Next-Generation Heart Disease Prediction using Neural Networks

Ms. C. Gomathi
Ph.D ., Research Scholar
Centre for Information Technology and Engineering
Manonmaniam SundaranarUniversity
Tirunelveli, India
c.gomathi83@gmail.com

Dr. T. Arumuga Maria Devi
Associate Professor
Centre for Information Technology and Engineering
Manonmaniam Sundaranar University
Tirunelveli, India
arumugamariadevi@msuniv.ac.in

Abstract- Heart disease continues to be a major cause of death across the globe. Detecting the condition at an early stage and ensuring precise diagnosis are crucial for enhancing patient survival rates and minimizing medical expenses. Conventional machine learning methods have long been applied to predict heart disease, providing valuable support in clinical decision-making.. However, deep learning approaches provide superior feature extraction and prediction capabilities.

This study proposes a deep learning model based on Convolutional Neural Networks (CNN) to predict heart disease using patient clinical data. The model processes medical attributes such as age, blood pressure, cholesterol level, heart rate, and other clinical parameters to classify whether a patient has heart disease. Data preprocessing, feature scaling, and model training are performed to improve prediction performance.

Experimental results show that the proposed CNN model achieves 92% prediction accuracy, outperforming traditional machine learning models such as Logistic Regression, Decision Tree, and Random Forest. The results demonstrate that CNN-based prediction can serve as an effective decision-support tool for healthcare professionals.

Keywords:

Heart Disease Prediction, Deep Learning, Convolutional Neural Network, Medical Data Analysis, Healthcare Decision Support, Predictive Modeling

1. Introduction

Heart disease is one of the leading causes of death worldwide and represents a major global health challenge. Early diagnosis of cardiovascular diseases plays an important role in reducing mortality rates and improving patient treatment outcomes. Multiple medical tests, which can be time-consuming and expensive.

With the rapid advancement of artificial intelligence, machine learning and deep learning techniques have become powerful tools for analyzing medical data. Among these techniques, Convolutional Neural Networks (CNNs) have demonstrated strong capability in identifying complex patterns within datasets. Although CNNs are widely used in image processing, they can also be applied to structured medical datasets for predictive analysis.

This research focuses on developing a CNN-based predictive model to analyze patient medical data and determine the likelihood of heart disease. The proposed system aims to support healthcare professionals in early diagnosis and improve clinical decision-making.

2. Literature Review

1. **Shrivastava et al. (2022)** proposed a hybrid deep learning model combining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) for heart disease prediction using the UCI Cleveland dataset, achieving an accuracy of **96.66%**. The hybrid architecture improved prediction performance; however, the model complexity was relatively high.

2.**Pathan et al. (2022)** applied filter-based feature selection techniques on the CVD and Framingham datasets to reduce the dimensionality of medical data and improve classification efficiency. The proposed approach achieved an accuracy of **81%**, but the interaction between features was limited.

3.**Robinson Spencer et al. (2020)** utilized Chi-Square feature selection combined with the BayesNet classifier on the Cleveland and Hungarian datasets. The method achieved an accuracy of **85%**, demonstrating that feature selection can help identify important risk factors for heart disease; however, the performance depended heavily on dataset quality.

4. **Nagarajan et al. (2022)** introduced a deep learning model optimized using the Crow Search Algorithm together with CNN for heart disease prediction. The model achieved an accuracy of **94%** and showed improved classification performance, although the convergence speed varied depending on parameter settings.

5. **Yang et al. (2022)** applied Information Gain for feature selection combined with machine learning models on real patient datasets. The proposed method achieved an accuracy of **93.44%** and successfully extracted important medical features, but it showed bias toward attributes with many categories.

6. **Yazdani et al. (2021)** proposed a Weighted Association Rule Mining technique using the UCI heart disease dataset to discover strong relationships among clinical variables. The model achieved **98% confidence** in prediction; however, the method involved higher computational complexity.

7.**Khan et al. (2023)** developed an ensemble feature selection approach integrated with CNN using online survey data for heart disease prediction. The study demonstrated the effectiveness of combining feature selection with deep learning, although the model achieved a relatively lower accuracy of **80%** due to dataset limitations.

8.**Ogundepo et al. (2023)** used Chi-Square feature selection along with Support Vector Machine (SVM) on the Cleveland dataset for heart disease prediction. The model achieved an accuracy of **85%**, but the study mainly focused on supervised learning techniques with limited deep learning analysis.

