Advances in Intelligent Systems and Computing 1164

Poonam Bansal
Meena Tushir
Valentina Emilia Balas
Rajeev Srivastava *Editors*

Proceedings of International Conference on Artificial Intelligence and Applications

ICAIA 2020

Springer

Advances in Intelligent Systems and Computing

Volume 1164

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Poonam Bansal · Meena Tushir ·
Valentina Emilia Balas · Rajeev Srivastava
Editors

Proceedings of International Conference on Artificial Intelligence and Applications

ICAIA 2020

Springer

Editors
Poonam Bansal
Department of Computer Science
and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India

Meena Tushir
Department of Electrical
and Electronics Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India

Valentina Emilia Balas
Department of Automation
and Applied Software
Aurel Vlaicu University of Arad
Arad, Romania

Rajeev Srivastava
Department of Computer Science
and Engineering
Indian Institute of Technology (BHU)
Varanasi, India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-981-15-4991-5 ISBN 978-981-15-4992-2 (eBook)
https://doi.org/10.1007/978-981-15-4992-2

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

Organization

ICAIA 2020 was organized at Maharaja Surajmal Institute of Technology, Janak Puri, New Delhi, India, during February 6–7, 2020.

ICAIA 2020 Organizing Committee

Patrons

Sh. Kaptan Singh, President, Surajmal Memorial Education Society, New Delhi
Dr. M. P. Poonia, Vice Chairman, AICTE, New Delhi
Prof. Prem Vrat, Pro-Chancellor, North Cap University, Gurugram
Mr. Karnal Singh, IPS, Chief Enforcement Directorate
Ms. Esha Jakhar, Vice President, Surajmal Memorial Education Society, New Delhi
Mr. Ajit Singh Chaudhary, Secretary, Surajmal Memorial Education Society, New Delhi
Mr. Raj Pal Solanki, Treasurer, Surajmal Memorial Education Society, New Delhi

General Chairs

Prof. S. K. Garg, Pro-Vice Chancellor, Delhi Technological University, New Delhi
Mr. Sachin Gaur, Local Coordinator, India EU IST Standards and Collaboration

Organizing Chair

Prof. K. P. Chaudhary, Director and Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi

Convener

Prof. Poonam Bansal, Dy. Director and Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi

Co-conveners

Prof. Meena Tushir, Head, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Naveen Dahiya, Head, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi

Core Committee Members

Dr. Puneet Azad, Head, Electronics and Communication Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Prof. Archana Balyan, Professor, Electronics and Communication Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Prabhjot Kaur, Associate Professor, Information Technology Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Rinky Dwivedi, Associate Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Yogendra Arya, Assistant Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Savita Ahlawat, Reader, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Sunil Gupta, Associate Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Amita Yadav, Reader, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Aashish Sobti, Accounts Officer, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Naresh Kumar, Associate Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Adeel Hashmi, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Tripti Sharma, Head, Information Technology Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Rekha Tripathi, Head, Applied Science Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Koyel Datta Gupta, Head, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Sapna Malik, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Kavita Sheoran, Reader, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Sitender Malik, Assistant Professor, Information Technology Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Bharti Sharma, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi

Ms. Pooja Kherwa, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Vimal Gaur, Reader, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Sonia Goel, Assistant Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Nistha Jatana, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Navdeep Bohra, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Rakhi Kamra, Assistant Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Shaily Malik, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Dr. Geetika Dhand, Reader, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Sachit Rathee, Assistant Professor, Electrical and Electronics Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Ms. Poonam Dhankar, Assistant Professor, Computer Science and Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Manoj Malik, Head, Information Technology Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Ajay Gahlot, Head, Applied Science Department, Maharaja Surajmal Institute of Technology, New Delhi
Mr. Sandeep Singh, Assistant Professor, Electronics and Communication Engineering Department, Maharaja Surajmal Institute of Technology, New Delhi

Advisory Committee

Prof. Sheng-Lung Peng, National Dong Hwa University, Taiwan
Mr. Sachin Gaur, Local Coordinator, India EU ICT Standards Collaboration Project
Prof. Valentina Emilia Balas, Aurel Vlaicu University of Arad, Romania
Prof. Chuan-Ming Liu, National Taipei University of Technology, Taiwan
Prof. Vikram Kumar, Emeritus Prof. IITD and Ex-Dir., NPL, New Delhi
Prof. S. D. Joshi, IIT Delhi, New Delhi
Prof. A. P. Mittal, AICTE and Netaji Subhas University of Technology, New Delhi
Prof. R. C. Bansal, University of Sharjah, Sharjah, UAE
Prof. Shunsaku Hashimoto, University of the Ryukyus, Japan
Prof. Xavier Fernando, Ryerson University, Canada
Dr. Kamal Hossain, International Director, NPL, UK
Dr. Toshiyuki Takastsuj, Director, NMIJ, Japan
Dr. Yi Hua Tang, Physicist, NIST, USA
Dr. D. K. Aswal, Director, CSIR-NPL, New Delhi

