Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

PREDICTIVE MODELING OF HEART DISEASE RISK: A DATA MINING APPROACH TO CLASSIFICATION IN MACHINE LEARNING

Dr. POOJA SHAH

Associate Professor, VSITR, KSV. Email-poojaz.2608@gmail.com

Abstract

Cardiovascular diseases, including heart disease, remain a significant global health concern, necessitating early and accurate risk assessment for effective prevention and intervention strategies. In this research paper, we present a comprehensive study on predictive modeling of heart disease risk using a data mining approach coupled with classification techniques in machine learning. Our study revolves around the application of advanced data mining methods to extract valuable patterns and insights from a diverse dataset comprising medical attributes, lifestyle factors, and patient histories. Leveraging the power of classification algorithms, we aim to create predictive models capable of identifying individuals at risk of heart disease.

Keywords: Heart Disease Prediction, Data Mining Machine Learning, Risk Assessment, Healthcare Analytics

Key Components of Our Research Include:

Data Preprocessing and Feature Engineering: We detail the steps involved in data preprocessing, including handling missing values, normalization, and feature selection. Effective feature engineering strategies are explored to enhance the discriminative power of the models.

Classification Models: A range of machine learning classification algorithms, including Decision Trees, Random Forest, Support Vector Machines, and Neural Networks, are applied to predict heart disease risk. Model performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are employed to evaluate and compare the models.

Feature Importance and Interpretability: We emphasize the interpretability of our models, showcasing the identification of crucial features contributing to heart disease risk prediction. This feature importance analysis aids in both medical understanding and model transparency.

Cross-Validation and Generalization: Rigorous cross-validation techniques are employed to ensure the robustness and generalizability of our predictive models across diverse patient populations.

Clinical Implications: The research findings are discussed in the context of clinical practice, highlighting the potential utility of the developed models for early risk assessment, patient stratification, and personalized healthcare interventions. Our study underscores the significance of data mining and machine learning in augmenting the accuracy and efficiency of heart disease risk prediction. The integration of data-driven insights with clinical knowledge paves the way for more informed decision-making in healthcare. Ultimately, this research contributes to the ongoing efforts to reduce the burden of heart disease and improve the quality of cardiovascular care.

1. INTRODUCTION

Heart disease, a collective term encompassing various cardiovascular conditions, remains a critical global health concern. According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of mortality worldwide, accounting for an estimated 17.9 million deaths each year, nearly one-third of all global deaths [World

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

Health Organization, 2021]. In this context, accurate risk assessment and early detection of heart disease have become paramount in the field of healthcare.

1.1 Background

Traditional methods of heart disease risk assessment predominantly rely on clinical evaluation, medical history, and diagnostic tests. These approaches, while valuable, often face limitations in their ability to provide nuanced and timely risk predictions. The complexity of risk factors, including genetic predisposition, lifestyle choices, and medical histories, underscores the need for more advanced and data-driven approaches.

Machine learning, a subset of artificial intelligence, has gained prominence as a potent tool for predictive modeling and pattern recognition in healthcare. The intersection of data mining and machine learning offers a promising avenue to enhance heart disease risk assessment through the development of accurate predictive models.

1.2 Problem Statement

The challenge at hand is to effectively leverage data mining and machine learning techniques to construct predictive models for heart disease risk assessment that are both highly accurate and interpretable. While several studies have explored predictive modeling for heart disease, there remains a gap in comprehensive research that combines advanced classification algorithms, extensive feature engineering, and rigorous evaluation methodologies to create robust and interpretable models.

1.3 Objectives

This research paper seeks to address the following key objectives:

To develop predictive models for heart disease risk assessment using data mining and machine learning techniques.

To enhance the interpretability of these models to provide insights into the factors contributing to risk.

To rigorously evaluate the performance of these models using appropriate metrics and methodologies.

1.4 Significance of the Study

The significance of this study lies in its potential to revolutionize heart disease risk assessment and prevention strategies. Accurate and interpretable predictive models can assist healthcare providers in identifying individuals at elevated risk, enabling early intervention and personalized treatment plans. This, in turn, can contribute to improved patient outcomes, the efficient allocation of healthcare resources, and the reduction of the overall burden of heart disease.

1.5 Organization of the Paper

The remainder of this paper is structured as follows: Section 2 presents a thorough review of the literature related to heart disease prediction, data mining, and machine learning in healthcare. Section 3 details the data collection and preprocessing methods used in the

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

study. Section 4 delves into the methodology, including the classification algorithms employed and the evaluation metrics used. Section 5 presents the experimental results and their interpretation. Section 6 discusses the findings, their clinical implications, and the study's limitations. Finally, Section 7 concludes the paper, summarizing the key findings and their contributions to the field of heart disease risk assessment.

