Classification and Ensemble Machine Learning Techniques to Improve Healthcare Decision Making For Heart Disease

Narendra Kumar Sharma[1], Dr. Alok Singh Chauhan[2*], Dr. Shahnaz Fatima[3], Dr. Osamah Ibrahim Khalaf[4], Dr. Swati Saxena[5], Dr. Sameer Algburi[6], Dr.Habib Hamam[7]

[1,3]Amity Institute of Information Technology, Amity University, Lucknow, Uttar Pradesh, India
[2]School of Computer Applications and Technology, Galgotias University, Greater Noida, Uttar Pradesh, India
[4]Department of Solar, Al-Nahrain Research Center for Renewable Energy, Al-Nahrain University, Jadriya, Baghdad, Iraq
[5]Department of Computer Application, Maharana Pratap Engineering College, Kanpur, Uttar Pradesh, India
[6] College of Engineering Techniques, Al-Kitab University, Iraq
[7] Uni de Moncton, NB, 1EA 3E9, Canada
*Correspondence: alok.chauhan@galgotiasuniversity.edu.in

Abstract

Cardiovascular disease is one of the main reasons for demise of people in the world today, whether it is a developed country or a developing country. It is not only affecting the people living in the urban but it has also affected the people of rural areas. If we know it at the primary stage, then its side effects can be avoided by reducing the chances of heart disease. So, correct prediction of heart disease is an imperative task to assist doctors and medical experts to take decision and make effective treatment policy to save the lives of people. In this paper, we use and combine multiple classification method of data mining and machine learning to perk up the precision of classifier. We intend an iterative ensemble approach to integrate various low-performance classifiers to form a strong classifier with high precision. We took dataset from IEEE data port for its implementation which contains around 1190 instances with 11 features of heart disease. We examine on the basis of initial symptoms whether the patient has heart disease or not. We explore the application of classification and ensemble machine learning techniques to augment healthcare decision-making for heart disease. By bridging the gap between data-driven insights and clinical decision-making, these techniques pave the way for a more proactive and patient-centric approach to cardiovascular health management.

Keywords: Healthcare, Heart Disease, Decision Making, Data Mining, Machine Learning, Ensemble Classifier

1. Introduction

Heart disease remains one of the leading causes of mortality globally, posing significant challenges to healthcare systems and clinicians. With the ever-growing volume of patient data and the complexity of heart disease diagnosis and treatment, there is a pressing need for advanced computational methods to aid in decision making. Machine learning (ML) techniques have emerged as powerful tools capable of extracting valuable insights from large datasets, thus revolutionizing healthcare decision making.

Heart disease is one of the foremost reasons of death as non-communicable diseases in the humankind today. According to the WHO, about 17.9 million people died of heart disease in 2019, which accounts for a possible 32% of all deaths worldwide. It is anticipated that 24% of the casualty owed to un-communicable ailment in India are due to heart disease. Cardiovascular disease prediction is a significant confront in the field of medical science and healthcare decision making. Machine learning is one such medium that can effectively facilitate in healthcare forecasting and decision making from the huge amount of data formed by the health sector [9]. Data mining or machine learning is unearthing method to analyze and encapsulate large data from a mixed approach to discover important decision [21-26]. Data mining provides wide range of disciplines including computational theory, pattern recognition, artificial neural network, genetic algorithms, statistics, data analytics and probabilistic modelling [10][27-33].

Heart disease continues to be a significant global health challenge, demanding innovative solutions to improve diagnostic accuracy and treatment outcomes. In this paper, we present a novel approach to enhance healthcare decision making for heart disease by leveraging four distinct classification algorithms: decision trees, k-nearest neighbors, logistic regression, and random forests. Additionally, we propose an iterative ensemble technique designed to boost precision and reliability in risk prediction and diagnosis.

The prevalence of heart disease underscores the urgency for more effective diagnostic tools and treatment strategies. Despite advances in medical technology, accurately assessing individual risk factors and determining optimal interventions remain complex tasks. Machine learning algorithms offer a promising avenue for extracting actionable insights from diverse patient datasets, potentially revolutionizing clinical decision making in cardiovascular healthcare. Our research focuses on harnessing the strengths of four classification algorithms: decision trees, k-nearest neighbors (KNN), logistic regression, and random forests. Decision trees provide intuitive interpretability, allowing clinicians to understand the underlying decision-making process. K-nearest neighbors excel in identifying patterns within data and are particularly useful for clustering similar cases. Logistic regression offers a straightforward probabilistic framework for binary classification tasks. Random forests, as an ensemble of decision trees, combine the predictive power of multiple models to improve accuracy and robustness.