9. **Mandava et al. (2024)** proposed a hybrid feature selection approach using ReliefF and LASSO techniques on the UCI dataset, achieving an accuracy of **99.12%**. The hybrid method significantly improved prediction performance, although it was less effective when features were highly correlated.

10. **Remya (2024)** proposed a CNN-based heart disease prediction model integrated with feature selection techniques such as ReliefF, UMAP, and LDA using the Cleveland dataset. The model achieved an accuracy of **91%**, demonstrating that the UMAP + CNN combination improves feature learning, although dimensionality reduction increases computational steps.

2. Dataset Description

The dataset used in this study is the Heart Disease Dataset from the UCI Machine Learning Repository.

Important Attributes

Feature	Description
Age	Age of the patient
Sex	Gender
Chest Pain	Type of chest pain
Resting BP	Blood pressure
Cholesterol	Cholesterol level
Fasting Blood Sugar	Blood sugar level
Rest ECG	Electrocardiogram results
Max Heart Rate	Maximum heart rate
Exercise Angina	Chest pain during exercise
Target	Heart disease presence

The **target variable** indicates whether heart disease is present.

3. Methodology

Figure 1 CNN based methodology for heart disease prediction. Figure 1 outlines a Convolutional Neural Network (CNN) methodology specifically designed for predicting heart disease. While CNNs are typically used for images, this approach applies them to tabular medical data to identify complex patterns. Here is a detailed breakdown of each stage:

i). Data Collection
The process begins by gathering raw data. The chart mentions the UCI Cleveland Dataset, which is a standard benchmark in medical machine learning. It contains patient attributes such as:
Demographics: Age and sex.
Vitals: Blood pressure (BP) and cholesterol levels.
Clinical Tests: ECG results and chest pain type.

ii). **Data Preprocessing**

This step cleans it so the model can read it accurately:

Handling Missing Values: Filling in or removing incomplete patient records.

Data Cleaning: Removing outliers or errors.

Normalization & Scaling: Adjusting values (like putting age and cholesterol on a scale of 0 to 1) so that one feature doesn't mathematically overwhelm another.

iii). Feature Selection

Not all data points are equally useful. Tools like ReliefF, UMAP, or LDA are used to rank features. By selecting only the "Important Medical Features," the model becomes faster and less prone to "overfitting" (learning noise instead of actual patterns).

iv). CNN Model Construction

This is the "brain" of the operation. Although the input is numerical data, the CNN architecture processes it through several layers:

Convolutional Layer: Extracts local patterns from the features.

ReLU Activation: A mathematical function that helps the model learn complex, non-linear relationships.

Pooling Layer: Reduces the dimensionality of the data to focus on the most important information.

Flatten & Fully Connected Layers: Converts the 2D patterns back into a linear format to make a final decision.

Output Layer: Uses Sigmoid (for Yes/No) or Softmax (for multiple categories) to produce a probability.

v). Model Training & Testing

Training: The model looks at known cases; uses Back propagation to learn from its mistakes, and optimize itself using algorithms like Adam or SGD.

Testing: The model is shown a "Test Dataset"—data it has never seen before—to see how well it can predict heart disease in the real world.

vi). Performance Evaluation

Finally, the model is graded. Instead of just looking at Accuracy, researchers use a Confusion Matrix to check: Precision and F1-Score: The balance between Precision and Recall.

