

Heart Disease Prediction Using Machine Learning and Deep Learning Techniques

Kanagasabai V M.E. Students

Department of Computer Science and Engineering, MRK Institute of technology, Kattumannarkoil, India

Ramalingam S Assistant Professor

Department of Computer Science and Engineering, MRK Institute of technology, Kattumannarkoil, India

Senthamaraiselvi M, HOD

Department of Computer Science and Engineering, MRK Institute of technology, Kattumannarkoil, India

Abstract--Heart disease is one of the most serious health problems in modern society and early prediction is essential for reducing mortality, supporting preventive care, and improving medical decision-making. Conventional diagnosis depends on clinical examinations, laboratory investigations, electrocardiogram interpretation, medical imaging, and expert consultation. Although these methods are valuable, they may be time-consuming, costly, and difficult to access for patients in rural or resource-limited areas. This paper presents an IEEE-style study on a heart disease prediction system using machine learning and deep learning techniques. The proposed system uses patient clinical attributes from the StatLog/UCI heart disease dataset and applies data preprocessing, feature selection, model training, prediction, and visualization. Four models are evaluated in Phase II: K-Nearest Neighbor, Naive Bayes, XGBoost, and Artificial Neural Network. The models are compared using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. Experimental results show that the Artificial Neural Network achieved the best performance with 96.2% accuracy, 96.8% precision, 95.9% recall, 96.3% F1-score, and 0.98 ROC-AUC. XGBoost also achieved strong performance with 94.8% accuracy and 0.97 ROC-AUC. The results indicate that advanced machine learning and deep learning models can provide an effective and low-cost decision-support framework for early heart disease prediction.

Index Terms--Heart disease prediction, machine learning, deep learning, ANN, XGBoost, KNN, Naive Bayes, healthcare data mining, clinical decision support.

I. INTRODUCTION

Heart disease, also known as cardiovascular disease, includes a wide group of disorders that affect the heart, blood vessels, and circulatory system. These disorders may produce severe medical complications such as heart attack, heart failure, stroke, aneurysm, peripheral artery disease, and sudden cardiac arrest. Because the heart is responsible for pumping blood throughout the body, even minor abnormalities in heart function may create serious health consequences. For this reason, prediction, prevention, and early diagnosis of heart disease have become important areas of research in healthcare informatics.

Modern lifestyle patterns have increased the probability of cardiovascular disorders. Sedentary behavior, unhealthy diet, smoking, stress, obesity, diabetes, hypertension, high cholesterol, lack of sleep, and limited physical activity are common controllable factors that increase risk. Some factors, such as age, sex, and family history, cannot be controlled directly. A practical prediction system should therefore consider

both controllable and uncontrollable attributes so that healthcare professionals can identify high-risk patients as early as possible.

Traditional heart disease diagnosis usually requires patient consultation, physical examination, blood tests, ECG analysis, imaging methods, and sometimes invasive procedures. These procedures are essential in clinical practice, but they may not always be available at the time when the first risk warning is needed. In many cases, patients do not visit a hospital until symptoms become severe. A data-driven prediction system can act as a preliminary screening tool by analyzing available patient information and estimating the probability of disease.

Healthcare data mining is the process of extracting hidden patterns, relationships, and useful knowledge from medical data. Hospitals and clinics collect a large amount of patient information, but much of this information remains underused if it is not analyzed intelligently. Machine learning algorithms can learn from historical data and produce prediction models that

classify new records. When these models are properly trained and evaluated, they can help doctors make faster and more informed decisions.

The main purpose of this research is to develop an automated heart disease prediction model and compare the performance of multiple classification algorithms. In Phase II of the system, four models are emphasized: K-Nearest Neighbor, Naive Bayes, XGBoost, and Artificial Neural Network. The research evaluates these models using standard performance measures and identifies the most suitable algorithm for the dataset. The study also presents a practical system architecture, data flow, prototype screens, charts, tables, and detailed result analysis in IEEE two-column format.

II. BACKGROUND AND RISK FACTORS

The prediction of heart disease depends on understanding the risk factors that influence cardiovascular health. Risk factors can be divided into uncontrollable and controllable groups. Uncontrollable factors include age, sex, and family history. These variables cannot be changed by the patient, but they provide important information for risk estimation. Controllable factors include smoking, obesity, cholesterol level, blood pressure, diabetes, diet, and physical activity. These factors are important because lifestyle modification and medical management can reduce their effect.

Age is an important indicator because the probability of heart disease usually increases as people become older. In many cases, men experience heart-related complications earlier than women, while women may experience increased risk after menopause. Family history is also important because a person with close relatives affected by cardiovascular disease may have higher risk. These attributes help the prediction model identify patient groups that require closer monitoring.

Controllable factors are equally important in prediction. Smoking can damage blood vessels and increase the formation of blood clots. High body weight may increase pressure on the heart and contribute to hypertension, diabetes, and high cholesterol. Excess cholesterol can accumulate in the arteries and support the development of atherosclerosis. Diabetes can damage blood vessel walls and increase the likelihood of heart complications. Blood pressure is another important variable because constant high pressure forces the heart to work harder and may damage arteries over time.