Prof. D. P. Kothari, Ex. Director, IIT Delhi
Dr. Krishan Lal, President, INSA and Ex-Director, NPL, India
Mr. Anil Relia, Director, NABL, India
Prof. S. S. Agrawal, Director General, KIIT, Gurgaon
Prof. Yogesh Singh, VC, Delhi Technological University, New Delhi
Prof. S. K. Garg, Pro-VC, Delhi Technological University, New Delhi
Prof. Asok De, Professor, ECE Department, Delhi Technological University, New Delhi
Prof. Smriti Srivastava, Netaji Subhas Institute of Technology, New Delhi
Prof. R. K. Sinha, Director, CSIR-CSIO, Chandigarh, Punjab
Prof. Sanjeev Sofat, Dy. Director, PEC University of Technology, Chandigarh, Punjab
Prof. Amita Dev, Vice Chancellor, IGDTUW, New Delhi
Prof. Shantanu Chaudhary, Director, CSIR-CEERI, Rajasthan, India
Prof. Vishal Bhatnagar, AIACT&R, New Delhi
Prof. S. K. Dhurander, Netaji Subhas University of Technology, New Delhi
Prof. Komal Kumar Bhatia, YMCA, Faridabad
Dr. Ravinder Singla, Dean of University Instruction, Panjab University, India
Prof. Narendar Kumar, Professor, Delhi Technological University, New Delhi
Prof. Zaheeruddi, Jamia Millia Islamia, New Delhi
Prof. Sangeeta Sabharwal, Netaji Subhas University of Technology, New Delhi
Prof. T. S. Bhatti, IIT, New Delhi
Prof. Manjeet Singh, YMCA, Faridabad
Dr. Mohit Gambhir, Director Innovation Cell, MHRD, New Delhi
Prof. Arvinder Kaur, GGSIPU, New Delhi
Prof. Navin Rajpal, USICT, GGSIPU, New Delhi
Prof. C. S. Rai, USCIT, GGSIPU, New Delhi
Prof. B. V. R. Reddy, USICT, GGSIPU, New Delhi
Ms. Sanchita Malik, Scientist F, DRDO
Prof. A. Q. Ansari, Jamia Millia Islamia, New Delhi, India
Prof. R. K. Sahu, Veer Surendra Sai University of Technology, Burla, Odisha, India
Dr. H. D. Mathur, BITS, Pilani, India
Prof. Nidul Sinha, NIT, Silchar
Dr. L. C. Sakia, NIT, Silchar
Dr. R. J. Abraham, IISST, Kerala
Dr. N. K. Gupta, Chairman, ISTE Delhi Section

Technical Committee

Dr. Jasdeep Kaur, IGDTUW, New Delhi
Dr. Mandeep Mittal, AMITY, Noida
Dr. Sapna Gambhir, J. C. Bose, YMCA University of Science and Technology, Faridabad, Haryana

