uniquely defined by its radius (denoted by  $r_c$ ) and its center coordinates ( $x_c$ ,  $y_c$ ). The statistical metric used to find the best hotspot for each heat map is log likelihood ratio (LLR).

For an activity area of size  $P \times P$ , all possible circular hotspots are considered where the radius  $r_c$  varies between 5 and  $\frac{P}{2}$  pixels. The center coordinates are also varied between 0 and  $P$ . Similarly, for ring-shaped hotspots, inner radius  $r_0$  varies from 5 to  $\frac{P}{2}$ , and outer radius  $r_1$  varies from  $r_0 + 5$  to  $\frac{P}{2}$ . The best hotspot for each shape is selected from this set of hotspots using the metric LLR.

The likelihood ratio expresses how many times more likely the data is under one model than the other. This likelihood ratio, or equivalently its logarithm, can then be compared to a critical value to decide whether or not to reject the null model. When the logarithm of the likelihood ratio is used, the statistic is known as a log likelihood ratio statistic.

The equation for computing log likelihood ratio is defined as

$$\text{LLR} = \log \left( \left( \frac{C}{B} \right)^C \times \left( \frac{A - C}{A - B} \right)^{A-C} \right) \quad (1)$$

where

$$B = \frac{A \times \text{area}(R)}{\text{area}(S)}$$

for a null hypothesis that the crime points(activities) are distributed uniformly across the activity area  $S$ . In the above equation,  $B$  denotes the expected number of activities within a given hotspot  $R$ ,  $C$  denotes the observed number of activities within the hotspot, and  $A$  denotes the total number of activity points present in the whole activity area  $S$  (refer Appendix for detailed derivation).

The likelihood ratio is the product of two terms; the first term  $\frac{C^C}{B^C}$  denotes the likelihood ratio of the crime activities inside the hotspot, and the second term denotes the likelihood ratio of the crime activities outside the hotspot. Thus, the product gives the likelihood ratio of actual distribution as against the distribution of the null hypothesis. The null hypothesis assumes that the distribution of crime points across the activity area follows uniform distribution, since no prior information about any clustering is available.

Higher values of log likelihood ratio for a given hotspot indicate that the distribution of crimes within the hotspot is higher when compared to the expected number of crimes according to the null hypothesis of uniform distribution. If the distribution exactly matches the null hypothesis, the LLR value computed would be 0. A negative LLR for a given hotspot indicates that the given hotspot has less number of crimes than the expected number of crimes. A hotspot is considered to be valid only if the given hotspot has a positive LLR value. The best hotspot is considered to be the hotspot for which the LLR is maximum for the given heat map. In this manner, the best hotspots are computed for all the heat maps. For each heat map, both ring-shaped and circular hotspots are computed.

## 2.3 Deep Learning Architecture

A deep neural network is a type of artificial neural network with a large number of hidden layers. The advantage of a deep neural network is that it can find very complex relations between input and output. The lower-level layers of a DNN identify low-level features like curves and edges that are more local in nature. The higher-level layers can identify increasingly global features which are more complex. Such a technique is particularly effective in cases where the dataset is huge and very complex.

From the results obtained, it was concluded that circular hotspots had a better performance compared to ring-shaped hotspots. So, circular hotspots were used in the training phase of the deep neural network.

The input layer to the DNN consists of a matrix that would contain the grayscale heat maps of an arbitrary number of consecutive intervals. The output layer of the DNN is to be a  $3 \times 1$  vector which contains the parameters (radius and center coordinates) that represent the best hotspot for the immediately next future interval.

### 2.3.1 Layers of DNN

- **2-D Convolutional Layer:** A 2-D convolutional layer applies sliding convolutional filters to the input. The layer convolves the input by moving the filters along the input vertically and horizontally and computing the dot product of the weights and the input. This result is called a feature map. The output of the convolution will be passed through the activation function. For example, a  $5 \times 5 \times 32$  convolutional layer indicates that the layer has 32 filters of size  $5 \times 5$  each. Also, the inputs are zero-padded on all sides to ensure that the size of input and output matrices of the convolutional layers is same, otherwise the output matrices would be slightly smaller than the input matrix depending on the size of the filter. Besides restricting outputs to a certain range, activation functions break the linearity of a neural network, allowing it to learn more complex functions than linear regression. The activation function used is the Rectified Linear Unit Activation Function (ReLU), defined as follows:

$$f(x) = x^+ = \max(0, x)$$

- **Maxpool Layer:** Max pooling is a sample-based discretization process. The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned. This is done to help overfitting by providing an abstracted form of the representation. As well, it reduces the computational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation. Max pooling is done by applying a max filter to (usually) non-overlapping sub-regions of the initial representation.

For example, assume a  $100 \times 100$  matrix representing the initial input and a  $2 \times 2$  filter that runs over the input. A stride of 2 means the  $(dx, dy)$  for stepping over the input will be  $(2, 2)$ , and will not overlap regions. Then, the resulting output will be a  $50 \times 50$  matrix. For each of the regions represented by the filter, we will take the maximum value of that region and create a new, output matrix where each element is the max of a region in the original input.

- **Dropout Layer:** Dropout refers to ignoring units (i.e., neurons) during the training phase of certain set of neurons which is chosen at random. These units are not considered during a particular forward or backward pass. At each training stage, individual nodes are either dropped out of the net with probability  $1-p$  or kept with probability  $p$ , so that a reduced network is left; incoming and outgoing edges to a dropped out node are also removed. Dropout layers are used to prevent overfitting in the neural network. A fully connected layer occupies most of the parameters, and hence, neurons develop codependency among each other during training which curbs the individual power of each neuron leading to overfitting of training data. Due to overfitting, the neural network will give a good performance for training data, but performs poorly for other data inputs, including testing data. In this experiment, the dropout layers are employed with a dropout probability of 20%. Also, the deep neural network employed in this project consists of repetition of 2-D convolutional layers, maxpool layers, and dropout layers.
- **Flattening Layer:** The flattening step is needed to make use of fully connected layers after some convolutional layers. Fully connected layers do not have a local limitation like convolutional layers (which only observe some local part of an image by using convolutional filters). This means we can combine all the found local features of the previous convolutional layers. Each feature map channel in the output of a CNN layer is a flattened 2-D array created by adding the results of multiple 2-D kernels (one for each channel in the input layer). For example, a flattening layer converts a  $25 \times 25 \times 8$  three-dimensional layer into a  $5000 \times 1$  one-dimensional layer. A flattening layer is usually used before a fully connected layer.
- **Fully Connected Layer:** A fully connected layer is placed just before the output layer. All neurons from the previous layer are connected to all neurons in the output layer. The activation function used is the rectified linear unit function. The output layer consists of three neurons, for the parameters  $(r_c, x_c$  and  $y_c)$  representing the predicted circular hotspot which is the expected output of the deep neural network.

### 2.3.2 Performance Metrics

The deep neural network modeled in this project tackles a regression problem where the target output is a  $3 \times 1$  vector. The performance of the deep neural network is analyzed by the predicted output for testing data as compared to the corresponding target outputs. The performance metrics used to evaluate the neural network in this project are defined below.

- **Mean Square Error (MSE):** The mean squared error (MSE) squares the difference of all corresponding elements of target vector and predicted vector before summing them all. The equation below defines the mean squared error.

$$\text{MSE} = \frac{1}{n} \sum (\mathbf{y} - \hat{\mathbf{y}})^2 \quad (2)$$

where  $\mathbf{y}$  is the target (actual) output vector,  $\hat{\mathbf{y}}$  is the predicted output vector, and  $n$  is the total number of data points. The effect of the square term in the MSE equation is most apparent with the presence of outliers in the data. Each residual in MSE contributes quadratically to the total mean squared error. This ultimately means that outliers in the data will contribute to much higher total error in the MSE, as compared to mean absolute error. Similarly, the model will be penalized more for making predictions that differ greatly from the corresponding actual value. This is to say that large differences between actual and predicted are punished more in MSE than in MAE.

- **Mean Absolute Error (MAE):** The mean absolute error (MAE) is the simplest regression error metric. The residual for every data point is calculated by taking only the absolute value of each so that negative and positive residuals do not cancel out. Then, the average of all these residuals is calculated. Effectively, MAE describes the typical magnitude of the residuals. The formal equation for mean absolute error is

$$\text{MAE} = \frac{1}{n} \sum |\mathbf{y} - \hat{\mathbf{y}}| \quad (3)$$

The MAE is also the most intuitive of the metrics since we only observe the absolute difference between the data and the model's predictions. Because we use the absolute value of the residual, the MAE does not indicate underperformance or overperformance of the model (whether or not the model under or overshoots actual data). Each residual contributes proportionally to the total amount of error, meaning that larger errors will contribute linearly to the overall error. A small MAE suggests that the model is great at prediction, while a large MAE suggests that the model may have trouble in certain areas. A MAE of 0 means that the model is a perfect predictor of the outputs. While the MAE is easily interpretable, using the absolute value of the residual often is not as desirable as squaring this difference. Depending on how the model should treat outliers, or extreme values, in your data, you may want to bring more attention to these outliers or downplay them. The issue of outliers can play a major role in which error metric you use. MAE requires more complicated tools such as linear programming to compute the gradient. MAE is more robust to outliers since it does not make use of square. On the other hand, MSE is more useful if concerning about large errors whose consequences are much bigger than equivalent smaller ones.

- **Mean Absolute Percentage Error (MAPE):** The mean absolute percentage error (MAPE) is the percentage equivalent of MAE. The equation looks just like that of MAE, but with adjustments to convert everything into percentages. The equation

for mean absolute percentage error is

$$\text{MAPE} = \frac{1}{n} \sum \left| \frac{\mathbf{y} - \hat{\mathbf{y}}}{\mathbf{y}} \right| \times 100\% \quad (4)$$

Just as MAE is the average magnitude of error produced by your model, the MAPE is how far the model's predictions are off from their corresponding outputs on average. Like MAE, MAPE also has a clear interpretation since percentages are easier for people to conceptualize. Both MAPE and MAE are robust to the effects of outliers thanks to the use of absolute value. However, for all of its advantages, MAPE is a weaker measure when compared to MAE. Many of MAPE's weaknesses actually stem from the use of division operation. Now that everything is to be scaled by the actual value, MAPE is undefined for data points where the value is 0. Similarly, the MAPE can grow unexpectedly large if the actual values are exceptionally small themselves. Finally, the MAPE is biased toward predictions that are systematically less than the actual values themselves. That is to say, MAPE will be lower when the prediction is lower than the actual compared to a prediction that is higher by the same amount.

- **Cosine Proximity:** Cosine proximity is same as cosine similarity, which is a measure of similarity between two nonzero vectors of an inner product space that measures the cosine of the angle between them. In this case, note that unit vectors are maximally similar if they are parallel and maximally dissimilar if they are orthogonal (perpendicular). This is analogous to the cosine, which is unity (maximum value) when the segments subtend a zero angle and zero (uncorrelated) when the segments are perpendicular. Cosine proximity loss function computes the cosine proximity between the predicted value and actual value, which is defined as follows:

$$\text{CP} = -\frac{\mathbf{y} \cdot \hat{\mathbf{y}}}{\|\mathbf{y}\| \cdot \|\hat{\mathbf{y}}\|} \quad (5)$$

The bounds between 0 and 1 apply for any number of dimensions, and the cosine similarity is most commonly used in high-dimensional positive spaces. One advantage of cosine similarity is its low-complexity, especially for sparse vectors: Only the nonzero dimensions need to be considered.

- **Percentage Deviation of Log Likelihood Ratio:** Another method used in this project to evaluate the performance of prediction of the deep neural network is to compare the log likelihood ratio (LLR) of the predicted hotspot and the actual target hotspot for all testing data. The target hotspot is the best hotspot computed using the LLR metric and represents the best hotspot for the given input, because that hotspot has the highest value of LLR among all possible locations. So, the LLR value of predicted hotspot would always be lesser than or equal to (best case) the LLR value of actual hotspot.

If the predicted hotspot is perfectly equal to the target hotspot, then both LLR values would be same, and the difference in LLR values would be zero. The difference in

LLR values gives a measure of how efficient the predicted hotspot is, as compared to the best possible hotspot. The error is measured as a percentage of the LLR of target hotspot, for better comparison. A low percentage in deviation of LLR values shows that the predicted hotspot is almost as efficient as the actual hotspot, even if the deviation in radius and center coordinates of hotspot is larger. To find the mean percentage deviation of LLR, the percentage deviation values for every testing data are assumed to be the sample outcomes from a Gaussian distributed random variable. Then, the mean of the above defined Gaussian random variable would give the mean percentage deviation of log likelihood ratio between actual and predicted hotspots.

### 3 Experiments and Results

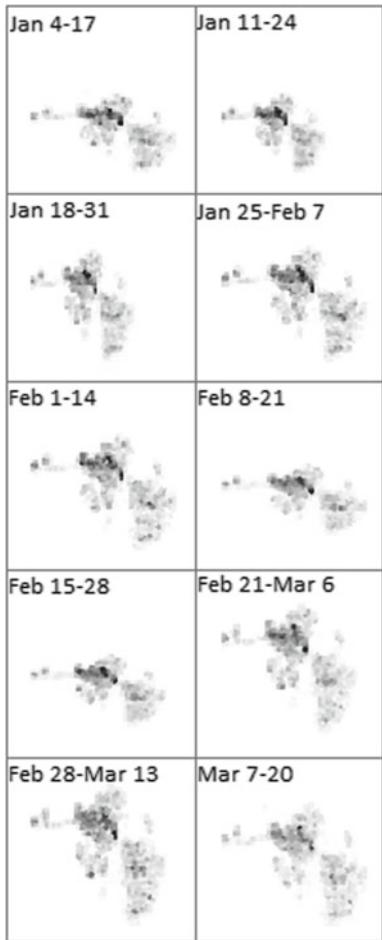
The dataset has been divided into 452 different sets based on the date on which each crime has occurred. The dataset contains crimes that were reported in the city of Los Angeles, California from January 1, 2010, to July 31, 2018. The dataset has been divided into sets where each set contains all crimes that have occurred over a 14-day period and with consecutive sets having an overlap of a 7-day interval. This timeline-based approach is useful in creating a deep learning network which will use past crime data to predict the future hotspots.

Figure 2 shows heat maps from ten consecutive sets over the time period between January 4, 2016, and March 20, 2016. It has been observed that there is an average of 8468 crimes within a two-week period. From the heat maps shown in Fig. 2, we can see that although there is an observable variation between distribution of crimes between different time intervals, there is also a significant correlation between heat maps of consecutive time intervals. This correlation is achieved as a result of the overlap of seven days introduced between consecutive heat maps. This ensures that around half of the crime locations would be same for consecutive heat maps, which result in a high correlation. This correlation is introduced to help the deep neural network to more efficiently find the time-related variances in locations of time, and subsequently to more efficiently predict the future hotspot which will have a strong correlation with the past heat maps.

With the 452 heat maps that have been generated from the dataset, the next step involved is to calculate the best hotspot for each heat map. In this experiment, both ring-shaped hotspots and circular hotspots have been investigated and compared. A ring-shaped hotspot is defined by its inner radius, outer radius, and center coordinates. Similarly, a circular hotspot is uniquely defined by its radius and center coordinates. The metric used to determine the best hotspot is log likelihood ratio. The best hotspot is determined by calculating the log likelihood ratio for all possible hotspots by scanning across the heat map.

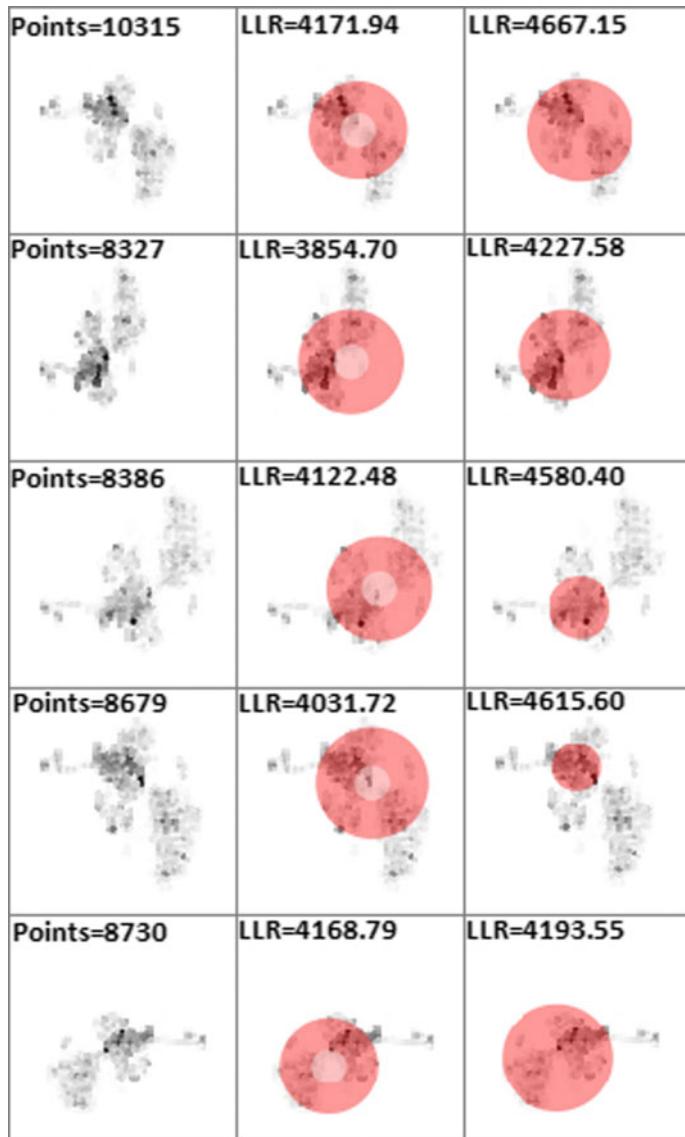
For a ring-shaped hotspot, the heat map is scanned for all possible combinations of inner and outer radii, and center coordinates. The inner radius is varied between 5 and 20, the outer radius is varied between 10 and 25, with the gap between inner

**Fig. 2** Grayscale heat maps generated for ten consecutive time intervals between January 4, 2016, and March 20, 2016



and outer radii varying between 5 and 20. Also, both  $x$  and  $y$  values of the center coordinates are varied between 10 and 40, thus ensuring that all possible hotspot locations are considered. Similarly, for a circular hotspot, the radius is varied for all values between 5 and 25, and the  $x$  and  $y$  coordinates of the center are varied independently between 5 and 45. The log likelihood ratio has been calculated for all these possible hotspots, and a hotspot is considered valid only if the LLR value is positive. The hotspot with the largest LLR value among the valid hotspots has been selected as the best hotspot for the corresponding heat map.

Figure 3 illustrates the best ring-shaped and circular hotspots for various heat maps, as computed using the log likelihood ratio. The figure also denotes the LLR value for the best hotspots. It is observed that the circular hotspot generally has higher LLR values as compared to ring-shaped hotspots. Table 1 shows a comparison of performance between the circular and ring-shaped hotspots. It is concluded that



**Fig. 3** Illustration for various heat maps and the corresponding best ring-shaped hotspots and circular hotspots

**Table 1** Hidden layers of the deep neural network used in case the number of past heat maps used to predict the future hotspot is fixed as 8, then the input matrix would have a size of  $100 \times 200$  (by concatenating 8 images of size  $50 \times 50$  each), along with the related parameters that define each layer, like the activation function

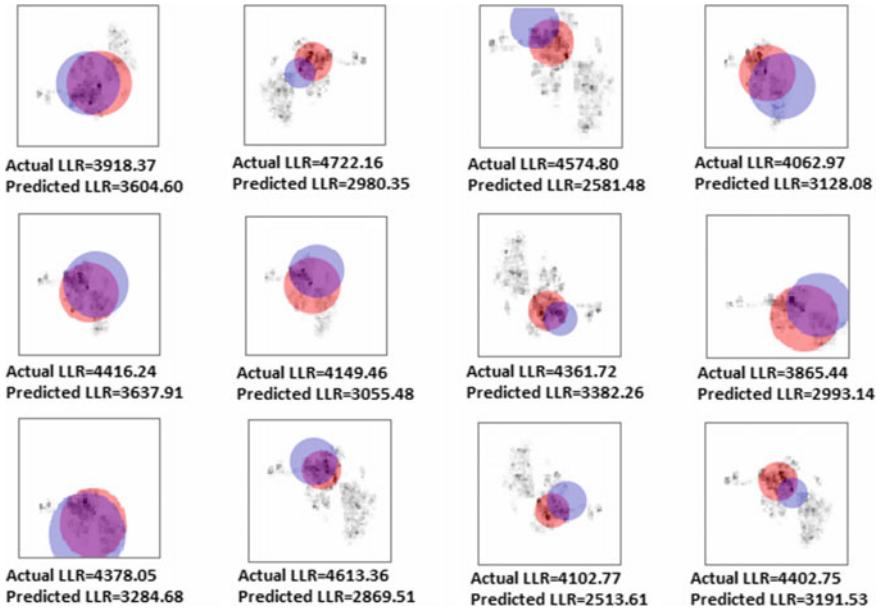
Layer No.	Layer name	Layer input size	Layer output size
1	2-D convolution ( $5 \times 5 \times 32$ )	$100 \times 200$	$100 \times 200 \times 32$
2	Maxpool ( $1 \times 2$ )	$100 \times 200 \times 32$	$100 \times 100 \times 32$
3	Dropout (20%)	$100 \times 100 \times 32$	$100 \times 100 \times 32$
4	2-D convolution ( $3 \times 3 \times 16$ )	$100 \times 100 \times 32$	$100 \times 100 \times 16$
5	Maxpool ( $2 \times 2$ )	$100 \times 100 \times 16$	$50 \times 50 \times 16$
6	Dropout (20%)	$50 \times 50 \times 16$	$50 \times 50 \times 16$
7	2-D convolution ( $3 \times 3 \times 8$ )	$50 \times 50 \times 16$	$50 \times 50 \times 8$
8	Maxpool ( $2 \times 2$ )	$50 \times 50 \times 8$	$25 \times 25 \times 8$
9	Dropout (20%)	$25 \times 25 \times 8$	$25 \times 25 \times 8$
10	Flatten	$25 \times 25 \times 8$	$5000 \times 1$
11	Fully connected ReLU	$5000 \times 1$	$3 \times 1$

the circular hotspots have a better performance than ring-shaped hotspots, based on the higher average LLR value, and also take significantly lesser training time as compared to ring-shaped hotspots. This helps in providing better real-time results as computational time is also a constraint in this application. So, only circular hotspots are considered for the subsequent phases of experiments.

The heat maps serve as input for the deep learning network implemented in the next phase of the experiment. The network is designed to take an arbitrary number of consecutive heat maps as input and give the hotspot parameters of the next future heat map as the output. So, the heat maps and corresponding circular hotspots are split into training data and testing data. Out of 452 images, 402 are used for training, and 50 are used for testing.

The deep neural network is trained using the 402 heat map images. Four different networks are trained, with varying number of heat maps used to predict the next hotspot. The number of heat maps used to predict the future hotspot is varied as 4, 6, 8, and 12, and the performance of the four networks is compared. The deep neural network has three neurons in the output layer corresponding to the radius and center coordinates of the output circular hotspot. So, the target outputs are similarly arranged as a  $3 \times 1$  vector.

Figure 4 shows some visual examples of the predictions made by the deep neural networks and a comparison of the performance based on log likelihood ratio values. Table 2 gives a comparison between the four different neural networks that were trained using different sample sizes of past data for every iteration of training. The performance of a neural network for regression is compared using various regression metrics like mean squared error, mean absolute error, mean absolute percentage error, cosine proximity, and mean percentage deviation of log likelihood ratio (Table 3).



**Fig. 4** Illustration for various heat maps with actual and predicted hotspots

**Table 2** Comparing performance of different hotspot shapes

	Min. LLR	Avg. LLR	Max. LLR	Min. time (in s)	Avg. time (in s)	Max. time (in s)
Ring-shaped	3634.50	4153.92	4808.09	225.19	276.76	323.41
Circular	4166.27	4978.64	5713.41	42.54	52.18	68.03

**Table 3** Comparing performance of different deep neural networks, with varying number of previous heat maps as inputs, denoted by  $N = 4, 6, 8, 12$

	MSE	MAE	MAPE (%)	Cosine proximity (%)	LLR % deviation (%)	Training time (in hours)
$N = 4$	13.092	4.412	56.6	36.9	42.51	1.85
$N = 6$	7.850	3.669	43.09	43.65	36.72	3.20
$N = 8$	7.197	3.200	41.18	46.5	35.48	4.36
$N = 12$	7.334	3.382	41.78	44.93	36.33	6.82

## 4 Conclusion and Future Scope

We have analyzed the performance of two different shapes of hotspots—circular and ring-shaped hotspots. Log likelihood ratio and hypothesis test for mean have been employed in the comparison, and it has been conclusively shown that circular hotspots have better performance than ring-shaped hotspots. Also, a completely new approach

based on deep learning has been proposed to predict the best circular hotspot for a future timeframe using the heat map distributions of crimes in past timeframes. The results show that deep learning approach can provide significantly good prediction results.

This paper provides a clear direction for future researchers in the domain of crime hotspot analysis and related topics. Metrics other than LLR can also be applied to compare different hotspots. Also, various other architectures for deep neural network can be tried to improve on the performance of the proposed network. Also, the heat map-based methodology can be adapted to related applications where activities can be represented as points on a geographic map. Some areas where this approach can be implemented are epidemiology, and natural disasters like forest fires and cyclones.

## Appendix

Let  $S$  be the activity area where the crime points are distributed, and let  $R$  be the subset of  $S$  which indicates the candidate hotspot area. Let  $A$  denote the total number of activities in the activity area. Let  $C$  denote the actual observed number of crime points within the hotspot. Assuming the null hypothesis  $H_0$  that the crime points are uniformly distributed within the activity area  $S$ , the expected number of activities within the hotspot  $R$  denoted by  $B$  can be defined as

$$B = \frac{A \times \text{Area}(R)}{\text{Area}(S)}$$

The probability that a given crime point lies within the hotspot if the null hypothesis were true, is given by  $\frac{B}{A}$ , and for the point to be outside the hotspot is  $\frac{A-B}{A}$ . Assuming all the points are distributed independently, and that null hypothesis is true, the total probability that exactly  $C$  points are present within the hotspot  $R$  is given by

$$\begin{aligned} P_0 &= \left(\frac{B}{A}\right)^C \times \left(\frac{A-B}{A}\right)^{(A-C)} \\ &= \frac{B^C \times (A-B)^{(A-C)}}{A^A} \end{aligned}$$

The alternate hypothesis  $H_1$  considers that the null hypothesis is not true. In this case, the probability that any point lies within the hotspot is given by  $\frac{C}{A}$ , and that the point is outside the hotspot is  $\frac{A-C}{A}$ . The probability that exactly  $C$  points are present within the hotspot  $R$  if the actual distribution is true is given by

$$P_1 = \left(\frac{C}{A}\right)^C \times \left(\frac{A-C}{A}\right)^{(A-C)}$$

$$= \frac{C^C \times (A - C)^{(A-C)}}{A^A}$$

The likelihood ratio is given by the expression  $\frac{P_1}{P_0}$ . Taking logarithm on both sides, we get the expression for log likelihood ratio.

$$\begin{aligned} \text{LLR} &= \log\left(\frac{P_1}{P_0}\right) \\ &= \log\left(\frac{C^C \times (A - C)^{(A-C)}}{A^A} \times \frac{A^A}{B^C \times (A - B)^{(A-C)}}\right) \\ &= \log\left(\left(\frac{C}{B}\right)^C \times \left(\frac{|A| - C}{|A| - B}\right)^{|A|-C}\right) \\ &= C \times \log\left(\frac{C}{B}\right) + (A - C) \times \log\left(\frac{A - C}{A - B}\right) \end{aligned}$$

## References

1. Eftelioglu E, Shekhar S, Kang JM, Farah CC (2016) Ring-shaped hotspot detection. *IEEE Trans Knowl Data Eng* 28:3367–3381
2. Eftelioglu E, Shekhar S, Oliver D, Zhou X, Evans MR, Xie Y, Kang JM (2014) Ring-shaped hotspot detection: a summary of results. In: Proceedings of IEEE international conference on data mining, pp 815–820
3. Xue Y, Brown DE (2003) A decision model for spatial site selection by criminals: a foundation for law enforcement decision support. *IEEE Trans Syst Man Cybern Part C: Appl Rev* 33(1):78–85
4. Turvey BE (2011) Criminal profiling: an introduction to behavioral evidence analysis. Elsevier, Amsterdam
5. Agarwal D, McGregor A, Phillips JM, Venkatasubramanian S, Zhu Z (2006) Spatial scan statistics: approximations and performance study. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, pp 24–33
6. Peterson John J (2009) Regression analysis of count data. *Technometrics* 41(4):371–371
7. Gardner W, Mulvey EP, Shaw EC (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol Bull* 118(3):392–404
8. Osgood DW (2000) Poisson-based regression analysis of aggregate crime rates. *J Quant Criminol* 16(1):21–43
9. Harries K (1999) Mapping crime: principle and practice. CDRC, NIJ. <https://www.ncjrs.gov/pdffiles1/nij/178919.pdf>
10. Bremer S (2000) An exploration of the methods for detecting hot spots and changes in hot spot locations. M.S. thesis, Univ. Virginia, Charlottesville, VA
11. Dalton J (1999) Bandwidth selection for kernel density estimation of geographic point processes. M.S. thesis, Univ. Virginia, Charlottesville, VA
12. Brown D, Liu H, Xue Y (2001) Mining preferences from spatial-temporal data. <https://doi.org/10.1137/1.9781611972719.26>
13. Cohen LE, Felson M (1979) Social change and crime rate trends: a routine activity approach. *Am Sociol Rev* 44(4):588–608
14. Rossmo DK (1999) Geographic profiling. CRC Press, Boca Raton, FL, USA
15. Brantingham PJ, Brantingham PL (1981) Environmental criminology. Sage Publications, Beverly Hills, CA, USA

16. Brantingham PL, Brantingham PJ (1993) Environment, routine and situation: toward a pattern theory of crime. *Routine Activ Ration Choice: Adv Criminol Theory* 5:259–294
17. Eck JE (2005) Crime hot spots: what they are, why we have them, and how to map them. In: *Mapping crime: understanding hot spots*. NIJ, pp 1–14. <http://discovery.ucl.ac.uk/11291/1/11291.pdf>
18. Chainey S, Dando J (2005) Methods and techniques for understanding crime hot spots. In: *Mapping crime: understanding hot spots*. NIJ, pp 15–34. <http://discovery.ucl.ac.uk/11291/1/11291.pdf>
19. Cameron JG, Leitner M (2005) Spatial analysis tools for identifying hot spots. In: *Mapping crime: understanding hot spots*. NIJ, pp 35–64. <http://discovery.ucl.ac.uk/11291/1/11291.pdf>

# Chapter 3

## Optimization Techniques for Machine Learning



Souad Taleb Zouggar and Abdelkader Adla

### 1 Introduction

Machine Learning (ML) is one of the areas of Artificial intelligence (AI). It aimed to extract and automatically exploit the crucial information present in large databanks. It refers to the development, analysis, and implementation of methods that enable a machine to evolve through a learning process and, thus, to perform tasks that are difficult or impossible to achieve by means of conventional algorithms.

ML algorithms draw on a variety of sources that combine different disciplines: statistics and data analysis [1], symbolic learning [2, 3], neural learning, inductive logic programming, reinforcement learning, statistical learning [4], support vector machines [5], expert committees, Bayesian inference and Bayesian networks [6], evolutionary algorithms (genetic algorithms, evolutionary strategies, genetic programming), databases, human–machine interfaces, etc. The optimization of learning methods saves storage space and prediction time by reducing the size of the obtained models obtained. This is essential for applications that require short response times.

In this study, we tried to address the following questions:

- To introduce the history, techniques, and application of machine learning to novice researchers;
- To provide a comprehensive review of machine learning methods;
- To identify the specific applications areas to which the commonly used learning methods are applied;
- To summarize the most popular optimization techniques used in machine learning;

---

S. T. Zouggar

Department of Economics, Oran 2 University, Oran, Algeria

e-mail: [Souad.taleb@gmail.com](mailto:Souad.taleb@gmail.com)

A. Adla (✉)

Department of Computer Science, Oran 1 University, Oran, Algeria

e-mail: [adla.abdelkader@univ-oran1.dz](mailto:adla.abdelkader@univ-oran1.dz)

- To discuss the strengths and the shortcomings of these techniques and highlight potential research directions.

The review presented here differs from previous works that in addition to describing the history and techniques of machine learning, it also gives a critical review of currently available selection measures.

The rest of the chapter is organized as follows: Sect. 2 provides an overview of machine learning. In Sect. 3, we present a general description of decision tree. Section 4 outlines ensembles methods, the different ways to generate them and their selection. Finally, concluding remarks and future work are given in Sect. 5.

## 2 Machine Learning

Machine learning becomes a major concern of artificial intelligence in the late 1970s, when expert systems face the challenge of acquiring existing expertise. It aims to build hypotheses from examples. The resulting hypotheses are judged according to two criteria: predictive efficiency (with respect to data) and intelligibility (with respect to the expert or the user) [7].

Machine learning is the development of programs that improve with experience. Its applications are numerous and concern a wide variety of fields. Examples include pattern recognition, in particular, speech and written word recognition, process control and fault diagnosis, etc. [8]. According to [9], the knowledge produced by the machine, in other words coming from machine learning, is not necessarily of a logical nature; it can take various forms: neural network, algebraic model, geometric model, etc. Simon [10] defines learning as “any change in the system that allows it to perform a task better the second time, when repeating the same task, or when another task occurs from the population.” Learning involves generalization from experience.

Why do we want a machine to learn to recognize an illness, for example, when so far the man has not done too badly? Various reasons may explain this need:

- The scarcity of specialists;
- The impossibility for humans to access certain hostile environments or difficult to access for reasons of cost or delay.

For example, in some clinical cases, the diagnosis of a disease is impossible without surgery. Developing an automatic diagnostic system would prevent some patients from having surgery wrongly and would allow the community to reduce health expenses and dispense patients from unnecessary acts.

There are two kinds of machine learning methods.

## 2.1 Empirical Learning Methods

Empirical learning methods are based on the acquisition of knowledge from examples. Empirical learning methods include case-based reasoning (CBR), artificial neural networks, decision trees, and genetic algorithms. These methods are divided between analog learning methods and induction learning methods.

**Learning Methods by Analogy.** Approaches based on analogy transfer knowledge on a well-known task to a less well-known one. Thus, it is possible to learn new concepts or to derive new solutions from similar known concepts and solutions. Two concepts become very important in the definition of learning by analogy: transfer and similarity, as for example case-based reasoning (CBR) systems [11].

**Induction Learning Methods.** In this approach, one seeks to acquire general rules representing the knowledge obtained from examples. The induction learning algorithm receives a set of learning examples and must produce classification rules allowing to classify the new examples. This algorithm can operate in a supervised or unsupervised manner [12, 13].

**Supervised Learning.** The goal is to find a general and featured description describing a class without having to enumerate all the examples of this class [14]. Learning strives toward two competing objectives:

1. An explanation of the studied concept, i.e., of the examples distribution in classes;
2. A decision or prediction function allowing assigning a class (insulin dependence, for example) to examples (patients) of which this one is unknown.

The goal of supervised learning is to construct a prediction model, also called classifier, which will allow to identify an attribute  $Y$  to predict, called endogenous variable, class, variable to explain or variable to predict, from a set of explanatory attributes  $X$ , called exogenous variables, explanatory variables or predictors, variables.

- The prediction model or classification function  $\varphi$  is built on a sub-ensemble of the population  $\Omega_a$ , called the learning sample;
- An individual  $\omega$  belonging to the sample;
- The attribute to predict  $Y$  associates with each individual of  $\Omega_a$  a class belonging to  $C$  {set of classes  $C = \{c_1, \dots, C_m\}$ }.

$$Y : \Omega_a \rightarrow C$$

$$\omega \rightarrow Y(\omega)$$

- $X$  the exogenous (explanatory) variable is defined by:

$$X : \Omega_a \rightarrow E_j$$

$$\omega \rightarrow X_j(\omega)$$

$$E_j = \{e_{1j}, e_{2j}, \dots, e_{pj}\} : \text{set of modalities (values) of } X_j.$$

**Unsupervised learning.** The unsupervised learning system considers a set of objects or examples without knowing if these objects belong or not to the same class. It tries to find the regularities between the examples while carrying out the best possible clusterings. These clusters of similar objects are called prototypes [15].

**Learning methods based on explanation.** The learning methods based on explanation (explanation-based methods—EBL) use preexisting knowledge and deductive reasoning to increase the information provided by sets of examples. These methods are known as analytical learning [16].

## 2.2 Illustrative Data

The examples in this section are used to illustrate the different concepts associated with the methods aforementioned. These are three bases of learning in the medical field: The DIABETES database which groups patients to be classified as Type I or Type II. The ULCERE base represents patients who are suffering or not from ulcer perforation and, the MONITDIAB database to detect classes of complications for diabetics.

**DIABETES Learning Base.** The DIABETES database contains 461 individuals  $\omega$  and 10 descriptive variables allowing to divide individuals (patients) in two class values. The 10 exogenous variables  $X_j$  ( $j = 1 \dots 9$ );  $X_j \in \{\text{Age, Weight, Anteced, State, ASSO, CDC, MR, IV, Sex, AST}\}$ ,  $E3 = \{0,1,2\}$  represents the set of values of the variable Anteced, the variable  $C$  to predict corresponds to the type of DIABETES, noted CLASS and takes its values in {0: type I diabetes, 1: type II diabetes} (Table 1).

**Table 1** Description of the diabetes base

Variable	Meaning	Possible values
Age	Age of discovery of diabetes	{0: > 35; 1: [15,35[; 2: other}
MR	Revealing mode	{0: Spontaneous, 1: Infectious home, 2: Glycemic imbalance} Récent
Poids	Patient weight	{0: Normal, 1: skinny, 2: Obese, 3: Overweight}
IV	Viral infection	{1: Yes, 0: No}
Etat	State	{1: emaciation, 0: no slimming}
Assoc	Association	{1: relationship with autoimmune diseases, 0: No}
CDC	Condition of discovery	{0: Diabetic feet (CDC0), 1: Fortuitous (CDC1), 2: Bacterial infection (CDC2), 3: Retinopathy (CDC3), 4: Hyperosmolar comas (CDC4), 5: Inaugural Diabetic Ketosis (CDC5), 6: Ketotic Comas (CDC6)}
AST	Asthenia	{1: Yes, 0: No}
Antéced	antecedents	{0: Family, 1: personal, 2: No history}
Classe	Type of diabetes	{0: Type I, 1: Type II}

**ULCERE Learning Base.** The base consists of 130 individuals and 12 descriptors and a two-valued class indicating the existence of ulcer perforation or not. The 12 exogenous variables  $X_j$  ( $j = 1, 12$ );  $X_j = \{\text{DEPIG, AGC, PYROSIS, DB, VOUM, DAP, BEPIG, DCR, CEPIG, FEVER, EMAT, DEDPP}\}$ ,  $E11 = \{0, 1\}$  represents the set of values of the variable EMAT, the variable to predict noted CLASS takes its values in {0: Unperforated Ulcer, 1: Perforated Ulcer} (Table 2).

**MONITDIAB Learning Base.** The MONITDIAB [17] application contains 353 patients showing various complications. The latter are described by 13 exogenous variables allowing to dissociate 5 clusters of complications representing the different class values. Table 3 describes the different exogenous variables with their possible values.

**Table 2** Description of the Ulcer base

Variable	Meaning	Possible values
DEPIG	Epigastric pain	{0: No existence of pain, 1: Existence of pain}
CAG	Abdominal contraction widespread	{0: No existence of contraction, 1: Existence of contraction}
Pyrosis	Heartburn	{0: No existence of Heartburn, 1: Existence of Heartburn}
DB	Brutal pain	{0: No existence of Brutal Pain, 1: Existence of Brutal Pain}
VOUM	Vomiting	{0: No existence of vomissement, 1: Existence of Vomiting}
DAP	Abdominal defense at the palpitation	{0: No existence de défense, 1: Existence Abdominal Defense}
DCR	Pain calmed by meals	{0: No existence, 1: Existence}
CEPIG	Epigastric cramp	{0: No existence of Cramp, 1: Existence of Cramp}
Fièvre	Fever	{0: Non existence of Fever, 1: Existence of Fever}
EMAT	Hematemesis and melena	{0: No existence of EMAT, 1: Existence of EMAT}
DEDPP	Epigastric pain triggered in postprandial	{0: No existence of DEDDP, 1: Existence of DEDDP}
Classe	Perforation of the ulcer	{0: No perforation, 1: Perforation}

**Table 3** Description of the MONITDIAB database

Variable	Code	Possible values
Type of Diabetes	TD	Type 1, Type 2 Type 2: (DNID)
State	Var	A pregnant woman (FE); An adult (Adult), If age $\leq$ 70; An old person (VP), If age $>$ 70
Body mass index	IMC	18 $<$ IMC $<$ 20 => Skinny (M); 20 $<$ IMC $<$ 25 => Normal; 25 $<$ IMC $<$ 30 => Overweight (SP); 30 $<$ IMC $<$ 35 => ObsG1; 35 $<$ IMC $<$ 40 => ObsG2; IMC $>$ 40 => ObsG3
Glycemia	Glyc	0.70G $\leq$ Glyc $<$ 1.80G => Normal; Glyc $<$ 0.70G => hypoglycemia (HypoG); 1.80G $\leq$ Glyc $\leq$ 6G => hyperglycemia (HyperG)
HB1NC	HBNC	Balanced (E); Imbalanced (D); Very unbalanced (TD)
Eye fundus examination	EOF	Retinopathy (R); No Retinopathy (PR)
creatinine	Crea	6 G/L $\leq$ Crea $\leq$ 13 G/L then Crea = Normal; Crea $>$ 13 G/L then Crea = Abnormal
Urea	Urée	0.30 G/L $\leq$ Urée $\leq$ 0.50 G/L then Urée = Normal; Urée $>$ 0.50 G/L then Urée = Abnormal (Renal failure)
Microalbuminuria	McrAlb	McrAlb = 20 mg/24 h then McrAlb = Normal; 30 $<$ McrAlb $<$ 100 then Stage 3A Diabetic Nephropathy (NDS3A); Si 100 $<$ McrAlb $<$ 300 then 3B Diabetic Nephropathy (NDS3B); McrAlb $>$ 300 then 4 Diabetic Nephropathy (NDS4); Si McrAlb $>$ 300 and High urea and high Crea then 5 Stage Renal failure (IR5)
Clearance of creatinine	Cc	70 $<$ Cc $<$ 100 Mild renal failure (IRL); 40 $<$ Cc $<$ 70 Moderate renal failure (IRM); 10 $<$ Cc $<$ 30 Strict renal failure (IRS); Cc $<$ 10 Very Stric renal failure (IRTS)
Neuropathy	Neuropath	Existence of Neuropathy (Neurpath); No Existence of Neuropathy (PNeurpath)
Electrocardiogram	ECG	Normal; Coronary failure (InsufCor); Heart failure (InsuffCar)
Arterial Doppler	DA	Existence of Arteriopathy (Art); No Arteriopathy (PArt)

### 3 Decision Trees

Decision trees emerged with the AID algorithm “Automatic Interaction Detection” [18]. Decision trees use regression trees for prediction. Among the improvements to AID, for example, the CHAID method “CHi-square AID” of [19] is used for classification.

The real success of these methods was resided in the development of CART and ID3 algorithms [2, 13] which laid down the theoretical and applied foundations of a new research field. Quinlan [3, 13] then proposes a set of heuristics to improve his system. He proposed C4.5 in 1993 and C5.0 implemented in commercial software.

#### 3.1 Measures of Partition Quality

To calculate the quality of a partition  $S$ , it is necessary to introduce quantities that allow comparing the different possible choices (according to the attributes). Functions and measures of quality of the partitions noted  $I(S)$  are thus defined. They allow measuring the degree of mixture of the examples between the different classes. This function must take:

- Its minimum value when all the examples are in the same class;
- Its maximum value when the examples are equi-distributed.

**Uncertainty variation (the gain):** When passing from a partition  $S_i$  to a partition  $S_{i+1}$ , we will maximize the quantity ( $\mathfrak{I}(S_{i+1}) = I(S_i) - I(S_{i+1})$ ) which is the uncertainty variation (gain) between the previous partition  $S_i$  and the following partition  $S_{i+1}$ .

**Distance Measures between Probability Distributions.** These measurements allow estimating the difference between two probability distributions, noted down from the frequencies observed in the sample.

**Distance from Kolmogorov-Smirnov.** Derived from the statistical domain, it calculates the maximum distance between two probability distributions. It is used as a partition criterion in [20] and has performances identical to those obtained with the gain ratio.

**Chi-square independence test.** It used in [21] directly as a segmentation criterion and to counterbalance the tendency of entropy-based criteria to favor multivalued attributes.

**Measures of the Information Theory.** Having  $n$  possible equiprobable messages (each having the probability  $p$  of  $1/n$ ), the amount of information provided by a message is  $-\log_2(p) = \log_2(n)$ . With a probability distribution  $P = (p_1, p_2, \dots, p_n)$ , the information provided by this distribution called entropy of  $P$  is:  $I(P) = -(p_1 * \log_2(p_1) + p_2 * \log_2(p_2) + \dots + p_n * \log_2(p_n))$ . If  $P = (0.5, 0.5)$ , then  $I(P) = 1$ ; if  $P = (0.67, 0.33)$ , then  $I(P) = 0.92$ ; and if  $P = (1, 0)$ , then  $I(P) = 0$ . If a set  $\Omega$  of records form a partition  $C_1, C_2, \dots, C_k$  based on the value of the target attribute, then the information needed to identify the class of an element of  $\Omega$  is  $\text{Info}(\Omega) = I(P)$  where  $P$  is the probabilistic distribution of the partition  $(C_1, C_2, \dots, C_k)$ :

$$P = \left( \frac{|C_1|}{|\Omega|}, \frac{|C_2|}{|\Omega|}, \dots, \frac{|C_k|}{|\Omega|} \right)$$

In the example DIABETE, we have  $\text{Info}(\Omega) = I(9/14, 5/14) = 0.64$ , whereas in the ULCERE example  $\text{Info}(\Omega) = I(6/10, 4/10) = 0.66$ . If we partition  $\Omega$  based on the values of a non-target attribute  $X$  in sets  $\Omega_1, \Omega_2, \dots, \Omega_n$ , then the information needed to identify the class of an element of  $\Omega$  becomes the weighted average of the information needed to identify the class of an element of  $\Omega_i$ , namely the weighted average of  $\text{Info}(\Omega_i)$ :

$$\text{Info}(X, \Omega) = \sum_{i=1,n} \left( \frac{|\Omega_i|}{\Omega} * \text{Info}(\Omega_i) \right) \quad (1)$$

In the case of the DIABETE database, the calculation of the information provided by the State variable is given by:

$$\begin{aligned} \text{Info}(\text{Etat}, \Omega) &= 8/14 * I(7/8, 1/8) + 6/14 * I(2/6, 4/6) \\ &= 0.205 + 0.268 = 0.47 \end{aligned}$$

In the case of the ULCERE database, the calculation of the information provided by the DEPIG variable is given by:

$$\text{Info}(\text{DEPIG}, \Omega) = 8/10 * I(6/8, 2/8) + 2/10 * I(0/2, 2/2) = 0.45$$

Consider the quantity Gain ( $X, \Omega$ ) defined as follows:  $\text{Gain}(X, \Omega) = \text{Info}(\Omega) - \text{Info}(X, \Omega)$ . The gain represents the difference between the information needed to identify an element of  $\Omega$  and the information needed to identify an element of  $\Omega$  after obtaining the value of the attribute  $X$ . This is the information gain due to the attribute  $X$ . In the DIABETES example, the gain for the *State* variable is:

$$\text{Gain}(\text{State}, \Omega) = \text{Info}(\Omega) - \text{Info}(\text{Etat}, \Omega) = 0.64 - 0.47 = 0.17$$

If we consider the Assoc attribute, we find  $\text{Info}(\text{Assoc}, \Omega)$  is equal to 0.36 and  $\text{Gain}(\text{Assoc}, \Omega)$  is 0.28. We deduce that the Assoc variable offers more information than the *State* attribute. The notion of gain is used to classify attributes and build a decision tree. At each node, there is the attribute that has the largest gain compared to the others. The advantage of this scheduling is to create a small decision tree which allows identifying a record with a small number of questions.

### 3.2 ID3 Method

ID3 is the first popular decision tree algorithm proposed by Quinlan in 1986 [13] for supervised classification. The tree is un-pruned, non-incrementally scalable and greedy, and where Shannon's entropy is used for data partitioning.

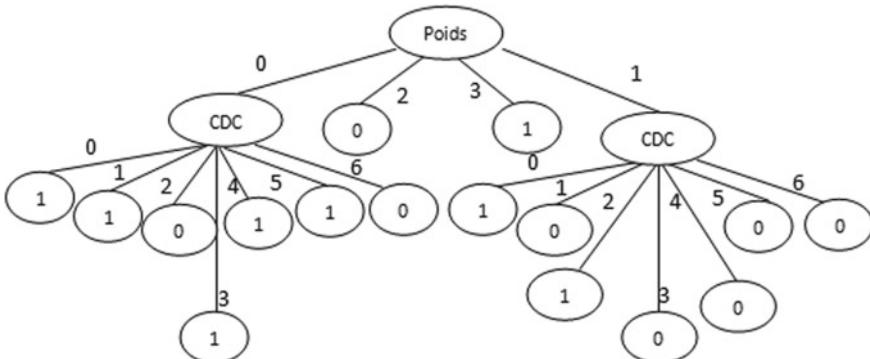
```

ID3 algorithm;
Input: X (exogenous variables), Y (class), learning
sample  $\Omega_a$ ;
If  $\Omega_a$  is empty then return a node of value failure;
If  $\Omega_a$  consists of similar values for the class then
return a node labeled by the value of that class;
If X is empty then return a simple node with as value
the most frequent value of the class values in  $\Omega_a$ ;
D  $\leftarrow \text{argmax}_{X_j} \text{gain}(X, \Omega_a)$  with  $X_j$  in X;
{ $d_{ji}$  with  $i = 1 \dots p$ } the values of the variable  $X_j$ ;
{ $\Omega_{ai}$  with  $i = 1 \dots p$ } the subsets of  $\Omega_a$  composed of
individuals having  $d_{ji}$  values of the variable  $X_j$ ;
Root tree D and arcs labeled by  $d_{j1}, \dots, d_{jp}$  going to
subtrees ID3( $X - D, Y, \Omega_{a1}$ ), ID3( $X - D, Y, \Omega_{a2}$ ), ..., ID3( $X - D, Y, \Omega_{ap}$ );
Output: ID3 decision tree;

```

By applying the ID3 algorithm, on an excerpt of the Diabetes base composed of 132 patients, we obtain the tree (see Fig. 1).

- From a decision tree, we can extract rules in the form:  
If <Condition> Then <Class ci> (degree of likelihood =  $\frac{\text{nbre\_instances\_ci}}{\text{effectif\_total\_feuille}}$ )
- From the tree of Fig. 1, we can extract sixteen rules. The extraction is done from the root and going toward the leaves of the tree. For example, we have the five following rules:
- If Weight = 0 and CDC = 0, then TD = 1  $\Leftrightarrow$  If Weight = ‘Normal’ and CDC = ‘Diabetic Foot,’ then Diabetes Type II.
- If Weight = 2, then TD = 0  $\Leftrightarrow$  If Weight = ‘Obese,’ then Diabetes Type I.
- If Weight = 1 and CDC = 3, then TD = 0  $\Leftrightarrow$  If Weight = ‘lean’ and CDC = ‘Retinopathy,’ then Diabetes Type I.
- If Weight = 1 and CDC = 4, then TD = 0  $\Leftrightarrow$  If Weight = ‘lean’ and CDC = ‘Coma hyperosmolar,’ then Diabetes Type I.



**Fig. 1** Tree partitions

- If Weight = 0 and CDC = 6, then TD = 0  $\Leftrightarrow$  If Weight = ‘Normal’ and CDC = ‘Ketotic Comas,’ then Diabetes Type I.

### 3.3 Measure of Segmentation

The decision trees are easily interpretable because of their graphical representation and have good prediction and generalization performance. Referring to the 2001 study conducted by Piatetsky-Shapiro on his site dedicated to the industrial market for extracting knowledge from data, decision trees are used by more than 50% of the population surveyed.

In the study conducted in 2007 and in response to the question “What are the most used data mining tools in the last 12 months? 62.2% of respondents cited the decision trees in the site <http://www.kdnuggets.com/polls/2007/dataminingmethods.htm>. These statistical studies confirm the importance of these methods, mainly because of their ease of use and their interpretability; these properties make these methods widely used in areas that require justification for decision making as in the medical field.

The most important constructing element of a decision tree classifier is the measure used to assess the quality of a partition. These measures belong to two main categories: those based on entropy and those based on the notion of distance. A partition quality calculation measure called distance-based new information measure (NIM) is proposed in [22]. It allows generating smaller-sized trees with high performances.

