Lecture Notes in Networks and Systems 868

Sandeep Kumar
Balachandran K.
Joong Hoon Kim
Jagdish Chand Bansal *Editors*

Fourth Congress on Intelligent Systems

CIS 2023, Volume 1

Springer

Lecture Notes in Networks and Systems

Volume 868

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose (aninda.bose@springer.com).

Sandeep Kumar · Balachandran K. ·
Joong Hoon Kim · Jagdish Chand Bansal
Editors

Fourth Congress on Intelligent Systems

CIS 2023, Volume 1

Springer

Editors
Sandeep Kumar
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India

Balachandran K.
Department of Computer Science and Engineering
Christ (Deemed to be University)
Bengaluru, India

Joong Hoon Kim
School of Civil, Environmental and Architectural Engineering, Anam-ro Anam-dong
Korea University
Seoul, Korea (Republic of)

Jagdish Chand Bansal
South Asian University
New Delhi, Delhi, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-99-9036-8 ISBN 978-981-99-9037-5 (eBook)
https://doi.org/10.1007/978-981-99-9037-5

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

This book contains outstanding research papers as the proceedings of the 4th Congress on Intelligent Systems (CIS 2023), held on September 4–5, 2023, at CHRIST (Deemed to be University), Bengaluru, India, under the technical sponsorship of the Soft Computing Research Society, India. The conference is conceived as a platform for disseminating and exchanging ideas, concepts, and results from academia and industry researchers to develop a comprehensive understanding of the challenges of intelligence advancements in computational viewpoints. This book will help in strengthening congenial networking between academia and industry. We have tried our best to enrich the quality of the CIS 2023 through the stringent and careful peer-review process. This book presents novel contributions to Intelligent Systems and is a reference material for advanced research in medical imaging and health informatics, agriculture, and engineering applications.

We have tried our best to enrich the quality of the CIS 2023 through a stringent and careful peer-review process. CIS 2023 received many technical contributed articles from distinguished participants from home and abroad. CIS 2023 received 870 research submissions from different countries. After a very stringent peer-reviewing process, only 104 high-quality papers were finally accepted for presentation and the final proceedings.

This book presents the first volume of 35 data science and applications research papers and serves as a reference material for advanced research.

Bengaluru, India	Sandeep Kumar
Bengaluru, India	Balachandran K.
Seoul, Korea (Republic of)	Joong Hoon Kim
New Delhi, India	Jagdish Chand Bansal

Contents

Editors and Contributors

About the Editors

Dr. Sandeep Kumar is a professor at CHRIST (Deemed to be University), Bengaluru. He recently completed his post-doctoral research at Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia, in sentiment analysis. He is an associate editor for Springer's *Human-Centric Computing and Information Sciences (HCIS)* journal. He has published over a hundred research papers in various international journals/conferences and attended several national and international conferences and workshops. He has authored/edited 12 books in the area of computer science. Also, he has been serving as General Chair of the Congress on Intelligent Systems (CIS), the International Conference on Communication and Computational Technologies (ICCCT), the International Conference on Sustainable and Innovative Solutions for Current Challenges in Engineering & Technology (ICSISCET) and the IEEE International Conference on Contemporary Computing and Communications (InC4). His research interests include nature-inspired algorithms, swarm intelligence, soft computing, and computational intelligence.

Dr. Balachandran K. is formerly Professor and Head of Computer Science and Engineering at Christ (Deemed to be University), Bengaluru, India. He has 38 years of experience in Research, Academia, and Industry. He served as Senior Scientific Officer in the Research and Development Unit of the Department of Atomic Energy for 20 years. His research interest includes Data Mining, Artificial Neural Networks, Soft Computing, and Artificial Intelligence. He has published more than 50 articles in well-known SCI/SCOPUS-indexed international journals and conferences and attended several national and international conferences and workshops. He has authored/edited four books in the area of computer science.

Prof. Joong Hoon Kim formerly Dean of Engineering College of Korea University, obtained his Ph.D. degree from the University of Texas at Austin in 1992 with the thesis title "Optimal replacement/rehabilitation model for water distribution systems". Professor Kim's major areas of interest include Smart Water Management Systems and Micro Water Grid Systems in Smart Cities, Design and Management of Water Distribution Systems, Maintenance and Management Techniques of Urban Drainage Systems, Real-time Management of Urban Water Systems Involving Smart Monitoring, Telemetered Systems, and Sensor Networks, Development of Alarm System against Urban Flood, Development of Optimal Decision Making Systems about Management of Water Supply and Sewage Networks. His publication includes "A New Heuristic Optimization Algorithm: Harmony Search", Simulation, February 2001, Vol. 76, pp. 60–68, cited over 7200 times by other journals of diverse research areas. He has been awarded the National Study Abroad Scholarship, Korean Ministry of Education, 1985.05; Quentin Mees Research Award, Arizona Water and Pollution Control Association 1994.03, Korea Water Resources Association Achievement Award 2005.12; National Emergency Management Agency Achievement Award 2007.11; Korea Minister of Government Administration and Home Affairs Award 2012.02, KSCE-Springer Award, 2013.12, Songsan Grand Prize (Academic) by KSCE, 2019.10, Order of Science and Technological merit, 2021.04, K-water Grand Academic Prize, 2022.12. He has been on the faculty of the School of Civil, Environmental, and Architectural Engineering at Korea University since 1993 and is now the professor emeritus. He has hosted international conferences, including APHW 2013, ICHSA 2014 and 2015, HIC 2016, HIC 2018, ICHSA 2019, ICHSA 2020, and ICHSA 2022, and has given keynote speeches at many international conferences, including AOGS 2013, GCIS 2013, SocPros 2014 & 2015, SWGIC 2017, RTORS 2017, ICHSA 2020, ICCIS 2020, and FSACE 2021. He is a member of the National Academy of Engineering of Korea since 2017.