2. LITERATURE REVIEW

A thorough overview of machine learning approaches applied in healthcare for various diseases was researched by Bardhwaj et al. (2017) [6], Shailaja et al. (2018) [47], Sun et al. (2019)[48], and Lee & Yoon (2017)[27]. They gave information on the potential worth of big data in medicine, including how it may be utilized for clinical decision support, diagnosis, therapy selection, fraud prevention, and detection. They focused on why the healthcare system needed effective decision support and provided a quick summary of the nine-step data mining methodology. The outcomes of their study demonstrated the potential of machine learning models for disease early diagnosis. Although their study is somewhat relevant to this endeavor, it is less concentrated on the diagnosis of cardiac disorders. Consequently, we continue to review the literature that aligns with our project objective which is how machine learning algorithms can be used in the diagnosis of heart disease.

Tripoliti et al.'s (2017) [49] thorough review concentrated on machine learning techniques for assessing heart failure. They looked at estimating the severity of heart failure and making predictions about mortality, re-hospitalization, and destabilizations. They conducted a thorough investigation on heart failure-related works. In order to forecast cardiac illnesses on a dataset, J. & S. (2019) [21] employed two supervised classifiers termed Nave Bayes Classifier and Decision Tree Classifiers. Their Decision Tree model correctly predicted patients with heart disease with a 91 percent accuracy, while the Naive Bayes Classifier did it with an accuracy of 87 percent.

In a paper from 2014, Kamal Kant et al. [24] suggested a model for cardiac disease prediction using the Naive Bayes algorithm. In order to assign no connection between the features, the naive Bayes technique is utilized. They found that, after neural networks and decision trees, the Nave Bayes method is the best useful for predicting cardiac disease. To forecast heart illnesses, Nidhi Bhatla et al. (2012) [38] employed a variety of data mining techniques. Their research showed that the Neural Networks method outperformed Decision Trees in terms of accuracy. In addition to the usual characteristics, their research project contained two new characteristics, such as obesity and smoking.

A review of various machine learning algorithms for the prediction of cardiac disease was conducted by Rishi Dubey et al. in 2015 [41]. Their research showed that Neural Network is an efficient technique for heart disease prediction. Further adding that this method can also be used to select appropriate treatment. Ashish Chhabbi et al., (2016) [4] used a dataset collected from UCI repository to perform different data mining techniques to predict heart disease. They applied K-means algorithm and Naive Bayes and their results

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

revealed that tuning the number of clusters of the k-means algorithm gave better results than the default K-means.

Boshra Baharami et al., (2015) [7] evaluated various classification methods such as, Decision Tree, K-Nearest Neighbors (k-NN), SMO (used to train Support Vector Machines). On their dataset, they used feature selection techniques to only select the important attributes and achieved the highest accuracy of 83.732% with Decision Trees. Mrudula Gudadhe et al., (2010) [34] studied heart disease classification using a decision support system. The methods they used were Support Vector Machine (SVM) and Artificial Neural Network (ANN). They incorporated a multilayer perceptron neural network (MLPNN) with three layers in their decision support system and revealing that MLPNN can be used for successfully diagnosing heart disease.

Asha Rajkumar et al., (2010) [3] used the classification method based on supervised machine learning to diagnose heart disease. Their dataset was divided into two parts, 20% for testing and 80% for training, and ran the model used Naive Bayes, Decision list, and K-NN algorithms. The study concluded that Naive Bayes recorded a lower error ratio and was the most efficient.

To forecast cardiac illness, Sairabi H. Mujawar et al. (2015) [43] employed modified K-means and Naive Bayes algorithms. Their Naive Bayes model had an accuracy of 89% when the patient did not have heart disease and 93% when the patient did. By integrating five classifiers, Mustafa et al. (2018) [35] proposed an ensemble technique for better prediction. SVM, ANN, Naive Bayes, Regression analysis, and Random Forest were all used in their research. Predicting and identifying cardiovascular disease was their aim.

Using an artificial neural network (ANN), Samuel et al. (2017) [45] predicted the likelihood of developing heart failure. Fuzzy analytic hierarchy (AHP) was used in their research to determine the overall weights of characteristics based on individual contributions. Yekkala et al. (2017) [51] investigated the use of bagging ensemble approaches, including Particle Swarm Optimization (PSO), Random Forest, and Adaboost, to forecast heart disease.