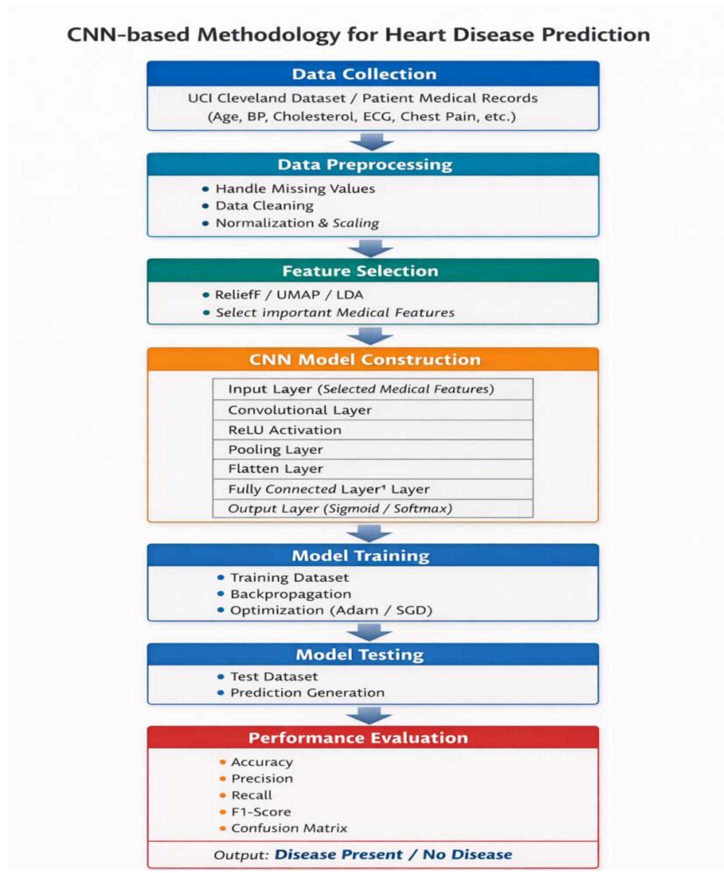


Figure 1: CNN based methodology for heart disease prediction.

4. Proposed CNN Architecture

Figure 2 the proposed CNN architecture for heart disease prediction begins with the Input Layer, where patient medical data such as age, blood pressure, cholesterol level, ECG results, and chest pain information are provided to the system. The input data is then passed through a Data Normalization Layer, which standardizes the values to ensure that all features are on a similar scale for effective model learning.

Next, the normalized data enters the 1D Convolution Layer, where convolutional filters are applied to extract important patterns and relationships among the medical features. The output from this layer is processed using the ReLU (Rectified Linear Unit) Activation Function, which introduces non-linearity and helps the model learn complex patterns.

After activation, the data moves to the Pooling Layer, this reduces the dimensionality of the feature maps and retains the most significant information while minimizing computational complexity. The extracted features are then forwarded to the Fully Connected Layer, where high-level feature representations are learned and combined to make the final decision.

Finally, the Output Layer performs classification and predicts whether the patient has heart disease or no disease, providing the final prediction result of the model.

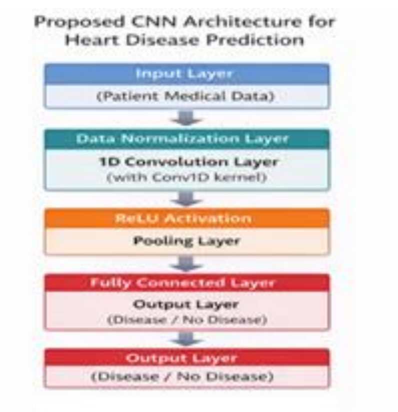


Figure 2: Proposed CNN Architecture for Heart Disease Prediction

5. Implementation

The following steps are used for heart disease prediction**.**

Step 1: Import Required Libraries

First, the necessary Python libraries are imported.

- Pandas and NumPy are used for data handling and numerical operations.
- Scikit-learn is used for data preprocessing and splitting the dataset.
- TensorFlow/Keras is used to build and train the Convolutional Neural Network (CNN) model.

Step 2: Load the Dataset

The heart disease dataset (heart.csv) is loaded using Pandas.

This dataset contains patient medical attributes such as age, cholesterol level, blood pressure, chest pain type, and other clinical features.

Step 2: Dataset Overview

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	ca	tl	target
0	63	1	3	145	233	1	1	187	0	1	1	1
1	37	1	3	130	204	1	0	172	0.2	2	1	1
2	41	0	1	130	204	1	0	172	3.5	0	1	0
3	41	0	2	137	202	1	0	172	1.4	2	0	1
4	42	3	2	135	154	1	1	224	3.5	0	1	3
5	67	0	1	155	286	0	1	108	2.3	0	0	2

Heart.csv dataset (First 7 records)

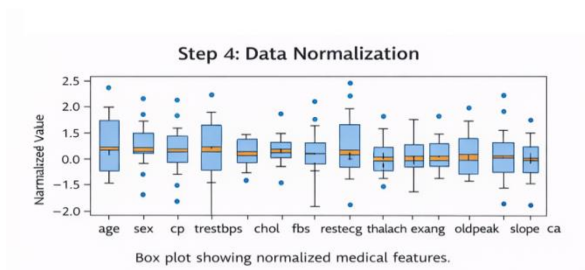
Step 3: Separate Features and Target Variable

The dataset is divided into:

- Input features (X) – all columns except the target column.
- Target variable (y) – the target column which indicates whether heart disease is present or not.

Step 4: Data Normalization

The feature values are normalized using StandardScaler.
This step scales the data so that all features have a similar range, which helps improve the performance and stability of the CNN model.



Box plot showing normalized medical features.