Dr. Jagdish Chand Bansal is an Associate Professor (Senior Grade) at South Asian University New Delhi and a Visiting Faculty at Maths and Computer Science Liverpool Hope University UK. He also holds a visiting professorship at NIT Goa, India. Dr. Bansal obtained his Ph.D. in Mathematics from IIT Roorkee. Before joining SAU New Delhi, he worked as an Assistant Professor at ABV-Indian Institute of Information Technology and Management Gwalior and BITS Pilani. His Primary area of interest is Swarm Intelligence and Nature Inspired Optimization Techniques. Recently, he proposed a fission-fusion social structure-based optimization algorithm, Spider Monkey Optimization (SMO), which is being applied to various problems in the engineering domain. He has published over 70 research papers in various international journals/conferences. He is the Section Editor (editor-in-chief) of the journal *MethodsX* published by Elsevier. He is the series editor of the book series *Algorithms for Intelligent Systems (AIS)*, *Studies in Autonomic, Data-driven and Industrial Computing (SADIC)*, and *Innovations in Sustainable Technologies and Computing (ISTC)* published by Springer. He is also the Associate Editor of Engineering Applications of Artificial Intelligence (EAAI) and ARRAY, published by Elsevier. He is the general secretary of the Soft Computing Research Society (SCRS). He has also received Gold Medal at UG and PG levels.

Contributors

K. A. Abdul Nazeer National Institute of Technology, Calicut, India

M. Abinesh Department of Aeronautical Engineering, ACS College of Engineering, Bangalore, India

Shajal Afaq Centre for Advanced Studies, AKTU, Lucknow, Uttar Pradesh, India

P. S. Ambika Center for Computational Engineering and Networking (CEN), Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India

T. Ambika Department of Aeronautical Engineering, ACS College of Engineering, Bangalore, India

Milos Antonijevic Singidunum University, Belgrade, Serbia

V. Aparna National Institute of Technology, Calicut, India

C. Arul Murugan Karpagam College of Engineering, Coimbatore, India

Senthil Kumar Arumugam Christ University, Bangalore, India

Shree Harsh Attri Department of Computer Science and Engineering, Sharda School of Engineering and Technology, Sharda University, Greater Noida, UP, India

Nebojsa Bacanin Singidunum University, Belgrade, Serbia

Abou Bakary Ballo LaMI, Université Felix Houphouët-Boigny, Abidjan, Côte d'Ivoire;
LMI Université Péléforo Gon Coulibaly, Korhogo, Côte d'Ivoire

Poonam Bansal Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, India

Daniel Bell Robotics Lab, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, Liverpool, UK

Kesari Eswar Bhageerath Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering (Autonomous), Madhurawada, Visakhapatnam, Andhra Pradesh, India

Rajesh Bhatia Punjab Engineering College (Deemed to be University), Chandigarh, India

Aleksandra Bozovic Technical Faculty "Mihajlo Pupin", University of Novi Sad, Zrenjanin, Serbia

Subhadeep Chakrarbarti Amity University Jharkhand, Ranchi, India

Lekshmy P. Chandran National Institute of Technology, Calicut, India

Kundan Kumar Chaubey School of Applied and Life Sciences, Uttaranchal University, Dehradun, Uttarakhand, India;
School of Applied and Life Sciences, Sanskriti University, Mathura, Uttar Pradesh, India

Ganesh Kumar Chellamani Department of Electronics and Communication Engineering, Amrita School of Engineering, Chennai, Amrita Vishwa Vidyapeetham, Chennai, India

Jyothsna Cherapanamjeri Jawaharlal Technological University, Anantapur, India

Adama Coulibaly LaMA Université Felix Houphouët-Boigny, Abidjan, Côte d'Ivoire

Sarat C. Dass School of Mathematical & Computer Sciences, Heriot-Watt University, Putrajaya, Malaysia

Subhojit Dawn Department of Electrical and Electronics Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India

Eleonora Desnica Technical Faculty "Mihajlo Pupin", University of Novi Sad, Zrenjanin, Serbia

K. Dhananjay Rao Department of Electrical and Electronics Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India

Ritika Dhaneshwar University Institute of Engineering and Technology, Panjab University, Chandigarh, India

R. Dharani Department of IT, Panimalar Engineering College, Chennai, India

Moustapha Diaby Lastic, Ecole Supérieure Africaine des Technologies de l'Information et de la Communication, Abidjan, Côte d'Ivoire

Lynette D'mello Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

Gauravkumarsingh Gaharwar Navrachana University, Vadodara, Gujarat, India

Prakash Garia School of Management, Graphic Era Hill University, Bhimtal, Uttarakhand, India

Georgi Georgiev South-West University "Neofit Rilski", Blagoevgrad, Bulgaria

Urvashi Goyal Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India

D. N. D. Harini Computer Science and Engineering, Gayatri Vidya Parishad College of Engineering (Autonomous), Madhurawada, Visakhapatnam, Andhra Pradesh, India

S. Harsha Department of Artificial Intelligence and Machine Learning, RNS Institute of Technology, Visvesveraya Technological University, Belagavi, India

K. V. Hridika National Institute of Technology, Calicut, India

Navin K. Ipe Computer Science and Engineering, M. S. Ramaiah University of Applied Sciences, Bangalore, India