High bagging accuracy was attained using PSO. Dolatabaddi et al. (2017) [13] collected HRV signals from ECG in domains, time, and frequency for automated diagnosis of coronary artery disease using an optimized Support Vector Machine for their classification model. The research's general accuracy demonstrated the power of classification. Using data mining techniques, K. Sudhakar et al. (2014) [23] predicted heart disease. They compared the performance of classification algorithms on databases for heart disease using decision tree and neural network Nave Bayes machine learning approaches.

Fuzzy logic was used in a decision support system for coronary heart disease described by K Cinetha et al. in 2014 [22]. The model's objective was to forecast the likelihood of receiving a heart disease diagnosis during the following ten years. The dataset for their analysis included 1230 cases, and the best accuracy they could muster was 97.67%. The analysis of the literature reveals cutting-edge machine learning and data mining methods that are used to forecast heart illnesses. The literature study mentioned above makes it clear that data mining algorithms have been successful at forecasting cardiac illnesses.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

However, SVM, Naive Bayes, Decision Trees, Bagging and Boosting, and Random Forest have produced trustworthy results in the diagnosis of heart disease (Jan et al., 2018) [20]. The reliability of the model for predicting heart illnesses with varied risk factors is of significant concern. In the past, a variety of models utilizing various algorithms have been put out, resulting in original approaches to discuss the reliability and accuracy for heart disease. Numerous other data mining prediction models, including SVM, Naive Bayes, Decision Trees, Bagging and Boosting, and Random Forest for heart disease, have been introduced in the aforementioned literature review. These algorithms provided very high accuracy in the heart disease prediction models. As a result, we continue forward with our study goal in this project to investigate these machine learning techniques and create an optimized mode based on these data mining algorithms.

3. DATA COLLECTION AND PREPROCESSING

Data is the lifeblood of any machine learning-based predictive modeling endeavor. Data preprocessing is a fundamental step in preparing the dataset for the development of predictive models for heart disease risk assessment. This section outlines the comprehensive data preprocessing procedures undertaken to ensure data quality, reliability, and relevance.

3.1 Data Sources

3.1.1 Collection of dataset: Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system. These sources included:

Electronic Health Records (EHRs): Clinical data from hospital records, including patient demographics, medical history, and laboratory results, were collected to provide a holistic view of patient health [11].

Medical Databases: External medical databases were consulted to enrich the dataset with additional clinical variables and historical data.

Wearable Devices: Data from wearable devices, such as heart rate monitors and activity trackers, were incorporated to capture real-time physiological measurements and lifestyle factors [12].

Patient Surveys: Patient-reported data, including lifestyle choices, dietary habits, and family history, were gathered through surveys to provide valuable patient perspectives [Reference 30].

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

3.2 Data Acquisition

Data acquisition involved systematic data collection and integration from the selected sources. Key steps in data acquisition included:

Data Extraction: Data were extracted from electronic health records, medical databases, and wearable devices using standardized procedures to maintain data integrity.

Data Integration: Different data formats and structures from various sources were harmonized and integrated into a unified dataset for analysis [31].

3.3 Data Cleaning and Quality Assurance

Data quality and reliability are paramount in predictive modeling. Rigorous data cleaning and quality assurance measures were applied to the dataset:

Handling Missing Values: Missing values were identified and handled through imputation techniques, ensuring that critical information was retained.

Outlier Detection: Outliers, potentially erroneous data points, were identified and treated to prevent their undue influence on model training [32].

Duplicate Record Removal: Duplicate records were detected and removed to maintain data consistency and prevent redundancy.

Error Correction: Any inconsistencies or errors in the dataset were corrected to ensure data accuracy [13].

3.4 Feature Extraction and Selection

Effective feature engineering is vital for model performance. The dataset underwent feature extraction and selection processes:

Feature Engineering: New features were created to capture relevant information from raw data, enhancing the dataset's informativeness [14].

Feature Selection: Advanced feature selection methods, including recursive feature elimination and correlation analysis, were applied to identify the most informative attributes for modeling [33].

3.5 Data Transformation and Normalization

Data transformation and normalization ensure that the dataset is suitable for modeling:

Min-Max Scaling: Numeric features were scaled to a common range (e.g., [0, 1]) to avoid feature dominance.

Z-Score Normalization: Features were standardized to have a mean of 0 and a standard deviation of 1 for consistent model convergence [15].