Dr. Saurabh Bhardwaj, Thapar Institute of Engineering and Technology, Punjab
Dr. Priyanka Jain, Delhi Technological University, New Delhi
Dr. Neeta Pandey, Delhi Technological University, New Delhi
Dr. Payal Pahwa, BPIT, New Delhi
Dr. D. K. Tayal, IGDTUW, New Delhi
Dr. Lalit Kumar, Shri Vishwakarma Skill University, Haryana
Dr. Nizamuddim, Noida Authority, Uttar Pradesh
Dr. Puja Dash, GVPCE, Visakhapatnam, Andhra Pradesh
Dr. Nikhil Pathak, Tsinghua University, Beijing, China
Dr. M. K. Debnath, Siksha 'O' Anusandhan University, Bhubaneswar, Odisha, India
Dr. C. K. Shiva, SR Engineering College, Warangal
Dr. Monika Gupta, MAIT, New Delhi
Dr. Pragya Varshney, Netaji Subhas University of Technology, New Delhi
Dr. Maîtreyee Dutta, NITTTR Chandigarh, Punjab
Dr. Anju Saha, USICT, GGSIPU, New Delhi
Dr. M. Bala Krishna, USICT, GGSIPU, New Delhi
Dr. Ravinder Kumar, HMRITM, New Delhi
Dr. Anil Ahlawat, KIIT, Ghaziabad, Uttar Pradesh
Dr. A. K. Mohapatra, IGDTUW, New Delhi
Dr. Rajesh Mehta, Thapar Institute of Engineering and Technology, Punjab
Prof. S. R. N. Reddy, IGDTUW, New Delhi
Dr. Tajinder Singh, NIT, Uttrakhand
Dr. Sujata Pandey, Amity School of Engineering and Technology, Noida, Uttar Pradesh
Dr. Manoj Kumar Pandey, Amity School of Engineering and Technology, Noida, Uttar Pradesh
Prof. Rekha Agarwal, Amity School of Engineering and Technology, Noida, Uttar Pradesh
Dr. Sanjay Kumar Malik, USICT, GGSIPU, New Delhi
Dr. Harish Kumar, UIET, Panjab University, Chandigarh
Dr. Hasmat Malik, Netaji Subhas University of Technology, New Delhi
Dr. V. P. Vishwakarma, GGSIPU, New Delhi
Prof. Zaheeruddin, Jamia Millia Islamia, New Delhi
Dr. I. Routray, C.V. Raman College of Engineering, Orissa
Prof. B. Pattnaik, Director, Amity School of Engineering and Technology, Kolkata
Dr. Pankaj Porwal, Techno India NJR Institute of Technology, Udaipur
Dr. Manoj Kumar, GGSIPU, New Delhi
Dr. Manosi Chaudhuri, BIMTECH, Noida, Uttar Pradesh
Ms. Neetu Bhagat, AICTE, New Delhi
Prof. Manisha Sharma, Bhilai Institute of Technology, Durg
Dr. Neelam Duhan, J. C. Bose, YMCA University of Science and Technology, Faridabad, Haryana
Dr. Manik Sharma, D.A.V. University, Punjab

Dr. Rajesh Jampala, P.B. Siddhartha College of Arts & Science, Siddhartha Nagar, Mogalrajpuram, Vijayawada
Dr. Urvashi Makkar, GLBIMR, Noida, Uttar Pradesh
Dr. Dipayan Guha, MNNIT, Allahabad

ICAIA 2020 Keynote Speaker

Dr. Sankar K. Pal, INSA Distinguished Professor Chair, Distinguished Scientist and Former Director, Indian Statistical Institute, Kolkata, India
Topic: **Granular Data Mining in Video Informatics**

ICAIA 2020 Invited Speakers

Dr. Toshiyuki Takatsuji, Director, Research Institute for Engineering Measurement, National Metrology Institute of Japan, Advanced Industrial Science and Technology, Japan
Topic: **Application of Information Technology for Metrology**

Dr. Vassil Vassilev, Reader in Artificial Intelligence and Cyber Security, London Metropolitan University, London, England
Topic: **AI on Demand Through Hybridization and Containerization**

Dr. Bhuvanesh Unhelkar, Ph.D., FACS, Associate Professor, University of South Florida Sarasota-Manatee, Florida, USA
Topic: **Challenges of Discovering "What to Ask?" of Big Data in Practice and the Role of Artificial Intelligence/Machine Learning in Identifying the Right Questions**

Prof. Smriti Srivastava, Professor, Instrumentation and Control Engineering, Netaji Subhas University of Technology, New Delhi, India
Topic: **Optimization Algorithms for Non-Linear Dynamical Systems**

Dr. Gaurav Gupta, Head, Department Chair School of Mathematical Sciences, College of Science and Technology, Wenzhou-Kean University, Wenzhou, China
Topic: **Exploratory Data Analysis (EDA) Using R Software**

Preface

The International Conference on "Artificial Intelligence and Applications" (ICAIA 2020) intended to provide an international forum for original research findings, as well as exchange and dissemination of innovative, practical development experiences in different fields of artificial intelligence. A major goal and feature of it were to bring academic scientists, engineers and industry researchers together to exchange and share their experiences and research results about most aspects of science and social research and discuss the practical challenges encountered and the solutions adopted.

The responses to the call for papers had been overwhelming—both from India and from overseas. ICAIA 2020 ensured to be both a stimulating and enlightening experience with numerous eminent keynote and invited speakers from all over the world. The event consisted of invited talks, technical sessions, paper presentations and discussions with eminent speakers covering a wide range of topics in artificial intelligence.

