Indonesian Journal of Electrical Engineering and Computer Science
Vol. 34, No. 2, May 2024, pp. 1122~1130
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v34.i2.pp1122-1130

Heart disease prediction using ML through enhanced feature engineering with association and correlation analysis

Annemneedi Lakshmanarao[1], Thotakura Venkata Sai Krishna[2], Tummala Srinivasa Ravi Kiran[3], Chinta Venkata Murali krishna[4], Samsani Ushanag[5], Nandikolla Supriya[6]
[1]Department of IT, Aditya Engineering College, Surampalem, India
[2]Department of CSE-Data Science, QIS College of Engineering and Technology (Autonomous), Ongole, India
[3]Department of Computer Science, P.B. Siddhartha College of Arts and Science Vijayawada, India
[4]Department of CSE (Data Science), NRI Institute of Technology, Pothavarappadu, India
[5]Department of CSE, University College of Engineering Kakinada, JNTUK Kakinada, India
[6]Department of CSE (CS), School of Engineering, Malla Reddy University, Hyderabad, India

Article Info

Article history:

Received Jan 3, 2024
Revised Feb 16, 2024
Accepted Feb 20, 2024

Keywords:

Association analysis
Correlation analysis
Feature engineering
Heart disease
Kaggle
Machine learning

ABSTRACT

Heart disease remains a prevalent and critical health concern globally. This paper addresses the critical task of heart disease prediction through the utilization of advanced machine learning techniques. Our approach focuses on the enhancement of feature engineering by incorporating a novel integration of association and correlation analyses. A heart disease dataset from Kaggle was used for the experiments. Association analysis was applied to the categorical and binary features in the dataset. Correlation analysis was applied to the numerical features in the dataset. Based on the insights from association analysis and correlation analysis, a new dataset was created with combinations of features. Later, newly created features are integrated with the original dataset, and classification algorithms are applied. Five machine learning (ML) classifiers, namely decision tree, k-nearest neighbors (KNN), random forest, XG-Boost, and support vector machine (SVM), were applied to the final dataset and achieved a good accuracy rate for heart disease detection. By systematically exploring associations and relationships with categorical, binary, and numerical features, this paper unveils innovative insights that contribute to a more comprehensive understanding of the heart disease dataset.

Corresponding Author:

Annemneedi Lakshmanarao
Department of IT, Aditya Engineering College
Surampalem, India
Email: laxman1216@gmail.com

1. INTRODUCTION

Heart disease, a prevalent and critical health concern worldwide, encompasses a range of conditions affecting the heart and blood vessels. It remains a leading cause of morbidity and mortality, necessitating advanced predictive models for timely diagnosis and intervention. Risk factors for heart disease are multifaceted, including lifestyle choices, clinical markers, and demographic variables. Coronary artery disease is the most common type of heart disease. It transpires when the blood vessels (coronary arteries) that supply blood to the heart muscle become narrowed or blocked, leading to a heart attack. Heart failure can result from various conditions, including coronary artery disease and hypertension. Valvular heart disease (VHD) involves damage to one of the four heart valves. VHD can disrupt the flow of blood through the heart. Some individuals may have a combination of all the conditions and risk factors, such as age, genetics, lifestyle choices, and other medical conditions. Early detection, lifestyle modifications, and medical management play crucial roles in the prevention and treatment of heart disease.

This paper addresses the imperative task of heart disease prediction through the integration of advanced machine learning (ML) techniques, emphasizing enhanced feature engineering. By systematically exploring associations among diverse features. This paper aims to develop an efficient heart disease prediction model.

García-Ordás *et al.* [1], the authors recommended applying deep learning methods along with feature augmentation techniques to evaluate the risk of cardiovascular disease in patients. The proposed methods demonstrated enhanced performance, exceeding other state-of-the-art approaches by 4.5%. This improvement was noteworthy, especially given its achievement of a precision of 91%, marking a significant advancement, particularly for a health condition affecting a substantial population. Chandrasekhar and Peddakrishna [2] utilized ML to enhance heart disease prediction. Employing six ML methods on Cleveland and IEEE Dataport datasets, the study optimized the model through GridsearchCV and 5-fold cross-validation. Logistic regression excelled with 90% accuracy in Cleveland, while AdaBoost achieved 91% accuracy in IEEE Dataport. The introduction of a soft voting ensemble classifier, incorporating all six methods, elevated accuracy to 94% for the Cleveland dataset and 95.5% for the IEEE Dataport dataset.