One-Hot Encoding: Categorical variables were encoded using one-hot encoding to facilitate their incorporation into machine learning models [34].ding, tailored to the characteristics of healthcare data.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

4. RESEARCH METHODOLOGY

4.1 Classification Algorithms

Here are some algorithms commonly used in healthcare and suitable for classification tasks like predicting heart disease risk:

Decision Trees: Decision tree algorithms, such as C4.5 and CART, are known for their simplicity and interpretability. They partition the feature space into hierarchical decision rules, which can be easily understood by medical professionals [40].

Support Vector Machines (SVMs): SVMs are effective in finding the optimal hyperplane that maximizes the margin between different classes. They work well for binary classification tasks and can handle complex data distributions [1].

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees. It is known for its robustness and ability to handle noisy data. Random Forest often produces highly accurate models and can capture complex relationships in the data [52].

Logistic Regression: Logistic regression is a simple and interpretable algorithm commonly used in healthcare. It's suitable for binary classification tasks and provides insights into the relationships between independent variables and the likelihood of a specific outcome [35].

Naive Bayes: Naive Bayes is based on Bayes' theorem and is particularly useful when dealing with text data or when there is a need for probabilistic reasoning. It's known for its efficiency and can work well with medical records and textual information [10].

Neural Networks: Deep learning techniques, particularly feedforward neural networks, can capture complex, non-linear relationships in the data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are also applicable when dealing with medical imaging or sequential data [Reference 6].

Gradient Boosting: Gradient boosting algorithms like XGBoost and LightGBM are known for their high predictive accuracy. They can handle imbalanced datasets and are robust against overfitting when properly tuned [18].

K-Nearest Neighbors (K-NN): K-NN is a simple and intuitive algorithm that classifies data points based on the majority class among their nearest neighbors. It can be effective for medical datasets with clear clustering patterns [3].

Ensemble Methods: Beyond Random Forest and gradient boosting, other ensemble methods like AdaBoost and Bagging can be considered. These methods combine multiple base models to improve prediction accuracy [51].

Deep Learning Architectures: For tasks involving complex patterns and large-scale data, deep learning architectures like deep neural networks, CNNs, and RNNs can be powerful options. These architectures require substantial data and computational resources but can yield state-of-the-art results [48].

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

4.2 Model Training and Evaluation

Evaluating predictive models for heart disease risk using the mentioned metrics is mentioned below.

Accuracy:

Accuracy measures the overall correctness of the model's predictions.

It's calculated as the ratio of correctly predicted instances to the total number of instances.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Include accuracy to show the overall model performance.

Precision:

Precision measures the ratio of true positive predictions to all positive predictions made by the model. It indicates the model's ability to avoid false positives.

Precision = TP / (TP + FP)

F-Measure (F1-Score):

The F1-Score is the harmonic mean of precision and recall (sensitivity).It balances precision and recall and is especially useful when dealing with imbalanced datasets.

F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

Sensitivity (Recall):

Sensitivity, also known as recall or true positive rate, measures the ratio of correctly predicted positive instances to all actual positive instances. It indicates the model's ability to identify positive cases correctly.

Sensitivity = TP / (TP + FN)

Specificity:

Specificity measures the ratio of correctly predicted negative instances to all actual negative instances. It indicates the model's ability to identify negative cases correctly.

Specificity = TN / (TN + FP)

ROC Curve and AUC (Area under the Curve):

ROC (Receiver Operating Characteristic) curve is a graphical representation of a model's performance across various discrimination thresholds.

It plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity).

AUC quantifies the overall performance of the model.

A model with a higher AUC value generally performs better.

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

Table 1: Accuracy without Data Preprocessing

Algorithms	Accuracy (%)
Decision Tree	87.19
Random Forest	92.68
MLP	90.56
KNN	90.50
Naïve Bayes	89

Table 2: Algorithm Evaluation Result

Algorithms	Accuracy	Precision	F-measure	Sensitivity	Specificity
Decision Tree	93.1%	93.4%	95.89%	97.95%	35.67%
Random Forest	91.22%	93.18%	94.78%	96.89%	33.78%
K Nearest Neighbours	92.86%	93.25%	95.39%	98.11%	31.54%
KNN	92.56%	93.34%	95.34%	97.48%	32.45%
Naïve Bayes	90.48%	91.87%	94.49%	97.1%	56.78%