**Description of the Measure.** The following notations are used:

- $n$ : The total number of individuals in the learning sample  $\Omega_a$ ;
- $n_i$ : The number of individuals of class  $i$ ;
- $e_{sj}$ : Modality  $s$  of the variable  $X_j$ ;
- $n_{sj}$ : Number of individuals associated with the modality  $s$  of the variable  $X_j$ ;
- $n_{isj}$ : Number of individuals class  $i$  associated with the modality  $s$  of the variable  $X_j$ ;
- $m$ : The number of modalities of the class,

The measure NIM uses two functions:

- The importance function denoted Imp which has as a parameter a variable modality. Let  $e_{sj}$  be the modality  $s$  of the variable  $X_j$ ,  $\text{Imp}(X_j = e_s) = \text{Imp}(e_{sj}) = \sum_{i=1,m} |n_{isj} - (n_{sj}/m)|$ .
  - The function  $f$ :
- Calculate the quantity  $f(\Omega_a) = \sum_{i=1,m} |n_i - (n/m)|$  associated with the sample  $\Omega_a$ .

- Calculate the quantity  $f(X_j)$ ,  $X_j$  is an exogenous variable of modalities  $e_{1j}, \dots, e_{sj}$ ,  $f(X_j) = [\sum_s \text{Imp}(e_{sj})] + \sigma^*$  number of leaves;  $\sigma$  is an empirically determined parameter that favors variables that generate the most leaves for the next partition.

**Tree Generation Process.** To generate a tree from the training sample using NIM, the following steps are performed:

Step 1:

- We calculate for the sample  $\Omega_a$  the quantity  $f(\Omega_a) = \sum_{i=1,m} |n_i - (n/m)|$ , we test if the individuals belong to the same class: If  $|f(\Omega_a) - n| \neq 0$ , then the individuals do not belong to the same class.

Step 2:

- To label the initial node, we compute for each descriptive variable  $X_j$  having  $e_{1j}, e_{2j}, \dots, e_{nj}$  as modalities the following quantities:  $\text{Imp}(e_{sj}) = \sum_{i=1,m} |n_{isj} - (n_{sj}/m)|$
- We calculate  $f(X_j) = [\sum_s \text{Imp}(e_{sj})] + \sigma^*$  number of leaves, for each variable  $X_j$  candidate to the segmentation;
- We choose the variable that maximizes the quantity  $f(X_j)$ .

Step 3:

- For the chosen variable  $X_j$  and for each of its modalities  $e_{1j}, e_{2j}, \dots, e_{nj}$ , if  $|\text{Imp}(e_{sj}) - n_{sj}| = 0$ , then the branch associated with the modality leads to a leaf.
- If  $|\text{Imp}(e_{sj}) - n_{sj}| \neq 0$ , then we repeat the step 1 by considering the remaining variables and considering only the subpopulation associated with the branch labeled by the modality  $e_{sj}$ .

Step 4:

- End the process when all nodes are “pure” leaves.

**Partitions’ Generation: Application on MONITDIAB.** To illustrate, we consider the example MONITDIAB for monitoring diabetics. We assign to  $\sigma$  the value 0:

Step 1: At the beginning of learning, the learning sample  $\Omega_a$  has the following initial configuration presented in Table 4:

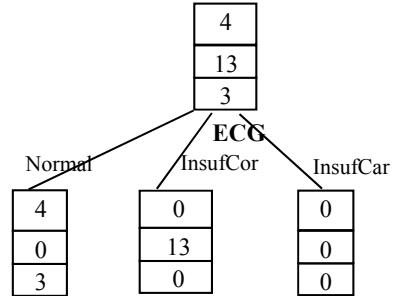
$$f(\Omega_a) = |4 - 20/3| + |13 - 20/2| + |3 - 14/2| = 2.66 + 6.34 + 3.66 = 12.66, \text{ la valeur } |f(\Omega_a) - 20| = |12.66 - 20| <> 0, \text{ then we deduce that the concerned node is not terminal.}$$

Step 2: We label the initial node by one of the 13 exogenous variables:

**Table 4** Initial distribution of individuals according to class values

Instances	DDTC	DDMC	DENC
20	4	13	3

**Fig. 2** Two first tree partitions



For the TD variable whose modalities are Type I and Type II:  $\text{Imp}(\text{TypeI}) = |4 - \frac{9}{3}| + |5 - \frac{9}{3}| + |0 - \frac{9}{3}| = 6$ ,  $\text{Imp}(\text{TypeII}) = |0 - \frac{11}{3}| + |8 - \frac{11}{3}| + |3 - \frac{11}{3}| = 8.66$ . Alors  $f(\text{TD}) = 6 + 8.66 = 14.66$ , the calculation is done in the same way for the remaining variables  $f(\text{Var}) = 14.66$ ,  $f(\text{IMC}) = 6.63$ ,  $f(\text{Glyc}) = 10.66$ ,  $f(\text{HBANC}) = 13.34$ ,  $f(\text{EFO}) = 18.65$ ,  $f(\text{Crea}) = 13.32$ ,  $f(\text{Urée}) = 12.66$ ,  $f(\text{McrAlb}) = 13.29$ ,  $f(\text{Cc}) = 12.66$ ,  $f(\text{Neuropath}) = 20.64$ ,  $f(\text{ECG}) = 21.97$ ,  $f(\text{DA}) = 13.97$ .

We choose the variable that maximizes the function  $f$ . So the ECG variable is chosen to split the root node (see Fig. 2).

The proposed algorithm named IDT\_NIM performs recursive partitioning as adopted by the ID3 method. The partitioning of the generated child nodes is done in the same way as the partitioning of the root node. The process stops when all the obtained nodes are homogeneous leaves. The different steps are described by the following pseudocode:

```

IDT_NIM algorithm;
Input:X(Exogenous Variables), Y(Class), Ωa(Learning Sample);
Calculate  $f(\Omega_a)$ ;
If  $|f(\Omega_a) - n| = 0$  Then "the tree is the root node";
D ← argmaxXj  $f(X, \Omega_a)$ , Xj in X;
{edj (d=1...k)} ensemble of k modalities;
{Ωaj (j=1...k)} sub ensembles of Ωa associated with the value edj of Xj;
If  $|f(e_{dj}) - f(\Omega_{aj})| \neq 0$  Then from D generate the sub-tree
IDT_NIM (X-D, Y, Ωaj) associated with the modality edj of Xj;
Otherwise from D generate a leaf associated with the modality edj of Xj and whose size is Ωaj;
Output: IDT_NIM tree.
  
```

An experimental study given in [22] shows the interest of NIM compared to Shannon's entropy and the gain ratio.

## 4 Ensembles' Methods

The problem of classification models instability, for example those based on decision trees, resides in that insignificant changes in the learning sample can cause large changes in the generated classification rules. Therefore, the rules generated from two similar samples with a few differences can be completely different. Different models or hypotheses ( $H$ ) are constructed from “almost” similar samples which complicates the decision-making process. The theoretical quality of a hypothesis  $H$  can be calculated by measuring the deviation, for each example  $x$  of  $X$ , between the result of  $H$  and that of  $y$ .

### 4.1 Aggregation of Models

An aggregated set contains different models obtained by perturbations of the initial sample. The error of an aggregate set is less than the error of each individual model provided that:

- The different models have uncorrelated errors. The error correlation between two models  $h_1$  and  $h_2$  is the probability that they make the same error knowing that one of them makes a mistake.

$$\begin{aligned} C_{\text{err}}(h_1, h_2) &= P(h_1(x_i) = h_2(x_i) \text{ tel que } h_1(x_i) \neq y_i \vee h_2(x_i) \neq y_i) \\ &= \sum_{i=1}^n \frac{I(h_1(x_i) = h_2(x_i))}{I(h_1(x_i) \neq y_i \vee h_2(x_i) \neq y_i)} \end{aligned} \quad (2)$$

- The models for which the prediction error is less than 0.5 are good enough. The probability of error of a set  $J$  of models is equal to the probability that  $J/2$  models are mistaken follows a binomial distribution.

Model aggregation goes through two stages: a diversification stage which allows different models to be selected to minimize error correlation. Diversification results in covering different regions in the instance space. This stage is followed by an integration that combines them to maximize the space covered. This integration can be static (vote or basic predictions average) or dynamic (use an adaptive process to integrate the basic predictions (meta-learning)).

**Diversification by Resampling.** There are four types of diversification by resampling:

**Bagging.** Bagging bootstrap aggregating is a resampling method introduced by Breiman in 1996 [23]. Given a learning sample  $\Omega_a$  and a prediction method called basic rule which builds on  $\Omega_a$  a predictor  $\hat{h}(., \Omega_a)$ , bagging consists to draw with reset several bootstrap samples  $(\Omega_a \theta_1, \dots, \Omega_a \theta_q)$ , apply on them the basic rule (decision tree) to generate a collection of predictors  $(\hat{h}(., \Omega_a^{\theta_1}), \dots, \hat{h}(., \Omega_a^{\theta_q}))$ , and finally,

combine (aggregate) these basic predictors. A bootstrap sample  $\Omega_a^l$  is obtained by randomly drawing with reset n observations in the sample  $\Omega_a$ . Each observation has a probability of  $\frac{1}{n}$  to be drawn. The random variable  $\theta_l$  represents this random draw. Initially, bagging was introduced with the basic rule of a decision tree, but the schema is general and can apply to other basic rules.

**Boosting.** In [24], Freund and Schapire introduce boosting concept which, theoretically, is able to significantly reduce the error of an algorithm generating a classifier that has no significant performance compared to a randomly constructed classifier. They also introduce the notion of “pseudo-loss” which forces a learning algorithm to focus on the most difficult labels to discriminate. Given a learning sample  $\Omega_a$  and a prediction method (basic rule), which builds on  $\Omega_a$  a predictor  $\hat{h}(\cdot, \Omega_a)$ , boosting consists of drawing a first bootstrap sample  $\Omega_a^{\theta_1}$  where each observation has a probability  $\frac{1}{n}$  of being drawn, then applying the basic rule to obtain a first predictor  $\hat{h}(\cdot, \Omega_a^{\theta_1})$  and, then calculating the error of  $\hat{h}(\cdot, \Omega_a^{\theta_1})$  on the learning sample  $\Omega_a$ . A second sample bootstrap  $\Omega_a^{\theta_2}$  is then drawn, but the drawing law of the observations is now no longer uniform.

The probability for an observation to be drawn depends on the prediction of  $\hat{h}(\cdot, \Omega_a^{\theta_1})$  on this observation. The principle is to increase the probability of drawing an incorrectly predicted observation and to decrease that of drawing a well-predicted observation. Once the new sample  $\Omega_a^{\theta_2}$  is obtained, we apply again the basic rule  $\hat{h}(\cdot, \Omega_a^{\theta_2})$ . We then draw a third sample  $\Omega_a^{\theta_3}$  which depends on the predictions of  $\hat{h}(\cdot, \Omega_a^{\theta_2})$  on  $\Omega_a$  and so on. The collection of predictors obtained is then aggregated using a weighted average.

**Randomizing Outputs.** In [25], Kodratoff introduces the randomizing output method which is an ensemble method of a different nature. It consists of constructing independent samples in which the outputs of the training sample are modified. The modifications that the outputs undergo are obtained by adding a noise variable to each  $Y_i$  of  $\Omega_a$ . A collection of randomized output samples is obtained, a basic rule is then applied to each sample, and finally, all the predictors obtained are aggregated.

**Random Subspace.** Another type of ensemble method is introduced in [26]. It is no longer to perturb the sample but rather to play on all the variables considered. Random subspace method consists of randomly draw a subset of variables and to apply a basic rule on  $\Omega_a$  that consider only the selected variables. We generate a collection of predictors, each of them is built using different variables, and then, we aggregate these predictors. The subsets of variables are drawn independently for each predictor. The idea behind this method is to construct several predictors, each of them is good in a particular subspace  $X$ , and then to reduce a predictor on the entire input space.

**Diversification by Hybridization.** This diversification technique consists of varying the learning algorithms:

**Stacking (Stacked Generalization).** It is also called generalization by stacking [27]. It is carried out in two levels: at the first level (level 0), a diversification by varying the learning methods; the second level (level 1) is integration phase by meta-learning.

*Multi-strategy methods.* Diversification is done by training M learning algorithms on a basic dataset and measuring their performance on a test set [28]. The models with a minimum error correlation are selected. The integration is done statically (continuous predictions: calculation of the average, median, linear combination, etc., and discrete predictions: uniform or continuous votes) or dynamically (by meta-learning).

## 4.2 Ensembles' Selection

The ensemble simplification of classifiers, called an ensemble pruning or an ensemble selection, allows to reduce the size of an ensemble before integration phase. Simplification of ensemble methods is important for two main reasons: prediction efficiency and performance. The less the ensemble consists of models, the shorter the execution time and the used memory space. Models with reduced performance negatively affect ensemble performance. Similar models reduce ensemble diversity, eliminate models with reduced performance while maintaining a high diversity within the remaining models, and allow good prediction performance [29].

A taxonomy of ensemble pruning methods of classifiers is proposed, and the main categories of methods are presented in [30]. According to the authors' contributions, the methods of ensemble pruning can be grouped into three basic categories:

**Ranking-Based Methods.** The models are first ordered based on an evaluation function, and then, the final number of models is chosen based on their order. One approach is to use a user-specified amount or a percentage of models [31–33].

**Clustering-Based Methods.** These methods consist of two steps: In a first step, they use a clustering algorithm in order to discover clusters of models that make similar predictions. In a second step, any cluster is simplified separately in order to increase the diversity of the ensemble. The main objective of these methods is to search clustering algorithms that are based on distance. The objective is to choose an adequate distance measure [34, 35].

**Optimization-Based Methods.** For these methods, the ensemble simplification is transformed into an optimization problem. It consists to find a sub-ensemble of the original ensemble that optimizes an indicative measure of its performance in generalization (precision on a validation ensemble). An exhaustive search in the sub-ensemble space is not feasible for an ensemble of moderate size. Three optimization approaches are considered for simplification: genetic algorithms using GASEN-b [36], semi-defined programming [37], and hill climbing [38]. We detail hereafter hill-climbing methods.

Hill-climbing methods allow you to replace an initial set of models with a subset using a greedy search procedure. Two paths can be used, forward selection (FS) or backward elimination (BE), and require the evaluation of  $T(T - 1)/2$  sub-ensembles ( $T$  the number of models). In FS, the starting sub-ensemble is initialized to the empty set. The algorithm progresses by adding to  $S$  a model  $m_i \in M \setminus S$  which optimizes a certain evaluation function  $f_{FS}(S, M_t, D)$  where  $S$  represents the current sub-ensemble

$t$ ,  $m_t$  the model to be added,  $D$  the evaluation or pruning ensemble. In a BE, the current sub-ensemble  $S$  is initialized to the full set  $M$  and the algorithm continues by iteratively eliminating from  $S$  the model  $m_t$  which optimizes the evaluation function  $f_{BE}(S, M_t, D)$ .

The search space is composed of the possible ensembles. The sub-ensembles are called states. The transition from one state to another is done using neighborhood. For example, the neighborhood of a sub-ensemble  $S = \{M_1, M_2\}$  is the ensemble  $\{\{M_1, M_2, M_3\}, \{M_1, M_2, M_4\}\}$  en FS et  $\{\{M_1\}, \{M_2\}\}$  en BE.

During a hill-climbing search, an evaluation function is used to judge the relevance of a sub-ensemble. Giving a sub-ensemble  $S$  and a model  $m$ , such a function allows estimating the possibility of inserting (eliminating)  $m$  at (of)  $S$ . These functions can be based on performance and/or diversity [39].

### 4.3 Selection Measures

The proposed measures are based on diversity and/or performance. The paths are hill-climbing ones or based on genetic algorithms.

**Multi-objective function.** This function [39, 40] allows a directed hill-climbing ensemble pruning (DHCEP) [38] search in a homogeneous ensemble of C4.5 trees [3]. The selected sub-ensemble must come to a compromise between diversity maximum and minimum error rate.

The motivation behind the joint use of the two criteria is that there is a relation between the individual performance of the classifiers and their diversity. The more precise the classifiers, the less they disagree. Using one of the two properties is not sufficient to find the best performing sub-ensemble. The multi-objective function is based on this compromise between tree individual performance and diversity of trees. A reduced number of trees allow a gain in memory space and computing time that can be very significant for large samples and real-time applications.

The function  $S$  is given by:

$$S = \frac{nX(\sum_{i=1}^n \theta_i^2) - X}{nk - X} + \alpha \frac{kn^2(\sum_{j=1}^k e_j^2) - X^2}{Xkn - X} \quad (3)$$

- $\alpha$ : Parameter determined empirically (usually is assigned the value of the learning sample size);
- $\theta_i = \frac{x_{i+}}{x}$  { $x_{i+}$  the total number of errors made for the individual  $i$ };
- $e_j = \frac{x_{j+}}{n}$  {the error rate associated with the model  $j$ };
- $X$  total number of errors made by a sub-ensemble of models at time  $t$ ;
- $n$  number of individuals in the selection sample;
- $k$  size of the current model sub-ensemble.

The Pruning Ensemble using Diversity and Accuracy (PEDA) algorithm given below summarizes the steps to simplify B-trees generated by bagging and using a hill-climbing path:

```

PEDA algorithm;
Entry: B = {A1, ..., Ak}
Eval: validation or pruning sample;
Neighborhood ( $\varphi_j$ ): function that returns the sub
ensembles of models obtained from  $\varphi_j$  by adding a model
(tree);
Initialize ( $\varphi_0$ );
Calculate S ( $\varphi_0$ , Eval);
If  $\exists \varphi_j$  such that  $S(\varphi_0, \text{Eval}) < S(\varphi_j, \text{Eval})$  où  $\varphi_j \in$ 
Neighborhood ( $\varphi_0$ ) then  $\varphi_0 = \text{argmin } \varphi_j (S(\varphi_j, \text{Eval}))$ ;
Go to 3;
Output: A sub ensemble  $\varphi_0$ ,  $\varphi_0 \subseteq B$ .

```

**Entropy function.** The entropy function is a diversity-based function presented in [41] and is used in [31] to simplify heterogeneous ensembles (an ensemble of different models). The function denoted  $f_E$  is given by:

$$f_E = \frac{1}{n} \sum_{j=1}^n \frac{1}{T - \frac{T}{2}} \min\{nc(x_j), T - nc(x_j)\} \quad (4)$$

- $n$ : Size of the learning sample;
- $T$ : The number of classifiers of the current ensemble;
- $nc(x_j) = \sum_{i=1}^T y_{ji}$ ;  $y_{ij} = 1$  if the classifier  $i$  correctly classifies the individual  $j$  and 0 otherwise;
- $f_E \in [0, 1]$  where 1 indicates a very large diversity and 0 an absence of diversity, so the goal is to maximize the function  $f_E$ .

The  $f_E$  measure was used with two paths: a hill-climbing path and a path based on genetic algorithms. For a search based on genetic algorithms, we suppose that we have an ensemble of four trees  $C = \{T_1, T_2, T_3, T_4\}$ , the chromosome  $ch_1 = (1 \ 0 \ 1 \ 0)$  corresponds to the fact that the trees  $T_1$ , and  $T_3$  are chosen in the sub-ensemble. To the two trees correspond classification vectors on  $\Omega_v$ . It is also assumed that  $|\Omega_v| = 2$ , we associates, for example, with  $T_1$  and  $T_3$  the vectors of classification  $(1 \ 0)^t$  and  $(0 \ 1)^t$  respectively. Calculating the fitness function  $ff_E$  for chromosome  $ch_1$  is equivalent to calculating  $f_E$ :

$$f_E = \frac{1}{n} \sum_{j=1}^n \frac{1}{T - \frac{T}{2}} \min\{nc(x_j), T - nc(x_j)\}$$

- $n = 2 = |\Omega_v|$ ,  $T = 2$  (the classifiers to which 1 corresponds to chromosomes  $T_1$  and  $T_3$ ),  $x_1$  et  $x_2$  are individuals of  $\Omega_v$  classified, respectively  $(1\ 0)^t$  et  $(0\ 1)^t$  par  $T_1$  et  $T_3$ .
- $nc(x_1) = 1$  (the number of trees that correctly classify instances  $x_1$ ).
- $nc(x_2) = 1$  (the number of trees that correctly classify instances  $x_2$ ).
- $f_E = \frac{1}{2} \left( \frac{1}{2-\frac{2}{2}} * \min(1, 2-1) + \frac{1}{2-\frac{2}{2}} * \min(1, 2-1) \right) = 1$  et  $ff_E = 1 - f_E = 0$ ,  
 $ff_E$  is minimum (equal to 0) when the trees disagree and is maximum 1 when they agree, so  $ff_E \in [01]$ .

## 5 Conclusion

Inference of classifiers from examples is an old but still active research field in machine learning community. Classification methods, particularly those based on decision trees, are of major interest given their application results obtained in different fields. Their major point, compared to any other classification method, resides in their intelligibility; they produce ranking functions that make sense of themselves. In addition, the methods have good prediction and generalization performance. However, these methods mainly suffer from the drawbacks of the generated models complexity and instability. Indeed, the complex models make these methods lose their property of interpretability which makes them the most widespread methods in the classification field. Instability reduces the credibility of the tool used which makes it highly dependent on the data.

Among the proposed measures segmentation variables selection, the new information measure (NIM) [22] is less complex than the information theory or distance measurements. NIM used in a greedy partitioning algorithm Induction of Decision Tree New Information Measure (IDT\_NIM) allows generating trees of reduced sizes with similar or even superior performances. For homogeneous or heterogeneous ensemble selection, diversity and/or performance-based [29, 30, 42] functions are used in hill climbing and algorithm genetic. The obtained sub-ensembles are smaller in size and more efficient than the initial ensemble.

Throughout this chapter, we have underlined several points of deepening and future work. First, NIM measurement can be used in sensitive areas where there is a class imbalance. Applications in these areas are very frequent where the imbalance resulting in the data which are scarce but critical may lead to serious economic and strategic consequences in case affectation error, for example, to diagnose a subject as healthy while suffering from cancer. The decentering proposed in [43] also may be used to favor scarce cases in a learning sample. As for the multi-objective function and the entropy function, they can be used in a random forest selection knowing that a random forest ensemble improves the performance of a bagging [44].

## References

1. Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning. *Nat Methods* 15(4)
2. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group
3. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann
4. Kodratoff Y (1998) Technique et outils de l'extraction de connaissances à partir de données. Université Paris-Sud, Revue SIGNAUX (92)
5. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: 5th annual workshop on computational learning theory. ACM, Pittsburgh, pp 144–152
6. Kim J, Pearl J (1987) Convice; a conversational inference consolidation engine. *IEEE Trans Syst Man Cybern* 17:120–132
7. Sebag M (2001) Apprentissage automatique, quelques acquis, tendances et défis. L.M.S: Ecole Polytechnique
8. Denis F, Gilleron R (1996) Notes de cours sur l'apprentissage automatique. Université de Lille
9. Kodratoff Y (1997) L'extraction de connaissance à partir de données: un nouveau sujet pour la recherche scientifique. Revue électronique READ
10. Simon H (1983) Why should machines learn? In: Machine learning: an artificial intelligence approach, vol 1
11. Carbonell JG (1962) Learning by analogy: formulating and generalizing plans from past experience. In: Michalak RS, Carbonell JG, Mitchell TM (eds) Machine learning, an artificial intelligence approach. Tioga Press, Palo Alto, CA
12. Langley P, Simon HA (1995) Applications of machine learning and rule induction. Technical Report 95-1, Institute for the Study of Learning and Expertise
13. Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106
14. Denis F, Gilleron R (1997) Apprentissage à partir d'exemples. Université Charles de Gaulle, Lille 3
15. Dayan P, Sahani M, Deback G (1999) Unsupervised learning. In: Wilson RA, Keil F (eds) The MIT encyclopedia of the cognitive sciences
16. Mitchell T (1997) Machine learning. McGraw-Hill Publishing Company, McGraw-Hill Series in Computer Science (Artificial Intelligence)
17. Taleb Zouggar S, Adla A (2013) On generating and simplifying decision trees using tree automata models. *INFOCOMP J* 12(2):32–43
18. Morgan JN, Sonquist JA (1963) Problems in the analysis of survey data, and a proposal. *J Am Stat Assoc* 58:415–434
19. Kass G (1980) An exploratory technique for investigating large quantities of categorical data. *Appl Stat* 29(2):119–127
20. Friedman JH (1977) A recursive partitioning decision rule for non parametric classification. *IEEE Trans Comput* 26(4):404–408
21. Partalas I, Tsoumakas G, Vlahavas I (2012) A study on greedy algorithms for ensemble pruning. Technical Report TR-LPIS-360-12, LPIS, Dept. of Informatics, Aristotle University of Thessaloniki, Greece
22. Taleb Zouggar S, Adla A (2017) Proposal for measuring quality of decision trees partition. *Int J Decis Support Syst Technol* 9(4):16–36
23. Breiman L (1996) Heuristics of instability and stabilization in model selection. *Ann Stat* 24(6):2350–2383
24. Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In: The 2nd European conference, EuroCOLT '95. Springer-Verlag, pp 23–37
25. Breiman L (2000) Randomizing outputs to increase prediction accuracy. *Mach Learn* 40:229–242
26. Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844

27. Wolpert D (1992) Stacked generalization. *Neural Netw* 5:241–259
28. Lewis-Beck MS, Bryman A, Liao TF (2004) Multi-strategy research. In: The SAGE encyclopedia of social science research methods
29. Partalas I, Tsoumakas G, Vlahavas I (2010) An ensemble uncertainty aware measure for directed hill climbing ensemble pruning. *Mach Learn* 81:257–282
30. Tsoumakas G, Partalas I, Vlahavas I (2009) An ensemble pruning primer. In: Okun, Valentino (eds) Applications of supervised and unsupervised ensemble methods. Springer-Verlag, pp 1–13
31. Margineantu DD, Dietterich TG (1997) Pruning adaptive boosting. In: The 14th international conference on machine learning. Morgan Kaufmann, San Francisco, pp 211–218
32. Yang Y, Korb K, Ting K, Webb G (2005) Ensemble selection for superparent-one-dependence estimators. In: AI 2005: advances in artificial intelligence, pp 102–112
33. Martínez-Muñoz G, Suárez A (2006) Pruning in ordered bagging ensembles. In: 23rd international conference in machine learning (ICML-2006). ACM Press, New York, pp 609–616
34. Bakker B, Heskes T (2003) Clustering ensembles of neural network models. *Neural Netw* 16(2):261–269
35. Fu Q, Hu SX, Zhao SY (2005) Clusterin-based selective neural network ensemble. *J Zhejiang Univ Sci* 6(5):387–392
36. Zhou ZH, Tang W (2003) Selective ensemble of decision trees. In: 9th International conference on rough sets, fuzzy sets, data mining, and granular computing. Chongqing, China, pp 476–483
37. Zhang Y, Burer S, Street WN (2006) Ensemble pruning via semi-definite programming. *J Mach Learn Res* 7:1315–1338
38. Partalas I, Tsoumakas G, Vlahavas I (2012) A study on greedy algorithms for ensemble pruning. Technical Report TR-LPIS-360-12, LPIS, Aristotle University of Thessaloniki, Greece
39. Taleb Zouggar S, Adla A (2018) A diversity-accuracy measure for homogenous ensemble selection. *Int J Interact Multimedia Artif Intell (IJIMAI)*
40. Taleb Zouggar S, Adla A (2018) A new function for ensemble pruning. In Dargam F, Delias P, Linden I, Mareschal B (eds) 4th International conference, ICDSST 2018, Heraklion, Greece, May 22–25, 2018, Proceedings. Decision support systems VIII: sustainable data-driven and evidence-based decision support, LNBP. Springer International Publishing AG
41. Kuncheva LI, Whitaker CJ (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach Learn* 51:181–207
42. Taleb Zouggar S, Adla A (2018) EMnGA: entropy measure and genetic algorithms based method for heterogeneous ensembles selection. *IDEAL* 2:271–279
43. Lallich S, Lenca P, Vaillant B (2007) Construction of an off-centered entropy for supervised learning. In ASMDA, 8
44. Breiman L (2001) Random forests. *Mach Learn* 45:5–32

# Chapter 4

## A Package Including Pre-processing, Feature Extraction, Feature Reduction, and Classification for MRI Classification



Alireza Balavand and Ali Husseinzadeh Kashan

### 1 Introduction

The classification of medical images is a diagnostic technique and pattern that classify different images based on some similar measurements in different categories. The identification of the type of tumor in abnormal brain images is considered as one of the important uses of the classification. Manual diagnosis of brain tumor tissues is time-consuming, due to the complexity of brain tissue, and it depends on the operator's condition. Also, there is a need for experts to examine the images to diagnose, which lead to the inefficiency of the common and old methods in the absence of these people. Therefore, the use of automatic methods will be very useful for the examination of tumors in a precise manner. Nowadays, the use of MRI images has attracted a lot of attention due to the simpler analysis to determine the tumor and its characteristics [1]. Relevant MRI images are usually used as proton density (PD), T1-Weighted, T2-Weighted, and FLAIR [2]. T2-W images have higher weights, denser textures, and their color tends to be white. This property causes cancer tissues are more easily detected because we will have more cell density due to the growth of cancer cells in the target area.

In the field of tumor diagnosis with computer-aided design (CAD), different classification algorithms have been created in MRI images, and different results have been obtained [3]. The methods of the classification MRI images can be divided into two categories of traditional methods and deep learning methods. In general, the steps involved in these algorithms can be divided into pre-processing, feature

---

A. Balavand

Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

e-mail: [a.balavand@srbiau.ac.ir](mailto:a.balavand@srbiau.ac.ir)

A. Husseinzadeh Kashan (✉)

Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran  
e-mail: [a.kashan@modares.ac.ir](mailto:a.kashan@modares.ac.ir)

extraction, dimension reduction, and classification. Pre-processing involves steps such as noise reduction, intensity values correction. A series of features are extracted from the image in the feature extraction method. These features usually include static features such as entropy, skewness, mean, energy, torque, correlation, etc., or the features derived from the application of other algorithms (Fourier transform, histogram, etc.). In the dimension reduction methods, the optimal effective features that make possible achieving the highest percentage of accuracy in sample detection are selected from the features obtained. Usually, in the classification method, the training features along with the classes are trained by supervised artificial intelligence methods, and prediction on classes is done in the test data. Some of the methods of data pre-processing, feature extraction, dimension reduction, and classification algorithms are examined in the literature review section.

In this study, a new package for classification is introduced for the classification of brain tumors in MRI images. This package includes four main steps, including data pre-processing, feature extraction, dimension reduction, and classification. The histogram equalization technique is used to pre-process the MRI images. The GLCM and GoogleNet techniques are considered for feature extraction, and the PCA technique is used for dimension reduction on GoogleNet features. In this study, the OVO-MV algorithm is used for classification. The classification using the OVO-MV algorithm includes two phases. In the first phase, binary classification is performed in which the data class is divided into  $(c \times (c - 1))/2$  binary subsets that the number of classes is  $c$ . Seven classification algorithms are used in heterogeneous groups to classify each binary subset. These classifiers including seven classifiers of decision tree (DT), K-nearest neighbors (K-NN), linear discriminant analysis (LDA), logistic regression (LR), Naive Bayes (NB), support vector machine (SVM), and SVM with radial basis function-based kernel (SVM-RBF). In the second phase, the final classes are calculated using the majority vote of classifiers. In this study, the three-fold cross-validation method is used to divide the training and test data.

## 2 Literature Review

There are various methods for pre-processing in MRI images. The intensity normalization is one of the fields that have many applications in the pre-processing of MRI images, which are essential for the analysis of quantitative textures and improving the contrast of the images. Six methods of intensity scaling, contrast stretch normalization, histogram normalization, histogram stretching, histogram equalization, Gaussian kernel normalization are introduced in [4] which is associated with the intensity normalization field. According to the results of [5], the histogram normalization method has better performance than other methods. In this method, the intensity values are generated based on the application of histogram normalization methods in the original images. But [5] says histogram equalization is more successful in medical images because it obliterates the small details.

Feature extraction is defined as the process of converting an image into a group of features. In recent years, the traditional method and deep learning method are used to feature extraction. A variety of methods, such as GLCM and DWT, can be pointed out in the field of the traditional method of feature extraction. GLCM uses the spatial relationship of two pixels to evaluate textures. The feature extraction was first introduced for the first time using the GLCM in [6]. In this study, they introduced 14 statistical techniques, such as contrast, entropy, sum variance, the sum of squares, and so on. Also, three new statistical features were introduced in [7] to increase the efficiency of the GLCM. Also, DWT is used to extract the feature in MRI images, which is a powerful tool based on mathematics [8], and it generates many features that analysis of all these features leads to increase computing time. In the field of deep learning, the pre-trained convolutional neural networks are used as feature extractors in recent years [9]. The various pre-trained models have been proposed which the most famous models including AlexNet [10] and GoogleNet [11]. Feature reduction techniques are used to reduce these features without losing important information. The various methods have been introduced to reduce the feature in the paper [12]. These methods include Pearson's correlation coefficients (PCC), principal component analysis (PCA), and independent component analysis (ICA). These methods have very little effect on classification [13].

Classification methods include supervised and unsupervised methods [3]. Some of the supervised methods include: decision tree (DT) [14], K-nearest neighbors (K-NN) [15], linear discriminant analysis (LDA) [16], logistic regression (LR) [17], Naive Bayes (NB) [18], support vector machine (SVM) [19], and SVM with radial basis function-based kernel (SVM-RBF) [20]. These methods have been used as a classification tools in many articles. According to the literature review, there are widespread usage of the artificial neural networks as supervised methods for various types of classification. Some of the most important of these are the generalized regression neural network (GRNN) [21], the probabilistic neural network (PNN) [22], the radial basis function (RBF) [23], and the back propagation neural network (BPNN) [24]. There are some problems in using these classification algorithms, in which one of the most important ones is not using some of these algorithms directly to classify several classes [25]. Using the ensemble methods with binary techniques such as decomposition method is considered as one of the solutions proposed in recent years to solve this problem. One of the decomposition methods is the One-vs-One (OVO) method [26], in which multi-class data is divided into the maximum binary subsets of classes, and the subsets are classified by binary algorithms. Finally, the majority vote of the classifiers is used to predict the final classes. Using the binary subsets of classes from two aspects can lead to improving classifications. First, it reduces the complexity of the classification. Because, the multi-class classification has more complexity due to the boundaries between classes, and in the second case, usually it increases the accuracy of classification [27]. In fact, a classification algorithm provides the most accurate classification, but this does not mean that it can provide the highest accuracy in each subset of classes related to the problem [28].

**Table 1** Comparison of the performance of classification methods

Pre-processing	Algorithm	Year
Histogram equalization, binarization, morphological operations	GLCM + GA + fuzzy rough set (20 features reduced to 7) + ANFIS [29]	1990
NA	DWT + SVM with linear kernel [20]	2006
NA	DWT + SVM with polynomial kernel [20]	2006
NA	DWT + SVM with radial basis function-based kernel [20]	2006
NA	DWT + PCA + K-NN [30]	2010
Median filter, pulse-coupled	GLCM + PCA + SVM [31]	2011
NA	DWT + PCA + ANN [32]	2011
NA	(DWT + spider web plot) + PNN [33]	2013
Median filtering, unsharp masking, histogram equalization, FLIRT	(Feature ranking using information gain, feature selection using ICA, extraction using Haar wavelet PCA, GA, 2D and 3D feature) + (SVM, ANN, K-NN) [34]	2013
Artifact removal and noise reduction	(Histogram based features + GLCM) + ANFIS [35]	2014
Gaussian filter	GLCM + decision tree algorithm [36]	2015
NA	(AlexNet, CoffeeNet, VGG-F) + gain ratio + SVM [9]	2018

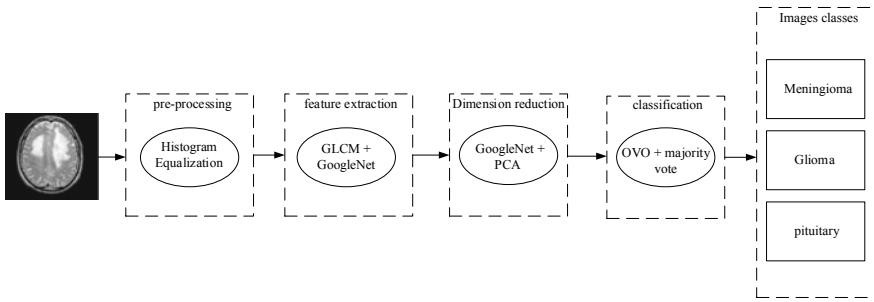
GA—Genetic Algorithm; DWT—Discrete Wavelet Transform; SVM—Support Vector Machine; PCA—Principal Component Analysis; ANN—Artificial Neural Network; K-NN—K-Nearest Neighbor; PNN—Probabilistic Neural Network; GLCM—Gray Level Co-occurrence Matrix; ICA—Independent Component Analysis; ANFIS—Adaptive Neuro-Fuzzy Inference System; NA—Not Applicable

Combination of pre-processing methods, feature extraction, dimension reduction, and classification algorithm leads to create different algorithms in the classification field. A summary of these algorithms is shown in Table 1.

The proposed algorithm is described in Sect. 3. In Sect. 4, the results of using the proposed method are presented, as well as the OVO-MV functions are examined, and the conclusion is presented in Sect. 5.

### 3 The Proposed Algorithm

In this section, a new algorithm is proposed for classifying brain tumors in MRI images, in which its flowchart is based on Fig. 1. This algorithm includes four main steps. In the first step pre-processing is performed on the images using the histogram equalization technique. In the second step, seven features using a GLCM technique, and 1000 features are extracted from MRI images using the GoogleNet technique,

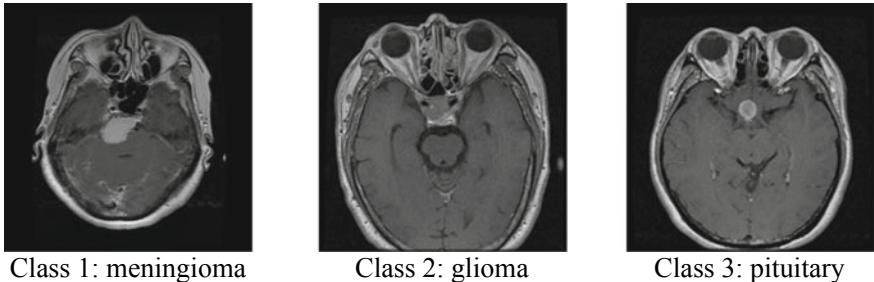


**Fig. 1** Proposed algorithm

and in the third step, due to the creation of many features by the GoogleNet method and the creation of high computational complexity in the classification, using the PCA technique, dimension reduction is performed in GoogleNet features, in which finally 100 important features of DWT are identified. Finally, in the fourth step, the OVO-MV method will be used for classification. In this algorithm, classes are divided into a maximum binary subset, and each binary subset is predicted by the majority vote of seven classification algorithms. In this study, the k-fold cross-validation method is used for dividing data into training and test data, and MSE is also used to calculate the error of classification. This study is carried out to increase the accuracy and reduce the prediction error in the classification of brain tumors in MRI images, which increased accuracy depends on these two factors. The first factor is the feature extraction, in which the use of appropriate feature extraction methods can have a great impact on classification accuracy, and the second factor is the use of an appropriate classifier. The combination of the OVO-MV classification algorithm and GoogleNet features can lead to an increase in the accuracy of the classification and reducing classification error in comparison with the single classifier, which is the main difference between our method and the state-of-the-art methods. There are three hypotheses: in the first hypothesis, it is expected that due to the efficiency of the GoogleNet feature, this technique leads to generate suitable features. In the second hypothesis, it is predicted that using the OVO-MV method can lead to an increase in the accuracy of the classification and reducing classification error in comparison with the single classifier, and in the third hypothesis, it is expected that no single classifier can have good results for all data. In the following, at first, the MRI images are described. Then histogram equalization and GLCM, GoogleNet methods with PCA methods are described. Finally, the OVO-MV algorithm is presented.

### 3.1 MRI Images

In this study, 900 MRI images in the form of T1-W have been collected from the southern medical university of Guangzhou website [37] to create a valid database.



**Fig. 2** Types of brain tumors

Three kinds of brain tumor of meningioma (900 slices), glioma (900 slices), and pituitary tumor (900 slices) have been detected in these images. An example of MRI images, as well as the type of tumor with their classes, is shown in Fig. 2. All images are resized to  $227 \times 227$  in size.

### 3.2 Histogram Equalization

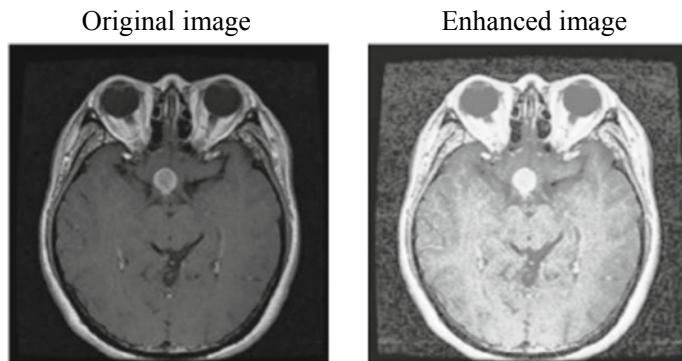
Histogram equalization is used for doing the adjustment process of the intensity values automatically. In this method, the histogram of the output image becomes uniform, and the image contrast will be increased as much as possible. Histogram equalization is calculated according to Eq. (1) for each pixel:

$$h(v) = \text{round}\left(\frac{\text{cdf}(v) - \text{cdf}_{\min}}{(w \times h) - 1}\right) \times (L - 1) \quad (1)$$

where,  $h(v)$  is the value of the histogram,  $\text{cdf}(v)$  is the value of the cumulative distribution function related to the pixel  $v$ ,  $\text{cdf}_{\min}$  is the minimum value of the cumulative distribution function,  $w$  is the image width,  $h$  is the image height, and  $L$  is the number of gray levels used which in most cases is 256. In Fig. 3, the left image represents the original image, and the right image is created after using the histogram equalization method. At this stage, the pre-processing operations are done using the histogram equalization method on all MRI images.

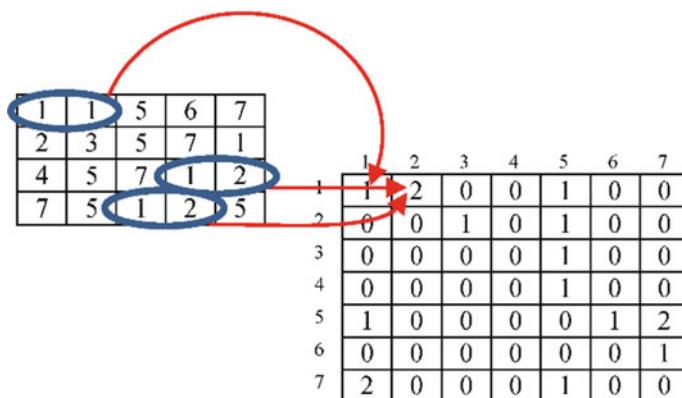
### 3.3 GLCM

GLCM uses the second-order statistical textural features. A GLCM is a matrix in which its rows and columns are equal to the number of gray levels in the used image. It means, if the number of degrees of gray in an image is  $G$ , then the dimension of the



**Fig. 3** Pre-processing by histogram equalization method

GLCM matrix is equal to a  $G \times G$  matrix. The GLCM matrix is created in accordance with Fig. 4. According to this figure, the left matrix indicates a  $4 \times 5$  image, and the right-hand matrix is transformed into a  $7 \times 7$  matrix using the co-occurrence matrix transform. The number 1 in row 1 and column 1 in the right matrix represents the number of repetitions of two numbers 1 together in the left matrix. Accordingly, number 2 on the right matrix indicates the number of repetitions of two numbers 1 and 2 together in the left matrix. At this step, using the GLCM, seven features are extracted from the MRI images that are shown, in Table 2.



**Fig. 4** Conversion of the co-occurrence matrix

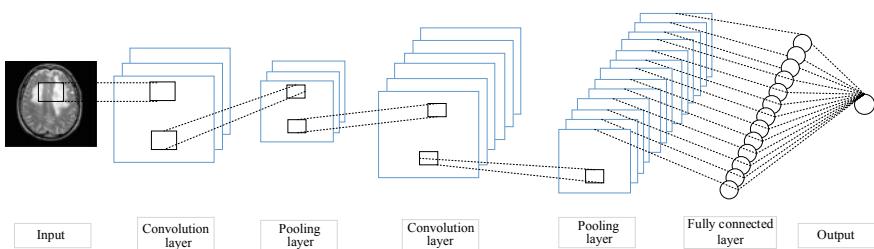
**Table 2** GLCM features equations

Features	Formula
Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j)  i - j  = n \right\}$
Correlation 1	$\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_{xy}$
Correlation 2	$\frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_{xy}}{\sigma_x \sigma_y}$
Dissimilarity	$\sum_i \sum_j  i - j  \cdot p(i, j)$
Energy	$\sum_i \sum_j p(i, j)^2$
Entropy	$-\sum_i \sum_j p(i, j) \log p(i, j)$
Homogeneity	$\sum_i \sum_j \frac{1}{1-(i-j)^2} p(i, j)$

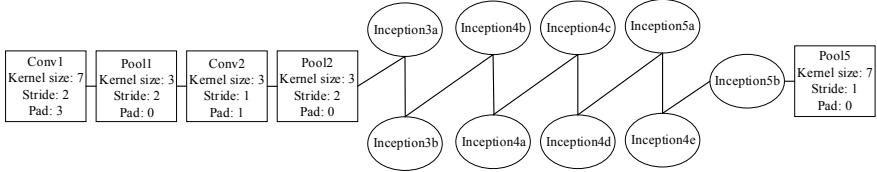
$p(i, j)$  :  $(i, j)$ th entry in a normalized gray-tones spatial-dependence matrix

### 3.4 Pre-trained Convolutional Neural Network

Nowadays, researchers that work with artificial intelligence use deep learning for creating powerful computing systems. Meanwhile, convolutional neural networks are used to feature extraction and classification [38, 39]. The purpose of the design of convolutional neural networks is detailed modeling of how the human visual system works. A convolutional neural network is the kind of deep learning which contains a large number of convolution and pooling layers. The input of the convolutional neural network is usually an image, and its output is a feature vector with high resolution and corresponding to one class. Hidden layers in the convolutional neural network include convolution layer, pooling layer, and fully connected layer [40]. The simple structure of the convolutional neural network is shown in Fig. 5. Convolution layer includes educable weights and biases which in the form of filters with different dimensions and depths are applied on input layers, and a feature map for each sample and filter is created. Connecting these feature maps to each other forms convolution layer. Pooling layer is a nonlinear sampling function along with scaling down which can be a function as maximizing, averaging, and even least square norm. Applying this layer to the input layer causes the input layer dimension to decrease gradually.



**Fig. 5** Simple structure of the convolutional neural network



**Fig. 6** Structure of the GoogleNet model

The fully connected layer is a final layer with high-level features, and each neuron in this layer connects with one of the feature maps in the previous layer.

There are two methods to use convolutional neural networks. In the first method, the process of training is done by using a large data set, and in the second method, the pre-trained convolutional neural networks are used to feature extraction [9]. In this study, a pre-trained method is used for feature extraction called GoogleNet. The GoogleNet is proposed in [11]. In this model, a new concept called inception is proposed. Each inception includes six convolutional layers and two pooling layers. Based on Fig. 6, this model includes two convolutional layers, three pooling layers, and nine inception layers.

### 3.5 PCA Technique

Analysis of the principal components is defined as an orthogonal linear transformation that takes data to a new coordinate system, so that the largest data variance be on the first coordinate axis, the second-largest variance be on the second coordinate axis, and so on. Principal components analysis can be used to reduce data dimensions, thus preserves components of the data set that have the greatest impact on variance. In order to examine the PCA technique, assume that there are  $P$  variables. The new linear composition of these  $P$  variables is based on Eq. (2) [41].

$$\begin{aligned}\epsilon_1 &= w_{11}x_1 + w_{12}x_2 + \cdots + w_{1j}x_p \\ \epsilon_2 &= w_{21}x_1 + w_{22}x_2 + \cdots + w_{2j}x_p \\ &\vdots \\ &\vdots \\ \epsilon_p &= w_{p1}x_1 + w_{p2}x_2 + \cdots + w_{pj}x_p\end{aligned}\tag{2}$$

where  $\epsilon_1, \epsilon_2$  and  $\epsilon_p$ , is the  $P$  principal component,  $w_{ij}$  is the weight of the variable  $j$  for the  $i$ th component, and  $x_p$  represents the variable p. In this study, the PCA technique is used to reduce the features of the GoogleNet technique, as well as to reduce the dependency and identify the principal features. The variance-covariance matrix is used for the principal component analysis to identify the principal features.

Using this technique, a total of 100 principal features of the GoogleNet features have been identified, in which these 100 features contain 84% of the variance.

### 3.6 Classification

In data with a large number of classes or large numbers of dimensions, classification algorithms may show different performance and provide different accuracy. This is because any classification algorithm cannot be the best option for solving all classification problems. In other words, a classifier algorithm provides a satisfactory solution in a particular problem, and the same algorithm may have poor performance in other problems. A new combination of OVO method and the majority vote of classifiers can be used to reduce this weakness. This algorithm is summarized in two main phases. In the first phase, the three-fold cross-validation method is used to divide the training and test data, and then, the classes are divided into maximum binary subsets, and each subset is classified by several heterogeneous algorithms. Also, prediction operations on classes are performed for each pair of classes in this phase. In the second phase, the majority vote method is used to increase the accuracy of the classification, and the error of the classifiers is calculated using the MSE method. Each of the phases will be addressed in the following.

#### 3.6.1 OVO-MV

Considering the pseudo-code shown in Fig. 7, in the first line, separating training and test data is performed by k-fold. In the following, the process of training is done for each classifier and each binary class ( $i,j$ ) in line six. Then, prediction on test data is done by the created model related to line six and test data. This process repeats for all classifiers of DT, K-NN, LDA, LR, NB, SVM, and SVM-RBF in each binary class. The process of majority vote is done from line 11 to line 19. In order to see how the calculation of the majority vote is done, we refer to Table 3. According to Table 3, we assume that the binary class (1,2) has four rows. Prediction of classifiers has been done from column two to column seven. The last column is obtained based on the majority vote of the columns of two to seven. For example, in the first row, all classifiers have predicted class 1. Therefore, the first element of the last column is equal to number one. Also in the last row, four classifiers have predicted class two, and three classifiers have predicted class one. Therefore, the last element of the last column is equal to number two. Also, we use three-fold cross-validation for dividing data to training and test data. In this method, all data are considered as a test once. The MSE is used to calculate the prediction error for all of the classifiers in each pair class, which is obtained based on the Eq. (3). In this equation,  $\hat{y}_i$  is the predicted class, and  $y_i$  is the real class. The average of MSE in three folds is considered as a final error of classification. In the last row of Table 3, the calculation of MSE has been shown. This row is obtained based on the difference between the real label (first

---

❖ First phase: binary classification

```

1 running k-fold with k=3
2 separating training and test data
3 for k=1 to 3
4   for each binary classes (i,j)
5     for each classifier (c)
6       model → training classifier(c) by training data in binary classes (i,j)
7       prediction → doing prediction of classifier(c) algorithm using model and test data in binary classes (i,j)
8       save predicted label of binary classes (i,j)
9   end for line 5
10 end for line 6

```

❖ Second phase: majority vote

```

11  for each binary classes (i,j) in fold k
12    for each row in binary classes (i,j)
13      if the number of i > number of j
14        row of the majority vote is equal i
15      if the number of i < number of j
16        row of the majority vote is equal j
17    end if
18  end for line 12
19 end for line 11
20 end for line 3
21 report the average of MSE in each binary class and predicted the labels

```

---

**Fig. 7** Majority vote function

**Table 3** Majority vote method and MSE calculation for binary classes (1,2)

Real classes	DT	K-NN	LDA	LR	NB	SVM	SVM-RBF	Majority vote
1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1
2	2	1	2	2	2	1	2	2
2	2	1	2	2	1	1	2	2
<b>MSE</b>	<b>0</b>	<b>0.5</b>	<b>0</b>	<b>0</b>	<b>0.25</b>	<b>0.5</b>	<b>0</b>	<b>0</b>

column) and predicted label of each classifier. In Table 3, classifiers of DT, LDA, LR, SVM-RBF, and majority vote have good performance, and all labels are predicted correctly.

$$\text{MSE} = \sum_{n=1}^c \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (3)$$

## 4 Data Analysis

In this section features of the GLCM and GoogleNet are classified using the OVO-MV algorithm to evaluate the proposed algorithm. Given that the OVO-MV algorithm uses the majority vote of the seven classification algorithms, the parameters of each classification algorithm are set according to Table 4.

### 4.1 MRI Images

This data includes seven extracted features from the GLCM and 100 important features from the GoogleNet, which are classified into three classes. Given that the GLCM features extract seven features of each MRI image, therefore, we have a database with 900 rows and 7 columns. Also, GoogleNet features have created a database with 900 rows and 100 columns. In the following of this section, the classification results of the GLCM and GoogleNet features will be shown and compared to each other.

