International Journal of Emerging Engineering and Technology (IJEET)
ISSN (e): 2958-3764
Volume: 2, Number: 1, Pages: 6- 12, Year: 2023

Heart Disease Prediction Using Hybrid Random Forest and Linear Model

Zeshan Khan, Gulbadan Sikandar, and Shahzad Anwar

Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar
Corresponding author e-mail: (zkhan@uetpeshwar.edu.pk)

Received: 10/04/2023, Revised: 25/06/2023, Accepted: 02/07/2023

Abstract— **Heart disease remains a significant global health concern, and accurate heart disease risk prediction plays a crucial role in its prevention. In recent years, machine learning techniques have shown promising results in cardiovascular disease prediction. This paper proposes a novel Hybrid Random Forest and Linear Model (HRFLM) that leverages diverse patient attributes to accurately predict the risk of developing heart disease. HRFLM utilizes a combination of feature engineering, feature selection, and an ensemble of machine learning algorithms to effectively capture the complex relationships between patient characteristics and heart disease outcomes. The model is trained and validated on a comprehensive dataset from the Framingham Heart Study, which includes a wide range of demographic, clinical, and lifestyle variables. The experimental results demonstrate that HRFLM outperforms several state-of-the-art machine learning models like K-nearest neighbour (KNN), support vector machine (SVM), decision tree (DT), multilayer perceptron (MLP) and random forest (RF) in terms of accuracy, precision, recall, and F1 score. The proposed model provides valuable insights into the risk factors associated with heart disease and can assist doctors in identifying individuals at high risk for preventive interventions. The proposed model has achieved an accuracy of 88%.**

***Index Terms*— Heart disease, HRFLM, Hybrid Algorithm, Machine Learning**

I. INTRODUCTION

The heart is the body's most vital organ since it pumps blood to all tissues and organs. The heart is linked to other circulatory system components, such as the arteries, veins, and capillaries. The heart is a vulnerable organ, although it is one of the most vital organs in the body [1]. A cardiovascular disease (CVD) is a medical condition that describes any issue inside the cardiovascular system, whether it be the heart or the blood vessels. Heart disease is often linked with a condition called atherosclerosis,
where there is an accumulation of fatty substances in the artery walls and a heightened chance of forming blood clots. It is also possible that it is linked to the deterioration of the arteries in several organs, such as the kidneys, eyes, heart, brain, and heart [2].
When cardiovascular disease occurs, the blood veins in the heart become obstructed, leading to severe chest discomfort, angina, and even a heart attack.

These symptoms manifest themselves after the disease has reached its last stage. At that point, the patient needs to undergo surgery. In medicine, making an accurate diagnosis at an early stage is of the utmost importance for the patient's health [3]. Approximately 80% of cardiovascular diseases (CVDs) are developed due to modifiable risk factors. Unhealthy diet choices, physical inactivity, tobacco use, high blood pressure, elevated cholesterol levels, obesity, and diabetes contribute to developing CVDs [4]. Poor dietary habits, characterized by high intake of saturated fats, trans fats, cholesterol, salt, and added sugars, increase the risk of vascular problems. Sedentary lifestyles and a lack of regular physical activity contribute to obesity, hypertension, high cholesterol levels, and diabetes, raising the likelihood of developing CVDs [5]. Tobacco use, whether through smoking or exposure to secondhand smoke, significantly damages blood vessels and promotes clot formation, leading to cardiovascular issues. Hypertension, or high blood pressure, stresses blood vessels, increasing the risk of heart attacks, strokes, and other heart-related problems.

Elevated cholesterol levels, especially low-density lipoprotein (LDL) cholesterol, contribute to plaque formation in the arteries, restricting blood flow and raising the risk of heart disease [6]. Obesity, particularly abdominal obesity, is associated with a higher likelihood of developing CVDs. Lastly, uncontrolled diabetes mellitus, especially type 2 diabetes, leads to complications such as atherosclerosis and coronary artery disease. Individuals can significantly reduce their risk of developing cardiovascular diseases by addressing these modifiable risk factors through lifestyle changes and medical interventions [7]. Several studies employed data from the Framingham heart disease dataset [8].

Machine learning prediction models need the right data for both training and testing. The precision of machine learning classifiers may be increased by training and testing them on a finely tuned dataset. Also, the predictive model's performance can be improved by including significant and associated data elements [9]. Therefore, for machine learning classifiers to be accurate, data normalization and feature selection are needed

[10]. Numerous researchers have employed various prediction algorithms in the literature; however, these methods do not reliably forecast cardiac problems. Data standardization is required to increase the accuracy of machine learning classifiers [11]. Many standardization approaches are utilized to eliminate the missing feature value instances from the dataset [12], including standard scalar (SS), min-max scalar, and others. The accessibility of the datasets has increased the accuracy of prediction models based on machine learning, and it has opened new doors for research into creating original algorithms that can estimate the risk of cardiovascular disease (CVD). These databases contain information on various risk factors and the patient's health. Because there is currently access to clinical datasets that are both inconsistent and redundant, preprocessing is required before the creation of prediction models for CVD [13]. In a similar vein, information is accessible regarding a large number of risk factors (features). The selection of an appropriate set of features is based on many characteristics, such as their prevalence in most populations, their influence on heart disease alone, and their potential to be managed or treated to minimize risk.