This book contains the research papers presented in the conference. Papers have been divided into the following tracks:

- Evolving machine learning and deep learning models for computer vision
- Machine learning applications in cyber security and cryptography
- Advances in signal processing and learning methods
- Social intelligence and sustainability
- Feature extraction and learning on image and speech data
- Optimization techniques and its applications in machine learning
- Recent Trends in Computational Intelligence and Data Science.

We express our sincere gratitude to the eminent keynote speakers, authors and the participants. Our earnest thanks to the Surajmal Memorial Education Society whose unconditional and munificent support has made the dream of hosting an international conference a reality. We would like to express our gratitude and appreciation

for all the reviewers who helped us maintain the high quality of manuscripts included in the proceedings. We are grateful to Springer, especially to Mr. Aninda Bose (Senior Publishing Editor, Springer India Pvt. Ltd.), and the entire Springer team for the excellent collaboration, patience and help during the evolvement of this volume.

New Delhi, India

Organizing Committee
ICAIA 2020

Contents

About the Editors

Dr. Poonam Bansal is currently a Professor at the Computer Science & Engineering Department & Deputy Director of the Maharaja Surajmal Institute of Technology, New Delhi, India. She has 30 years of teaching, industry and research experience. She received her B.Tech. and M.Tech. from Delhi College of Engineering, and her Ph.D. in Computer Science and Engineering from GGSIP University, Delhi. Her research interests include neural networks, speech technology, digital image processing and pattern recognition, and she has published more than 50 research papers in respected international journals and at conferences. She is a member of IEEE, ASI, ISTE and CSI.

Dr. Meena Tushir is a Professor and Head of the Electrical & Electronics Engineering Department at Maharaja Surajmal Institute of Technology, New Delhi, India. She holds a Ph.D. degree in Instrumentation and Control Engineering from Delhi University, and she has over 25 years of teaching experience. Her research areas include fuzzy logic and neural networks, pattern recognition and image segmentation. She has published more than 50 research papers in various respected international journals and at conferences.

Dr. Valentina Emilia Balas is a Professor at the Department of Automatics and Applied Software at the Faculty of Engineering, Aurel Vlaicu University of Arad, Romania, where she is also Head of the Intelligent Systems Research Centre. Her research interests include intelligent systems, fuzzy control, soft computing, smart sensors, information fusion, modeling and simulation, and she has published more than 300 research papers in refereed journals and at international conferences. She is a member of EUSFLAT and SIAM and a senior member IEEE, where she is a member of TC – Fuzzy Systems (IEEE CIS), Emergent Technologies (IEEE CIS), and Soft Computing (IEEE SMCS).

Dr. Rajeev Srivastava is currently a Professor and Head of the Department of Computer Science & Engineering, IIT (BHU), Varanasi, India. He has over 20 years of teaching experience and has published more than 100 research papers in respected international journals and at conferences. His research interests are in the field of image processing, computer vision, pattern classification, video surveillance, medical image processing and pattern recognition, mathematical image modeling and analysis, artificial intelligence and machine vision and processing. He is a Fellow of IETE, IEI and ISTE (India).

Evolving Machine Learning and Deep Learning Models for Computer Vision

Analysis of Breast Cancer Detection Techniques Using RapidMiner

Adhish Nanda and Aman Jatain

Abstract One of the most widely spread diseases among women is the breast cancer. In past few years, the incidences of breast cancer kept on rising. At this point, diagnosis of the cancer is crucial. However, breast cancer is treatable if it is identified during the earlier stages. Classification of tumor in case of breast cancer is done using machine learning algorithms, viz. decision trees, regression, SVM, k-NN, and Naïve Bayes. Then, accuracy of these algorithms is compared to predict the class (benign, malignant) of tumor, and the most appropriate algorithm is suggested based on the results. Wisconsin Breast Cancer Diagnostic Dataset is used. The data has been preprocessed, split, and applied to respective models. Tenfold cross-validation is applied to determine the accuracies.