Anamika Jain Dr. Vishwanath Karad, MIT-World Peace University, Kothrud, Pune, Maharashtra, India

R. Jeena Department of CSE, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

Saju P. John Department of Computer Science and Engineering, Jyothi Engineering College, Cheruthuruthy, Kerala, India

Luka Jovanovic Singidunum University, Belgrade, Serbia

N. Juber Rahman Computer Science, PSG College of Arts and Science, Coimbatore, India

A. Kalavani Rajalakshmi Engineering College, Chennai, India

Soumen Kanrar Amity University Jharkhand, Ranchi, India

Mandeep Kaur University Institute of Engineering and Technology, Panjab University, Chandigarh, India

S. Kiruthiga Department of ECE, Saranathan College of Engineering, Trichy, India

R. Kottaimalai Kalasalingam Academy of Research and Education, Krishnankoil, India

Anjan K. Koundinya Department of Information Science and Engineering (Cyber Security), BMS Institute of Technology and Management, Visvesveraya Technological University, Belagavi, India

Raghavendra V. Kulkarni Computer Science and Engineering, Graphic Era Deemed to be University, Dehradun, India

K. R. Pruthvi Kumar Department of Computer Science and Engineering, BMS Institute of Technology and Management, Visvesveraya Technological University, Belagavi, India

Ritika Kumari Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, India;
USICT, Guru Gobind Singh Indraprastha University, New Delhi, India

K. Likitha Shree Department of Aeronautical Engineering, ACS College of Engineering, Bangalore, India

Diarra Mamadou LaMI, Université Felix Houphouët-Boigny, Abidjan, Côte d'Ivoire

Joseph P. Mani Modern College of Business and Science, Muscat, Sultanate of Oman

Ashapurna Marndi Council of Scientific and Industrial Research-Fourth Paradigm Institute, Bangalore, Karnataka, India;
Academy of Scientific and Innovative Research, Ghaziabad, Uttar Pradesh, India

S. Marulkar MATS School of IT, MATS University, Raipur, CG, India

R. Mente School of Computational Sciences, PAH Solapur University, Solapur, India

J. Midhunchakkaravarthy Lincoln University College, Petaling Jaya, Selangor, Malaysia

Amit Mittal Department of Allied Science, Graphic Era Hill University, Bhimtal, Uttarakhand, India

R. Mukesh Department of Aerospace Engineering, ACS College of Engineering, Bangalore, India

Lalasa Mukku CHRIST (Deemed to be University) Kengeri, Bangalore, India

S. Mythili Department of ECE, PSNA College of Engineering and Technology, Dindigul, India

G. S. Nagaraja Department of Computer Science and Engineering, RV College of Engineering, Visvesveraya Technological University, Belagavi, India

Pooja S. Nair National Institute of Technology, Calicut, India

Nikitha Reddy Nalla Department of Electronics and Communication Engineering, Amrita School of Engineering, Chennai, Amrita Vishwa Vidyapeetham, Chennai, India

B. Narain MATS School of IT, MATS University, Raipur, CG, India

B. Narendra Kumar Rao Sri Vidyanikethan Engineering College, Tirupathi, India

Ivan Nedyalkov South-West University “Neofit Rilski”, Blagoevgrad, Bulgaria

P. Nithya Networking and Mobile Application, PSG College of Arts and Science, Coimbatore, India

P. S. Niveditha APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala, India

Ganesh Pai NMAM Institute of Technology-Affiliated to NITTE (Deemed to be University), Nitte, Karnataka, India

Sharnil Pandya Symbiosis International University, Pune, Maharashtra, India

Pooja Department of Aeronautical Engineering, ACS College of Engineering, Bangalore, India

Mukul Prabhu Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

Tatiparti B. Prasad Reddy SENSE, Vellore Institute of Technology, Chennai, India

M. Prem Kumar Department of Electrical and Electronics Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India

Shefali Raina Vasantdada Patil Pratisthan's College of Engineering and Visual Arts Sion, Mumbai, Maharashtra, India

Ranjana Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India

Harish Rawat Department of Botany, Dhanauri P. G. College Dhanauri, Haridwar, Uttarakhand, India

Vidhi Ruparel Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

Shruti Sahu Department of Artificial Intelligence and Data Sciences, IGDTUW, New Delhi, India

N. Sai Dhanush Center for Computational Engineering and Networking (CEN), Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore, India

Aarti Sardhara Vishwakarma University Pune, Pune, Maharashtra, India

Madhvi Saxena Vishwakarma University Pune, Pune, Maharashtra, India

Emanuele Lindo Secco Robotics Lab, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, Liverpool, UK

N. Senthamilarasi Department of CSE, Sathyabama Institute of Science and Technology, Chennai, India

Sonia Setia Department of Computer Science and Engineering, Sharda School of Engineering and Technology, Sharda University, Greater Noida, UP, India

Rushabh Shah Dwarkadas J. Sanghvi College of Engineering, Mumbai, India

Aarti Mehta Sharma Symbiosis Centre for Management Studies, Bengaluru Campus, Symbiosis International (Deemed University), Pune, India

Shweta Sharma College of Biotechnology, DUVASU, Mathura, Uttara Pradesh, India

M. Sharmila Kumari P. A. College of Engineering-Affiliated to VTU, Mangalore, Karnataka, India

Suman Shit Amity University Jharkhand, Ranchi, India

Chintala Rahulsai Shrinivas Punjab Engineering College (Deemed to be University), Chandigarh, India

