INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Heart Disease Prediction Using CNN with Various Feature Selection Approaches

S V Remya[1]*

1 Process Associate, Nyeste Venture Technologies Pvt. Ltd., Trivandrum, Kerala, India

OPEN ACCESS

Received: 23-04-2024

Accepted: 03-08-2024

Published: 08-10-2024

Citation: Remya SV (2024) Heart Disease Prediction Using CNN with Various Feature Selection Approaches. Indian Journal of Science and Technology 17(38): 3960-3968. https://doi.org/10.17485/IJST/v17i38.1360

* **Corresponding author**.

remya.anetta@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To evaluate the performance of CNN models with feature selection methods like Relief, Uniform Manifold Approximation and Projection (UMAP), and Linear discriminant Analysis (LDA) for forecasting heart diseases. **Methods:** The present research for heart disease prediction compares the performance of feature selection algorithms like ReliefF, Uniform Manifold Approximation and Projection (UMAP), and Linear discriminant Analysis (LDA) with Convolution Neural networks (CNN) for prediction. The study is conducted in a dataset with 303 records collected from patients with 14 attributes. It is also validated with the publicly available Cleveland dataset (Kaggle). The software environment used for implementation is Jupyter Notebook, which uses Python. The dataset consists of 303 records collected from the patients with 14 attributes. It is validated with the publicly available Cleveland dataset. **Findings:** The study examines how these feature selection techniques affect CNN's accuracy. According to experimental findings, the CNN-UMAP hybrid model outperforms with an accuracy of 91%, precision of 88%, and recall of 85% compared to ReliefF and LDA. UMAP shows up as the most successful feature selection method among the studied techniques when utilized alongside CNNs. **Novelty:** ReliefF, UMAP, and LDA allow CNN to learn more significant and discriminative features by minimizing the dimensions of the input data while preserving its underlying pattern. Previous studies also attempted to identify the key contributing characteristics to heart disease prediction, but less emphasis was placed on these feature selection methods in determining the effectiveness of the features for heart disease prediction.

Keywords: Heart Disease; Feature Selection Methods; Convolutional Neural Network; ReliefF; UMAP and LDA

1 Introduction

The rate of cardiovascular diseases is rising internationally, and the World Health Organization (WHO) projects that 17 million people die each year from heart disorders, mainly heart attacks and strokes. It is anticipated that the mortality rate from heart disease will increase to 22 million individuals by 2030 [(1)]. Patient's survival rate can be significantly improved with quick and precise diagnosis of specific conditions that cause

heart diseases.

Machine Learning (ML) approaches have recently demonstrated promising results in predicting cardiac disorders with the help of numerous patient data [2]. Complicated patterns can now be learned directly from raw data using Deep Learning (DL) methods, mainly using Convolutional Neural Networks (CNN), which makes them ideal for tasks like image-based analysis and diagnosis. To determine the crucial features, feature selection methods are employed to exclude redundant and unnecessary features, which influence the outcome of the CNN prediction. Recent studies show various feature selection methods for the prediction of heart diseases, and they need to be improved further for accurate prediction [3].

The goal of this study is to explore and examine the existing feature models in order to contribute to the development of robust feature selection algorithms. This study focuses on analyzing sophisticated feature selection approaches such as ReliefF [4], UMAP [5], and LDA [6] with CNN to improve the model's efficacy by selecting relevant features. However, the research does not include a comparison examination of feature selection methods and their performance to determine the adequacy of the selected models for heart disease prediction. This study intends to close these research gaps by assessing the accuracy, precision, and recall of the ReliefF, UMAP, and LDA models in the context of heart disease prediction.

CNN, with feature selection methods, is suitable for developing a prediction-based approach that can provide relevant and meaningful data to serve as a resource for academics and radiologists in the detection, treatment, and avoidance of cardiovascular diseases [7]. The feature selection methods considered in this study are as follows.

a. ReliefF: It is a type of relief algorithm which addresses multiclass classification problems as opposed to two-class problems. It measures the quality of features based on their ability to effectively discriminate between instances that are close together [8].

b. UMAP: A fast and scalable dimensionality technique which preserves both local and global structure of data [9].

c. LDA: LDA is a computationally efficient ML training approach for identifying cardiac diseases in the early stage by preserving the class separation and improves model training [10].