Subramani *et al.* [3] study presented a suite of ML models designed to address a specific problem, incorporating diverse data observation methods and training procedures from various algorithms. To validate the efficacy of the strategy, the Heart Dataset was amalgamated with other classifiers. The proposed method exhibited an accuracy of nearly 95%, outperforming existing methods, and underwent a thorough analysis across multiple metrics. Srinivasan *et al.* [4] used eight ML classifiers to improve heart disease prediction. Neural network models achieved 95%, NB given, and 91% accuracy using various feature combinations and well-known methods. Clinical test data was used to predict heart failure using the ML metamodel [5]. A metamodel was created with random forest, decision tree, k-nearest neighbors (K-NN) as the final estimator. A heart disease dataset with 11 standard features was used in the experiments. A metamodel with 88% accuracy beats other ML models. Rajkumar *et al.* [6] used the Hungarian heart disease dataset from IoT sensor devices to offer an upgraded deep learning framework for heart disease prediction. The dataset was preprocessed using median studentized residual and feature selected using harris hawk optimization (HHO). Modified deep long short-term memory (MDLSTM) identified characteristics as normal or abnormal, while improved spotted hyena optimization (ISHO) changed long short-term memory (LSTM) output. Using Python, the findings showed better prediction accuracy (98%) and lower error rate than previous methods. Gopalakrishnan *et al.* [7], the authors evaluated the accuracy of cardiac diagnosis accuracy. Recursive partitioning greedily reordered partitions without considering optimality. Through preprocessing, feature extraction, and classification, a CNN predicted sickness. Reduced dimension for more accurate dataset predictions. AI classification systems often lagged behind Lasso or Ridge regression, which regularly outperformed them.

Vayadande *et al.* [8], various ML and deep learning (DL) algorithms were implemented and tested on a Kaggle dataset with 15 features. The algorithms included logistic regression, K-NN, SVC, multilayer perceptron network (MLP), DT, and RF. Nagavelli *et al.* [9] mentioned heart disease detection using ML technology. The research looked at Naïve Bayes with a weighted approach, frequency, time, and information theory classifiers for finding ischemic heart disease and better support vector machine (SVM) for finding heart failure. The research used XGBoost to help medics diagnose cardiac problems early. Rindhe *et al.* [10], authors predicted cardiac illnesses using data mining, ML, neural networks, random forest and SVC. Heart disease prediction and diagnosis were tough for physicians and hospitals in India and overseas. A fast and effective diagnostic method was needed to reduce heart disease fatalities. Lakshmanarao *et al.* [11], the authors applied ensemble learning techniques for heart disease prediction and obtained good results. Two types of ensembles, namely stacking and voting classifiers, are used in their work. A logistic regression classifier, grid-based solver hyperparameter tweaking, and tenfold cross-validation were used to enhance heart disease prediction in [12]. The dataset was taken from the UCI Machine Learning repository to test the model. Experimental findings indicated a 91% performance rate, outperforming numerous models. Mohan *et al.* [13] predicted heart diseases using ML models. Data on various human health factors is utilized for training and testing. Many AI and ML systems anticipate heart diseases. The ML algorithm's performance is compared after implementation. A diagnostic assistance system for analyzing the main cardiovascular risk factors, such as age, gender, and high blood pressure, was suggested in [14]. It was built on optimal machine learning algorithms, including ANN, SVC, KNN, Naïve Bayes, and decision tree. The machine learning models were trained and tested with a medical dataset including over 550 individuals diagnosed with atherosclerosis. The 97% accuracy was found to be a promising threshold for atherosclerosis prediction using ANN. Krittanawong *et al.* [15] tested ML models for cardiovascular disease prediction. A complete MEDLINE, Embase, and Scopus search found 103 cohorts with more than 3,00,000 people. SVM had a pooled AUC of 0.93 for stroke prediction and 0.89 for coronary artery disease. Pasha *et al.* [16] examined a Kaggle dataset with heart disease-related variables including age, gender, blood pressure, and cholesterol. They tested SVM, KNN, and decision tree accuracy. When run on a huge dataset, these methods performed poorly.