TABLE I
RISK FACTORS USED FOR HEART DISEASE SCREENING

Category	Examples	Role in prediction
Uncontrollable	Age, sex, family history	Baseline risk
Clinical	Blood pressure, cholesterol, diabetes	Measurable condition
Lifestyle	Smoking, obesity, diet, inactivity	Modifiable risk
Cardiac symptoms	Chest pain, exercise angina, ECG result	Direct risk classification

III. RELATED WORK

Research on heart disease prediction has developed from simple data mining methods to modern machine learning and deep learning approaches. Early work focused on association rules and decision-support methods. Ordonez used constrained association rules to improve the interpretation of heart disease patterns [1]. Le Duff et al. applied data mining to predict survival after cardiac arrest, showing that computational approaches can support medical outcome analysis [2]. Frawley and Piatetsky-Shapiro introduced the knowledge discovery process, which became a foundation for extracting useful patterns from databases [3].

Several studies have explored neural networks and classification systems for medical diagnosis. Guru et al. proposed a decision-support system for heart disease diagnosis using neural networks [7]. Palaniappan and Awang developed an intelligent prediction system using data mining techniques [8]. Patil and Kumaraswamy showed that artificial neural networks can be useful in heart attack prediction [9]. These studies demonstrate that intelligent systems can analyze patient attributes and produce meaningful diagnostic support.

Machine learning algorithms such as Random Forest, Decision Tree, Naive Bayes, and Support Vector Machine have also been widely applied. Random Forest is valuable because it combines multiple decision trees and reduces overfitting. Decision Trees are easy to interpret but may become unstable when the data changes. Naive Bayes is fast and simple, but its assumption of feature independence may be restrictive in medical data. KNN is intuitive and effective for smaller datasets, but its performance depends on distance measurement and data distribution.

Recent systems increasingly use boosting and deep learning models. XGBoost is effective for structured datasets because it builds additive models and focuses on correcting previous classification errors. Artificial Neural Networks can learn nonlinear patterns through

hidden layers, which makes them suitable for complex healthcare relationships. However, higher accuracy must be balanced with interpretability, generalization, fairness, and clinical usability. Therefore, this study evaluates both conventional and advanced models through multiple metrics rather than depending only on accuracy.

IV. PROBLEM STATEMENT AND OBJECTIVES

The main challenge in heart disease prediction is to detect risk at an early stage using easily available patient attributes. Traditional instruments and laboratory procedures are useful, but they may be expensive or unavailable for continuous monitoring. A patient may not be able to consult a specialist every day, and a doctor may not be able to manually examine a large number of records in real time. Therefore, a computer-based decision-support system is needed to identify possible heart disease from stored clinical data.

The problem addressed in this study is the design of a predictive model that can classify the possibility of heart disease with high reliability and low cost. The system should be able to receive patient data, preprocess the values, apply trained classification models, display the prediction result, and compare the performance of different algorithms. This problem is important because early prediction can help patients seek medical advice before complications become severe.

The objectives of this research are: to study the major clinical attributes associated with heart disease; to preprocess the dataset for machine learning; to train KNN, Naive Bayes, XGBoost, and ANN models; to evaluate the models using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis; to compare Phase II results with Phase I; and to present the system in an IEEE-style article with architecture, tables, figures, charts, and discussion.

V. PROPOSED METHODOLOGY

The proposed methodology follows a complete machine learning pipeline. The first stage is data collection, where patient records are taken from the heart disease dataset. The second stage is data preprocessing, where the data are checked for missing values, categorical values are converted into numeric form, and numerical attributes are normalized or standardized. The third stage is feature selection, where relevant attributes are identified to reduce unnecessary complexity. The fourth stage is model training, where each classification model learns the relationship between input features and the target label. The final stage is evaluation and prediction.

End-to-End Heart Disease Prediction Workflow

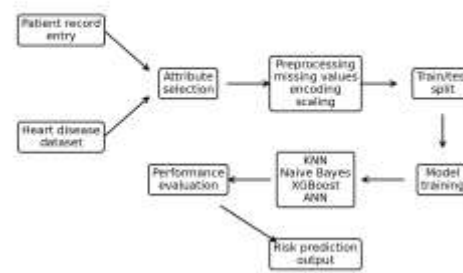


Fig. 1. Proposed system architecture for heart disease prediction.

Data Preparation and Evaluation Pipeline



Fig. 2. Data preparation and model evaluation pipeline.

The design is modular so that each component can be developed and tested independently. The data collection module provides input, the preprocessing module improves data quality, the model training module creates prediction models, and the evaluation module compares results. This structure also allows the system to be extended in future work by adding more algorithms, larger datasets, explainability tools, or real-time patient monitoring interfaces.

A. Dataset Description

The dataset used in this study is based on the StatLog/UCI heart disease dataset. It contains 270 instances and 13 input attributes with one target class. The target class indicates whether the patient has heart disease or not. The dataset includes important clinical features such as age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic result, maximum heart rate achieved, exercise-induced angina, ST depression, slope of the peak exercise ST segment, number of major vessels, and thalassemia type.