Various research has been done for the identification of heart diseases using appropriate feature selection methods for CNN. Shrivastava et al. [11] explored extra tree classification techniques for feature selection by calculating the significance of each feature with CNN and Binary Long short-term Memory (BiLSTM) for cardio-related disease prediction. However, the extra tree classifiers are largely unstable, and any modification in the data can lead to a significant change in the prediction outcomes. Another research [12] employed a filter-based selection method to identify the top risk variables from extremely detailed database datasets to accurately classify cardiac disorders, and the efficiency of the model is enhanced significantly with the reduced dataset. The filter-based techniques have highly restricted interactions with the model, and there is a risk of neglecting the interactions that are extremely essential for prediction.

The benefit of employing feature selection differs according to the ML technique utilized for the heart datasets. Robinson Spencer et al. [13] suggested a combination of Chi-squared feature selection and the BayesNet approach for feature selection and prediction. Moreover, the Chi-Square test can be utilized to determine whether there is a significant relationship between every feature and the variable being examined. Nagarajan et al. [14] propose a crow search approach for feature selection and classification utilizing deep CNN with a classification accuracy of 94%. Meta-heuristic algorithms' convergence rates and capacity to avoid local minima vary between algorithms, necessitating careful consideration and evaluation, and the performance also depends on the specific algorithms used.

Yang et al. [15] examined an information gain-based feature selection method to extract the critical features in the dataset to enhance the results of the technique. However, features having a large number of classes may have more information gain because of their level of detail, resulting in a bias in selecting features. Yazdani et al. [16] predicted heart disease using scores of the important features with weighted associative rule mining. Ensemble-based feature selection algorithms with CNN are used in the study by Khan et al. [17] for the detection of cardiovascular diseases. The research by Saba et al. [18] employs various feature selection algorithms to classify the data using SVM. A statistical feature selection method is employed to find the association between the data and SVM, which is employed to predict heart diseases [19]. Table 1 lists some of the feature selection methods with their evaluation metrics, scores, and datasets used in the recent studies.

Table 1. Studies on Features Selection Methods for Heart Disease Prediction

Authors	Feature Selection method	Evaluation metric	Score	Dataset
Shrivastava et al., 2022 [11]	Extra tree classifier	Accuracy	96.66%	Cleveland UCI
Pathan et al., 2022 [12]	Filter based technique	Accuracy	81%	Cardiovascular disease (CVD) and Framingham

Continued on next page

Table 1 continued

Robinson Spencer et al., 2020 [13]	Chi-Square test	Accuracy	85%	Cleveland, Long-Beach –VA, Hungarian, Switzerland
Nagarajan et al., 2022 [14]	Crow Search Algorithm	Accuracy	94%	Medical data set Ten different sets of data
Yang et al., 2022 [15]	Information gain	Accuracy	93.44%	HDD –real patient data
Yazdani et al., 2021 [16]	Association Rule Mining	Confidence	98%	UCI
Khan et al., 2023 [17]	Ensemble with CNN	Accuracy	80%	Online Survey data
Saba et al., 2022 [18]	27 different feature selection methods(filter, wrapper, embedded techniques)	Accuracy Sensitivity Specificity F-Measure	94.45%(Avg) 91.% (Avg)	Microarray and Cleveland
Ogundepo et al., 2023 [19]	Chi-Square test and SVM	Accuracy Sensitivity Specificity Precision The area under the ROC curve Log loss value	85% 82% 88% 87% 91% 38%	Cleveland

Relief and Least Absolute Shrinkage and Selection Operator (LASSO) methods for selecting suitable features were used by Mandava et al. for the design of cardiovascular prediction systems. This combined method uses all features and has no feature selection limits, unlike the other methods [20]. Moreover, LASSO is not appropriate for datasets containing associated attributes. The work by Kilicarslan et al. [21] presents hybrid algorithms that use relief and stacked auto-encoder approaches for feature selection to reduce the dimensions and support vector machines (SVM) and CNN for classification with an increase in accuracy.