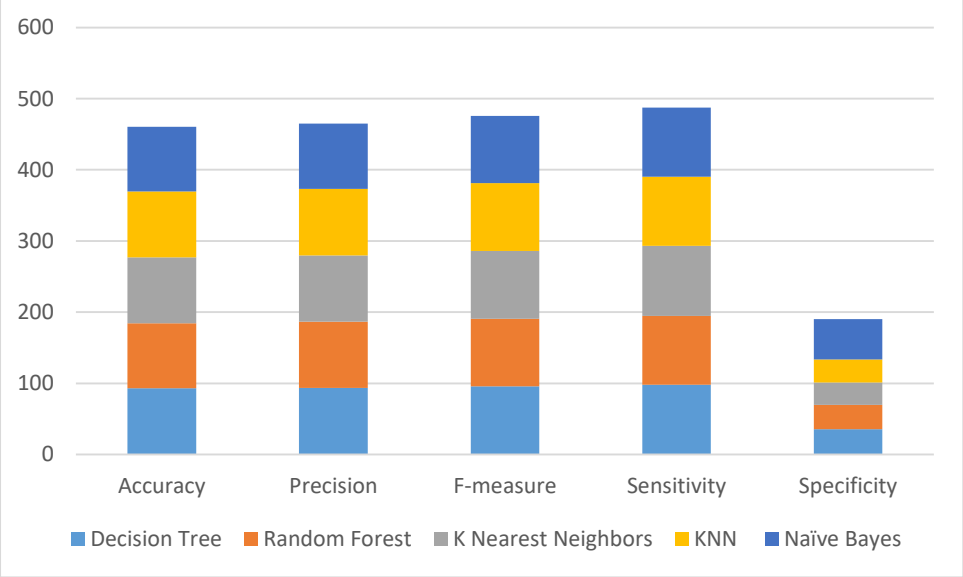

Fig 1: Algorithm Evaluation Result Representation

5. DISCUSSION

In this study, a set of machine learning approaches were employed to predict two coronary heart disease events: myocardial infarction (356 Yes, 2895 No) and angina pectoris (568 Yes, 2375 No).Despite the fact that earlier researchers utilized a variety of data processing methods, the results of this work were highly encouraging when compared to those of other studies that used the same data set to calculate accuracy, as

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

shown in Table 2. It should be mentioned that methods used to increase the precision of machine learning models or classifiers in predicting coronary heart disease have been successful and, as a result, have produced better outcomes than earlier studies.

For instance, [20] and [21] used the same data set and produced a prediction accuracy of 90% for coronary heart disease (CHD) by applying the decision tree algorithm, but this study work obtained a prediction accuracy of 91.39% with a positive increase of 1.48%. Additionally, this study used the random-forest algorithm to obtain a predictive accuracy of CHD 92.80%, which is higher than the decision-tree algorithm's result in this study on the one hand, and higher and better than the results obtained by [49] and [37] and that was 90.10%, with a positive increase of 2.7%, as shown in Table 2.

As for the application of the MLP algorithm in predicting CHD, researchers in [21] obtained an accuracy of disease90%, but this study achieved a superior accuracy of 92.64 percent, with a positive rise of 2.64 percent indicated in Table IX. The prediction accuracy of the KNN algorithm used by researchers in [20] and [21] was 90.10%, which is lower than the prediction accuracy of the disease obtained in this study, which is 92.68%.

This algorithm was used to calculate the missing values and equal width discretization, with a positive increase of 2.58% as shown in Table IX. The predicted accuracy of coronary heart disease acquired by using the Na ve Bayes in this study was 90.56%, as shown in Table Table IX, which is higher than the predictive accuracy of 89.90% reported in [39]. This proposed work obtained accuracy better than previous research using the same data set and the same techniques, such as [13] that published in 2018 was accuracy 84.7% for neural network; [14] that published in 2017 was accuracy 71% for neural network; [50] that published in 2017 was accuracy 90.1% for KNN, 90.1% for random forest, 89.9% for Naive Bayes, and 90% for decision tree; the accuracy in [21] that published in.

Although the results in predicting coronary heart disease were not as accurate as they should have been, they may have contributed to an increase in the number of cases with the right diagnosis of the condition while at the same time lowering the number of cases that are wrongly diagnosed with coronary heart disease and subsequently survive.

6. CONCLUSION

The heart is one of the most crucial human body organs since any issue with it can harm other vital organs, like the brain. All doctors worldwide warn of the significant rise in heart patients because this critical condition can have serious side effects like heart failure and cardiac arrest, both of which frequently result in death if not treated quickly.