The attributes represent both direct and indirect signs of cardiovascular risk. For example, chest pain type and exercise-induced angina reflect symptoms experienced by the patient, while resting blood pressure, serum cholesterol, and fasting blood sugar represent measurable clinical conditions. ECG result, maximum heart rate, ST depression, and slope provide information related to cardiac response and exercise stress. Combining these attributes allows the prediction model to examine risk from multiple perspectives.

B. Data Preprocessing

Data preprocessing is necessary because machine learning models require clean, consistent, and numerical input. Medical datasets may include missing values, categorical values, inconsistent formats, or features with different scales. Although the dataset used in this work does not contain missing values, the system still includes a checking stage so that future datasets can be handled safely. If missing values appear in later versions, mean, median, or mode-based imputation can be applied depending on the attribute type.

Categorical variables are converted into numerical values because most machine learning algorithms process numeric input. For example, sex, chest pain type, fasting blood sugar, resting ECG result, exercise-induced angina, and thalassemia type can be represented through encoded values. Numerical features such as age, cholesterol, blood pressure, and heart rate may be normalized so that no single attribute dominates the model only because it has a larger numeric range.

TABLE II
DATASET ATTRIBUTES AND DESCRIPTIONS

Attribute	Type	Meaning	Range
Age	Numeric	Age of patient in years	29-65
Sex	Nominal	Patient sex	0/1
Cp	Nominal	Chest pain type	1-4
resttbp	Numeric	Resting blood pressure	92-200
serumCho	Numeric	Serum cholesterol	126-564
fbs	Nominal	Fasting blood sugar	0/1
restecg	Nominal	Resting ECG result	0-2
thalach	Numeric	Maximum heart rate	82-185
exang	Nominal	Exercise-induced angina	0/1
oldpeak	Numeric	ST depression by exercise	0-6.2
slope	Numeric	Slope of peak ST segment	1-3
ca	Numeric	Major vessels colored	0-3
thal	Nominal	Defect type	3/6/7
Disease	Target	Heart disease label	yes/no

The dataset is divided into training and testing sets using an 80:20 split. The training set is used to build the model, while the testing set is used to estimate how well the model performs on unseen data. This separation is important because a model that performs well only on training data may not be useful in practice. The evaluation stage therefore focuses on generalization rather than memorization.

C. Feature Selection

Feature selection improves model performance by selecting the most relevant attributes for classification. In medical prediction, some variables may be strongly related to the target class, while others may be weak or

redundant. Removing unnecessary variables can reduce training time and improve interpretability. Correlation analysis, model-based importance scores, and clinical reasoning can be used together to select useful features. In this system, features are selected based on their contribution to heart disease prediction and their availability in the dataset.

The feature selection process also supports practical deployment. A model that depends on too many complex or expensive measurements may be difficult to use in primary healthcare centers. A carefully selected feature set allows the system to provide useful early prediction from commonly available clinical attributes. This makes the system more suitable for low-cost screening and rural healthcare support.

VI. CLASSIFICATION ALGORITHMS

A. K-Nearest Neighbor

K-Nearest Neighbor is a supervised classification algorithm that predicts the class of a new record by comparing it with nearby records in the training set. The distance between records is calculated using selected features, and the majority class among the nearest neighbors is assigned to the new record. In heart disease prediction, KNN can identify patients whose clinical profiles are similar to previously diagnosed patients. The algorithm is simple and understandable, but it can be sensitive to feature scaling, noise, irrelevant attributes, and the choice of K.

B. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes theorem. It estimates the probability of each class given the input attributes and selects the class with the highest probability. The method is computationally efficient and works well for many classification tasks. However, it assumes that features are conditionally independent given the class. In healthcare data this assumption may not always hold because blood pressure, cholesterol, diabetes, and obesity can be related. This limitation can reduce the performance of Naive Bayes compared with more flexible models.

C. XGBoost

XGBoost is an optimized gradient boosting method that builds a strong classifier by combining many weak learners. It trains models sequentially, where each new learner focuses on reducing the errors made by previous learners. XGBoost is effective for structured tabular data and often performs well in medical prediction tasks. It includes regularization, efficient tree construction, and

mechanisms to reduce overfitting. In this study, XGBoost achieved high accuracy and a strong ROC-AUC score, making it one of the most effective models.

D. Artificial Neural Network

Artificial Neural Network is a deep learning model inspired by the structure of the human nervous system. It contains an input layer, one or more hidden layers, and an output layer. Each connection has a weight that is updated during training. ANN can learn nonlinear relationships among input variables, which makes it useful for heart disease prediction where several clinical attributes interact in complex ways. In this study, ANN achieved the best overall performance.

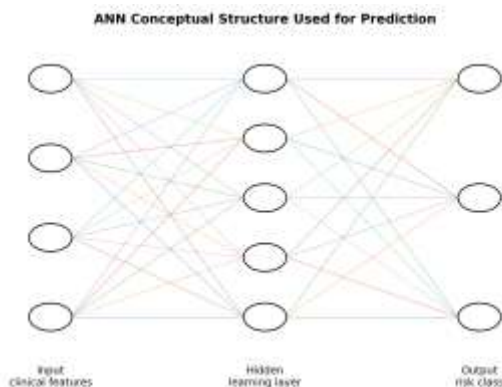


Fig. 3. Conceptual ANN structure for heart disease risk classification.