In addition to these individual classifiers, we introduce a novel iterative ensemble approach tailored specifically for heart disease diagnosis. This iterative ensemble method iteratively combines predictions from multiple base classifiers, iteratively refining the ensemble to optimize precision while maintaining high recall. By dynamically adjusting the contribution of each classifier based on its performance, our approach adapts to the complexity and variability of heart disease data, enhancing diagnostic accuracy and clinical utility.Throughout this paper, we detail the implementation of each classification algorithm and describe our iterative ensemble approach in depth. We present experimental results demonstrating the efficacy of our method compared to traditional standalone classifiers and existing ensemble techniques. Moreover, we discuss the practical implications of our approach for real-world healthcare decision making, emphasizing its potential to improve patient outcomes and reduce healthcare costs.

In conclusion, our research represents a significant advancement in leveraging machine learning for heart disease management. By integrating multiple classification algorithms and introducing an innovative iterative ensemble technique, we offer a comprehensive framework for enhancing diagnostic precision and clinical decision making in cardiovascular healthcare. Our findings pave the way for future research and implementation of data-driven approaches to mitigate the burden of heart disease on individuals and healthcare systems worldwide.

1.1 Data Mining and Machine Learning in Healthcare Decision Making

Data mining (DM) and machine learning (ML) both are very useful technique in healthcare and medical field. These techniques are very helpful in accurate prediction and quick diagnosis of diseases [34-38]. These are very useful to predict and analyze the particular health condition of patient and guide them to medical professional and decision maker to take effective decision regarding particular medical condition of patient and their illness [39-41]. Hence, these techniques save a lot of time of doctors and medical experts and can be helpful in correct treatment by working out individual imprecision. [11].

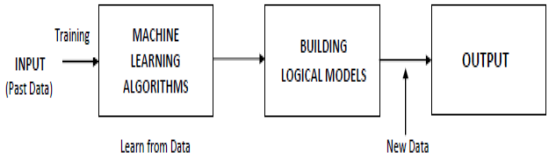

Figure 1. Machine Learning Process

1.2 Ensemble Learning in Decision Forecasting

Ensemble learning is machine learning technique which combines multiple models into single predictive model. Ensemble learning utilizes the advantages of multiple base models usually called weak learners to compensate each models weakness.

The main principle behind ensemble learning is to group weak learners together to form on strong learner that achieves better performance than any individual weak learner [8]. This approach permits the better predictive ability as contrast to single model.

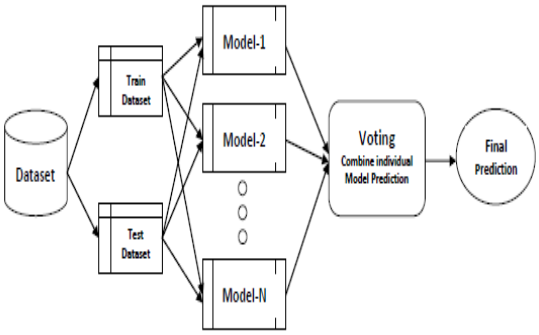

Figure 2. Ensemble Learning

Eventually, Ensemble methods are meta-algorithms that combine multiple ML technique into one predictive model to increase the performance of model. The advance ensemble learning method possibly will use bagging to diminish change, boosting to predisposition and stacking approach for improve prediction. We can categorize ensemble learning models into two category based on the choice of weak learners. Homogeneous ensemble model utilize single base learning and heterogeneous ensemble model use different base learning algorithms [10].

2. Problem Definition and Motivation

There is an invaluable endowment of healthy life, but in today's busy life, we are all affected by it. There are many such diseases but among them heart disease and diabetes are mainly found in today's people. These diseases are not only affecting the people living in the urban but they have also affected the people of rural areas. In such a situation, we need a kind of health system which can reduce the risks of heart disease by making accurate and correct predictions. Thus, accurate prediction of heart disease is an imperative task to assist doctors and medical experts to take effective decision and make valuable treatment policy to save the lives of people.

3. Related Work

H.S. Niranjana Murthy et al. [1] explore and employ different data mining approaches along with machine learning technique on diabetes and analyze the performance of ML technique. The study concern with SVM model for diabetes prediction and identification. Satish Ranbhise et al. [2] introduce a wireless based health monitoring system using data mining and IoT that can monitor the health status of patient and predict the heart disease based on symptoms. Basma Saleh et al. [3] worked on cardiovascular disease and predict the heart disease using data extraction algorithms and data mining to identify the main risk factor involves in heart disease. They explore the integration of different data extraction algorithms through data framework to predict heart disease. H Benjamin Fredrick David et al. [4] present the comparative analysis on ensemble learning algorithms that explores the future of medications and state of patient and identifying the crucial risk involve in heart disease.