Lakshmanarao *et al.* [17], the authors applied conventional ML algorithms with sampling techniques and achieved good results. Peteti and Nandan [18] summarized past studies on heart disease prediction, highlighting improved performance using KNN, ANN, and GA algorithms in various scenarios and achieved good detection rate chronic heart failure (CHF), which causes significant mortality and medical expenses, was diagnosed by ECG in [19]. A unique adaptive filter using a delayed error normalized LMS method was devised for efficient preprocessing. R-peak detection, HRV feature extraction, and CNN-GRU-AM model training achieved 99.8% accuracy, beating CNN (94%) and GRU (92%). Srinivas *et al.* [20] developed an ACLS-RCNN and ICSOA-based intelligent heart disease prediction system. Feature extraction and dimensionality reduction improved prognosis accuracy, while preprocessing fixed missing values and skewed data. In accuracy, precision, recall, and f1-score, the suggested approach surpassed known methods. The authors predicted heart disease risk variables using K-means using public data in [21]. Pre-processing, classifier performance, and assessment metrics were reviewed for 209 records and 8 variables, including age and blood pressure. The displayed results showed correct predictions. Boukhatem *et al.* [22], MLP, SVM, RF, and NB were used for data preprocessing and feature selection in prediction models. SVM's accuracy was 91.67%, demonstrating its exceptional performance. A weighted associative rule mining approach was presented in [23] to quantify heart disease prediction feature strength. The research predicted heart disease using these major factors and found satisfactory results. Data mining improves heart disease diagnosis and therapy [24]. The correlation-based feature subset selection technique identified heart disease predictors such as age, gender, smoking, obesity, food, and more. Random forest with selected characteristics provided the highest accuracy (90%), suggesting early prediction potential.

2. METHOD

The proposed method for heart disease prediction was shown in Figure 1. In this paper, a robust methodology was devised to improve heart disease detection by synergistically combining association and correlation analyses. Initially, a dataset is collected from the Kaggle repository. The dataset comprises categorical features, binary indicators, and numerical features. The Apriori algorithm was used to find common item sets and rules within categorical and binary features. This showed complex relationships that were indicative of heart disease. Concurrently, correlation analysis was employed on numerical features, unraveling associations between pairs of variables crucial for predictive insights. Later, an innovative feature engineering process was applied. Based on the revelations from association and correlation analyses, a set of new features was crafted. Categorical features involved the creation of combined indicators that captured joint occurrences, while averages were used to turn numerical features into categorical features that captured behaviors that were related. These newly engineered features were seamlessly integrated with the original dataset, resulting in a comprehensive, feature-rich dataset ready for machine learning model development. In the next phase, several classification algorithms, namely k nearest neighbors, support vector classifier, decision tree, random forest and XGBoost applied to the final dataset and evaluated the performance in heart disease prediction.

2.1. Data collection

In this paper, a dataset from Kaggle [25] was used for experiments. The dataset contains 11 input features. The features with types are Age (Numerical) sex (categorical), ChestPain_Type (Catogorical), RestingBP (Numerical), Cholesterol (Numerical), FastingBS (Numerical), RestingECG (Categorical), MaxHR (numerical), ExerciseAngina (Categorical), Oldpeak (Numerical), ST_Slope (Categorical). So, there are six numerical attributes and five categorical attributes in the given dataset.

2.2. Applying association analysis for categorical and binary features

There are five categorical attributes in the given dataset. The categorical attributes are sex, ChestPain_Type, RestingECG, Exercise Angina (exercise-induced angina), ST_Slope (the slope of the peak exercise ST segment). These five attributes are applied with Apriori algorithm. After applying apriori algorithm, it generates several association rules. The setting for generating rules in apriori is constrained by several conditions. Here, min_support=0.1 is used as association rule for generating rules. With this condition, it produced 70 rules. But we considered only few rules. The rules are considered based on the criteria used for creating new features. In this paper, new features from association analysis created in two different ways namely creation of binary features, creation of count-based features.

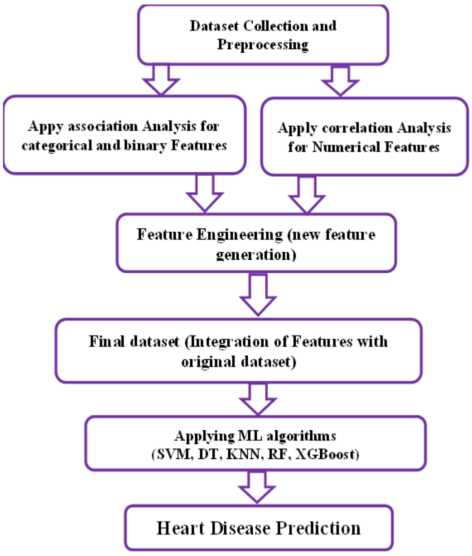

Figure 1. Proposed methodology

2.2.1. Creating binary features from association rules

Binary features are created that represent the presence or absence of specific combinations of categorical or binary features. For example, if an association rule indicates that when A and B are present, C is likely to be present, you can create a new binary feature 'A_and_B' that is 1 when both A and B are present and 0 otherwise. Table 1 shows the details of the association rules used for creating binary features.