Keywords Breast · Cancer · Decision trees · KNN · Machine learning · Naïve Bayes · Prediction · RapidMiner · Regression · SVM

1 Introduction

The single most common disease responsible for high number of deaths in women is the breast cancer. However, by classifying the type of tumor in a women's breast, we can diagnose whether the nature of tumor is benign or malignant. Doctors and scientists are in search of various methods and techniques that may help in distinguishing the tumors. There are various factors that pose as risks in developing breast cancer like being a woman, or being obese, having less or no physical exercise, family history, alcohol and drug abuse, late pregnancy, among several others. According to reports from the American Cancer Society, breast cancer is treatable if detected early by providing the required treatment before the cancer reaches its maximum growth phase. Also, the accurate classification of tumors into one of the

A. Nanda · A. Jatain (✉)
Amity School of Engineering and Technology, Amity University Gurgaon, Gurgaon, Haryana 122413, India
e-mail: amanjatainsingh@gmail.com

P. Bansal et al. (eds.), *Proceedings of International Conference on Artificial Intelligence and Applications*, Advances in Intelligent Systems and Computing 1164,
https://doi.org/10.1007/978-981-15-4992-2_1

two categories (benign or malignant) can prevent the patient from going through unnecessary treatments which may result in the wastage of both time and money. Hence, the proper identification and classification of tumors in and around breasts are a concern of much research.

Classification is the process of assigning a new observation to one of the categories in a set based on its characteristics and properties. This is done with the help of a training set of data whose variable assignments are already predetermined or known [1]. Classification is considered an instance of supervised learning, i.e., learning with the help of a training set where correctly identified observations are available. The corresponding unsupervised procedure is known as clustering and involves grouping data into categories based on some measure of inherent similarity or distance. In this work, the dataset has been divided into test and train data, respectively. The training dataset is used to train the different models, while test dataset is used to verify their functionality.

2 Related Work

Every year doctors diagnose around 14 million new people with cancer. But research shows that the prognosis rate after detection is only 60%. This is where machine learning comes into play. There are various researches available on the topic. Some of them are mentioned below.

Mittal et al. proposed [2] a hybrid strategy for breast cancer investigation which gave a compelling certainty over the dataset. This proposed technique consolidates unsupervised self-arranging maps with a regulated classifier called stochastic gradient descent. The test results are led by contrasting their outcomes with three machine learning algorithms, namely SVM, random forest, and decision trees. Vasantha et al. [3] focused on characterizing the mammogram pictures into three classifications (ordinary, benign, and malignant). Then, Halawani et al. have practiced diverse clustering algorithms on them so as to recognize a breast tumor. Trials were directed utilizing digital mammograms in the University of Erlangen-Nuremberg somewhere in the range of 2003–2006.

A review done by Saranya and Satheeskumar [4] has broken down a noteworthy number of investigations done in the field of breast cancer identification. Their emphasis was on analyzing the diverse data mining procedures applied in breast cancer classification alongside their advantages and inconveniences. Especially, this review talks about C4.5 and ID3 algorithms and their utilization in the examination and characterization of breast cancer.

Kourou et al. [5] examined various machine learning algorithms in cancer forecast and its treatment. In the introduced survey of around 70 approaches, they arrived at a resolution, that in the last years, the exploration is centered around the improvement of prescient models utilizing regulated ML strategies and grouping calculations where the combination of heterogeneous data that is multidimensional in nature is joined with the utilization of various methods for highlighting choices. Then, Cruz and

Wishart [6] studied the execution of various AI algorithms that are being connected to breast cancer forecast, and prognosis has been clarified, analyzed, and evaluated. They recognized various patterns concerning the sorts of AI strategies being utilized, the sorts of training data being incorporated, the sorts of endpoint forecasts being made, the kinds of diseases being contemplated, and the general execution of these techniques in foreseeing malignancy vulnerability or results.

Syed Shajahaan et al. [7] took a shot at anticipating the presence of breast cancer with the help of data mining. They used the decision tree algorithm for this. Information gathered contained 699 examples (understanding records) with ten qualities and the class label as benign or malignant to determine its severity. Info utilized contained observation ID, thickness of its cluster, consistency in the shape of the tumor and the size of the cells, its development and different outcomes of other physical examination. Results of the applied supervised algorithms demonstrated that the random forest had the most elevated exactness of 100% with a blunder rate of 0, whereas CART had the least precision with an estimation of 92.99%, yet Naive Bayes had a precision of 97.42% with an error rate of 0.0258.

Mangasarian et al. [8] applied classification techniques on demonstrative data of breast cancer. The grouping strategy received by them for demonstrative information is called Multisurface Method Tree (MSM-T) which utilizes a programming model which continually puts a progression of isolating planes in the component space of the precedents. On off chance that the two arrangements of focuses are straightly distinguishable, the main plane is inserted between them. When the sets are not directly distinct, MSM-T will build a plane which limits the normal separation wrongly classified values to the plane, in this manner almost limiting the quantity of nonclassified or wrong-classified focuses. The training division and the forecast precision with the MSM-T approach were 97.3 and 97% separately, while the RSA approach had the capacity to give exact expectation just for every individual patient. But the inalienable linearity of the prescient models was a major issue relating to their downside.