UMAP dimensionality reduction technique uses topological analysis of data to identify the information's underlying topology while taking into account the data's local and global structure, resulting in more accurate embedded data for manifold input. Wang et al. proposed a UMAP-based feature selection method to eliminate duplicate and irrelevant features [22]. Due to its capacity to preserve both local and global data structures, computational effectiveness, and scalability with massive data sets, UMAP is employed by Paplomatas et al. to predict metabolic syndrome with ML techniques [23].

LDA, along with other feature selection methods like Principal Component Analysis (PCA) and kernel PCA (KPCA) algorithms, is investigated. The findings by Mutinda et al. [24] show that LDA consistently emerged as a powerful approach, significantly improving the algorithm's efficiency across all evaluated criteria in heart disease prediction.

LDA is used with a Genetic algorithm (GA) for the prediction of liver diseases, and the study by Suryaningrum et al. [25] reveals that LDA with GA outperforms the conventional methods. Table 2 lists the feature selection methods like ReliefF, UMAP, and LDA that are considered in this study, and it is evident from the literature that very few classification algorithms use these methods for feature selection.

Table 2. Studies in the literature employing Relief, UMAP, and LDA feature selection methods

Authors	Feature Selection method	Evaluation metric	Score	Dataset
Mandava et al. 2024 [20]	Relief and Least Absolute Shrinkage and Selection Operator	Accuracy	99.12%	UCI
Kilicarslan et al. 2020 [21]	Relief and stacked AutoEncoder With SVM and CNN	Accuracy Ovarian dataset Leukemia dataset Central Nervous System(CNS)	96.14% 4.83% 65%	Microarray datasets
Wang et al. 2024 [22]	UMAP & LASSO	Accuracy	96%	Benchmark datasets
Paplomatas et al. 2024 [23]	UMAP	AUC, Recall Precision F1 Score Kappa MCC T-Sec	**	Unpublished dataset (2017 to 2022)
Mutinda et al. 2024 [24]	LDA	Accuracy, F1-Score, Precision, Recall, Specificity	Refer Table 3	Kaggle
Surayaningrum et al. 2024 [25]	LDA	Average Forecast Error Rate (AFER)	0.0435%	UCI

**https://www.mdpi.com/eng/eng-05-00075/article_deploy/html/images/eng-05-00075-g001-550.jpg (Values for all the metrics for evaluating 14 different algorithms are provided).

Table 3 presents the classifier performance for LDA feature selection algorithms, which represents the scores field in Table 2 for the study by Mutinda et al.

Table 3. Classifier performance –LDA feature selection (Mutinda et al.)

Algorithm	Accuracy	F1-Score	Precision	Recall	Specificity
Logistic Regression	84.07	81.16	84.80	78.29	88.36
SVM	65.19	53.95	65.56	46.30	80.71
KNN	62.69	50.66	61.14	44.35	78.14
Naïve Bayes	85.19	82.78	84.76	81.68	87.52
DNN	76.05	64.57	67.01	63.13	75.07

Recent research has provided a critical understanding of the complex relationship in the feature reduction approaches and provides essential insights for developing more precise and efficient heart disease prediction methods. Although numerous feature selection approaches are available, selecting the optimal feature selection technique is critical, especially in medical applications.

This study fills the research gap by providing some insights into feature selection techniques like ReliefF, UMAP, and LDA through comparative analysis using evaluation measures like accuracy, precision, and recall. The implication of this research is to know the effect of Relief, UMAP, and LDA feature selection to improve heart disease prediction.

The main contribution of the study is to assess the CNN model for predicting heart disease by employing features selected from ReliefF, UMAP, and LDA methods. This comparative analysis will provide an understanding of the feature selection approaches for ML/DL models for heart disease prediction. This research work seeks to contribute to the ongoing enhancement of feature selection algorithms for CNN.

2 Methodology

The comparative study for the prediction of heart disease using three different feature selection methods, Relief, UMAP, and LDA, is examined in this work. The components of the suggested framework are shown in Figure 1.