2.2.2. Creating count-based features from association rules

Count-based features are created by utilizing the support values from association rules. If the association rule had high support, then we created a feature counting the occurrences of that itemset. In this way, four features are created. The details of features created along with association rules are in Table 2.

2.3. Applying correlation analysis for numerical features

Next, Correlation analysis applied with numerical features in the dataset. There are five categorical attributes in the given dataset. Numerical features in the dataset are Age, RestingBP, Cholesterol, FastingBS, MaxHR, Oldpeak.

2.3.1. Creating features through interaction terms from correlation analysis

After applying correlation analysis, it is observed that Age and Cholesterol have a moderate positive correlation, indicating that as Age increases, Cholesterol levels also tend to increase. MaxHR and Oldpeak have a moderate negative correlation, indicating that as MaxHR increases, Oldpeak tends to decrease. For correlated numerical features, we created new features representing interactions between them. This can capture nonlinear relationships. These features are shown in Table 3.

Table 1. Binary feature creation from association rules

Antecedents	Consequents	Confidence	Created binary feature
({'RestingECG_Normal', 'ST_Slope_Up', 'ChestPainType_ATA'})	({'ExerciseAngina_N'})	0.96	C1
({'ST_Slope_Up', 'ChestPainType_ATA'})	({'ExerciseAngina_N'})	0.95	C2
({'ST_Slope_Up', 'Sex_F'})	({'ExerciseAngina_N'})	0.927	C3
({'ST_Slope_Up', 'ChestPainType_NAP'})	({'ExerciseAngina_N'})	0.923	C4
({'FastingBS', 'ST_Slope_Flat'})	({'Sex_M'})	0.921	C5
({'RestingECG_Normal', 'ChestPainType_ATA'})	({'ExerciseAngina_N'})	0.91	C6
({'RestingECG_ST', 'ChestPainType_ASY'})	({'Sex_M'})	0.90	C7
({'ChestPainType_ASY', 'FastingBS'})	({'Sex_M'})	0.90	C8

Table 2. Count based feature creation from association rules

Antecedents	Consequents	Confidence	Created binary feature
({'RestingECG_Normal'})	({'Sex_M'})	0.4727	1.RestingECG_Normal_count 2. Sex_M_count
({'ChestPainType_ASY'})	({'Sex_M'})	0.464	3.ChestPainType_ASY_count
({'ExerciseAngina_N'})	({'Sex_M'})	0.927	4.ExerciseAngina_N_Sex_M_count

Table 3. Creating features from interacted terms in correlation analysis

F1	F2	Created Features
Age	Cholesterol	interaction_Age_Clho
MaxHR	Oldpeak	interaction_MaxHR_Oldpeak

2.3.2. Creating features by applying transformations to numerical features based on their correlation

In this step, correlation is considered for creating features. For instance, if features X and Y are positively correlated, consider creating a feature that represents their average. In this dataset, Age and Cholesterol are positively correlated. So, a new feature is created by taking the average of these two. Similarly, MaxHR and Oldpeak are negatively correlated. So, a new feature is created by taking the average of these two. Table 4 shows new features created from this analysis.

Table 4. Creating features by transform numerical features in correlation analysis

F1	F2	Created features
Age	Cholesterol	average _Age_Clho
MaxHR	Oldpeak	average_MaxHR_Oldpeak

2.3.3. Creating features ration features based on their correlation

Here, features are created by taking the ratio of two correlated numerical features. In this dataset two new features namely ratioMaxHR_Oldpeak, ratio_Age_Clho are created in this manner. The details are shown in Table 5.

Table 5. Creating features from interacted terms in correlation analysis

F1	F2	Created features
Age	Cholesterol	Ratio_Age_Clho
MaxHR	Oldpeak	Ratio_MaxHR_Oldpeak

3. RESULTS AND DISCUSSION

After creating a final dataset from association analysis and correlation analysis, several ML classification algorithms proposed on final dataset. All the experiments are conducted in google Collaboratory environment. Google colab provides sufficient resources for executing Ml experiments.