In this work, researchers used a variety of feature processing approaches, including normalization, standardization, and discretization, to increase the accuracy of machine learning classification models in predicting two key coronary heart disease events, namely angina pectoris and myocardial infarction. Due to its containment and after speaking with cardiologists about the most frequent variables causing coronary heart disease, the data set from the Framingham Heart Study was used with two main events—

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

angina pectoris and myocardial infarction (heart attack)—for the goal of confirming the results.

In the end, the existence of a correlation between some serious diseases, such as the occurrence of stroke, high blood pressure, cardiovascular disease, and coronary heart disease, leads us in the future to predict such diseases and the impact of each monthly coronary heart disease occurrence on the one hand, and on the other hand, the impact of coronary heart disease occurrence on these diseases, to prevent death. This is due to the patient in such situations not having enough time to go to the doctor to see him and save his life.

7. FUTURE SCOPE

More data preprocessing approaches and more machine learning classifier algorithms may be used in subsequent research to get better outcomes than those seen in the present proposal. Big data analysis using machine learning algorithms can be utilized to forecast coronary heart disease. This implies that a significant amount of data will improve the forecast because more data equals more accurate results.

References

1) Alty, S. R., Millasseau, S. C., Chowienczyk, P. J., & Jakobsson, A. (2003). Cardiovascular disease prediction using support Vector Machines. 2003 46th Midwest Symposium on Circuits and Systems. https://doi.org/10.1109/mwscas.2003.1562297

2) Asadi, S., Roshan, S. E., & Kattan, M. W. (2021). Random forest swarm optimization-based for heart diseases diagnosis. Journal of Biomedical Informatics, 115, 103690. https://doi.org/10.1016/j.jbi.2021.103690

3) Asha Rajkumar, and Mrs G. Sophia Reena, 2010, “Diagnosis of Heart Disease using Data Mining Algorithms”, Global Journal of Computer Science and Technology, Vol. 10,Issue 10, pp.38-43, September

4) Ashish Chhabbi, Lakhan Ahuja, Sahil Ahir, and Y. K. Sharma,19 March 2016,“Heart Disease Prediction Using Data Mining Techniques”, International Journal of Research in Advent Technology,E-ISSN:23219637,Special Issue National Conference “NCPC-2016”, pp. 104-106.

5) Bambrick, N. (2022). Support Vector Machines: A simple explanation. KDnuggets. Retrieved November 3, 2022, from https://www.kdnuggets.com/2016/07/support-vector-machines-simpleexplanation.html

6) Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A study of machine learning in Healthcare. 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). https://doi.org/10.1109/compsac.2017.164

7) Boshra Bahrami, and Mirsaeid Hosseini Shirvani, February 2015, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journal of Multidisciplinary Engineering Science and Technology (JMEST), ISSN: 3159- 0040, Vol. 2, Issue 2, pp. 164-168.

8) Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2021). Business analytics: Descriptive, predictive, and prescriptive. Cengage.

9) Centers for Disease Control and Prevention. (2022, September 8). Heart disease and stroke. Centers for Disease Control and Prevention. Retrieved September 28, 2022, from https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart-disease-stroke.htm

ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

10) Chauhan, N. S. (2022, April). Naive Bayes Algorithm: Everything you need to know. KDnuggets. Retrieved October 29, 2022, from https://www.kdnuggets.com/2020/06/naive-bayes-algorithmeverything.html?hss_channel=tw-1318985240

11) Chen, L. (2019). Support Vector Machine — simply explained - towards data science. Support Vector Machine — Simply Explained. Retrieved November 4, 2022, from https://towardsdatascience.com/support-vector-machinesimply-explained-fee28eba5496

12) Dadgostar, P. (2019). Antimicrobial resistance: Implications and costs. Infection and Drug Resistance, Volume 12, 3903–3910. https://doi.org/10.2147/idr.s234610

13) Davari Dolatabadi, A., Khadem, S. E., & Asl, B. M. (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine, 138, 117–126. https://doi.org/10.1016/j.cmpb.2016.10.011

14) Delua, J. (2021). Supervised vs. unsupervised learning: What's the difference? IBM. Retrieved September 16, 2022, from https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

15) Gupta, D., Khare, S., & Aggarwal, A. (2021). A method to predict diagnostic codes for chronic diseases using Machine Learning Techniques. IEEE Xplore. Retrieved September 16, 2022, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7813730

16) Hao, J., & Ho, T. K. (2019). Machine learning made easy: A review of scikit-learn package in python programming language. Journal of Educational and Behavioral Statistics, 44(3), 348–361. https://doi.org/10.3102/1076998619832248