Ahmad Mousa Altamimi et al. [5] worked on children diabetic dataset and predict the diabetic condition in children using different data mining algorithms to analyze the performance and make early prediction on children diabetic condition. Sinkon Nayak et al. [6] emphasis on the heart disease diagnosis at premature phase and used data mining classification and frequent item mining approach to filter the attribute and determine the heart disease at primitive stage. Anjan Nikhil Repaka et al. [7] carries out heart disease forecasting by using naïve bayesian and complex encryption system to find the risk factor involve in heart disease and designs a smart heart disease prediction based on mobile health technology. Rahma Atallah et al. [8] use ensemble technique to mingle several machine learning methods on the way to envisage the heart disease. They determine the majority voting accomplished by ensemble method and predict the probable heart disease in peoples.

This review explores the application of classification and ensemble machine learning techniques in improving healthcare decision making for heart disease. By leveraging these techniques, clinicians can enhance risk prediction, diagnosis, prognosis, and treatment planning, ultimately leading to better patient outcomes.

4. Proposed Approach

In proposed approach, manifold classifiers are promoted to increase the precision of classifier. We intend an iterative ensemble approach to integrate various low-performance classifiers to form a strong classifier with high precision. The main insight of this proposal is to situate classifier weights and prepare a test of the data at every iteration that produce correct prediction from heterogeneous interpretation. Any ML approach are often used as a base classifier as elongated because it accepts the instruction set weights. The proposed approach is anticipated to meet the subsequent conditions:

i. The classifier training should be carried out interactively with different weighted learning examples.
ii. At every iteration, the classifier attempts to render the finest for those examples at the same time as minimizing getting to know errors.

4.1 Working Steps

i. Primarily, a training separation is selected randomly.
ii. Iteratively train the proposed ML model through deciding on training set mainly pedestal on the accurate predictions of final training.
iii. Give higher significance to missorted instances in order to the ones instances are much more probable to be categorized in a next recurrence.
iv. Based at the accuracy of the classifier, additionally allot weights to the classifier trained on each iteration. The higher classifier receives greater weight.
v. We replicate the same process until the whole training data is matched with no miscalculation otherwise until the maximum number of evaluators individual is achieved.
vi. Toward rank, you 'vote' for each of the learning algorithms that you create.

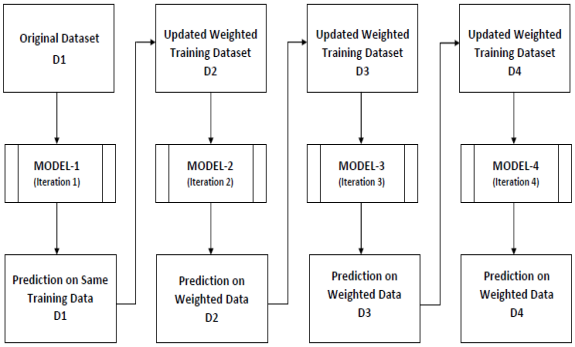

Figure 3. Proposed Iterative Approach

Algorithm of proposed approach is shown below:

Input

Dataset D = { (x_1, y_1), (x_2, y_2)…., (x_M, y_M) };
L = {Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), KNN}
T = 4 (#Number of learning round)

Algorithm

Step 1: Initialize the weight distribution

$$D_1 = \frac{1}{M}$$

Step 2: for t= 1, 2..….T

- Train a base learner ht from D using Dt
  Ht = L (D, Dt)
- Calculate Error which is nothing, but the summation of all the sample weights of misclassified data points.

$$e_t = \sum_{i=1}^{M} D_t(i)[h_t(x_i) \neq y_i)]$$

- Determine the weight of h_t

$$\alpha_t = \frac{1}{2}\ln\frac{1-e_t}{e_t}$$

- Calculate normalization factor that enables D_{t+1} to be a distribution

$$Z_t = \sum_{i=1}^{m} D_t(i) \times \begin{cases} e^{-\alpha t} & if\ h_t(x_i) = y_i \\ e^{\alpha t} & if\ h_t(x_i) \neq y_i \end{cases}$$

- Update the distribution

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha t} & if\ h_t(x_i) = y_i \\ e^{\alpha t} & if\ h_t(x_i) \neq y_i \end{cases}$$

Step 3: End

Output

$$H(x) = sign \sum_{t=1}^{T} \alpha_t\, h_t(x)$$

Proposed method transforms “weak” individual models in a “strong” ensemble of models.