3.1. Apply ML algorithms with original dataset with 11 features

Five ML classifiers namely KNN, support vector machine, decision tree, random forest and XGBoost applied on final dataset. To compare the performance of the proposed model, these five classifiers also applied with original dataset without new features. Table 6 and Figure 2 shows results of experiments with ML models on original dataset with 11 features.

From Table 6 and Figure 2, it is evident that, except for KNN, all the remaining algorithms given more than 80% accuracy. KNN given only 74% accuracy. Random forest and XGboost algorithms given good accuracy values. The precision, recall and f1-score values also good for XGboost and RF. After these two models, Decision tree performed well with an accuracy of 85% and good precision, recall and f1-score measures. Next, SVC given 80.4% accuracy with moderate recall value of 81%.

Table 6. Results with original dataset

Algorithm	Class	Precision	Recall	F1-Score	Accuracy
KNN	NO	73%	76%	75%	74%
	YES	76%	73%	74%	
SVC	NO	86%	73%	79%	80.4%
	YES	77%	88%	82%	
Decision tree	NO	86%	82%	84%	85%
	YES	84%	87%	85%	
Random forest	NO	90%	87%	88%	89%
	YES	88%	90%	89%	
XGBoost	NO	90%	87%	88%	89%
	YES	80%	90%	89%	

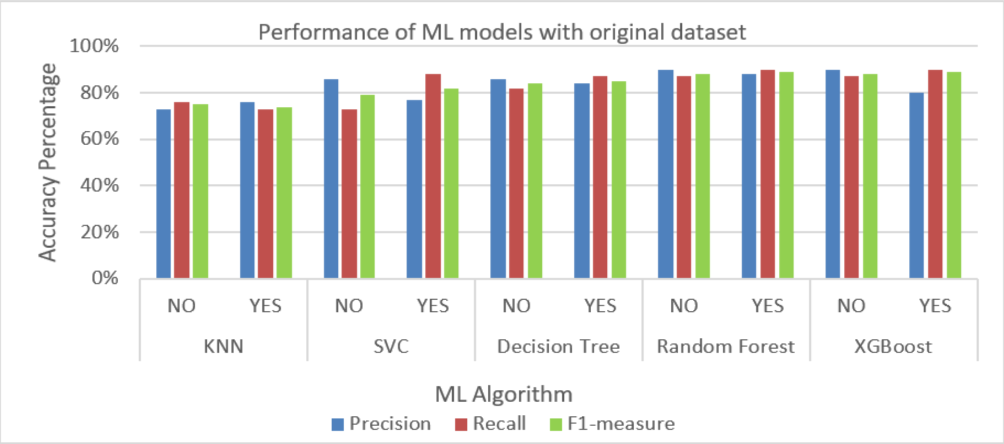

Figure 2. Precision, recall, F1-score of ML algorithms

3.2. Apply ML models with modified dataset

After applying the ML model to the original dataset, it is identified that RF and XGboost given best accuracy of 89%. Later, we applied these five classifiers with new dataset generated from association and correlation analysis. The dataset contains 29 features with 918 samples. It is divided into train and test sets in a split ratio of 80:20. The number of samples in train and test sets are 734, 184 respectively. Later, the same ML models are applied and the results are tabulated. The precision, recall, and f1-score results with proposed dataset are shown in Table 7 and Figure 3. After applying the ML model to the proposed dataset, it was identified that all the algorithms improved significantly with the new dataset.

Table 7. Results with proposed dataset

Algorithm	Class	Precision	Recall	F1-score	Accuracy
KNN	NO	85%	76%	83%	84%
	YES	84%	87%	84%	
SVC	85%	84%	87%	85%	85%
	86%	85%	81%	84%	
Decision Tree	NO	90%	89%	91%	90%
	YES	89%	91%	90%	
Random forest	NO	94%	95%	94%	95%
	YES	95%	96%	92%	
XGBoost	NO	95%	94%	94%	96%
	YES	95%	94%	93%	