Rishi Pal Singh Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India

Aashna Sinha School of Applied and Life Sciences, Uttaranchal University, Dehradun, Uttarakhand, India

G. Soniya Priyatharsini Department of CSE, Dr.MGR Educational and Research Institute, Chennai, India

Gummadi Srinivasa Rao Department of Electrical and Electronics Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada, India

Sasidhar Babu Suvanam Department of Computer Science and Engineering, Presidency University, Bangalore, India

Arnav Taya University Institute of Engineering and Technology, Panjab University, Chandigarh, India

M. Thilagaraj MVJ College of Engineering, Bengaluru, India

Jyothi Thomas CHRIST (Deemed to be University) Kengeri, Bangalore, India

Meena Ugale Lincoln University College, Petaling Jaya, Selangor, Malaysia

Naramula Venkatesh SR UNIVERSITY Warangal, Warangal, India

M. Vijay Department of Aerospace Engineering, ACS College of Engineering, Bangalore, India

Vydeki SENSE, Vellore Institute of Technology, Chennai, India

Shruti Wadhwa Punjab Engineering College (Deemed to be University), Chandigarh, India

Suhail Rashid Wani Department of Computer Science and Engineering, Sharda School of Engineering and Technology, Sharda University, Greater Noida, UP, India

Miodrag Zivkovic Singidunum University, Belgrade, Serbia

A Simple Way to Predict Heart Disease Using AI

Soumen Kanrar, Suman Shit, and Subhadeep Chakrarbarti

Abstract Early diagnosis of cardiovascular diseases in high-risk patients will help them make decisions about lifestyle changes and, in turn, minimise their complications. Due to asymptomatic illnesses like cardiovascular diseases, healthcare costs are exceeding the average national medical treatment cost and corporate budgets. The need for early identification and treatment of such diseases is critical. One of the developments in machine learning is the technology that has been used for disease prediction in many fields around the world, including the healthcare industry. Analysis has been attempted to classify the most influential heart disease causes and to reliably predict the overall risk using homogeneous techniques of data mining. This paper uses machine learning algorithms and selects the best one based on its classification report to find a simple way to predict heart disease. It helps to develop cost-effective software to predict heart disease for the betterment of mankind.

Keywords Heart disease · Chest pain · Cardiovascular · Machine learning · AdaBoost · Random forest

1 Introduction

Recent research has delved into amalgamating these techniques using methods such as hybrid data mining algorithms. To provide an approximate approach for the heart disease prediction, this paper has considered a rule-based model to compare the performance of applying regulations to the individual effects of classification and regression machine learning techniques. This work will follow these steps in order to develop a heart disease classifier. The well-known UCI heart disease dataset is used to develop the heart disease classifier. The purpose of this study is to classify the most important cardiac disease predictors and use classification and regression algorithms to predict the overall risks [1]. The classification algorithms in machine learning are thus used to classify the predictors [2]. Furthermore, to verify models,

S. Kanrar (✉) · S. Shit · S. Chakrarbarti
Amity University Jharkhand, Ranchi, India
e-mail: dr.soumen.kanrar@gmail.com

S. Kumar et al. (eds.), *Fourth Congress on Intelligent Systems*, Lecture Notes in Networks and Systems 868, https://doi.org/10.1007/978-981-99-9037-5_1

researchers conduct data analysis in Python using Jupyter Notebook [3]. Health policies and programs are frequently recommended based on the doctor's instructions rather than knowledge-rich data. Sometimes, this results in errors and high expenses, both of which have an impact on the quality of medical treatments. Clinical choices can be improved by using analytic tools and data modelling. As a result, the goal is to create a web application that will guide doctors in diagnosing cardiac disorders [4]. The main advantage is that if the doctors can make a proper diagnosis at the exact time, they can provide proper treatment at a considerable cost. This paper presents the exhibitor's analysis of the predictions of heart disease using algorithms for classification. These secret trends in medical data can be used for health diagnosis. Healthcare management may use the knowledge to get better treatment. For big data concerns and data science projects, the data science lifecycle has to be constructed first. Problem formulation, data collection, data preparation, data exploration, data modelling, model evaluation, and model deployment are the seven major steps normally used in data science. The human heart beats one hundred thousand times every day, pumping 7570.824 L of blood into the body. There are 96,560.64 km of blood veins inside the body. Men are at higher risk of heart attack than women, based on the basic symptoms. Generally, women may feel squeezing, pressure, fullness, or discomfort in the middle of their chest during a heart attack. It can also cause arm, back, neck, jaw, or stomach pain, as well as shortness of breath, nausea, and other symptoms. The basic symptoms observed for a heart attack in humans are chest pain, filling uneasiness, and mental stress. Sometimes, it is observed that humans have pain in their hands, arms, and shoulders, as well as in their neck, back, and jaw, as well as shortness of breath. To make the classification process more effective, the heart disease database is pre-processed. Classification and regression categories are the pre-processed data. In order to diagnose heart diseases early, it is important that health complications be reduced. In the current healthcare industry, machine learning has been commonly used to diagnose and predict the existence of diseases using data models. One of these relatively common machine learning algorithms for studies involving risk assessment is the classification technique for complex illnesses. Therefore, the research aims to classify the most relevant cardiovascular disease predictors and predict the overall risk using classification algorithms. On the Kaggle website, the dataset used for support vector machine, random forest, AdaBoost, and gradient boost analysis is available for various projects. To develop a classifier for the early stage of heart disease prediction, the researcher must collect and pre-process the data as prerequisites at the initial steps. Researchers used the UCI heart disease dataset for this purpose. The aim is to decide if the patient has a risk of future heart disease. The UCI dataset consists of 442 patient data records and 14 attributes. We have conducted the test through the Python program, which is run in a Jupyter Notebook. The Jupyter Notebook is equipped with a very powerful data science software package. For the last two decades, predicting heart attacks using data mining methods has gained importance. Much of the article has applied techniques to various patient datasets from around the world, such as SVM, neural networks, regression, decision trees, Naive Bayesian classifiers. The multiple regression model was also suggested in the research paper to forecast heart disease [5].