5. Methodology

The main intention of this research work is to successfully predict probable heart disease problems from medical dataset. A model is established using a prediction technique to access the heart disease facility by certain characteristics. Ensemble learning is used to construct class prediction based on selected features in this work to combine multiple classification approach to perk up the precision of classifier. In this research work, dataset is obtained from IEEE data port which contains around 1190 instances with 11 features of heart disease. We implemented proposed approach in Python with various libraries like pandas, numpy, matplotlib, seaborn, scikit-learn etc. The work execution is represented within following flow diagram [fig. 4].

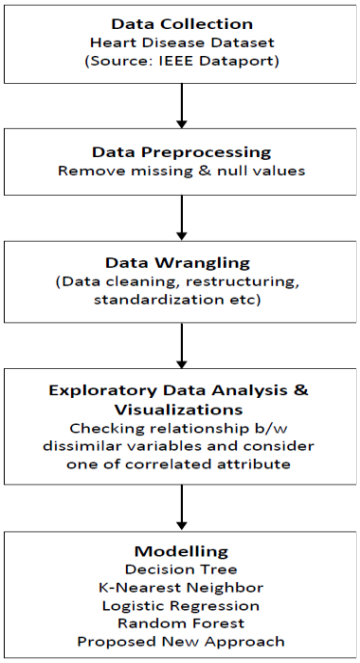

Figure 4. Flow Diagram of Execution

5.1 Data Set

In this research work, dataset is obtained from IEEE data port which contains around 1190 instances with 11 features of heart disease patient and their health condition. The dataset are combined with five popular heart disease dataset resembling Cleveland, Hungarian, Switzerland, Long Beach VA and Statlog dataset which makes it prevalent and comprehensive heart disease dataset. The attribute include in dataset are age, sex, chest_pain_type, resting_blood_pressure, serum_cholesterol, fasting_blood_sugar, resting_electrocardiogram_results, max_heart_rate achieved, exercise_induced_angina, oldpeak_ST,the slope_of_the_peak_exercise_ST_segment etc.

5.2 Data Preprocessing

In this process we have to check that dataset which have any missing or null value are available or not. This is a significant step when building a machine learning model. After the preprocessing of dataset no null value found.

5.3 Data Wrangling

Data wrangling is a process which provides the facility to restructure, clean, enrich, discover, publish and validate the available raw data into a more standard format. It is also known as data munging. We perform data wrangling in python with different libraries numpy, pandas, matplotlib, plotly etc.

5.4 Exploratory Data Analysis

We carry out exploratory data analysis toward obtain vital information about the data. To understand the data examine and relationship between different variables, we ensure that there is any strong relationship among two variables or not, after that we can consider one of them. After examine the dataset we find that there is no predicament of multi-alignment.

5.5 Modelling

A model represents what has been learned by a machine learning algorithm. In our experiment, we employ Decision Tree, K-nearest neighbor, Logistic Regression, Random Forest and proposed new approach. We considered and put into subsequent features of dataset:
Sex 1 = male, 0= female; Chest Pain Type -- Value 1: typical angina, -- Value 2: atypical angina,-- Value 3: non-anginal pain,-- Value 4: asymptomatic; Fasting Blood sugar (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false); Resting electrocardiogram results -- Value 0: normal, -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria; Exercise induced angina; 1 = yes, 0 = no; the slope of the peak exercise ST segment -- Value 1: upsloping, -- Value 2: flat, -- Value 3: downsloping and class 1 = heart disease, 0 = Normal.
The method was trained to recognize the target class from heart disease features and evaluate the performance metrics of model as accuracy, precision, recall, f1-score, support, macro and weighted average. The performance metrics used for the research work is described as:

Accuracy: In this, we predict how many times the classifier is correct in total and determined as:

$$Accuracy = \frac{Number\_of\_Correct\_Predictions}{Total\_Predictions}$$

Accuracy can be computed as TP, TN, FP and FN [13]

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

Where,
TP: These are the cases in which we predicted yes who have the heart disease

TN: In this we predicted whether not, and they do not have this disease.
FP: In this we predicted yes, but they do not actually have the heart disease.
FN: We predicted that they did not, but they actually have the disease.

Precision: In this, we predict yes and check how many times is it correct.

$$\Pr ecision = \frac{TP}{FP + TP}$$

Recall: It is defined out of the total number of positive classes that our model predicted correctly. The recall should be as high as possible.

$$\text{Re}\, call = \frac{TP}{FN + TP}$$

F1 Score: In this, recall and precision average weighted are calculated as:

$$\text{F1 score} = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$$

Macro-average: It is analyzed for each performance and calculates separately.
Weighted-Average: This is the weighted average and estimated as F1 score as follows:

$$\text{weighted average (F1 Score)} = \frac{\sum_{i=1}^{n} F_i \times N_i}{N}$$

Where,

F_i = F1 Score for i^{th} class,

N_i = calculation of instances in i^{th} class,

N = entire calculation of occurrences and n is calculation of classes in dataset.