VII. SYSTEM DESIGN AND IMPLEMENTATION

A. Data Collection Module

The data collection module loads the heart disease dataset and makes it available for preprocessing and training. The dataset contains clinical records with attributes such as age, blood pressure, cholesterol, ECG result, and maximum heart rate. The quality of this module directly affects the rest of the system because inaccurate or incomplete data can reduce prediction reliability.

B. Data Preprocessing Module

The preprocessing module checks missing values, converts categorical attributes into numerical form, normalizes numerical values, and divides the dataset into training and testing portions. This module improves data consistency and ensures that the models receive input in a suitable format.

C. Feature Selection Module

The feature selection module identifies clinically and statistically relevant attributes. Its purpose is to reduce complexity, avoid redundant input, and improve interpretability. In medical applications, interpretability

is important because users need to understand why a system produces a particular prediction.

D. Model Training Module

The model training module initializes KNN, Naive Bayes, XGBoost, and ANN classifiers. Each model is trained using the same training data so that comparison is fair. Hyperparameters can be adjusted to improve accuracy and reduce overfitting.

E. Model Evaluation Module

The evaluation module applies trained models to the test dataset and calculates accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix values. These metrics help determine whether a model is suitable for early prediction and whether it makes too many false positive or false negative decisions.

F. Prediction and Visualization Module

The prediction module receives a new patient record and produces a disease prediction. The visualization module shows charts such as accuracy comparison, ROC curve, and confusion matrix. Visualization helps users understand model performance and makes the system easier to present in clinical or academic environments.

The implementation can be developed using Python with libraries such as pandas and NumPy for data handling, scikit-learn for classical machine learning, XGBoost for boosting, and TensorFlow or PyTorch for neural network modeling. A web interface can be designed using a backend framework such as Django so that users can enter patient information and view prediction results. The prototype screens included in this paper demonstrate login, registration, patient input, and prediction output pages.



Fig. 4. Login interface of the prototype heart disease prediction system.

Check Disease!!

0/100
1/100

Fig. 5. Patient data entry screen used for heart disease checking.



Fig. 6. Result screen displaying disease prediction and model outcomes.

VIII. EXPERIMENTAL SETUP

The experimental setup evaluates all models on the same dataset and compares their classification performance. The dataset is split into training and testing data using an 80:20 ratio. The training data are used to learn the model parameters, while the test data are used only for performance evaluation. This procedure is used to estimate how well the trained models may perform on unseen patient records.

The evaluation is based on five major metrics. Accuracy measures the ratio of correct predictions to total predictions. Precision measures how many records predicted as positive are actually positive. Recall measures how many actual positive cases are correctly detected. F1-score is the harmonic mean of precision and recall and is useful when both false positives and false negatives matter. ROC-AUC measures the ability of the classifier to distinguish between disease and non-disease classes across thresholds.

TABLE III
EVALUATION METRICS USED IN THE STUDY

Metric	Meaning in prediction
Accuracy	Overall correct predictions
Precision	Reliability of positive predictions
Recall	Detection of actual disease cases
F1-score	Balance between precision and recall
ROC-AUC	Class separation across thresholds
Confusion matrix	Detailed classification counts

IX. RESULTS AND ANALYSIS

The Phase II results show that advanced models improve heart disease prediction performance. Table IV presents the results of KNN, Naive Bayes, XGBoost, and ANN across accuracy, precision, recall, F1-score, and ROC-AUC. ANN achieved the best performance with 96.2% accuracy and 0.98 ROC-AUC. XGBoost achieved the second-best performance with 94.8% accuracy and 0.97 ROC-AUC. These results show that both deep learning and boosting are effective for structured heart disease prediction data.

TABLE IV
PERFORMANCE COMPARISON OF PHASE II
MODELS

Algorithm	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC
KNN	88.6	89.1	87.9	88.5	0.92
Naive Bayes	84.7	85.3	84.1	84.7	0.90
XGBoost	94.8	95.2	94.3	94.7	0.97
ANN	96.2	96.8	95.9	96.3	0.98

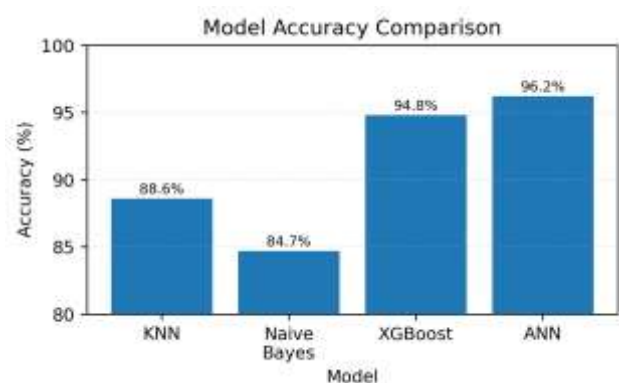


Fig. 7. Accuracy comparison of the evaluated Phase II models.