Maarten et al. (2022) addressed the issue of artificial intelligence needed to solve the targeted medical problem of cardiovascular disease [6]. It has been observed from the research papers that various linear regression models present the perfect conditions for predicting the chance of heart disease. The classification accuracy of the regression algorithm is good compared to other algorithms [7]. S. Das et al. have proposed the local discovery method of frequent disease prediction in the medical dataset [4]. S. Das et al. focus on examining information mining processes that are needed to classify for medical information mining, such as heart disease, lung malignancy, and bosom disease. The prediction algorithm used Naive Bayes analytic model, which is developed based on Baye's theorem [8]. Therefore, Naive Bayes could make independent assumptions [9]. Huang et al. have applied artificial intelligence to Wearable Sensor Data to identify and forecast cardiovascular disease [10]. V. Jackins et al. used a random forest classifier and Naive Bayes in the efficient prediction of clinical disease in the heart [11]. Big data tools such as Hadoop Distributed File System, MapReduce, and support vector machine are also considered for heart disease prediction [12]. Maarten et al. used the concept of data mining to forecast heart disease [6]. It is recommended to use HDFS on various nodes for storing a large volume of data and performing the prediction algorithm using SVM simultaneously. Concurrently, these results are obtained in faster computation time compared to other machine learning models. Another repository from which we can collect the sourced data is one of the leading institutes for diabetic research in Chennai, Tamil Nadu, India [13]. The dataset consists of more than 500 patients' data. In another dataset, the instrument implemented in Weka considers a seventy percent break for classification. Here, 86.419% of the data is the precision provided by Naive Bayes. Neurological computing is a little bit helpful for analysing the electric signal from the heart to determine the cardiovascular issue [14]. Ghrabat et al. [15] discussed the use of a nonlinear classification algorithm for the estimation of heart disease [16]. On the Kaggle website, the dataset uses support vector machine, random forest, AdaBoost, and gradient boost analysis, which are available from ongoing Kaggle projects. The aim of our study is to determine if the patient has a risk of future heart disease. We have conducted the study considering a training dataset consisting of 3000 cases with thirteen unique attributes. The dataset is separated into two clusters. One cluster contains seventy percent of the data that is used for training purposes, and another cluster contains thirty percent of the data that is considered for testing purposes.

2 Materials and Methods

In the body, the heart is like any other muscle. For the muscle to contract and pump blood to the rest of the body, it requires enough blood flow to deliver oxygen. The heart pumps blood into the coronary arteries. Arteries derive from the aorta (the largest artery of the body, which contains the heart's oxygenated blood) and then spread out along the heart wall. Table 1 presents the different types of heart disease with their common symptoms and causes. If humans change certain habits in their lives, they can reduce the chance of heart disease.

These include:

- Consume fibre-rich, nutritious diets.
- Regular exercise: These will help to improve the heart condition and circulatory system, decrease cholesterol, and maintain blood pressure.
- Control the body weight with respect to height.
- Control alcohol consumption: Per day may be consumed according to the prescription of the doctor.
- Managing underlying issues: Finding care for problems such as elevated blood pressure, obesity, and diabetes that impair heart well-being.
- Quitting or avoiding smoking: Smoke is an important risk factor for cardiovascular and heart disease.

Machine learning is commonly used in nearly all sectors of the healthcare industry around the world. ML technology helps to adopt new formations from experience. In comparison, the process of considering various methodologies to predict, forecast, and explore something in the universe based on machine learning is relatively simple [17]. Sometimes the complex heart disease problem can be predicted by machine learning, i.e. regression and classification. Regression algorithms are primarily used for numeric results and binary and multi-category problems. Algorithms for machine learning were further split into two groups, such as supervised learning and unsupervised learning. Supervised learning is basically carried out using prior knowledge of output values. Unsupervised learning does not possess fixed tags; consequently, it infers the spontaneous build-up in the dataset. Therefore, machine learning algorithm selection needs to be properly analysed. AdaBoost is an ensemble strategy in which a group of weak learners are joined to generate a powerful learner. The AdaBoost model is present in Fig. 1. Typically, each weak learner is created as a "decision stump" (a tree with only one major split). In the random forest procedure, different