6. Experimental Evaluation

Here we evaluate the performance of manifold machine learning models based on the chosen input dataset. The method of evaluation mainly focuses on the accuracy of the model during predicting the final results. To employing ML and the proposed approach, we must sure that data is suited as follows:

1. First perform cleaning on data,
2. Converting null values in integer value by 0,
3. Replace entity data by unknown,
4. Substitute 0 and unknown,
5. Perform encoding on data,
6. Leave unrelated feature,
7. Perform normalization on data,
8. Split data into training and validation datasets,
9. The training data contain 80% and validation contains remaining 20% of data.

Analyses of machine learning method are performed by Python tool with various libraries numpy, pandas, scikit-learn, seaborn, matplotlib, plotly etc. and explored categorical variable.

6.1 Result Analysis

We can examine the existing approach and obtained result of the proposed approach on the basis of its correctness. The performance of current approach and proposed can be tested through its accuracy, precision, recall, f1-score and support. The results are shown as follows [5]:

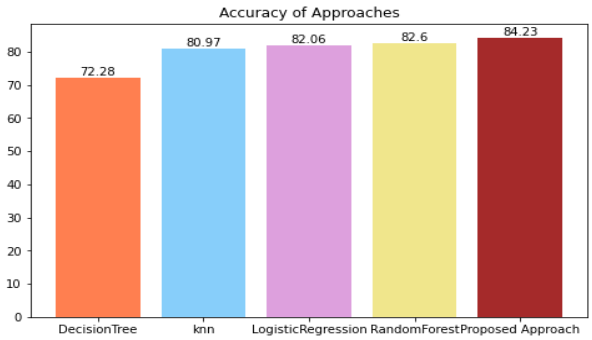

Figure 5. Accuracy of Approaches

The performance of all classifier is graphically represented by ROC (Receiver Operating Characteristics) curve in figure 6. An ROC curve is a graphical illustration that indicates the presentation of a classification model on every categorization thresholds. These curve plots on the basis of two parameters one is TP rate and other is FP rate [6].

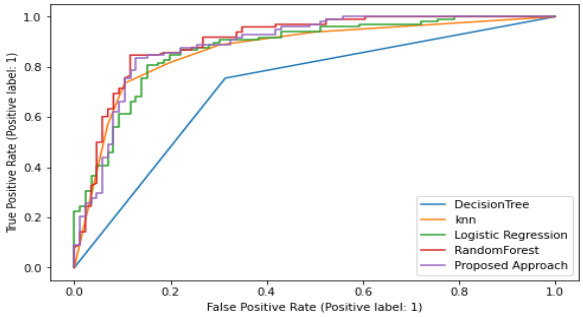

Figure 6. ROC curve of proposed and other approaches

The comparison of accuracy of existing approach and proposed approach is graphically represented in figure [5]. It clearly shows that the obtained accuracy after performing the experiment using machine learning tool python on heart disease. Proposed approach obtained highest 84.23% accuracy as compared to other approaches like random forest get 82.60%, logistic regression get 82.06%, K-nearest neighbor get 80.97% and decision tree get 72.28% of accuracy.

Thus, the end result shows that proposed approach will capable to assist and forecast the possibility involved in heart disease and effectively take decision to prevent from heart disease. After the experiment we found that 55.3% have heart disease and 44.7% does not have heart disease.

6.2 Discussion:

The results demonstrate the efficacy of the proposed iterative ensemble approach, which outperformed individual classifiers and traditional ensemble techniques in terms of predictive accuracy. Several key points emerge from these findings:

- **Precision of Predictive Models:** The superior performance of the proposed approach highlights the importance of leveraging ensemble techniques to enhance the precision of heart disease prediction. By iteratively refining the ensemble based on the performance of individual classifiers, our approach achieves higher accuracy levels compared to standalone models.
- **Clinical Utility:** The high accuracy of predictive models is of paramount importance in clinical decision making. Healthcare professionals can use these models to identify individuals at high risk of heart disease and intervene early with preventive measures or targeted treatments. The ability to accurately predict heart disease can lead to improved patient outcomes and reduced healthcare costs.