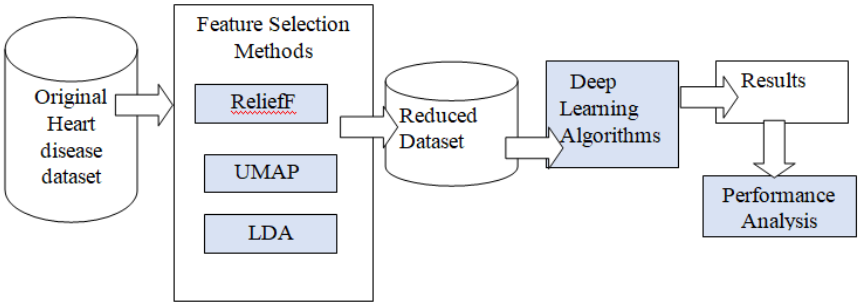

Fig 1. Block Diagram for the overall flow of the Prediction Process

2.1 Dataset Description

The data was collected from a group of people who have the possibility of heart disease. Three hundred three distinct records were extracted with 14 attributes. Age, sex, chest pain type, RBP (Resting Blood Pressure), serum cholesterol, FBS (Fasting Blood Sugar), troponin, history, diabetes, smoking, edema, diet, risk factors, and resting ECG are the features considered in this study. Highest Heart Rate Recorded, Exercise-induced angina, peak exercise ST segment, activity-induced ST depression compared to rest, Number of main vessels (0–3). It ensures that the data is accurate and handles the missing and unwanted data in the right way with the help of clinicians. Finally, it was validated with the Cleveland dataset downloaded from Kaggle for the same Number of records, and the results were recorded. The original data is divided into testing and training phases in a proportion of 80% to 20% to minimize overfitting and measure the accuracy of the output.

2.2 Preprocessing

The MinMax scaling method is used for the preprocessing of the heart disease dataset. A common data preprocessing method for the normalization of a dataset's features is called min-max scaling. Each trait is turned into a range, frequently between 0 and 1. This formula is used on each feature in a dataset.

$$x - scaled = \frac{x-(x)}{(x)-(x)} \tag{1}$$

For several important reasons, min-max scaling is crucial for heart disease dataset preprocessing. Age, cholesterol levels, and blood pressure are just a few examples of the varied variables that can be found in datasets on heart disease. These features are converted into a consistent range (often 0 to 1) using min-max scaling, which guarantees that each feature contributes proportionately to the predictive model. For the algorithm to treat all characteristics equally, this uniform scaling is essential since it prevents qualities with higher numerical values from controlling the learning process. Min-max scaling guarantees that the model catches key patterns without being biased towards any particular feature in heart disease prediction, where the link between several variables might be subtle and complex. Additionally, it improves the stability and speed of the training process, facilitating a more rapid and effective convergence of ML algorithms, such as neural networks. Min-max scaling helps in the preprocessing pipeline for heart disease datasets because it encourages equitable and balanced feature contributions, which eventually results in more precise, trustworthy, and understandable predictions.

2.3 Feature Selection Techniques

2.3.1 ReliefF

ReliefF[26], a multivariate selection of features, takes features based on their physical location. The mathematical calculation of feature weights presents a convex optimization challenge. The following equation is used to determine relief and feature group selection.

$$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2 \tag{2}$$

where the nearest instance of the same class is near it, and the closest instance of a different class is nearness. W stands for the weight, and X is a feature vector[18].

2.3.2 UMAP

Euclidean metric computes the continuous attributes, and the two vectors' ED (Euclidean distance) is calculated using Equation (3).

$$d(x,y) = \sum_{i=1}^{n} (x_i - y_i)^2 \tag{3}$$

Hamming is employed as the metric for the nominal characteristics. Distance is calculated as shown in Equation (4).

$$d(x,y) = \sum_{i=1}^{n} \delta(x_i, y_i) \tag{4}$$

Where $\delta(x_i, y_i)$ = 1, if $x_i = y_i$ and $\delta(x_i, y_i)$ =0 otherwise. Hamming distance, which measures how similar two data points are, is frequently employed for such features[27].