The rest of the paper is organized as follows; section 2 provides a literature review and analysis of the algorithms' limitations. Section 3 provides the methodology of the proposed model. Section 4 shows the results of using different algorithms and their comparison with our proposed model. In section 5, we concluded our work and outlined future research.

II. LITERATURE REVIEW

Across several sectors, including the medical and healthcare fields, there is significant interest in machine learning. According to the findings of experts [14], there are a range of successful methods that machine learning has been applied to diagnose heart diseases. CVDs are the major cause of death and morbidity worldwide, burdening healthcare systems and individuals. Preventive interventions, early intervention, and personalized treatment require CVD risk prediction [15]. Due to their ability to analyze huge, diverse datasets, discover complicated patterns, and make accurate predictions [16], machine learning (ML) techniques have gained popularity in healthcare. ML approaches combined with conventional risk factors and clinical variables can provide a comprehensive and personalized prediction model that considers many parameters affecting CVD development and progression [17]. Clinical practice has relied on risk assessment models like the Framingham Risk Score to estimate a patient’s 10-year CVD risk. These models may not fully describe CVD pathogenesis due to their narrow risk factor collection [18].

ML algorithms can analyze vast datasets and find complicated correlations between genetic, lifestyle, and environmental risk variables, enabling more accurate prediction models.

This work proposes combining traditional risk assessment models with ML algorithms to use large-scale dataset’s rich information [19-25]. The hybrid paradigm integrates electronic health records, genetic profiles, lifestyle data, and medical imaging to assess CVD risk holistically and individually [20, 26-30]. This method may improve risk classification, treatment, and patient outcomes. The hybrid approach can also uncover new risk factors and patterns. Decision trees, random forests, support vector machines, and deep learning models can identify high-risk subgroups and generate tailored preventative treatments by uncovering complicated nonlinear risk factor correlations and interactions [21]. As a direct consequence of this, numerous studies focusing on developing medical applications by utilizing a variety of machine-learning algorithms and approaches have been published [22]. It is mentioned in [23] that particle swarm optimization and the feedforward back-propagation of neural networks can be used to detect cardiac disease. The goal of the research was to use data mining methodologies to make accurate diagnoses of heart issues. Classification techniques, like neural networks and decision trees, can be used to predict cardiac disease and determine which elements are most significant in causing it. The authors of this study attempted to mitigate these challenges by using many methods. Given the necessity, this study [24] proposes to design an intelligent system for anticipating and correcting heart disease diagnoses, eliminating unintended mistakes, saving medical expenses, and enhancing the level of care.

In [25], the authors examine the factors considered while predicting the condition. After that, the in-depth analysis included in each contribution presents the performance attainments. In this work [26], authors explored and used a variety of classifiers, including the Support vector machine, K-nearest neighbour, Random Forest, Decision Tree, etc. Based on the provided dataset, the Random Forest classifier is shown to be the most effective. Doctors can forecast the likelihood of CVD by entering patient data into a web program. The characteristics in the cardiac disease dataset are extracted in [27] using the feature extraction method gradient boosting-based sequential feature selection (GBSFS), which is then used to generate useful data for medical services. The research paper's authors used data mining techniques and two prediction models, the XGBoost algorithm and logistic regression, to predict chronic heart disease (CHD) [28]. The authors in [29, 31-35] used an ensemble-based machine learning model for heart disease prediction, evaluated on UCI and Framingham datasets. SVM and LR demonstrate the highest accuracies on UCI and Framingham datasets, respectively, while RF excels in sensitivity. The Voting approach shows superior performance across multiple metrics on the UCI dataset.

III. METHODOLOGY

This section outlines the methodology utilized in the study, encompassing a comprehensive depiction of the characteristics of the heart dataset, techniques employed for data preprocessing, methods for feature extraction and selection, evaluation of the performance of machine learning classification algorithms, and prediction. Fig. 1 illustrates the experimental process of the proposed methodology.

A. Data Collection

The Framingham heart dataset (FHD) is presented, which contains information on 4240 samples and 16 risk factors for cardiovascular disease.

B. Data Preprocessing

After data preprocessing, 3401 and 557 samples belong to class 0 and class 1, respectively. As shown in Fig. 2. Class 0 represents the absence of heart disease, and class 1 represents the presence of heart disease.