- **Comparative Analysis:** The comparative analysis provides valuable insights into the strengths and weaknesses of different classification approaches. While random forest, logistic regression, and k-nearest neighbor methods demonstrate respectable accuracy, the proposed iterative ensemble approach surpasses them all. Understanding the relative performance of these methods can guide researchers and clinicians in selecting appropriate algorithms for heart disease prediction tasks.
- **Dataset Distribution**: The distribution of individuals diagnosed with heart disease versus those without heart disease highlights the prevalence of the condition within the dataset. This distribution reflects the real-world scenario where heart disease remains a significant public health concern. Predictive models trained on such datasets need to accurately capture the complexities of heart disease risk factors and manifestations to yield clinically relevant predictions.
- **Future Directions:** Future research endeavors may focus on further optimizing predictive models, exploring alternative ensemble techniques, and incorporating additional data sources to improve accuracy and generalizability. Moreover, efforts to validate predictive models on diverse patient populations and evaluate their impact on clinical outcomes are essential for real-world deployment.

In conclusion, the results and discussion underscore the potential of machine learning approaches to enhance heart disease prediction and support clinical decision making. By leveraging advanced computational techniques and robust datasets, we can improve our ability to identify individuals at risk of heart disease and implement targeted interventions to reduce morbidity and mortality associated with this prevalent condition.

7. Conclusion

In the presented paper, we have attempted to correctly assess the future health challenges taking into relation the severity of heart disease. For which we have attempted to combine multiple data mining and machine learning methods to create a more correct and accurate predictive model, so that the probability of heart disease can be forecast correctly. The proposed approach uses ensemble learning technique that combines multiple machine learning classifiers to produce a strong classifier with high accuracy. The foremost proposal of this approach is to put the classifier weights and trained the sample data on every iteration thereby the accurate prediction can be achieved. After evaluating the accuracy with different present approaches, we can say that the precision of proposed approach is improved than others, which can be useful in prediction and diagnosis of heart disease. In summary, the utilization of classification and ensemble machine learning techniques represents a promising avenue for enhancing healthcare decision-making in the realm of heart disease. Through the comprehensive analysis of diverse datasets and the deployment of advanced algorithms, our research has demonstrated the potential of these methods to improve diagnostic accuracy and prognostic prediction. It can be further helpful for clinicians and medical experts to take decision and make effective treatment policy to save the lives of people.

Data availability

Data will be made available on reasonable request.

Funding

The authors thank Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project. These granting agencies did not contribute in the design of the study and collection, analysis, and interpretation of data.

Corresponding author

Correspondence to Dr. Alok Singh Chauhan alok.chauhan@galgotiasuniversity.edu.in

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] H.S. Niranjana Murthy, Early Prognosis of Diabetes Using Supervised Learning Techniques: A Comparison of Performance, Revista Geintec-Gestao Inovacao E Tecnologias, ISSN: 2237-0722, Vol. 11 No. 4 (2021)

[2] Abhishek Lakshman Singh, Ashok Yadav, Satish Ranbhise, Patient Health Monitoring using IoT and Disease Prediction using Data Mining, International Research Journal of Engineering and Technology, Volume: 08 Issue: 04, Apr 2021, ISSN: 2395-0056

[3] Basma Saleh, Ahmed Saeidi, Ali al-Aqbi, Lamees Salman, Analysis of Weka Data Mining Techniques for Heart Disease Prediction System, International Journal Of Medical Reviews, Int J Med Rev 2020 Jan;7(1): 15-24

[4] H Benjamin Fredrick David, Impact Of Ensemble Learning Algorithms Towards Accurate Heart Disease Prediction, ICTACT Journal On Soft Computing, April 2020, Volume: 10, Issue: 03, ISSN: 2229-6956, DOI: 10.21917/ijsc.2020.0296

[5] Ahmad Mousa Altamimi and Mohammad Hijjawi, Performance Analysis of Supervised Classifying Algorithms to Predict Diabetes in Children. Journal of Xi'an University of Architecture & Technology, Volume XII, Issue III, 2020, ISSN No : 1006-7930

[6] Sinkon Nayak, Manjusha Pandey, Mahendra Kumar Gourisaria Siddharth Swarup Rautaray, Prediction of Heart Disease by Mining Frequent Items and Classification Techniques, International Conference on Intelligent Computing and Control Systems, ISBN: 978-1-5386-8113-8, (ICICCS 2019) IEEE

[7] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, Design And Implementing Heart Disease Prediction Using Naives Bayesian, Third International Conference on Trends in Electronics and Informatics ISBN: 978-1-5386-9439-8, (ICOEI 2019) IEEE

[8] Rahma Atallah, Amjed Al-Mousa, Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method, 2nd International Conference on new Trends in Computing Sciences ISBN 978-1-7281-2882-5(ICTCS 2019) IEEE

[9] Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam, Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System, World Journal of Engineering and Technology, 2018, 6, 854-873, ISSN: 2331-4249

[10] Narendra Kumar Sharma, Shahnaz Fatima, Swati Saxena, Improvising Healthcare Decision Making By Employing Ensemble Technique, Journal of Mathematical and Computational Science 11 (2021), No. 4, ISSN: 1927-5307.