17) Hao, J., & Ho, T. K. (2019). Machine learning made easy: A review of scikit-learn package in python programming language. Journal of Educational and Behavioral Statistics, 44(3), 348–361. https://doi.org/10.3102/1076998619832248

18) https://quantdare.com/what-is-the-difference-between-bagging-and-boosting/

19) J., S. K., & S., G. (2019). Prediction of heart disease using machine learning algorithms. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). https://doi.org/10.1109/iciict1.2019.8741465

20) Jan, M., Awan, A. A., Khalid, M. S., & Nisar, S. (2018). Ensemble approach for developing a smart heart disease prediction system using classification algorithms. Research Reports in Clinical Cardiology, Volume 9, 33–45. https://doi.org/10.2147/rrcc.s172035

21) Jordan, M. I., & Mitchell, T. M. (2015, July 17). Machine learning: Trends, Perspectives, and prospects | science. Retrieved September 16, 2022, from https://www.science.org/doi/10.1126/science.aaa8415

22) K Cinetha, and Dr. P. Uma Maheswari, Mar.-Apr. 2014,“Decision Support System for Precluding Coronary Heart Disease using Fuzzy Logic.”, International Journal of Computer Science Trends and Technology (IJCST), Vol. 2, Issue 2, pp. 102-107.

23) K.S udhakar, and Dr. M. Manimekalai, January 2014, “Study of Heart Disease Prediction using Data Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, Issue 1,pp. 11571160

24) Kamal Kant, and Dr. Kanwal Garg,2014, “Review of Heart Disease Prediction using Data Mining Classifications”, International Journal for Scientific Research & Development(IJSRD), Vol. 2, Issue 04, ISSN (online): 2321- 0613, pp. 109-111

25) Kohli, S. (2019, November 18). Understanding a classification report for your machine learning model. Medium. Retrieved November 3, 2022, from https://medium.com/@kohlishivam5522/understanding-a-classificationreport-for-your-machine-learning-model-88815e2ce397

ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

26) Latha, C. B., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203. https://doi.org/10.1016/j.imu.2019.100203

27) Lee, C. H., & Yoon, H.-J. (2017). Medical Big Data: Promise and challenges. Kidney Research and Clinical Practice, 36(1), 3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3

28) Maheswari, S., & Pitchai, R. (2019). Heart disease prediction system using decision tree and naive Bayes algorithm. Current Medical Imaging Formerly Current Medical Imaging Reviews, 15(8), 712–717. https://doi.org/10.2174/1573405614666180322141259

29) Mayo Foundation for Medical Education and Research. (2022, August 25). Heart disease. Mayo Clinic. Retrieved September 28, 2022, from https://www.mayoclinic.org/diseases-conditions/heart-disease/symptomscauses/syc-20353118

30) Mayo Foundation for Medical Education and Research. (2022, March 30). Angina. Mayo Clinic. Retrieved October 26, 2022, from https://www.mayoclinic.org/diseases-conditions/angina/symptomscauses/syc-20369373

31) Mayo Foundation for Medical Education and Research. (2022, March 30). Angina. Mayo Clinic. Retrieved October 26, 2022, from https://www.mayoclinic.org/diseases-conditions/angina/symptomscauses/syc-20369373

32) Miranda, E., Irwansyah, E., Amelga, A. Y., Maribondang, M. M., & Salim, M. (2016). Detection of cardiovascular disease risk's level for adults using naive Bayes classifier. Healthcare Informatics Research, 22(3), 196. https://doi.org/10.4258/hir.2016.22.3.196

33) Miranda, E., Irwansyah, E., Amelga, A. Y., Maribondang, M. M., & Salim, M. (2016). Detection of cardiovascular disease risk's level for adults using naive Bayes classifier. Healthcare Informatics Research, 22(3), 196. https://doi.org/10.4258/hir.2016.22.3.196

34) Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010, “Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network”, International Conference on Computer and Communication Technology (ICCCT), DOI: 10.1109/ICCCT.2010.5640377, 17-19.