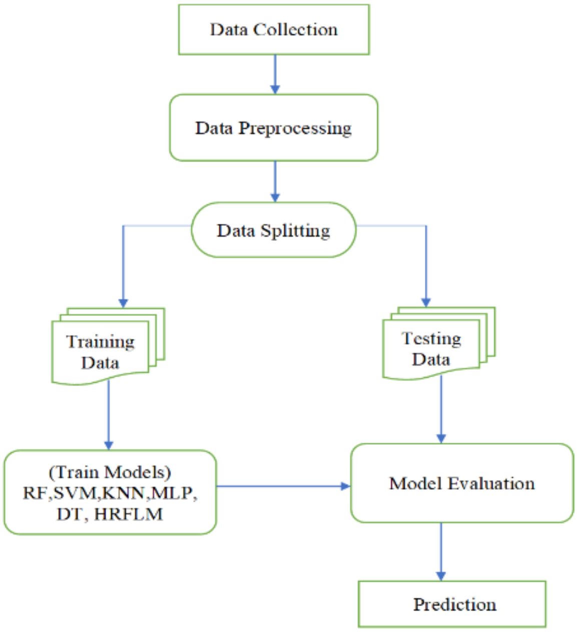

Figure 1. Flowchart of the proposed methodology.

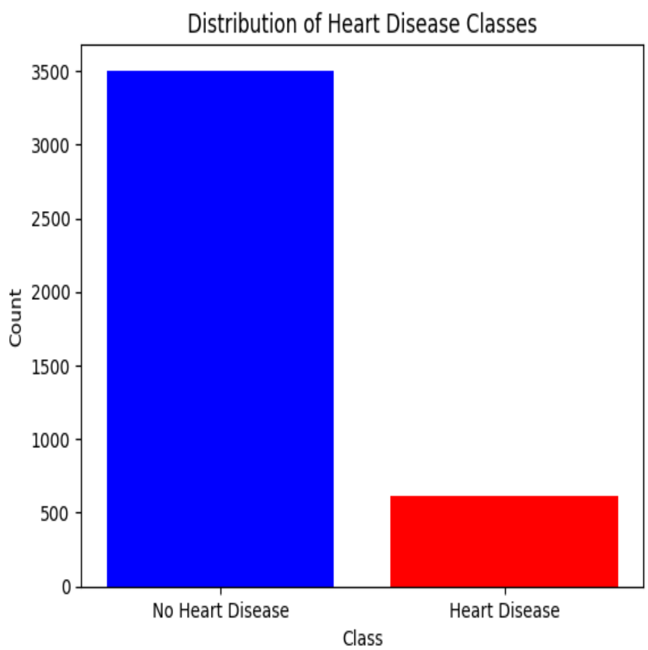

Figure 2. Distribution of heart disease classes

The dataset has people from different age groups ranging from 30 to 70 years old. Figure 3 shows a graphical representation of individuals from different age groups in FHD.

C. Outlier Detection

Outlier detection is a technique used in data analysis to identify data points that deviate significantly from the majority of We used statistical methods to calculate mean, standard deviation, and z-scores to identify data points that fall outside a certain range. These outlier values are then replaced with standard deviation values. Figure 4 shows multiple variables with and without outliers.

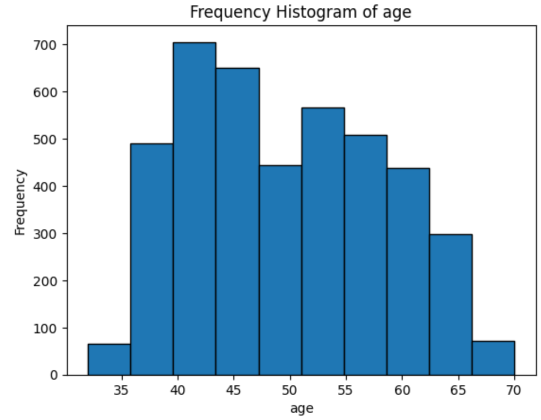

Figure 3. Frequency histogram of age

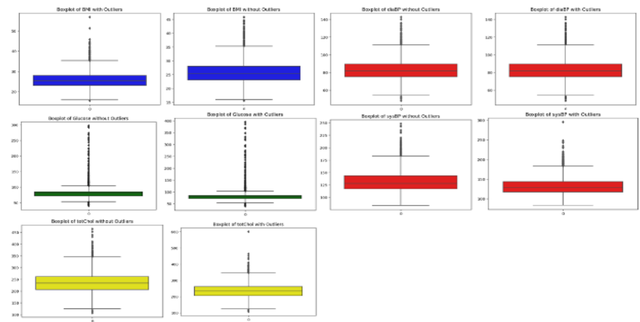

Figure 4. Visual representation of different features with & and without outliers.

D. Data Correlation

Correlation analysis determines how changes in one variable relate to changes in another. A positive correlation indicates that the other variable also tends to increase as one variable increases, while a negative correlation indicates an inverse relationship. A correlation matrix is shown in Fig. 5.