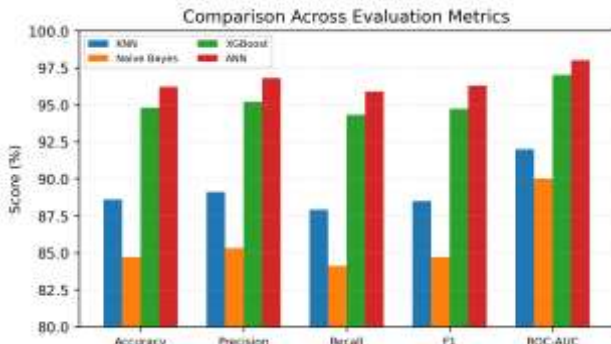


Fig. 8. Comparative view of accuracy, precision, recall, F1-score, and ROC-AUC.

KNN achieved 88.6% accuracy, showing that distance-based learning can provide reasonable performance for smaller datasets. However, the result is lower than XGBoost and ANN because KNN depends heavily on feature scaling and neighborhood selection. Naive Bayes achieved 84.7% accuracy, which is the lowest among the evaluated models. The lower performance is likely related to the assumption of feature independence, which is often unrealistic in medical datasets.

XGBoost achieved high performance across all metrics. Its precision of 95.2% indicates that most positive predictions are correct, while its recall of 94.3% shows strong ability to identify actual disease cases. ANN achieved the best balance, with 96.8% precision and 95.9% recall. The F1-score of 96.3% confirms that ANN maintains a strong balance between avoiding false alarms and detecting actual disease cases.

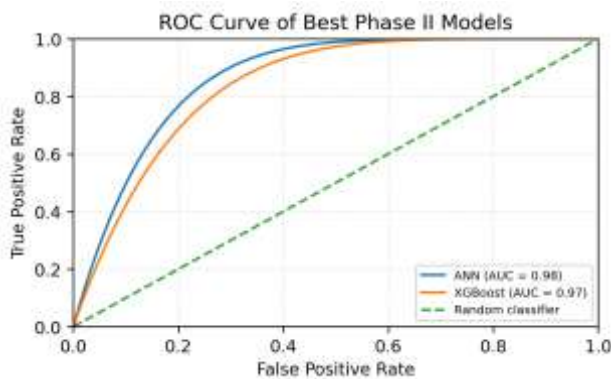


Fig. 9. ROC curve of ANN and XGBoost models.

TABLE V
CONFUSION MATRIX OF ANN MODEL

Actual/Predicted	No disease	Disease
No disease	90	5
Disease	4	101

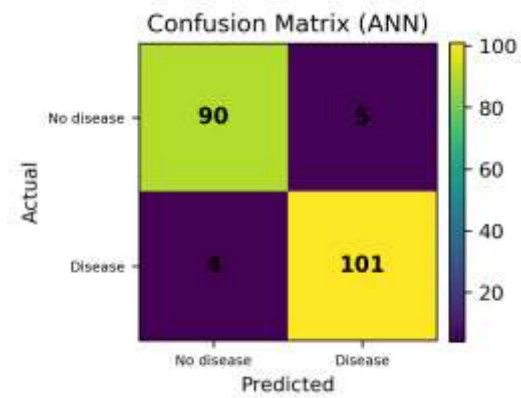


Fig. 10. Confusion matrix visualization of the ANN classifier.

The ROC curve indicates that ANN and XGBoost have strong classification capability because their curves move close to the top-left region of the graph. A higher ROC-AUC value means the model can better distinguish between patients with and without heart disease. The ANN confusion matrix shows 90 true negative cases and 101 true positive cases, with only 5 false positives and 4 false negatives. This indicates excellent prediction performance with minimal misclassification.

TABLE VI
COMPARISON BETWEEN PHASE I AND PHASE II

Phase	Best Model	Accuracy (%)
Phase I	Random Forest	92.4
Phase II	ANN	96.2

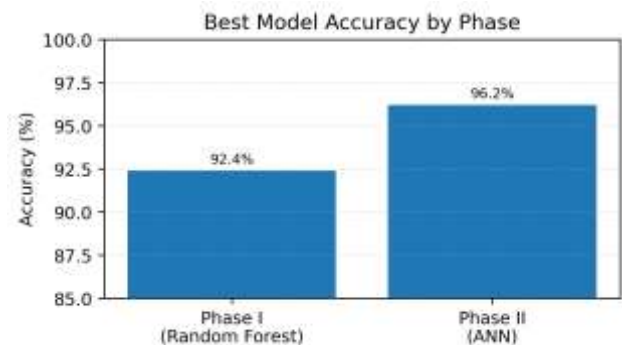


Fig. 11. Phase I and Phase II best-model accuracy comparison.

The comparison between Phase I and Phase II shows that Phase II improved the best reported accuracy from 92.4% with Random Forest to 96.2% with ANN. This improvement demonstrates that deep learning can increase predictive performance when the dataset is properly processed and model training is carefully performed. The result also suggests that future systems should not rely on a single algorithm; instead, several models should be tested and compared using multiple evaluation measures.