Nalini and Meera [9] displayed an investigation about breast cancer dependent on data mining techniques to find a successful method to foresee it by recognizing an exact model to anticipate the occurrence of cancer dependent on patients' clinical records. Information mining models that were utilized are Naive Bayes, SVMs, ANNs, and AdaBoost tree. Foremost the dimensionality reduction method of principal component analysis (PCA) alongside the proposed models was utilized to decrease the element space. The execution assessment of this model was controlled by utilizing the Wisconsin Breast Cancer Database (1991) and Wisconsin Diagnostic Breast Cancer (1995). After this, Asri et al. [10] performed correlation between various AI algorithms such as SVM, C4.5, Naive Bayes (NB), and K-Nearest Neighbor (k-NN) on the Wisconsin Breast Cancer Datasets to survey the rightness in characterizing information as for efficiency and adequacy of every algorithm regarding exactness, accuracy, affectability, and explicitness. All investigations are executed inside a reenactment situation and led in WEKA information mining apparatus.

3 Comparative Study of Machine Learning Algorithms

There are various machine learning algorithms and statistical approaches that are used by the computer systems to perform specific tasks like pattern identification, inference, etc., from data. Here, we have chosen the most popular and effective ones and applied them to determine the best among them in case of predicting a medical condition. The dataset used is Wisconsin Breast Cancer Dataset which has around 32 real-valued attributes, zero missing values, and class distribution with 357 benign and 207 malignant.

3.1 Decision Trees

A decision tree is an algorithm that makes use of a tree-like structure of decisions or choices and their conceivable results, including arbitrary occasions, asset expenses, and utility. It is a single approach that shows a calculation which consists of only restrictive control edicts. Decision tree is also known as C4.5 algorithm. It makes use of a decision tree as a prescient model to go from impressions of an object (represented as branches) to decisions about the object's unprejudiced value (represented as leaves). It is also the predictive modeling method utilized in measurements, mining of useful information from data and AI. Structures in which variable can take a discrete value of qualities are used characterization or classification; in these, leaves represent respective labels of a class, and branches represent conjunctions of attributes that lead to those class labels. Decision trees where the variables accept continuous and real values or qualities are known as regression trees. Learning in a decision tree refers to the development of a tree-like structure designed from class-marked preparing tuples. It has a stream outline structure, in which each inward hub (nonleaf) represents a test on a property, each branch represents result of the test, and each leaf (or terminal) hub represents a class mark (Fig. 1).

Utilizations in a decision tree is a white box display because on the off chance that when any given circumstance is recognizable in the model, the clarification for the condition is effectively underlined using Boolean rationale. On the other hand, in the black box display, the clarification of the outcomes is normally hard to comprehend, for instance, with a counterfeit neural system. But decision trees are very little robust in nature. A little change in the training data can result in a substantial change in results of predictions (Table 1).

As per the results, we can see that the mean accuracy is around 93.67%. There are seven instances of data where it should have been classified as Benign, but rather it is classified as malignant and 29 instances where it should have been classified as malignant but rather is classified as benign. However, the class precisions are quite high.

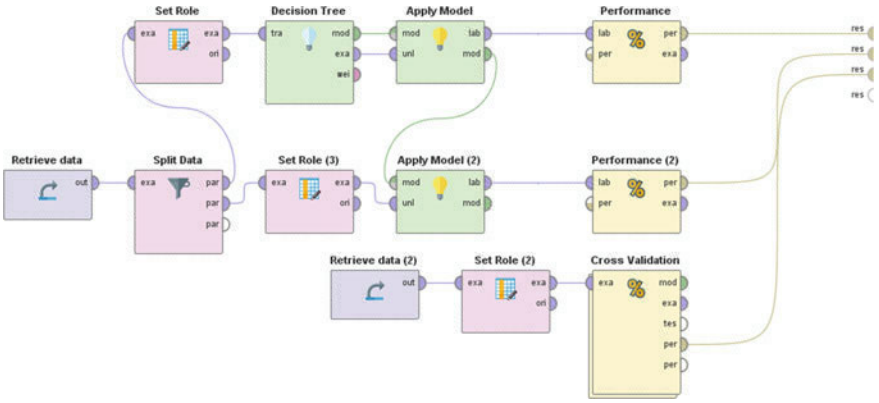

Fig. 1 Decision tree classifier in RapidMiner

Table 1 Classification and accuracy results of decision tree

	True M	True B	Class precision (%)
Pred. M	183	7	96.32
Pred. B	29	350	92.35
Class recall (%)	86.32	98.04	