Canberra is used as the measure for the ordinal characteristics. It is the Manhattan metric weighted variant, and it is calculated as follows:

$$d(x,y) = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|} \tag{5}$$

2.3.3 LDA

The LDA method computes a projection vector in two-class cases that reduces the intra-class scatter matrix in the feature space while increasing the inter-class scatter matrix. LDA seeks to improve the ability to distinguish the various forms of CVD by shifting data to a lower-dimensional space. To record class-specific biased information, it selects parameters that enhance inter-class variability by reducing intra-class variation. For n instances, the between-class matrix, S_B, and within-class matrix, S_W, are measured as follows:[24]

$$S_B = \sum_{i=1}^{C} n_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{6}$$

$$S_W = \sum_{j=1}^{C} \cdot \sum_{i=1}^{n_j} (x_{ij} - \mu_j)(x_{ij} - \mu_j)^T \quad (7)$$

In Equations (6) and (7), x_i is $the\ i^{th}$ sample of the j^{th} class and $\mu_i\ and\ \mu$ represent the class means to value and whole mean value, respectively. C Denotes total Number of classes.[25]

2.3.4 CNN

CNNs are an effective tool for predicting cardiac disease since they have demonstrated great performance in several domains, including the processing of medical images. A modified back propagation training method is used to train the CNN. Testing revealed that CNN is more accurate in predicting both the absence and presence of heart diseases. CNN, combined with feature selection methods including ReliefF, UMAP, and Linear Discriminant Analysis (LDA and the Min-Max Scaling, produced convincing findings. The CNN model augmented by UMAP and Min-Max Scaling stood out as the best performance among these combinations, displaying superior precision, recall, and accuracy[17].

3 Results and Discussion

A combination of CNN and feature selection approaches, including ReliefF, UMAP, and LDA coupled with Min-Max scaling, were used in this study to predict heart diseases. The efficacy of the suggested feature selection methods was evaluated with metrics like Accuracy, Precision, and Recall in order to show the efficiency of the CNN model.

3.1 Model Evaluation

3.1.1 Accuracy:

The overall accuracy of the model's predictions is assessed, and the proportion of samples that are correctly classified is compared to all samples. It provides a thorough analysis of the model's effectiveness[20–22].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Table 4. Comparison of accuracy values (ReliefF, UMAP, L DA)

No.of Iterations	CNNAccuracy (%) without feature Selection		LDA+ CNN Accuracy (%)		ReliefF+CNN Accuracy (%)		UMAP+CNN Accuracy (%)	
			With Feature Selection					
	Dataset (Collected)	Cleveland dataset	Dataset (Collected)	Cleveland dataset	Dataset (Collected)	Cleveland dataset	Dataset (Collected)	Cleveland dataset
10	86..8	86.5	87.70	86.70	89.95	88.56	90.75	90.30
20	85.3	84.98	88.17	87.98	90.80	90.10	91.11	90.00
30	84.6	84.30	88.48	88.10	90.67	90.02	91.51	91.41
40	86.7	86.20	88.94	88.40	90.85	90.40	91.77	91.56
50	85.2	85.00	89.91	89.23	90.91	90.67	91.88	91.76

Table 4 represents the accuracy values for CNN without and with feature selection for the dataset collected from the heart disease patients and the Cleveland dataset downloaded from Kaggle. According to the results, the accuracy of CNN is 86% to 85%, LDA+CNN is 87% to 90%, ReliefF+CNN is 89% to 91%, and UMAP+CNN is 90% to 92%, which is higher than the accuracy of other algorithms. Then, the CNN with and without feature selection is validated for correctness against the Cleveland dataset. The tabulated values almost match the values of the dataset collected.

3.1.2 Precision

Relative to all samples predicted to be positive, the proportion of positively recognized positive samples is known as precision[24].