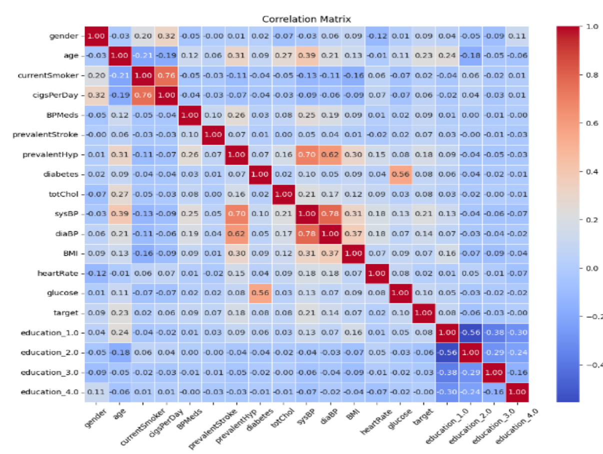

Figure 5. Correlation matrix

IV. MACHINE LEARNING MODELS

This section discusses different machine learning models like KNN, SVM, DT, MLP, RF and HRFLM. A brief discussion about the models is stated in subsections.

E. Support Vector Machines

Support Vector Machines is a supervised machine learning algorithm to solve classification problems based on the theory of problems. In the context of our heart disease prediction, SVM classified patients into those with heart disease and those without, based on the features of FHD. The SVM model leverages its ability to find complex decision boundaries and make accurate predictions, as shown in Fig. 6.

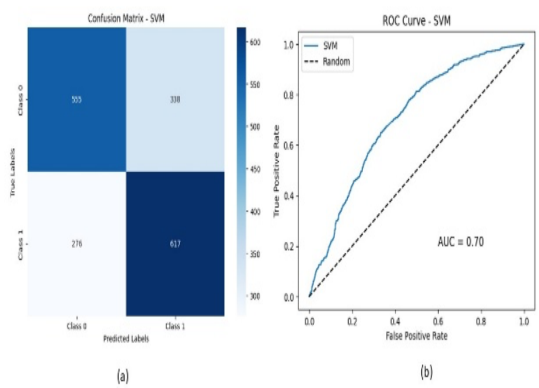

Figure 6. (a) Confusion Matrix (b) ROC curve of SVM

F. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is used for classification and regression tasks. The K-Nearest Neighbors (KNN) algorithm operates by computing the Euclidean distance between a novel data point and all pre-existing data points within the training set. Subsequently, the algorithm recognizes the K closest neighbours by utilizing computed distances. In the context of classification, the K-Nearest Neighbors (KNN) algorithm assigns a new data point to the class most frequently observed among its k-closest neighbours as shown in Fig. 7.

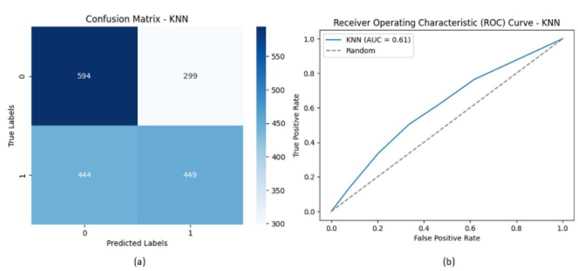

Figure 7. (a) Confusion Matrix of KNN (b) ROC Curve of KNN

G. Decision Tree

Decision Tree was used to predict the presence or absence of heart disease based on the available features and risk factors. The decision tree model was trained using the dataset, where each internal node represented a feature or attribute, each branch represented a decision rule, and each leaf node represented the predicted outcome. By analyzing the decision tree, important risk factors and their relationships with heart disease were identified. The trained Decision Tree model provided a clear decision-making process, allowing for predicting heart disease based on specific combinations of feature values. The model's performance was evaluated, as shown in Fig. 8 to assess its effectiveness in predicting heart disease.

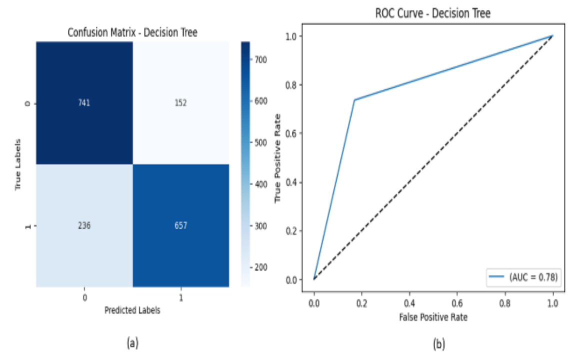

Figure 8. (a) Confusion Matrix of DT (b) ROC curve of DT

H. Multi-layer Perceptron

The Multilayer Perceptron (MLP) is a class of artificial neural networks characterized by multiple layers of interconnected nodes (neurons) that can acquire intricate patterns and relationships within the input data.