Table 1 Heart disease

S. No.	Disease types		
	Heart disease types	Symptoms	Reason
1	Coronary artery disease	Chest pain or discomfort (angina)	Plaque build-up in the arteries
2	Congenital heart defects	Rapid heartbeat and breathing	Several genetic health conditions
3	Arrhythmia	A fluttering in the chest and a racing heartbeat	Low supply of oxygen and nutrients to the heart
4	Heart failure	Out of breath. Feel tired always. Swollen ankles and legs	High blood pressure and high cholesterol
5	Myocardial infarction	Chest pain, fatigue, and heartburn	Smoking and high cholesterol

weights are assigned to increase efficiency. When we need high accuracy, we need to assign more weight. Tree is created from the consecutive training set. AdaBoost can improve its efficiency by anticipating the updated weights. Gradient Boosting Machine (GBM), like AdaBoost, combines a number of weak learners to generate a powerful learner. The classifier's residual is used as the input for the next classifier, on which the trees are created, making this an additive model. The residuals are acquired by the classifiers in a step-by-step manner to bag major variation in the data, which is accomplished through massive learning. The Gradient Boosting Machine is present in Fig. 2. It is gradually moving in the correct direction towards improved prediction using this strategy. As a result, based on the number of classifiers, we arrive at a prediction value that is very close to the observed value. Initially, a single-node tree is constructed that predicts the aggregated value of Y in the case of regression or the log (odds) of Y in the case of classification issues, following which trees with greater depth are developed on the residuals of the preceding classifier. Olsen et al. [18] provide a good overview of machine learning procedures for the prediction of heart failure [18]. Joon-myoung et al. (July 2019) used a deep learning-based artificial intelligence algorithm to predict acute heart failure [19]. Akbilgic et al. [20] developed an electrocardiographic artificial intelligence model to predict heart failure [20]. They have used the electrocardiogram artificial intelligence model for the prediction of heart failure. Manimurugan, S., et al. used the Two-Stage Classification model for the identification of heart disease [21].

In GBM, each tree has 8–32 terminal nodes, and approximately, 100 trees are produced. For each tree, learning rates are set to a constant so that the model may take modest steps in the appropriate direction to capture variance and train the classifier. Unlike in AdaBoost, where all trees are given equal weights. The dataset used

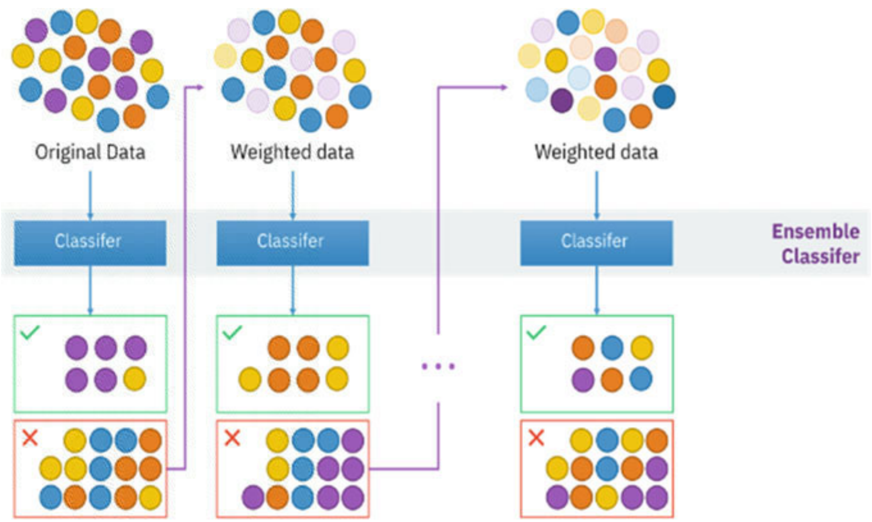

Fig. 1 AdaBoost model

Fig. 2 Gradient boosting machine model

in the study of classification and regression algorithms is from an ongoing UCI cardiovascular study and is available on the Kaggle website.

The object of this analysis for classification is to predict that the patient has a chance of potential heart failure from the infections. With 442 pieces of information, the Framingham dataset consists of patient data and 14 attributes. Reviewing the Framingham data set and using Jupyter Notebook in Python programming, which is a more flexible and efficient technology tool for data science. Figure 3 indicates the steps followed to build the classification and regression models in machine learning.

In the data pre-processing stage, the initial process involves cleaning the data first to remove noise, identify inconsistent data, and obtain missing data. The initial dataset attributes are presented in Table 2. The regression model is created using the UCI Machine Learning Repository.

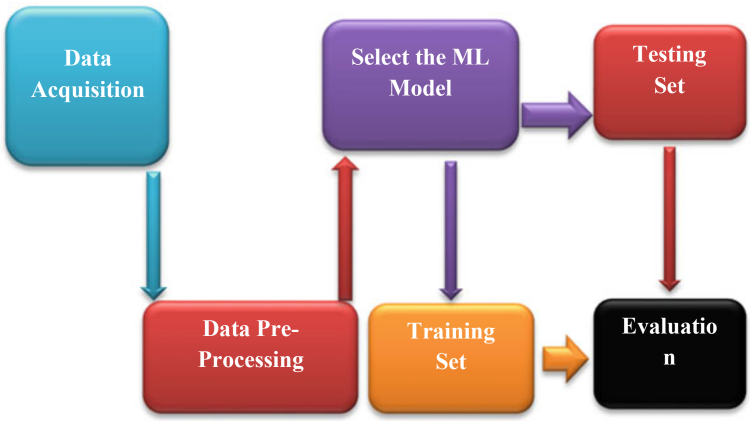

Fig. 3 Workflow of classification and regression model

Table 2 Dataset attributes

Attribute name	Attribute description, type, value			
	Description	Type/value		
Age	Age in years	Integer		
CP	Chest pain	Common angina		Value = 0
		Anomalous angina		Value = 1
		Non-angina pain		Value = 2
		No symptoms		Value = 3
trestbps	Resting blood pressure	In mm Hg		
chol	Serum cholesterol	In mg/dl		
fbs	Fasting blood sugar	More than 120 mg/dl	True	Value = 1
			False	Value = 0
restecg	Resting electrocardiographic	Normal		Value = 0
		ST-T wave abnormality		Value = 1
		Left ventricular hypertrophy		Value = 2
thalach	Heart rate	Heart rate/minimum		
Exang	Angina	Yes		Value = 1
		No		Value = 0
oldpeak	Depression	Exercise relative to rest		
slope	Slope of the peak exercise	Upsloping		Value = 0
		Flat		Value = 1
		Downsloping		Value = 2
ca	Number of major vessels	Coloured by fluoroscopy		Value range (0–3)
thal	Thalassemia	Normal		Value = 3
		Fixed defect		Value = 6
		Reversible defect		Value = 7
Gender	Male	Value = 1		
	Female	Value = 0		