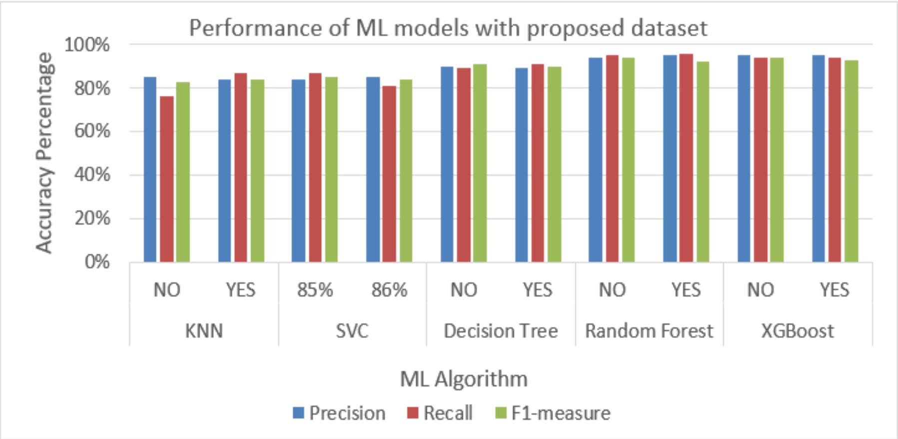

Figure 3. Precision, recall, F1-measure with proposed dataset with 29 features

3.3. Comparison of proposed dataset with original dataset

After applying ML models with original and proposed datasets, it is evident that association analysis and correlation analysis played a vital role in improving the accuracy of the ML algorithm. All five algorithms improved performance in all performance metrics. The accuracy comparison between two datasets with all five algorithms is shown in Figure 4 and Table 8.

Table 8. Comparsion with previous work

Algorithm	Accuracy with original dataset	Accuracy with new dataset
KNN	74%	84%
SVC	80.4%	85%
DTC	85%	90%
RF	89%	95%
XGB	89%	96%
Algorithm	Accuracy with original dataset	Accuracy with new dataset

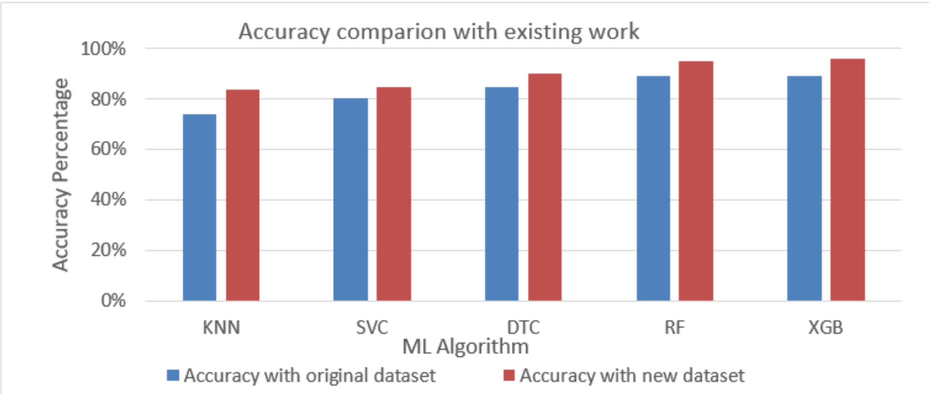

Figure 4. Comparison between two datasets

4. CONCLUSION

This paper pioneers a comprehensive approach to heart disease prediction, leveraging advanced ML techniques and a novel feature engineering strategy. The original dataset, featuring 11 attributes, underwent a transformative enhancement through association and correlation analyses. Association analysis contributed 8 binary features and 4 count features, illuminating intricate relationships among categorical and binary variables. Correlation analysis further enriched the dataset with six features, capturing nuanced interactions within numerical variables. The impact of these enhancements on predictive modeling was substantial. When employing the classic machine learning algorithms KNN, SVC, DTC, RF, and XGB on the original dataset, respectable accuracy was achieved, ranging from 74% to 89%. However, the proposed dataset, crafted through association and correlation analyses, significantly elevated predictive performance. The refined dataset yielded notable improvements, with accuracies reaching 84% to 96% across the same algorithms. By synergizing association and correlation analyses within a ML framework, our methodology not only elevates predictive accuracy but also deepens understanding of the intricate factors contributing to heart disease. The addition of novel features through association and correlation analyses empowers the models to capture subtle patterns, showcasing the potential for improved clinical insights.

REFERENCES

[1] M. T. García-Ordás, M. Bayón-Gutiérrez, C. Benavides, J. Aveleira-Mata, and J. A. Benítez-Andrades, "Heart disease risk prediction using deep learning techniques with feature augmentation," *Multimedia Tools and Applications*, vol. 82, no. 20, pp. 31759–31773, Aug. 2023, doi: 10.1007/s11042-023-14817-z.