#### 4.1.1 GLCM

The results of OVO-MV algorithm based on GLCM features are shown in Tables 5, 6, and 7. Given that a total of seven features are extracted from 900 MRI images using techniques GLCM, a database with 900 rows and seven columns is obtained that rows show the number of images, and columns indicate the features that are considered as inputs to the classification algorithm. Given that this data has three classes, therefore, the binary classes are equal to three classes. MSEs of classifiers in fold 1 in Table 5

**Table 4** Parameter setting for each classification algorithm

Algorithm	Parameter setting
DT	Minimum number of leaf = 1 Prune = true
K-NN	K = 1 Distance type = Euclidean
LDA	No parameter
LR	No parameter
NB	Kernel = normal
SVM	C = 1.0 Kernel type = linear Epsilon = 1.0E-12
SVM-RBF	C = 1.0 Kernel type = RBF Epsilon = 1.0E-12

**Table 5** MSEs of classifiers in binary subsets in fold 1 for GLCM features

Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority vote
(1,2)	0	0	0	0	0	0.5	0.125	0
(1,3)	0	0.5	0	2	0	2	0.5	0
(2,3)	0	0.125	0	0.5	0.125	0.5	0.4375	0.125
Average	0	0.2	0	0.83	0.04	1	0.35	0.04

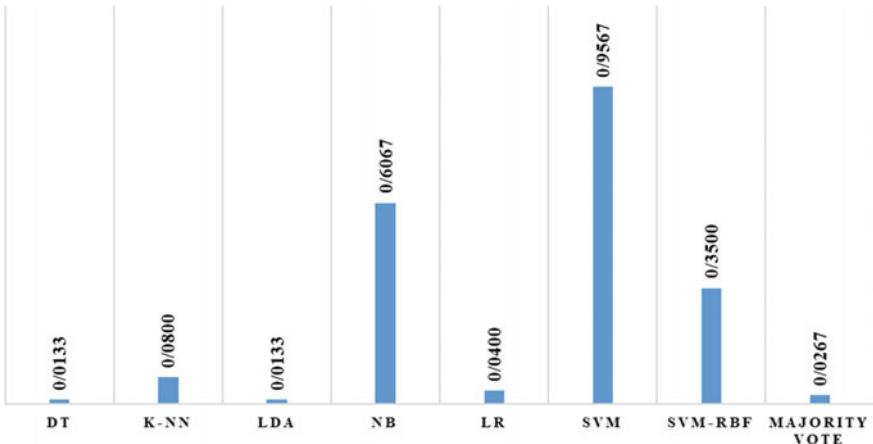
**Table 6** MSEs of classifiers in binary subsets in fold 2 for GLCM features

Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority vote
(1,2)	0	0	0	0	0	0.5	0.125	0
(1,3)	0.125	0	0	0.5	0.125	2	0.5	0
(2,3)	0	0.125	0	0.5	0.125	0.5	0.4375	0.125
Average	0.04	0.04	0	0.33	0.08	1	0.35	0.04

**Table 7** MSEs of classifiers in binary subsets in fold 3 for GLCM features

Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority vote
(1,2)	0	0	0	0	0	0.5	0.125	0
(1,3)	0	0	0	2	0	2	0.5	0
(2,3)	0	0	0.125	0	0	0.125	0.4375	0
Average	0	0	0.04	0.66	0	0.87	0.35	0

shows that among the classifiers, LDA and DT have the minimum, and SVM has the maximum average of MSE. The last column shows MSE of the majority vote of classifiers. Results of MSE in majority vote show that there is only classification error in binary class (2,3). Average of MSE in majority vote is reported 0.04 which shows that majority vote has appropriate performance in the classification. Table 6 shows the MSEs of classifiers in fold 2. Average of classifiers error in the last row shows that the best performance and the worst performance are related to LDA and SVM, respectively. The error of the majority vote in the last column is 0.04 which shows that most classifiers have had good performance. MSEs of classifiers in fold 3 present in Table 7. DT, K-NN, LR, and majority vote have had the best performance. Also, SVM has not had good performance compared with other algorithms. The average of MSEs of last rows in Tables 5, 6, and 7 is shown in Fig. 8. In this figure, LDA, DT, majority vote, LR, K-NN, SVM-RBF, NB, and SVM have had minimum MSE error, respectively.



**Fig. 8** Comparison of the average of MSEs in three folds in classifiers for GLCM features

#### 4.1.2 GoogleNet

The results of OVO-MV algorithm based on GoogleNet features are shown in Tables 8, 9, and 10. 1000 features have been extracted from each MRI image by using the GoogleNet method. These features have been decreased by the PCA technique. Therefore, by using the PCA, 100 important features have been identified in which a database with 900 rows and 100 columns is created. The results of the MSEs of classifiers in fold 1 are shown in Table 8. DT, K-NN, LDA, LR, and majority vote have had the best performance in classification. Other classifiers have also had

**Table 8** Error of classification in binary subsets in fold 1 for GoogleNet features

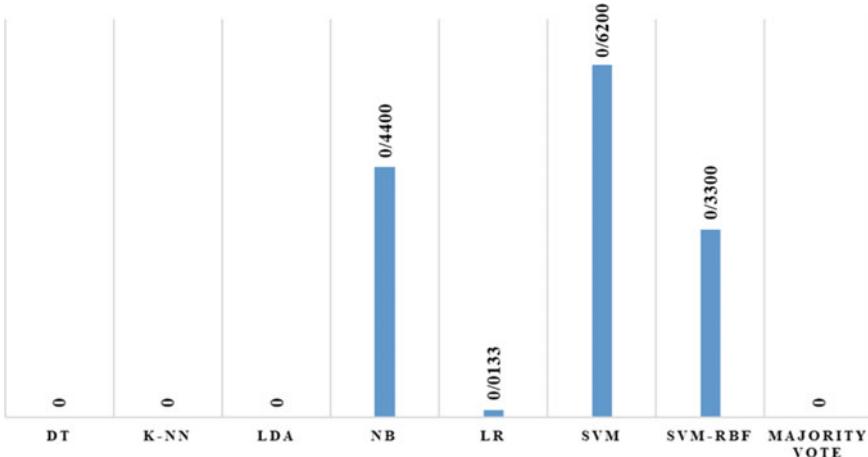
Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority vote
(1,2)	0	0	0	0	0	0.5	0	0
(1,3)	0	0	0	1	0	1	0.5	0
(2,3)	0	0	0	0.5	0	0.5	0.5	0
Average	0	0	0	0.5	0	0.66	0.33	0

**Table 9** Error of classification in binary subsets in fold 2 for GoogleNet features

Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority Vote
(1,2)	0	0	0	0	0	0.5	0	0
(1,3)	0	0	0	0	0.125	1	0.5	0
(2,3)	0	0	0	0.5	0	0.5	0.5	0
Average	0	0	0	0.16	0.04	0.66	0.33	0

**Table 10** Error of classification in binary subsets in fold 3 for GoogleNet features

Binary subsets	DT	K-NN	LDA	NB	LR	SVM	SVM-RBF	Majority vote
(1,2)	0	0	0	0	0	0.5	0	0
(1,3)	0	0	0	2	0	1	0.5	0
(2,3)	0	0	0	0	0	0.125	0.5	0
Average	0	0	0	0.66	0	0.54	0.33	0

**Fig. 9** Comparison of the average of MSEs in three folds in classifiers for GLCM features

low MSEs. This is also evident in Table 9. In this table, most classifiers have good performance in fold 2. The MSEs of classifiers in Table 10 show that GoogleNet features have good quality and cause the proper separation of classes. The average of MSEs of last rows in Tables 8, 9, and 10 is shown in Fig. 9. Minimum MSE in this figure is related to DT, K-NN, LDA, and majority vote, and maximum MSE is related to SVM. Also, LR classifier has had good performance.

## 5 Discussion and Conclusion

This study introduced a new algorithm for classifying brain tumors in MRI images, including 900 MRI images. Four steps including pre-processing, feature extraction, dimension reduction, and classification using the OVO-MV algorithm were defined in order to classify MRI images. In the first step, the pre-processing operations were performed on the images using the histogram equalization method. In the second step, seven features using the GLCM method and 100 features were extracted using

the GoogleNet method, in which the PCA method was used to reduce the dimensions and dependence due to having many features using the GoogleNet method, and finally, 100 main features are identified from the GoogleNet features. In the fourth step, the OVO-MV algorithm with two phases was introduced. In the first phase, the three-fold cross-validation method is used to divide the training and test data, then, the binary classification was performed in which the data class was divided into maximum binary subsets, and seven classification algorithms were used in heterogeneous groups in order to classify each binary subset. Classification algorithms consisted of seven classifiers of DT, K-NN, LDA, LR, NB, SVM, and SVM-RBF. According to the results, the proposed method achieved the high accuracy in the classification of brain tumors in GoogleNet features which in the classification of GoogleNet features, most of the classifiers were better than GLCM features. Although a highly accurate classification was achieved by OVO-MV algorithm, this method may have some limitations including increasing run time of classification and decreasing classification accuracy in some problems. Also, according to the comparative results, no classification could provide the appropriate results in all data, and in more times, better results can be achieved using the majority vote method. For future works, the clustering phase could be added in current work for segmentation of MRI images. Metaheuristic search methods such as league championship algorithm [42, 43], optics inspired optimization [44, 45], and find-fix-finish-exploit-analyze [46] can be used for clustering method. Also, there are efficient algorithms such as ACDEA [47] for determining the optimum number of the cluster for increasing performance of clustering.

## References

1. Ramakrishnan T, Sankaragomathi B (2017) A professional estimate on the computed tomography brain tumor images using SVM-SMO for classification and MRG-GWO for segmentation. *Pattern Recogn Lett*
2. Zhang N et al (2011) Kernel feature selection to fuse multi-spectral MRI images for brain tumor segmentation. *Comput Vis Image Underst* 115(2):256–269
3. Mohan G, Subashini MM (2018) MRI based medical image analysis: survey on brain tumor grade classification. *Biomed Signal Process Control* 39:139–161
4. Nabizadeh N, Kubat M (2015) Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. *Comput Electr Eng* 45:286–301
5. Loizou CP et al (2009) Brain MR image normalization in texture analysis of multiple sclerosis. In: 9th International Conference on Information technology and applications in biomedicine, 2009. ITAB 2009. IEEE
6. Haralick RM, Shanmugam K (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* 6:610–621
7. Soh L-K, Tsatsoulis C (1999) Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans Geosci Remote Sens* 37(2):780–795
8. Daubechies I (1992) Ten lectures on wavelets. SIAM
9. Vogado LH et al (2018) Leukemia diagnosis in blood slides using transfer learning in CNNs and SVM for classification. *Eng Appl Artif Intell* 72:415–422
10. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Advances in neural information processing systems
11. Szegedy C et al (2015) Going deeper with convolutions. *Cvpr*

12. Zöllner FG, Emblem KE, Schad LR (2012) SVM-based glioma grading: optimization by feature reduction analysis. *Zeitschrift für medizinische Physik* 22(3):205–214
13. Dash M, Liu H (1997) Feature selection for classification. *Intel Data Anal* 1(1–4):131–156
14. Coppersmith D, Hong SJ, Hosking JR (1999) Partitioning nominal attributes in decision trees. *Data Min Knowl Disc* 3(2):197–217
15. Fletcher-Heath LM et al (2001) Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artif Intell Med* 21(1):43–63
16. Guo Y, Hastie T, Tibshirani R (2006) Regularized linear discriminant analysis and its application in microarrays. *Biostatistics* 8(1):86–100
17. Freedman DA (2009) Statistical models: theory and practice. Cambridge University Press
18. Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning, vol 1. Springer series in statistics, New York
19. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press
20. Chaplot S, Patnaik L, Jagannathan N (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed Signal Process Control* 1(1):86–92
21. Wasserman PD (1993) Advanced methods in neural computing. Wiley
22. Specht DF (1990) Probabilistic neural networks. *Neural Netw* 3(1):109–118
23. Du K-L, Swamy M (2014) Radial basis function networks. In: Neural networks and statistical learning. Springer, pp 299–335
24. Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
25. Kang S, Cho S, Kang P (2015) Constructing a multi-class classifier using one-against-one approach with different binary classifiers. *Neurocomputing* 149:677–682
26. Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. In: Neurocomputing: algorithms, architectures and applications, vol 68(41–50), p 71
27. Galar M et al (2011) An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn* 44(8):1761–1776
28. Kang S, Cho S (2015) Optimal construction of one-against-one classifier based on meta-learning. *Neurocomputing* 167:459–466
29. Dean BL et al (1990) Gliomas: classification with MR imaging. *Radiology* 174(2):411–415
30. El-Dahshan E-SA, Hosny T, Salem A-BM (2010) Hybrid intelligent techniques for MRI brain images classification. *Digit Signal Proc* 20(2):433–441
31. Marshkole N, Singh BK, Thoke A (2011) Texture and shape based classification of brain tumors using linear vector quantization. *Int J Comput Appl* 30(11):21–23
32. Zhang Y et al (2011) A hybrid method for MRI brain image classification. *Expert Syst Appl* 38(8):10049–10053
33. Saritha M, Joseph KP, Mathew AT (2013) Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recogn Lett* 34(16):2151–2156
34. Ortiz A et al (2013) Improving MRI segmentation with probabilistic GHSOM and multiobjective optimization. *Neurocomputing* 114:118–131
35. Zahran B (2014) Classification of brain tumor using neural network. *Int Rev Comput Softw (IRECOS)* 9(4):673–678
36. Gaikwad SB, Joshi MS (2015) Brain tumor classification using principal component analysis and probabilistic neural network. *Int J Comput Appl* 120(3)
37. *School of Biomedical Engineering.* Jun Cheng: Southern Medical University, Guangzhou, China
38. LeCun Y, Kavukcuoglu K, Farabet C (2010) Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE international symposium on circuits and systems (ISCAS). IEEE
39. Nagi J et al (2011) Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: 2011 IEEE international conference on signal and image processing applications (ICSIPA). IEEE

40. Liu T et al (2015) Implementation of training convolutional neural networks. arXiv preprint [arXiv:1506.01195](https://arxiv.org/abs/1506.01195)
41. Subhash S (1996) Applied multivariate techniques. Wiley, Canada
42. Kashan AH (2011) An efficient algorithm for constrained global optimization and application to mechanical engineering design: League championship algorithm (LCA). Comput Aided Des 43(12):1769–1792
43. Kashan AH (2014) League Championship Algorithm (LCA): an algorithm for global optimization inspired by sport championships. Appl Soft Comput 16:171–200
44. Kashan AH (2015) An effective algorithm for constrained optimization based on optics inspired optimization (OIO). Comput Aided Des 63:52–71
45. Kashan AH (2015) A new metaheuristic for optimization: optics inspired optimization (OIO). Comput Oper Res 55:99–125
46. Kashan AH, Tavakkoli-Moghaddam R, Gen M (2017) A warfare inspired optimization algorithm: the Find-Fix-Finish-Exploit-Analyze (F3EA) metaheuristic algorithm. In: Proceedings of the tenth international conference on management science and engineering management. Springer
47. Balavand A, Kashan AH, Saghaei A (2018) Automatic clustering based on Crow Search Algorithm-Kmeans (CSA-Kmeans) and Data Envelopment Analysis (DEA). Int J Comput Intell Sys 11(1):1322–1337

# Chapter 5

# Predictive Analysis of Lake Water Quality Using an Evolutionary Algorithm



Mrunalini Jadhav, Kanchan Khare, Sayali Apte and Rushikesh Kulkarni

## 1 Introduction

One of the preconditions for the existence of a living organism and the sustainability of the planet earth is water. It plays a crucial role in socio-economic development, ecological sustainability and economic growth. The exponential increase in population has resulted in stress on the limited natural resources, and water is one of this overstressed natural resource. Over 3.6 billion people worldwide are already living in potential water-scarce areas for at least one month per year, and this might increase to 4.8–5.7 billion in 2050. The world economic forum's global risk report 2018 states that among the most pressing environmental challenges dealing with us are extreme climate occasions and temperatures; accelerating biodiversity loss and pollution of air, soil and water [33]. As per the World Water Vision Report, the crisis is no longer about having too little water to satisfy our wants. But the disaster is about suited administration of available water [35]. Lakes are one of the vital sources of fresh water. They can also provide us with prime opportunities for recreation, tourism, and cottage or residential living. They have historical and traditional values and also serve to be a source of raw drinking water for a municipality, industry and an irrigation

---

M. Jadhav  
SVC Polytechnique, Pune, Maharashtra 411041, India  
e-mail: [mrunaljadhav@gmail.com](mailto:mrunaljadhav@gmail.com)

K. Khare (✉) · S. Apte · R. Kulkarni  
Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune,  
Maharashtra 412115, India  
e-mail: [kanchan.khare@sitpune.edu.in](mailto:kanchan.khare@sitpune.edu.in)

S. Apte  
e-mail: [sayali.apte@sitpune.edu.in](mailto:sayali.apte@sitpune.edu.in)

R. Kulkarni  
e-mail: [rushikeshk@sitpune.edu.in](mailto:rushikeshk@sitpune.edu.in)

source for agriculture, and also work to replenish groundwater. Positively influence the water quality of downstream watercourses and prevent flooding. Global water resource situation shows that, out of the whole accessible water, fresh water is solely 3%. Out of this 3%, surface water is 0.3%; out of this 0.3% surface water, 87% water is in natural lakes or artificial reservoirs—11% is in swamps and only 2% is in rivers [30]. It is therefore worthwhile taking efforts to save the water in our lakes.

## 1.1 *Issues and Challenges*

Water quality is described by physiochemical, biological and microbiological parameters that reflect the abiotic and biotic status of the ecosystem. Water quality testing helps to determine the trends in pollutant concentration and their effects on human and aquatic life. It also helps to identify source contribution to pollution, its development and further to decide the controlling strategies.

The leading causes for lake pollution entering from fixed point sources are discharge of nutrients from wastewater from municipal and domestic effluents, discharge of organic, inorganic and toxic pollutant loading due to disposal of industrial effluents and biodegradable wastes, and discharge of storm water run-off. Nutrients through fertilisers, toxic pesticides and other chemicals mainly coming from agricultural run-off, deforestation and denudation in the catchment areas, causing soil erosion and consequent siltation, are the essential causes of non-point source pollution. Organic pollution loading from human settlement spread over a city in the immediate surroundings of the lakes. Organic pollution loading from human settlement spread over an area in the immediate surroundings of the lakes. The other problems of lake pollution are silting of lakes, land disturbances happening in a diversion of rivers which feed the lakes, the drainage basin, cultural siltation in the form of immersion of idols during specific festivals has been a source of severe metallic pollution of lakes [1].

Throughout the world, the water quality of lakes, natural or human-made, has been deteriorating because of these urban, agricultural, industrial and other impacts. Widespread eutrophication of lakes leads to the overgrowth of plants and algae; the bacterial degradation of their biomass consumes more oxygen from water resulting in a state of hypoxia [1–3]. The reason responsible for the growth of algae is phosphate, which causes a severe reduction in water quality. During the last fifty years, the demand for scientific and sustainable management of lakes that includes prevention and restoration. Substantial research has been carried out to control and reverse the degradation of lake water. Various methods and techniques have been evolved for lake restoration. For effective water management, monitoring lake water quality for potential use, therefore, has become very vital [13]. Problems of lakes vary, depending upon their morphology, the climate of the catchment, land use in watershed, etc. However, specific issues, which are more or less familiar to most of the lakes, are pollution, water quality deterioration, eutrophication and sedimentation. Therefore, it has become essential to assess the water pollution of these water bodies systematically so that suitable corrective actions could be recommended for conservation.

## 1.2 *Lake Water Quality Assessment and Monitoring*

The quality of water may be delineated in terms of the concentration and dissolved or particulate state of organic and inorganic material present in water. Physical characteristics of water add to this quality assessment. Long-term, standardised measurement of water quality may be termed as monitoring. Monitoring is carried out to estimate nutrient fluxes discharged by rivers or groundwater to lakes and other water bodies. It is also used to check whether any unexpected change is occurring in water quality. Monitoring helps to determine trends in the quality of water or aquatic environment. We can even understand how the quality is affected by the release of contaminants, other anthropogenic activities and by waste treatment operations.

Monitoring of water quality is based on the collection of data. Data collection points are selected at given geographical locations in the water body. Water quality variables are described by the longitude and latitude of the sampling or measurement site ( $x$  and  $y$  coordinates). They are characterised by the depth at which the sample is taken (vertical coordinate  $z$ ). Monitoring data must also be recorded at the time  $t$  at which the sample is taken. Thus,  $c = f(x, y, x, t)$ , where  $c$  is a concentration of any physical, chemical or biological variable. Monitoring data must, therefore, provide a precise determination of these parameters to be used for data interpretation and water quality assessments.

All assessment programs start with scrutinising the real need for water quality information critically since we use water resources to several competing beneficial uses. There are two types of monitoring programs, depending on how many assessment objectives have to be met. Single-objective monitoring is set up to address one problem area only.

This process involves a set of variables, such as pH, alkalinity and some cations for acid rain; nutrients and chlorophyll pigments for eutrophication; various nitrogenous compounds for nitrate pollution; or sodium, calcium, chloride and a few other elements for irrigation. Multi-objective monitoring may cover multiple water uses and provide data for more than one assessment program such as drinking water supply, industrial manufacturing, fisheries or aquatic life, thereby involving a large set of variables. The assessment objectives may focus on the spatial distribution of quality (high station number), on trends (high sampling frequency) or pollutants. Full coverage of all three requirements is virtually impossible, or very costly.

Water quality monitoring can help researchers predict and learn from natural processes in the environment and determine human impacts on an ecosystem. These measurement efforts can also assist in restoration projects or ensure environmental standards are being met. Many researchers have worked on prediction/forecasting of water quality. However, more work needs to be done in terms of effectiveness, reliability, accuracy, as well as usability of the current water quality management methodologies.

## 2 Forecasting of Lake Water Quality Parameters

Modelling and forecasting of water quality parameters involve a variety of approaches. Traditionally, water quality forecast was carried out using hard computing approaches. They include deterministic, stochastic, statistical or numerical models. Mathematical models that are available for the prediction of water quality are plenty in number. These models are complex in the structure and require detailed information about source and receptor, which is a costly and challenging task that leaves a scope to try alternative approaches [34]. Water quality is affected by many factors. Traditional data processing methods are not good enough for solving the problem as such factors show a complicated nonlinear relation to the variables of water quality forecasting [38]. For process-based models, there is a requirement of a lot of input data and model parameters are often unknown and are computationally expensive, while evolutionary algorithms provide an effective alternative to conventional process-based modelling [6, 8, 19, 24]. These models are computationally high speed and require fewer input parameters than process-based models [26, 27, 38].

Many real-life problems do not lend themselves to precise solutions; hence, hard models are insufficient for such issues. On the other hand, evolutionary algorithms are based on the guiding principle of being tolerant of imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness and low solution cost [18]. Researchers, nowadays, have a wealth of data to use for analysis and data mining because there is extensive use of in situ hydrological instrumentation [38]. They are often robust under noisy input environments and have a high tolerance for imprecision in the data on which they operate. Neural networks (NN or ANN) [28], evolutionary computation (EC) [28], model tree (MT) [27] and fuzzy systems (FS) [17] have been used in water quality prediction. Neural networks trained with small data sets often demonstrate unstable behaviour in performance, i.e. random fluctuations due to the sensitivity of neural networks to initial parameter values and training order [14, 16, 31].

But for “data-rich, theory poor” instances, GP may offer advantages over all other techniques since GP can self-modify, through the genetic loop, a population of function trees to finally generate an “optimal” and physically interpretable model [20].

### 2.1 Evolutionary Algorithm: Genetic Programming

There are many meta-heuristic algorithms known today in computer science, including random optimisation, simulated annealing and even greedy algorithm. One of the algorithms is evolutionary algorithms. Evolutionary algorithms are used to discover solutions to problems free of human preconceptions or biases. The adaptive natures of evolutionary algorithms do generate solutions which are comparable to, and often better than the best human efforts. They use mechanisms inspired by biological

evolution, for example reproduction, mutation, recombination and natural selection. Usually, a set of genome described the problem space then-candidate solution is created by the use of various operators like mutation, reproduction and finally, a cost function determines which solution to retain (fitness). These operations are repeated several times, and due to natural selection, candidate solutions improve over time.

An evolutionary algorithm is divided into several categories based on implementation details. One of them is genetic algorithms, one other genetic programming among a few more.

Genetic programming gives solutions which are in the form of a computer program. Ability to solve a computational problem decides the fitness of the program.

The nature-inspired technique of genetic programming (GP) evolves the best individual (program) through the combination of cross-over, mutation and reproduction processes. GP can also be used to discover a functional relationship between features in data (symbolic regression), to group data into categories (classification). It works on the Darwinian principle of “survival of the fittest” [15].

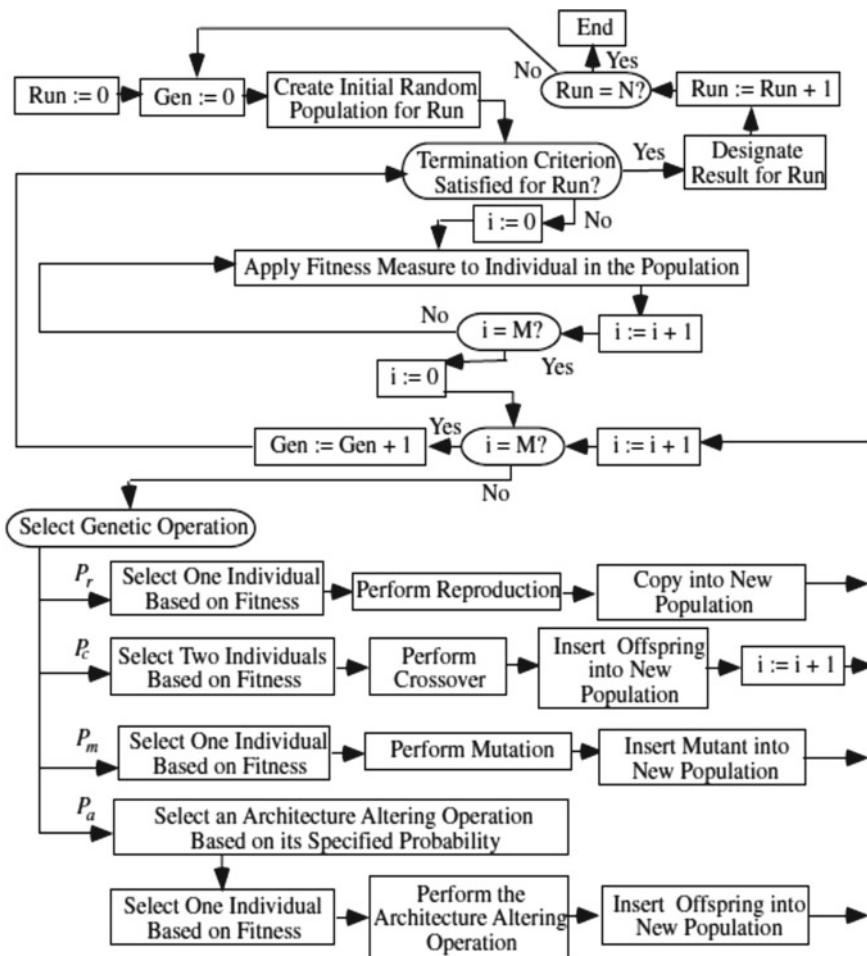
Typical flow chart of genetic programming is shown in Fig. 1, and typical steps followed in GP are as follows.

- Creation of an initial population of individuals (i.e. programs or equations)
- Evaluation of fitness of individuals
- Selection of the fittest individuals as parents
- Creation of new individuals (also called the children or offspring) through the genetic operations of cross-over, mutation and reproduction
- Replacing the weaker parents in the population by the stronger ones
- Repetition of steps until the user-defined termination criterion is satisfied
- Minimum error or number of generations decides the termination criterion.

Knowledge of the underlying physical process is not a prerequisite for GP models. Substantial exogenous metrological and bathymetric data sets are not required. The user even does not have to specify the overall functional form of the model in advance, and still GP models can provide a better approximation of the complex natural processes and more insight into the functional relationship between the input variables [2, 4, 9, 10, 15, 25]. They find the optimal model structure and its coefficients through appropriate learning. Many variants of GP have emerged due to continuous advancement in the areas of computer software and hardware.

## 2.2 Case Study in India

The study presents faecal coliform, biochemical oxygen demand and chemical oxygen demand forecasting models one month in advance for Gangapur reservoir located in state of Maharashtra. Models are developed with 18 input parameters, viz. temperature (Temp °C); electrical conductivity general (EC\_GEN); electrical conductivity field (EC\_FLD;  $\mu\text{mho}/\text{cm}$ ); pH (general and field) (pH\_GEN, pH\_FLD) (pH units); dissolved oxygen (DO; mg/L); total dissolved solids (TDS; mg/L); total coliforms



**Fig. 1** Genetic programming flow chart. (Source <http://www.geniqmodel.com/KozaGPs.html>)

(T-col-MPN; MPN/100 mL); total phosphorus (P-Tot; mg P/L), total nitrogen oxidised ( $\text{NO}_2 + \text{NO}_3$ ; mg N/L); ammonia nitrogen ( $\text{NH}_3\text{-N}$ ; mg N/L); sodium (Na; mg/L); chemical oxygen demand (COD mg/L); carbonate ( $\text{CO}_3$ ; mg/L); chloride (Cl; mg/L); biochemical oxygen demand ( $\text{BOD}_{3-27, 3 \text{ days}}$ ; mg/L); total alkalinity (ALK-Tot; mg  $\text{CaCO}_3/\text{L}$ ); and faecal coliform F-col (MPN/100 mL).

Monthly water quality data collected by the Maharashtra Water Resources Department, Hydrological Data Users Group (HDUG), from March 2001 to January 2015, is used in the present study. The number of sampling points is generally equal to the rounded value of the log of the lake area in square kilometres. The surface area of the lake under consideration is about  $22.86 \text{ km}^2$ . Therefore, the data from a single sampling point is sufficient to represent the lake water quality [12]. Nashik (20°

02' N, and 73° 50' E) is situated on both banks of Godavari River, extending in an east–west direction along its banks and that of its tributaries Nasardi, Waghadi and Darna. Nashik is famous for a religious gathering “*Kumbh Mela*” which adversely affects the environment and public health. The problems arising out of such activities are mainly associated with mass bathing, cloth washing, idol immersion, *nirmalyav-isarjan*, etc. Domestic waste generated is disposed of in the river through *nallas* in unsewered areas. Outside the municipal boundaries, agricultural activities are carried out at a massive scale on both the banks of Godavari River. Because of social events and farming activities of wine yards, there is an adverse impact on river ecosystem, and therefore calls for regular monitoring water quality of the river, define the level of pollution and take immediate remedial measures to restore the quality. Gangapur dam ( $22.86 \text{ km}^2$ ) is an earthen dam constructed on the Godavari River. Water from the reservoir is used for drinking purposes, irrigation and pisciculture. Gangapur dam headwork's on river Godavari, which supplies piped water for almost 1.6 million residents of Nashik Municipal Corporation area [5]. It is the primary source of water for domestic and industrial use in Nashik city. Nashik has sewage treatment plants having a combined capacity of  $270.5 \text{ m}^3/\text{day}$ . About 78 effluent-generating industries from MIDC Satpur are just 18 km away from Gangapur reservoir, and most of them are the industrial sectors having water pollution index score of 60 and above [29].

Various water quality parameters that must be monitored for assessment and prediction of Gangapur lake water quality are physical parameters like temperature, pH and turbidity, DO, DO saturated, TDS, total coliform, P,  $\text{NO}_2$  and  $\text{NO}_3$ , Na, COD, hardness, chlorides BOD and alkalinity.

### 2.3 Experimentation Using Genetic Programming

We have used software developed under the talent project N° 9800463 entitled “Data to Knowledge—D2K” funded by the Danish Technical Research Council (STVF) and the Danish Hydraulic Institute (DHI).

F-col, BOD and COD are core water quality parameters given in water quality criterion 2002 [33]. The presence of faecal coliform bacteria in the aquatic environment indicates water contamination with the faecal material of man or other animals. BOD measures an approximate amount of biodegradable organic matter present in water. It serves as an indicator parameter for the extent of water pollution. Monthly water quality data collected by the Maharashtra Water Resources Department, Hydrological Data Users Group (HDUG), from March 2001 to January 2015, is used in the present study for model verification.

Significant input parameters for all models have been found by genetic programming which is used for cause-effect models. The previous concentration of all significant parameters (t to t-6) is used for hybrid cause-effect models (cause-effect with time step models).

Mathematically, the cause-effect models can be written as presented in equation numbers 1, 2 and 3, respectively.

$$F - \text{col}(t + 1) = f(\text{EC\_FL}, \text{pH\_GEN}, \text{DO}, T - \text{col}, P - \text{Tot}, \text{COD}, \text{BOD}, F - \text{col}) t \quad (1)$$

$$\text{BOD}(t + 1) = f(\text{EC\_FLD}, \text{pH\_GEN}, \text{pH\_FLD}, \text{NO}_2 + \text{NO}_3, \text{Na}, \text{COD}, \text{ALK} - \text{Tot}, \text{BOD}) t \quad (2)$$

$$\text{COD}(t + 1) = f(\text{Temp}, \text{Na}, \text{COD}, \text{Cl}) t \quad (3)$$

GP equations were evolved to develop relationship between outputs at time  $t + 1$  with input variables with time steps from  $t$  to  $t-6$  for hybrid cause-effect models. Refer equation numbers 4, 5 and 6, respectively.

$$\begin{aligned} F - \text{col}(t + 1) &= f(\text{EC\_FLD}(t - 2), \text{pH\_GEN}(t - 1), \text{DO}(t - 4), T - \text{col}(t - 6), P - \text{Total}(t - 2) \\ &\quad \text{COD}(t - 1), \text{BOD}(t - 4), F - \text{col}(t - 1)) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{BOD}(t + 1) &= f(\text{BOD}(t), \text{BOD}(t - 1), \text{BOD}(t - 3), \text{BOD}(t - 4), \text{Cl}(t - 2), \text{pH\_FLD}(t - 1) \\ &\quad \text{pH\_GEN}(t), \text{NO}_2 + \text{NO}_3(t), \text{Na}(t - 2)) \end{aligned} \quad (5)$$

$$\begin{aligned} \text{COD}(t + 1) &= f(\text{COD}(t), \text{COD}(t - 1), \text{COD}(t - 2), \text{COD}(t - 4), \text{COD}(t - 5), \text{Cl}(t - 2) \\ &\quad \text{Cl}(t - 6), \text{Temp}(t - 6), \text{Temp}(t - 3)) \end{aligned} \quad (6)$$

From these equations, the parameters having more than 2% recurrence are treated as significant parameters. The values of significant parameters at time from  $t$  to  $t-6$  may influence the forecasting process [21, 22]. With these significant parameters, GP equations were evolved to develop a relationship between output at time  $t + 1$  and significant input parameters with time steps from  $t$  to  $t-6$ . The control parameters and function sets used for GP runs are summarised in Tables 1 and 2, respectively.

**Table 1** Control parameters used in GP

Sr. No.	Parameter used	Value
1.	Maximum initial tree size	45
2.	Maximum tree size	15
3.	Population size	500
4.	No. of children produced	500
5.	Mutation	0.05
6.	Cross-over rate	0.04–1.00
7.	Objective type	$R^2$ , RMSE

**Table 2** Function set used in GP

Trial No.	Function set
1.	+, -, *, /
2.	+, -, *, /, sqrt

Flow chart for F-col is presented in Fig. 2 based on Eq. 4, and control parameters and functions sets are shown in Tables 1 and 2, respectively. The flow charts for BOD and COD can be developed using Eqs. 5 and 6, respectively. Out of the available data sets, 75% data is used for training and 25% for testing for all runs.

The maximum initial tree size was restricted to 45, and maximum tree size was selected to be 15 because GP tends to evolve uncontrollably large trees if the tree size is not limited [22].

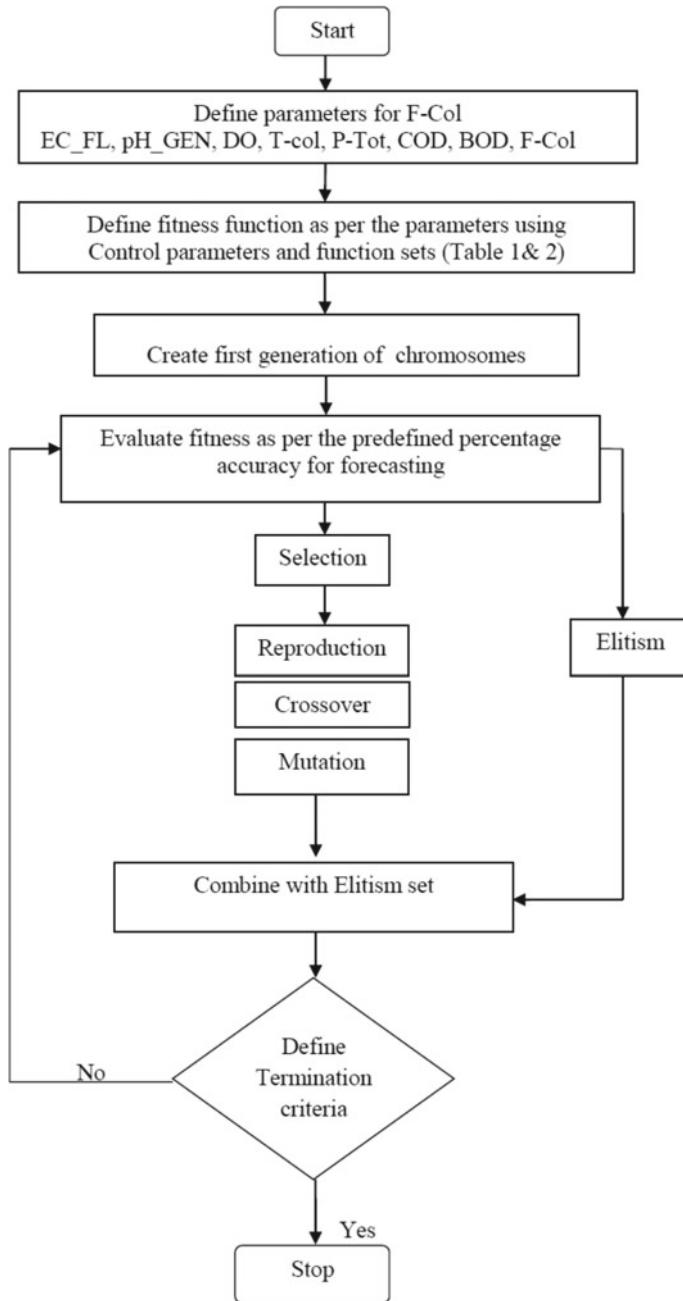
Maximum tree size 15 has another advantage; restricting to this size evolves simple expressions that are easy to interpret and contain only 4–8 variables which are most significant and comfortable to handle [20–22].

The values of population size, no. of children to be produced, objective type, cross-over rate and the mutation were fixed by trial and error and by referring earlier researchers' work [12, 21, 22, 35]. For GP runs, different simple mathematical operations are used as function sets. Small and simple function sets, as represented in Table 2, are used. GP is very creative at choosing simple functions and creating what it needs by combining those [22]. A simple function set also leads to the evolution of simple GP models, which are easy to interpret. With complex functions, the models are difficult to understand and therefore avoided in the present study.

## 2.4 Cross-validation

The data is procured from 2001 to 2015, wherein there is one value per month of each parameter. The sample size is comparatively small. Training data and validation data are prepared in advance. Training data is used for learning. Validation data is never used for learning. The final evaluation is done by validation data, which gives a fair judgement whether the program has acquired an acceptable level of generalisation without overfitting.

There are noise and errors generally involved in learning data because of various reasons. Therefore, when the user conducts the training, until fitness value has reached a minimum (forecasting error), the program has learnt not only what was required to model, what is the phenomenon of interest also it has learned the errors in the particular set of data used for training. If the user can forecast the validation data, it is no longer possible to measure the robustness of the solution. Therefore, it is considered as an excellent practice to prepare the third validation data set (an independent data set) to assess the model which is entirely separate from the training set. More significantly, the data set more robust is the solution obtained. Cross-validation is the process used when data sets are small. Part of the data is excluded, and learning is performed with the remainder of the set. The excluded part is then used for the test. This procedure is repeated with different portions excluded from the original data until all of the data has been eliminated. Ten trials with different sets of data were taken for testing, and mean values of the scores are the index of robustness [11]. Trials are carried for F-col cause-effect model with and without k-fold validation. Since GP-evolved equations relating to input and output variables might shed



**Fig. 2** Genetic programming flow chart for forecasting F-col

physical insight into the ecological processes involved, they are used to identify the significant variables [20, 21, 23].

### 3 Results and Discussion

We have demonstrated the application of genetic programming, which is one of the evolutionary algorithms to forecast lake water quality parameters few time steps in advance. Cause-effect models and hybrid cause-effect models (cause-effect with time steps) are presented.

#### 3.1 Selection of Significant Input Parameters

Selection of significant input parameter is one of the most critical steps. A large number of inputs may lead to the curse of dimensionality [20–22]. Computational complexity and memory requirement of the model increase, due to increase in input dimensionality, which results from an increase in time to build the model. As the input parameters increase, there is an increase in training samples. Addition of irrelevant input increases the local minima present in the error surface, which results in poor model accuracy. Interpreting complex models is complicated, and if simple models achieve comparable results, one should select those. In time series, as lag length increases, the complexity of the model also increases. Thus, the selection of an appropriate set of significant inputs plays an important role. Since GP-evolved equations relating to input and output variables might shed physical insight into the ecological processes involved, they are used to identify the significant variables [21, 22]. For F-col, BOD and COD models, 47, 65, and 54 GP equations, respectively, are evolved for 30 days ahead forecasting as shown in Table 3. A similar exercise for hybrid cause-effect models is shown in Table 5.

**Table 3** Number of equations evolved to find significant inputs for cause-effect models

Trial No.	Function set	% of training	F-col	BOD	COD
			No. of equations evolved		
1.	+, -, *, /, sqrt	75	10	12	10
2.	+, -, *, /, sqrt	80	05	16	10
3.	+, -, *, /, sqrt	85	11	14	13
4.	+, -, *, /, sqrt	90	10	10	11
5.	+, -, *, /	75	11	13	10
Total number of equations evolved			47	65	54

GP evolves equations which contain most significant variables out of the total 18 input parameters.

It is measured by considering number of times the variable is selected in equations. Table 4 shows a summary of the recurrence of several input variables. These parameters are those whose number of terms is more than 2% of the total number of terms in GP equations (Table 5).

For faecal coliform model, eight significant input parameters are identified, e.g. EC\_FL, pH\_GEN, DO, T-col, P-Total, COD, BOD and F-col (t). For BOD model, eight significant input variables are determined, e.g. EC\_FLD, pH\_GEN, pH\_FLD, NO<sub>2</sub> + NO<sub>3</sub>, Na, COD, ALK-Tot, BOD, whereas for COD, four significant input variables are selected, viz. Temp, Na, COD and Cl. Summary for hybrid cause-effect models is presented in Table 6.

### **3.2 Models Developed**

Training data and validation data are prepared in advance for model runs. The final evaluation is performed with validation data, providing a reasonable judgement of whether the program has acquired an acceptable level of generalisation without overfitting. Cross-validation is executed by excluding part of the data, with learning performed with the remainder of the data set. The excluded part is then used for the test. This procedure is repeated with different portions excluded from the original data until all the data have been excluded. Thus, the trials were executed ten times with different data sets for testing and for cross-validation [11].

#### **3.2.1 Cause-Effect Models**

Results of all cause-effect models are shown in Table 7 and Fig. 3a, b, c. Correlation coefficients (CC), root-mean-square error (RMSE), coefficient of determination ( $R^2$ ) and coefficients of efficiency (CE) of forecasted and observed values are presented.

#### **3.2.2 Hybrid Cause-Effect Models (Cause-Effect with Time Steps)**

The values of significant parameters at time from t to t-6 may influence the forecasting process [20–22]. With significant parameters, GP equations were evolved to develop a relationship between output at time t+1 and significant variables with time steps from t to t-6. Table 8 and Fig. 2 show the results of forecasted models.

For both models, performance evaluation of the correlation coefficient (CC), coefficient of determination ( $R^2$ ), root-mean-square error (RMSE) and coefficient of efficiency (CE) are used in the present study to test the performance of various models generated by GP. The correlation coefficient (CC) is selected as the degree of collinearity criterion of forecasting level. It has been widely used for model evaluation.

**Table 4** Recurrence and the contribution factor of each parameter in all equations for BOD and COD cause-effect models

BOD	COD					F-col	Contribution factor in %
	Input variables	Recurrence	Contribution factor in %	Input variables	Recurrence		
Temp	5	1.19	1.19	Temp	63	23.07	EC_FLD
EC_GEN	5	1.19	1	EC_GEN	0.36	pH_GEN	29
EC_FLD	9	2.14	0	EC_FLD	0	pHFLD	7.53
pH_GEN	22	5.23	2	pH_GEN	0.73	DO	1.50
pH_FLD	9	2.14	0	pH_FLD	0	DO_Sat.%	9
T-col	2	0.47	1	T-col	0.36	TDS	3.20
P-Tot	4	0.95	0	P-Tot	0	T-col	0.90
NO <sub>2</sub> + NO <sub>3</sub>	27	6.42	5	NO <sub>2</sub> + NO <sub>3</sub>	1.83	P-Tot	0.90
NH <sub>3</sub> -N	1	0.23	0	NH <sub>3</sub> -N	0	NO <sub>2</sub> + NO <sub>3</sub>	26.50
Na	27	6.42	6	Na	2.19	Na	9.93
F-col	0	0	1	F-col	0.36	COD	1.20
COD	62	14.76	170	COD	62.27	C1	0.60
CO <sub>3</sub>	1	0.23	0	CO <sub>3</sub>	0	BOD	2.71
C1	38	0.95	Cl	24	8.79	F-col	4.20
ALK-Tot	1	47.38	ALK-Tot	0	0		
TDS	4	0.95	TDS	0	0		
BOD	199	47.3809	BOD	0	0		
DO	4	0.9523	DO	0	0		
Total No. of terms	<b>420</b>		Total no. of terms	<b>273</b>		Total No. of terms	<b>281</b>

**Table 5** Number of equations evolved to find significant input for hybrid cause-effect models

Trial No.	Function set	% of training	No. of equations evolved		
			F-col	BOD	COD
1.	+,-,*,/,	75	10	12	10
2.	+,-,*,/,sqrt	80	12	11	11
3.	+,-,*,/,	85	10	10	10
4.	+,-,*,/,pow(x, 2)	90	10	11	10
Total number of equations evolved			42	44	41

Although it is oversensitive to high extreme values (outliers) and insensitive to additive and proportional differences between model forecasting values and measured data [33], it has been identified as an inappropriate measure in hydrological model evaluation. A complete model performance should include at least one absolute error measure (e.g. RMSE) as a necessary supplement to a relative error measure [36].  $R^2$  describes the proportion of the total variance in the observed data that can be explained by the model, and CE provides useful comparisons between studies since they are independent of the scale of data used. It measures the goodness of fit of modelled data concerning observed data [7].

### 3.2.3 Time Lag Correction

Water quality data is periodic data consisting of physical–chemical and biological parameters. Data may be seasonal or yearly, showing monthly or fortnightly periodicities in time series. By plotting time series data with the period domain, we can understand how the signal series changes over time, whereas plotting time series with frequency domain shows the frequency of data to change over time. Frequency domain graph contains information about a signal, amplitude and phase, which is required to regenerate the original data from the frequency spectrum.

Spectral analysis is used for describing the structure of the time series and explaining the main components that contribute to the total variance in the observed data and defining the behaviour using spectral density function. Analysis of data in the frequency domain is useful because it explains the periodicities in input data as well as periodicities of time series analysis. It allows transforming a time series into its coordinates in the space of frequencies and then analyses its characteristics [33].

From the time series plots shown in Figs. 4a and 5a, it can be seen that there is a lag between observed and forecasted parameters for hybrid cause-effect models of BOD and COD. It may be due to autocorrelation because previous values of dependent parameter like (BOD (t-1)) are used to forecast (BOD (t+1)). It introduces error when it is applied in real-time forecasting.

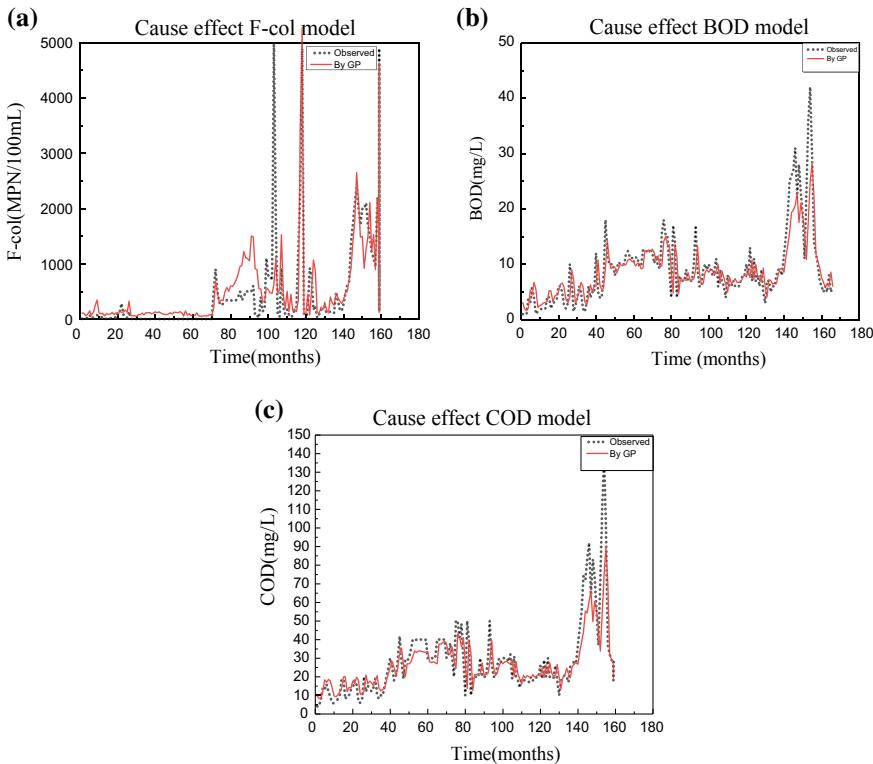
Spectral analysis (SA) is one of the tools which allow transforming a time series into its coordinates in the space of frequencies. It helps to analyse its characteristics also. SA has been applied as a data pre-processing technique to improve neural

**Table 6** Significant input parameters for hybrid cause-effect models

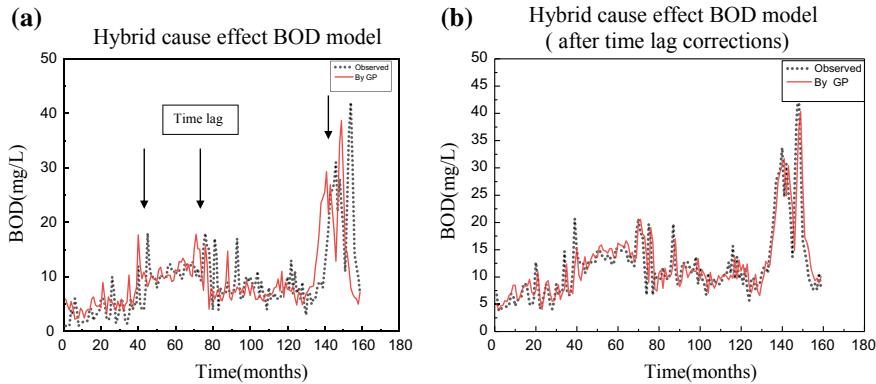
BOD	COD					F-col		
	Input variables	Recurrence in all equations	Contribution factor in%	Input variables	Recurrence in all equations	Contribution factor in%	Input variables	Recurrence in all equations
BOD(t)	117	44.83	COD (t)	123	43.30	EC_FLD(t-1)	48	5.42
BOD(t-1)	6	2.30	Cl (t-6)	10	3.52	EC_FLD(t-2)	55	6.21
BOD(t-3)	19	7.28	COD (t-5)	14	4.93	EC_FLD(t-6)	22	2.48
BOD(t-4)	19	7.28	COD (t-1)	35	12.33	pH_GEN(t-1)	36	4.06
Cl(t-2)	29	11.11	Cl (t-2)	24	8.45	pH_GEN(t-5)	30	3.39
pH_FLD(t-1)	6	2.30	COD (t-4)	15	5.28	DO(t)	31	3.50
NO <sub>2</sub> + NO <sub>3</sub> (t)	21	8.05	COD (t-2)	13	4.58	T-col(t-1)	264	29.80
pH_GEN(t)	6	2.30	Temp (t-6)	12	4.22	P-Tot(t-3)	50	5.64
Na(t-2)	7	2.68	Temp (t-3)	9	3.16	COD(t-4)	54	6.09
						BOD(t-4)	88	9.93
						BOD(t-5)	32	3.61
						P-Tot(t)	30	3.39
Total No. of terms	261			284			886	

**Table 7** Results of cause-effect models with significant parameters

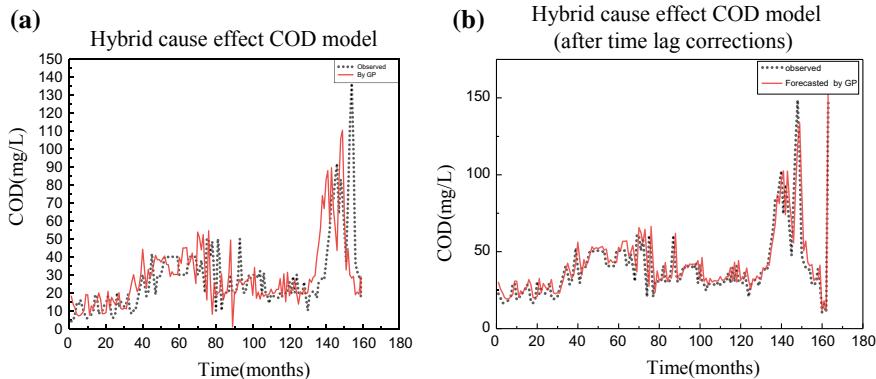
Performance evaluator	Model	Forecasting parameters		
		F-col	BOD	COD
Correlation coefficient	CC	0.85	0.84	0.84
Root-mean-square error	RMSE	489.77	2.45	6.53
Coefficient of determination	$R^2$	0.76	0.61	0.67
Coefficient of efficiency	CE	0.76	0.67	0.66

**Fig. 3 a, b, c** Comparison of observed and predicted values (cause-effect models)**Table 8** Performance evaluation of hybrid cause-effect models

Performance evaluator		Forecasted parameters		
		F-col	BOD	COD
Correlation coefficient	CC	0.85	0.84	0.87
Root-mean-square error	RMSE	498.28	3.58	10.05
Coefficient of determination	$R^2$	0.81	0.71	0.75
Coefficient of efficiency	CE	0.76	0.18	0.16



**Fig. 4** **a** Time series plot of hybrid BOD model (with an observed time lag), **b** time series plot of hybrid BOD model (after executing time lag correction)



**Fig. 5** **a** Time series plot of the hybrid COD model (with an observed time lag), **b** time series plot of the hybrid COD model (after executing time lag correction)

network performance in daily flow predictions [36]. In the present study, SA is used to estimate the error frequency of all BOD models with time steps and the COD model with time steps. Table 9 indicates the average repetition time of error cycle for hybrid cause-effect models of BOD and COD. Trial version of XLSTAT is used to perform SA.