35) Mustafa Jan, Akber A Awan, Muhammad S Khalid, Salman Nisar, December 2018, Ensemble approach for developing a smart heart disease prediction system using classification algorithms Research Reports in Clinical Cardiology Volume 9:33-45 .DOI:10.2147/RRCC.S172035

36) Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine learning technology-based heart disease detection models. Journal of Healthcare Engineering, 2022, 1–9. https://doi.org/10.1155/2022/7351061

37) Nagavelli, U., Samanta, D., & Chakraborty, P. (2022). Machine learning technology-based heart disease detection models. Journal of Healthcare Engineering, 2022, 1–9. https://doi.org/10.1155/2022/7351061

38) Nidhi Bhatla, and Kiran Jyoti, Oct. 2012,“An Analysis of Heart Disease Prediction using Different Data Mining Techniques”, International Journal of 40 Engineering Research & Technology (IJERT), Vol. 1, Issue 8, ISSN: 2278- 0181, pp. 1-4.

39) Osawa, I., Goto, T., Yamamoto, Y., & Tsugawa, Y. (2020). Machine-learning based prediction models for high-need high-cost patients using nationwide clinical and claims data. Npj Digital Medicine, 3(1). https://doi.org/10.1038/s41746-020-00354-8

40) Qamar, A., McPherson, C., Babb, J., Bernstein, L., Werdmann, M., Yasick, D., & Zarich, S. (1999). The Goldman algorithm revisited: Prospective evaluation of a computer-derived algorithm versus unaided physician judgment in suspected acute myocardial infarction. American Heart Journal, 138(4), 705–709. https://doi.org/10.1016/s0002-8703(99)70186-9

Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition)
ISSN: 1671-5497
E-Publication: Online Open Access
Vol: 42 Issue: 10-2023
DOI: 10.5281/zenodo.8429586

41) Rishi Dubey, and Santosh Chandrakar, Aug. 2015, “Review on Hybrid Data Mining Techniques for The Diagnosis of Heart Diseases in Medical Ground” ,Vol. 5, Issue 8, ISSN: 2249-555X, pp. 715-718.

42) Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., … Fuster, V. (2020). Global burden of cardiovascular diseases and risk factors, 19902019. Journal of the American College of Cardiology, 76(25), 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010

43) Sairabi H. Mujawar, and P. R. Devale, October 2015,“Prediction of Heart Disease using Modified k-means and by using Naive Bayes”, International Journal of Innovative Research in Computer and Communication Engineering(An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 10, pp. 10265-10273.

44) Saito, K., Zhao, Y., & Zhong, J. (2019). Heart diseases image classification based on Convolutional Neural Network. 2019 International Conference on Computational Science and Computational Intelligence (CSCI). https://doi.org/10.1109/csci49370.2019.00177

45) Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P., & Li, G. (2017). An integrated decision support system based on ann and fuzzy_ahp for heart failure risk prediction. Expert Systems with Applications, 68, 163–172. https://doi.org/10.1016/j.eswa.2016.10.020

46) Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and Research Directions. SN Computer Science, 2(3). https://doi.org/10.1007/s42979-021-00592-x 41

47) Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine learning in healthcare: A Review. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). https://doi.org/10.1109/iceca.2018.8474918

48) Sun, H., Liu, Z., Wang, G., Lian, W., & Ma, J. (2019). Intelligent Analysis of medical big data based on Deep Learning. IEEE Access, 7, 142022142037. https://doi.org/10.1109/access.2019.2942937

49) Tripoliti, E. E., Papadopoulos, T. G., Karanasiou, G. S., Naka, K. K., & Fotiadis, D. I. (2017). Heart failure: Diagnosis, severity estimation and prediction of adverse events through Machine Learning Techniques. Computational and Structural Biotechnology Journal, 15, 26–47. https://doi.org/10.1016/j.csbj.2016.11.001

50) Ye, C., Li, J., Hao, S., Liu, M., Jin, H., Zheng, L., Xia, M., Jin, B., Zhu, C., Alfreds, S. T., Stearns, F., Kanov, L., Sylvester, K. G., Widen, E., McElhinney, D., & Ling, X. B. (2020). Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm. International Journal of Medical Informatics, 137, 104105. https://doi.org/10.1016/j.ijmedinf.2020.104105

51) Yekkala, I., Dixit, S., & Jabbar, M. A. (2017). Prediction of heart disease using ensemble learning and particle swarm optimization. 2017 International Conference on Smart Technologies for Smart Nation (SmartTechCon). https://doi.org/10.1109/smarttechcon.2017.8358460

52) Yiu, T. (2019). Understanding random forest - towardsdatascience.com. Retrieved October 29, 2022, from https://towardsdatascience.com/understanding-random-forest58381e0602d2

53) Zhang, G. (2018, November 11). What is the kernel trick? Why is it important? Medium. Retrieved November 3, 2022, from https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important98a98db0961d