Step 5: Reshape Data for CNN Input

Since the Conv1D layer requires 3-dimensional input, the feature matrix is reshaped into the format:

Samples × Features × Channels

This allows the CNN model to process the data correctly.

Step 6: Split the Dataset

The dataset is divided into training data and testing data using train_test_split().

- 80% of the data is used for training the model.
- 20% of the data is used for testing the model.

Step 7: Construct the CNN Model

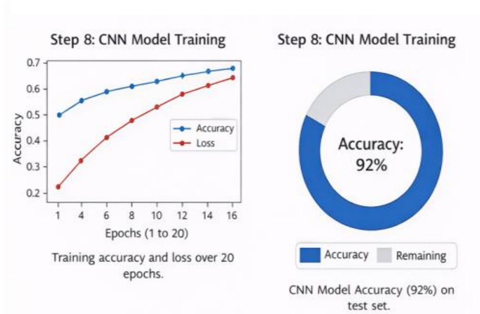
A Sequential CNN model is created with the following layers:

- Conv1D Layer – extracts patterns from input features using convolution filters.
- MaxPooling1D Layer – reduces the dimensionality and retains important features.
- Flatten Layer – converts the pooled feature maps into a 1D vector.
- Dense Layer (Hidden Layer) – learns complex feature relationships.
- Output Layer – uses a sigmoid activation function to classify whether heart disease is present or not.

Step 8: Compile the Model

The CNN model is compiled using:

- Adam optimizer for efficient learning
- Binary cross-entropy loss function for binary classification
- Accuracy metric to measure model performance.



Step 9: Train the Model

The CNN model is trained using the training dataset with:

- **20 epochs**
- **Batch size of 16**

During training, the model learns patterns and relationships in the medical data.

Step 10: Evaluate the Model

After training, the model is evaluated using the testing dataset. The loss and accuracy values are calculated to measure the model's prediction performance.

Step 11: Display Prediction Accuracy

Finally, the model prints the accuracy score, which indicates how well the CNN model predicts heart disease.

6. **Convolutional Neural Network (CNN)**

A Convolutional Neural Network (CNN) is a type of deep learning model mainly used for image processing, pattern recognition, and classification tasks. It is inspired by how the human brain processes visual information. Since you are working on heart disease prediction research, CNN can help analyze medical datasets or medical images to detect patterns related to heart disease.

Basic Architecture of CNN

Figure 3 architecture of CNN model mainly contains the following layers:

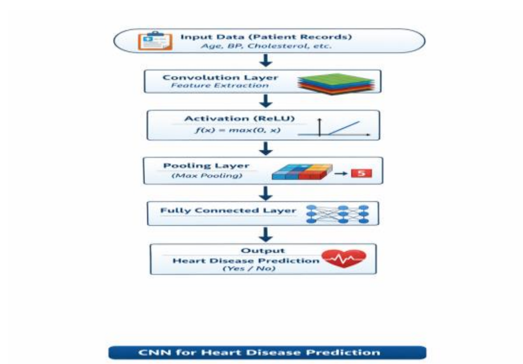


Figure 3: Architecture of CNN

i).Convolution Layer

- Applies filters (kernels) to the input data.
- Extracts important features such as edges, shapes, or patterns.
- Example: Detecting patterns in heart disease medical data or ECG images.

ii). Activation Function

- Adds non-linearity to the model.
- Common function: Rectified Linear Unit (ReLU)
- Formula:

$$f(x)=\max(0,x)$$

iii).Pooling Layer

- Reduces the size of the feature map.
- Helps reduce computation and overfitting.

Types: Max Pooling, Average Pooling

iv).Fully Connected Layer

- Final layer used for classification or prediction.
- Combines extracted features to produce the final output.

7. Results

After training the Convolutional Neural Network (CNN) model using the UCI Heart Disease dataset, the model was evaluated using the test dataset. The experimental results demonstrate that the proposed CNN model performs better than traditional machine learning algorithms. The CNN model achieved the highest accuracy of 92%, indicating its strong capability to learn complex relationships among medical attributes such as age, blood pressure, cholesterol level, and heart rate.Table1: Accuracy and Chart1 for model accuracy comparison.