3 Data Handle

The regression model is created using the UCI Machine Learning Repository. We have developed Python language code and tested it on the Jupiter notebook platform after removing some unnecessary fields from the collected data. In addition, the number of missed values for cleaning up current databases has been identified. A sample snapshot of the data header is presented in Table 3. The dataset is error-free and contains all of the data required for each variable. There are no issues regarding missing values or inconsistencies that are identified using the info (), describe (), and isnull () functions as defined in the Pandas Library. As a result, the dataset is fairly balanced. From the collected data, the proposed techniques approximate the (0.56) probability portion of the human population suffering from heart disease in the sample dataset presented in Table 3. Figure 4 presents the heart disease of the people with respect to the small sample dataset present in Table 3.

To build the attribute correlation, we consider the heatmap, which depicts the relationships between the dataset's attributes as well as how they interact. Attribute correlation is presented in Fig. 5.

Generally, from the observed prediction, we draw the inference that older people suffer from heart disease. The heart rate becomes lower at old age as he or she suffers from heart disease. The correlation between age and heart rate is presented in Fig. 6. The heatmap shows that the types of chest pain (CP), exercise-induced angina (Exang), exercise-induced ST depression, atypical, and non-anginal are all strongly linked to heart disease (target). We also see that cardiac illness and maximum heart rate have an inverse relationship (thalamus). It is presented in Fig. 7.

Table 3 Sample dataset

	Age	Sex	CP	trestbps	chol	fbs	restecg	thalach	Exang	oldpeak	slope	ca	thal	Target
0	70	1	4	130	322	0	2	109	0	2.4	2	3	3	2
1	67	0	3	115	564	0	2	160	0	1.6	2	0	7	1
2	57	1	2	124	261	0	0	141	0	0.3	1	0	7	2
3	64	1	4	128	263	0	0	105	1	0.2	2	1	7	1
4	74	0	2	120	269	0	2	121	1	0.2	1	1	3	1