[11] Senthil kumar Mohan, Chandrasegar Thirumalai, And Gautam Srivastava, Effective Heart Disease Prediction using Hybrid Machine Learning Techniques, DOI: 10.1109/ACCESS.2019.2923707 (IEEE 2019), ISSN: 2169-3536

[12] Hasnain, M., Pasha, M.F., Ghani, I., Imran, M., Alzahrani, M.Y. and Budiarto, R., "Evaluating trust prediction and confusion matrix measures for web services ranking." IEEE Access, 8, pp.90847-90861, 2020.

[13] Patro, V.M. and Patra, M.R., Augmenting weighted average with confusion matrix to enhance classification accuracy. Transactions on Machine Learning and Artificial Intelligence, 2(4), pp.77-91, 2016.

[14] Ghiasi, M.M., Zendehboudi, S. and Mohsenipour, A.A., "Decision tree-based diagnosis of coronary artery disease: CART model." Computer methods and programs in biomedicine, 192, p.105400, 2020

[15] Bhandari, S., Shaktawat, A.S., Tak, A., Patel, B., Shukla, J., Singhal, S., Gupta, K., Gupta, J., Kakkar, S. and Dube, A., "Logistic regression analysis to predict mortality risk in COVID-19 patients from routine hematologic parameters." Ibnosina Journal of Medicine and Biomedical Sciences, 12(2), p.123, 2020.

[16] Iwendi, C., Bashir, A.K., Peshkar, A., Sujatha, R., Chatterjee, J.M., Pasupuleti, S., Mishra, R., Pillai, S. and Jo, O., "COVID-19 patient health prediction using boosted random forest algorithm." Frontiers in public health, 8, p.357, 2020.

[17] Janssens, A.C.J., "ROC curves for clinical prediction models part 2. The ROC plot: the picture that could be worth a 1000 words." Journal of Clinical Epidemiology, 126, pp.217-219, 2020.

[18] Sohaib Latif, Xian Wen Fang, Kaleem Arshid, Abdullah Almuhaimeed, Azhar Imran and Mansoor Alghamdi, Analysis of Birth Data using Ensemble Modeling Techniques, APPLIED ARTIFICIAL INTELLIGENCE 2023, VOL. 37, NO. 1, e-ISSN:2158273

[19] Sashikala Mishra, Kailash Shaw, Debahuti Mishra, Shruti Patil, Ketan Kotecha, Satish Kumar and Simi Bajaj, Improving the Accuracy of Ensemble Machine Learning Classification Models Using a Novel Bit-Fusion Algorithm for Healthcare AI Systems, Frontiers in Public Health, Vol. 10, 04 May 2022.