The MLP model was trained on the FHD using the available features and risk factors associated with heart disease. The algorithm acquired the ability to identify patterns and correlations within the dataset using both forward and backward propagation of signals. The Multilayer Perceptron's weights and biases underwent iterative adjustments throughout the training procedure to minimize the discrepancy between the predicted output and the predicted target values. Upon completing the training process, the Multilayer Perceptron (MLP) demonstrated the capability to generate predictions on novel, unobserved data by transmitting the input characteristics through the network and producing the corresponding output.

The MLP model's performance is assessed in Fig. 9 using suitable metrics, including accuracy, precision, recall, and F1 score, to evaluate its efficacy in precisely predicting heart disease in the FHD.

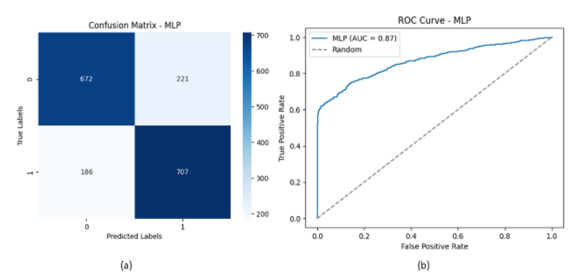

Figure 9. (a) Confusion Matrix of MLP (b) ROC curve of MLP

I. Random Forest

The Random Forest algorithm is a powerful machine learning methodology employed for predictive modelling. The RF algorithm utilizes decision trees to attain greater levels of predictive precision. The RF algorithm employs an ensemble technique that combines the predictions of multiple decision trees, each of which has been trained on a unique subset of the feature space. This results in robust and reliable outcomes, as depicted in Fig. 10. The algorithm employs the concept of feature randomness and voting to improve generalization and handle complex associations in the data.

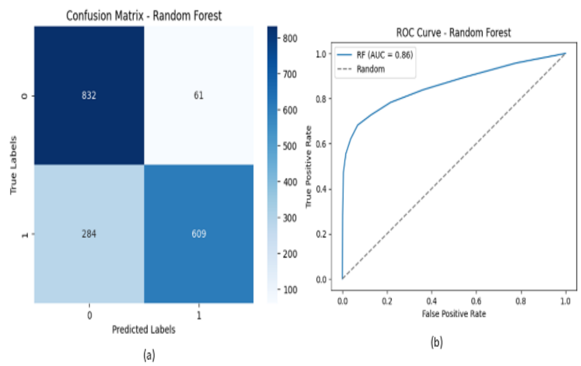

Figure 10. (a) Confusion Matrix of RF (b) ROC curve of RF

J. Hybrid Random Forest Linear Model

The HRFLM methodology is a hybridized approach that integrates the advantageous features of Random Forest and linear regression models to predict the incidence of heart disease outcomes in individuals.

The Random Forest component in the HRFLM is tasked with capturing complicated non-linear relationships and interactions among the dataset's features. Predictions are based on subsets of the data, and an ensemble of decision trees is used. This approach effectively enables the model to manage nonlinearity, feature interactions, and variable importance.

The HRFLM incorporates a linear regression component that employs a linear model, which integrates a set of chosen features from the Random Forest component. Utilizing a linear model facilitates the capture of linear associations and affords explication to the prognostications.

The HRFLM methodology initially conducts training of the Random Forest model on the FHD dataset to capture intricate patterns and discern significant features. Subsequently, a subset of pertinent features is chosen through a process of importance evaluation. These features have been utilized in constructing a linear regression model, as illustrated in Fig. 11.

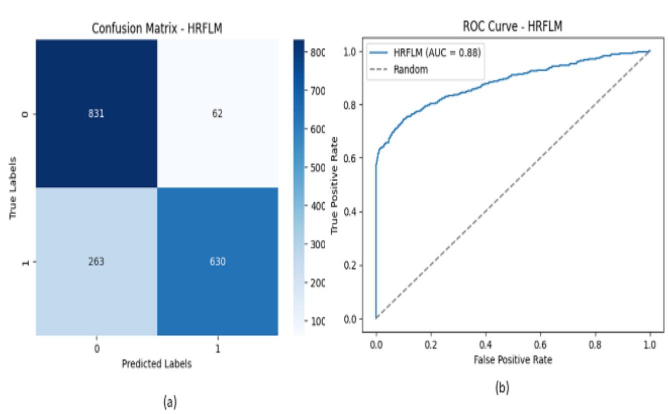

Figure 11. (a) Confusion Matrix of HRFLM (b) ROC curve of HRFLM

V RESULTS

The assessment of various algorithms on the FHD dataset revealed that the Hybrid Random Forest Linear Model (HRFLM) exhibited superior performance compared to the other algorithms in predicting heart disease outcomes. Table I, compares performance metrics, namely accuracy, precision, recall, and F1 score.