[2] N. Chandrasekhar and S. Peddakrishna, "Enhancing heart disease prediction accuracy through machine learning techniques and optimization," *Processes*, vol. 11, no. 4, p. 1210, Apr. 2023, doi: 10.3390/pr11041210.

[3] S. Subramani *et al.*, "Cardiovascular diseases prediction by machine learning incorporation with deep learning," *Frontiers in Medicine*, vol. 10, Apr. 2023, doi: 10.3389/fmed.2023.1150933.

[4] S. Srinivasan, S. Gunasekaran, S. K. Mathivanan, B. A. M. M. B, P. Jayagopal, and G. T. Dalu, "An active learning machine technique-based prediction of cardiovascular heart disease from UCI-repository database," *Scientific Reports*, vol. 13, no. 1, p. 13588, Aug. 2023, doi: 10.1038/s41598-023-40717-1.

[5] I. Mahmud, M. M. Kabir, M. F. Mridha, S. Alfarhood, M. Safran, and D. Che, "Cardiac failure forecasting based on clinical data using a lightweight machine learning metamodel," *Diagnostics*, vol. 13, no. 15, p. 2540, Jul. 2023, doi: 10.3390/diagnostics13152540.

[6] G. Rajkumar, T. Gayathri Devi, and A. Srinivasan, "Heart disease prediction using IoT based framework and improved deep learning approach: Medical application," *Medical Engineering & Physics*, vol. 111, p. 103937, Jan. 2023, doi: 10.1016/j.medengphy.2022.103937.

[7] S. Gopalakrishnan, M. S. Sheela, K. Saranya, and J. J. Hephzipah, "A novel deep learning-based heart disease prediction system using convolutional neural networks (CNN) algorithm," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 10s, pp. 516–522, 2023.

[8] K. Vayadande *et al.*, "Heart disease prediction using machine learning and deep learning algorithms," in *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, IEEE, May 2022, pp. 393–401. doi: 10.1109/CISES54857.2022.9844406.

[9] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine learning technology-based heart disease detection models," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–9, Feb. 2022, doi: 10.1155/2022/7351061.

[10] B. U. Rindhe, N. Ahire, R. Patil, S. Gagare, and M. Darade, "Heart disease prediction using machine learning," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 267–276, May 2021, doi: 10.48175/IJARSCT-1131.

[11] A. Lakshmanarao, A. Srisaila, and T. S. R. Kiran, "Heart disease prediction using feature selection and ensemble learning techniques," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, Feb. 2021, pp. 994–998. doi: 10.1109/ICICV50876.2021.9388482.

[12] N. Narisetty, A. Kalidindi, M. V. Bujaranpally, N. Arigela, and V. V. Ch, "Ameliorating heart diseases prediction using machine learning technique for optimal solution," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 16, pp. 156–165, Nov. 2023, doi: 10.3991/ijoe.v19i16.42071.

[13] N. Mohan, V. Jain, and G. Agrawal, "Heart disease prediction using supervised machine learning algorithms," in *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, IEEE, Oct. 2021, pp. 1–3. doi: 10.1109/ISCON52037.2021.9702314.

[14] A. El-Ibrahimi, O. Terrada, O. El Gannour, B. Cherradi, A. El Abbassi, and O. Bouattane, "Optimizing machine learning algorithms for heart disease classification and prediction," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 15, pp. 61–76, Oct. 2023, doi: 10.3991/ijoe.v19i15.42653.

[15] C. Krittanawong *et al.*, "Machine learning prediction in cardiovascular diseases: a meta-analysis," *Scientific Reports*, vol. 10, no. 1, p. 16057, Sep. 2020, doi: 10.1038/s41598-020-72685-1.

[16] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, "Cardiovascular disease prediction using deep learning techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 981, no. 2, p. 022006, Dec. 2020, doi: 10.1088/1757-899X/981/2/022006.

[17] A. Lakshmanarao, Y. Swathi, and P. S. S. Sundareswar, "Machine learning techniques for heart disease prediction," *International Journal of Scientific and Technology Research*, vol. 8, no. 11, pp. 374–377, Nov. 2019.

[18] B. S. Peteti and D. Nandan, "Heart disease classification/prediction: a review," *Revue d'Intelligence Artificielle*, vol. 37, no. 2, pp. 347–377, Apr. 2023, doi: 10.18280/ria.370213.