**Table 9** Repetitive time of error cycle

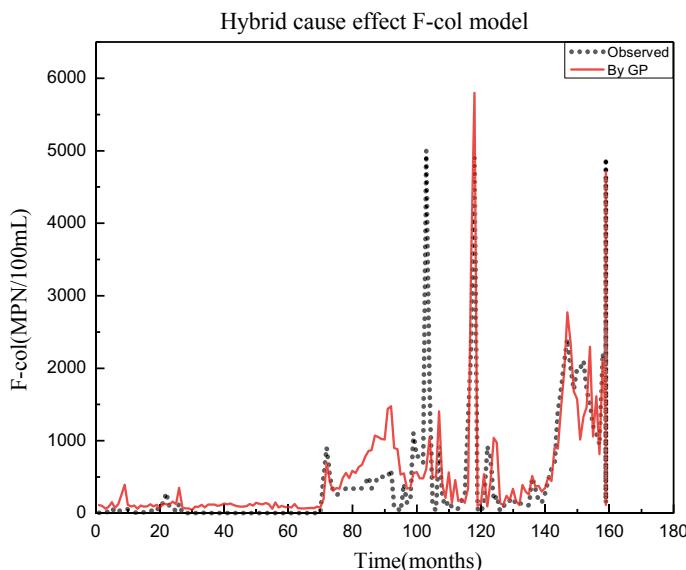
Model	Error cycle repetition time
Hybrid cause-effect model of BOD	2.69
Hybrid cause-effect model of COD	10.66

Runs were executed with transformed data set by differencing and converting each  $i$ th element into its difference from the  $(i-k)$ th element. The correction is applied for the respective data set. New models were developed with corrected data sets for GP.

It was observed that the time lag is removed, and the relationship between input and output is better mapped. Time series plot before and after removal of time lag for hybrid cause-effect BOD and COD models is presented in Figs. 4a, b and 5a, b. Results of both models before and after time lag corrections are shown in Table 10. Time series plot for hybrid cause-effect models for F-col is presented in Fig. 6.

**Table 10** Comparison of results of hybrid cause-effect models (after lag correction)

		Before lag correction		After lag correction	
		Forecasted parameters	Forecasted parameters	BOD	COD
Performance evaluator					
Correlation coefficient	CC	0.84	0.87	0.86	0.88
Root-mean-square error	RMSE	3.58	10.05	3.33	10.05
Coefficient of determination	$R^2$	0.71	0.75	0.74	0.75
Coefficient of efficiency	CE	0.18	0.16	0.22	0.19



**Fig. 6** Time series plot of hybrid F-col model

## 4 Conclusion and Future Directions

We have presented the application of genetic programming to forecast lake water quality parameters 30 days in advance. Significant water quality parameters are site-specific, and genetic programming models are capable of finding them. The sample size is comparatively small. To ensure that models are not overfitted, cross-validation was performed.

In water quality modelling, researcher always prefers cause-effect models than time series models. In this study, cause-effect models work better than hybrid cause-effect models. In developing countries like India, there is a problem of availability of data. In such situations, hybrid models are preferred if significant parameter data is not available. We can observe that the results of both types of models are comparable. In hybrid cause-effect models of BOD and COD, spectral analysis is used to remove the time lag. Error cycle repetition time for each model was found. It is found that the performance of all models is improved after removing time lag.

The major challenge is the small data sets used. In the present study, 14 years of data, with only one value per month of each parameter, is used. When a sufficient number of samples are available, the efficiency of learning by data-driven techniques of the interrelationships in the data is expected to correlate with its test performance. We have tried the performance only for a few case studies and have also ensured that the models are not overfitted by executing cross-validation. To handle such a situation in a better way and to evaluate the performance in the presence of random effects, a surrogate data test is proposed in such studies [32]. Surrogate data mimics the statistical properties of the original data set independently for each component of the input vector. Such an exercise is planned as future work.

## References

1. Azhagesan R (1999) Water quality parameters and water quality standards for different uses. National Water Academy Report
2. Banzhaf W, Nordin P, Keller RE, Francone FD (1998) Genetic Programming: an introduction, an automatic evolution of computer programs and its applications. Morgan Kaufmann Publishers, Inc., San Francisco, California
3. Bartram J, Ballance R, World Health Organization & United Nations Environment Programme (1996). Water quality monitoring: a practical guide to the design and implementation of fresh-water quality studies and monitoring programs. In: Bartram J, Ballance R (eds). E & FN Spon, London. <http://www.who.int/iris/handle/10665/41851>
4. Brameier M (2004) On linear genetic programming; Ph.D. thesis, University of Dortmund <https://pdfs.semanticscholar.org/31c8/a5e106b80c07c1c0f74bcf42de6d24de2bf1.pdf>
5. Chavan A, Sharma MP, Bhargava R (2009) Water quality assessment of the Godavari river National conference on hydraulics. HydroNepal J Water Energy Environ 1:31–34. <https://doi.org/10.3126/hn.v5i0.2483>
6. Coppola E, Rana A, Poultney M, Szidarovszky F, Uhl V (2005) A neural network model for predicting aquifer water level elevations. Ground Water 43(2):231–243
7. Dawson CW, Wilby RL (1999) Hydrological modelling using artificial neural networks. Prog Phys Geogr Earth Environ 25(01):80–108. <https://doi.org/10.1177/030913330102500104>

8. Dogan E, Koklu R, Sengorur B (2009) Modelling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J Environ Manage* 90(2):1229–1235
9. Francone FD, Markus C, Banzhaf W, Nordin P (1999) Homologous crossover in genetic programming. *Proc Genet Evol Comput Conf* 2:1021–1026
10. Guven A (2009) Linear genetic programming for time-series modelling of daily flow rate. *J Earth Syst Sci* 118(02):137–146
11. Hitoshi I, Yoshihiko H, Topon KP (2009) Applied genetic programming and machine learning. CRC Press International Series on Computational Intelligence, Boca Raton
12. Jadhav MS, Khare KC, Warke AS (2015) Water quality prediction of Gangapur reservoir (India) using LS-SVM and genetic programming. *Lakes Reservoirs Res Manag* 20(04):275–284. <https://doi.org/10.1111/lre.12113>
13. Jadhav MS, Khare KC, Warke AS (2014) Selection of significant input parameters for water quality prediction—a comparative approach. *Int J Res Advent Technol* 2(03):81–90
14. Khovanova NA, Shaikhina T, Mallick KK (2015) Neural networks for analysis of trabecular bone in osteoarthritis. *Bioinspired, Biomimetic Nanobiomaterials* 4(1):90–100
15. Koza JR (1992) Genetic programming: on the programming of computers using natural selection. A Bradford book. MIT Press, Cambridge, Massachusetts, London, England
16. Lebaron B, Weigend AS (1998) A bootstrap evaluation of the effect of data splitting on financial time series, *IEEE Trans Neural Networks* 213–220
17. Lermontov A, Yokoyama L, Lermontov M, Machado MAS (2009) River quality analysis using fuzzy water quality index: Ribeira do Iguaçu river watershed, Brazil. *Ecol Ind* 9(6):1188–1197
18. Londhe SN, Dixit PR (2012) Genetic programming—new approaches and successful applications. In: Soto SV (ed) 8/12. In Tech Publications
19. Londhe S, Charhate S (2010) Comparison of data-driven modelling techniques for river flow forecasting. *Hydrol Sci J* 55(7):1163–1174
20. Muttill N, Chau K (2007) Machine learning paradigms for selecting ecologically significant input variable. *Eng Appl Artif Intell* 20(06):735–744. <https://doi.org/10.1016/j.engappai.2006.11.016>
21. Muttill N, Chau K (2006) Neural network and genetic programming for modelling coastal algal blooms. *Int J Environ Pollut* 28(3–4):223–238. <https://doi.org/10.1504/IJEP.2006.011208>
22. Muttill N, Lee JHW (2005) Genetic programming for analysis and real-time prediction of coastal algal blooms. *Ecol Model* 189(03):363–376. <https://doi.org/10.1016/j.ecolmodel.2005.03.018>
23. Muttill N, Lee JHW, Jayawardena AW (2004) Real-time prediction of coastal algal blooms using genetic programming. In: 6th international conference on hydro informatics. Singapore, pp 890–897. [https://doi.org/10.1142/9789812702838\\_0110](https://doi.org/10.1142/9789812702838_0110)
24. Najah A, Elshafie A, Karim OA, Jaffar O (2009) Prediction of johor river water quality parameters using artificial neural networks. *Eur J Sci Res* 28(3):422–435
25. Nordin JP (1997). Evolutionary program induction of binary machine code and its application. Ph.D. dissertation, Department of Computer Science, University of Dortmund
26. Palani S, Liong S-Y, Tkalich P (2008) An ANN application for water quality forecasting. *Mar Pollut Bull* 56:1586–1597
27. Preis A, Ostfeld A (2008) A coupled model tree–genetic algorithm scheme for flow and water quality predictions in watersheds. *J Hydrol* 349:364–375
28. Recknagel F, Cao H, Kim B, Takamura N, Welk A (2006) Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecol Inform* 1(2):133–151
29. Sawant R (2015). A comprehensive study of polluted river stretches and preparation of action plan of river Godavari from Nashik downstream to Paithan. The report, Aavanira Biotech P. Ltd., Maharashtra Pollution Control Board. [http://mpcb.gov.in/ereports/pdf/GodavariRiver\\_ComprehensiveStudyReport.pdf](http://mpcb.gov.in/ereports/pdf/GodavariRiver_ComprehensiveStudyReport.pdf)
30. Shiklomanov I (1993) Water in crisis: a guide to the world's freshwater resources. In: Gleick PH (ed). Oxford University Press, New York, pp 13–25, [https://www.academia.edu/902661/Water\\_in\\_Crisis\\_Chapter\\_2\\_Oxford\\_University\\_Press\\_1993](https://www.academia.edu/902661/Water_in_Crisis_Chapter_2_Oxford_University_Press_1993)

31. Tikhe SS, Khare KC, Londhe SN (2015) Multicity seasonal air quality index forecasting using soft computing techniques. *Adv Environ Res* 4(02):83–104. <https://doi.org/10.12989/aer.2015.4.2.083>
32. Shaikhina T, Khovanova NA (2017) Handling limited datasets with neural networks in medical applications: a small-data approach. *Artif Intell Med* 75:51–63. <https://doi.org/10.1016/j.artmed.2016.12.003>
33. US Environmental Protection Agency (2009). Technical assistant document for reporting of daily air quality—air quality index. Research Triangle Park, North Carolina
34. Wang W, Chau K, Xu D, Chen X (2015) Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour Manage* 29(08):2655–2675. <https://doi.org/10.1007/s11269-015-0962-6>
35. Water Quality Criteria (2002) <https://www.epa.gov/sites/production/files/2018-12/documents/national-recommended-hh-criteria-2002.pdf>
36. Whigham PA, Recknagel F (1999) Predictive modelling of plankton dynamics in freshwater lakes using genetic programming. The Information Science Discussion Paper Series. Department of Information Science, University of Otago, Dunedin, New Zealand, pp 1–7
37. Wu CL, Chau KW, Li YS (2009) Methods to improve neural network performance in daily flows. *J Hydrol* 372(1–4):80–93
38. Xiang, Y, Jiang L (2009) Water quality prediction using LS-SVM and particle swarm optimization. In: Conference proceedings of the second international workshop on knowledge discovery and data mining, WKDD 2009, Moscow, Russia. <https://doi.org/10.1109/wkdd.2009.217>

## Chapter 6

# A Survey on the Latest Development of Machine Learning in Genetic Algorithm and Particle Swarm Optimization



Dipti Kapoor Sarmah

## 1 Introduction and Literature

Due to the industrial importance, optimization always attracts attention of several researches from industry and academia. Nowadays, the complexity to solve real-world problems has increased due to the increased dependencies of external factors, resulting many times the classical optimization techniques fail to solve them. At the same time, the area of optimization is becoming richer with new concepts, ideas and algorithms due to the continuous innovations by the researchers. As there is no common method to solve a particular problem (free lunch theorem [117]), there is always a scope of improvement towards the existing algorithms. Due to which more and more efficient optimization algorithms are getting developed and used in various fields on regular basis, rapid development of technologies and increased computational powers of the computers are also contributing to this. The efficiency of any optimization algorithm can be judged and observed by maintaining a good balance between two functions, exploration [109] and exploitation [109], related to global search and local search, respectively. Similar to optimization field, a new emergent concept of machine learning (ML) such as deep learning (DL) [15] is continuously progressing to solve such problems. It has been observed that by applying the concept of ML along with various optimization algorithms, particularly along with nature-inspired algorithms, the performance of the computational power is increasing while finding solution to various challenging problems.

---

D. K. Sarmah (✉)

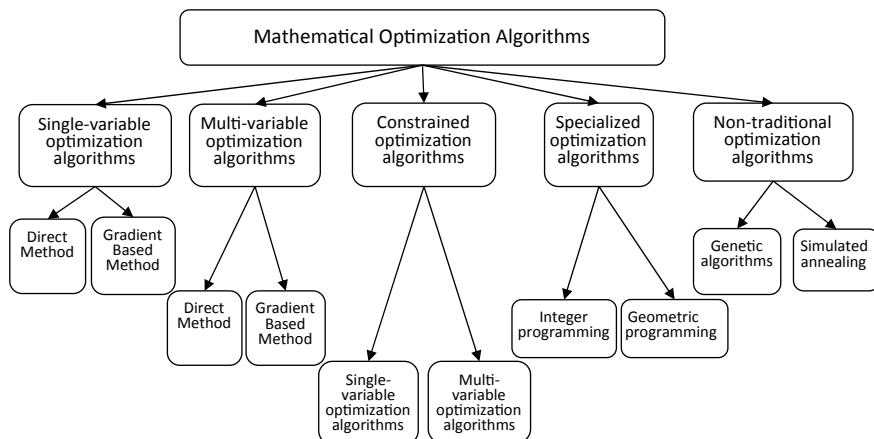
Symbiosis Institute of Technology, Symbiosis International (Deemed University), 412115 Pune, MH, India

e-mail: [diptik.sarmah@gmail.com](mailto:diptik.sarmah@gmail.com)

## 1.1 Literature on Optimization Algorithms

As depicted in Fig. 1, there are five categories of optimization algorithms. The first two categories are single-variable optimization algorithm [69] and multi-variable optimization algorithms [69]. Single-variable optimization algorithms are classified into two groups: (a) direct methods [113] and (b) gradient-based methods [113]. Direct methods always take the values of objective function to analyse the search process. There is no derivative information of the objective function associated to execute the process. However, the gradient-based methods utilize the first-order/second-order derivative functions to guide the search process. Very few single-variable optimization problems exist in the real scenario; thus, multi-variable optimization algorithms are demonstrated. These algorithms are also partitioned into two techniques: (a) direct and (b) gradient-based techniques. The third category is defined as constrained optimization algorithms [7]. These algorithms frequently make an effort to identify the optimal solution in the feasible search region. They are most often used to solve engineering optimization problems. The fourth category is considered as specialized optimization problems classified as integer programming [71] and geometric programming [44]. Integer programming deals with the integer design variables. However, the geometric programming entertains the objective function and constraints written in particular form. The last category is described as non-traditional optimization algorithms. They are referred to as (a) genetic algorithm (GA) [73] and (b) simulated annealing [27].

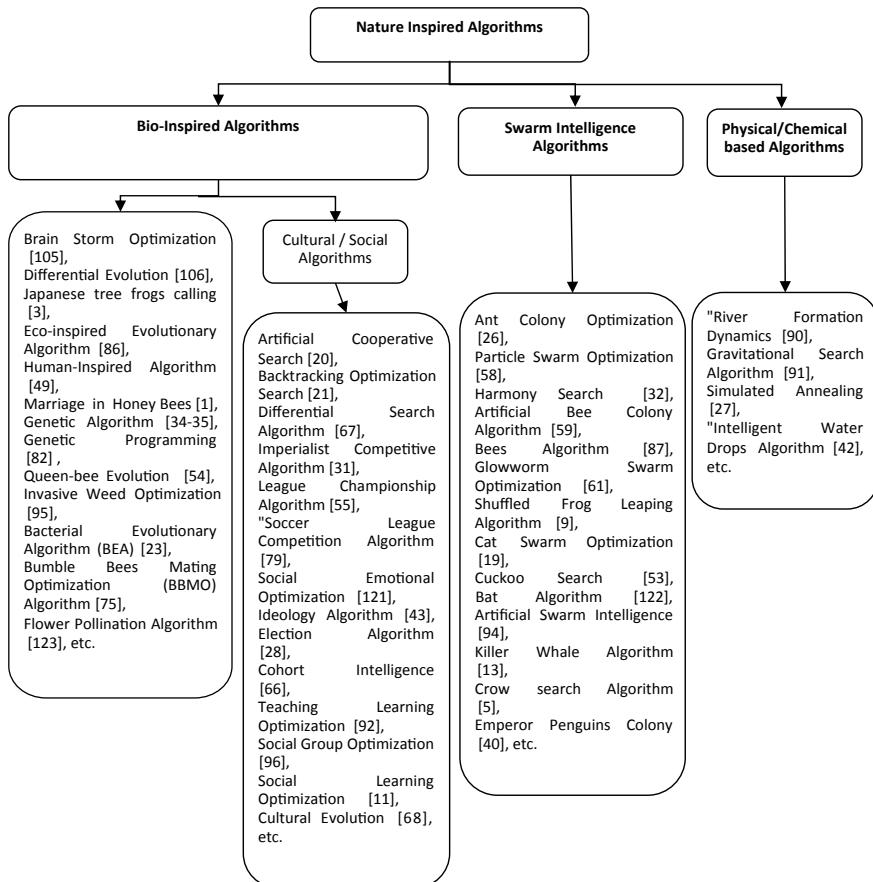
The broad category of solving optimization engineering problem is described as heuristics [63] and metaheuristics [45]. Heuristics techniques most often fall into local optima as they are very much problem dependent and try to utilize all the problem parameters and its specification. On the other side, the metaheuristics algorithms are not at all problem dependent. Such techniques explore the solution space



**Fig. 1** Categories of optimization algorithms

more thoroughly to get a better solution, and it can be used as black boxes. There is no assurance to achieve global optimal solution from metaheuristics algorithms in comparison with iterative methods. They are very useful to identify the optimal solution to real-world combinatorial problems as it is very simple for them to search a solution for a large set of feasible solutions. These algorithms are generally referred to as nature-inspired optimization techniques [126] which are quite popular among researchers nowadays. Such techniques consist of collections of algorithms which seek inspiration from various occurrence perceived in nature. As shown by Kumar et al. [65], the broad category of nature-inspired optimization algorithms is partitioned into three groups: (a) bio-inspired, (b) swarm intelligence and (c) physical–chemical systems. These techniques are successfully used to solve NP-hard problems.

The promising algorithms under each category are listed in Fig. 2. The recognition of these algorithms has increased due to their approach in finding the optimal solutions to complex and real-world computational problems. The limitations of such



**Fig. 2** Classification of nature-inspired optimization algorithms

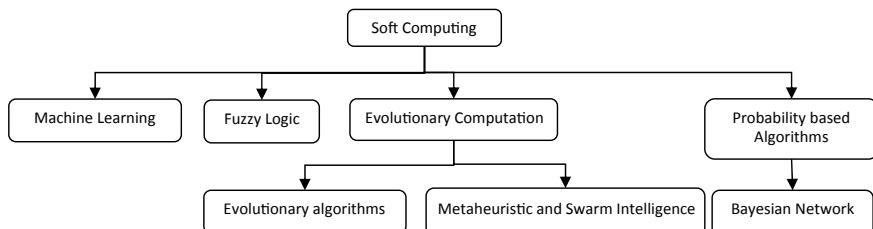
algorithms are tuning of the parameters, handling large-scale diverse applications having millions of variables, possibility of hybridization of the mentioned algorithms and the capability of self-adaptivity to solve a complex computational problem in a quick manner. These boundaries create a room for the interested researchers to develop new optimization algorithms or blending with new science/concept to solve the complex combinatorial problems in an effective manner.

## 1.2 Literature on Machine Learning Algorithms

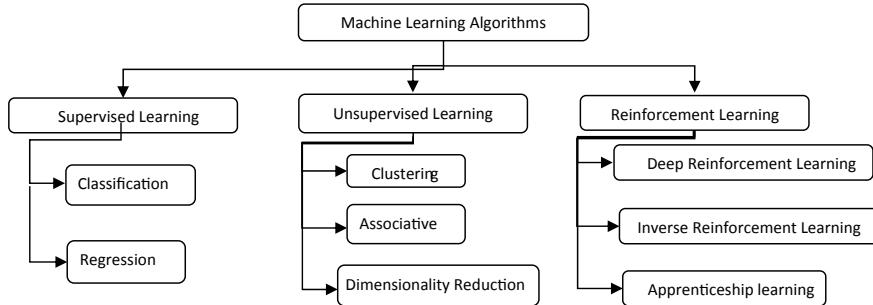
On the other side, there is an emergent concept of ML, showing potential to solve computational thinking natural world problems. DL algorithm is one of the strongest developments of this concept which is radically gaining the importance and transforming the real-world scenario by enhancing the performance of the computer-based procedures. ML is a sub-branch of computational intelligence/soft computing which works on the principles of human mind. It tries to solve the NP-hard problems to compute the exact solution in polynomial time which otherwise is sometimes challenging for the existing algorithms. As shown in Fig. 3, the science of soft computing is divided into four components: (a) ML, (b) fuzzy logic, (c) evolutionary computation and (d) methods involving probability computations.

ML is also referred to as predictive algorithms which designs a mathematical model centred on certain training data. The nature of these algorithms varies based on the allotted task, problem, input and output and is categorized into three groups: (a) supervised learning [119], (b) unsupervised learning [33] and (c) reinforcement learning [78] as depicted in Fig. 4.

The common functionality of any of the group of ML algorithms is to mimic the human common sense to identify the hidden characteristics or features for analysing the new data. Supervised learning algorithms work on the pair of input data and required output(s), referred to as training examples which helps to prepare mathematical model to predict new data or to improve precision of its outputs. Supervised



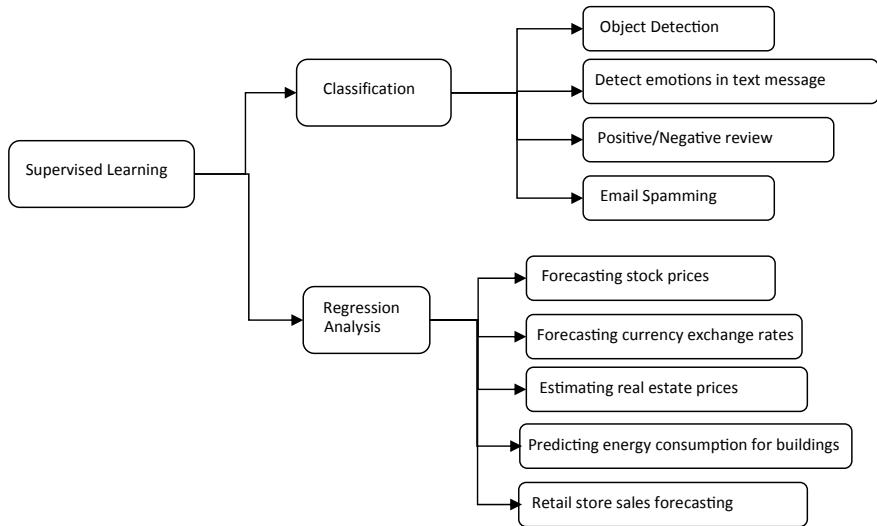
**Fig. 3** Soft computing components



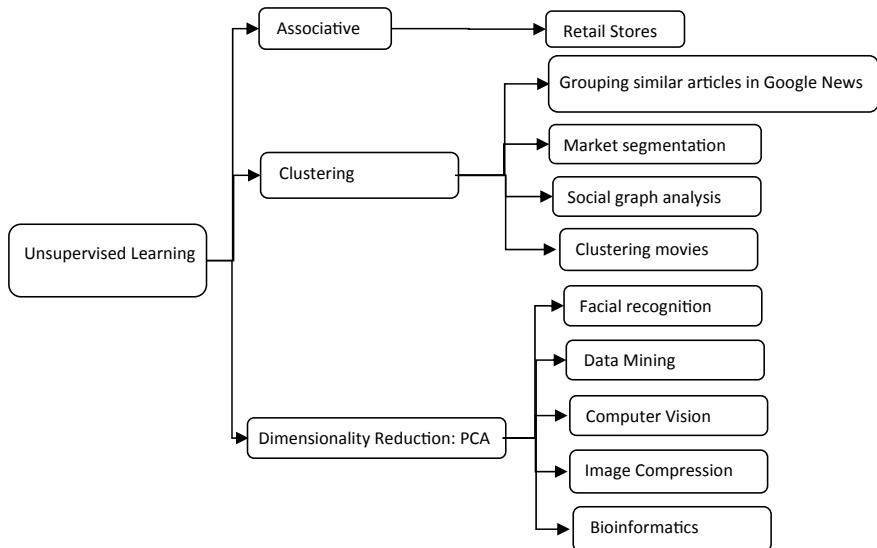
**Fig. 4** Classification of ML algorithms

learning is classified into two subgroups, i.e. classification and regression. The second group of unsupervised learning algorithms deals with the input data. There are no respective output values associated with it. Such algorithms build a mathematical model by identifying some common features of the raw data and learn based on the occurrence of that feature in each new bit of data. Unsupervised learning algorithm is partitioned into three subgroups, i.e. clustering, associative and dimensionality reduction. Clustering is used when there is a need to group similar type of data in one cluster. K-means [38] and K-nearest neighbours (KNNs) [130] are the well-known algorithms used for clustering. The second subgroup of unsupervised learning, i.e. associative, is used to find the closeness or togetherness of frequently used items. Apriori algorithm is the acknowledged algorithm under this category. The third subgroup, i.e. dimensionality reduction, is employed to solve complex problems where thousands of input parameters are involved. The well-known algorithm under this subgroup is principal component analysis (PCA) [52] which transforms the two-dimensional input parameters to one dimension. The third group of reinforcement learning focuses to maintain a balance between exploration and exploitation in which the decision is taken by the system based on the preceding performed action. It is divided into three subgroups: (a) deep reinforcement learning [80], (b) inverse reinforcement learning [2] and (c) apprenticeship learning [85]. The practical and most common use cases of these three groups are depicted in Figs. 5, 6 and 7.

The widely used ML algorithms are linear regression, logistic regression, clustering/K-means, support vector machine (SVM) [128], decision trees [62], Naïve Bayes [116], etc. The complex and advance form of ML is referred to as DL which employs the concept of neural network (NN) [118]. A NN is a model used in ML which solves the complex problems by modelling the data using neurons. They take the intelligent decisions by themselves by structuring the NN in a layered form. One of the simplest forms of NN is referred to as artificial neural network (ANN) [128, 129] which consists of three layers of neurons: (a) input layer, hidden layer and output layer. DL is also considered as a subset of ML where multiple layers can be used by DL models in order to extract the high-level features. Such neural networks are recognized as deep neural networks (DNNs) [15]. DL algorithms play a very



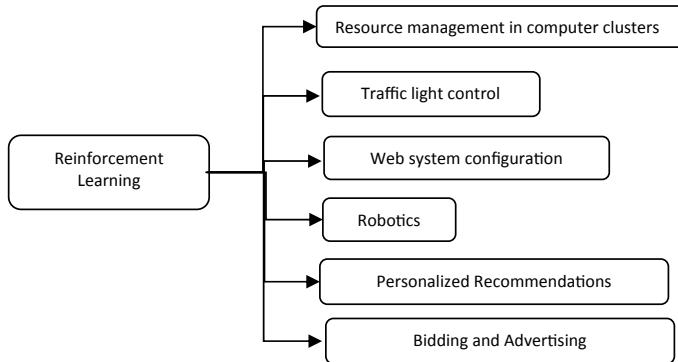
**Fig. 5** Practical use cases for supervised learning algorithms



**Fig. 6** Practical use cases for unsupervised learning algorithms

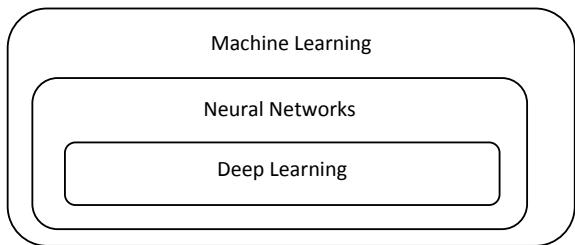
important role in solving real-world NP-hard problems. The relationship of ML, DL and NN is exhibited in Fig. 8.

The broad classification of learning algorithms using NN is shown in Fig. 9. The most commonly used architectures are DNN, convolution neural network (CNN)

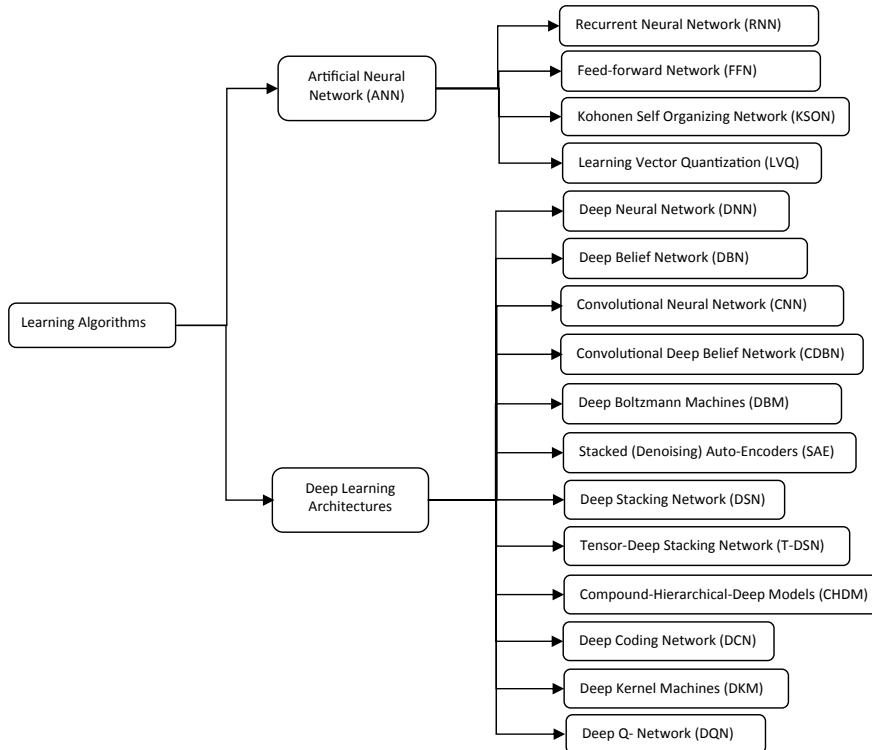


**Fig. 7** Practical use cases of reinforcement learning

**Fig. 8** Relationship between ML, DL and NN



[47] and recurrent neural network (RNN) [81]. DNN typically follows a strategy used in feed-forward network (FFN) where the movement of data is from input layer to output layer without looping back. In RNN, the data movement is in either direction forward/backward. RNN can be mainly used to solve sequential problems such as (a) one-to-many, (b) many-to-one, (c) many-to-many. The most common applications used for RNN are handwriting recognition, speech recognition, natural language processing, sentiment analysis, question answering, anomaly detection in time series, log data analysis (Web data), sensor data analysis (time series), video classification, etc. On the other side, CNN is mostly used for image data. This architecture can be applied to any of the prediction problems such as classification prediction or regression prediction, where image data is used as an input. The variety of processes/algorithms can be applied to single/multiple types of ML algorithms to improve their performance. However, each algorithm cannot be used to solve all types of problems as there are certain pros and cons associated with every algorithm. Thus, there is a need to explore further in recognizing ML algorithms as per the real-world applications. Also, the efficiency and the optimized solution of the algorithms could be observed for solving a particular application which opens a new direction for the researchers to hybridize nature-inspired optimization algorithms with ML.



**Fig. 9** Learning algorithms for neural networks (NNs)

In this section, we have discussed the categories of optimization algorithms and various ML techniques which are effectively used in real-world challenging problems. The next section explains the well-known nature-inspired optimization algorithms used in ML. Conclusion and future scope are mentioned in Sect. 3. References are drawn at the end.

## 2 Nature-Inspired Optimization Algorithms Used in Machine Learning

As discussed in Sect. 1 about the various well-known nature-inspired optimization algorithms, most commonly used algorithms with ML are categorized as GA, particle swarm optimization (PSO), cuckoo search (CS), ant colony optimization (ACO), artificial bee colony (ABC), etc. They are widely used with NN to optimize the solutions in different applications. However, the special focus is given to GA and PSO based on their popularity and problem-solving approach in an efficient way. In

this section, these two nature-inspired optimization algorithms employed with ML are studied and explored.

## 2.1 *Genetic Algorithms with Machine Learning*

Most of the organizations are employing GA in neural networks (NNs) to make NN more efficient in terms of learning and providing a better solution. GA is one of the popular metaheuristic, stochastic and nature-inspired optimization algorithms which works on the principle of survival of the fittest. GA relies on three bio-inspired operators, i.e. mutation, crossover and selection. On the other side, NN has also solved a variety of real-world challenging problems. However, there is still an issue of knowing the correct hyperparameters for a NN as described in the article published by Suryansh [110]. In this paper, GA is combined with NN to identify the accurate parameters. The continuous learning of these parameters is done through GA. Table 1 demonstrates various applications where GA is successfully applied with ML to solve a variety of optimization problems. GA is effectively used for designing NN as proposed by several researchers [39, 51, 93]. In the year 1994, Koehn solved the encoding problem by combining GA and NN. As described by Watanabe et al. [115], Hopfield neural network (HNN) and a GA are combined to solve a combinatorial optimization problem. HNN is a type of RNN which assures to converge the solution to a local minimum, and therefore, sometimes the solution may get converged to a wrong local minimum instead of the expected actual solution. Thus, the solutions obtained from HNN are passed to GA to get the actual global optimal solution. In this paper, the applied combination of GA and HNN is investigated and validated for solving three NP-complete problems, i.e. the maximum clique problem, the node cover problem and the travelling salesman problem. Musical composition is also developed by hybridizing the two techniques of ML and GA [16]. The authors identified a novel procedure for which NN is used as a fitness evaluation parameter and proposed a method in which the NN fitness evaluation is applied to GA. In the proposed work, the concept of an adaptive resonance theory (ART) is used as fitness evaluator. Experimental results show successful results in order to generate rhythmic patterns. The review paper is published by Shapiro [103] by merging the concept of GA and NN with fuzzy logic. The author addressed the limitations and advantages of the mentioned technology and presented a study on insurance-related applications.

Further, Shafti et al. [104] worked on multi-feature extraction and proposed a novel method of constructive induction (CI) where greedy search is applied to identify the new features from a given attribute set. In the research work, the two approaches of ML and GA are amalgamated to address the limitations of CI. The experimental results show a success towards this approach. Taiwanese banking industry is also considered as one of the applications for financial warning system by the authors Hsieh et al. [41] where GA and NN are integrated to compare the performance with other four early warning systems, i.e. case-based reasoning (CBR), back-propagation

**Table 1** A list of applications as per state of the art by combining GA and ML

Application	Author	Year	References
Designing neural network	Robbins et al.	1993	[93]
Designing neural network	Jone	1993	[51]
Encoding problem	Koehn	1994	[64]
To solve a combinatorial optimization problem	Watanabe et al.	1998	[115]
Musical composition	Burton et al.	1998	[16]
Insurance applications	Shapiro	2002	[103]
Multi-feature extraction	Shafti et al.	2004	[104]
Designing neural network	Harpham et al.	2004	[39]
Financial warning system	Hsieh et al.	2006	[41]
To estimate electrical energy consumption	Azadeh et al.	2007	[4]
Job shop scheduling	Lee et al.	2010	[72]
Stock price prediction	Kaboudan	2010	[57]
To optimize topology and neural weights of feed-forward network	Vizitiu et al.	2010	[114]
To estimate the quality of a river	Ding et al.	2014	[25]
Profile identification	Carbone et al.	2015	[17]
Optimize NN through GA	Chiroma et al.	2017	[18]
Solve Benchmark problems [12, 14, 77, 99, 100, 111]	Such et al.	2018	[107]
Optimize the CNN model for different visual data sets	Tian et al.	2018	[112]
Identify the best travel route	Lazovskiy	2018	[70]
English character recognition	Kaur et al.	2019	[56]
Identify the suitable architecture of CNN based on the image classification problem	Sun	2019	[108]
Decision structure management	Serrano	2018, 2019	[101, 102]

neural network (BPNN), logistic regression analysis (LR) and quadratic discriminant analysis (QDA). In order to validate the results, the financial information of different banks of Taiwanese banking industry is collected from 1998 to 2002. Furthermore, Azadeh et al. [4] proposed a new technique to estimate electrical energy consumption by integrating GA and ANN. Researchers have considered the case study of Iranian agriculture sector from 1981 to 2005. In order to predict electricity demand, few parameters have been used in this paper. GA is applied to tune these parameters, and ANN is used for forecasting the electricity consumption rate. The results are validated in comparison with regression analysis and time series approach. One of the combinatorial NP-hard optimization problems of job shop scheduling is efficiently solved by hybridizing the concept of GA and ML. Lee et al. [72] proposed a system by considering the strengths of GA and ML to build a system to solve this

optimization problem. The obtained results are quite satisfactory in comparison with the contemporary methods. Further, genetic programming (GP) is used to predict the stock prices.

Kaboudan [57] proposed a profitable prediction approach where GP is utilized to develop regression models which direct to build up a single day-trading strategy (SDTS). The proposed work is validated on six stocks for 50 successive trading days and experimentally produced high returns on investment in comparison with the similar approaches. The combination of GA and NN further helped to optimize both topology and neural weights of feed-forward network [114]. A hybrid intelligent algorithm is developed by Ding et al. [25] to estimate the quality of a river by combining three techniques, i.e. principal component analysis (PCA) [125], GA [6] and back-propagation neural network (BPNN) [10]. It is observed that the merging of these three techniques predicts accurately the water quality which further helps to reduce the real-time associated risk [60]. Carbone et al. [17] worked on the limitation of profile identification where sorting a number of profiles contextually and recognizing a profile for a particular individual are quite challenging. The authors developed a framework by customizing the concept of vector space model. GA is further applied to train this model and to identify and compare the similarity of two profiles. Experimentally, it is proved to consider this method for profile clustering or finding a match similarity between numbers of profiles.

Further, a review study is done by Chiroma et al. [18] to optimize NN through GA. In this paper, authors analysed several NN design issues and limitations to solve complex problems by employing GA and presented a state of the art. Furthermore, Such et al. [107] worked on a query “Is GA suitable to solve a problem in deep artificial neural network (DANN)?” The authors considered a population-based GA to gradually develop the weights of DNN. This paper validated the said question by applying GA on deep reinforcement learning (DRL) benchmark problems such as Atari 2600 [12, 14, 77] and Humanoid Locomotion in the MuJoCo simulator [14, 99, 100, 111]. The satisfactory results are obtained in comparison with the existing algorithms. GA is also applied with ML to recognize profile of a person. Tian et al. [112] proposed a research work which optimized the CNN model for different visual data sets. The optimizing is done using GA by considering pre-trained CNN models as population. Experimental results prove the efficiency of the proposed framework in comparison with the contemporary techniques. GA is further applied with ML to optimize the travel time between each pair of location to identify the best travel route [70]. In this article, a tree-based ML model, i.e. standard XGBoost model, is applied to a huge data set to manage various categorical features. GA is then applied to this trained model to plan the optimal journey. However, this work could be further extended by incorporating Google Maps API for route planning.

Further, Kaur et al. [56] developed a system for English character recognition by hybridizing the concept of NN and GA where back-propagation algorithm is used with GA to work on the extracted features of characters. On the other side, Sun [108] presented a novel method to identify the suitable architecture of CNN based on the image classification problem. GA is applied to discover the suitable CNN architecture. The proposed work is validated on benchmark data sets by comparing with

different CNN architectures such as manually designed, semiautomatically designed and four automatically designed. The proposed work demonstrated exceedingly better classification accuracy in comparison with the existing methods. Serrano [101] has proposed a new GA where the concept of genome is combined with reinforcement and DL. It is used for management decision structures where learning is completed during transmission of information to new generations. This research is based on combining the concepts of GA, DL cluster algorithms and random neural network (RNN) to imitate the behaviour of human brain. The proposed genetic learning algorithm is validated in Fintech, a smart investment application and an intelligent banker application associated with buy and sell of the products involving the market risks, and satisfactory results are obtained.

In this section, we have observed several real-time applications where the combination of GA and ML is applied to produce efficient results. This study helps the researchers to identify the most popular practical applications of the combination of GA and ML. On the other hand, PSO, a well-known nature-based optimization technique, integrates with ML to solve a variety of optimization problems. Blending of these two dominant techniques is explained in the next section. Various realistic applications are considered where the amalgamation of PSO and ML is successfully applied.

## 2.2 *Particle Swarm Optimization Algorithm with Machine Learning*

Gradient descent, a well-known and a popular optimization algorithm, produces good results for convex functions and low-dimensional space. However, PSO, as considered one of the accepted algorithms, generates fantastic results for such problems. In swarm, there is a collection of particles which interacts with each other as an agent. They have a complete freedom to move in their search space as there is no central control on each other. In today's world, swarm-based algorithms are very popular to solve complex and NP-hard problem as this algorithm is very efficient to produce high-quality results in less computational time. Further, its popularity is getting increased after combining PSO with ML in order to solve real-world challenging problems. Table 2 reflects most of the latest work in ML domain employed with PSO. PSO algorithm is used to train an ANN to diagnose unexplained syncope in the paper discussed by Gao et al. [30]. The results demonstrate the better accuracy rate for diagnosis and fast convergence in comparison with GA and BP-based learning ANN. PSO is further combined with reinforcement learning algorithm in the research as presented by Lima et al. [48] where a revised procedure of PSO is identified and is applied to shortest path problems. In this paper, the agents in a swarm interact with each other. At the same time, the state action values are revised based on personal best and global best. Experimentally, the results are also validated.

**Table 2** A list of applications as per state of the art by combining PSO and ML

Application	Author	Year	References
To diagnose unexplained syncope	Gao et al.	2006	[30]
Shortest path problems	Lima et al.	2009	[48]
Time series forecasting	Neto et al.	2009	[84]
To optimize the input weights and hidden biases of single-hidden-layer feed-forward neural networks (SLFN)	Han et al.	2012	[37]
To detect breast cancer	Zhang et al.	2012	[127]
Gender classification of real-world face images	Nazir et al.	2014	[83]
Representations of feature construction	Dai et al.	2014	[22]
Accuracy detection for intrusion attacks	Bamakan et al.	2015	[8]
Designing artificial neural network (ANN)	Garro et al.	2015	[29]
Detecting travel mode	Xiao et al.	2015	[120]
To optimize DL parameters using PSO	Qolomany et al.	2017	[89]
High performing robot controllers	Mario et al.	2017	[74]
Hyperparameter selection method	Ye	2017	[124]
Twitter application	Jayasekara	2018	[50]
Web spamming	Singh et al.	2018	[97]
Predicting voltage instability	Ibrahim et al.	2018	[46]
To diminish lung nodule false positive on computed tomography scans	Silva et al.	2018	[98]
To develop a multi-criteria recommender system	Hamada et al.	2018	[36]
Real-world NP-hard problem	Ding et al.	2019	[24]
Tunnel settlement forecasting	Hu et al.	2019	[76]

Further, Neto et al. [84] proposed a model by integrating PSO and ANN to identify the solution to time series forecasting problem as ANN works better in forecasting systems where decision-making is involved. In this research, ANN parameters are adjusted efficiently by using PSO. Six real-world time series are considered for testing purpose to analyse the results. In the year 2012, Han et al. presented an improved PSO which is applied on extreme learning machine (ELM) to optimize the input weights and hidden biases of single-hidden-layer feed-forward neural networks (SLFN). The proposed work is found more efficient in comparison with the traditional ELM methods. Further, Zhang et al. [127] developed a novel NN classifier to detect breast cancer. The authors have improved the efficiency of the traditional classifier methods by combining floating centroid methods with PSO. Testing is accomplished using UCI ML data set. Also, a novel method is developed and proposed by Nazir et al. [83] for gender classification of real-world face images in an unconstrained manner. The local features of an image are extracted through local binary pattern (LBP). The classification accuracy rate is improved by merging the extracted features with clothing features. PSO algorithm and GA are selected to identify the optimal

number of features which are treated as an input for support vector machine (SVM). Experimentally, there is an improvement observed in classification accuracy rate by comparing with the existing methods.

Further, feature extraction is considered an important parameter by Dai et al. [22]. The authors proposed a novel technique of representations of feature construction and developed two representation techniques to overcome the limitations of the traditional approaches, i.e. PSOFCPair and PSOFCArray. Experimentally, their classification performance is improved by identifying a new high-level feature in contrast to the existing methods. Furthermore, PSO is combined with a technique of ML to improve the detection accuracy for intrusion attacks in Bamakan et al. [8]. A novel model is proposed by merging multiple criteria linear programming, a classification method with PSO. The proposed work is evaluated by considering the data set of KDD CUP 99. The results demonstrated better performance in comparison with two benchmark classifiers as mentioned in the paper. An automatic ANN is designed using PSO algorithm as described by Garro et al. [29]. Three PSO algorithms are employed in the research work, namely basic PSO, second-generation particle swarm optimization (SGPSO) and a new model of PSO called NMPSO. Experimentally, the proposed work exhibited better results in terms of efficiency as compared with conventional methods. Xiao et al. [120] proposed a model for detecting travel mode by merging the concept of NN with PSO algorithm. A travel survey is conducted based on a smartphone which considers four travel modes, specifically walk, bike, bus and car. The positioning data is collected through GPS for testing purpose which results in the improved accuracy in comparison with the contemporary methods. A research work is proposed by Qolomany et al. [89] to optimize DL parameters using PSO. In any DL network, two parameters are considered, i.e. number of layers in the network and number of neurons in each layer. The experimental results showcased the optimized tuning of these parameters in comparison with the grid search method.

Further, Mario et al. [74] proposed a research work on high performing robot controllers. The authors considered multi-robot obstacle avoidance as a benchmark optimization problem and compared the results of PSO with Q-learning. The results are exhibited good results for PSO for certain testing scenario and different evaluation parameters such as performance efficiency, total evaluation time and their overall behaviours. In recent work, Ye [124] worked on the hyperparameter selection method in order to optimize the values for network training phase. One of the important hyperparameters, namely learning rate, is considered in his research. An efficient way is proposed by integrating the advantages of PSO and steepest gradient descent algorithm which allows the model to automatically identify the optimized network structure for DNN and enables to optimize the hyperparameters as well. Several experiments are considered in this work which showcased the better results in contrast to the existing frameworks. In one of the latest reports by Jayasekara [50], ML and PSO are applied on a Twitter application. The important concept to consider for this application is feature selection as an optimization problem. There are several applications of feature selection such as data classification, image classification, cluster analysis, data analysis, image retrieval, opinion mining and review analysis.

Different methods such as wrapper method and filters are applied to solve this optimization problem. However, the best optimal results are obtained by PSO. Further, Tweet data clustering is completed by applying PSO algorithm after pre-processing. Experimental results validate the performance of PSO clustering quite satisfactory in comparison with hierarchical and partitioning clustering techniques.

There are different applications in which DL is applied. Singh et al. [97] also worked on the similar lines by considering an important application of Web spamming, one of the major challenges in search engines. Optimal feature selection method plays a significant role to reduce Web spamming. The authors described a novel method where PSO is used with the properties of correlation-based feature selection (CFS) technique to identify the relevant and optimal features. During experimentation, five classifiers are considered in Web spam-2006. Results indicate success towards the size of features and accuracy. One of the challenging applications of predicting voltage instability is considered in the paper by Ibrahim et al. [46]. In this paper, a powerful algorithm, namely recurrent neural network (RNN) [88], is investigated and PSO algorithm is applied to train RNN for projecting voltage instability. In order to validate the effectiveness of the proposed work, back-propagation (BP) [10] algorithm is applied to train RNN and results are compared. PSO also worked efficiently with CNN to diminish lung nodule false positive on computed tomography scans as proposed by Silva et al. [98]. The efficiency of the presented work is validated by considering two databases, i.e. Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI). Further, PSO is employed with ANN to develop a multi-criteria recommender system by Hamada et al. [36]. A multi-criteria data set of movie recommendation to users is selected for experimental purpose which demonstrates the high prediction accuracy in comparison with the recent approaches. Also, Ding et al. [24] projected the limitations of asynchronous and traditional reinforcement learning algorithm to solve a real-world problem. In order to address the limitations, the authors applied PSO to asynchronous reinforcement learning algorithm to generate the optimal solution to the problem. This novel version of the algorithm is referred to as asynchronous PSO. Further, the authors developed a new algorithm based on asynchronous PSO and backward Q-learning which is referred to as APSO-BQSA. The effectiveness of these algorithms is also evident in this paper. Tunnel settlement forecasting, one of the major challenges for construction companies to avoid unexpected disasters, is pointed out by Hu et al. [76] in their research work. By identifying the limitations of traditional forecasting methods, namely model-based methods and artificial intelligence (AI) enhanced methods, the authors extended the approach of AI by integrating the concept of PSO with support vector regression (SVR), back-propagation neural network (BPNN) and extreme learning machine (ELM). This work is validated experimentally in two large cities of China by forecasting the exterior completion of tunnel structure.

It could be observed through the state of the art that PSO is a powerful technique in order to solve the real-world optimization problems. However, by joining with the techniques of ML, it outperforms in several real-time applications and shows improvement in many ways. By observing the latest applications solved by PSO

and ML as depicted in Table 2, it opens a new path for researchers in terms of understanding, applying and solving the real-world complex problems.

### 3 Conclusion and Future Scope

The study reveals various real-world and practical applications where GA or PSO is applied with ML techniques to enhance the solution efficiency or to identify the optimized solution to a complex problem. In this discussion, the advantages of both the algorithms are evaluated and elaborated for solving a problem. As explained in the methodologies, Sect. 2, by combining ML either with PSO or GA, the drawbacks of each algorithm are sheltered; thus, improved solution is extracted. In this paper, a sincere effort is made to make the researchers to understand the concepts more thoroughly by citing and explaining the most of the modern research work related to these fields. In today's world, there are thousands of applications in banking and financial services, government services, education, health care, transportation, etc., which directly or indirectly uses optimization techniques and ML. Therefore, demand for improved optimization techniques along with security is always there. By harnessing the new power of ML along with highly scalable computing powers of today's computers, researchers can give a new direction to the world. Also, researchers can focus on security aspect of any algorithm and also other nature-inspired optimization techniques along with ML so that new possibilities can be explored.