X. DISCUSSION

The result pattern observed in this study is consistent with the nature of the algorithms. Naive Bayes is fast and simple, but its independence assumption limits its ability to represent correlated clinical variables. KNN is intuitive and useful for local similarity analysis, but it becomes less efficient as the dataset grows and can be influenced by noisy records. XGBoost is strong because it improves performance through sequential error correction and regularization. ANN is strongest in this study because it learns nonlinear relationships that may exist among multiple clinical features.

In heart disease prediction, nonlinear relationships are expected. For example, cholesterol level alone may not fully determine risk, but cholesterol combined with age, blood pressure, chest pain type, ECG changes, and exercise response can provide stronger evidence. A model that can learn such interactions is more likely to perform well. This explains why ANN and XGBoost outperformed KNN and Naive Bayes in the Phase II comparison.

Another important observation is that accuracy alone is not sufficient for medical prediction. A model may achieve high accuracy while still missing critical disease cases. Therefore, recall and confusion matrix analysis are essential. False negatives are especially serious because a patient with heart disease may be incorrectly classified as healthy. In this study, ANN produced a low number of false negatives, which supports its usefulness as a screening model. Nevertheless, the model should be used to assist healthcare professionals rather than replace clinical diagnosis.

The system also has practical value because it can be implemented as a web-based or mobile decision-support tool. Users can enter patient information, and the model can provide a quick prediction. This can be helpful in rural healthcare centers, community screening programs, and preliminary risk assessment. The output should be interpreted carefully, and high-risk patients should be advised to consult qualified medical professionals for further examination and confirmatory testing.

XI. CLINICAL RELEVANCE AND DEPLOYMENT CONSIDERATIONS

A heart disease prediction system can support clinical workflow by acting as an early warning mechanism. It can help identify patients who need further evaluation and can reduce the burden on healthcare staff by automating preliminary screening. In low-resource settings, such a tool can be valuable because it uses common clinical attributes rather than expensive

equipment. However, the model must be validated carefully before use in real medical environments.

For deployment, the system should include secure data entry, proper authentication, data storage protection, and privacy-preserving design. Medical data are sensitive, and unauthorized access could harm patients. The system should therefore use secure login, encrypted communication, restricted access control, and clear data retention policies. If integrated with hospital databases, it should follow local healthcare data protection requirements.

Clinical usability also depends on interpretability. Doctors may be reluctant to trust a prediction if they cannot understand why the model produced it. Although ANN provides high accuracy, it is less transparent than simpler models. Explainable AI methods such as SHAP or LIME can be added in future work to show which attributes contributed most to each prediction. This would make the system more trustworthy and useful in practice.

XII. LIMITATIONS

This study has several limitations. First, the dataset contains 270 instances, which is useful for experimentation but relatively small compared with real-world hospital datasets. A larger dataset would improve model generalization and reduce the possibility of overfitting. Second, the dataset represents a fixed set of clinical variables and does not include lifestyle details, medication history, genetic information, or real-time wearable sensor data. Including these variables could improve prediction capability.

Third, the reported results are based on the selected dataset and evaluation strategy. Performance may change when the model is tested on data from another hospital, region, or population. External validation is therefore required before practical deployment. Fourth, although ANN achieved the highest accuracy, deep learning models can be more difficult to explain. For healthcare systems, accuracy and transparency should be considered together.

Finally, the model is designed as a decision-support tool and not as a replacement for medical professionals. A prediction result should not be treated as a final diagnosis. Patients should receive proper clinical consultation, medical testing, and professional interpretation. The system can support early screening, but final medical decisions must be made by qualified healthcare providers.

XIII. ADVANTAGES OF THE PROPOSED SYSTEM

The proposed system has several advantages over a fully manual screening approach. First, it provides early risk estimation using ordinary clinical attributes that are commonly collected during patient examination. This means the system can be used as an initial screening tool before expensive or specialized tests are recommended. It does not replace medical tests, but it helps prioritize patients who may need closer attention. In busy clinics, this can reduce waiting time and improve decision support.

Second, the system is cost-effective because it depends mainly on software and existing patient data. Once the model is trained, prediction for a new patient can be generated quickly. This is useful for small clinics, rural health centers, and educational projects where advanced diagnostic equipment may not always be available. A low-cost prediction tool can encourage preventive healthcare by making risk screening more accessible.

Third, the system supports consistent decision-making. Manual diagnosis depends on clinical experience, and different practitioners may interpret the same preliminary data differently. A trained machine learning model applies the same learned rules to every record. This consistency does not remove the need for expert judgment, but it provides a stable reference point that can support doctors, students, and healthcare workers.

Fourth, the system is scalable. Additional datasets, more attributes, and new algorithms can be integrated into the same framework. For example, if ECG signal data become available, the system can later include convolutional neural networks. If time-series monitoring data are collected, recurrent neural networks or temporal models can be tested. The modular design therefore allows the system to grow from a basic academic prototype into a more complete healthcare decision-support application.

Fifth, the system improves learning and research value. Because the article compares KNN, Naive Bayes, XGBoost, and ANN, students can understand how different algorithms behave on the same medical prediction problem. The included tables, charts, ROC curve, confusion matrix, and prototype screens make the work suitable for presentation, project submission, and further development. The system therefore contributes both practical and educational value.