TABLE I
PERFORMANCE METRICS OF ALL ALGORITHMS

Algorithm	Accuracy	Precision	Recall	F1 Score
SVM	0.67	0.71	0.60	0.65
DT	0.71	0.72	0.70	0.71
MLP	0.75	0.75	0.75	0.75
KNN	0.68	0.69	0.68	0.68
RF	0.81	0.93	0.68	0.78
HRFLM	0.88	0.91	0.70	0.78

Figure 12 depicts the Receiver Operating Characteristic (ROC) curves of various algorithms employed to predict heart disease, presenting a comparative analysis of their respective performances. The Hybrid Random Forest Linear Model (HRFLM) algorithm is notable for its exceptional performance concerning the ROC curve compared to other algorithms. The curve of the HRFLM exhibits a greater degree of elevation and closer proximity to the top left corner of the plot, which suggests that it possesses superior proficiency in accurately categorizing positive and negative instances compared to the other algorithms. This shows that HRFLM exhibits a superior ability to correctly identify positive instances and a reduced tendency to incorrectly identify negative instances, thereby establishing its efficacy as a reliable predictive model for heart disease.

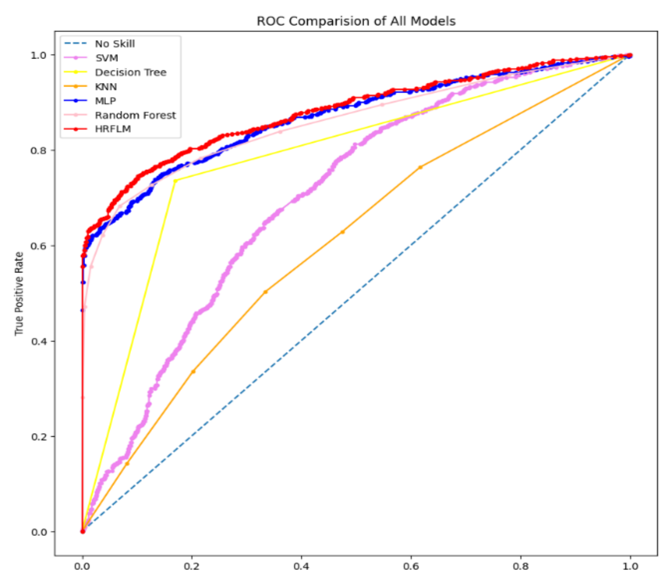

Figure 12. ROC Curve Comparison of all Algorithms

The ROC curve analysis reveals the potential of HRFLM as a valuable tool for heart disease prediction. Figure 13 depicts that the HRFLM algorithm outperformed other algorithms in accuracy, demonstrating its proficiency in categorizing individuals into their respective heart disease categories. Furthermore, the precision score of the HRFLM model demonstrated its proficiency in reducing the occurrence of false positive predictions, thereby increasing the certainty level in identifying individuals who are susceptible to heart disease. The recall metric indicated that HRFLM exhibited an outstanding ability to accurately detect positive instances, thereby reducing the likelihood of overlooking individuals affected with heart disease.

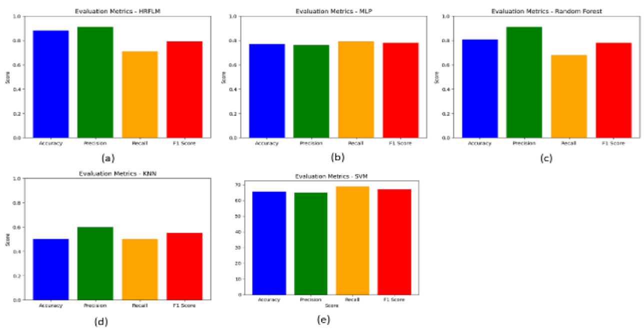

Figure 13. (a) Evaluation Metrics of HRFLM (b) Evaluation Metrics of MLP (c) Evaluation Metrics of RF (d) Evaluation Metrics of KNN (e) Evaluation Metrics of SVM

VI. CONCLUSION

Overall, the HRFLM algorithm performed exceptionally well predicting heart disease outcomes in the FHD. These results, as depicted in Fig. 13, demonstrate the HRFLM approach's potential as an effective tool for early detection and risk assessment of cardiovascular disease, which can contribute to enhanced patient care and prevention. Using the FHD, this study sought to develop a predictive model for heart disease.

Several machine learning algorithms, including SVM, KNN, Decision Tree, MLP, and Random Forest, were implemented and evaluated. According to the findings, the HRFLM algorithm outperformed other models regarding accuracy, precision, recall, and F1 score. Combining Random Forest and linear models yielded good results for predicting cardiovascular disease. In addition, feature selection and preprocessing techniques were utilized to enhance the model's efficacy.

The findings of this study have significant implications for the early detection and prevention of heart disease. Accurately predicting heart disease risk enables doctors to provide patients with appropriate treatments. Based on a person's health parameters and other relevant factors, the HRFLM algorithm provides a reliable method for predicting their risk of heart disease.

Nonetheless, it is essential to recognize the limitations of this investigation. The efficiency of the models may vary based on the used dataset and the specific population characteristics. Additional research and validation on larger and more diverse data sets are required to generalize the findings.