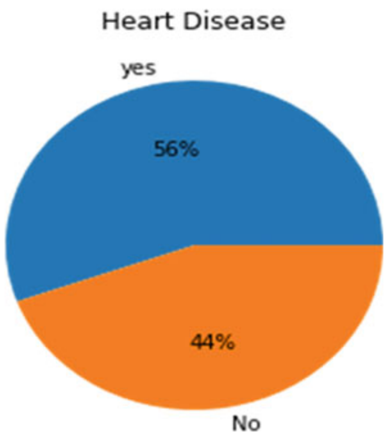

Fig. 4 Heart disease with respect to sample data

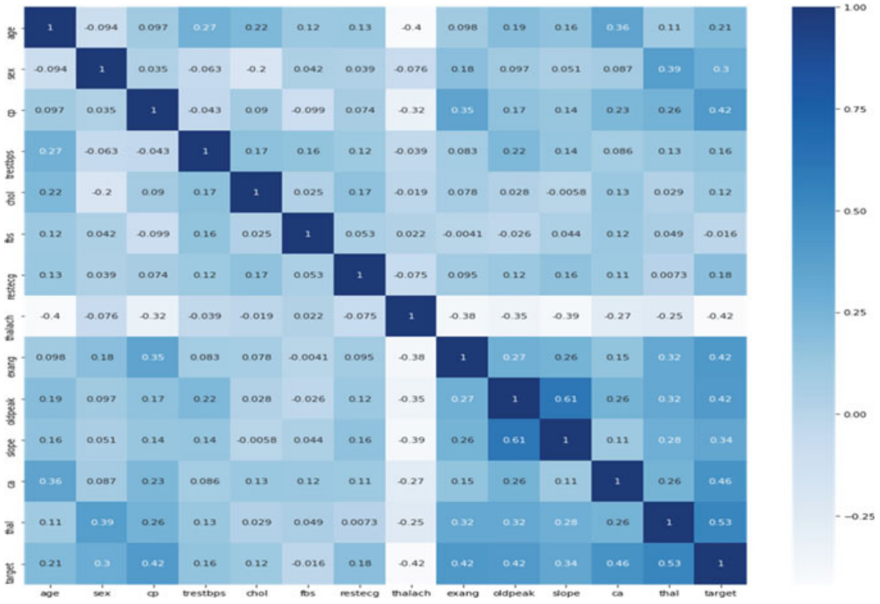

Fig. 5 Attributes correlation

Fig. 6 Correlation between age and heart rate

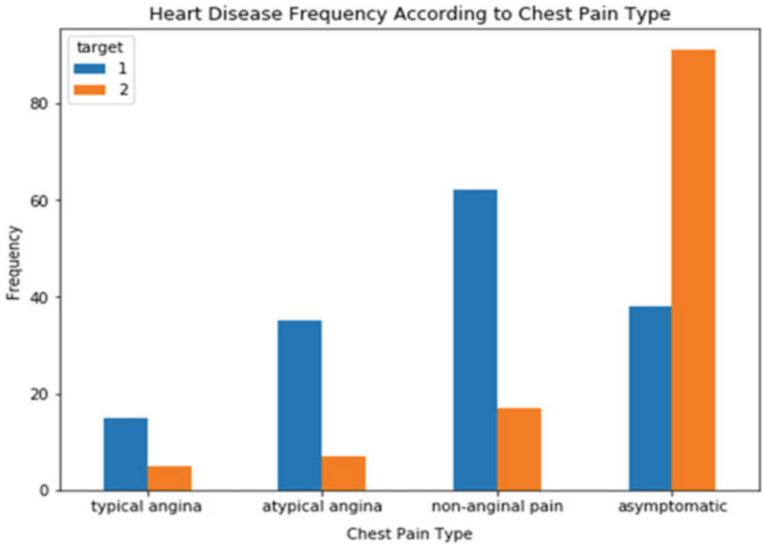

Fig. 7 Heart disease based on chest pain

4 Experiment and Results

From the result analysis, we get four different types of chest pain as the main cause of heart disease. The majority of heart disease patients have asymptomatic difficulty breathing.

The majority of patients with heart disease are elderly and have one or more main arteries coloured by Fluoroscopy. It is presented in Fig. 8.

We have logically subdivided the dataset. As some columns treat it as the "target" column, it is considered to serve as the class. A training set and a test set are created from the data. Seventy percent of the data is used for training purposes, and the remaining amount is used for testing purposes. Finally, the test results for the models for support vector machine, random forest, AdaBoost, and gradient boosting are exhibited in Figs. 9, 10, 11 and 12. Inference can be considered based on the best classification report.

Based on the aforementioned obtained results, with component-wise comparison based on precision, we can consider that gradient boosting is one of the best classifiers for heart disease prediction.

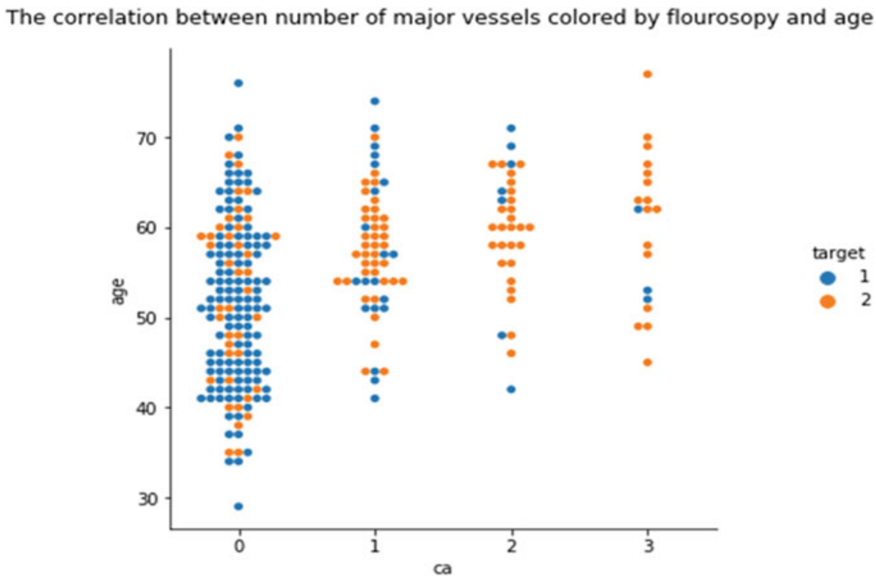

Fig. 8 Correlation between fluoroscopy and age

```
classification_report :
              precision    recall  f1-score   support

           1       0.71      0.90      0.79        30
           2       0.81      0.54      0.65        24

    accuracy                           0.74        54
   macro avg       0.76      0.72      0.72        54
weighted avg       0.76      0.74      0.73        54
```

Fig. 9 Snap of the support vector machine used in the dataset

```
RandomForestClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.76      0.73      0.75        30
           2       0.68      0.71      0.69        24

    accuracy                           0.72        54
   macro avg       0.72      0.72      0.72        54
weighted avg       0.72      0.72      0.72        54
```

Fig. 10 Snap of the random forest classifier used in the dataset

```
AdaBoostClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.75      0.80      0.77        30
           2       0.73      0.67      0.70        24

    accuracy                           0.74        54
   macro avg       0.74      0.73      0.73        54
weighted avg       0.74      0.74      0.74        54
```

Fig.11 Snap of the AdaBoost classifier used in the dataset

```
GradientBoostingClassifier

classification_report :
              precision    recall  f1-score   support

           1       0.79      0.73      0.76        30
           2       0.69      0.75      0.72        24

    accuracy                           0.74        54
   macro avg       0.74      0.74      0.74        54
weighted avg       0.74      0.74      0.74        54
```

Fig.12 Snap of the gradient boosting classifier used in the dataset

5 Conclusion

In conclusion, we learned how to build a suitable application for drawing predictions about heart diseases from medical datasets. We began with the definition of the problem and the moderate amount of accumulated data. Then we worked on data preparation, investigation, and building models. To create a classification model, we used a standardised dataset of heart disease patients. We began with data exploration and then moved on to data preparation. After that, we tested four heart disease classification models. Finally, we chose the finest model and saved it for later usage. The limitations of these works focus primarily on the use of classification strategies and algorithms for the prediction of heart disease. In this domain, researchers use different mechanisms for data cleaning and plan to construct a dataset that is suitable for their desired algorithms.