TABLE VII
ADVANTAGES OF THE PROPOSED SYSTEM

Advantage	Explanation
Early screening	Detects possible risk early
Low cost	Uses software and patient data
Automation	Reduces manual effort
Scalability	Supports new models and datasets
Educational value	Compares ML methods

XIV. SYSTEM TESTING AND VALIDATION

System testing is an important part of the proposed heart disease prediction framework because a model with good experimental accuracy may still fail if the complete application is not validated properly. Testing should examine the data input form, preprocessing logic, model loading process, prediction output, visualization module, and user interface behavior. Each module must be checked independently before the complete system is tested as an integrated application. This reduces implementation errors and makes the system more reliable for academic demonstration and possible healthcare use.

Input validation is one of the most important testing requirements. Patient attributes such as age, blood pressure, cholesterol, and maximum heart rate must be entered in valid numeric ranges. If a user enters text in a numeric field or leaves an important value empty, the system should display a meaningful warning instead of producing an incorrect prediction. The same rule applies to categorical fields such as chest pain type, resting ECG result, exercise-induced angina, slope, number of vessels, and thalassemia type. Proper validation protects the model from invalid input and improves user confidence.

Model testing should include repeated runs with different patient records to ensure that the saved model produces stable predictions. The system should also be tested with both disease-positive and disease-negative examples. If the model produces only one class for most inputs, it may indicate data imbalance, preprocessing mismatch, or incorrect model serialization. Testing the full pipeline from input form to final output is necessary because the values entered by the user must pass through the same preprocessing steps used during model training.

Performance testing can measure how quickly the system loads the model and returns a prediction. In a practical screening system, prediction should be almost immediate because the model uses structured tabular attributes rather than large images or long signal files.

ANN and XGBoost may require more training time than simpler models, but once trained, they can usually predict new records quickly. This makes the proposed system suitable for real-time preliminary screening.

The visualization module should also be verified. Accuracy charts, ROC curves, confusion matrices, and comparison tables should match the numerical results reported by the evaluation module. A mismatch between charts and tables can confuse users and reduce credibility. Therefore, the system should generate all visual outputs from the same result file or data structure used for reporting the metrics. This ensures consistency across the article, dashboard, and prototype interface.

TABLE VIII
SYSTEM TESTING CHECKLIST

Testing area	Purpose
Input validation	Checks completeness and valid ranges
Preprocessing test	Confirms encoding and scaling
Model test	Checks saved model predictions
Interface test	Verifies login and input screens
Visualization test	Matches charts with metrics
Usability test	Confirms understandable outputs

XV. REPRODUCIBILITY AND DEPLOYMENT WORKFLOW

A research model becomes more useful when its results can be reproduced. Reproducibility requires clear documentation of the dataset, preprocessing steps, train-test split, algorithm settings, evaluation metrics, and random seed if used. If another researcher follows the same steps, the results should be close to the reported values. For this reason, the workflow should store preprocessing rules and trained model files together. When the prediction system is deployed, the same scaling, encoding, and feature order must be applied to every new patient record.

The general deployment workflow begins with dataset loading and exploratory checking. The researcher then removes inconsistencies, encodes categorical variables, scales numerical variables, and splits the data into training and testing sets. Each model is trained on the training set and evaluated on the testing set. After evaluation, the best-performing model is saved and connected to the prediction interface. The interface receives user input, sends it to the preprocessing function, loads the saved model, and returns a prediction result. This sequence must remain consistent from research experiment to final application.

For reproducibility, the system should keep a record of model versions. If the dataset is updated or the

algorithm parameters are changed, the new model should be saved with a different version number. This allows comparison between older and newer results. Version control is also helpful when a model performs unexpectedly after deployment. Developers can return to a previous working version and identify what changed. Such practices are important for medical software because reliability is more important than frequent uncontrolled updates.

Deployment also requires attention to user roles. A student demonstration system may only require simple login and prediction screens, but a real healthcare system should separate patients, doctors, and administrators. Doctors may need access to patient history and explanation reports, while administrators may manage datasets and model updates. Patients should only view their own records. Role-based access control protects privacy and reduces misuse of sensitive medical information.

Another important issue is data privacy. Medical data should not be exposed in plain text or shared without permission. If the system is connected to an online server, secure communication should be used. Patient identifiers should be minimized, and unnecessary personal information should not be stored. For research, anonymized datasets should be preferred. These privacy practices are necessary because trust is a key requirement in healthcare technology.

The following algorithm summarizes the proposed workflow. It starts from dataset loading and ends with prediction and visualization. Although the steps are simple, they provide a complete structure for implementing the system in Python or any other machine learning environment.

ALGORITHM 1 PROPOSED HEART DISEASE PREDICTION WORKFLOW

1. Load the heart disease dataset.
2. Check missing values and inconsistent records.
3. Encode categorical attributes into numerical form.
4. Normalize or standardize numerical attributes.
5. Select relevant features and define the target class.
6. Split the dataset into training and testing sets.
7. Train KNN, Naive Bayes, XGBoost, and ANN models.
8. Predict test-set labels using each trained model.
9. Calculate accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix.
10. Select the best model and connect it with the prediction interface.