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Table 5. Comparison of precision values (ReliefF, UMAP, LDA)

No.of Iterations	CNNPrecision Without feature Selection		LDA+CNN Precision		ReliefF+ CNN Precision		UMAP+CNN Precision	
			With Feature Selection					
	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset
10	0.81	0.79	0.82	0.80	0.83	0.81	0.85	0.84
20	0.80	0.78	0.82	0.79	0.84	0.85	0.86	0.85
30	0.83	0. 79	0.84	0.81	0.84	0.86	0.88	0.90
40	0.82	0.80	0.84	0.82	0.85	0.87	0.89	0.90
50	0.82	0.80	0.84	0.80	0.85	0.86	0.89	0.85

Table 5 represents precision values for CNN without and with feature selection for the dataset collected from the heart disease patients and the Cleveland dataset downloaded from Kaggle. The findings show that the precision of LDA+CNN is 0.82 to 0.84, ReliefF+CNN is 0.83 to 0.85, and UMAP+CNN is 0.85 to 0.90, which is higher than other algorithms in terms of precision. It is validated with the existing Cleveland dataset, and the performance does not differ much from the dataset used.

3.1.3 Recall (Sensitivity)

Recall, also referred to as sensitivity, is the percentage of positive samples correctly recognized as such from the total Number of samples that are positive.[23]

$$Recall = \frac{TP}{TP + FN} \tag{10}$$

Table 6. Comparison of recall values (ReliefF, UMAP, LDA)

No.of Iterations	CNNRecall Without feature Selection		LDA+CNN Recall		ReliefF+CNN Recall		UMAP+CNN Recall	
			With Feature Selection					
	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset	Dataset (Collected	Cleveland Dataset
10	0.82	0.80	0.79	0.80	0.81	0.80	0.83	0.79
20	0.81	0.79	0.80	0.79	0.82	0.81	0.84	0.80
30	0.82	0.78	0.81	0.78	0.83	0.82	0.85	0.81
40	0.83	0.81	0.80	0.81	0.83	0.81	0.85	0.80
50	0.84	0.83	0.82	0.78	0.83	0.84	0.85	0.84

Table 6 represents the recall values of CNN without and with feature selection methods. The results show that the recall of LDA+CNN is 0.79 to 0.82, ReliefF+CNN is 0.81 to 0.83, and UMAP+CNN is 0.83 to 0.85, all of which are higher than ReliefF and LDA. It is also validated using the Cleveland dataset. Through careful testing, it was evident that the CNN model combined with UMAP feature selection and Min-Max scaling outperformed alternative designs, displaying improved precision, recall, and accuracy metrics.

The dataset's internal structure remains intact due to UMAP's ability to reduce the dataset's dimensionality, which was critical for CNN in recognizing complex patterns required for accurate predictions. Furthermore, Min-Max scaling makes sure that all features contribute equally, preventing any one attribute from monopolizing the learning process. The combination of UMAP and Min-Max scaling dramatically improved recall. Furthermore, the improved accuracy underlines the validity of this method in identifying those with and without heart disease. The usefulness of UMAP and the significance of consistent feature scaling via Min-Max scaling are highlighted by these results, underscoring their crucial roles in improving the precision and efficacy of CNN-based heart disease prediction models.

Table 7. Comparison with other works in the literature

Authors	Accuracy (Existing Study)	Accuracy (Collected dataset)
Shrivastava et al., 2022 [11]	96.66%	95.%
Pathan et al., 2022 [12]	81%	78%
Robinson Spencer et al., 2020 [13]	85%	80%
Nagarajan et al., 2022 [14]	94%	92%
Yang et al., 2022 [15]	93.44%	90%

Further, the heart disease prediction using CNN with and without feature selection methods for reducing the dimensions is also studied. Table 3 lists the accuracy values for the existing feature selection methods. As a result, the suggested feature selection methods (Relief, UMAP, LADA) perform on par with the recent existing studies in terms of accuracy. This study evaluates and selects efficient feature selection strategies for developing an accurate prediction system for real-time deployments.

The results of this study are essential for developing predictive analytics and diagnosing diseases. The diagnosis of heart disease and possibly other medical problems may be revolutionized if CNN models with UMAP-selected features are shown to be superior with regard to accuracy, precision, and recall. Early-stage accurate forecasts can result in prompt actions that eventually save lives and ease the strain on healthcare systems.