## References

1. Abbass HA (2001) MBO: marriage in honey bees optimization-a Haplometrosis polygynous swarming approach. In: Proceedings of the 2001 Congress on evolutionary computation (IEEE Cat. No.01TH8546), 27–30 May 2001, IEEE, Seoul, South Korea
2. Abbeel P, Ng AY (2010) Inverse reinforcement learning. In: Sammut C, Webb GI (eds) Encyclopedia of machine learning. Springer, Boston, MA
3. Aihara, I. (2009): “Modeling synchronized calling behavior of Japanese tree frogs”, Physical Review E 80, 011918, pp 1–7
4. Azadeh A, Ghaderi SF, Tarverdian S, Saberi M (2007) Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Appl Math Comput 186(2):1731–1741
5. Askarzadeh A (2016) A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. Comput Struct 169:1–12
6. Baese AM, Schmid V (2014) Chapter 5—genetic algorithms. In: Pattern recognition and signal analysis in medical imaging 2nd edn. pp 135–149
7. Bhattacharya S, Prasad H, Prashanth L (2013) Algorithms for constrained optimization. In: Stochastic recursive algorithms for optimization, lecture notes in control and information sciences, 434, Springer, London, pp 167–186
8. Bamakan SMH, Amiric B, Mirzabagheri M, Sh Y (2015) A new intrusion detection approach using PSO based multiple criteria linear programming. Information technology and quantitative management (ITQM 2015). Procedia Computer Science 55:231–237

9. Baghmisheh MTV, Madani K, Navarbaf A (2011) A discrete shuffled frog optimization algorithm. *Artif Intell Rev* 36–267
10. Baughman DR, Liu YA (1995) 2-Fundamental and practical aspects of neural computing. In: Neural networks in bioprocessing and chemical engineering, pp 21–109
11. Bandura A (1962) Social learning through imitation. In: Jones MR (ed), Nebraska symposium on motivation, University of Nebraska Press, Lincoln
12. Bellemare MG, Naddaf Y, Veness J, Bowling M (2015) An evaluation platform for general agents. In: Proceedings of the twenty-fourth international joint conference on artificial intelligence, pp 4168–4152
13. Biyanto TR, Matradji, Irawan S, Febrianto HY, Afdanny N, Rahman AH, Gunawan KS, Pratama JAD, Bethiana TN (2017) Killer whale algorithm: an algorithm inspired by the life of killer whale. *Procedia Computer Science* 124:151–157
14. Brockman G, Cheung V, Pettersson L, Schneider J, Schulman J, Tang J, Zaremba W (2016) OpenAI Gym. *Machine Learning*, arXiv: 1606.01540
15. Bengio Y (2009) Learning deep architectures for AI. *Found Trends Mach Learn* 2:1–127
16. Burton A, Vladimirova T (1997) Genetic algorithm utilising neural network fitness evaluation for musical composition. Artificial neural nets and genetic algorithms. In: ICANNGA 97, Proceedings of the 3rd international conference in Norwich, GB, April 2–4
17. Carbone Y, Jacob C (2015) Genetic algorithm as machine learning for profiles recognition. In: 7th international joint conference on computational intelligence (IJCCI), 12–14 Nov. 2015, IEEE, Lisbon, Portugal
18. Chiroma H, Noor ASM, Abdulkareem S, Abubakar AI, Hermawan A, Qin H, Hamza MF, Herawa T (2017) Neural networks optimization through genetic algorithm searches: a review. *Appl Mathe Info Sci* II(6):1543–1564
19. Chu S, Tsai P, Pan J (2006) Cat swarm optimization. In: Pacific rim international conference on artificial intelligence, part of the lecture notes in computer science, 4099, pp 854–858
20. Civicioglu P (2013) Artificial cooperative search algorithm for numerical optimization problems. *Inf Sci* 229:58–76
21. Civicioglu P (2013) Backtracking search optimization algorithm for numerical optimization problems. *Appl Math Comput* 219(15):8121–8144
22. Dai Y, Xue B, Zhang M (2014) New representations in PSO for feature construction in classification. In: European conference on the applications of evolutionary computation, lecture notes in computer science. Springer, 8602, pp 476–488
23. Das S, Chowdhury A, Abraham A (2009) A bacterial evolutionary algorithm for automatic data clustering. IEEE congress on evolutionary computation, IEEE, Trondheim, Norway
24. Ding S, Xing WD, Zhao X, Wang L, Jia W (2019) A new asynchronous reinforcement learning algorithm based on improved parallel PSO. *Appl Intell* 1573–7497
25. Ding YR, Cai YJ, Sun PD, Chen B (2014) The use of combined neural networks and genetic algorithms for prediction of river water quality. *J Appl Res Technol* 12(3):493–499
26. Dorigo M, Blum C (2005) Ant colony optimization theory: a survey. *Theoret Comput Sci* 344(2–3):243–278
27. Eglese RW (1990) Simulated annealing: a tool for operational research. *European J Oper Res* 46(3):15, 271–281
28. Emami H, Derakhshan F (2015) Election algorithm: a new socio-politically inspired strategy. *AI Commun* 28(3):591–603
29. Garro BA, Vázquez RA (2015) Designing artificial neural networks using particle swarm optimization algorithms. *Comput Intell Neurosci* (<https://doi.org/10.1155/2015/369298>)
30. Gao L, Zhou C, Gao HB, Shi YR (2006) Combining particle swarm optimization and neural network for diagnosis of unexplained syncope. *Int Conf Intell Comput Part Lecture Notes Comput Sci Book Series* 4115:174–181
31. Gargari EA, Lucas C (2007) Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: Evolutionary computation, CEC, 2007 IEEE Congress, IEEE, Singapore, pp 4661–4667

32. Geem ZW (2010) State-of-the-art in the structure of harmony search algorithm. In: Recent advances in harmony search algorithm, studies in computational intelligence. Springer, Berlin, Heidelberg
33. Ghahramani Z (2003) Unsupervised learning. Summer school on machine learning. In: Advanced lectures on machine learning, lecture notes in computer science, 3176, pp 72–112
34. Goldberg DE, Holland JH (1988) Genetic algorithms and machine learning. *Mach Learn* 3(2–3):95–99
35. Goldberg, D.E., Deb, K. (1991): “A comparative analysis of selection schemes used in genetic algorithms”, in: Foundations of Genetic Algorithms, 1, Morgan Kaufmann Publishers Inc, pp. 69–93
36. Hamada M, Hassan M (2017) Artificial neural networks and particle swarm optimization algorithms for preference prediction in multi-criteria recommender systems. *informatics* 5(25):1–16
37. Han F, Yao HF, Ling QH (2011) An Improved extreme learning machine based on particle swarm optimization. *Int Conf Intell Comput Bio-Inspired Comput Appl* 6840:699–704
38. Harbi SH, Smith VJR (2006) Adapting k-means for supervised clustering. *Appl Intell* 24(3):219–226
39. Harpham C, Dawson CW, Brown MR (2004) A review of genetic algorithms applied to training radial basis function networks. *Neural Comput Appl* 13(3):193–201
40. Harifi S, Khalilian M, Mohammadzadeh J, Ebrahimnejad S (2019) Emperor Penguins colony: a new metaheuristic algorithm for optimization. *Evol Intel* 12(2):211–226
41. Hsieh JC, Chang PC, Chen SH (2006) Integration of genetic algorithm and neural network for financial early warning system: an example of Taiwanese Banking Industry. In: First international conference on innovative computing, information and control, 1, 30 Aug.–1 Sept. 2006, IEEE, Beijing, China
42. Hosseini HS (2009) The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm. *Int J Bio-Inspired Comput* 1(1/2):71–79
43. Huan TT, Kulkarni AJ, Kanesan J (2016) Ideology algorithm: a socio-inspired optimization methodology. *Neural Comput Appl* 1–32. (<https://doi.org/10.1007/s00521-016-2379-4>)
44. Huang CH (2013) Engineering design by geometric programming. *Mathematical problems in engineering*, 2013, Article ID 568098, pp 1–8
45. Hussain K, Salleh MNM, Cheng S, Shi Y (2018) Metaheuristic research: a comprehensive survey. *Artif Intell Rev* (<https://doi.org/10.1007/s10462-017-9605-z>)
46. Ibrahim AM, El-Amry NH (2018) Particle Swarm Optimization trained recurrent neural network for voltage instability prediction. *J Electr Syst Inform Technol* 5(2):216–228
47. Indolia S, Goswami AK, Mishra SP, Asopa P (2018) Conceptual understanding of convolutional neural network-a deep learning approach. *Procedia Comput Sci* 132:679–688
48. Iima H, Kuroe Y (2009) Swarm reinforcement learning algorithm based on particle swarm optimization whose personal bests have lifespans. In: International conference on neural information processing, part of the lecture notes in computer science book series, 5864, pp 169–178
49. Javid AA (2011) Anarchic society optimization: a human-inspired method. In: Evolutionary computation, CEC, 2011 IEEE Congress, IEEE, New Orleans, USA, pp 2586–2592
50. Jayasekara D (2018) Machine learning—particle swarm optimization (PSO) and Twitter, <https://medium.com/pythondatasciencezerotohero/machine-learning-particle-swarm-optimization-pso-and-twitter-c952a9ace499>
51. Jones AJ (1993) Genetic algorithms and their applications to the design of neural networks. *Neural Comput Appl* 1(1):32–45
52. Jolliffe IT (2002) Introduction. In: Principal component analysis, Springer series in statistics. Springer, New York, NY, pp 1–9
53. Joshi AS, Kulkarni O, Kakandikar GM, Nandedkar VM (2017) Cuckoo search optimization-a review. *Mater Today: Proc* 4(8):7262–7269
54. Jung SH (2003) Queen-bee evolution for genetic algorithms. *Electron Lett* 39(6):575–576

55. Kashan AH (2009) League championship algorithm: a new algorithm for numerical function optimization. In: International conference on soft computing and pattern recognition, SOCPAR09, IEEE, Singapore, pp 43–48
56. Kaur R, Singh B, Gobindgarh I, Sahib BF, Sahib F (2011) A Hybrid neural approach for character recognition system. *Int J Comput Sci Inform Technol* 2(2):721–726
57. Kaboudan MA (2000) Genetic programming prediction of stock prices. *Comput Econ* 16(3):207–236
58. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks, 4, IEEE, pp 1942–1948
59. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *J Global Optim* 39:459–471
60. Krausmann E, Cruz AM, Salzano E (2017) Chapter 14—reducing natech risk: organizational measures. *Natech Risk Assessment and Management, Reducing the Risk of Natural-Hazard Impact on Hazardous Installations*, pp 227–235
61. Krishnanand KN, Ghose D (2006) Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications. *Multiagent Grid Syst Int J* 2:209–222
62. Kotsiantis SB (2013) Decision trees: a recent overview. *Artif Intell Rev* 39(4):261–283
63. Kleining G, Witt H (2000) The Qualitative Heuristic approach: a methodology for discovery in psychology and the social sciences. Rediscovering the method of introspection as an example. *Forum Q Soc Res* 1(1), Article 13
64. Koehn, P. (1994): “Combining Genetic Algorithms and Neural Networks: The Encoding Problem”, A Thesis Presented for the Master of Science Degree The University of Tennessee, Knoxville, pp 1–67
65. Kumar M, Kulkarni AJ, Satapathy SC (2018) Socio evolution & learning optimization algorithm: a socio-inspired optimization methodology. *Fut Generation Comput Syst* 81:252–272
66. Kulkarni AJ, Durugkar IP, Kumar M (2013) Cohort intelligence: a self supervised learning behavior. In: Systems, man, and cybernetics, SMC, IEEE international conference. IEEE, Manchester, UK, pp 1396–1400
67. Kumar V, Chhabra JK, Kumar D (2015) Differential search algorithm for multiobjective problems. *Procedia Comput Sci* 48:22–28
68. Kuo HC, Lin CH (2013) Cultural evolution algorithm for global optimizations and its applications. *J Appl Res Technol* 11(4):510–522
69. Lindfield GR, Penny JET (2012) 8—optimization methods. In: Numerical methods 3rd edn. Science Direct, pp 371–432
70. Lazovskiy V (2018) Travel time optimization with machine learning and genetic algorithm. Towards Data Science (<https://towardsdatascience.com/travel-time-optimization-with-machine-learning-and-genetic-algorithm-71b40a3a4c2>)
71. Louveaux Q, Skutella M (2016) Integer programming and combinatorial optimization. In: 18th international conference, IPCO 2016, Liège, Belgium, June 1–3, 2016, proceedings, part of the lecture notes in computer science, 9682
72. Lee CV, Piramuthu S, Tsai YK (2010) Job shop scheduling with a genetic algorithm and machine learning. *Int J Prod Res* 35(4):1171–1191
73. Man KF, Tang KS, Kwong S (1996) Genetic algorithms: concepts and applications [in engineering design]. *IEEE Trans Industr Electron* 43(5):519–534
74. Mario ED, Talebpour Z, Martinoli A (2013) A Comparison of PSO and reinforcement learning for multi-robot obstacle avoidance. IEEE Congress on Evolutionary Computation, Cancún, México, June 20–23, 2013, pp 149–156
75. Marinakis Y, Marinaki M (2014) A bumble bees mating optimization algorithm for the open vehicle routing problem. *Swarm Evol Comput* 15:80–94
76. Min Hu M, Li W, Yan K, Ji Z, Hu H (2019) Modern machine learning techniques for univariate tunnel settlement forecasting: a comparative study. In: Mathematical problems in engineering, Hindawi, 2019, pp 1–12

77. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G et al (2015) Human-level control through deep reinforcement learning. *Nature* 518:529–541
78. Moerland TM, Broekens J, Jonker CM (2018) Emotion in reinforcement learning agents and robots: a survey. *Mach Learn* 107(2):443–480
79. Moosavian N (2015) Soccer league competition algorithm for solving knapsack problems. *Swarm Evol Comput* 20:14–22
80. Mousavi SS, Schukat M, Howley E (2016) Deep reinforcement learning: an overview. In: Proceedings of SAI intelligent systems conference, Lecture notes in networks and systems, 16, pp 426–440
81. Mulder WD, Bethard S, Moens MF (2015) A survey on the application of recurrent neural networks to statistical language modelling. *Comput Speech Lang* 30(1):61–98
82. Nanda SJ, Panda G (2014) A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm Evol Comput* 16:1–18
83. Nazir M, Majid-Mirza A, Khan SA (2014) PSO-GA based optimized feature selection using facial and clothing information for gender classification. *J Appl Res Technol* 12(1):5–163
84. Neto PSG, Petry GG, Aranildo RLJ, Ferreira TAE (2009) Combining artificial neural network and particle swarm system for time series forecasting. In: International joint conference on neural networks, IEEE, 14–19 June 2009, Atlanta, GA, USA
85. Ng AY (2006) Reinforcement learning and apprenticeship learning for robotic control. In: International conference on algorithmic learning theory, lecture notes in computer science, 4264, pp 29–31
86. Parpinelli RS, Lopes HS (2011) An eco-inspired evolutionary algorithm applied to numerical optimization. In: Third world congress on nature and biologically inspired computing, 19–21 Oct. 2011, IEEE, Salamanca, Spain
87. Pham D, Ghanbarzadeh A, Koç E, Otri S, Rahim S, Zaidi M (2005) The Bees algorithm technical note. Manufacturing Engineering Centre, Cardiff University, UK, pp 1–57
88. Poznyak TI, Oria IC, Poznyak AS (2019) Chapter 3—Background on dynamic neural networks. Ozonation and Biodegradation in Environmental Engineering, Dynamic Neural Network Approach, pp 57–74
89. Qolomany B, Maabreh M, Al-Fuqaha A, Gupta A, Benhaddou D (2017) Parameters optimization of deep learning models using particle swarm optimization. In: 13th international wireless communications and mobile computing conference (IWCMC), IEEE, 26–30 June 2017, Valencia, Spain
90. Rabanal P, Rodríguez I, Rubio F (2017) Applications of river formation dynamics. *J Comput Sci* 22:26–35
91. Rashedi E, Pour HN, Saryazdi S (2009) GSA: a gravitational search algorithm. *Inf Sci* 179(13):2232–2248
92. Rao RV, Savsani VJ, Vakharia DP (2012) Teaching–learning-based optimization: an optimization method for continuous non-linear large scale problems. *Inform Sci* 183(1):1–15
93. Robbins GE, Plumbley MD, Hughes JC, Fallside F, Prager R (1993) Generation and adaptation of neural networks by evolutionary techniques (GANNET). *Neural Comput Appl* 1(1):23–31
94. Rosenberg L (2016) Artificial Swarm Intelligence vs human experts. In: International joint conference on neural networks (IJCNN), 24–29 July 2016, IEEE, Vancouver, BC, Canada
95. Sang HY, Duan PY, Li JQ (2018) An effective invasive weed optimization algorithm for scheduling semiconductor final testing problem. *Swarm Evol Comput* 38:42–53
96. Satapathy S, Naik A (2016) Social group optimization (SGO): a new population evolutionary optimization technique. *Complex Intel Syst* 2(3):173–203
97. Singh S, Singh AK (2018) Web-spam features selection using CFS-PSO. *Procedia Comput Sci* 125:568–575
98. Silvaa GLF, Valente TLA, Silvaa AC, Paivaa ACD, Gattass M (2018) Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. *Comput Methods Programs Biomed* 162:109–118

99. Schulman J, Levine S, Abbeel P, Jordan M, Moritz P (2015) Trust region policy optimization. In: Proceedings of the 32nd international conference on machine learning, PMLR, 37, pp 1889–1897
100. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. arXiv preprint arXiv: 1707.06347
101. Serrano W (2019) Genetic and deep learning clusters based on neural networks for management decision structures. *Neural Comput Appl* 1–25 <https://doi.org/10.1007/s00521-019-04231-8>
102. Serrano W (2018) The random neural network with a genetic algorithm and deep learning clusters. In: Fintech: Smart Investment, Imperial College London
103. Shapiro AF (2002) The merging of neural networks, fuzzy logic, and genetic algorithms. *Insurance Mathe Econ* 31(1):115–131
104. Shafti LS, Pérez E (2004) Machine learning by multi-feature extraction using genetic algorithms. In: Ibero-American Conference on Artificial Intelligence, Part of the Lecture Notes in Computer Science, 3315, pp 246–255
105. Shi Y (2011) Brain storm optimization algorithm. In: International conference in Swarm Intelligence, advances in swarm intelligence, part of the lecture notes in computer science, 6728, pp 303–309
106. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359
107. Such FP, Madhavan V, Conti E, Lehman J, Stanley KO, Clune J (2018) Deep Neuroevolution: genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. *Neural Evol Comput* 1–14
108. Sun Y, Zhang M, Yen GG (2019) Automatically designing CNN architectures using genetic algorithm for image classification. arXiv: 1808.03818 v2, pp 1–12
109. Sun J, Zhang H, Zhang Q, Chen H (2018) Balancing exploration and exploitation in multiobjective evolutionary optimization. In: Proceedings of the genetic and evolutionary computation conference companion, GECCO’18, Kyoto, Japan—July 15–19, 2018
110. Suryansh S (2018) Genetic algorithms + neural networks = best of both worlds. Towards Data Science (<https://towardsdatascience.com/gas-and-nns-6a41f1e8146d>)
111. Todorov E, Erez T, Tassa Y (2012) MuJoCo: A physics engine for model-based control. In: international conference on intelligent robots and systems (IROS) (<https://doi.org/10.1109/iros.2012.6386109>)
112. Tian H, Pouyanfar S, Chen J, Chen SC, Iyengar SS (2018) Automatic convolutional neural network selection for image classification using genetic algorithms. In: IEEE international conference on information reuse and integration (IRI), 6–9 July 2018, IEEE, Salt Lake City, UT, USA
113. Venter G (2010) Review of optimization techniques. In: Encyclopedia of aerospace engineering. Wiley
114. Vizitiu I, Popescu F (2010) GANN system to optimize both topology and neural weights of a feed forward neural network. (<https://doi.org/10.1109/icomm.2010.5509105>)
115. Watanabe Y, Mizuguchi N, Fujii Y (1998) Solving optimization problems by using a Hopfield neural network and genetic algorithm combination. *Syst Comput Japan* 29:68–74
116. Webb GJ (2010) Naïve Bayes. In: Encyclopedia of machine learning, pp 30–45
117. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67–82
118. Wu B (1992) An introduction to neural networks and their applications in manufacturing. *J Intell Manuf* 3(6):391–403
119. Xia X, Lo D, Wang X, Yang X, Li S, Sun J (2013) A comparative study of supervised learning algorithms for re-opened bug prediction. In: 17th European conference on software maintenance and reengineering, 5–8 March 2013, IEEE, Genova, Italy
120. Xiao G, Juan Z, Gao J (2015) Travel mode detection based on neural networks and particle Swarm optimization. *Information* 6:522–535

121. Xu Y, Cui Z, Zeng J (2010) Social emotional optimization algorithm for nonlinear constrained optimization problems. In: Swarm, evolutionary, and memetic computing, SEMCCO 2010. In: Lecture notes in computer science, 6466, Springer, Berlin, Heidelberg, pp 583–590
122. Yang X (2010) A new metaheuristic bat-inspired algorithm. *Nature Inspired Cooperative Strategies for Optimization* (NICSO 2010). *Stud Comput Intell* 284:65–74
123. Yang, X.S. (2012): “Flower Pollination Algorithm for Global Optimization”, International Conference on Unconventional Computing and Natural Computation, Part of the Lecture Notes in Computer Science, 7445, pp 240–249
124. Ye F (2017) Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data. *Plos One*, 12(12) (<https://doi.org/10.1371/journal.pone.0188746>)
125. Youcai Z, Sheng H (2017) Chapter four—pollution characteristics of industrial construction and demolition waste. In: *Pollution control and resource recovery, industrial construction and demolition wastes*, pp 51–101
126. Zang H, Zhang S, Hapeshi K (2010) A review of nature-inspired algorithms. *J Bionic Eng* 7(Supplement):S232–S237
127. Zhang L, Wang L, Wang X, Liu K, Abraham A (2012) Research of neural network classifier based on FCM and PSO for breast cancer classification. In: *International conference on hybrid artificial intelligence systems*, part of the lecture notes in computer science book series, 7208, pp 647–654
128. Zhang Y (2012) Support vector machine classification algorithm and its application. *Int Conf Info Comput Appl Commun Comput Info Sci* 308:179–186
129. Zou J, Han Y, So SS (2008) Overview of artificial neural networks. In: DJ Livingstone (eds) *Artificial neural networks, methods in molecular biology™*, 458, Human Press
130. Zhang S, Zong M, Sun K, Liu Y, Cheng D (2014) “Efficient kNN algorithm based on graph sparse reconstruction. In: *International conference on advanced data mining and applications*, ADMA 2014, lecture notes in computer science, 8933, pp 356–369

# Chapter 7

# A Hybridized Data Clustering for Breast Cancer Prognosis and Risk Exposure Using Fuzzy C-means and Cohort Intelligence



Meeta Kumar, Anand J. Kulkarni and Suresh Chandra Satapathy

## 1 Introduction

Cancer is a type of disease in which cells in the body grow and mutate in an unorderly and uncontrollable manner triggered due to certain genetic abnormalities. This uncontrollable growth of cells may result in the formation of a mass of cells called as tumor; which may be malignant or a benign tumor. Cancer as a disease is one of the leading causes for deaths in the world and can be classified into different types depending on the area of the body where they origin from and the type of the cell they comprise (or resemble). Breast cancer, the most diagnosed cancer types in females, is a result of abnormal and unruly growth of the cells in the breast tissues. The cells speedily segregate from a bulge of extra tissue, called tumor. This tumor can be either malignant (cancerous) or benign (non-cancerous) in nature. As per a recent article on global cancer statistics by Bray et al. [7], breast cancer is the leading causes of cancer death in many countries, and around 2.1 million women between the age group 40–55 year of age were diagnosed worldwide with this disease in 2018 alone. Studies also establish that the major contributing factors for breast cancer are

---

M. Kumar · A. J. Kulkarni (✉)

Symbiosis Institute of Technology, Symbiosis International University, Pune, MH 412115, India  
e-mail: [anand.kulkarni@sitpune.edu.in](mailto:anand.kulkarni@sitpune.edu.in); [kulk0003@uwindsor.ca](mailto:kulk0003@uwindsor.ca)

M. Kumar

e-mail: [meeta.kumar@sitpune.edu.in](mailto:meeta.kumar@sitpune.edu.in)

A. J. Kulkarni

Odette School of Business, University of Windsor, 401 Sunset Avenue, Windsor, ON N9B3P4,  
Canada

S. C. Satapathy

Department of Computer Science and Engineering, PVP Siddhartha Institute of Technology,  
Vijayawada, AP, India  
e-mail: [sureshsatapathy@ieee.org](mailto:sureshsatapathy@ieee.org)

gender and the age, with genetic mutations set off as a result of the aging process and lifestyle changes rather than from hereditary factors.

Fortunately, the mortality rate has deteriorated in the recent years, with improved prognostic and diagnostic techniques and more effective treatment and medicines. Earlier, clinical approaches like mammography, surgical biopsy, magnetic resonance imaging (MRI) and fine needle aspiration (FNA) were practiced, among which mammography is highly recommended in order to predict and diagnose breast cancer [12]. Detection of seeded region (tumor mass) and distinguishing them from the background tissues using morphological operators with image processing and image segmentation methods were used. Image processing and image segmentation methods [16, 34] focus primarily on abnormality detection from a mammogram. These methods rely on some form of preprocessing techniques which filter the noise from a mammogram. Various researches propose different image segmentation techniques which separate out the region of interest (RoI) (i.e., the probable area where the tumor could be concentrated). This is then followed by image processing techniques to identify/classify abnormalities in the ROI [17]. Over the past years, researchers have augmented the techniques of data mining and/or machine learning with clinical methods to improve the quality of results obtained for breast cancer prediction/detection. Machine learning [18, 27] is a branch of artificial intelligence, where machines generate the competence to learn on its own without being programmed. Machine learning methods [2, 3, 33] like Support Vector Machine, Decision Tree, Naïve Bayes, Artificial Neural Network [1, 5], Bayesian Network are being extensively used for applications focusing on early detection and prognosis of cancer [19] and its type. Enrichment in the data collected from different sources, and robust data mining methods play a crucial part in medical field. Machine learning algorithms extensively use data mining techniques for building models to predict the future outcome of given data. Data mining [13] is analytical process of exploring and retrieval of useful hidden pattern or relationships between the variables on a given dataset. These consistent patterns and findings are then validated by applying the detected pattern to new subset of data.

Clustering and classification are popular data mining methods that have been used on cancer datasets for early detection of cancer in patients on the basis of parameters in the datasets [29]. A popular classification technique K-nearest neighbor (KNN) is used for cancer prognosis [26, 28] which is dependent on the number of neighbors and percentage of data used. KNN is highly dependent on distance measure (for example, Euclidean and Manhattan distance). They are effective in terms of classification and performance but are often time consuming in nature. Data clustering techniques partition the set of unstructured data objects into clusters [31] where the clusters are formed in such a way that the data objects within one cluster exhibit more similarities than to the objects of another cluster. Data clustering may be categorized as hierarchical or partitional clustering. Hierarchical clustering [11] groups objects into connected tree-like structure in which clusters are connected to each other. The hierarchical clustering is further divided as top-down (divisive) and bottom-up (agglomerative) approach. The article by Jain et al. [15] summarizes the advantages of hierarchical clustering. Firstly, no advance initialization of number of

clusters and secondly the cluster formation is not dependent of initial conditions. However, hierarchical clustering is passive in nature, i.e., a data object appoints to a cluster cannot move to another. Also, the overlapping of clusters cannot be eliminated due to lack of knowledge about clusters initial shape and size. With partitional clustering [24], only one set of cluster is created, where the data is grouped into several disjoint groups of cluster. Thus, partitional clustering technique is best suited for larger datasets. The advantages of hierarchical clustering happen to be the disadvantages of partitional clustering and vice versa. A commonly used partitional algorithm K-means [14], which is a hard clustering technique, fast and simple in nature, is used for partitioning the data for the minimization of mean square error measure. Although it is a popular method for clustering, it faces few limitations. Firstly, its performance is dependent on initial centroid choice, secondly, the objective function is not convex, and hence, it contains local minima as well as local maxima. Also, it is highly sensitive toward noises and outliers. K-means clustering [10] was used to assess the impact of clustering on the Breast Cancer Wisconsin (BCW) diagnostic dataset, which is a popular dataset used across various researches focusing on detection and prognosis of breast cancer.

Another clustering algorithm is fuzzy C-means (FCM) [6, 32] is an iterative method of clustering, most frequently used in pattern recognition and image processing. It is based on the concept of minimization of objective function to achieve the least-squared error. FCM is an unsupervised soft clustering method, in which data objects on the limits between classes may not fully belong to a single class. Every data point is assigned with a membership value between 0 and 1 conforming to each cluster on the basis of the distance between the cluster center and data point [9, 35]. It employs fuzzy partition such that the data point can belong to all groups with this degree, and after every iteration, the membership degree and cluster centers get updated. Factors like distance measures, cluster shape and scattering of data points in 2D space make fuzzy C-means more suited to larger datasets. Fuzzy C-means is a better classification method in comparison with k-means [8, 30] because of its unsupervised soft clustering nature. FCM is also dependent on center initialization, and it is more convoluted in computation than k-means. It also takes higher computation time. To overcome certain limitations of FCM, in the recent years, hybrid version of this clustering techniques is being proposed. Certain heuristic approaches are also used to optimize the performance of traditional FCM. For example, genetic algorithm (GA) is hybridized with various machine learning and data mining methods. An article by Jain et al. [23] discusses a system using the concept of fuzzy with genetic algorithm referred as genetic fuzzy rule-based system (GFRBS) is proposed. The system attains high performance and provides the interpretable feature. We present a Table 1 to enlist a few relevant researches on breast cancer prognosis which use various machine learning/data mining approaches and hybrid methods.

The current work proposes a hybrid fuzzy C-means method compounded with an optimization technique for robust and superior data clustering to overcome possible limitations of FCM. The cohort intelligence (CI) optimization algorithm [20, 21] is used in the current work to hybridize the traditional FCM, and then, the performance of this hybrid methodology is validated using the Wisconsin breast cancer dataset.

**Table 1** Literature on different methods for breast cancer prediction

Authors	Year	Methodology proposed in article for breast cancer prognosis/diagnosis
Pawlovsky et al.	2017	GA is used for component selection to improve the accuracy of a KNN method when using it for breast cancer prognosis
Alzubaidi et al.	2016	Uses hybrid approach to detect breast cancer which uses GA for feature selection and machine learning classifiers, KNN and support vector machine
Dubey et al.	2016	K-means clustering is applied on the breast cancer Wisconsin dataset, and analysis was done using different settings for parameters like centroid, distance, split method, number of iterations, etc.
Pourmandia and Addeh	2015	Article presents a diagnostic system using hybrid of fuzzy C-means and optimized neural network (ONN)
Bethapudi	2015	Uses GA with a 3-fold cross validation to classify benign and malignant breast cancer
Pawlovsky and Nagahashi	2014	KNN method is used for prognosis; GA is used for component selection
Tintu	2013	Breast cancer diagnosis on Wisconsin prognostic breast cancer datasets using fuzzy C-means; also dimensionality reduction of the features is used
Muhic	2013	Breast cancer detection is done using FCM for classification of the data and a pattern recognition model
Suganya and Shanthi	2012	Classify the benign and malignant breast cancer using fuzzy C-means algorithm and pattern recognition model
Banu et al.	2012	Prediction of breast cancer in mammogram image using support vector machine classifier; continuous wavelet transform is used as feature selection technique
Akay	2009	Support vector machine used for breast cancer diagnosis, which works on a selected subset of relevant features
Basha and Prasad	2009	Uses morphological operators for image segmentation; segmented regions are then processed using FCM for breast cancer diagnosis
Kermani et al.	1995	Uses neural networks for prognosis of breast cancer; feature selection and extraction are done by GA

In current research, a new novel hybridized data clustering model referred as fuzzy cohort intelligence algorithm (FCI) is proposed. The projected algorithm converges more swiftly and attains more precise solutions avoiding getting trapped in the local minima. The rest of the article is organized as follows: Sect. 2 discusses the algorithmic framework of the proposed technique, also focusing on the basic FCM method and the working of the CI optimizer. Section 3 details the experimental findings and compares the performance of the proposed hybridized FCI with the traditional FCM method. Finally, Sect. 4 concludes the study.

## 2 The Methodology: Fuzzy-CI

### 2.1 The Fuzzy C-means Clustering Algorithm

The FCM is an autonomous partitional clustering method in which data objects lying on the borderline in between the classes may not entirely belong to a single class. The data points located on the cluster boundaries are not mandated to belong to a particular cluster. They may be members of multiple clusters, and a fuzzy membership value determines their degree of association with a certain cluster. Every data object is assigned a membership value between [0, 1] based on the distance of the data point with a cluster center.

Let  $X = [x_1, x_2, \dots, x_N]$ , where  $x_i \in X^D$ , be the set of  $N$  data objects that are to be clustered and  $C = [c_1, c_2, \dots, c_A]$  be the set of  $A$  clusters represented by their centers and  $U = [u_1, u_2, \dots, u_N]$  be the set of fuzzy membership function for  $A$  clusters. In this procedure, each data in set  $X$  will be assigned in one of the  $A$  cluster in such a manner that it minimizes the objective function. The objective function is the summation of membership function  $u_{ij}$  and squared Euclidean distance between each  $x_i$  and  $c_j$ . This objective function is defined as

$$S = F(U, C) = \sum_{i=1}^N \sum_{j=1}^A \min \left\{ u_{ij}^m \|x_i - c_j\|^2 \right\}, 1 \leq m < \infty \quad (1)$$

where

- $m$  is a real number that works like the fuzziness index,  $m \in [1, \infty]$ .
- $c_j \neq \emptyset, \forall j \{1, 2, \dots, A\}$  is the center of cluster  $j$ .
- $u_{ij}$  is the fuzzy membership function which represents the membership of the ‘ $i$ th’ data point to the ‘ $j$ th’ cluster center.
- $k = \{1, 2, \dots, n\}$ , where  $n$  is number of iterations.

The method uses fuzzy partition such that the data point is acceptable to all clusters with this degree and the membership degree, and cluster centers get updated with every iteration. Thus, the algorithm performs fuzzy partitioning through iterative optimization, and the partitions will become fuzzier with increasing  $m$ . This iterative process continues till there is improvement in the computed values of the objective function. The process stops when this improvement between the current and the previous iteration is below a threshold value  $\varepsilon$  (where  $0 < \varepsilon < 1$ ). The fuzzy membership function  $u_{ij}$  can be defined as

$$u_{ij} = \frac{1}{\sum_{k=1}^A \left( \frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

and

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m * x_i}{\sum_{i=1}^N u_{ij}^m}, \quad c_j \text{ represents } j\text{th cluster center.} \quad (3)$$

## 2.2 Cohort Intelligence Optimization Algorithm

Cohort intelligence (CI) is a promising optimizer algorithm belonging to the class of socio-inspired metaheuristics. It is based on the idea where candidates (agents) in a cohort may interact and/or compete with other candidates in the cohort to evolve and achieve certain shared goals. The strength of the algorithm lies in its decentralized behavior where the candidates may choose to learn and evolve by observing the behavior of other possible better behaving candidates. The CI algorithm begins by initializing algorithmic parameters and then generates an initial population randomly.

Every candidate in the population is represented using his qualities (the problem variables) and the associated behavior of that candidate (solution vector or the objective function). The optimization process begins with every candidate calculating its own behavior; it may then choose to follow a better behaving candidate. This choice of which candidate behavior is to be followed is simulated using the probabilistic Roulette wheel selection approach (RWS). Once a candidate decides that it will follow a certain candidate from the cohort, it updates its qualities in the close neighborhood of this candidate. The iterative process of learning continues till no significant improvement is seen in the cohort behavior, or other terminating conditions are met. The optimizer is then said to have saturated and converged.

## 2.3 Hybridized Fuzzy Cohort Intelligence

The current work presents a hybridized algorithm referred to as fuzzy cohort intelligence (fuzzy-CI) which hybridizes the basic fuzzy C-means algorithm with the optimizer CI algorithm with an aim to improve the clusters and hence generated. The hybrid methodology attempts to optimize and thus minimize the objective function of FCM resulting in improved data cluster formation for a given dataset/clustering problem at hand. An optimized objective function indicates optimized centroids and better partitioning of data and thus better recognition of patterns in a dataset. The hybridized algorithmic approach may be used to optimize cluster formation for augmenting the prediction accuracy. The amalgamation allows the proposed algorithm to converge more rapidly and attain a more precise solution by avoiding getting stuck in local minima.

### Steps of Fuzzy-CI Algorithm

Consider the objective function of basic FCM which needs to be minimized, given as Eq. (1). The CI optimizer attempts to optimize this equation. In the current study,

**Table 2** Initial value of parameters for fuzzy-CI

Control parameters	Initial values
Maximum number of iterations (max_iterations)	3000
Number of candidates ( $Z$ )	05
Number of clusters formed for each candidate ( $C$ )	03
Fuzzy exponent ( $m$ )	02
Minimum improvement ( $\varepsilon$ )	1e-5
Sampling interval factor ( $r$ )	0.9500–0.9995

when CI is being applied to a data clustering problem, the set of clusters  $C = [c_1, c_2, \dots, c_A]$  represent the features/qualities of every candidate  $z$  and the objective function  $F(U, C)$  (Eq. 1) represents the behavior of a candidate. The CI begins with the initialization of its parameters listed in Table 2.

**Step 1:** Initialize the number of candidates  $Z$ , number of iterations  $n$ , sampling interval factor  $r \in [0, 1]$ , convergence parameter or the minimum improvement parameter  $\varepsilon$ . The values of the control parameters in Table 2 have been chosen based on initial trials carried out on the algorithm. In the next step, every candidate then generates and computes its set of clusters.

**Step 2:** Randomly generate the initial candidates  $Z$  as described below:

$$\text{Candidates} = \begin{bmatrix} S^1 \\ S^2 \\ S^3 \\ \vdots \\ S^Z \end{bmatrix} \quad (4)$$

where

$$S^z = F(u^z, c^z) = [c_1^z, c_2^z, \dots, c_j^z, \dots, c_A^z] \quad (5)$$

$$c_j^z = [x_1^z, x_2^z, \dots, x_j^z, \dots, x_D^z] \quad (6)$$

where  $c_j^z$  represents the ‘jth’ cluster center of a candidate  $z$  ( $z = [1, 2, 3, \dots, Z]$ ),  $A$  is the number of clusters,  $j = [1, 2, 3, \dots, A]$ , and  $D$  is the dimension of cluster center  $c_j^z$ . Therefore,

$$S^z = [x_1^z, x_2^z, \dots, x_j^z, \dots, x_D^z]_{1 \times b}, \text{ where } b = A \times D \quad (7)$$

For each candidate, randomly generate the initial cluster centers  $A$ , described in Eq. (2), where  $C_A = [c_1, c_2, \dots, c_A]$ .

**Step 3:** Every candidate then calculates its fuzzy membership measure  $u_{ij}$ , using Eq. (2), for  $k = 1, 2, 3 \dots, n$ ,  $i = 1, 2, 3 \dots, N$  and  $j = 1, 2, 3 \dots, A$ .

**Step 4:** Each candidate determines its new cluster centers, using Eq. (3), described as  $[c_1^Z, c_2^Z, \dots, c_j^Z]$ , where  $z = 1, 2, 3, \dots, Z$  and  $j = 1, 2, 3, \dots, A$ . For example, if  $A = 3$ , new clusters formed by candidate  $z(1)$  may be represented as  $z_1 = [c_1^1, c_2^1, c_3^1]$  and for candidate  $z(2)$  as  $z_2 = [c_1^2, c_2^2, c_3^2]$ .

**Step 5:** At each iteration, every candidate computes its objective function  $S^z$  using Eq. (1), where  $z = 1, 2, 3, \dots, Z$ , which represents the overall behavior of a specific candidate in the cohort at an iteration  $n$ .

**Step 6:** Every candidate in the cohort instinctively attempts to enhance its behavior by updating its behavior. This is done by a candidate by observing the behavior of other candidates in the cohort as well as itself. It may then choose to follow the behavior of a better behaving candidate. The probability  $p^z$  of choosing the behavior  $S^z$  of a candidate  $z$  is calculated using

$$p^z = \frac{1/S^z}{\sum_{z=1}^Z 1/S^z} \quad (8)$$

**Step 7:** Every candidate may pursue certain behavior, and this behavior is selected using Roulette wheel selection approach in the FCI. Using the Roulette wheel, each candidate may choose which corresponding behavior  $S^{Z[*]} = f[C_1^{Z[*]}, C_2^{Z[*]}, \dots, C_A^{Z[*]}]$  is to be followed. Roulette wheel selection method is a probabilistic selection approach which is used in current study to recommend a fitter/better behavior. This approach improves the chance of every behavior to be selected based purely on its quality at least once. This process helps each candidate to preferred better behavior with the help of correlated probability  $p^z$  using Eq. (8) seeing that  $p^z$  is directly proportional to the characteristics of the behavior  $S^z$ .

**Step 8:** Every candidate shrinks its sampling interval  $\alpha_i^{z[*]}$  for every feature represented as  $c_i^{Z[*]}$  to its nearest neighbor and forms a new sampling interval using Eq. (9). Following or learning from a certain behavior means that current sampling interval associated with every  $S^z$  is updated to the close neighborhood of the candidate to be followed.

$$\alpha_i^{z[*]} \in \left[ c_i^{z[*]} - (\|\alpha_i\|/2), c_i^{z[*]} + (\|\alpha_i\|/2) \right] \quad (9)$$

where  $\alpha_i = (\|\alpha_i\|) \times r$ . Here, '\*' illustrates that the behavior is selected at random by the candidates and not known previously.

**Step 9:** After having updated its features (i.e., the cluster centers), each candidate computes the updated objective function according to Eq. (1).

**Step 10:** This iterative process continues (between steps 3 and 9) till the cohort converges, i.e., if any of the under-mentioned conditions become true.

- if no significant improvement is noticed in the behavior  $S^z$  of every candidate in cohort

- $\|\max(S^Z)^{cn} - \max(S^Z)^{cn-1}\| \leq \varepsilon$ , where  $cn$  is the current iteration
- if maximum number of iterations (max\_iterations) is reached.

**Step 11:** If either of the two conditions is fulfilled, then acquire any of the best behavior from  $S^Z$  from the set of candidate behaviors (best objective function) as the concluding objective function value and end. If not, then continue to step 3.

### 3 Results and Discussion

The section discusses the dataset used to test the proposed hybridized data clustering algorithm, and the findings are reported. To evaluate the functioning of FCI, a comparative analysis is conducted with traditional FCM approach tested on the same dataset. In the experiment, the computations are executed in MATLAB R2013a on Mac OS platform with 1.6 GHz Intel Core i5 processor with 8 GB RAM.

#### 3.1 Dataset Used

The Wisconsin Breast cancer (WBC) dataset [4, 25] from UCI Machine Learning Repository is used to validate the proposed algorithm. The dataset contains 699 instances (amounting to a single clinical case) including 16 missing values. It includes nine features (as shown in Table 3) each of which is assigned an integer value between 1 and 10 and a class output attribute. This class output may report/classify a benign or malignant breast cancer diagnoses for a particular data object.

**Table 3** Attributes: the Wisconsin breast cancer dataset

Feature No.	Attribute or feature	Values
1	Clump thickness	1–10
2	Uniformity of cell size	1–10
3	Uniformity of cell shape	1–10
4	Single epithelial cell size	1–10
5	Marginal adhesion	1–10
6	Bare nuclei	1–10
7	Normal nucleoli	1–10
8	Bland chromatin	1–10
9	Mitoses	1–10
10	Class	B—Benign, M—Malignant

### 3.2 Analysis of Results

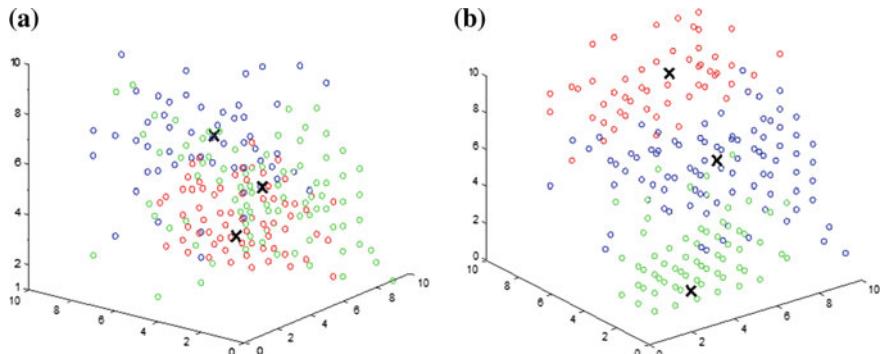
In the current study, the WBC dataset was used to validate the performance of the proposed FCI. The clustering performance of fuzzy C-means was also tested on the WBC dataset and then compared with hybridized FCI. A total of 40 trials were carried out with each method, and the number of clusters ‘A’ is known prior to solving the clustering problem. The simulation tries to optimize cluster centers of fuzzy C-means clustering algorithm using CI. For every trial, the input to the system included: the dataset as a csv file ( $N = 683$ ,  $A = 3$ ) and random initial cluster centers. Parameters like the best solution produced (Best), the worst value (Worst) recorded for the objective function, mean value of solution (Average), standard deviation (Std. Deviation), average running time (R.T.) and number of function evaluations (FE) across the trials for each FCI and FCM algorithms were recorded. The simulation results given are presented in Table 4. The results indicate that the hybridized fuzzy C-means, i.e., the FCI is superior in performance to FCM. The optimizer definitely aids in improved data clustering as can be seen in Table 4 that the objective function is minimized in FCI for all the criterion (best case, average case and even in worst trial run). It can also be seen that the hybridized fuzzy C-means shows a consistent performance than the traditional FCM even though the optimizer itself is heuristic nature and has a aspect of randomness to it; and also, the traditional FCM initializes with random seed (i.e., random cluster centers at the start).

Thus, it can be inferred that the optimized FCI lends a more consistent performance to the fuzzy C-means making it more robust with a smaller value of standard deviation. This may be attributed to the strength of CI which has strong capabilities of reaching better and accurate solutions by avoiding getting stuck in the local minima and also leading the algorithm to converge much more quickly.

Figure 1 illustrates the cluster formation graphically for both FCM and hybridized FCI with selected attributes on the WBC dataset for three clusters. Figure 1a shows the cluster formation after the traditional FCM was applied to the said dataset. It shows that the centroids as suggested by FCM result in overlapping clusters, thus hinting at weaker cluster formation due to the cluster centers. This may also lead to weaker predictive qualities if these clusters were used for further classification of unknown data. Figure 1b shows well-formed clusters as the objective function also

**Table 4** Simulation result

Criteria	Fuzzy C-means (FCM)	Fuzzy-CI (FCI)
Best	5442.042	5440.601
Average	5442.397	5440.605
Worst	5443.420	5440.617
Standard deviation	5.324982	0.0043814
FE	112	2503
R.T. (s)	0.18023173	0.179574305

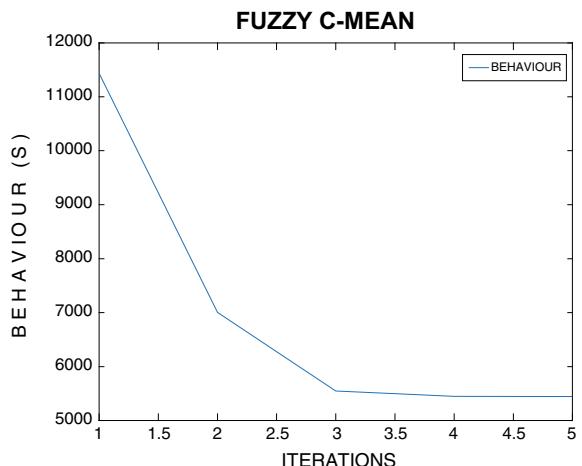


**Fig. 1** Cluster formation on WBC dataset using FCM (a) and hybridized FCI (b), respectively

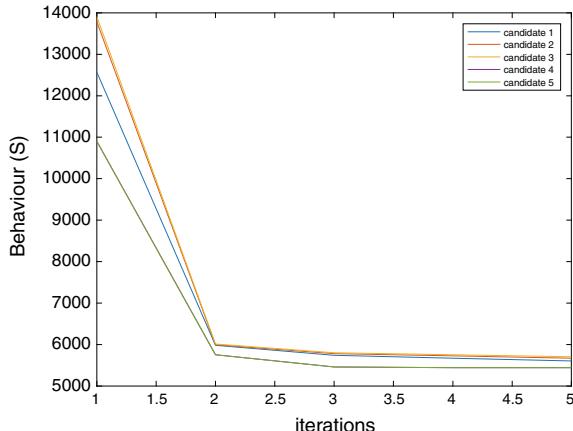
had achieved a better and minimized value with the fuzzy C-means hybridized with the CI optimizer. It may also be noted that there is a large difference in the FE values for FCM and the hybrid FCI Table 4. This is due to the reason that the hybrid version uses CI, where multiple agents (known as candidates of the cohort) are exploring the solution space, with each candidate making its own function evaluation or calculation of objective function at every iteration. However, even with higher FE, the hybrid model of FCI runs much faster also yielding improved cluster formations (as seen from the running time taken by both the clustering algorithms in Table 4).

Figures 2 and 3 illustrate behavior plots of FCM and FCI, respectively. Figure 2 shows that how the behavior ‘S’ is progressing and steadily moves toward convergence. On the other hand, the hybrid model of FCI (Fig. 3) with five different candidates in the cohort, each of which have their own certain set of behaviors, shows all the candidates approaching convergence much faster and exploiting the solution space more gradually as they near convergence.

**Fig. 2** Behaviour plot for FCM



**Fig. 3** Behaviour plot for FCI



## 4 Conclusion

This paper presents a hybrid fuzzy-CI procedure for data clustering. And the hybrid algorithm tries to combine the advantages of two algorithms, where fuzzy C-means is hybridized with the optimizer CI to enhance the cluster formation capabilities of traditional FCM. The proposed method is tested on Wisconsin Breast Cancer (WBC) dataset. The blend of fuzzy C-means and the stochastic CI allows the proposed algorithm to converge faster with improved and more accurate clustering. The results of the hybridized FCI were then compared with traditional Fuzzy C-means. The empirical result indicates the algorithmic outcome produces greater quality clusters with a much lower standard deviation on the particular dataset. In the future, performance of the traditional FCM could be improved and validated by comparing with concurrent metaheuristics. A very recent and promising class of optimization frameworks includes the socio-inspired metaheuristics which are evolutionary algorithms inspired from the social behavior of humans seen in various societal setups. Another scope for research could be to use a modified CI algorithm which would be self-adaptive in nature, which will aid in further optimization of the traditional FCM.