In conclusion, this study contributes to the field of heart disease diagnosis by demonstrating the accuracy of machine learning algorithms, specifically the HRFLM method, in accurately identifying individuals at risk for heart disease. The proposed model may aid doctors in making better decisions and enhancing patient outcomes. Future research can further investigate additional feature engineering techniques, ensemble methods, and larger datasets to improve the models' predictive capabilities.

REFERENCES

[1] A. Pierron, L. Fond-Harmant, A. Laurent, and F. Alla, "Supporting parenting to address social inequalities in health: a synthesis of systematic reviews," *BMC Public Health*, vol. 18, no. 1, p. 1087, Dec. 2018, doi: 10.1186/s12889-018-5915-6.

[2] *Cardiovascular Disability*. Washington, D.C.: National Academies Press, 2010. doi: 10.17226/12940.

[3] W. Frąk, A. Wojtasińska, W. Lisińska, E. Młynarska, B. Franczyk, and J. Rysz, "Pathophysiology of Cardiovascular Diseases: New Insights into Molecular Mechanisms of Atherosclerosis, Arterial Hypertension, and Coronary Artery Disease," *Biomedicines*, vol. 10, no. 8, p. 1938, Aug. 2022, doi: 10.3390/biomedicines10081938.

[4] H. S. Buttar, T. Li, and N. Ravi, "Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity and smoking cessation.," *Exp Clin Cardiol*, vol. 10, no. 4, pp. 229–49, 2005.

[5] A. Al-Jawaldeh and M. M. S. Abbass, "Unhealthy Dietary Habits and Obesity: The Major Risk Factors Beyond Non-Communicable Diseases in the Eastern Mediterranean Region," *Front Nutr*, vol. 9, Mar. 2022, doi: 10.3389/fnut.2022.817808.

[6] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, "A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms," *Mobile Information Systems*, vol. 2022, pp. 1–9, Mar. 2022, doi: 10.1155/2022/1410169.

[7] C. Cercato and F. A. Fonseca, "Cardiovascular risk and obesity," *Diabetol Metab Syndr*, vol. 11, no. 1, p. 74, Dec. 2019, doi: 10.1186/s13098-019-0468-0.

[8] V. G. Krishnan, Dr. M. V. V. Saradhi, Dr. S. S. Kumar, G. Dhanalakshmi, P. Pushpa, and Dr. V. Vijayaraja, "Hybrid Optimization based Feature Selection with DenseNet Model for Heart Disease Prediction," *International Journal of Electrical and Electronics Research*, vol. 11, no. 2, pp. 253–261, Apr. 2023, doi: 10.37391/ijeer.110203.

[9] S. E. Awan, M. Bennamoun, F. Sohel, F. M. Sanfilippo, and G. Dwivedi, "Machine learning-based prediction of heart failure readmission or death: implications of choosing the right model and the right metrics," *ESC Heart Fail*, vol. 6, no. 2, pp. 428–435, Apr. 2019, doi: 10.1002/ehf2.12419.

[10] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: 10.1016/j.neucom.2017.11.077.

[11] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, IEEE, Aug. 2020, pp. 729–735. doi: 10.1109/ICSSIT48917.2020.9214160.

[12] R. TR, U. K. Lilhore, P. M, S. Simaiya, A. Kaur, and M. Hamdi, "PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES," *Malaysian Journal of Computer Science*, pp. 132–148, Mar. 2022, doi: 10.22452/mjcs.sp2022no1.10.

[13] C. Militello, F. Prinzi, G. Sollami, L. Rundo, L. La Grutta, and S. Vitabile, "CT Radiomic Features and Clinical Biomarkers for Predicting Coronary Artery Disease," *Cognit Comput*, vol. 15, no. 1, pp. 238–253, Jan. 2023, doi: 10.1007/s12559-023-10118-7.

[14] N. L. Fitriyani, M. Syafrudin, G. Alfian, C. Yang, J. Rhee, and S. M. Ulyah, "Chronic Disease Prediction Model Using Integration of DBSCAN, SMOTE-ENN, and Random Forest," in *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, IEEE, Jun. 2022, pp. 289–294. doi: 10.1109/ICETSIS55481.2022.9888806.

[15] Naseer, Fawad, Muhammad Nasir Khan, Akhtar Rasool, and Nafees Ayub. "A novel approach to compensate delay in communication by predicting teleoperator behaviour using deep learning and reinforcement learning to control telepresence robot." Electronics Letters 59, no. 9, e12806, 2023.

[16] Naseer, Fawad, Muhammad Nasir Khan, and Ali Altalbe. "Intelligent Time Delay Control of Telepresence Robots Using Novel Deep Reinforcement Learning Algorithm to Interact with Patients." Applied Sciences 13, no. 4, 2462, 2023.

[17] Naseer, Fawad, Muhammad Nasir Khan, and Ali Altalbe. "Telepresence Robot with DRL Assisted Delay Compensation in IoT-Enabled Sustainable Healthcare Environment." Sustainability 15, no. 4, 3585, 2023.