4 Conclusion

The study on the prediction of heart disease using CNN combined with feature selection algorithms like ReliefF, UMAP, and LDA was studied, and it was validated with and without feature selection algorithms. It has shown convincing evidence of UMAP's superior influence with 91%, 88%, and 85% of accuracy, precision, and recall, respectively. Extensive testing revealed that UMAP, as a feature selection strategy, considerably improves the CNN model's performance by successfully identifying important patterns in the data and decreasing noise with better interpretability and generalization abilities. The model's performance can still be enhanced with additional discriminatory feature sets and datasets. Finally, feature selection may successfully boost the predictive value of a dataset while also significantly improving model accuracy. These results represent a significant development in the field of medical diagnostics and offer a viable route to more accurate and effective heart disease forecasts, which will eventually lead to early interventions and better patient outcomes.

References

1) Noroozi Z, Orooji A, Erfannia L. Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific reports* . 2023;13:1–15. Available from: https://doi.org/10.1038/s41598-023-49962-w.
2) Vanessa, Nadoo A, Ogala E, Gbaden T. Machine Learning Model for the Prediction of Cardiovascular Diseases. *Procedia Computer Science* . 2024;3(2):430–443. Available from: https://www.researchgate.net/publication/378153091_Machine_Learning_Model_for_the_Prediction_of_Cardiovascular_Diseases.
3) Najafi A, Nemati A, Ashrafzadeh M, Zolfani SH. Multiple-criteria decision making, feature selection, and deep learning: A golden triangle for heart disease identification. *Engineering Applications of Artificial Intelligence*. 2023;125. Available from: https://doi.org/10.1016/j.engappai.2023.106662.
4) Das P, Dobhal DC, Dobhal M. Heart disease detection using feature optimization and classification. In: Automation and Computation. CRC Press. 2023;p. 1–10. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003333500-1/heart-disease-detection-using-feature-optimization-classification-purushottam-das-dinesh-dobhal-manika-manwal.
5) Jain R, Betrabet PR, Rao BA, Reddy NVS. Classification of Cardiac Arrhythmia using improved Feature Selection methods and Ensemble Classifiers. In: 1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021) ;vol. 2161 of Journal of Physics: Conference Series. IOP Publishing. ;p. 12003–12003. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012003/meta.
6) Wang G, Lauri F, Hassani AHE. Feature Selection by mRMR Method for Heart Disease Diagnosis. *IEEE Access*. 2022;10:100786–100796. Available from: https://doi.org/10.1109/ACCESS.2022.3207492.
7) Sharma A, Pal T, Jaiswal V. Chapter 12 - Heart disease prediction using convolutional neural network. In: Cardiovascular and Coronary Artery Imaging;vol. 1. 2022;p. 245–272. Available from: https://doi.org/10.1016/B978-0-12-822706-0.00012-3.
8) Balasubramaniam S, Joe C, Manthiramoorthy C, Kumar KS. ReliefF based feature selection and Gradient Squirrel search Algorithm enabled Deep Maxout Network for detection of heart disease. *Biomedical Signal Processing and Control*. 2024;87(Part A). Available from: https://doi.org/10.1016/j.bspc.2023.105446.
9) Wang G, Zheng S, Yang X, Song Y, Tang Z, Jiang Y, et al. Convolutional Neural Network-Based ECG Signal Classification Model: A Study on Addressing Class Imbalance and Enhancing Model Interpretability. *Preprintsorg*. 2024. Available from: https://doi.org/10.20944/preprints202405.1290.v1.
10) Isnanto RR, Rashad I, and CEW. Classification of Heart Disease Using Linear Discriminant Analysis Algorithm. *E3S Web of Conferences*. 2023;448:1–11. Available from: https://doi.org/10.1051/e3sconf/202344802053.
11) Shrivastava PK, Sharma M, sharma P, Kumar A. HCBiLSTM: A hybrid model for predicting heart disease using CNN and BiLSTM algorithms. *Measurement: Sensors*. 2023;25:1–7. Available from: https://doi.org/10.1016/j.measen.2022.100657.
12) Pathan MS, Nag A, Pathan MM, Dev S. Analyzing the impact of feature selection on the accuracy of heart disease prediction. *Healthcare Analytics*. 2022;2:1–9. Available from: https://doi.org/10.1016/j.health.2022.100060.