Accuracy: 93.68% ± 3.69% (micro average: 93.67%)

3.2 k-NN

The k-NN or K-Nearest Neighbor algorithm is a nonparametric technique used for characterization or regression purposes [11]. In both the scenarios, the input data comprises the k number of nearest training variables in the component space. The results depend upon whether k-NN is used for classification or for regression analytics. In classification utilizing the k-NN, the output is a class enrollment. An item is characterized by the majority of votes it receives from its neighbors, with the article being assigned to the class which is the most normal among its k closest neighbors (k is a small positive number). On other hand, if $k = 1$, at that point the item is basically assigned to the class of single closest neighbor. In k-NN regression, the output is the property estimation for the article which is the mean of the estimations of its k number of neighbors nearest to it. It is a sort of occasion-based algorithm for learning, or apathetic realizing, in which capacity is computed and approximated locally, and all other calculations are conceded until arrangement. The calculations involved in k-NN algorithms are the least complex when compared to calculations that are involved in other AI algorithms. The preparation models are vectors in a multidimensional component space, each with a class mark. Training period of the model comprises putting away the element vectors and class marks of the train sets. Here, k is an unlabeled vector which is consistently characterized by client and is arranged by allotting the mark which is most sequential among the k training sets closest to

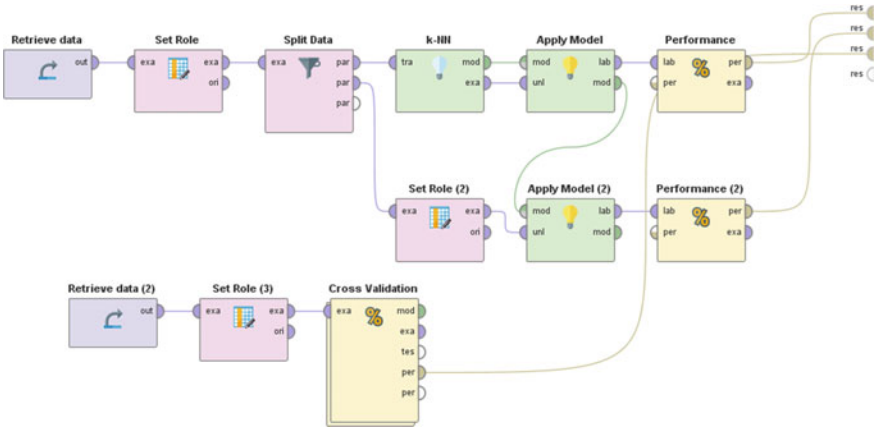

Fig. 2 Classifier using k-NN in RapidMiner

Table 2 Classification and accuracy results of k-NN

	True M	True B	Class precision (%)
Pred. M	106	17	86.18
Pred. B	106	340	76.23
Class recall (%)	50.00	95.24	

Accuracy: 78.40% ± 4.11% (micro average: 78.38%)

that query point. Most common separation metric for continuous factors is Euclidean distance. For discrete factors like content arrangement, another measurement can be utilized, for example, the cover metric (or Hamming distance) (Fig. 2).

Obtained results indicate that k-NN is not suitable for classifying or predicting the type of tumor in the case of breast cancer (Table 2).

3.3 Neural Networks

The neural nets are a collection of algorithms, constructed loosely afterward hominoid cerebrum [12], which is envisioned to identify designs and patterns. They decipher palpable data over a kind of machine recognition application responsible for marking or bunching unpolished data [13]. The illustrations they observe are numerical in nature, mostly confined in vectors, into which all true data like pictures, sound, or content classification can be interpreted. Systems based on neural networks help us cluster and arrange. Neural systems can likewise extract and fetch topographies that are fed to different processes for clustering and classification; so, you can consider profound neural systems as parts of bigger AI bids including calculations for reinforcement learning, characterization, and relapse. Neural nets or neural networks

form the basis or the major part of deep learning. Deep learning maps contributions (inputs) toward yielding results (outputs). It notices relations. It is known as a "Universal Approximator," on the grounds that it can fathom how to surmise an obscure capacity $f(a) = b$ between any input a and output b, expecting they are connected by any stretch of the imagination (by relationship or causation, for instance). During the time spent learning, a neural system finds the correct f, or the right way of changing "a" into "b", irrespective of whether that be $f(a) = 43a + 121$ or $f(a) = 15a - 0.3$. All grouping tasks rely on datasets that are already labeled; explicitly, individuals should exchange their insight with the dataset unanimously for a neural system to become acquainted with the relationship among labels and data. This regulated form of learning is called supervised learning (Fig. 3).