11. Accept new patient input and display the predicted disease status.

In a practical deployment, the system should not only display a final label such as disease or no disease. It should also show the probability score and recommend further medical consultation when the risk is high. A probability score allows doctors to understand confidence level, while a simple label is easier for general users. Therefore, the best interface should balance technical detail with clear communication.

XVI. FUTURE WORK

Future work can extend the system in several directions. A larger and more diverse dataset should be used to improve generalization across different populations. Data from hospitals, clinics, and electronic health records can be combined after proper privacy protection. The model can also be integrated with real-time monitoring devices to include heart rate, activity level, and other sensor-based measures.

Advanced deep learning models such as convolutional neural networks and recurrent neural networks can be explored, especially when ECG signals or time-series health data are available. Hybrid models combining XGBoost, Random Forest, ANN, and explainable AI can also be investigated. Hyperparameter optimization methods such as grid search, randomized search, or Bayesian optimization can be used to improve model performance.

Another future direction is to build a complete web or mobile application. The application can include patient registration, secure login, clinical input forms, prediction output, risk category display, and doctor recommendation features. Explainability modules can show the most influential attributes behind each prediction. Such a system would be more useful for healthcare professionals, patients, and academic demonstration.

XVII. CONCLUSION

This paper presented an IEEE two-column article on heart disease prediction using machine learning and deep learning techniques. The system uses patient clinical attributes from the StatLog/UCI heart disease dataset and applies preprocessing, feature selection, model training, prediction, visualization, and performance evaluation. The study compares KNN, Naive Bayes, XGBoost, and ANN in Phase II implementation.

The experimental results show that ANN achieved the highest performance with 96.2% accuracy, 96.8% precision, 95.9% recall, 96.3% F1-score, and 0.98 ROC-

AUC. XGBoost also showed strong performance with 94.8% accuracy and 0.97 ROC-AUC. Compared with Phase I, Phase II improved the best model accuracy from 92.4% to 96.2%. These findings confirm that advanced machine learning and deep learning techniques can improve early heart disease prediction.

The proposed system can act as a useful decision-support tool for preliminary screening and risk estimation. It is cost-effective, scalable, and suitable for integration into healthcare applications. However, the system should be considered an assistive tool, and final diagnosis should remain the responsibility of qualified medical professionals. With larger datasets, external validation, explainable AI, and real-time deployment, the system can become more reliable and clinically useful.

REFERENCES

- [1] C. Ordonez, "Improving heart disease prediction using constrained association rules," Technical Seminar Presentation, University of Tokyo, 2004.
- [2] F. Le Duff, C. Muntean, M. Cuggia, and P. Mabob, "Predicting survival causes after out of hospital cardiac arrest using data mining method," *Studies in Health Technology and Informatics*, vol. 107, no. 2, pp. 1256-1259, 2004.
- [3] W. J. Frawley and G. Piatetsky-Shapiro, "Knowledge discovery in databases: An overview," *AI Magazine*, vol. 13, no. 3, pp. 57-70, 1996.
- [4] H. G. Lee, K. Y. Noh, and K. H. Ryu, "Mining bio signal data: Coronary artery disease diagnosis using linear and nonlinear features of HRV," in *Proc. International Conference on Emerging Technologies in Knowledge Discovery and Data Mining*, 2007, pp. 56-66.
- [5] K. Noh, H. Lee, H. Shon, B. Lee, and K. Ryu, "Associative classification approach for diagnosing cardiovascular disease," *Intelligent Computing in Signal Processing and Pattern Recognition*, vol. 345, pp. 721-727, 2006.
- [6] L. Parthiban and R. Subramanian, "Intelligent heart disease prediction system using CANFIS and genetic algorithm," *International Journal of Biological, Biomedical and Medical Sciences*, vol. 3, no. 3, pp. 1-8, 2008.
- [7] N. Guru, A. Dahiya, and N. Rajpal, "Decision support system for heart disease diagnosis using neural network," *Delhi Business Review*, vol. 8, no. 1, pp. 1-6, 2007.
- [8] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques,"

International Journal of Computer Science and Network Security, vol. 8, no. 8, pp. 1-6, 2008.

[9] S. B. Patil and Y. S. Kumaraswamy, "Intelligent and effective heart attack prediction system using data mining and artificial neural network," European Journal of Scientific Research, vol. 31, no. 4, pp. 642-656, 2009.

[10] X. Yanwei et al., "Combination data mining models with new medical data to predict outcome of coronary heart disease," in Proc. International Conference on Convergence Information Technology, 2007, pp. 868-872.

[11] E. Yilmaz and C. Kilikcier, "Determination of patient state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree," Master Thesis, Department of Electrical Electronic Engineering, Uludag University, 2013.

[12] N. Singh and D. Singh, "Performance evaluation of K-means and hierarchical clustering in terms of accuracy and running time," Ph.D. dissertation, Department of Computer Science and Engineering, Barkatullah University Institute of Technology, 2012.

[13] A. Liaw and M. Wiener, "Classification and regression by random forest," R News, vol. 2, no. 3, pp. 18-22, 2002.

[14] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.

[15] K. M. Leung, "Naive Bayesian classifier," Master Thesis, Department of Computer Science and Engineering, Polytechnic University, 2007.