13) Spencer R, Thabtah F, Abdelhamid N, Thompson M. Exploring feature selection and classification methods for predicting heart disease. *Digital Health*. 2020;6:1–10. Available from: https://doi.org/10.1177/2055207620914777.
14) Nagarajan SM, Muthukumaran V, Murugesan R, Joseph RB, Meram M, Prathik A. Innovative feature selection and classification model for heart disease prediction. *Journal of Reliable Intelligent Environments*. 2022;8:333–343. Available from: https://doi.org/10.1007/s40860-021-00152-3.
15) Yang J, Guan J. A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm. *Information*. 2022;13(10):1–15. Available from: https://doi.org/10.3390/info13100475.
16) Yazdani A, Varathan KD, Chiam YK, Malik AW, Ahmad WAW. A novel approach for heart disease prediction using strength scores with significant predictors. *BMC Medical Informatics and Decision Making*. 2021;21(1):1–16. Available from: https://doi.org/10.1186/s12911-021-01527-5.
17) Mamun MMRK, and TE. Detection of Cardiovascular Disease from Clinical Parameters Using a One-Dimensional Convolutional Neural Network. *Bioengineering*. 2023;10(7):1–29. Available from: https://doi.org/10.3390/bioengineering10070796.
18) Bashir S, Khattak IU, Khan A, Khan FH, Gani A. A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and Embedded Approaches. *Complexity*. 2022;2022:1–12. Available from: https://doi.org/10.1155/2022/8190814.
19) Ogundepo EA, Yahya WB. Performance analysis of supervised classification models on heart disease prediction. *Innovations in Systems and Software Engineering*. 2023;19:129–144. Available from: https://doi.org/10.1007/s11334-022-00524-9.
20) Mandava M, vinta SR. MDensNet201-IDRSRNet: Efficient cardiovascular disease prediction system using hybrid deep learning. *Biomedical Signal Processing and Control*. 2024;93. Available from: https://doi.org/10.1016/j.bspc.2024.106147.
21) Kilicarslan S, Adem K, Celik M. Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network. *Medical Hypotheses*. 2020;137. Available from: https://dx.doi.org/10.1016/j.mehy.2020.109577.
22) Wang Y, Huang M, Zhou L, Che H, Jiang B. Multi-cluster nonlinear unsupervised feature selection via joint manifold learning and generalized Lasso. *Expert Systems with Applications*. 2024;255(Part A). Available from: https://dx.doi.org/10.1016/j.eswa.2024.124502.
23) Paplomatas P, Rigas D, Sergounioti A, Vrahatis A. Enhancing Metabolic Syndrome Detection through Blood Tests Using Advanced Machine Learning. *Eng*. 2024;5(3):1422–1434. Available from: https://dx.doi.org/10.3390/eng5030075.
24) Mutinda JK, Langat AK. Exploring the Role of Dimensionality Reduction in Enhancing Machine Learning Algorithm Performance. *Asian Journal of Research in Computer Science*. 2024;17(5):157–166. Available from: https://dx.doi.org/10.9734/ajrcos/2024/v17i5445.
25) Firmansyah F. Penyusunan Program Semester dalam Pembelajaran: Analisis Teoretis dan Praktis untuk Meningkatkan Efektivitas Pembelajaran. LPPM Universitas Singaperbangsa Karawang - Research Department in Indonesia University. 2024. Available from: https://dx.doi.org/10.35706/azzakiy.v2i1.11122. doi:10.35706/azzakiy.v2i1.11122.
26) Nadheer I. Heart Disease Prediction System using hybrid model of Multi-layer perception and XGBoost algorithms. In: Fifth International Scientific Conference of Alkafeel University (ISCKU 2024);vol. 97 of BIO Web of Conferences. EDP Sciences. 2024;p. 1–9. Available from: https://doi.org/10.1051/bioconf/20249700047.
27) Murad N, Melamud E. Global patterns of prognostic biomarkers across disease space. *Scientific Reports*. 2022;12(1):1–13. Available from: https://dx.doi.org/10.1038/s41598-022-25209-y.