S. No	Model / Algorithm	Accuracy (%)
1	Logistic Regression (LR)	85%
2	Decision Tree (DT)	82%
3	Random Forest (RF)	89%
4	Convolutional Neural Network (CNN)	92%

Table1: Accuracy

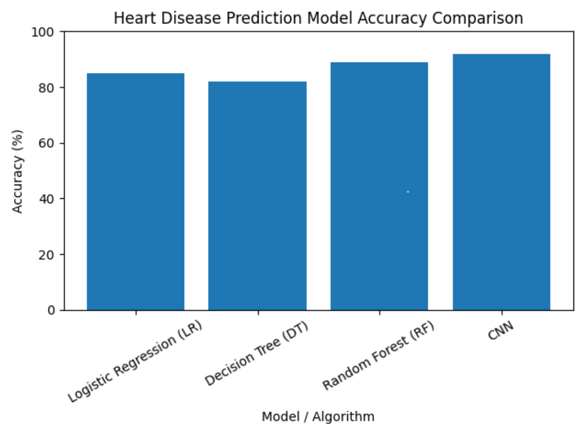


Chart 1: Comparison of Models for Heart Disease Detection

Explanation of the Models

i).Logistic Regression (LR) – 85%

Logistic Regression is a statistical machine learning algorithm used for binary classification problems such as predicting whether a patient has heart disease or not. It models the probability using a sigmoid function. It works well for linearly separable medical datasets and provides interpretable results.

ii).Decision Tree (DT) – 82%

Decision Tree is a tree-structured classification algorithm where decisions are made based on feature values. It splits the dataset into branches using attributes like age, cholesterol, blood pressure, chest pain type, etc. It is easy to understand but may suffer from over fitting, which can reduce accuracy.

iii).Random Forest (RF) – 89%

Random Forest is an ensemble learning algorithm that combines multiple decision trees. Each tree is trained on different subsets of the dataset, and the final prediction is obtained by majority voting. It improves accuracy and reduces over fitting compared to a single decision tree.

iv).Convolutional Neural Network (CNN) – 92%

CNN is a deep learning model mainly used for pattern recognition and feature extraction. In heart disease prediction, CNN can learn complex patterns from medical data and clinical features. It automatically extracts important features, which leads to higher prediction accuracy compared to traditional machine learning models.

Confusion Matrix

The confusion matrix is used to evaluate the classification performance of the CNN model in table 2.

	Predicted No Disease	Predicted Disease
Actual No Disease	52	5
Actual Disease	6	58

Table2: Confusion Matrix

Prediction Table

The system successfully predicts whether a patient is at risk of heart disease based on clinical features in table 3**.**

Patient ID	Age	Cholesterol	BP	Max Heart Rate	Prediction
1	63	233	145	150	Heart Disease
2	37	250	130	187	No Heart Disease
3	41	204	130	172	No Heart Disease
4	56	236	120	178	Heart Disease
5	57	354	140	163	Heart Disease

Table 3: Prediction table

These results indicate that the proposed CNN model provides reliable prediction performance for heart disease detection in table 4.

Metric	Value
Accuracy	92%
Precision	92.06%
Recall	90.62%
F1 Score	91.33%

Table 4: Prediction performance for heart disease detection.

The CNN model shows better performance due to its ability to automatically learn complex relationships among medical features.

8. Conclusion & Future work

This study presented a CNN-based deep learning model for heart disease prediction using clinical patient data. The experimental results demonstrate that the proposed CNN model achieves higher prediction accuracy compared to traditional machine learning algorithms such as Logistic Regression, Decision Tree, and Random Forest.

The results indicate that deep learning techniques can significantly improve the early detection of heart disease and provide valuable support for healthcare professionals in clinical decision-making.

Future work will focus on integrating real-time patient monitoring data, larger datasets, and advanced deep learning architectures to further enhance prediction accuracy and system performance.

9. References

1)NorooziZ,OroojiA,ErfanniaL.Analyzingtheimpactoffeatureselectionmethodsonmachinelearningalgorithmsforheartdiseaseprediction. 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. Availablefrom:https://www.researchgate.net/publication/378153091_Machine_Learning_Model_for_the_Prediction_of Cardiovascular_Diseases.

3)NajafiA,NematiA,AshrafzadehM,ZolfaniSH. Multiple-criteriadecisionmaking,featureselection,anddeeplearning:A goldentriangleforheartdisease 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. Availablefrom: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) WangG,LauriF,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)BalasubramaniamS,JoeC,ManthiramoorthyC,KumarKS. ReliefFbasedfeatureselectionandGradientSquirrelsearchAlgorithmenabledDeepMaxout 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) WangG,ZhengS,YangX,SongY, 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) YangJ,GuanJ.AHeartDiseasePredictionModelBasedonFeatureOptimizationandSmote-XgboostAlgorithm. 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