By a similar token, presented to enough of the correct kind of data, deep learning can expand connections between present occasions and future occasions. It can run a regression between previous times and what's to come. Deep learning is the name we use for "stacked neural systems," viz. systems made from several layers. Table 3 indicates that although class precisions are not perfect, the overall accuracy of the classification is still very much high.

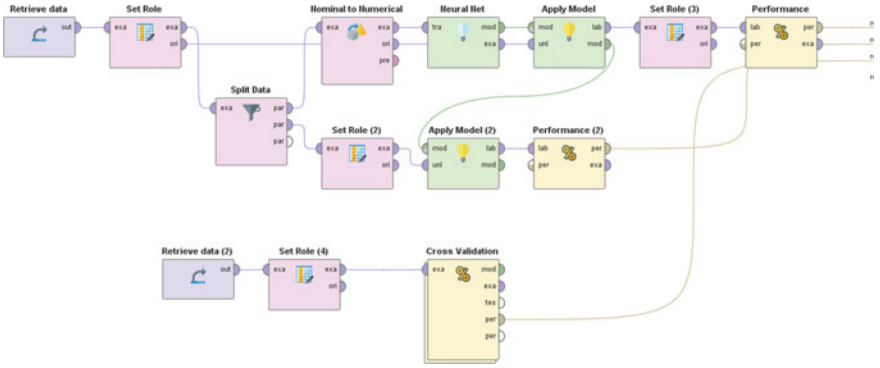

Fig. 3 Simple neural network classifier in RapidMiner

Table 3 Classification and accuracy results of neural nets

	True M	True B	Class precision (%)
Pred. M	202	7	96.65
Pred. B	10	350	97.22
Class recall (%)	95.28	98.04	

Accuracy: 97.01% ± 2.24% (micro average: 97.01%)

3.4 Linear Regression

In stats, linear regression is a straightforward way to demonstrate the connection between scalar values, i.e., dependent variable and the independent variables. The occurrence of one rational variable is called a simple linear regression. For over one descriptive variable, the process is called multiple linear regression [14]. This word is different from multivariate linear regression, where different corresponded secondary factors are estimated, instead of a single scalar variable [15]. In linear regression, networks are displayed employing direct indicator mechanisms whose vague model parameters are calculated from input data. These are known as linear models [16]. Utmost usually, the limiting mean of the reaction assumed by the estimations of the logical factors (or indicators) is thought as a relative capacity of them qualities; irregularly, the contingent medium or some other quantile is employed (Fig. 4).

Corresponding types of regression analytics, linear regression centers around the conditional probability distribution of the reaction given the estimations of the predictors, as opposed to on the joint probability distribution of these factors in space of multifarious investigation. Linear regression was the main kind of regression analysis to be considered thoroughly, which is also utilized widely in down-to-earth solicitations [17]. This is accordingly to the models which depend directly on their incomprehensible structures and simpler to fit than models which are nondirectly acknowledged with their structures and due to that the realistic properties of the following estimators are less challenging to decide. Least squares approach is generally used to fit the linear models. From the results, we can observe that class precisions are quite high, and this means that the classification done is genuine, but the accuracy of prediction is not that great (Table 4).

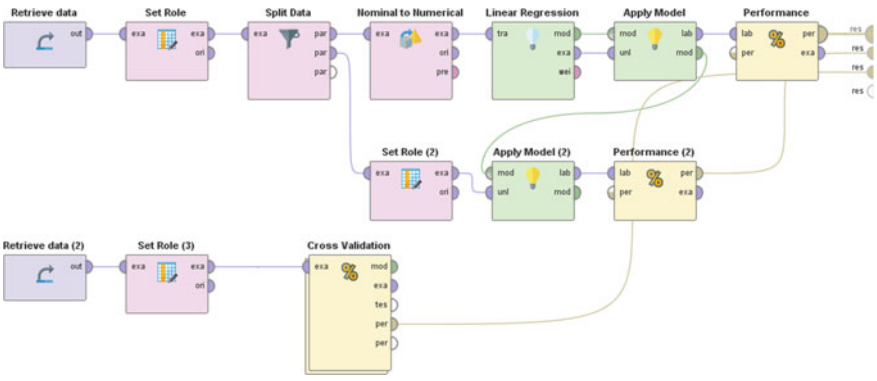

Fig. 4 Classification with linear regression in RapidMiner