## References

1. Agrawal S, Agrawal J (2015) Neural network techniques for cancer prediction: a survey. Proc Comput Sci 60:769–774
2. Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR (2013) Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform 4(124):3
3. Asri H, Mousannif H, Al Moatassime H, Noel T (2016) Using machine learning algorithms for breast cancer risk prediction and diagnosis. Proc Comput Sci 83:1064–1069

4. Asuncion A, Newman DJ (2007) UCI machine learning repository. University of California, School of Information and Computer Science, Irvine, CA. <http://www.ics.uci.edu/~mlearn/MLRepository.html>
5. Ayer T, Alagoz O, Chhatwal J, Shavlik JW, Kahn CE Jr, Burnside ES (2010) Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration. *Cancer* 116(14):3310–3321
6. Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. *Comput Geosci* 10(2–3):191–203
7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 68(6):394–424
8. Cebeci Z, Yildiz F (2015) Comparison of K-means and fuzzy C-means algorithms on different cluster structures. *Agrarinformatica/J Agric Inform* 6(3):13–23
9. Chattopadhyay S, Pratihar DK, Sarkar SCD (2012) A comparative study of fuzzy c-means algorithm and entropy-based fuzzy clustering algorithms. *Comput Inform* 30(4):701–720
10. Dubey AK, Gupta U, Jain S (2016) Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. *Int J Comput Assist Radiol Surg* 11(11):2033–2047
11. Frigui H, Krishnapuram R (1999) A robust competitive clustering algorithm with applications in computer vision. *IEEE Trans Pattern Anal Mach Intell* 21(5):450–465
12. Gayathri BK, Raajan P (2016) A survey of breast cancer detection based on image segmentation techniques. In: International conference on computing technologies and intelligent data engineering (ICCTIDE). IEEE, pp 1–5
13. Han J, Pei J, Kamber M (2011) Data mining: concepts and techniques. Elsevier
14. Hartigan JA, Wong MA (1979) Algorithm AS 136: a k-means clustering algorithm. *J R Stat Soc Ser C (Appl Stat)* 28(1):100–108
15. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput Surv (CSUR)* 31(3):264–323
16. Kamalakkannan J, Krishna PV, Babu MR, Mukeshbhai KD (2015) Identification of abnormality from digital mammogram to detect breast cancer. In: 2015 international conference on circuits, power and computing technologies (ICCPCT-2015). IEEE, pp 1–5
17. Kashyap KL, Bajpai MK, Khanna P (2015) Breast cancer detection in digital mammograms. In: 2015 IEEE international conference on imaging systems and techniques (IST). IEEE pp 1–6
18. Kotsiantis SB, Zaharakis I, Pintelas P (2007) Supervised machine learning: a review of classification techniques. *Emerg Artif Intell Appl Comput Eng* 160:3–24
19. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI (2015) Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 13:8–17
20. Krishnasamy G, Kulkarni AJ, Paramesran R (2014) A hybrid approach for data clustering based on modified cohort intelligence and K-means. *Expert Syst Appl* 41(13):6009–6016
21. Kulkarni AJ, Durugkar IP, Kumar M (2013) Cohort intelligence: a self supervised learning behavior. In: 2013 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, pp 1396–1400
22. Kumar M, Kulkarni A (2019) Socio-inspired optimization metaheuristics: a review. In: Socio-cultural inspired metaheuristics, pp 1–28. Springer International Publishing (In Press)
23. Lafta HA, Ayoob NK (2013) Breast cancer diagnosis using genetic fuzzy rule based system. *J Univ Babylon* 21(4):1109–1120
24. Leung Y, Zhang JS, Xu ZB (2000) Clustering by scale-space filtering. *IEEE Trans Pattern Anal Mach Intell* 22(12):1396–1410
25. Mangasarian OL, Setiono R, Wolberg WH (1990) Pattern recognition via linear programming: theory and application to medical diagnosis. *Large-scale Numer Opt* 22–31
26. Medjahed SA, Saadi TA, Benyettou A (2013) Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules. *Int J Comput Appl* 62(1)
27. Michalski RS, Carbonell JG, Mitchell TM (eds) (2013) Machine learning: an artificial intelligence approach. Springer Science & Business Media

28. Odajima K, Pawlovsky AP (2014) A detailed description of the use of the kNN method for breast cancer diagnosis. In: 2014 7th international conference on biomedical engineering and informatics (BMEI). IEEE, pp 688–692
29. Ojha U, Goel S (2017) A study on prediction of breast cancer recurrence using data mining techniques. In: 2017 7th international conference on cloud computing, data science and engineering-confluence. IEEE, pp 527–530
30. Panda S, Sahu S, Jena P, Chattopadhyay S (2012) Comparing fuzzy-C means and K-means clustering techniques: a comprehensive study. In: Advances in computer science, engineering and applications. Springer, Berlin, Heidelberg, pp 451–460
31. Ramani R, Valarmathy S, Vanitha NS (2013) Breast cancer detection in mammograms based on clustering techniques—a survey. *Int J Comput Appl* 62(11)
32. Suganya R, Shanthi R (2012) Fuzzy c-means algorithm—a review. *Int J Sci Res Publ* 2(11):1
33. Suthaharan S (2016) Machine learning models and algorithms for big data classification. *Integr Ser Inf Syst* 36:1–12
34. Verma A, Khanna G (2016) A survey on image processing techniques for tumor detection in mammograms. In: 2016 3rd international conference on computing for sustainable global development (INDIACom). IEEE, pp 988–993
35. Yang MS (1993) A survey of fuzzy clustering. *Math Comput Model* 18(11):1–16

## Chapter 8

# Development of Algorithm for Spatial Modelling of Climate Data for Agriculture Management for the Semi-arid Area of Maharashtra in India



Vidya Kumbhar and T. P. Singh

## 1 Introduction

Agriculture is the backbone of Indian economy. It not only provides the food grains and other raw material but also it provides employment opportunities to more than 50% of the population [22]. It acts as a major source of income and also provides the food and fodder to the livestock. It is the major contributor to the national income and brings the foreign exchange to the country [12, 17, 19]. The semiarid and arid regions, contribute 67% of the net sown area in India. The semiarid region of India extends over 218 districts across 14 states [23]. The states in the Northern region include Rajasthan, Punjab, Gujarat, Haryana and the southern regions include Maharashtra, Karnataka and Tamilnadu, Telangana. Out of the 174 million hectares cropped area in India, 131 million hectare lies under semiarid regions [10]. In spite of the major contribution of rainfed agriculture in Indian agriculture, the region is facing problems such as low productivity of the major rainfed crops and the degradation in the socioeconomic conditions of the small and marginal farmers. The agricultural crop production system in this region is greatly influenced by climatic parameters such as rainfall, temperature and evapotranspiration [15, 16, 24]. Increase in temperature due to climate change increases the potential evapotranspiration and thus increases the crop water requirement by 10% in semiarid and arid regions of India [18]. The uncertainty in the rainfall and limited irrigation facilities affects the crop yield in this region. The variations in the climate affect the crop management activities and credit investment management of the farm. It becomes difficult for the farmers to adjust their farm management activities and amount of investment to be done in the crop

---

V. Kumbhar (✉) · T. P. Singh

Symbiosis Institute of Geoinformatics, Symbiosis International (Deemed University), Pune, Maharashtra, India

e-mail: [vidya@sig.ac.in](mailto:vidya@sig.ac.in)

production inputs. It affects the season cycle and because of this the gap between crop yield and investment affects the income prospects for the farmers [13]. The variability in the rainfall also have major effect on crop yield in the region. The distribution of rainfall during the crop growth cycle is uneven, receiving scarce amount when in need, and high amounts when already in abundance, thereby, adversely affecting the crop yield in the respective regions [1, 3–8, 18]. The mid season growth for the semi arid crops get affected due to these prolonged rainless spells [21].

The area selected for the study is a semiarid region of Satara district, Maharashtra. The area selected for the current study covers eastern part of Satara district. Administratively, it covers five talukas of the district namely, Khandala, Koregaon, Phaltan, Man, Khatav. The geographical location of the study area covers the area between  $17^{\circ}22'54.807''$  N to  $18^{\circ}10'57.579''$  N and  $73^{\circ}52'14.2566''$  E to  $74^{\circ}54'35.0238''$  E, which corresponds to an area of  $5454.80\text{ km}^2$ . The agriculture pertaining to our study area is completely dependent on the rainfall. The region suffers from a climate change and it is classified as drought region for 20% of the years between 1991 and 2011 [9, 14, 26]. The maximum number of continuous dry days affects the crop growth and irrigation scheduling in the study area [3]. There is strong need to propose a model which will provide the early warnings to the farmers in this study area about the spatial variation of climate parameters. The current study has proposed an algorithm for spatial modelling of the climate data provided by Indian Meteorological department (IMD).

## 2 Methodology

This section explains the step by step method to design the spatial modelling of climate data. The spatial data generated with the mentioned algorithm is validated with the real time satellite data of Tropical Rainfall Monitoring Mission (TRMM).

### 2.1 Climatic Data

The climatic parameters such as rainfall, temperature and evapotranspiration are considered for the current study. The grid wise temperature and rainfall data has been collected from National Climate Centre, India Meteorological Department (IMD), Pune for the years 2010–2015.

Table 1 shows the details of climatic data collected.

**Table 1** Details of climatic data collected

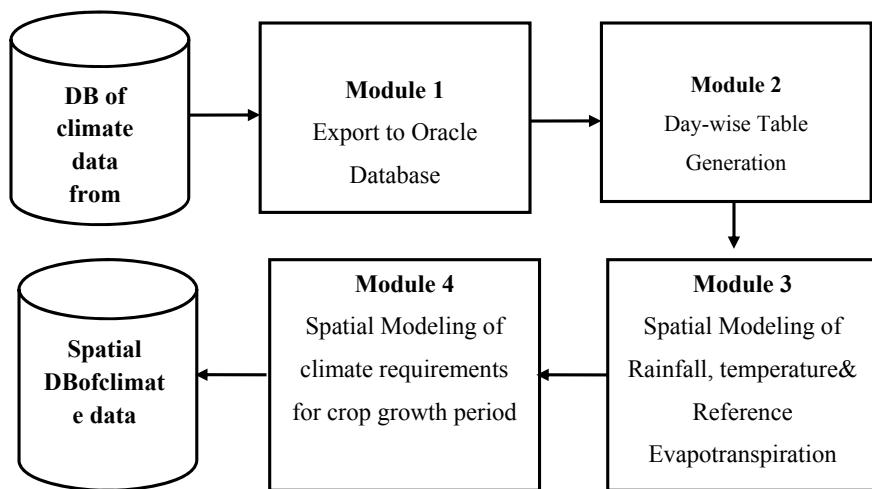
S. no	Data collected	Details
1	Daily minimum and maximum temperature	Grid size $1^{\circ} \times 1^{\circ}$
2	Daily rainfall	Grid size $0.25^{\circ} \times 0.25^{\circ}$

## 2.2 Proposed System for Climatic Data Process

Figure 1 shows the flow of the proposed architecture for climate data processing system which we have named as “Day wise Spatial Climate Data Generation Process (DSCDPG)” consists of processes like export to database, day-wise table generation, spatial data modeling of climate data and spatial modeling of climate requirements for crop growth period.

### 2.2.1 Database of Climate Data

The grid wise climate data collected from Indian Meteorological department was imported to Microsoft Excel Workbook by using text to columns utility in Excel. For both the type of climate data, rainfall and temperature, a separate workbook was created and stored in one common folder named as “Input”. This folder was given as an input to the next process for exporting to Oracle database. Table 2 shows the



**Fig. 1** Overview of proposed day wise spatial climate data generation process (DSCDPG)

**Table 2** Sample of the DB from IMD for rainfall

1012012	74.5	75.5	76.5	77.5
17.5	18.37	18.5	18.81	19.53
18.5	18.06	17.41	18.79	18.52
19.5	15.01	17.53	17.22	19.19
20.5	15.87	17.04	17.86	19.09
21.5	14.12	15.29	17.42	17.23

**Table 3** Table structure for store day wise rainfall data

S. no	Column name	Data type
1	as_on_date	Date
2	X	Number
3	Y	Number
4	Rainfall	Number

sample of the database collected from IMD for rainfall data, for the date 01st January 2012. In the Table 2, the top left cell shows the date for which data was collected and then remaining column headings show the longitude values and row headings shows the latitude values.

### 2.2.2 Export to Oracle DB

The objective of this module was to read all the workbook files from the “Input” folder specified and export the data to Oracle database. The algorithm for this process, for which code was written using Java as a programming language is as follows:

- Step 1: Import Apache Poor Obfuscation Implementation (POI) libraries in the java file. The Apache POI is the Java Application Programming Interface (API) to access the Microsoft by Apache.
- Step 2: Create an object of HSSF Workbook class from Apache POI was used to represent the workbook and locate a sheet.
- Step 3: Read a number of rows and columns from the entire worksheet.
- Step 4: Locate the cell values for the selected study area from the worksheet and read those values from the total rows and columns.
- Step 5: Generate the Oracle “Insert table” query script (.sql) by concatenating cell values.
- Step 6: Execute the “Insert table” .sql script in Oracle database.

After Step 6, the tables are created in the Oracle database. Table 3, shows the sample structure of the table created in Oracle database.

### 2.2.3 Day Wise Table Generation

The objective of this module was to read the day wise climate data from the master table created in Module-2 and create separate tables (views) for each date, for the selected latitude and longitude. Thus for the year, 365/366 views will be created for the selected area at the end of this module. For this, the Oracle script was written to read the data from the master table and create day wise separate Oracle views. For this process execution, in an Oracle script file, the concept of cursors in SQL was used. Oracle cursor is nothing but a memory area or a context area which holds the

results of the SQL query. The algorithm for this process, which was written using Structured Query Language (SQL) as a programming language is as follows:

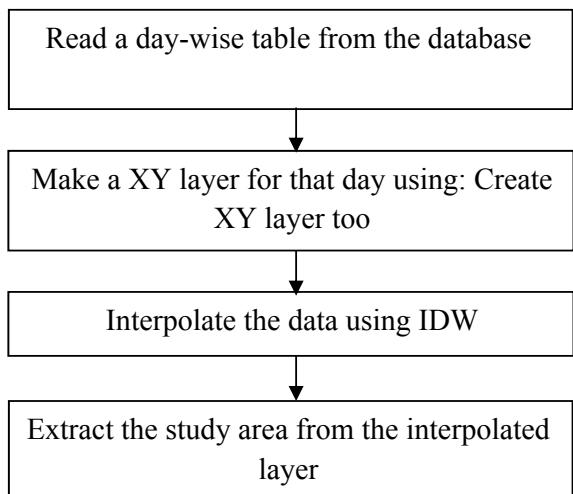
- Step 1: Create a cursor which finds the distinct dates from the base table.
- Step 2: Open the cursor, and for each date from the cursor, retrieve the data from the base table and create a view.

#### 2.2.4 Spatial Modeling of Climate Data

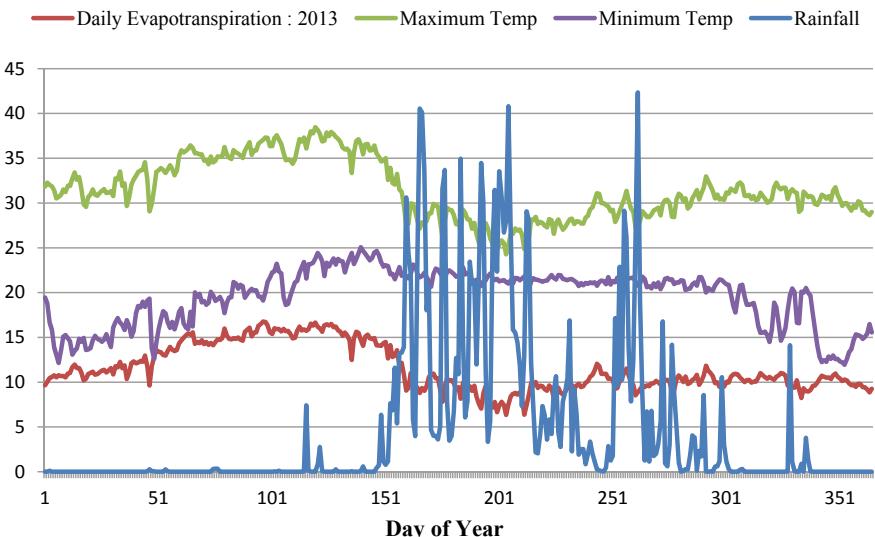
This module was designed to develop the spatial representation of day wise tabular representation of rainfall and temperature data, which was created in the previous step. For this process, a model was written in ARCGIS. The model iteratively reads the table from the database, and then creates an XY layer from the same. Daily average temperature and rainfall data were calculated using Inverse Distance Weighted (IDW) Interpolation method from each day wise XY layer. The mathematical model of IDW is based on the basic principle that, the values which are closest to the prediction location will have the highest weight than the values which are far away from the prediction location. The value of the weight will go on reducing as the measured value moves away from the value to be predicted [6, 20, 25]. From the interpolated image the study area was extracted to generate day wise climate parameters maps for the study area (Fig. 2).

In this next step of this module, the reference evapotranspiration was calculated from the spatial data prepared in the previous module for minimum and maximum temperature. Reference evapotranspiration is the amount of water evaporated by soil surface [2]. The reference evapotranspiration was calculated using the Hargreaves Potential Evapotranspiration (PET) method (Eq. 1) [11].

**Fig. 2** Flowchart for the spatial modeling of climate data



### Rainfall, Temperature and Evapotranspiration : 2013



**Fig. 3** Relationship between rainfall, temperature and evapotranspiration-2013

Hargreaves Potential Evapotranspiration (PET)

$$ET_0 = 0.0023R_a(T_{\max} - T_{\min})^{0.5} \left( \frac{T_{\max} + T_{\min}}{2} + 17.78 \right) \quad (1)$$

where

$R_a$  The total incoming extraterrestrial Solar radiation

$T_{\max}$  Daily Maximum Temperature

$T_{\min}$  Daily Minimum Temperature.

The algorithm for this process, for which code is written using Python as a programming language is as follows:

- Step 1: Import the arcpy library.
- Step 2: Set the workspace for the code, as the path where the Spatial data of temperature is stored.
- Step 3: For each date, read the minimum and maximum temperature spatial data from the database.
- Step 4: Calculate the reference evapotranspiration using Eq. 1.
- Step 5: Repeat the step for all the days of the year.

## 2.2.5 Spatial Modeling of Climate Requirements for Crop Growth Period

At the end of this proposed system, “Day wise Spatial Climate Data Generation Process (DSCDGP)”, from the 365/366 days spatial data of the year, climate requirements for crop growth period were calculated. The calculating of climate requirement includes, total rainfall, average minimum and maximum temperature and total crop water requirement. The algorithms for this process, for which code was written using Python as a programming language is as follows:

- **Algorithm for total rainfall during the crop growth cycle:**

- Step 1: Read the start date of the crop growth cycle.
- Step 2: Read the end date of the crop growth cycle.
- Step 3: Iterate through the database of spatial data created in Module 3 between the start date to end date and calculate the total of all the maps.
- Step 4: Save the result of Step 3 spatial data.

- **Algorithm for calculating average minimum and maximum temperature during the crop growth cycle:**

- Step 1: Read the start date of the crop growth cycle.
- Step 2: Read the end date of the crop growth cycle.
- Step 3: Iterate through the database of spatial data created in module 3 between the start date to end date and calculate the total of all the maps.
- Step 4: Divide the total by number of days of crop growth period.
- Step 5: Save the result of step 4 spatial data.

- **Algorithm for calculating crop evapotranspiration during the crop growth cycle:**

- Step 1: Read the start date of the crop growth cycle.
- Step 2: Read the end date of the crop growth cycle.
- Step 3: Declare the crop coefficient variables and their values as per the crop selected for the study and growth stage of the crop.
- Step 4: Iterate through the database of reference evapotranspiration spatial data and check the date of that spatial data.
- Step 5: Multiply the data with crop coefficient as per the stage in which that date falls (Initial/development/mid/late).
- Step 6: Calculate the total of all the spatial data calculated in Step 5.
- Step 7: Save the result of step 6 spatial data.

### 3 Results and Discussion

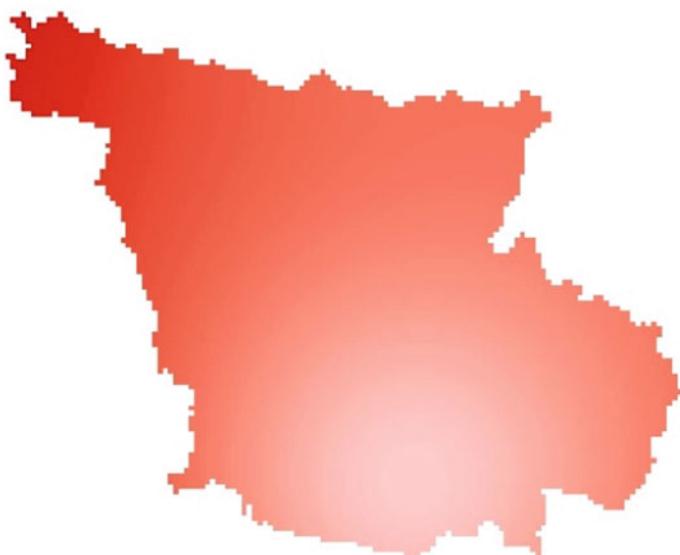
#### 3.1 Climate Data Analysis

The spatial representation of climate data was done for the years 2010–2013 by applying DSCDGP (Figs. 4, 5, 6 and 7). The analysis of the spatial data generated includes the study of the variation of climate parameters such as rainfall, temperature, reference evapotranspiration for the study area. For the study area, the reference evapotranspiration derived from rainfall and temperature shows that, the daily minimum reference evapotranspiration ranges between 6.34 to 7.57  $\text{mm d}^{-1}$  and maximum reference evapotranspiration ranges between 16.33 and 17.33  $\text{mm d}^{-1}$ . The analysis also



**Fig. 4** Rainfall Kharif season 2013

Minimum Temperature:Kharif-2013 



#### Legend

Value
High : 21.8637
Low : 21.0282

Kilometers  
0 5 10 20 30 40

**Fig. 5** Minimum temperature Kharif 2013

shows that, there was a decrease in rainfall and temperature and thus the reference evapotranspiration also decreased from the year 2010–2011. There was not much variation in evapotranspiration for the year 2011–2012. The results also revealed that, as the temperature has increased and rainfall has decreased, the reference evapotranspiration has increased from year 2012–2013. The trend analysis of reference evapotranspiration shows that reference evapotranspiration is more from January to May and then from June onwards it decreases as the temperature reduced and rainfall increases (Table 4). Figure 3 shows the variation of relationship between climate parameters for the year 2013.

The analysis of climate data for the study area concludes that, there is uneven distribution of the rainfall and continuing increase in the maximum temperature. The

Maximum Temperature:Kharif-2013 



#### Legend



**Fig. 6** Maximum temperature Kharif 2013

study revealed that average maximum temperature for the study area has gradually increased above 38 °C. Because of increase in temperature and unusual rainfall, the reference evapotranspiration has increased for the study area and this has affected the soil moisture contents and fertility of the soil for the region. The studies also show that increase in reference evapotranspiration has increased the crop evapotranspiration and because of increase in water evaporated by crop, increased the crop water requirement for the crop. The lack of water availability and rainfall dependent agriculture has affected the crop yield for the region.

### Crop Evapotranspiration:Kharif-2013



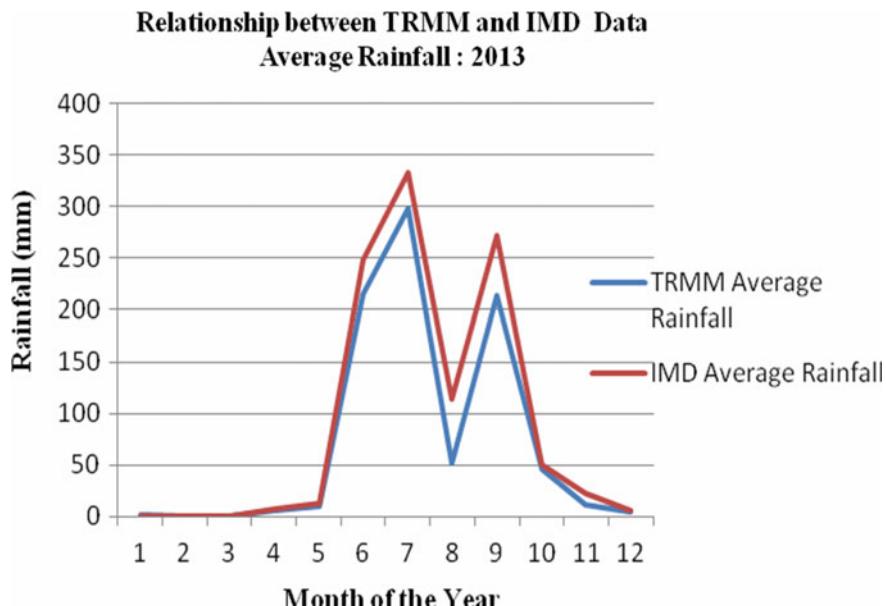
#### Legend



**Fig. 7** Crop evapotranspiration Kharif 2013

**Table 4** Details of daily average climate data

Type of climate data	2010		2011		2012		2013	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
Rainfall (mm)	0	42.09	0	30.07	0	57.85	0	42.34
Min. Temp. (°C)	10.98	26.15	9.98	24.60	10.99	23.81	11.94	25.05
Max. Temp. (°C)	24.62	39.96	25.10	38.13	26.40	37.24	24.27	38.46
ET <sub>0</sub> (mm d <sup>-1</sup> )	6.58	17.33	7.15	16.67	7.57	16.533	6.34	16.76



**Fig. 8** Validation of DSCDGP algorithm with TRMM data

### 3.2 Validation of DSCDGP

The results of DSCDGP were validated with the Tropical Rainfall Monitoring Mission (TRMM) data for the years 2012 and 2013. The monthly  $0.025 \times 0.25$  (degree) TRMM data product 3B43 was collected for the study area from Goddard Earth Sciences Data and Information Services Center (GES DISC). The study area was extracted from the TRMM data. The monthly average rainfall, maximum rainfall and minimum rainfall were found from the extracted data. The results of correlation analysis for average rainfall between TRMM and IMD for the year 2012 was observed to be 0.865 and for the year 2013 was 0.990 (Fig. 8). This validates the proposed DSCDGP system.

## 4 Conclusion

The suggested method has proposed a system for climate data process named as “Day wise Spatial Climate Data Generation Process (DSCDGP)” which has automatized the process of generating spatial representation of climate data. This process has offered the agricultural experts an easy technique to study the spatial variation of climate parameters and helps them for contingency planning of the study area. The current research has also validated the grid wise Indian Meteorological Department

(IMD) rainfall data with the Tropical Rainfall Monitoring Mission (TRMM) satellite rainfall data. The model will have predicted climate data from IMD for the upcoming season and soil data for the farmer from the selected taluka and village. From the daily spatial climate data the crop growth stage wise variation of climate parameters will help farmers for micro level planning of agriculture of study area. With the real time availability of IMD data the model will provide early warnings of climate variation can help farmers to decide the crops. If monsoon is delayed then contingency planning of crop can be done as per the rainfall.

## References

1. Aggarwal et al (2010) Managing climatic risks to combat land degradation and enhance food security: key information needs. *Proc Environ Sci* 1:305–312
2. Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56. FAO, Rome 300(9):D05109
3. Atal KR, Zende AM (2015) Wet and dry spell characteristics of semi-arid region, western Maharashtra, India E-proceedings of the 36th IAHR world congress, deltas of the future and what happens upstream, pp 1–7. International Association for Hydro-Environment Engineering and Research-IAHR, The Hague, The Netherlands
4. Balaghi et al (2010) Managing climatic risks for enhanced food security: key information capabilities. *Proc Environ Sci* 1:313–323
5. Bantilan MCS, Aupama KV (2006) Vulnerability and adaptation in dryland agriculture in India's SAT: experiences from ICRISAT's village-level studies. *J SAT Agric Res* 2(1):1–14
6. Burrough PA, McDonnell RA (1998) Principles of geographical information systems. Oxford University Press Inc., New York, pp 333–340
7. Coe R, Stern RD (2011) Assessing and addressing climate-induced risk in sub-Saharan rainfed agriculture: lessons learned. *Exp Agric* 47(02):395–410
8. Cooper PJM, Dimes J, Rao KPC, Shapiro B, Shiferaw B, Twomlow S (2008) Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: an essential first step in adapting to future climate change? *Agric Ecosyst Environ* 126(1):24–35
9. Dawane PR (2015) A comparative study of dairy co-operative unions in Satara district. Doctoral dissertation. <http://hdl.handle.net/10603/34954>
10. Gautam R, Rao J (2007) Integrated water management-concepts of rainfed agriculture. Central Research Institute of Dryland Agriculture (CRIDA), IARI. <http://nsdl.niscair.res.in/jspui/bitstream/123456789/554/1/Conceptsofrainfedagriculture-Formatted.pdf>
11. Hargreaves GH (1994) Simplified coefficients for estimating monthly solar radiation in North America and Europe, departmental paper, Department of boiler and irrigation engineering, Utah State University, Logan, Utah
12. Himani (2014) An analysis of agriculture sector in Indian economy. *IOSR J Human Soc Sci* 19(1):47–54
13. Hochman Z, Horan H, Reddy DR, Sreenivas G, Tallapragada C, Adusumilli R, Roth CH (2017) Smallholder farmers managing climate risk in India: 1. Adapting to a variable climate. *Agric Syst* 150:54–66
14. Jagannath B (2014) Rainfall trend in drought prone region in eastern part of Satara district of Maharashtra, India. *Euro Acad Res* 2(1):329–340
15. Krishna Kumar K, Rupa Kumar K, Ashrit RG, Deshpande NR, Hansen JW (2004) Climate impacts on Indian agriculture. *Int J Climatol* 24(11):1375–1393
16. Mall RK, Singh R, Gupta A, Srinivasan G, Rathore LS (2006) Impact of climate change on Indian agriculture: a review. *Clim Change* 78(2–4):445–478

17. Mathur AS, Das S, Sircar S (2006) Status of agriculture in India: trends and prospects. *Econ Polit Weekly* 41(52):5327–5336
18. Meinke H, Nelson R, Kokic P, Stone R, Selvaraju R, Baethgen W (2006) Actionable climate knowledge: from analysis to synthesis. *Clim Res* 33(1):101–110
19. Pandey MM (2009) Indian agriculture—an introduction [Country Report]. Asian and Pacific centre for agricultural engineering and machinery (APCAEM). Thailand, Country Report, India
20. Philip GM, Watson DF (1982) A precise method for determining contoured surfaces. *Aust Pet Explor Assoc J* 22(1):205–212
21. Sarker RP, Biswas BC (1978) Agricultural meteorology in India: a status report. In: Agroclimatological research needs of the semi-arid tropics, proceedings of the international workshop on the agroclimatological research needs of the semi-arid tropics. International Crops Research Institute for the Semi-Arid Tropics
22. Sharma VP (2011) India's agricultural development under the new economic regime: policy perspective and strategy for the 12th five year plan. Indian Institute of Management
23. Singh HP, Venkateswarlu B, Vittal KPR, Ramachandran K (2000) Management of rainfed agro-ecosystem. In: Proceedings of the international conference on managing natural resources for sustainable agricultural production in the 21st century, pp 14–18
24. Sinha SK, Singh GB, Rai M (1998) Decline in crop productivity in Haryana and Punjab: myth or reality. Indian Council of Agricultural Research, New Delhi, India
25. Watson DF, Philip GM (1985) A refinement of inverse distance weighted interpolation. *Geo-processing* 2(4):315–327
26. Zende AM, Nagarajan R, Atal KR (2012) Rainfall trend in semi arid region-Yerala river basin of western Maharashtra, India. *Int J Adv Technol* 3:137–145

# Chapter 9

# A Survey on Human Group Activity Recognition by Analysing Person Action from Video Sequences Using Machine Learning Techniques



Smita Kulkarni, Sangeeta Jadhav and Debashis Adhikari

## 1 Introduction

In computer vision, there is a plethora of techniques, which focuses on single-person or complex human activities [1–3] in video using machine learning (ML). The areas of group activity recognition are comparatively unexplored. In real-life applications like video surveillance, human–computer interface and sports video analytics, it requires significant group activity and interrelation between people, which is a challenging task. Analysis of group activity presents numerous real-life applications involving social role in understanding and anticipating social events. Event understanding in video surveillances is a significant module of computer vision structure. In video surveillance, GAR is important for video summarization and retrieval. In sports, GAR is more inspiring than video surveillance due to variation in relative location of players. Sports activity recognition is demanding because of the quick changeover among the actions, occlusions and speedy activities of the players, diverse camera positions, and camera movements. Additionally, spatiotemporal formation of events varies a lot in different sports. Large volumes of video data need an automatic GAR for dynamic scenes.

---

S. Kulkarni  
D.Y. Patil College of Engineering, Akurdi, Pune, India  
e-mail: [skulkarni@entc.maepune.ac.in](mailto:skulkarni@entc.maepune.ac.in)

S. Jadhav  
Army Institute of Technology, Dighi, Pune, India  
e-mail: [djsangeeta@rediffmail.com](mailto:djsangeeta@rediffmail.com)

S. Kulkarni · D. Adhikari (✉)  
MIT Academy of Engineering, Alandi (D), Pune, India  
e-mail: [dadhibari@entc.mitaoe.ac.in](mailto:dadhibari@entc.mitaoe.ac.in)

In videos, localizing group activities is challenging. It requires understanding of spatiotemporal scales about persons' actions along with group interactions transversely. GAR involves interpretation regarding interaction between individual and their relations. This is tricky due to uncertainty in features for interpreting relations between people. Modelling of GAR is complicated, as persons involved in an interaction keep moving.

To extract valuable information and make appropriate decisions from video data, machine learning techniques have been observed as a powerful solution. A group's activity is strongly associated with contribution of an individual's actions, considering their interrelations. For identifying collective/group activity, interaction between person based on sequential dynamics over time offer significant indications. To understand group activity, the important cue is to capture a person actions and the interaction between people, jointly. It is essential to consider only those person whose interaction is significant for the group activity recognition. Machine learning (ML) representations [4] encourage automatic adaptive framework for modelling interaction between people. Surveys in [4–8] look at the enormous literature in activity recognition. This review paper discusses various ML approaches and recent advances in deep learning for GAR.

The rest of the paper is presented as follows: Sect. 2 illustrates probabilistic structure modelling for GAR. Section 3 presents ML in GAR surveyed by an overview of HMM frameworks, along with their rewards and shortcomings based on related works in literature. This paper distinguishes among the more conventional schemes constructed on handcrafted features in Sect. 4 review followed by person action context model, person–person interaction graphical model and support vector machine (SVM) technique for group context model. Section 5 summarizes the state-of-the-art learned features based on deep learning model for GAR. Section 6 summarizes GAR for the discussed techniques and provides concluding remarks.

## 2 Probabilistic Structure Modelling for GAR

The majority of previous work [2, 3] on GAR is modelled on a small group of actions with comprehensible structural level information. In [9], the authors modelled the scenes using 2D polygonal shape, and each person in the group is considered over a period. The model is functional for malfunction detection in surveillance. Application of rigidity 3D polygonal formation to represent the parade group activity is discussed in [10]. They use an entire group as a complete activity instead of considering every person individually.

In [11], authors employ probabilistic, highly structured techniques for recognizing actions such as the way American football is played. In [12], recognized multi-player games and player strategies are discussed. Specific group behaviour activity is recognized in [13] with the help of multiple cameras. Two hierarchical clustering approaches are anticipated in [14] for real-time surveillance in a challenging environment. The major problem of these frameworks is its design for particular type

of activities with firm strategy, and as a result cannot apply to additional general activities. Stochastic representation is present in [15, 16] which describes equally spatial and temporal engagements between group people intended for more general group activities.

However, many research experts encode the illustration of actions manually. The above-mentioned approaches are able to recognize group activities automatically and those which are important for surveillance and sport analytics applications.

### 3 Probabilistic Graphical HMM Machine Learning Model for GAR

Various ML techniques in [4] have been employed for automatic human activity recognition modelling. Interactions between people are approximated based on probability distribution using hidden Markov model (HMM) with distribution of sequence learning. The HMM framework in [17] is able to be used to model stochastic methods wherever the non-observable state of the scheme is directed by a Markov process. The human-recognizable sequences of the system have an essential probabilistic dependency. An HMM computing the probability model is used for recognizing activities.

A layered probabilistic representation of HMM is successfully applied for sequence learning of actions. Single-layer HMM faces the overfitting problem due to limited training data. In [17], a proposed two-layer HMM structure that had benefited over previous works to discriminate group actions from individual actions is being discussed. Layer I-HMM represents the individual person action, and G-HMM designates the group action. Most of these previous methods are developed for a fixed number of group members. They cannot handle a changing number of group members. For automatic group activity detection, asynchronous HMM is implemented in [18] that handle changing number of group members. In [19], symmetry group activity is captured by HMM model for recognizing symmetry activities by computation of probability.

Though HMM integrates temporal information, it has a drawback which requires large-scale training data [20]. HMM is less efficient for relating and differentiating complicated temporal interactions along with several trajectories in-group activities. Besides in-group activity, recognition handling motion uncertainties of an individual person is an essential issue. In view of the fact that uncertain motion nature of persons differs inherently in-group activities by this, the recognition accuracy may be significantly affected. Thus, it is essential to build up a more flexible recognition framework [21].

However, these approaches had restrictions for identifying the scene-related actions due to the negligence of relationship between persons in addition to their surrounding scene. In unpredictable situations, HMM model becomes complex and restricted to represent interaction between people.

## 4 Handcrafted Feature-Based Machine Learning Model for GAR

Group activity can discriminate through spatiotemporal appearance/motion properties of an individual and their relations. GAR based on handcrafted visual features can be resolved by considering person action context model, person-to-person interaction model and group activity classification model by implementing SVM on group activity.

### 4.1 Person Action Context Model

Person-level action feature descriptors are generated by using handcrafted features [22–31] such as histogram of gradients (HOG) [26, 29], spatiotemporal local (STL) descriptor [22], motion-based STIP features [23], scale-invariant feature transform (SIFT), shape context descriptor and principal component analysis (PCA). Additionally, these techniques are considered as context model like action context (AC) descriptor [27], spatiotemporal volume [24, 25], pair-wise interaction model or their combined approach. These are subsequently used to learn interaction between people by graphical model.

### 4.2 Person–Person Interaction Graphical Model

ML graphical model [22–38] encourages automatic adaptive structures to model the interaction at the structure level. Various researchers [31–37] investigated graphical modelling for understanding interaction between people and their role in-group activity. In [26], adaptive graphical modelling framework is proposed which automatically understands the optimum configuration of person-to-person interaction.

A graphical model is a probabilistic model. The graph articulates the conditional dependence framework between individual actions and interactions with other people in a scene. Graphical structures construct higher-order correlation between persons in the scene to identify a relationship in a group activity. In graphical model, nodes correspond to individual variables, and edges represent statistical dependencies connections between individual actions. The probability function of graphical model denotes the person action feature, and the graph  $G$  is context model of person interaction to represent group activity. Graphical models allow message-passing algorithms that execute probabilistic group activity inference efficiently.

Modelling interactions between people and their role in activity recognition have been investigated for graphical model, such as AND/OR graphs in [23–25], hierarchical graphical model in [27–30] and dynamic Bayesian networks [24], have been proposed. In [23], AND/OR graphs allow for spatiotemporal constraints of action

relationships and [24, 25] formulate ST-AOG as Monte Carlo Tree Search (MCTS) cost-sensitive inference between persons' action. This model leads to more challenging learning problems. In [31], authors have implemented fully connected graph to discover subgroups of interacting people. In [27, 33], authors have modelled latent adaptive structures and grouping nodes [30] to discriminate interactions in a scene.

These frameworks are trained using handcrafted features and cannot straightforwardly be adopted in deep learning learned feature models. Effectively combined graphical model with the deep neural network [36, 38] captures dependencies between persons and gains competitive good accuracy than latent max-margin graphical method [27]. These methodologies for GAR is not feasible because these approaches frequently involve high computational cost. It is very difficult to generalize higher-order interactional framework, using the graphical method. Various graphical structures [39] have been discovered to model pair-wise interaction context; however, they cannot indicate the entire interaction context adequately and efficiently. In [39], this group activity optimized multi-target tracking interaction between pair-wise people using hypergraph Bayesian technique. This hypergraph solution is efficiently applicable to real-world application as camera calibration is not essential.

In recent studies, developed contextual information model for individual person and nearby person does not adequately represent the spatial and temporal reliability in group actions. To overcome this problem, Kaneko et al. [40] illustrated a technique to assimilate the individual recognition information through fully connected conditional random fields (CRFs), which describe every relation among the people in a video frame and adjust the relations strength by means of the amount of their similarity.

### 4.3 Group Context Model by SVM

Support vector machine (SVM) is a statistical machine learning algorithm which is selected to learn and classify group activities of high-dimensional space [22–40]. SVM classifier established on person descriptor (HOG) [27], STV [30], context (AC) descriptor [32] and their related action labels through fixed graph structure [26–31] is able to capture the group activities automatically.

The main scheme behind SVM is finding the optimal hyperplane for separation of the group activity categories. Loss function for group activity is specified by

$$c(x, y, f(x)) = (1 - y * f(x)) \quad (1)$$

where  $x$  is person action labels,  $y$  is group action label, and  $f(x)$  is predicted label. The inference problem is solved by optimizing the model parameters  $w$  to find the best group action label  $y$  for a person action labels  $x$ . Maximizing the distance between the hyperplanes requires minimizing  $\|w\|$  which is an optimization problem and can be written as in Eq. (2).

$$\min \lambda \|w\| + \sum_{i=1}^n (1 - y_i(x_i, w)) \quad (2)$$

In Eq. (2), the first term is a regularizer of the SVM; the second term is the loss. The regularizer  $\lambda$  balances between margin maximization and loss. For a misclassified sample, update the weight vector  $w$  using the gradients, else if classified properly.

The improved discriminative capacities of SVMs are robust and thus are appropriate for GAR in noisy surroundings. The kernel matrix involved in SVM is proficient for handling high-dimensional data in the optimization process.

The major computational challenge in SVM learning is loss-augmented inference or finding the most complicated group activity. In SVM, kernel selection is tricky task on which output accuracy depends for a given task. The input vectors of an SVM require fixed dimensions, whereas in GAR each sequence can have variable intervals.

## 5 Learned Feature-Based Deep Model for GAR

Hand-engineered and static feature human activity models [22–40] are not suitable for automatic high-level learning. The state-of-the-art method for GAR consists of handcrafted feature extractor as densely or sparse significant points (e.g. HOG and SIFT) in a Bag-of-Words static feature model, which are not suitable for continuous learning. These are then used to learn interaction between people. Manually selected features and static interaction models require independent design for each application. These models are incapable of handling dynamic environments due to the static nature of the feature model. Thus, it is essential to develop techniques for online activity recognition based on automatic learning of the feature models for GAR recognition, from the unlabelled data in unsupervised manner for newly arriving instances.

Recently, deep learning has been implemented effectively into several regions such like computer vision, natural language processing, audio recognition and bioinformatics. In [41], the authors implemented automatically selected deep hybrid features for continuous active learning. These deep learning methods ensure significant improvement in the performance of action recognition in computer vision [42].

The deep model needs to learn spatiotemporal relations between the persons [36]. GAR is a higher-level representation that captures scene-level actions. Spatiotemporal relations are changed for different group activity. For complex group activity, handcrafted feature approach is limited for representation as it uses linear mode. The most recent subfield of ML, deep learning, is able to act as an association involving big video data and intelligent group activity learning. Deep learning approaches are capable to represent high-level video data and classify pattern by assembling multiple layers of statistics segments during hierarchical structural design [43]. Most of

the earlier group activity recognition approaches do not deal with high-order interactional framework and are restricted to offer flexible and scalable structure. Deep learning-based methods have an end-to-end effectiveness within trainable model for higher-level reasoning [44].

### **5.1 CNN for Person Action**

Deep learning-based convolution neural networks (CNNs) [43] extract impressive individual personal features from the scene which perform better than hand-engineered features such as STIP [23] and HOG [26]. Additionally, CNN features extract useful information from the scene, both supervised and unsupervised learning, for scene classification. CNN provides visual classification probability distribution over person action on the entire image which is used to directly estimate the group activity in the scene.

### **5.2 Group Activity Recognition with Recurrent Neural Network (RNN)**

Most previous works take the approach of indirect modelling structure of frame-level classifiers successively over a video at multiple temporal scales which do not satisfy accuracy as well as computational efficiency.

Group activity recognition needs sequential nature frames by means of individual person action and interaction among persons with dynamic temporal information. Recurrent neural network (RNN) handles variable length space-time inputs and dynamic temporal behaviour as it contains nonlinear units. RNN is broadly appropriate for video analysis tasks such as activity recognition [44]. To model person-level dynamics to entire group dynamics, deep model by assembling several layers of RNN recommended. Visual recognition approaches emphasize on deep learning methodologies associating the reasonably low-level model's output to interpret higher-level compositional scenes. This remains a challenging task. In [38], graphical models are integrated with deep neural networks. Additionally, RNN models highlight the dynamics of human interaction as collective group activity. RNN model gets deteriorated from vanishing gradients which neglect human interaction dynamics.

### 5.3 *Group Activity Recognition by Long Short-Term Memory (LSTM) RNN*

Recurrent neural networks based on the long short-term memory (LSTM) models have accomplished decent achievement in a great variety of applications having temporal sequencing data. Sequence learning represented by RNN/LSTM from video frames signifies improved performance to describe group-level dynamics in spite of demonstrating group action from a video frame.

In recent times, LSTM has turned out to be excellent in modelling dynamics of individual person action identification. This is owing to its capability of capturing the sequential temporal motion facts. LSTM includes additional ‘memory cell’ modules for keeping information over longer periods, which permits them to learn long-term dependencies of human interaction dynamics [42].

In [44], the authors present an end-to-end mode by means of a combination of back propagation and reinforce methodology for action recognition in video which motivates directly to predict sequential bounds of actions. In multi-person event occurrence, though many persons are acting, only a small group of persons contribute to a definite event in the scene. In [45], the authors proposed a method which acquires time-varying features at every time instant and are processed using RNN to indicate responsible people for the event classification. The bidirectional LSTM hidden states are then used by an attention model to recognize the ‘key’ player at each instant.

Recursive network including LSTM accepts orderly input sequences. However, in GAR the position of person-level features is without order. To recognize group activity, hierarchical deep model is assembled [46, 47]. Additionally, the model requires clear tags for person actions which are exclusive and rigid to recognize activities in sports like ice hockey.

For classifying group activity in ice hockey as suggested in [48], a deep learning model by feature aggregation of a person’s data is combined in the context of the activities in ice hockey games. In [49], hierarchical relational deep network model learns relational feature illustrations between persons in a scene that can efficiently classify person and group activity.

In [48], authors proposed bidirectional LSTM network, for group interaction prediction that incorporates both global motion and detailed local action dynamics of each individual. In [50], GAR is implemented by confidence-energy recurrent network (CERN) by minimization of the energy and maximization of the confidence measure of predictions.

In-group activities, recognizing multi-person interaction and information of every individual is a challenging problem. Group activity analysis is required in several applications such as societal incident prediction. Semantics-based GAR structure is proposed in [51] which uses two-stage LSTM model that accomplishes higher accuracy and effectiveness. CNN features perform well in the task of scene classification. It is extremely essential to be able to predict a group activity in real time for some application scenarios, e.g. sport analytics.

Hierarchical long short-term concurrent memory (HLSTCM) is proposed in [52] for human interaction. Single-person LSTM is employed to learn single-person dynamics, which is provided to Concurrent LSTM (Co-LSTM) unit. Concurrent LSTM unit assimilates interrelated movement information and recognizes the dynamics of interaction among all people. For understanding team sport activity, hierarchical LSTM recurrent network is recommended in [53]. Multiple persons' features extracted using CNN explicitly integrate over a LSTM model whose outputs are based on temporal sequence and improved robustness against instability in the number of observed players.

Though graphical structures or RNN approaches express high-order relationship between people in the scene, it ignores essential characteristic of group activity where all persons' activities are not contributed to group activity, homogenously. In [54], the authors offer a participation-contributed temporal dynamic model (PC-TDM) for GAR which models the significant dynamics of main individuals while escaping the inappropriate dynamics of outlier individuals.

The overall process proposed in [55] describes multi-stream convolution neural network framework. Each stream operates on a different modality, and predictions of all streams are combined to estimate human body posture heat map which include in depth, facts about the human body parts. Person-level features are extracted from multiple layers of CNN for GAR. Scene-level GAR is also created with the output of the last layer of the CNN. In [56], it proposes a distinct structural design that mutually localizes multiple persons and classifies the actions of collective activity. This model does not require pre-computed detection and tracking task for estimation in a distinct forward pass. Structural recurrent neural networks (SRNNs) [57] clearly models associations between individuals and all RNNs which can be trained simultaneously by means of a single loss function that outperforms hierarchical LSTMs. In [58], authors discuss effective control of the inaccurate captions generation in GAR in semantic domain by deep network.

In [59, 60], collective activities of the long-range temporal inconsistency and uniformity are controlled by a two-stage gated recurrent units (GRUs) network. Individual actions and group actions recognized in videos by semantic RNN, namely StagNet, are used in [61] which are capable of extracting discriminative and useful spatiotemporal illustrations and capturing interperson interactions. In recent advancement, generative adversarial networks (GANs) are competent in learning real-life example in which output is difficult to distinguish. In [62], GANs proposed recurrent semi-supervised model with capability to learn losses automatically for GAR.

## 6 Conclusion

In this review article, a brief introduction regarding GAR using ML models is given to present an insight to a reader interested in this domain. ML in GAR was initiated with probabilistic structure modelling and followed by layered HMM model for

sequence learning of action. However, these approaches had limitations for complexity of inference in unexpected circumstances due to the negligence of relationship between persons in addition to their surrounding scene. GAR established on hand-crafted feature-based machine learning model resolves complexity in scene-related actions by considering individual action context model, person-to-person interaction graphical model along with SVM classifier for GAR. Graphical model cannot indicate the entire group interaction context adequately and efficiently. This complexity has been recently achieved with the modern developments in learned features using deep learning model. Deep learning-based methods have effectiveness within higher-level reasoning as GAR.