[20] R. Priyadarshini, Abdul Quadir Md, Senthilkumar Mohan, Abdullah Alghamdi, Mesfer Alrizq, Ummul Hanan Mohamad, Ali Ahmadian, Novel framework based on ensemble classification and secure feature extraction for COVID-19 critical health prediction, Engineering Applications of Artificial Intelligence, Volume 126, Part D, November 2023, 107156.
[21] Chauhan, A.S.; Lilhore, U.K.; Gupta, A.K.; Manoharan, P.; Garg, R.R.; Hajjej, F.; Keshta, I.; Raahemifar, K. Comparative Analysis of Supervised Machine and Deep Learning Algorithms for Kyphosis Disease Detection. Appl. Sci. 2023, 13, 5012. https://doi.org/10.3390/app13085012
[22] Chauhan, A. S.; Singh, J.; Kumar, S.; Saxena, N.; Gupta, M.; Verma, P. Design and Assessment of Improved Convolutional Neural Network Based Brain Tumor Segmentation and Classification System. J Integr Sci Technol 2024, 12, 793. https://doi.org/10.62110/sciencein.jist.2024.v12.793
[23] Naeem, A. B. .; Senapati, B. .; Chauhan, A. S. .; Kumar, S.; Orosco Gavilan, J. C. .; Abdel-Rehim, W. M. F. . Deep Learning Models for Cotton Leaf Disease Detection With VGG-16. Int J Intell Syst Appl Eng 2023, 11, 550-556.
[24] Mall, P. kumar .; Kumar, M. .; Kumar, A. .; Gupta, A. .; Srivastava, S. .; Narayan, V. .; Chauhan, A. S. .; Srivastava, A. P. . Self-Attentive CNN+BERT: An Approach for Analysis of Sentiment on Movie Reviews Using Word Embedding. Int J Intell Syst Appl Eng 2024, 12, 612.
[25] Novita, M., Chauhan, A. S., Ujianti, R. M. D., Marlina, D., Kusumo, H., Anwar, M. T., Piasecki, M., & Brik, M. G. (2024, May). Exploring deep learning and machine learning for novel red phosphor materials. Journal of Luminescence, 269, 120476. https://doi.org/10.1016/j.jlumin.2024.120476
[26] Naeem, A. B. .; Senapati, B. .; Chauhan, A. S. .; Makhija, M. .; Singh, A. .; Gupta, M. .; Tiwari, P. K. .; Abdel-Rehim, W. M. F. . Hypothyroidism Disease Diagnosis by Using Machine Learning Algorithms. Int J Intell Syst Appl Eng 2023, 11, 368-373.
[27] Y. Singh, A.S. Chauhan, Neural networks in data mining, J. Theor. Appl. Inf. Technol. 5 (1) (2009).
[28] A.S. Chauhan, Modeling and predicting student academic performance in higher education using data mining techniques, Int. J. Software Innovat. 10 (1) (2022).
[29] Chauhan, Alok Singh, et al. "Effective Decision Making in Higher Educational Institutions Using Data Warehousing and Data Mining." Journal of Computer Science Engineering and Information Technology Research (JCSEITR) 2.1 (2012).
[30] Chauhan, A. S. (2020). Healthcare decision making using k-means clustering technique. Pranjana: The Journal of Management Awareness, 23(2), 28-33.
[31] Singh, J. .; Fatima, S. .; Chauhan, A. S. . Design and Performance Study of Improved Fuzzy System With Genetic Algorithm. IJRITCC 2023, 11, 242-248.
[32] Singh, J. .; Fatima, S. .; Chauhan, A. S. . Multi-Objective Travel Route Optimization Using Non-Dominated Sorting Genetic Algorithm. Int J Intell Syst Appl Eng 2023, 11, 785-794.
[33] Latoria, A.; Chauhan, A.; Anand, S. Saxena Fuzzy Approach for Pattern Recognition Using Classification Algorithms. Int. J. Comput. Technol. 2012, 3, 458–462.
[34] M. Fatmawati et al., "Maternal Health Risk Classification - A Comparison between Algorithm Decision Tree and k-Nearest Neighbor (kNN)," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 507-511, doi: 10.1109/ICSEIET58677.2023.10303507.
[35] S. Saxena, R. Sharma and A. S. Chauhan, "Review of Brain Tumor MRI Detection Using CNN," 2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET), Ghaziabad, India, 2023, pp. 548-554, doi: 10.1109/ICSEIET58677.2023.10303561.
[36] Abdul Mustaan Madni, Awad Bin Naeem, Abdul Majid Soomro, Biswaranjan Senapati, Alok Singh Chauhan, Fridous Ayub, & Khuram Shahzad. (2023). Breast Cancer Diagnosis Comparative Machine Learning Analysis Algorithms. Journal of Computing & Biomedical Informatics, 4(02), 49–65.
[37] D. Anand, O. I. Khalaf, F. Hajjej, W. K.Wong, S. H. Pan, G. R. Chandra, "Optimized swarm enabled deep learning technique for bone tumor detection using histopathological image," SINERGI, vol.23, no. 3, 2023, pp. 451-466, doi:10.22441/sinergi.2023.3.016.
[38] Mangalampalli, S., Karri, G.R., Kumar, M. et al. DRLBTSA: Deep reinforcement learning based task-scheduling algorithm in cloud computing. Multimed Tools Appl 83, 8359–8387 (2024). https://doi.org/10.1007/s11042-023-16008-2
[39] Ghaida Muttashar Abdulsahib et al., " A Modified Bandwidth Prediction Algorithm for Wireless Sensor Networks," Journal of Information Science and Engineering, Vol. 40 No. 1, pp. 177-188.
[40] Xingsi Xue et al., "Soft computing approach on estimating the lateral confinement coefficient of CFRP veiled circular columns," in Alexandria Engineering Journal , Volume 81, 15 October 2023, Pages 599-619,doi.org/10.1016/j.aej.2023.09.053.
[41] S. Sadesh, O. I. Khalaf, M. Shorfuzzaman, A. Alsufyani, K. Sangeetha et al., "Automatic clustering of user behaviour profiles for web recommendation system," Intelligent Automation & Soft Computing, vol. 35, no.3, pp.3365–3384, 2023. https://doi.org/10.32604/iasc.2023.030751.