[19] I. Khan, S. P. Maniraj, K. S. Reddy, V. Balaji, K. Kalaivani, and M. Singh, "Congenital heart disease prediction based on hybrid approach of CNN-GRU-AM," in *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, May 2023, pp. 613–618. doi: 10.1109/ICICCS56967.2023.10142526.

[20] R. Srinivas, R. kiran Bagadi, T. Rama Reddy, N. Praveen, and G. Aparanjini, "An efficient hybrid optimization algorithm for detecting heart disease using adaptive stacked residual convolutional neural networks," *Biomedical Signal Processing and Control*, vol. 87, p. 105522, Jan. 2024, doi: 10.1016/j.bspc.2023.105522.

[21] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart disease prediction using exploratory data analysis," *Procedia Computer Science*, vol. 173, pp. 130–139, 2020, doi: 10.1016/j.procs.2020.06.017.

[22] C. Boukhatem, H. Y. Youssef, and A. B. Nassif, "Heart disease prediction using machine learning," in *2022 Advances in Science and Engineering Technology International Conferences (ASET)*, IEEE, Feb. 2022, pp. 1–6. doi: 10.1109/ASET53988.2022.9734880.

[23] A. Yazdani, K. D. Varathan, Y. K. Chiam, A. W. Malik, and W. A. Wan Ahmad, "A novel approach for heart disease prediction using strength scores with significant predictors," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 194, Dec. 2021, doi: 10.1186/s12911-021-01527-5.

[24] M. I. Hossain *et al.*, "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison," *Iran Journal of Computer Science*, vol. 6, no. 4, pp. 397–417, Dec. 2023, doi: 10.1007/s42044-023-00148-7.

[25] Fedesoriano, "Heart failure prediction dataset." [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction. Accessed on November 05, 2023.

BIOGRAPHIES OF AUTHORS

Annemneedi Lakshmanarao currently working as Associate Professor in Aditya Engineering College, Surampalem. He completed his B. Tech in CSIT and M. Tech in Software Engineering. He is pursuing Ph.D. in Andhra University, Vishakapatnam. His areas of interest are machine learning, cyber security, and deep learning. He is a life member of Computer Society of India (CSI). He can be contacted at email: laxman1216@gmail.com.

Dr. Thotakura Venkata Sai Krishna currently working as Professor in QIS College of Engineering and Technology, Ongole. He completed his B. Tech in CSE and M. Tech in CSE. He completed Ph.D. in JNTUK, Kakinada. His areas of interest are ML, AI, and deep learning. He is a Life Member of Computer Society of India (CSI). He can be contacted at email: tvsai.kris@gmail.com.

Dr. Tummala Srinivasa Ravi Kiran currently working as Associate Professor & HOD in Department of Computer Science, PB Siddhartha College of Arts and Science, Vijayawada. He completed his Ph.D. in Acharya Nagarjuna University. His areas of interest are machine learning, databases, and cyber security. He has published research papers in various conferences and journals. He can be contacted at email: tsravikiran@pbsiddhartha.ac.in.

Chinta Venkata Murali krishna currently working as Associate Professor and HOD in CSE (Data Science) department at NRI Institute of Technology. He is a member of IAENG, IFERP, and INSC. He completed his M.Tech. in Computer Science & Engineering in 2009 and is currently pursuing a Ph.D. in Computer Science & Engineering at GITAM (Deemed to be University), Vishakapatnam. He has published research papers in various conferences and journals and has been granted three patents with ten others in the pipeline for the grant. Four of his books have been published by international and national publishing agencies. He was awarded the “Best Researcher Award” from IOSRD in 2018. He can be contacted at email: muralikrishna_chinta2007@yahoo.co.in.

Samsani Ushanag currently working as Associate Professor(c) in CSE Department, University College of Engineering (UCEK), JNTUK Kakinada. She completed her B. Tech in CSE and M. Tech in Computer Networks from Andhra University, Vishakhapatnam. She is pursuing Ph.D. in University College of Engineering (UCEK), JNTUK Kakinada. Her areas of interest are ML, data science, computer networks, cyber security, and deep learning. She can be contacted at email: ushavasi582010@gmail.com.

Dr. Nandikolla Supriya currently working as Associate Professor, in the department of Computer Science and Engineering (CS&IOT), Malla Reddy University, Hyderabad. She obtained her M.Tech (CSE) and Ph.D (CSE) from ANU, Guntur. She published several papers in National and International Journals. She is active member of various professional bodies. Her current research is focused on software reliability, predictive analytics and cloud computing. She can be contacted at email: nsupriyacse@gmail.com.