## References

1. Blank M, Gorelick L, Shechtman E, Irani M, Basri R (2005) Actions as space-time shapes. In: ICCV
2. Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: a local SVM approach. In: ICPR
3. Wang H, Schmid C (2013) Action recognition with improved trajectories. In: Proceedings of the IEEE international conference on computer vision, pp 3551–3558
4. Turaga P, Chellappa R, Subrahmanian VS, Udrea O (2008) Machine recognition of human activities: a survey. *IEEE Trans Circuits Syst Video Technol* 18(11):1473
5. Aggarwal JK, Ryoo (2011) MS human activity analysis. *ACM Comput Surv* 43(3):1–43
6. Ke S-R, Thuc H, Lee Y-J, Hwang J-N, Yoo J-H, Choi K-H (2013) A review on video-based human activity recognition. *Computers* 2(2):88–131
7. Vahora SA, Chauhan NC (2017) A comprehensive study of group activity recognition methods in video. *Indian J Sci Technol* 10(23):1–11
8. Stergiou A, Poppe R (2018) Understanding human-human interactions: a survey. [arXiv:1808.00022](https://arxiv.org/abs/1808.00022)
9. Vaswani N, Roy Chowdhury A, Chellappa R (2003) Activity recognition using the dynamics of the configuration of interacting objects. In: 2003 IEEE computer society conference on computer vision and pattern recognition, proceedings, vol 2, pp II-633
10. Khan SM, Shah M (2005) Detecting group activities using rigidity of formation. In: Proceedings of the 13th annual ACM international conference on multimedia, pp 403–406
11. Intille SS, Bobick AF (2001) Recognizing planned multiperson action. *Comput Vis Image Underst* 81(3):414–445
12. Moore D, Essa I (2002) Recognizing multitasked activities from video using stochastic context-free grammar. In: AAAI/IAAI, pp 770–776
13. Cupillard F, Brémont F, Thonnat M (2002) Group behavior recognition with multiple cameras. In: Sixth IEEE workshop on applications of computer vision, proceedings, pp 177–183
14. Chang M-C, Krahmstoever N, Lim S, Yu T (2010) Group level activity recognition in crowded environments across multiple cameras. In: 7th IEEE international conference on advanced video and signal based surveillance, pp 56–63
15. Ryoo MS, Aggarwal JK (2009) Stochastic representation and recognition of high-level group activities: describing structural uncertainties in human activities. In: 2009 IEEE computer society conference on computer vision and pattern recognition workshops, pp 11–11
16. Ryoo MS, Aggarwal JK (2011) Stochastic representation and recognition of high-level group activities. *Int J Comput Vision* 93(2):183–200
17. Zhang D, Gatica-Perez D, Bengio S, McCowan I (2006) Modeling individual and group actions in meetings with layered HMMs. *IEEE Trans Multimedia* 8(3):509–520

18. Lin W, Sun M-T, Poovendran R, Zhang Z (2010) Group event detection with a varying number of group members for video surveillance. *IEEE Trans Circuits Syst Video Technol* 20(8):1057–1067
19. Zaidenberg S, Boulay B, Brémont F (2012) A generic framework for video understanding applied to group behavior recognition. In: IEEE ninth international conference on advanced video and signal-based surveillance. Beijing, pp 136–142
20. Guo P, Miao Z, Zhang X, Shen Y, Wang S (2012) Coupled observation decomposed hidden markov model for multiperson activity recognition. *IEEE Trans Circuits Syst Video Technol* 22(9):1306–1320
21. Lin W, Chu H, Wu J, Sheng B, Chen Z (2013) A heat-map-based algorithm for recognizing group activities in videos. *IEEE Trans Circuits Syst Video Technol* 23(11):1980–1992
22. Choi W, Shahid K, Savarese S (2009) What are they doing? Collective activity classification using spatio-temporal relationship among people. In: IEEE 12th international conference on computer vision workshops, ICCV workshops, pp 1282–1289
23. Gupta A, Srinivasan P, Shi J, Davis LS (2009) Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos. In: IEEE conference on computer vision and pattern recognition, pp 2012–2019
24. Amer MR, Xie D, Zhao M, Todorovic S, Zhu S-C (2012) Cost-sensitive top-down/bottom-up inference for multiscale activity recognition. European conference on computer vision. Springer, Berlin, Heidelberg, pp 187–200
25. Amer MR, Todorovic S, Fern A, Zhu S-C (2013) Monte carlo tree search for scheduling activity recognition. In: Proceedings of the IEEE international conference on computer vision, pp 1353–1360
26. Lan T, Wang Y, Yang W, Mori G (2010) Beyond actions: discriminative models for contextual group activities. In: Advances in neural information processing systems, pp 1216–1224
27. Lan T, Wang Y, Yang W, Robinovitch SN, Mori G (2012) Discriminative latent models for recognizing contextual group activities. *IEEE Trans Pattern Anal Mach Intell* 34(8):1549–1562
28. Lan T, Sigal L, Mori G (2012) Social roles in hierarchical models for human activity recognition. In: IEEE conference on computer vision and pattern recognition, pp 1354–1361
29. Choi W, Savarese S (2012) A unified framework for multi-target tracking and collective activity recognition. In: European conference on computer vision. Springer, Berlin, Heidelberg, pp 215–230
30. Amer MR, Lei P, Todorovic S (2014) Hirf: hierarchical random field for collective activity recognition in videos. In: European conference on computer vision. Springer, pp 572–585
31. Choi W, Chao YW, Pantofaru C, Savarese S (2012) Discovering groups of people in images. In: European conference on computer vision. Springer, pp 417–433
32. Khamis S, Morariu VI, Davis LS (2012) Combining per-frame and per-track cues for multi-person action recognition. European conference on computer vision. Springer, Berlin, Heidelberg, pp 116–129
33. Hajimirsadeghi H, Mori G (2015) Learning ensembles of potential functions for structured prediction with latent variables. In: Proceedings of the IEEE international conference on computer vision, pp 4059–4067
34. Zhu Y, Nayak NM, Roy-Chowdhury AK (2013) Contextaware modeling and recognition of activities in video. In: Computer vision and pattern recognition (CVPR), IEEE conference, pp 2491–2498
35. Tran KN, Gala A, Kakadiaris IA, Shah SK (2014) Activity analysis in crowded environments using social cues for group discovery and human interaction modeling. *Pattern Recogn Lett* 44:49–57
36. Deng Z, Zhai M, Chen L, Liu Y, Muralidharan S, Roshtkhari MJ, Mori G (2015) Deep structured models for group activity recognition. In: British machine vision conference, pp 179.1–179
37. Hajimirsadeghi H, Yan W, Vahdat A, Mori G (2015) Visual recognition by counting instances: a multi-instance cardinality potential kernel. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2596–2605

38. Deng Z, Vahdat A, Hu H, Mori G (2016) Structure inference machines: recurrent neural networks for analyzing relations in group activity recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4772–4781
39. Li W, Chang MC, Lyu S (2018) Who did what at where and when: simultaneous multi-person tracking and activity recognition. [arXiv:1807.01253](https://arxiv.org/abs/1807.01253)
40. Kaneko T, Shimosaka M, Odashima S, Fukui R, Sato T (2014) A fully connected model for consistent collective activity recognition in videos. *Pattern Recogn Lett* 43:109–118
41. Hasan M, Roy-Chowdhury AK (2015) A continuous learning framework for activity recognition using deep hybrid feature models. *IEEE Trans Multimedia* 17(11):1909–1922
42. Bisagno N, Zhang B, Conci N (2018) Group LSTM: group trajectory prediction in crowded scenarios. In: Proceedings of the European conference on computer vision (ECCV)
43. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L (2014) Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1725–1732
44. Yeung S, Russakovsky O, Mori G, Fei-Fei L (2016) End-to-end learning of action detection from frame glimpses in videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2678–2687
45. Ramanathan V, Huang J, Abu-El-Haija S, Gorban A, Murphy K, Fei-Fei L (2016) Detecting events and key actors in multi-person videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3043–3053
46. Ibrahim MS, Muralidharan S, Deng Z, Vahdat A, Mori G (2016) A hierarchical deep temporal model for group activity recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1971–1980
47. Ibrahim MS, Muralidharan S, Deng Z, Vahdat A, Mori G (2016) Hierarchical deep temporal models for group activity recognition. [arXiv:1607.02643](https://arxiv.org/abs/1607.02643)
48. Tora MR, Chen J, Little JJ (2017) Classification of puck possession events in ice hockey. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 147–154
49. Ibrahim MS, Mori G (2018) Hierarchical relational networks for group activity recognition and retrieval. In: Proceedings of the European conference on computer vision (ECCV), pp 721–736
50. Shu T, Todorovic S, Zhu S-C (2017) CERN: confidence-energy recurrent network for group activity recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5523–5531
51. Li X, Chuah MC (2017) Sbgar: semantics based group activity recognition. In: Proceedings of the IEEE international conference on computer vision, pp 2876–2885
52. Shu X, Tang J, Qi G-J, Liu W, Yang J (2018) Hierarchical long short-term concurrent memory for human interaction recognition. [arXiv:1811.00270](https://arxiv.org/abs/1811.00270)
53. Tsunoda T, Komori Y, Matsugu M, Harada T (2017) Football action recognition using hierarchical LSTM. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 99–107
54. Yan R, Tang J, Shu X, Li Z, Tian Q (2018) Participation-contributed temporal dynamic model for group activity recognition. In: ACM multimedia conference on multimedia conference, pp 1292–1300
55. Azar SM, Atigh MG, Nickabadi A (2018) A multi-stream convolutional neural network framework for group activity recognition. [arXiv:1812.10328](https://arxiv.org/abs/1812.10328)
56. Bagautdinov T, Alahi A, Fleuret F, Fua P, Savarese S (2017) Social scene understanding: end-to-end multi-person action localization and collective activity recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4315–4324
57. Biswas S, Gall J (2018) Structural recurrent neural network (SRNN) for group activity analysis. In: IEEE winter conference on applications of computer vision (WACV), pp 1625–1632
58. Tang Y, Wang Z, Li P, Lu J, Yang M, Zhou J (2018) Mining semantics-preserving attention for group activity recognition. In: 2018 ACM multimedia conference on multimedia conference, pp 1283–1291

59. Lu L, Di H, Yao L, Zhang L, Wang S (2019) Spatio-temporal attention mechanisms based model for collective activity recognition. *Sig Process Image Commun* 74:162–174
60. Vahora SA, Chauhan NC (2019) Deep neural network model for group activity recognition using contextual relationship. *Eng Sci Technol Int J* 22(1):47–54
61. Qi M, Wang Y, Qin J, Li A, Luo J, Van Gool L (2019) StagNet an attentive semantic RNN for group activity and individual action recognition. *IEEE Trans Circuits Syst Video Technol*
62. Gammulle H, Denman S, Sridharan S, Fookees C (2018) Multi-level sequence GAN for group activity recognition. [arXiv:1812.07124](https://arxiv.org/abs/1812.07124)

# Chapter 10

## Artificial Intelligence in Journalism: A Boon or Bane?



Santosh Kumar Biswal and Nikhil Kumar Gouda

### 1 Introduction

With the introduction of artificial intelligence news anchor by China's state news agency *Xinhua*, the world of journalism has witnessed the adoption of the next level of technology [2]. The ongoing transformations in media landscape remain unabated across the globe. The radical digital advancements and innovations could be attributed to the sea changes in information and communication technologies (ICTs) [3]. Such kind of digital revolution is instrumental for the development of a nation. However, the perception and implementation of ICTs differ from a technologically advanced nation to technologically marginalized nation. Moreover, it has invited numerous deliberations, which are diverse from a sector to a sector in which technology is being utilized.

Since technology is one of the key factors for development, its positioning by international agencies for development carries worthy discussions. The United Nations asserts that the use of technology is required to minimize poverty, which can drive society towards sustainable development. Hence, the use of technology and human development cannot be isolated from each other. However, such technological solutions should be judiciously used for societal development.

---

S. K. Biswal (✉)

Symbiosis Institute of Media and Communication (SIMC), Symbiosis International (Deemed University), Pune, Maharashtra 412115, India  
e-mail: [santoshkumar.biswal@gmail.com](mailto:santoshkumar.biswal@gmail.com)

N. K. Gouda

Department of Media and Communication, School of Communication, Central University of Tamil Nadu, Thiruvarur, Tamil Nadu 610101, India  
e-mail: [nikhilkumar@cutn.ac.in](mailto:nikhilkumar@cutn.ac.in)

## 2 Technology and Artificial Intelligence

Artificial intelligence (AI) has been an important part of the technology industry. As an academic discipline, AI came into existence in 1956 [4], and since then, it has been experiencing a series of optimism and pessimism. AI, an area of computer science, stresses on the creation of intelligent machines to work and react like human beings. For this, computers with AI cover the aspects of speech recognition, learning, planning, and problem solving. AI can be divided into analytical, human-inspired, and humanized artificial intelligence [5].

In this twenty-first century, AI is being used in the field of health care, automotive, finance and economics, video games, military, audit including advertising, journalism, and various other branches of media and communication. It has become an instrumental to resolve the issues in computer science, software engineering and operations research [1, 6]. Hence, AI can be associated with all sorts of area in which the efficiency of a human being can be enhanced.

Machine learning (ML), a subset of AI, is the scientific study of algorithms and statistical models that computer systems perform various assignments without using clear instructions [7]. The use of a machine is immense in the field of agriculture, banking, communication, sentiment analysis, software engineering, user behavior analytics, search engines, and the like. Even though its application is very important in various fields, machine learning suffers from certain shortcomings. Lack of suitable data, biases in choosing the data set, wrong algorithms and lack of resources, and evaluation could be the reasons for the underperformance of such kind of technology.

## 3 Technology, Medium, and Communication

Taken literally, Marshall McLuhan's famous quote 'medium is the message' [8], we can find that the medium is getting more emphasis since the applications of AI are in practice. As a result, the process of communication including the source, message, and receiver is also influenced by the technology-driven by AI. On the contrary, the stand of Manuel Castells is different.

In the context of discussing the medium of communication and AI, digital journalism and online activism come into the picture. Digital journalism and online activism are interrelated. Online activism which is technology-driven is mobilizing social movements [9]. However, going beyond the information society, Manuel Castells [10] opines that it is not the technology, which is the key to social structure and social movement, but social networks which manage the technologies used for information dissemination. Along with the importance of the message, the vitality of the medium cannot be avoided.

The ICTs used for internal communication play a vital role in any organization [11]. They are also being utilized in the process of teaching and learning. Both teaching and learning are forms of communication. The power of technology is immense

in the domain of business communication. Ump teen forms of business communication—advertisements, user-generated contents, contents circulated from business establishments in social media—are proved to be more effective in the field of marketing [12]. Emphasizing more on business communication driven by technologies, Chakraborty and Bhat [13, 14] assert that digital communication in the area of brand communication is doing rounds these days. Consumers are getting empowered with the interactive tools of communication which are used for brand assessment. In this context, technology has major bearings on medium and communication, in shaping and reshaping society.

## 4 What is Journalism?

Journalism is the collection, preparation, and distribution of news through print, electronic and digital media blogs, webcasts, podcasts, and social networking sites (Encyclopædia Britannica [15]). However, the process of journalism has undergone sea changes with the passage of time and the adoption of newer technologies.

Journalism is a dynamic field. The print and electronic media have their own space in the production, distribution, and consumption of news. However, to some scholars, the dominance of print media is declining in developed nations. News in social media, blog, *WhatsApp*, and other forms of digital media has become a major landmark in journalism. In this context, [16] finds that *Facebook* is a digital medium for the dissemination of news. This medium has been touted as a potential digital platform which can engage the audiences on the messages on health and fitness in India [17]. Besides, [18] asserts that social media as an alternative media platform which has become a tool of protest against oppression.

Blogging, a kind of digital media, is making rounds in the space of journalism for a long time. Bulatova et al. [19] highlight that it has the right to communicate like the traditional media. Journalism practitioners have accepted the space of blogging to enhance their audiences. Across the globe, social media has become a popular platform for news consumption [20] despite the issues of the veracity of content and other news values. The regular technology adoption in newsrooms has seemingly become a professional practice to add value to the dissemination of information to the audiences. Broadly, the process of digitalization has brought significant changes in news consumption pattern among young generation [21]. Therefore, the medium of dissemination of news remains vital in journalism.

Journalism in India has departed from a creative sphere to a business entity. Journalism has created a place for its own. However, rampant defamation against media persons, unnecessary pressure on whistleblowers and RTI (Right to Information), poor level of media activism, and political infotainment has made news media outlets dysfunctional in this largest democratic nation of the world [22]. Development journalism required for the nation is in a fragile condition. However, it has opened up doors for other alternative means of journalism. Digital media has become an instrument in mobilizing social-political movements in India [23]. The advent of

alternative digital platforms like *The Quint*, *The Wire*, *Firstpost*, and *Daily O* is considerably free from government and corporate interferences in terms of disseminating information.

The rise of community media, a source of alternative media, could be one solution to give voice to the voiceless [24–26]. Citizen journalism, one form of alternative media, is proliferating its space in the field where mainstream and business-driven media outlets dominate [27]. However, it cannot replace trained journalists, as they are part and parcel of news ecology. Everybody will believe in one thing that the world is gripping with the issue to differentiate between truth and myth [4, 28].

Similarly, media education in India is also not free from flaws. Journalism education is at a crossroads as there is a swift pace of digitization and globalization of media [29]. Unfortunately, classroom teaching in several places is not able to accommodate these changes. Hence, the pedagogy about media education should consider the technological improvements including the foray of artificial intelligence.

## 5 Journalism Before Artificial Intelligence

Before discussing the impact of AI in the field of journalism, it is essential to assess the impact of technology in the given field in a chronological order. In this context, understanding and deliberating on virtual reality (VR) and augmented reality (AR) are of utmost importance. Since the medium is the message and medium often undergoes a series of changes due to technological advancements, understanding the medium and the impact of the technology remains a critical area to probe the journalism practice and journalism education.

The use of VR and AR has been immense. The VR is an innovative way of experience taking place within a simulation. Such kind of simulation can be related to or entirely different from reality. It can also be described as a specific type of reality emulation [30]. The use of VR has been witnessed in the field of education and entertainment. On the other hand, AR is an interactive way of taking the experience of a real-world environment. It is accomplished by computer-generated perceptual information, multiple sensory modalities, and the like. The uses could be constructive or destructive.

The research and development reveal that VR has made the journalistic practice exciting. It enables the audiences to come closer to a news story than any other previous format of storytelling. The 360-deg, stereoscopic video and updated formats of headsets have pushed the journalism profession to a newer height. The Tow Center for Digital Journalism finds that a blend of technology, narrative structure, and journalistic determination plays a decisive role in measuring and determining the degree of agency given to users in a VR experience [31]. Certain prominent media conglomerates like *The New York Times* and the *BBC* continue to experiment with newer technologies [32]. However, sometimes it is not advisable to use the technology recklessly at the cost of accuracy, creativity, and human employment.

Max Boenke, Head of video, *Berliner Morgenpost*, has stated that nowadays news organizations are frequently using 360 for stories, which sometimes may not be interesting. It may deter the audiences to watch the news content further. On the contrary, previously associated with *BBC Research & Development* has opined that if VR contents are perfectly made, it can empower the journalism field. It can make many wonders [33]. Such type of technology is very much useful for science communication.

A study has found that journalism has become a driving force for taking and executing VR mainstreaming. The scope of journalism has enhanced in terms of topic, style, and scope. However, the use of VR has brought challenges in terms of journalistic norms and practices [34]. Another study has found that in the domain of VR, journalism remains a minor section. However, VR has enabled the emergence of immersive journalism which has fueled the media industry and media education. The advent of the theoretical and conceptual framework is providing a Philip to future academic and industry endeavors through immersive journalism. Hence, the impact of VR on journalism remains a mixed bag of advantages and disadvantages.

Similarly, the AR has a significant impact on journalism. The content of journalism has undergone multiple changes with the advent of AR. It has enhanced the audience engagement which is not available in the traditional form of disseminating information. Moreover, this technology has provided more contextualized information in the age of fast-paced journalism [35]. Hence, one more aspect of AR could be that it has fueled citizen journalism and user-generated contents which are being produced, distributed, and consumed by the citizens themselves. There is no surprise that VR and AR are the change agents in the field of journalism. However, AI has started proving more influential than VR and AR.

## 6 Journalism and Artificial Intelligence

When news stories are produced automatically by computers instead of human reporters, it is called as automated journalism, algorithmic journalism, or robot journalism. By the virtue of AI, the news is interpreted, organized, and presented in human-readable ways. It involves algorithm which processes the huge amount of data, picks from pre-programmed article structures, places crucial points, and inserts the requirements like names, statistics, figures, and the like [36].

For digital news projects (2019) by Reuters Institute and University of Oxford, Ritu Kapur from *The Quint*, a digital platform from India, has stated that there is a need of AI and human intelligence. *The Quint* is a leading online platform, which disseminates the news at a faster pace. In the same report, Lisa Gibbs from *Associated Press* has opined that the requirement of journalists will be there all the time. In addition, the use of newer technologies will assist these journalists to be more efficient in varied dimensions. Moreover, with the help of AI, the news industry will serve the audiences better. It can also debunk false information to maintain the norms of ethical journalism [32].

Chinese news apps like *Jinri Toutiao*, *Qutoutiao*, and *Kuaibao* are immensely used to provide personalized news from a range of news providers [32]. AI enables to personalize the media contents in order to recommend better to its audiences. By the virtue of robot journalism, more and more stories and videos can be incorporated. AI provides technological support to journalists in the age of information overload syndrome.

The journalistic practice has gone trendy in the light of understanding, researching, and implementing AI. The following can be summarized.

## **6.1 Quantitative Getting in Journalism**

The quantitative formats have become the new phenomena in modern-day journalism [37]. This new kind of journalism has created a special space in academic literature and media practice. With the functioning of AI, the quantitative format of journalism has transformed to the next level. As a result, the production, distribution, and consumption of media contents have been redefined.

## **6.2 Data Journalism**

Data journalism is a newer format of journalism in which there is a use of an increased amount of numerical data. Data are being used in the production of news stories. They are used to make the stories easier to understand for the audiences. With the help of AI, data journalism facilitates the audiences to understand the complex concepts used in news stories [38]. The proliferation of digital outlets has intensified the availability of data for the journalistic process. However, all-time data journalism does not necessarily require AI. AI could be part of such type of journalism to make the communication process more effective.

The areas covered in data journalism are—cybercrime reporting, computer-assisted reporting, and data-driven journalism, infographics, data visualization, interactive visualization, serious games accommodating the interaction in advanced levels and information management system [39]. In India, data journalism is getting popular with the emergence of alternative digital platforms like *Newslaundry*, *The Quint*, and *IndiaSpend*. The news stories like ‘Shoddy Sanitary Napkins Impact Menstrual Hygiene Drive’ and ‘Why Mumbai Fire Brigade Gets 1 Structural Collapse Call Each Day’ in *IndiaSpend* can be cited in this context.

### 6.3 Algorithm Journalism

Algorithm journalism, a newer format of journalism, involves digital processing which come into play where there is an intersection between journalism and data technology. Moreover, in the process, a combination of algorithms, data, and knowledge is the major ingredients to enhance the credibility of journalism [40]. In such type of journalistic practices, the use of AI is inevitable. When it comes to the journalistic practice in India, AI is yet to be utilized. The media organizations are in the mode of researching its applications in light of cost, speed, and employment.

### 6.4 Automated Journalism

When there is an increased volume of news contents produced and distributed for the consumption of audiences automatically, it is called automated journalism. It is an algorithm process which enables the data set to be converted into news stories for human interest and readability [41, 42]. This can only be possible when AI is used in newsrooms. AI mobilizes the newsroom in varied manners—streamlining the media production process, automating the routinized tasks, crunching more data, exploring media insights, minimizing the fake news, and delivering the requirements.

The leading media houses like *The New York Times*, Reuters, *The Washington Post*, *Quartz*, *Yahoo*, *Associated Press*, *The Guardian*, and *The BBC* have adopted AI in their newsrooms. In an experimental mode, *The New York Times* has executed its AI project ‘Editor’ in 2015 in order to simplify the journalistic production process. The aim of the project was to simplify the journalistic process in the newsroom. When writing an article, a journalist can use tags to highlight phrase, headline, or main points of the text. By using various tools through AI, *The New York Times* has attempted to moderate the readers’ comments and encouraged constructive discussions and at the same time overcoming the abusive remarks. Needless to say, *The BBC* has a huge amount of data comprising news, features, and videos. Since 2012, it has been using *Juicer*, a data extraction tool to link all the data more accessible and more meaningful. Since 2016, *Reuters* has been using AI with assistance from semantic technology company Graphiq. With the help of AI, it is able to provide data-driven news stories, which are visually stimulating and easy to understand. Apart from providing speedy access to data, AI also allows the publishers to get the information in terms of simple tables or charts [43].

The use of Heliograf smart software in *The Washington Post*; Automative Insights—a prominent natural language generation vendor in *Yahoo*; Semantic Discovery and News Whip in *Associated Press*; and Chatbot Media Interfaces in *The Guardian* and *Quartz* have been the indicators of adopting AI in newsrooms worldwide. In India, this format of journalism is yet to take off. Leading media houses like *The Times of India*, *Hindustan Times*, *The Hindu*, *The Telegraph*, *The Indian Express*, *NDTV*, and *India Today* may experiment AI to speed up the journalistic process.

## 7 Artificial Intelligence Becoming Fruitful

The use of AI in journalistic practice has several advantages. Firstly, AI has overcome certain contemporary journalistic issues. Journalists are able to analyze the data from several sources. Apart from analyzing the images, they can convert the spoken words into texts, texts to audio and video. They are able to overcome the issues of information overload, lack of credibility, and shoddy journalism. Secondly, today the journalists are facing the issues of fake news and misinformation. Professor Kalina Bontcheva has further identified the prevalence of fake information on social media [44]. With the help of AI, they can deliver enhanced news quality and accuracy by identifying and dismantling the fake news. Journalists are getting benefitted quickly by automated fact-checking [45]. Thirdly, AI has quickened the news editing process as per given editorial policy. It has brought relief for the journalists who boringly slug in the newsroom. Software is available to collect news and later to rephrase it, according to the prescribed editorial policy without any human interventions. *The Associated Press* uses urbs to distribute news stories to various media houses [36]. Fourthly, AI has facilitated a personalized news agenda, which differs from a media house to another. By the virtue of content personalization, it can provide news services in multiple languages, keeping the larger audiences across the globe in mind. Fifthly, AI has propelled the speed of journalistic practice. Robot reporters are able to produce news stories [1] at a faster pace. *The Associated Press* has confirmed that AI has enhanced customer services by more than ten times. Lastly, AI has fetched a robust defense against such manipulation and propaganda that can endanger a nation's security. The Chinese government is using AI to track objectionable contents, dissent, and propaganda messages. Certain countries are using AI to probe foreign interference in elections by understanding the contents on *Facebook* and other social media outlets [1, 46].

## 8 Challenges Before Journalism

Even though AI has brought revolution in the profession of journalism, it is not free from shortcomings. Since these issues are critical, it could be detrimental to the profession. Ethical challenges before journalism are major hurdles which need to be suitably handled.

Firstly, there can be a lack of credibility and quality in AI-driven journalism. The automated news stories may not render the journalistic credibility. It is also commented from several quarters that machines cannot replace human capabilities. The space for creativity, humor, and critical thinking will remain forever in the field of journalism. Secondly, there are confusions over-crediting the authorship of news stories which are automated by AI. Who will get the credit among the reporter or the participants in the algorithmic process? Thirdly, AI has been a potential threat to the volume of jobs in the profession. News organizations are more interested

to adopt AI in order to avoid costs in human resources apart from fastening the newsroom processing. Fourthly, legitimate concerns can grip the journalistic practice driven by AI. As of now, technological developments have no solutions to legal problems emanated from algorithm-generated content about private citizens. News organizations may not be able to defend the legal issues which could be because of algorithm-driven news stories in *Google* and other similar digital news platforms. Fifthly, data utilization has been an issue in AI [47, 48]. The security and privacy of data have often been an issue to overcome for developers and governments. To bring correct, objective, and accurate data, news organizations using AI should shoulder ethical duties for the time being.

## 9 Way Forward

In the age of science, technology plays a vital role in society. Technology keeps on changing with the pace of time. Therefore, in the context of AI and machine learning in journalism, what needs to automated should be automated. It has reorganized the newsroom as never before in several developed countries. Participatory culture is getting exercised in newsroom setup across the globe. However, the adoption of technologies should not push this professional field into a tailspin. The prediction that the newsroom of 2025 to be run by AI will be witnessed in the years to come.

To some, in the future, larger media contents will be produced with the help of AI. In a study report, 78% of the industries has agreed on the fact that it is high time to invest in artificial intelligence in the field of journalism [32]. Technology through AI can act as an enabler for better journalism and more impactful journalism. It can pave the professional way in aligning media contents with social good [46].

As the domain of journalism is technology-driven, the industry will shift from time to time with changes in AI. However, as AI is more into play, it would not pose a threat to the profession and employment [38]. It can further add values to the journalists in the digital age. The machine would not completely replace the journalists. Rather machines will enhance the journalistic skills in more sophisticated manners. The presence of human journalists is inevitable no matter how much technology changes.

The use of AI in the field of journalism in India will be a learning and experimental curve. Ramesh Menon, an author, and award-winning journalist asserts, AI is already being experimented by the Chinese in newsrooms to write news stories and features. Also, other countries are testing it. It is just a matter of time when AI would be dominant in Indian newsrooms and even media management systems that will use it to figure out consumer profiles and needs to keep up with the changing times and stiff competition. We do not know what the next five or ten years are going to be and are at a loss in the classroom how to prepare media students for the future. We do not know how penetrative AI is going to be and how it will affect jobs.

Will Indian media houses invest in AI writing news stories? Of course, they will. And, why not? After all if you feed in the required information, the robot would figure out how to pick up relevant information, the kind of intro to writing, how to

structure the story logically, what conclusion it should have based on the research it does from the Internet, the graphs and illustrations and the photographs to be secured for the story that will not have copyright issues. Who will say no to this? However, the fact is that the best stories will come from writer-journalists who can put in fine details, empathy, drama, color, and analysis into their stories. What is really good in the changing scenario where AI will come in is that tomorrow we can get robots to do the routine stuff that today takes 80% of the journalists' time. This can help the reporters and editors concentrate on big-ticket stories that require a lot of footwork in terms of getting to the right people, getting them to talk, analyzing the present and even talk of the way forward. Their time can be better utilized if they have robots to help them do the normal sundry work. Dynamic changes are happening, and we as journalists can see that. AI and robots will write stories in the future, and they will get better at doing it as humans will fine-tune it. Fine-tuning has to be done as there have been instances of AI going completely wrong in figuring out the news story. Instead of being overexcited, we must be very cautious. The human interface, therefore, cannot be completely ruled out as human intelligence to tell the right from the wrong will be dominant. Imagine what will happen if AI gives a wrong headline or a wrong interpretation?

It will just be a matter of time when AI-assisted automated reporting systems and machine learning techniques to sift through massive data to write news reports. Whether we like it or not, it is going to affect media jobs. In another five years, we would know. That is why media schools must start teaching techniques that will equip them for the future and not get stuck on teaching the inverted pyramid style of writing which the robot will do. They will have to have different skills, and media schools will do well to stress on ethics which robots will not do.

Interestingly, *Google* has coughed up \$805,000 to build software that will gather, automate, and write nearly 30,000 local stories every month to British news agency Press Association. Labeled as reporters and data and robots, the software will automate local reporting with large public databases from government agencies or the local police.

*Yonhap*, a news agency in South Korea, has introduced an automated reporting system to produce news on football games. Machine learning algorithms are being already employed to write stories by *Thomson Reuters* and *Associated Press* and *The New York Times*. Others are using it to beef up their research. Web sites hungry for content and news Web sites eager to be the first with news and analysis are going to use AI shortly. Very soon, you will not be able to even think of quality content generation and the speed with which it is required without AI. Menon concludes that we might be apprehensive of losing jobs, but in the final analysis, no one will be able to replace a good journalist who can write stylistically and turn phrases into very readable copy or even sit down and use his or her knowledge to analyze the turn of historic events.

Suffice to say, AI will have an immense impact on the ecosystem of media market round the globe. One the one hand, the technology can have enough scope to create social good where it can assist the human to navigate the required data out of a huge pool of data by personalized recommendations. On the other hand, AI can

manufacture the media contents as human needs which could not be beneficial to humankind. It could only happen by deceiving media audiences. By the path of business model with utter manipulations, it may reduce the social good to business good which can only be a bubble of business for a short period. Hence, ethical challenges need to be amicably resolved. Its use and human resource should strike a chord in the industry. Later, there will a clarion call to use AI in the field of journalism for the greater interest of humankind.

## References

1. Peiser J (2019, February 5) The rise of the robot reporter. The New York Times. Retrieved from May 26, 2019. <https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html>
2. Kuo L (2018, November 9) World's first AI news anchor unveiled in China. The Guardian. Retrieve from July 20, 2019. <https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china>
3. Wölker A, Powell TE (2018) Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, 1–18
4. Simon HA (1965) The shape of automation for men and management. Harper & Row, New York
5. Kaplan A, Michael H (2018) Siri, Siri in my hand, who's the fairest in the land? on the interpretations, illustrations and implications of artificial intelligence. *Bus Horiz* 62(1):15–25
6. Clark J (2015, 8 December) Why 2015 Was a breakthrough year in artificial intelligence. Bloomberg News. Retrieved from July 18, 2019. <https://www.bloomberg.com/news/articles/2015-12-08/why-2015-was-a-breakthrough-year-in-artificial-intelligence>
7. Salathé M, Vu DQ, Khandelwal S, Hunter DR (2013) The dynamics of health behavior sentiments on a large online social network. *EPJ Data Science*, 2(4). DOI:<https://doi.org/10.1140/epjds16>
8. McLuhan M (1967) Understanding media: the extensions of man. Sphere Books, London
9. Basu P, De S (2016) Social media and social movement: contemporary online activism in Asia. *Media Watch* 7(2):226–243
10. Castells M (2012) Networks of outrage and hope: social movements in the internet age. Polity Press, Cambridge, UK
11. O'Donovan T (1998) The impact of information technology on internal communication. *Educ Inf Technol* 3(1):3–26
12. Trivedi J (2014) Effectiveness of social media communications on Gen Y's attitude and purchase intentions. *J Manag Outlook* 4(2):30–41
13. Chakraborty U, Bhat S (2017) Credibility of online reviews and its impact on brand image. *Manag Res Rev* 41(1):148–164
14. Chakraborty U, Bhat S (2018) Online reviews and its impact on brand equity. *Int J Internet Mark Advertising* 12(2):159–180
15. Encyclopædia Britannica (2019) Journalism. Retrieved from July 10, 2019. <https://www.britannica.com/topic/journalism>
16. Patankar S (2015) Facebook as platform for news dissemination, possibilities of research on Facebook in Indian context. *Amity J Media Commun Stud* 6(2):49–56
17. Jaggi R, Ghosh M, Prakash G, Patankar S (2017) Health and fitness articles on Facebook-a content analysis. *Indian J Public Health Res Dev* 8(4):762–767. <https://doi.org/10.5958/0976-5506.2017.00428.4>
18. Kusuma KS (2018) Media, technology and protest: an Indian experience. *Language in India*, 18(7). Retrieved from August 12, 2019. <http://languageinindia.com/july2018/kusumamediatechnologyprotest.pdf>

19. Bulatova M, Kungurova O, Shtukina E (2019) Recognizing the role of blogging as a journalistic practice in Kazakhstan. *Media Watch* 10(2):374–386
20. Yusuf Ahmed IS, Idid SA, Ahmad ZA (2018) News consumption through SNS platforms: extended motivational model. *Media Watch* 9(1):18–36
21. Ghosh M (2019) Understanding the news seeking behavior online: a study of young audiences in India. *Media Watch* 10:55–63
22. Biswal SK (2017) Role of the media in a democracy revisited. *Vidura* 9(1):19–20
23. Pradhan A, Narayanan S (2016) New media and social-political movements. In Narayan SS, Narayanan S (eds) *India connected*. Sage, New Delhi
24. Dash B (2015) Community radio movement: an unending struggle in India. *J Dev Commun* 26(1):88–94
25. Dash B (2016) Media for empowerment: a study of community radio initiatives in Bundelkhand. (PhD Thesis). Tata Institute of Social Sciences, Mumbai
26. Pavarala V, Malik K (2007) Other voices: the struggle for community radio in India. Sage, Thousand Oaks, California
27. Biswal SK (2019) Exploring the role of citizen journalism in rural India. *Media Watch* 10:43–54
28. Simons M (2017, April 15) Journalism faces a crisis worldwide—we might be entering a new dark age. *The Guardian*. Retrieved from July 15, 2019. <https://www.theguardian.com/media/2017/apr/15/journalism-faces-a-crisis-worldwide-we-might-be-entering-a-new-dark-age>
29. Raman U (2015, December 1) Failure of communication: India must face up to the rift between its newsrooms and classrooms. *The Caravan*. Retrieved from July 19, 2019. <https://caravanmagazine.in/perspectives/failure-of-communication-rift-between-india-newsrooms-clasrooms>
30. Virtual Reality Society (2017) What is virtual reality? Retrieved from July 9, 2019. <https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html>
31. Owen T, Pitt F, Aronson-Rath R, Milward J (2015) Virtual reality journalism. Retrieved from July 10, 2019. [https://www.cjr.org/tow\\_center\\_reports/virtual\\_reality\\_journalism.php](https://www.cjr.org/tow_center_reports/virtual_reality_journalism.php)
32. Newman N (2019) Journalism, media, and technology trends and predictions 2019. Digital News Project, 2019. Retrieved from May 31, 2019. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Newman\\_Predictions\\_2019\\_FINAL\\_2.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Newman_Predictions_2019_FINAL_2.pdf)
33. BBC Academy (2017, November 7) Virtual reality journalism: is it the new reality for news? Retrieved from July 10, 2019. <https://www.bbc.co.uk/academy/en/articles/art20171107112942639>
34. Mabrook R, Singer JB (2019) Virtual reality, 360° video, and journalism studies: conceptual approaches to immersive technologies. *Journal Stud*. DOI: <https://doi.org/10.1080/1461670x.2019.1568203>
35. Pavlik JP, Bridges F (2013) The emergence of augmented reality (AR) as a storytelling medium in journalism. *Journal Commun Monogr* 15(1):4–59
36. Graefe A (2016) Guide to automated journalism. Columbia Journalism Review, New York. Retrieved from July 19, 2019. [https://www.cjr.org/tow\\_center\\_reports/guide\\_to\\_automated\\_journalism.php](https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php)
37. Coddington M (2015) Qualifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digit Journal* 3(3):331–348
38. Veglis A, Bratsas C (2017) Towards a taxonomy of data journalism. *J Media Critiques* 3(11):109–121
39. Williams II D (2017, December 6) The history of augmented reality (Infographic). Huff Post. Retrieved from July 9, 2019. [https://www.huffpost.com/entry/the-history-of-augmented\\_b\\_9955048](https://www.huffpost.com/entry/the-history-of-augmented_b_9955048)
40. Hamilton JT, Turner F (2009) Accountability through algorithm: developing the field of computational journalism. Behavioral Sciences Summer Workshop, Stanford. Retrieved from July 16, 2019. <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf/>

41. Carlson M (2015) The robotic reporter: automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digit Journal* 3(3):416–431
42. Galily Y (2018) Artificial intelligence and sports journalism: is it a sweeping change? *Technol Soc* 54:47–51
43. Underwood C (2019, January 31) Automated journalism—AI applications at New York Times, Reuters, and Other Media Giants. *Emerj Artificial Intelligence Research*. Retrieved from July 20. <https://emerj.com/ai-sector-overviews/automated-journalism-applications/>
44. Ali W, Hassoun M (2019) Artificial intelligence and automated journalism: contemporary challenges and new opportunities. *Int J Media Journal Mass Commun* 5(1):40–49
45. Graves L (2018) Understanding the promise and limits of automated fact-checking. Retrieved from July 16, 2019. [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves\\_factsheet\\_180226%20FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf)
46. Sullivan D (2016, December 24) Google's top results for 'did the Holocaust happen' now expunged of denial sites. Retrieved July 15, 2019. <https://searchengineland.com/google-holocaust-denial-site-gone-266353>
47. Monti M (2019) Automated journalism and freedom of information: ethical and juridical problems related to AI in the press field. Retrieved from July 10, 2019. <http://www.opinioruriscomparatione.org/opinio/article/view/126>
48. Wang W, Siau K (2018) Ethical and moral issues with AI: a case study on healthcare robots. In: Twenty-fourth Americas conference on information systems. Retrieved from July 16, 2019. file:///C:/Users/HP/Downloads/EthicalandMoralIssueswithAI.pdf

# Chapter 11

## The Space of Artificial Intelligence in Public Relations: The Way Forward



Santosh Kumar Biswal

### 1 Introduction

The industry of media and communication keeps on evolving and is ceaselessly moving forward. Various types of mass communication—journalism, advertising, Public Relations (PR), social media, audio, film and television, and photography—have been witnessing sea changes with technological interventions. Imparting education, itself a form of communication, is influenced by certain technologies. Certain technologies have made their marks in the field of educational pedagogies [15]. Going further, technology is being used in varied other services including banking sector. Banks offer chatbots to improve customer service. Chatbots form as an information system, which is essential for examining customer experiences [18]. Mobile communication and other means of information and communication technologies are being used for healthcare facilities [11, 12]. E-governance is able to meet the requirements of citizens in building a progressive nation [10]. Moreover, with the explosion of the Internet, E-commerce sector is expanding in which the optimum utilization of big data can be possible for bigger business possibilities [6]. Since consumers are active on digital platforms, it is imperative to understand online reviews on functional and hedonic brand images, which are required for the promotion of business and branding [2]. Moreover, in the Indian context, Pandey [13] finds that it is the Internet, a type of technological innovation which could speed up the developmental process.

Understanding the pattern of communication in journalism, film, or business by tapping the big data is essential. Overall, communication can be art-oriented or business-oriented. Business communication, an applied form of communication, remains an essential characteristic of the management of a business. It is the information disseminating among people within and outside an organization. Such kind

---

S. K. Biswal (✉)

Symbiosis Institute of Media and Communication (SIMC), Symbiosis International (Deemed University), Pune, Maharashtra 412115, India

e-mail: [santoshkumar.biswal@gmail.com](mailto:santoshkumar.biswal@gmail.com)

of communication is executed for the sake of commercial interests of an individual, group, or organization. Business communication can be internal or external, upward or downward, formal or informal, lateral or interactive, and mass or grapevine [8]. Such communication is goal-oriented and tries to resort to suitable channels of communication. It could be advertising PR and the like.

PR, a form of business communication, is the dissemination of information between an individual or an organization and its public. PR professionals attempt to build and maintain relationships between the organization and its target audience, the media, and other opinion leaders. Such kind of business communication uses the subject or topic or news items, which may not demand direct payment. On this line, PR is separated from advertising and other forms of business communication [17]. PR, tied with historical roots, keeps on updating with the pace of time and requirements. Technological advancements are the factors, which have been renewing the PR activities from time to time. In this context, understanding and discussing artificial intelligence (AI) and Machine Learning (ML) are important.

## 2 Artificial Intelligence, Machine Learning, and Communication

ML is the scientific study of algorithms and statistical models that computer systems execute certain assignments without using explicit instructions. ML algorithms are being used in various fields. It is a subset of artificial intelligence (AI). With the development of AI, ML and natural language processing along with new technological platforms, it is feasible to dehumanize the processing of large quantity of publicly available data [16]. Such kind of machine application has changed the process of communication. In contrast, communication has been considered as a human process often mediated by technology [3]. Adopting the mode of AI, the *Associated Press* has changed the pattern of production and distribution of news. The technology is being used from interpersonal interlocutor to content producer. Amazon's Alexa is programmed to meet human queries and needs. AI is automating and fastening the pace of the communication, and subsequently, social processes are getting reliant on it [5]. Therefore, there is a departure from historical role of media to a new emerging role of business and social communication. Business communication, a form of communication with technological interventions, is getting more efficient for the dissemination of consumer information [9]. Such kind of machine application is also beneficial in the hiring process and the recruitment industry, which are beneficial for clients and candidates as well [19]. Therefore, such kind of machine utilizations can be commercially oriented and utilized.

### 3 PR Activities and Artificial Intelligence

Needless to say, AI has slowly come into play in communications industry. In the domain of PR, AI has the capacity to frame the data-driven contents and handle the crisis. It also understands the upcoming media trends. As of now, only prominent PR agencies have been able to tap the power of AI in their daily works. It is being used to enhance the capabilities of people. As a result, people working in PR firms are able to spend their time on creative activities [14]. Bourne [1] finds that due to ignorance of AI, the level of diversity in the PR functions may be lowered. Therefore, the role of such technologies has become essential to make PR activities effective.

#### 3.1 Data-Driven PR Campaigns

With the inputs from AI, the creation of new campaigns can be possible. It can also help a PR firm to get rid of the guess works. The automation and ML assist the professionals to understand which elements will pay the success to PR campaigns. Since a machine does faster than human beings, it is easy to take fast and accurate decisions, which are beneficial for the client concerned. It helps to understand and foresee the trend, which is ultimately required for decision-making process [14]. Such kind of machine inputs has been proved fruitful for qualitative and quantitative decisions. It assists in sorting out the time, content, medium, and audience of the campaign. By employing AI, PR persons can produce hyper-specific materials, which will be best suitable for their clients' requirements. It can lessen the time wastes on content creations for specified audience.

#### 3.2 Automating Routine Works

AI has attempted to bring relief to the PR professionals from mundane tasks. Routinized or repetitive works are being easily accomplished with the intervention from AI. By the virtue of this technology, Robotic Process Automation (RPA) is making several regular works possible. Scheduling calendars, structuring meeting notes, and other similar works are done with the machine used in the firm. The technology is bringing relief to PR persons from the works like administration, crunching numbers, and organizing files. Empowered with the technology, they are able to create, organize, and prioritize task in their firms to meet their clients' requirements. Certain instances have revealed that PR firms have started engaging with automating things. Earnings' reports can be of one such instance, which maximizes their creative assignments. Since several of the works are completed with AI, PR persons are more engaged with project ideation and venturing into newer avenues.

### **3.3 *Sentiment Analysis and Crisis Management***

According to Mentionlytics Report, online crisis is a severe problem for a company. In this context, AI is being used in sentiment analysis. Sentiment analysis is also known as opinion mining or emotion. AI indicates the use of natural language processing, text analysis, computational linguistics, and subjective information. This analysis is being applied to the voice of the customer materials. Its uses are from marketing to customer service to clinical medicine. With the interventions from the machine, sentiment analysis uses the natural language processing to separate vocabulary use, tone, and language settings. AI facilitates the PR companies to address the press that may arise not considering sentiment. It enables the PR persons to analyze several factors including social listening, which ultimately helps the clients to keep their brand values intact. Moreover, it assists the PR agencies to handle the clients' adverse situations. Nowadays, AI systems are interpreting the context and are able to attribute the true meaning.

With AI comes more into the picture, the technology proves beneficial in creating smarter chatbots which are being used to interact with consumers as brands [4]. In order to mobilize social interaction, chatbots are acting as instrumental to follow the relevant hashtags or respond to messages. They are also being used for effective communication. Enhancing the social interactions is the job of PR professionals. During the time of crisis, generally PR persons become more knee-jerk instead of proactive. However, with ample information delivered by AI, they become more sensitive and pragmatic.

They are able to create ready-made messages for their audience. Admittedly, AI is the development of computer systems to carry out number of jobs, which need human intelligence. Broadly, this technology enables the PR persons to allocate press releases, form media lists, transcribe audio and video into text, forecast media trends, and observe the social media.

## **4 PR, AI, Media Education, and Research: Growing Perspectives**

Media education on business communication and the use of AI are interrelated. Imparting the uses of AI to the students in the field of media communication is essential to tap its maximum utility, which will prove beneficial for the development of business.

A PR educator emphasized that AI can make PR easier and effective in several ways. When handling routine customer queries and grievances especially on Web-mediated applications such as e-mail, online chat support, social media, and on telephone (including voice calls and SMS/Voice messages), AI can become useful as separate human resources need which not be diverted for this purpose [7]. Providing timely updates, the dissemination of information to the media, updating company's

PR tools such as its Web sites, social media pages on *Facebook*, *Instagram*, and *Twitter* can all be carried out efficiently by AI.

PR research is another area where AI would be extremely effective. Real-time sourcing of information related to the company being posted on the Web by media outlets or other users can be carried out instantly. Analysis of opinion polls, consumers' feedback, and monitoring of various platforms can be entrusted to AI. With the growth of such technology, it is not difficult to use AI to draft press releases. It is possible to present a technically correct press release based on the information fed into the system and the situation at hand. However, this is one area where this technology may fail to make a mark. Only humans can truly guess the pulse of the audience/general public at large. The press release from an AI-operated robot may be transparent, simple, and direct. However, it may miss the human touch of understanding and conveying the message emotionally. Secondly, while AI may take over efficiency, trust between an organization and its stakeholders would suffer. People trust a brand that delivers the promise of 'customer service' but not by opting the easy way out with AI. Also, when it comes to disaster management and crisis PR, it is only humans who can truly take decisions and win back trust from the company's stakeholders.

Another media educator of eminence who has been associated with researches in PR finds that the use of AI in PR is very less talked about topic and practically abysmally used phenomenon in India. However, at global platforms, its importance has been acknowledged in the form of research for PR campaigns, automation of routine yet important tasks and analysis of people's sentiments and crisis communication. Since PR is understood as image-building exercise by an organization, the effective use of media has become highly automated. Eventually, PR has to depend upon AI for many types of groundworks. Also, PR tries to mold public opinion in the direction of favorable image of the organization, so human acumen and wit are essential to handle any difficult situation. Therefore, human skills supported by AI to fasten the procedure are best way to conceptualize, execution, and completion of PR activities. One cannot depend completely on it for such activities since AI does not differentiate between people and machine and does not know how to handle the emotions of the human beings. On the other side, humans are not as fast as machines and many times, they can be biased with the data. So, AI should be used to complement the lacunae of human beings, and human beings should be used to complement the shortcomings of AI. Then, only we can get the best results.

A researcher in the field of PR and AI explores that data crunching is a big issue, which requires ample amount of time. AI and ML cater to the needs of millennial consumers. It provides a unique user experience in terms of business communication to PR persons. The services like 'speech to text conversion', 'sentiment analysis', 'massive data analysis', and 'identification of common problems' are worth-mentioning. Though the need of AI is increasing, the requirement of human does not decrease. Only the approach is changing that for our benefit, some strata of the business world consider AI is as a threat, if we look back into technological help to us. We have transitioned from letters to E-mails and *WhatsApp*. We can not deny the fact that life

has become much easier and communication is growing strong with each passing day.

Mudita Mishra, a media educator in the field of PR, asserts that the business of PR is to craft an image for an organization or a person, and then help sustain it, by way of managing reputation and then crisis, if one might arise. In an era dominated by technological evolution, the field of Corporate Communication (CorpComm) in PR has not been left untouched by deliberations on AI. If one is able to look at CorpComm as a model wherein communication is directed at internal and external publics, who are the stakeholders of the organization, then there could not have been a better time to appreciate the possibility of integration of AI and CorpComm. This is understood better once we acknowledge that PR in general has always had to deal with a certain public distrust, in that the public has found it difficult since the early days to believe that corporate communicators would be telling the truth. On the contrary, corporate communicators have been looked at as quite the defenders of wrong-doings of organizations or any other entities that they might be defending, by way of presenting such information in a manipulated way.

The possible role of AI and its applications can be a revolution that might help in contesting this widely held disbelief in two broad ways. Firstly, by helping the public in finding comfort and belief in the authenticity of communication being meted out by corporate houses, since the communications will no longer be the sole proprietary of humans, so as to say, but be validated by ‘machines’. Secondly, and more importantly, the reason why the former reasoning will be able to stand is that the primary data for crafting such communications itself will be crafted with the help of AI. This technology will be able to sense and pick up the various data points being generated by countless conversation points among consumers, audience, and citizens—the external public, in general. All in all, CorpComm integrated with AI makes for a very robust case for the image of PR in itself—a case that the PR industry has been fighting solely with the aid of its human representatives until now. AI will help bring in that unbiased perspective to this mix—at least, hopefully, a perception of unbiased communication coming from corporations.

Professor Pradeep Nair, a media educator and researcher, opines that the AI technologies have revolutionized the methods of teaching PR as a subject and as a practice. It brought a paradigm shift in PR education by making the teaching pedagogy more approachable. It makes the learning process more collaborative by engaging both the teachers and the students in real-life corporate situations. Today, AI is used in teaching PR for designing a teaching module and for engaging the students in assignments, assessment, and evaluation of students’ projects. It is used to assess the subjective understanding of the students by designing instructional contents as per the immediate needs of the students. It provides multiple digital platforms to interact and instruct the students about emerging PR practices, thus making PR as an academic discipline more structured and streamlined. By producing smart audio-visual contents, a teacher has an opportunity to help the students to understand the PR industry and can help them to improve their insights on the need of consumers and creating fine-tuned PR messages for them. The use of AI in teaching PR helps the media educators to adopt a utilitarian approach by analyzing the most prevalent

trend among the students and to address it accordingly. It also helps the media educators to teach the students about how PR companies are improving their services with the help of high-speed data to understand the digital DNA so that tailored and customized PR messages could be designed as per the requirements of the market.

## 5 Concluding Remarks

The PR persons or PR firms have started believing that AI to be a massive game changer which will enhance the work culture. They should not be scared of AI implications. It does not mean that they will be the experts in the implications of technologies rather they should develop the understanding of such technologies to provide informed counsel to their clients. John Bara, President and CMO of a leading company, states that savvy PR professionals would understand that big data and AI can provide their readership with amazing, data-rich research on a myriad of topics. Companies should not fear big data and AI. They should, instead, embrace the trend and experiment with new stories that match big data analysis and messages to their audience [20]. It is predicted that AI will soon optimize PR works than earlier. It is going to become a driving force in the PR industry. The PR and marketing persons are striving hard for better outcomes. The researchers are on for the most in-depth and have comprehensive look to resolve the issues. Stephen Waddington, formerly associated with Ketchum, underlines research, shows that in five years' time, AI is likely to have a stronger grip on PR functions.

Researches need to be conducted to understand the space and execution of AI in public and private organizations. There is no doubt that PR research is an integral part of an academic discipline. However, there is a dearth of research on the use of AI in this domain. The bottom line is that a PR person can be curious about AI and PR but should not afford to make the mistake of overlooking it.

**Acknowledgements** The researcher is sincerely thanking Sneha Verghese, Archana Kumari, and Prerona Sengupta for their insightful comments on the topic.

## References

1. Bourne C (2019) AI cheerleaders: public relations, neoliberalism and artificial intelligence. *Publ Relat Inquiry* 8(2):109–125
2. Chakraborty U, Bhat S (2017) Credibility of online reviews and its impact on brand image. *Manag Res Rev* 41(1):148–164
3. Dance F (1970) The “concept” of communication. *J Commun* 20:201–210
4. Dods P (2018, August 27). The impact of artificial intelligence on the PR industry. Retrieved from August 10, 2019. <https://learn.g2.com/pr-industry-artificial-intelligence>
5. Gehl RW, Bakardjieva M (2017) Socialbots and their friends: digital media and the automation of sociality. Routledge, NY

6. Ghosh M (2017) Significance of big data in E-commerce: the case of Amazon India. *Media Watch* 8(2):61–66
7. James SB (2018, May 23). Humans still needed: AI use in PR to treble in three years, report suggests. Retrieved from 16 Aug 2019. <https://www.prweek.com/article/1465483/humans-needed-ai-use-pr-treble-three-years-report-suggests>
8. Kaul A (2015) Effective business communication. Prentice Hall India, New Delhi
9. Naidoo J, Dulek RE (2018) Artificial intelligence in business communication: a snapshot. *Int J Bus Commun* 1–22
10. Nair P (2009) An IT technical framework for e-government: based on case study in Indian context. *Electron Gov Int J* 6(4):391–405
11. Nair P (2014) ICT based health governance practices: the Indian experience. *J Health Manag* 16(1):25–40
12. Nair P, Bhaskaran H (2015) The emerging interface of healthcare system and mobile communication technologies. *Health Technol* 4(4):337–343
13. Pandey US (2016) The Internet in India: crystallizing the historical inequalities. In: Narayan SS, Narayanan S (eds) *India connected: mapping the impact of New Media*. Sage, New Delhi, pp 221–236
14. Peterson A (2019, January 16). The past, present & future of artificial intelligence in PR. Retrieved from 16 Aug 2019. <https://www.cision.com/us/2019/01/artificial-intelligence-PR/>
15. Rego R (2017) New Media technologies in teaching and learning in higher education. *Media Watch* 8(1):75–88
16. Salathé M, Vu DQ, Khandelwal S, Hunter DR (2013) The dynamics of health behavior sentiments on a large online social network. *EPJ Data Sci* 2(1):4. <https://doi.org/10.1140/epjds16>
17. Seitel FP (2007) *The practice of public relations*. Pearson Prentice Hall, Upper Saddle River
18. Trivedi J (2019) Examining the customer experience of using banking Chatbots and its impact on brand love: the moderating role of perceived risk. *J Internet Commer* 18(1):91–111
19. Upadhyay AK, Khandelwal K (2018) Applying artificial intelligence: implications for recruitment. *Strateg HR Rev* 17(5):255–258. <https://doi.org/10.1108/SRH-07-2018-0051>
20. Whitaker A (2017, March 20) How advancements in artificial intelligence will impact public relations. Retrieved from 16 Aug 2019. <https://www.forbes.com/sites/thexec/2017/03/20/how-advancements-in-artificial-intelligence-will-impact-public-relations/#1b84ba8941de>

# Chapter 12

## Roulette Wheel Selection-Based Computational Intelligence Technique to Design an Efficient Transmission Policy for Energy Harvesting Sensors



**Shaik Mohammad, E. S. Gopi and Vineetha Yogesh**

### 1 Introduction

Internet of Things (IoT) and machine learning are getting much attention in recent years. Besides the connectivity of computers and mobile phones, Internet of Things empowers the connectivity among billions of ‘things’ and devices through Internet or local area networks (LAN). Multifarious applications of IoT include but not limited to household needs, industrial applications, wireless sensors, etc. Most of these applications need gathering and transmission of sensed data round the clock. Enabling these billions of devices requires a continuous supply of energy for their uninterrupted functioning. Conventional power supply may not be feasible for all applications, especially those of wireless domain, and the usage of battery requires timely monitoring and replacement. As a result of unprecedented growth in IoT-enabled devices, maintenance of these power resources becomes a hefty exercise and led to the evolution of energy harvesting (EH) sensors as a viable option [1]. EH sensors harvest the energy from natural resources in small amounts, store it in a rechargeable battery and use it instantaneously for all the needs [2, 3, 18]. EH sensors contribute to green communication and can operate independently over long periods of time. These EH devices are finding a considerable applications in wireless sensor networks (WSN) because of the benefits mentioned [17–26]. EH sensors are relatively low-cost devices and operate with minimal amount of energy. So their prevalent presence can be seen in many applications like monitoring and controlling the environment, especially in remote and dangerous areas [22]. In EH applications, harvesting process and phenomenon need to be properly analysed and adapted; trans-

---

S. Mohammad · V. Yogesh  
Communication Systems, NIT, Trichy, India

E. S. Gopi (✉)  
Department of Electronics and Communication Engineering, NIT, Tiruchirappalli, India  
e-mail: [esgopi@nitt.edu](mailto:esgopi@nitt.edu)

mission energy management with deterministic harvesting process has been studied in [14]. However, energy harvesting is being done from the environment and is least likely to be deterministic. Due to the sporadic nature of resource availability, chance of energy harvesting in a given time interval can be treated as a stochastic process with some harvesting probabilities [30, 31]. Quantity of harvested energy also depends on various factors and varies time to time, and this should also be considered while setting up the simulation environment [13, 14, 20]. Different power management schemes have been studied in [14, 29]. In a communication model, either transmitting node or receiving node or both of them can be capable of energy harvesting. Considerable research work had been carried out with transmitter nodes alone being energy harvesting capable [31, 32], receiving nodes alone being harvesting capable [27, 28] and both transmitting and receiving nodes harvesting capable [23–26]. In this paper, we consider that the transmitting nodes are capable of harvesting energy from the environment. To evaluate the performance of any transmission policy, a communication model with performance metric needs to be considered [6, 9, 15, 16, 31]. Performance metric can aim at optimization of any single key parameter or overall energy utilization such that it defines the efficiency of communication model with all constraints.