[18] Altalbe, Ali, Muhammad Nasir Khan, Muhammad Tahir, and Aamir Shahzad. "Orientation Control Design of a Telepresence Robot: An Experimental Verification in Healthcare System." Applied Sciences 13, no. 11, 6827, 2023.

[19] Khan, Muhammad Nasir, Syed K. Hasnain, and Mohsin Jamil. Digital Signal Processing: A Breadth-first Approach. Stylus Publishing, LLC, 2016.

[20] Naseer, Fawad, Muhammad Nasir Khan, Akhtar Rasool, and Nafees Ayub. "A novel approach to compensate delay in communication by predicting teleoperator behaviour using deep learning and reinforcement learning to control telepresence robot." Electronics Letters 59, no. 9, e12806, 2023.

[21] A. Javaid *et al.*, “Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology,” *Am J Prev Cardiol*, vol. 12, p. 100379, Dec. 2022, doi: 10.1016/j.ajpc.2022.100379.

[22] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, “An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators,” *Healthcare Analytics*, vol. 2, p. 100118, Nov. 2022, doi: 10.1016/j.health.2022.100118.

[23] S. Uddin, S. Wang, H. Lu, A. Khan, F. Hajati, and M. Khushi, “Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics,” *Expert Syst Appl*, vol. 205, p. 117761, Nov. 2022, doi: 10.1016/j.eswa.2022.117761.

[24] L. Jahangiry, M. A. Farhangi, and F. Rezaei, “Framingham risk score for estimation of 10-years of cardiovascular diseases risk in patients with metabolic syndrome,” *J Health Popul Nutr*, vol. 36, no. 1, p. 36, Dec. 2017, doi: 10.1186/s41043-017-0114-0.

[25] M. Dehghan-Bonari, M. Alipour-Vaezi, M. M. Nasiri, and A. Aghsami, “A diagnostic analytics model for managing post-disaster symptoms of depression and anxiety among students using a novel data-driven optimization approach,” *Healthcare Analytics*, vol. 4, p. 100238, Dec. 2023, doi: 10.1016/j.health.2023.100238.

[26] I. H. Sarker, “AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems,” *SN Comput Sci*, vol. 3, no. 2, p. 158, Mar. 2022, doi: 10.1007/s42979-022-01043-x.

[27] J. Abdollahi and B. Nouri-Moghaddam, “A hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation,” *Iran Journal of Computer Science*, vol. 5, no. 3, pp. 229–246, Sep. 2022, doi: 10.1007/s42044-022-00104-x.

[28] B. Ayshwarya, M. Dhanamalar, and V. R. Sasikumar, “Heart Diseases Prediction Using Back Propagation Neural Network with Butterfly Optimization,” in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, IEEE, Feb. 2023, pp. 01–06. doi: 10.1109/ICECCT56650.2023.10179742.

[29] Y. Yan *et al.*, “Application of back propagation neural network model optimized by particle swarm algorithm in predicting the risk of hypertension,” *The Journal of Clinical Hypertension*, vol. 24, no. 12, pp. 1606–1617, Dec. 2022, doi: 10.1111/jch.14597.

[30] P. Mikalef and M. Gupta, “Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance,” *Information & Management*, vol. 58, no. 3, p. 103434, Apr. 2021, doi: 10.1016/j.im.2021.103434.

[31] A.-D. Samaras, S. Moustakidis, I. D. Apostolopoulos, N. Papandrianos, and E. Papageorgiou, “Classification models for assessing coronary artery disease instances using clinical and biometric data: an explainable man-in-the-loop approach,” *Sci Rep*, vol. 13, no. 1, p. 6668, Apr. 2023, doi: 10.1038/s41598-023-33500-9.

[32] S. Modak, R. K. Ghosh, and D. Sarkar, “Predictive Analysis of Heart Disease Using Machine Learning Algorithms,” 2023, pp. 151–165. doi: 10.4018/978-1-6684-7561-4.ch011.

[33] R. Kapila, T. Ragunathan, S. Saleti, T. J. Lakshmi, and M. W. Ahmad, “Heart Disease Prediction Using Novel Quine McCluskey Binary Classifier (QMBC),” *IEEE Access*, vol. 11, pp. 64324–64347, 2023, doi: 10.1109/ACCESS.2023.3289584.

[34] S. M. Muhammed, G. Abdul-Majeed, and M. S. Mahmoud, “Early Prediction of Chronic Heart Disease Based on Electronic Triage Dataset by using Machine Learning,” in *2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, IEEE, Jul. 2023, pp. 131–136. doi: 10.1109/AICCIT57614.2023.10218241.

[35] D. Yewale, S. P. Vijayaragavan, and V. K. Bairagi, “An Effective Heart Disease Prediction Framework based on Ensemble Techniques in Machine Learning,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.0140223.