## 1.1 *Background*

Because of their replenishing abilities and prolonged lifetime, energy harvesting sensors found a place in communication models [4–11]. Significant research work has been carried out on different factors aiming at achieving better performance of the system. An efficient multi-stage energy transfer system, which has the relation among various components of the system and their optimal selection according to the needs, is presented in [13]. Adapting those guidelines in hardware components considerably increases the capacity of an EH sensor. Multi-parametric programming approaches with adjustments to different crucial parameters such as buffer size, sampling rate, timing and routing are studied in [21, 29]. In [4], a directional water filling algorithm to minimize the transmission completion time of the communication session while maximizing the throughput has been introduced. An online dynamic programming framework to control the admissions into data buffer is derived in [5]. Energy management policies stabilizing the data queues and optimizing the delay properties in a single-user communication model under a liner approximation is studied in [6]. Throughput optimal energy allocation with a time-constrained slotted setting in energy harvesting system is studied in [8]. Some other performance metrics of an EH sensor that have been studied in the literature include the minimization of transmission time [9], improving the quality of coverage [15], maximization of short-term throughput [16], optimizing throughput and minimizing delay [6]. Apart from these, main aim of any communication model also includes the faithful transmission of collected data. Packet drop probability or packet outage probability gives a measure of successful transmission. In [31, 32], packet drop probability has been considered

as performance metric. In this paper, we did consider packet drop probability as a performance metric to compare the performance of transmission policies underlined with different computational intelligence techniques. These computational intelligence techniques play a vital role in determining the state of communication system at a given instant by estimating the possible channel gain. An accurate channel gain estimation helps in finding adequate transmission power to successfully communicate with the receiver, which in turn results in an efficient system with low packet drop ratio and higher life span.

Contribution of this paper includes the simulation of wind energy harvesting as per the real-time governing scenarios to harvest the EH node. We are proposing a new intuitive but effective algorithm to estimate the channel gain, in turn the next possible state. Performance of the transmission policy employing this technique has been compared with other popular computational techniques, artificial neural network and extreme learning machine. Performance of these techniques has been evaluated under varying conditions of key parameters. We are also proposing a novel collaborative transmission policy to improve the performance of wireless sensor network (WSN) nodes. A comparative study on performance between the WSNs employing and not employing the collaborative transmission policy is also presented.

Rest of the paper is organized as follows. System model is presented in Sect. 2. Transmission energy assignment strategy is explained in Sect. 3. Collaborative transmission policy is detailed in Sect. 4. Simulated numerical results under multiple environmental conditions are presented in Sect. 5, followed by conclusions in Sect. 6.

## 2 System Model

In this paper, we consider a communication model with EH sensors employing the proposed transmission policies. Energy harvesting from natural resources at EH sensors is considered as a stochastic process. Amount of harvested energy is randomly assigned from one of the possible set of values, which are derived based on a practical survey to comply with randomness of natural resources. Upon harvesting from natural resources, EH sensors store the energy in a rechargeable battery and use it for transmitting the data. The communication model considered in this paper is similar to that of [31, 32]. Data has been divided into equal length packets, and being a resilient technique to channel variations, automatic repeat request (ARQ) transmission is considered [12, 31]. In ARQ model, the transmitting EH node receives either a positive acknowledgement (ACK) or a negative acknowledgement (NACK) for each packet of data it transmits, which helps in acquiring channel state information (CSI). Concept of re-transmission has been adapted, where each packet of data is attempted for a maximum of K transmissions. Time required for transmission of data and receiving the corresponding acknowledgement is defined as slot time ( $T_{slot}$ ). K times the slot time is defined as one frame time ( $T_{frame}$ ). Every packet of data is allotted with one frame time for its transmission. System at any instant can be described by its state

**S**, which constitutes of estimated transmission power ( $E$ ), channel gain ( $\gamma$ ), battery level ( $B$ ), feedback signal from receiver ( $R_{m,n}$ ) and the packet transmission attempt number ( $k$ ).

## 2.1 Harvesting Model

As mentioned, EH sensors harvest the energy from natural resources such as solar, vibration, mechanical and wind [13, 14, 18, 33]. Due to the sporadic nature of resource availability, it cannot be deterministic, it can only be treated as a stochastic process with probability of energy being harvested  $P_{harv}$  at the beginning of every slot time. In this paper, we considered wind energy as the source of harvesting. Intensity of the wind usually varies from period to period adhering to the environmental conditions [35, 36]. Wind power density changes from place to place. Wind power of any selected site is proportional to the cube of wind speed; therefore, wind power density (WPD) can be written as [35, 37, 38]:

$$\text{WPD} = \frac{1}{2n} \rho \sum_{i=1}^n v^3 = \frac{1}{2} \rho \bar{v}^3 \quad \left( \frac{\text{W}}{\text{m}^2} \right) \quad (1)$$

where WPD is the wind power density in  $\text{W/m}^2$ ,  $\rho$  is the air density in  $\text{kg/m}^3$ ,  $\bar{v}$  is the mean wind speed in  $\text{m/s}$  and  $n$  is the number of observations in the specific time period. This wind power density is estimated closely using the Weibull distribution function in [35] as

$$p(v) = \frac{1}{2} \rho A \bar{v}^3 \quad (2)$$

$$\text{WPD} = \frac{p(v)}{A} = \int_0^\infty \frac{1}{2} \rho A \bar{v}^3 f(v) dv = \frac{1}{2} \rho c^3 \Gamma \left( 1 + \frac{3}{g} \right) \quad (3)$$

where  $p(v)$  denotes the power available in watts,  $A$  denotes the root swept area in  $\text{m}^2$ ,  $\Gamma$  denotes the mathematical gamma function and  $f(v)$  denotes the two parameter Weibull function with  $c$  as Weibull scale parameter in  $\text{m/s}$  and  $g$  as Weibull shape parameter.

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx \quad (4)$$

$$f(v) = \frac{g}{c} \left( \frac{v}{c} \right)^{g-1} \exp \left( - \left( \frac{v}{c} \right)^g \right) \quad (5)$$

For simulating the harvesting environment, Weibull scale ( $c$ ) and shape ( $g$ ) parameters of Taralkatti area, Karnataka, have been adopted [35]. To confirm the randomness in wind and to evaluate the rigidity of designed policies,  $g$  and  $c$  values for four months are considered. Therefore, the quantity of harvested energy can be any of the values formed by all possible combinations of  $g$  and  $c$ . The performance of transmission policies has been evaluated and presented for a range of harvesting probabilities ( $P_{harv}$ ).

## 2.2 Battery Model

EH sensor consists of a rechargeable battery of limited capacity ( $B_{max}$ ), and harvested energy will be stored in it and is used for all the needs of EH sensor. At the beginning of every time slot, EH nodes check whether the data needs to be transmitted, and if it needs to, it checks if the required amount of energy is available in battery. If sufficient energy is available in battery, EH node transmits the data, else EH node waits till it acquires sufficient energy level through harvesting. Transmission attempts will be made at the beginning of every slot if the packet has not received a positive acknowledgement (ACK) and is within its frame time. A maximum of  $K$  transmission attempts will be made owing to the availability of sufficient energy level in battery.

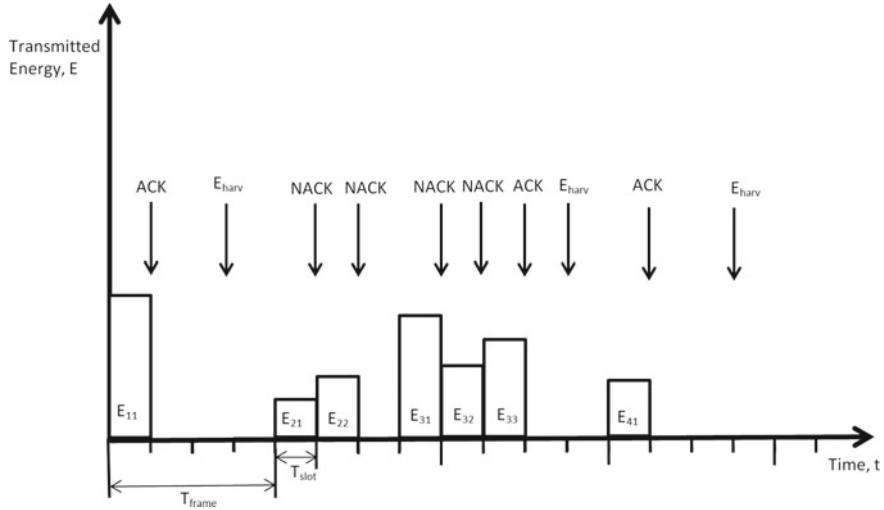
As mentioned, energy harvesting happens with a probability of  $P_{harv}$  at the beginning of every time slot. Each time harvesting happens, and the battery level increases by an amount,  $E_{harv}$  corresponding to the values of  $k$  and  $c$  in that time slot (3). Therefore, the battery level at the beginning of any time slot  $n + 1$  of time frame  $m$  will be

$$B_{m,n+1} = \begin{cases} \min(B_{m,n} + E_{harv} - E_{m,n}, B_{max}) & \text{with probability } P_{harv} \\ B_{m,n} - E_{m,n} & \text{with probability } 1 - P_{harv} \end{cases} \quad (6)$$

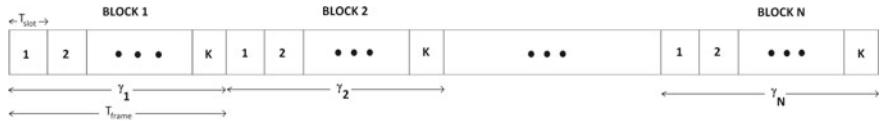
where  $B_{m,n}$  represents the battery level and  $E_{m,n}$  represents the energy spent in transmitting a data packet at slot  $n$  of time frame  $m$ . Packet re-transmission model with re-transmission index (K) of four is shown in Fig. 1.

## 2.3 Channel Model

To imitate the wireless channel and its fading effects, a Rayleigh fading channel with additive white Gaussian noise has been considered [32, 34]. A discrete channel model is used for covering the fading gains of the wireless channel. All the possible fading gains or channel gains are considered and covered by a discrete channel gain set,  $\mathcal{G} = \{\gamma_1, \gamma_2, \dots, \gamma_N\}$ . Besides reduction in memory, discretization of channel gains also results in all the benefits of quantization [31]. These states can be computed based on



**Fig. 1** Packet re-transmission model with energy harvesting for re-transmission index,  $K = 4$



**Fig. 2** Block fading channel model

the underlying Doppler frequency and fading distribution, following the procedures mentioned in [39, 40]. The channel is considered to be block fading channel [31], which implies that the coherence channel time is larger than the frame duration and it changes for every frame time, i.e.  $T_{frame} \leq T_{coherence}$  [34]. So, the channel characteristics remain constant for one frame time, as shown in Fig. 2. Rayleigh distribution of channel gain ( $\gamma$ ) with scale parameter ‘ $\alpha$ ’ is defined as:

$$R(\alpha) = \frac{\gamma}{\alpha^2} \exp\left(\frac{-\gamma^2}{2\alpha^2}\right) \quad (7)$$

$$R_{mean}(\alpha) = \alpha \times \frac{\pi}{2} \quad (8)$$

$$R_{sd}(\alpha) = \alpha^2 \left( \frac{4 - \pi}{2} \right) \quad (9)$$

Performance variations of proposed transmission policies for a range of variations in channel gain mean are computed and presented in numerical results section.

### 3 Transmission Energy Assignment

State of the communication system at a time instant,  $T_{m,n}$  (Frame  $m$ , slot  $n$ ), is denoted as  $\mathbf{S}_{m,n} = (E_{m,n}, \gamma_m, B_{m,n}, R_{m,n}, k)$ , where  $R_{m,n}$  is the corresponding acknowledgement of the transmitted packet. It is ‘1’ if an ACK is received and is ‘0’ if an NACK is received.  $E_{m,n}$  is the energy spent by EH node in transmitting data packet at  $T_{m,n}$  slot. It can be a positive quantity if the packet has not received a positive acknowledgement ( $R_{m,n-1} = 0$ ) and its value is less than the battery reserve,  $B_{m,n}$ . If the packet has already received an ACK or if the battery level is less than the demand value,  $E_{m,n}$  will be zero as no transmission takes place. In a BPSK modulation technique,  $E_{m,n}$  and  $\gamma_m$  are related to each other by probability of error ( $P_e$ ) [31] as

$$P_e(\gamma_m, E_{m,n}) = 1 - \left(1 - Q\left(\sqrt{\frac{2\gamma_m E_{m,n}}{N_0}}\right)\right)^l \quad (10)$$

where  $N_0$  represents the power spectral density of additive white Gaussian noise (awgn) and  $l$  represents the packet length, i.e. number of bits per packet.

Therefore, once we find out or fix the desired accuracy of the communication system in terms of its probability of error, we can derive a relation between channel gain and transmission energy as

$$\gamma_m \times E_{m,n} = \frac{\left(Q^{-1}\left(1 - (1 - P_e)^{\frac{1}{l}}\right)\right)^2 \times N_0}{2}. \quad (11)$$

From (11), it is evident that if we estimate the channel gain value,  $\gamma_m$ , we can estimate the amount of transmission energy required for packets’ first transmission attempt ( $k = 1$ ). Therefore, in this paper, we are estimating the channel gain to determine the next possible state  $S_{m+1,1}$ , in turn  $E_{m+1,1}$  to transmit the data packet with an optimal transmission energy. State of the EH node at  $K$ th slot of  $m$ th frame will be  $\mathbf{S}_{m,K} = (E_{m,K}, \gamma_m, B_{m,K}, R_{m,K}, K)$ . Transition of state from  $T_{m,K}$  to  $T_{m+1,1}$  will be

$$S_{m+1,1} = \begin{cases} (E_{m+1,1}, \gamma_{m+1}, B_{m+1,1}, 0, 1) & \text{if } B_{m,K} > E_{m+1,1} \\ (0, \gamma_{m+1}, B_{m+1,1}, 0, 1) & \text{if } B_{m,K} < E_{m+1,1} \end{cases} \quad (12)$$

$\gamma_{m+1}$  will be estimated by the computational intelligence techniques mentioned in later sections. Its not likely to get the perfect estimation every time, which leads to an NACK. If the EH sensor receives an NACK, it has to reassess the required energy and attempt to transmit with energy,  $E_{m,n+1}$ . On receiving an NACK, EH node tries to retransmit with an increased transmission energy. Instead of incrementing the energy by a fixed amount, which may require adjustments with changes in channel conditions and estimation techniques, a common principle is used throughout the policies discussed in this paper. During the channel gain estimation using any technique, a

buffer  $D$  of size  $K$  will be maintained, and whenever the EH node receives an NACK in first attempt and an ACK in its subsequent attempts, the deviation between the two channel gain estimates is recorded and stored in this buffer. So, the buffer,  $D$ , holds the most recent  $K$  deviations of faulty estimations and these deviations are used in finding the energy for next transmission attempt. Incremental energy is taken as the root mean square value of deviations stored in  $D$ .

$$E_{m,n+1} = E_{m,n} + \delta E \quad (13)$$

$$\delta E = \sqrt{\sum_{i=1}^K \frac{D^2(i)}{K}} \quad (14)$$

The state transfer from slot  $T_{m,n}$  ( $S_{m,n}$ ) to slot  $T_{m,n+1}$  ( $S_{m,n+1}$ ) can be observed as :

$$S_{m,n+1} = \begin{cases} (0, \gamma_m, B_{m,n+1}, 0, K) & \text{if } k = K \\ (0, \gamma_m, B_{m,n+1}, 1, k+1) & \text{if } k < K \& R_{m,n} = 1 \\ (0, \gamma_m, B_{m,n+1}, 0, k+1) & \text{if } k < K \& R_{m,n} = 0 \& B_{m,n+1} < E_{m+1,1} \\ (E_{m,n+1}, \gamma_m, B_{m,n+1}, 0, k+1) & \text{if } k < K \& R_{m,n} = 0 \& B_{m,n+1} > E_{m+1,1} \end{cases} \quad (15)$$

Our main aim is to reduce the packet drop probability to as low as possible. Therefore, in addition to the above-mentioned state transitions, a special case is introduced when  $0.7 * B_{m,n+1} \leq E_{m+1,1} < B_{m,n+1}$  and  $k = K$ . In this scenario, a transmission attempt is made as it may result in  $R_{m,n} = 1$ .

### 3.1 Channel Gain Estimation

Channel gain estimation plays a vital role in energy estimation, in turn the state estimation. An accurate channel gain estimation reduces the number of re-transmissions and leads to an efficient transmission policy design. In this paper, we considered the well-known artificial neural network (ANN) and extreme learning machine (ELM) techniques in addition to the proposed maximum matched distribution (MMD) technique. We compare the packet outage probability of all these techniques. Initial sample value for channel gain estimate is taken as twice the mean channel gain value. From the next sample onwards, policies use respective estimation techniques. All these computational intelligence techniques require a lead on channel gain history to estimate the next possible sample value. As mentioned in Sect. 2.3, Rayleigh dis-

tribution is used for simulating the wireless channel. While evaluating performance for a Rayleigh fading channel of mean gain ' $\gamma_{mid}$ ', these models are trained with a sequence of mean gain ' $\gamma_{mid}$ ', length equal to one-third of the number of total slots. However, the channel gains to be estimated are simulated as a Rayleigh sequence with mean value varying arbitrarily between ' $\gamma_{mid} - 2$ ' and ' $\gamma_{mid} + 2$ ' to measure the robustness to variations from the trained values. A two-hidden-layer ANN with resilient back propagation mechanism is considered [44, 45]. Additive hidden nodes with log sigmoid function are considered for ELM and estimate the next channel gain sample [41–43]. Most recent history of ten channel gain samples is considered while estimating the next channel gain sample in ANN and ELM. In MMD model, a transition probability distribution matrix,  $\mathbf{T}$  is constructed using the history of channel gain samples. In  $\mathbf{T}$ , each row and each column correspond to the discretized channel gains. The elements represent the transition probability from one state to the other. Therefore, selected row of  $\mathbf{T}$  represents the transition probabilities from that channel gain to all other channel gain values. Algorithm for constructing  $\mathbf{T}$  is as follows:

---

**Algorithm 1** Formation of MMD matrix,  $\mathbf{T}$ 


---

```

1: procedure Form_MMD_Matrix_T
2:   H  $\leftarrow$  history of channel gain samples
3:   L  $\leftarrow$  length of H
4:    $\delta \leftarrow$  selected step size value
5:    $\gamma \leftarrow$  Possible digitized channel gain states
6:   i  $\leftarrow$  1
7:   j  $\leftarrow$  1
8: Digitize the channel gain samples
9:
10:  while i  $\leqslant$  L do ▷ Total history of samples
11:    Digi_H  $\leftarrow$  H(i) Digitized to closest element of  $\gamma$ 
12:    i  $\leftarrow$  i+1
13: Form the matrix T
14:
15:  while j  $\leqslant$  L do ▷ Total history of samples
16:    if j  $\geqslant$  2 then
17:      m  $\leftarrow$  Digi_H(j - 1)
18:      n  $\leftarrow$  Digi_H(j)
19:      T(m, n) = T(m, n) + 1
20:    j  $\leftarrow$  j+1

```

---

Once  $\mathbf{T}$  is formed, normalize each individual row so that it represents the transition probabilities. To estimate the next channel gain sample, the latest channel gain estimate will be considered. This sample is then digitized, and its corresponding row from  $\mathbf{T}$ , which represents the transition probabilities, will be taken. A roulette wheel is formed with the ten most probable states of transition probabilities, and one of them will be selected as next channel estimate based on roulette wheel selection mechanism.

## 4 Collaborative Transmission Policy

From the observations of Table 1, it can be understood that with the increase in harvested energy which can be through increase in harvesting probability or the re-transmission index, performance of all the policies has been improved. It can also be observed that the improvements are a bit more significant in the case of transmission policy which uses MMD-RW mechanism to estimate the channel gain. Another important observation is when the channel gain is less, the performance of MMD-RW is not as prominent as compared to that in higher channel gain scenarios, whereas the performance of policy employing ANN has been improved moderately but the change in outage probability is not as remarkable as that employing MMD-RW technique. This clearly indicates that the transmission policy employing MMD-RW method is more effective but is consuming a little extra energy for transmitting a packet. Therefore, if that little extra requirement is fulfilled, its performance could be even better.

Understandably, the extra energy cannot be assured from natural resources as it is not in the control of EH node. Increasing re-transmission index helps but also increases the delay in transmitting subsequent packets from source to destination. Therefore, we are proposing the concept of collaborative transmission policy. According to this policy, when an EH node runs out of energy, i.e. required estimated energy for transmitting a packet is more than the battery reserve, it seeks the help of other nearby EH nodes in the network. EH node, which has sufficient energy for transmitting its own packet as well as the requesting nodes packet, will then transmit both the packets. In this manner, more efficient data transmission can be achieved by means of collaboration. Selection of EH node which aids the requesting node depends on reward factor. The node with higher reward factor will be given priority.

To illustrate the collaborative transmission policy, a wireless sensor network with EH sensor nodes is considered. At each EH node, a metric space of its surrounding nodes is maintained. In this work, three-dimensional euclidean space is considered. So, every EH node has a prior knowledge of its adjacent nodes and respective euclidean distances. Euclidean distance between two points  $(x_1, y_1, z_1)$  and  $(x_2, y_2, z_2)$  is measured as

$$d_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (16)$$

Battery insufficiency occurs when the required transmission energy of the packet is higher than the battery reserve.

$$B_{m,n} < E_{m,n} \quad (17)$$

In this case, it looks for the status of its neighbour nodes by seeking their status vector  $S$  and evaluates the reward factor associated with each of them. Rewarding factor  $\Re$  varies inversely with distance as path loss varies directly with the distance.

**Table 1** Packet drop probability or outage probability for re-transmission index  $K = 4$  and  $K = 6$  with different energy harvesting probabilities,  $P_{harv}$  and mean channel gains,  $\gamma_{mid}$ 

Attempt	MMD-RW		ELM		ANN	
	$K = 4$	$K = 6$	$K = 4$	$K = 6$	$K = 4$	$K = 6$
<i>Harvesting probability, <math>P_{harv} = 0.3</math> and <math>\gamma_{mid} = 6</math></i>						
1	0.4115	0.2295	0.3915	0.2642	0.3895	0.2695
2	0.4218	0.2320	0.3870	0.2697	0.3835	0.2738
3	0.4128	0.2298	0.3860	0.2585	0.3832	0.2538
4	0.4190	0.2385	0.3972	0.2662	0.3935	0.2600
5	0.4205	0.2375	0.3907	0.2750	0.3872	0.2730
Avg	0.4171	0.2335	0.3905	0.2667	0.3874	0.2660
<i>Harvesting probability, <math>P_{harv} = 0.3</math> and <math>\gamma_{mid} = 8</math></i>						
1	0.2807	0.1235	0.2635	0.1930	0.2767	0.1817
2	0.2840	0.1242	0.2550	0.1943	0.2715	0.1883
3	0.2860	0.1225	0.2612	0.1963	0.2737	0.1923
4	0.2830	0.1270	0.2520	0.1875	0.2697	0.1870
5	0.2908	0.1258	0.2590	0.1983	0.2750	0.1975
Avg	0.2849	0.1246	0.2581	0.1939	0.2733	0.1894
<i>Harvesting probability, <math>P_{harv} = 0.5</math> and <math>\gamma_{mid} = 6</math></i>						
1	0.1832	0.0742	0.2050	0.1805	0.2013	0.1030
2	0.1742	0.0775	0.1940	0.2000	0.1960	0.1060
3	0.1710	0.0762	0.2060	0.1935	0.2000	0.0980
4	0.1772	0.0788	0.2010	0.1890	0.1990	0.1055
5	0.1810	0.0757	0.2023	0.1835	0.2025	0.1030
Avg	0.1773	0.0765	0.2017	0.1893	0.1998	0.1031
<i>Harvesting probability, <math>P_{harv} = 0.5</math> and <math>\gamma_{mid} = 8</math></i>						
1	0.0843	0.0288	0.1708	0.1667	0.1643	0.1625
2	0.0887	0.0318	0.1752	0.1590	0.1730	0.1485
3	0.0840	0.0300	0.1812	0.1570	0.1742	0.1512
4	0.0910	0.0275	0.1782	0.1745	0.1718	0.1690
5	0.0890	0.0280	0.1812	0.1600	0.1752	0.1557
Avg	0.0874	0.0292	0.1773	0.1634	0.1717	0.1574
<i>Harvesting probability, <math>P_{harv} = 0.7</math> and <math>\gamma_{mid} = 6</math></i>						
1	0.0890	0.0340	0.1660	0.1457	0.1598	0.1427
2	0.0870	0.0333	0.1737	0.1593	0.1663	0.1400
3	0.0830	0.0338	0.1817	0.1552	0.1730	0.1740
4	0.0862	0.0348	0.1802	0.1532	0.1745	0.1462
5	0.0860	0.0320	0.1750	0.1562	0.1668	0.1353
Avg	0.0862	0.0336	0.1753	0.1539	0.1681	0.1476
<i>Harvesting probability, <math>P_{harv} = 0.7</math> and <math>\gamma_{mid} = 8</math></i>						
1	0.0293	0.0090	0.1633	0.1368	0.1608	0.1330
2	0.0285	0.0085	0.1603	0.1312	0.1585	0.1258
3	0.0225	0.0088	0.1600	0.1292	0.1555	0.1217
4	0.0290	0.0083	0.1530	0.1418	0.1507	0.1380
5	0.0293	0.0090	0.1683	0.1332	0.1650	0.1288
Avg	0.0277	0.0087	0.1610	0.1344	0.1581	0.1295

Let the EH node  $i$  is in battery shortage and is trying to evaluate the reward factor associated with EH node  $l$ . Then,

$$\mathfrak{R}_{m,n}^{i,l} = \left( \frac{B_{m,n}^l}{d_{i,l}} + B_{m,n}^l - E_{m,n}^i \right) \delta_{R_{m,n-1}^l, 1} \quad (18)$$

where  $\mathfrak{R}_{m,n}^{i,l}$  is the rewarding factor of EH node  $l$  w.r.t EH node  $i$ ,  $E_{m,n}^i$  denotes the estimated energy required for transmitting the packet of EH node  $i$ .  $B_{m,n}^l$  denotes the battery reserves of EH node  $l$  during slot  $n$  of time frame  $m$ .  $R_{m,n-1}^l$  is the acknowledgement of EH node  $l$  at the instant  $T_{m,n-1}$  and  $\delta_{i,j}$  is Kronecker delta function.

After evaluating the rewarding factor of all the EH nodes in its vicinity, EH node  $i$  requests the node with highest reward factor to transmit the packet.

$$L = \max (\mathfrak{R}_{m,n}^{i,l}) \quad \forall l \in \mathbb{V} \quad (19)$$

where  $L$  denotes the selected node to request the packet transmission and  $\mathbb{V}$  denotes the neighbourhood of EH node  $i$ . Once the EH node  $L$  transmits the packet of node  $i$ , its battery reserve will be updated to

$$B_{m,n}^L = B_{m,n}^L - E_{m,n}^i \quad (20)$$

Battery reserve of EH node  $i$  will be updated to

$$B_{m,n}^i = B_{m,n}^i - d_{i,L} \times E_u \quad (21)$$

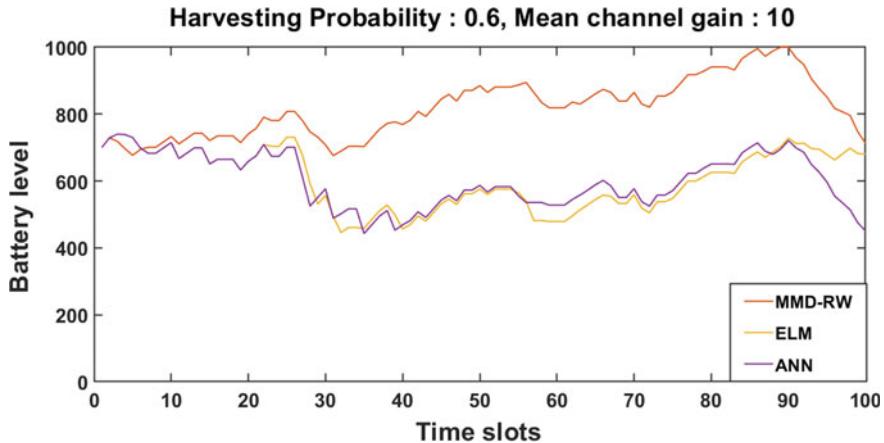
where  $E_u$  is the energy spent in transmitting a packet per unit distance among the nodes. If none of the neighbouring nodes have a positive  $\mathfrak{R}_{m,n}^{i,l}$  or if the EH node  $i$  does not have battery reserve even to transmit the packet to EH node  $L$  or if the packet receives an NACK ( $R_{m,n} = -1$ ) after transmission, it results in packet drop or outage once the packet runs out of maximum transmission attempts ( $K$ ).

## 5 Numerical Results

Experiments are carried out to evaluate and compare the performances of all the proposed transmission policies under various effecting factors such as harvesting probability ( $P_{harv}$ ), mean channel gain ( $\gamma_{mid}$ ) and re-transmission index ( $K$ ). The results are summarized and presented in Table 1. Further, the impact of collaborative transmission policy in achieving even better results is studied, and results are tabulated in Table 2. MMD-RW represents the transmission policy which employed maximum matched distribution model-based technique with roulette wheel selection to estimate the channel gain. ELM represents the transmission policy with extreme

**Table 2** Packet drop probability or outage probability for collaborative re-transmission policy with re-transmission index  $K = 4$ , different energy harvesting probabilities,  $P_{harv}$  and mean channel gains,  $\gamma_{mid}$

Attempt	Without collaboration				With collaborative policy			
	RW	ANN	RW	ANN	RW(RW)	ANN(RW)	RW(ANN)	ANN(ANN)
<i>Harvesting probability, <math>P_{harv} = 0.3</math> and <math>\gamma_{mid} = 6</math></i>								
1	0.4158	0.3842	0.4105	0.3890	0.2542	0.2705	0.2452	0.2690
2	0.3960	0.3693	0.3937	0.3665	0.2375	0.2630	0.2243	0.2610
3	0.4030	0.3563	0.4083	0.3688	0.2452	0.2675	0.2430	0.2660
4	0.4020	0.3625	0.4065	0.3703	0.2437	0.2650	0.2362	0.2590
5	0.3995	0.3655	0.4073	0.3795	0.2540	0.2717	0.2370	0.2715
Avg	0.4033	0.3676	0.4053	0.3748	0.2469	0.2675	0.2371	0.2653
<i>Harvesting probability, <math>P_{harv} = 0.3</math> and <math>\gamma_{mid} = 8</math></i>								
1	0.2880	0.2732	0.2855	0.2925	0.1600	0.2268	0.1557	0.2225
2	0.2940	0.2692	0.2960	0.2750	0.1570	0.2253	0.1520	0.2200
3	0.2815	0.2652	0.2858	0.2637	0.1565	0.2308	0.1525	0.2238
4	0.3040	0.2913	0.3090	0.3010	0.1745	0.2490	0.1658	0.2432
5	0.2968	0.2695	0.3035	0.2863	0.1593	0.2288	0.1510	0.2253
Avg	0.2929	0.2737	0.2960	0.2837	0.1615	0.2321	0.1554	0.2270
<i>Harvesting probability, <math>P_{harv} = 0.5</math> and <math>\gamma_{mid} = 6</math></i>								
1	0.1757	0.2200	0.1767	0.2182	0.1050	0.2065	0.0988	0.2050
2	0.1742	0.2188	0.1735	0.2172	0.1075	0.2100	0.1035	0.2062
3	0.1875	0.2235	0.1802	0.2255	0.1118	0.2075	0.1045	0.2115
4	0.1822	0.2238	0.1730	0.2190	0.1160	0.2092	0.1047	0.2047
5	0.1643	0.2150	0.1740	0.2097	0.1037	0.2040	0.1003	0.2020
Avg	0.1768	0.2202	0.1755	0.2179	0.1088	0.2074	0.1024	0.2059
<i>Harvesting probability, <math>P_{harv} = 0.5</math> and <math>\gamma_{mid} = 8</math></i>								
1	0.0770	0.1305	0.0783	0.1375	0.0503	0.1275	0.0520	0.1260
2	0.0785	0.1305	0.0808	0.1435	0.0450	0.1278	0.0510	0.1280
3	0.0777	0.1393	0.0725	0.1445	0.0465	0.1340	0.0473	0.1355
4	0.0732	0.1365	0.0760	0.1390	0.0437	0.1313	0.0418	0.1315
5	0.0650	0.1123	0.0675	0.1288	0.0370	0.1100	0.0413	0.1108
Avg	0.0743	0.1298	0.0750	0.1387	0.0445	0.1261	0.0467	0.1264
<i>Harvesting probability, <math>P_{harv} = 0.7</math> and <math>\gamma_{mid} = 6</math></i>								
1	0.0808	0.1747	0.0803	0.1760	0.0540	0.1745	0.0542	0.1747
2	0.0858	0.1742	0.0860	0.1787	0.0610	0.1732	0.0597	0.1725
3	0.0717	0.1792	0.0775	0.1735	0.0510	0.1777	0.0490	0.1777
4	0.0757	0.1742	0.0745	0.1797	0.0525	0.1742	0.0460	0.1742
5	0.0777	0.1787	0.0765	0.1727	0.0503	0.1790	0.0515	0.1767
Avg	0.0783	0.1762	0.0790	0.1761	0.0538	0.1757	0.0521	0.1752
<i>Harvesting probability, <math>P_{harv} = 0.7</math> and <math>\gamma_{mid} = 8</math></i>								
1	0.0285	0.1462	0.0222	0.1462	0.0205	0.1460	0.0138	0.1457
2	0.0280	0.1485	0.0238	0.1490	0.0187	0.1485	0.0165	0.1482
3	0.0272	0.1560	0.0270	0.1568	0.0182	0.1557	0.0175	0.1555
4	0.0195	0.1475	0.0265	0.1455	0.0140	0.1467	0.0182	0.1475
5	0.0290	0.1552	0.0270	0.1497	0.0182	0.1550	0.0203	0.1545
Avg	0.0264	0.1507	0.0253	0.1494	0.0179	0.1504	0.0173	0.1503



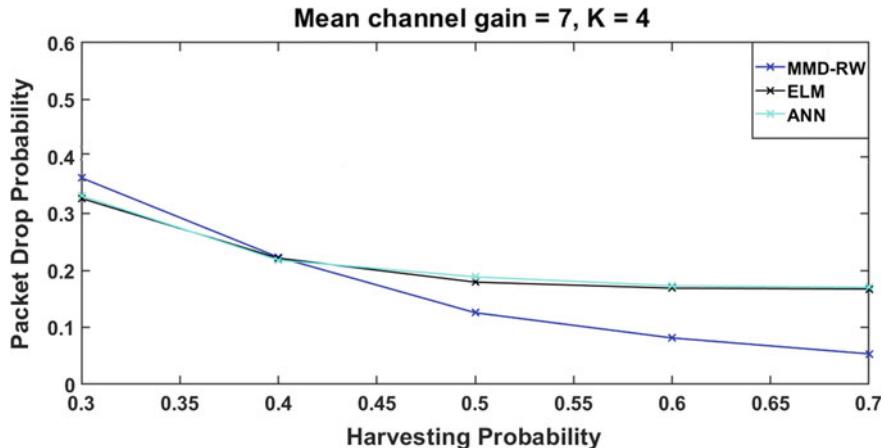
**Fig. 3** Battery utilization and energy harvesting for the first 100 time slots of transmission. Environmental conditions, harvesting probability,  $P_{harv} = 0.6$ , mean channel gain,  $\gamma_{mid} = 10$  and re-transmission index,  $K = 4$

learning machine technique to estimate the channel gain, and ANN represents the transmission policy that utilizes artificial neural network to estimate the channel gain.

Weibull scale and shape parameters corresponding to Taralkatti, Karnataka, region for four months have been adopted. So, the number of probable combinations which decide the amount of harvested energy is sixteen. This helps in predicting the suitability of proposed policies to tough environmental conditions. Frequency of energy harvesting is governed by the harvesting probability. Initial battery level at EH node is taken as 70% of the total capacity. Battery levels of all the transmission policies for the first hundred time slots are shown in Fig. 3. Supporting environmental conditions are harvesting probability,  $P_{harv} = 0.6$ , re-transmission index,  $K = 4$ , and mid-channel gain  $\gamma_{mid} = 10$ . The increase in battery level indicates energy harvesting, and the dips in battery level over the slots indicate the energy spent in transmitting the packets.

### 5.1 Effect of Harvesting Probability on Packet Drop Probability

Harvesting probability indicates the frequency of energy getting harvested from the natural resources. The higher the probability, the higher will be energy availability in battery to spend in transmitting the packets. This results in lesser non-transmissions due to the lack of sufficient energy (15), in turn reduction in packet outage probability. Probability of harvesting and amount of energy harvested each time totally depends on environmental conditions. Reduction in packet drop probability with increase in



**Fig. 4** Variation in packet drop probability against harvesting probability under a fixed environment of mean channel gain,  $\gamma_{mid} = 7$  and re-transmission index,  $K = 4$

harvesting probability for a particular environmental condition is shown in Fig. 4. The detailed variations in packet drop probability with changes in harvesting probability are presented in Table 1.

## 5.2 Effect of Channel Gain on Packet Drop Probability

From (11), it can be understood that the channel gain or fading gain directly effects the amount of energy required in transmitting a packet from EH node. If the channel gain is more, the energy required will be less, doesn't exhaust much battery. This results in lesser packet outages due to lack of energy, which directly reduces the packet drop probability. Simulations are carried out for a range of channel gain variations. As mentioned earlier, for a considered  $\gamma_{mid}$ , the computational intelligence techniques are trained with a history of channel gain samples with mean channel gain of  $\gamma_{mid}$ , whereas the actual channel gain samples to be estimated vary arbitrarily from  $\gamma_{mid} - 2$  to  $\gamma_{mid} + 2$ . From Table 1, a significant reduction in packet drop probability with higher  $\gamma_{mid}$  can be observed. For a good channel with higher gains, packet outages will be lesser and performance will be higher.

## 5.3 Effect of Re-transmission Index on Packet Drop Probability

Re-transmission index,  $K$ , indicates the maximum number of transmission attempts allowed for a packet, and it also decides the number of slots per one frame time.

The higher the  $K$ , the higher the number of slots, more chance for energy harvesting as well as transmission attempts. Higher harvesting increases the battery reserve and reduces the packet drop probability. It may happen that the EH node transmits the packet with energy closer to the required value in  $(K - 1)$ th attempt. If another attempt is given, it may result in an ACK. Therefore, the chance of reaching the actual required energy increases with increase in re-transmission index. Though higher  $K$  gives better performance, it significantly increases the delay. Effect of  $K$  on packet drop probability is quiet evident from the results presented in Table 1.

#### **5.4 Impact of Collaborative Transmission Policy**

To illustrate the benefit of collaborative transmission policy, two exactly similar clusters of four nearby EH nodes each in a wireless sensor network are considered: one with individual transmission policies, not using collaborative transmission policy, and the other with collaborative transmission policy among the nodes. Performance of EH nodes in both the clusters is given in Table 2. In each cluster, out of four EH nodes considered, two are employing MMD-RW technique and other two are employing ANN technique. The EH nodes are situated in such a way that one of the two EH nodes employing MMD-RW technique is dimensionally closer to a node employing ANN and the other one is closer to the node employing MMD-RW node. Similar arrangement is made for EH nodes that are employing ANN in their transmission policy design. This arrangement actually helps in finding the energy utilization and the combination that yields better results. Though rewarding factor doesn't depend merely on euclidean distance, from the observations it does make a difference. In Table 2, last four columns of each observation set indicate the performance of WSN with collaborative transmission policy. At every EH node, channel gain estimation technique used in the EH node dimensionally closer to it is represented in parenthesis. MMD-RW(ANN) indicates the performance of EH node which is using MMD-RW technique in its transmission policy and ANN is employed in the EH node metrically closest to it.

## **6 Conclusion**

Design of an efficient transmission policy for an EH sensor has been attempted. An ARQ packet-based communication model with energy harvesting transmitter node has been considered for evaluating the performance of policies. Different transmission policies employing different computational intelligence techniques to predict the channel gain are attempted and their performances are compared. Environment has been simulated to check the robustness of policy and its adaptiveness towards variations in actual environmental conditions. Random and sporadic nature of natural resources has been taken care by having randomness in amount of harvested energy

and frequency of harvesting. The actual channel gain to be estimated is considered to be different from that used for training to comply with variations in channel gain. Results for different set of governing conditions are presented in Table 1. From which, it can be understood that ANN is performing comparatively better in low harvesting conditions and the performance of policies getting improved with increase in available battery level. This can be due to either higher harvesting probability or the better channel gain or increase in number of slots per frame. It can also be observed that the performance of transmission policy with proposed channel gain estimation algorithm, MMD-RW, is better than ANN and ELM, especially when the battery abundance is high, which motivated us to introduce collaborative transmission policy among the EH nodes in a WSN, and from Table 2, a clear improvements in the performance can be observed.

## References

1. J.M. Rabaey, M.J. Ammer, J.L. da Silva, D. Patel, S. Roundy, "PicoRadio supports ad hoc ultra-low power wireless networking", *IEEE Computer Society*, pp.42 – 48, Jul 2000
2. J.A. Paradiso, T. Starner, "Energy scavenging for mobile and wireless electronics", *IEEE Pervasive Computing*, pp. 18 – 27, 2005
3. Sravanthi Chalasani, James M. Conrad, "A survey of energy harvesting sources for embedded systems", *IEEE South east Conf.*, pp.442 – 447, 2008
4. Ozel O, Tutuncuoglu K, Yang J, Ulukus S, Yener A (Sep. 2011) Transmission with energy harvesting nodes in fading wireless channels: Optimal policies. *IEEE J. Sel. Areas Commun.* 29(8):1732–1743
5. Lei J, Yates R, Greenstein L (February 2009) A generic model for optimizing single-hop transmission policy of replenishable sensors. *IEEE Trans. Wireless Commun.* 8:547–551
6. Sharma V, Mukherji U, Joseph V, Gupta S (April 2010) Optimal energy management policies for energy harvesting sensor nodes. *IEEE Trans. Wireless Commun.* 9:1326–1336
7. Gatzianas M, Georgiadis L, Tassiulas L (February 2010) Control of wireless networks with rechargeable batteries. *IEEE Trans. Wireless Commun.* 9:581–593
8. C. Ho and R. Zhang, "Optimal energy allocation for wireless communications powered by energy harvesters", in *IEEE ISIT*, June 2010
9. J. Yang and S. Ulukus, "Transmission completion time minimization in an energy harvesting system", in *CISS*, March 2010
10. J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system", *IEEE Trans. Commun.*, Jan 2012
11. K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes", *IEEE Trans. Wireless Commun.*, Mar 2012
12. I. Stanojev, O. Simeone, Y. Bar-Ness, and D. Kim, "On the energy efficiency of hybrid-ARQ protocols in fading channels", in *Proc. ICC*, 2007, pp. 3173–3177
13. X. Jiang, J. Polastre, and D. Culler, "Perpetual environmentally powered sensor networks", *Proc. 4th ACM/IEEE IPSN*, 2005, pp. 463–468
14. Kansal A, Hsu J, Zahedi S, Srivastava MB (Sep. 2007) Power management in energy harvesting sensor networks. *ACM Trans. Embedded Comput. Syst.* 6(4):32–66
15. Seyed Alireza, Sikdar Biplab (2010) Energy Efficient Transmission Strategies for Body Sensor Networks with Energy Harvesting. *IEEE Transactions on Communications*
16. Tutuncuoglu Kaya, Yener Aylin (2011) Short-Term Throughput Maximization for Battery Limited Energy Harvesting Nodes

17. Shenqiu Z, Seyed A, Sikdar B (Aug. 2013) An analytical approach to the design of energy harvesting wireless sensor nodes. *IEEE Trans. Wireless Commun.* 12(8):4010–4024
18. S. Roundy, D. Steingart, L. Frechette, P.K. Wright, and J.M.Rabaey, “Power Sources for Wireless Sensor Networks”, Proc. First European Workshop Wireless Sensor Networks (EWSN '04), pp. 1–17, Jan. 2004
19. Shaobo Mao, Man Hon Cheung and Vincent W. S. Wong, “Joint Energy Allocation for Sensing and Transmission in Rechargeable Wireless Sensor Networks”, *IEEE Trans. Vehicular Tech.*, vol. 63, no. 6, pp. 2862–2875, Jul 2014
20. B. Zhang, R. Simon, and H. Aydin, “Maximal utility rate allocation for energy harvesting wireless sensor networks”, in *Proc. ACM Int. Conf. Model., Anal., Simul. Wireless Mobile Syst.*, 2011, pp. 7–16
21. Ren-Shiou Liu ; Prasun Sinha ; Can Emre Koksal “Joint Energy Management and Resource Allocation in Rechargeable Sensor Networks”, *Proc. IEEE INFOCOM*, March 2010
22. Stankovic JA, Abdelzaher TE, Lu C, Sha L, Hou JC (Jul. 2003) Real time communication and coordination in embedded sensor networks. *Proc. IEEE* 91(7):1002–1022
23. Zhou S, Chen T, Chen W, Niu Z (Mar. 2015) Outage minimization for a fading wireless link with energy harvesting transmitter and receiver. *IEEE J. Sel. Areas Commun.* 33(3):496–511
24. Sharma MK, Murthy CR (2014) “Packet drop probability analysis of ARQ and HARQ-CC with energy harvesting transmitters and receivers”, in Proc. Atlanta, GA, USA, Dec, IEEE Global Signal Inf. Process., pp 148–152
25. Doshi J, Vaze R (2014) “Long term throughput and approximate capacity of transmitter-receiver energy harvesting channel with fading”, in Proc. Macau, China, Nov, IEEE Int. Conf. Commun. Syst., pp 46–50
26. Yadav A, Goonewardena M, Ajib W, Elbiaze H (2015) “Novel retransmission scheme for energy harvesting transmitter and receiver”, in Proc. London, U.K., Jun, IEEE Int. Conf. Commun., pp 4810–4815
27. Mahdavi-Doost H, Yates RD (2013) “Energy harvesting receivers: Finite battery capacity”, in Proc. Istanbul, Turkey, Mar, IEEE Int. Symp. Inf. Theory, pp 1799–1803
28. Yates RD, Mahdavi-Doost H (2013) “Energy harvesting receivers: Optimal sampling and decoding policies”, in Proc. Austin, TX, USA, Dec, IEEE Global Signal Inf. Process., pp 367–370
29. Moser C, Thiele L, Brunelli D, Benini L (Apr. 2010) Adaptive Power Management for Environmentally Powered Systems. *IEEE Trans. Comput.* 59(4):478–491
30. Lei J, Yates R, Greenstein L (Feb. 2009) A generic model for optimizing single-hop transmission policy of replenishable sensors. *IEEE Trans. Wireless Commun.* 8(2):547–551
31. Aprem A, Murthy CR, Mehta NB (Oct. 2013) Transmit power control policies for energy harvesting sensors with retransmissions. *IEEE J. Sel. Topics Signal Process.* 7(5):895–906
32. Animesh Yadav, Mathew Goonewardena, Wessam Ajib, Octavia A. Dobre and Halima Elbiaze, “Energy Management for Energy Harvesting Wireless Sensors With Adaptive Retransmission”, *IEEE Transactions on communications*, Dec. 2017
33. Nicholas Rosevare and Balasubramaniam Natarajan, “An Alternative Perspective on Utility Maximization in Energy-Harvesting Wireless Sensor Networks”, *IEEE Trans. Vehicular Tech.*, Vol. 63, no. 1, Jan. 2014
34. Bhargav Medepally, Neelesh B. Mehta, Chandra R. Murthy, “Implications of Energy Profile and Storage on Energy Harvesting Sensor Link Performance”, *IEEE Global Telecommunications Conference*, 2009. GLOBECOM 2009
35. Akanksha Sharma, Bharat Kumar Saxena and K. V. S. Rao “Comparison of Wind Speed, Wind Directions, and Weibull Parameters for Sites Having Same Wind Power Density”, *IEEE Intl. Conf. on Technological Advancements in Power and Energy*, 2017
36. Kong Fanxin, Dong Chuansheng (Nov. 2014) Xue Liu and Haibo Zeng “Quantity Versus Quality: Optimal Harvesting Wind Power for the Smart Grid”. Proc. of the IEEE 102(11):1762–1776
37. Masseran N (Mar. 2015) Evaluating wind power density models and their statistical properties. *Energy* 84:533–541

38. Mohammadi K, Alavi O, Mostafaeipour A, Goudarzi N, Jalilvand M (Nov. 2016) Assessing different parameters estimation methods of Weibull distribution to compute wind power density. *E. Convers. Manag.*, 108:322–335
39. Zhang Q, Kassam S (1999) Finite-state Markov model for Rayleigh fading channels. *IEEE Trans. Commun*
40. H. Wang, N. Moayeri, “Finite-state Markov channel– A useful model for radio communication channels”, *IEEE Trans. Veh. Technol.*, 1995
41. Guang-Bin Huang, Qin-Yu Zhu, Chee-Kheong Siew, “Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks”, *IEEE International Joint Conference on Neural Networks*, Jul. 2004
42. Guang-Bin Huang, Qin-Yu Zhu, K.Z. Mao, Chee-Kheong Siew, P. Saratchandran, N. Sundararajan “Can Threshold Networks be Trained Directly?”. *IEEE Tran. on Circuits and Systems II: Express Briefs*, Mar. 2006
43. Nan-Ying Liang, Guang-Bin Huang, P. Saratchandran, N. Sundararajan, “A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks”, *IEEE Tran. on neural networks*, Nov. 2006
44. E S Gopi, “Algorithm Collections for Digital Signal Processing Applications Using Matlab”, *Springer publications*, 2007
45. Christopher M. Bishop, “Pattern Recognition and Machine Learning”, *Springer publications*, 2006

# Author Index

## A

Adhikari, Debasish, 1, 141  
Adla, Abdelkader, 31  
Apte, Sayali, 69

## B

Balavand, Alireza, 51  
Bansode, Nutan V., 1  
Biswal, Santosh Kumar, 155, 169

## G

Gopi, E. S., 13, 177  
Gouda, Nikhil Kumar, 155

## H

Husseinzadeh Kashan, Ali, 51

## J

Jadhav, Mrunalini, 69  
Jadhav, Sangeeta, 141

## K

Khare, Kanchan, 69  
Kulkarni, Anand J., 113  
Kulkarni, Rushikesh, 69

Kulkarni, Smita, 141

Kumar, Meeta, 113  
Kumbhar, Vidya, 127

## M

Mahammad, Shaik, 177

## N

Nair, Sankar N., 13

## R

Rajarapollu, Prachi R., 1

## S

Sarmah, Dipti Kapoor, 91  
Satapathy, Suresh Chandra, 113  
Singh, T. P., 127

## Y

Yogesh, Vineetha, 177

## Z

Zouggar, Souad Taleb, 31