



Performance Assessment of a Clinical Support System for Heart Disease Prediction Using Machine Learning



Koteswara Rao Kodepogu^{1*}, Eswar Patnala², Jagadeeswara Rao Annam³, Shobana Gorintla⁴,
Veerla Vijaya Rama Krishna⁵, Vipparla Aruna⁴, Vijaya Bharathi Manjeti⁶

¹ Department of CSE, PVP Siddhartha Institute of Technology, 520007 Vijayawada, India

² Department of CSE, Koneru Lakshmaiah Education Foundation, 522301 Vaddeswaram, India

³ Department of CSE, CVR College of Engineering, 501510 Ibrahimpatnam, India

⁴ Department of CSE, NRI Institute of Technology, 521212 Agiripalli, India

⁵ Department of Electronics & Communication, Lakireddy Bali Reddy College of Engineering, 521230 Mylavaram, India

⁶ Department of CSE, GITAM School of Technology, 530045 Visakhapatnam, India

* Correspondence: Koteswara Rao Kodepogu (kkrao@pvpiddhartha.ac.in)

Received: 07-24-2024

Revised: 09-15-2024

Accepted: 09-23-2024

Citation: Kodepogu, K. R., Patnala, E., Annam, J. R., Gorintla, S., Krishna, V. V. R., Aruna, V., & Manjeti, V. B. (2024). Performance assessment of a clinical support system for heart disease prediction using machine learning. *Healthcraft. Front.*, 2(3), 159-168. <https://doi.org/10.56578/hf020305>.



© 2024 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Heart disease remains a leading cause of mortality worldwide, necessitating early and accurate detection to improve clinical outcomes. Traditional diagnostic approaches relying on conventional clinical data analysis often encounter limitations in precision and efficiency. Machine learning (ML) techniques offer a promising solution by enhancing predictive accuracy and decision-making capabilities. This study evaluates the performance of a clinical support system (CSS) for heart disease prediction using a hybrid classification approach that integrates support vector machine (SVM) and k-nearest neighbor (KNN). Patient data were stratified by age group and gender to assess the model's performance across diverse demographic profiles. Key performance metrics, including accuracy, recall, precision, F1-score, and area under the curve (AUC), were employed to quantify predictive efficacy. Experimental results demonstrated that the combined SVM-KNN model achieved superior classification performance, yielding an accuracy of 97.2%, recall of 97.6%, precision of 96.8%, AUC of 97.1%, and an F1-score of 98.2%. These findings indicate that the integration of SVM and KNN enhances heart disease prediction accuracy, thereby reinforcing the potential of CSS in improving early diagnosis and patient management.

Keywords: Clinical data; Support vector machine; Machine learning; Clinical support system; Heart disease; K-nearest neighbor

1. Introduction

In most countries, heart disease is the leading cause of death (Li et al., 2020). The related risk factors include obesity, stress, unhealthy lifestyle choices, and medical history. Complications from cardiac disease include heart attacks, heart failure, strokes, etc. (Mohan et al., 2019). Every other part of the body receives blood flow from the heart, which is the most important organ. A disruption in the heart's function can result in the rapid failure of other critical organs, particularly the brain, leading to irreversible damage or death within minutes (Sarkar & Koehler, 2012). Thus, the maintenance of optimal cardiac function is essential for survival and systemic homeostasis.

The World Health Organization (WHO) estimated that 17.9 million individuals worldwide lost their lives due to cardiovascular diseases (CVDs) in 2016 (Wang et al., 2019). Heart disease, which is frequently diagnosed as coronary heart disease (CHD), is one of the syndromes that fall under the category of CVDs, which damage the heart and blood vessels (Rahim et al., 2021). However, by leading a healthy lifestyle and avoiding risk factors, heart disease can be avoided. Therefore, understanding what causes these alarming factors can help in heart disease prediction and prevention. The main diagnostic technique is usually angiography, which is used to determine the

location of cardiac vessel stenosis (Jain & Bhaumik, 2017). As a result, increasing efforts have been directed toward the development of automated, data-driven diagnostic systems. These systems have been built upon structured medical datasets, incorporating historical treatment outcomes and leveraging recent advancements in clinical research and biomedical informatics (Kumar et al., 2023).

There are certain diseases that harm the heart. Heart attacks, arrhythmias, CHD, myocarditis, cardiomyopathy, angina pectoris, congenital heart disease, congestive heart failure, and heart cancer are some of the conditions that are taken into consideration. This condition is especially dangerous for those who have CHD or CVD (Schmidt et al., 2015). Some significant causes of heart disease are thought to exist, such as blood pressure, cholesterol, age, smoking, sugar, obesity, depression, poor diet, family history, and physical inactivity (Rezaeieh et al., 2014).

A person's potential risk for heart disease is determined by a shared set of basic risk factors, despite the fact that there are many different forms of heart disease (Cenitta et al., 2022). Information from several sources might be gathered to reach a conclusion, categorizing it under the proper headings and then doing an analysis to obtain the information that is required. The prediction of cardiac disease can be effectively performed using this technique. As the adage "prevention is better than cure" suggests, the death rate from heart disease can be prevented and decreased with early detection and treatment (Shokouhmand et al., 2023). Clinicians can utilize a prediction model built into clinical support system (CSS) to help them identify a patient's heart disease risk and recommend appropriate treatments to further decrease that risk.

In order to create screening tools with pattern recognition and classification, machine learning (ML) approaches are being used more often. The accuracy of these tools is higher than that of traditional techniques. ML is used to extract medical data for hidden truths. ML is a multidisciplinary field (Bertsimas et al., 2021), including knowledge analytics, data processing, statistics, and algebra, among many more, and it seeks to enable machines to learn (Ahmad et al., 2022). ML methods include unsupervised learning, supervised learning, and reinforcement learning. Supervised algorithms analyze input data and create functions based on known output values, thereby enabling accurate prediction of outputs for unseen inputs.

This study evaluates the performance of CSS for heart disease prediction using ML. This research uses a combined model that includes both support vector machine (SVM) and k-nearest neighbor (KNN) classifiers for prediction and employs several measures, such as area under the curve (AUC), F1-score, accuracy, recall, and precision, to analyze the classifiers' performance.

The rest of the study is organized below. Section 2 summarizes the literature review. Section 3 introduces the methodology for evaluating the performance of CSS using ML. Section 4 presents the experimental results and performance analysis of the hybrid model. Section 5 presents the conclusion.

2. Literature Review

In the study by Su et al. (2021), the primary controller for the internet of medical things is provided by the STM32. It is integrated with Internet of Things (IoT) devices such as a temperature sensor, a pulse sensor, and a sphygmomanometer cuff for data collection and instrument control. Using deep learning to create appropriate models and analyze information, this assembly is utilized to create a screening method for valvular heart disease. Patients' characteristic signal values can be properly analyzed and evaluated by this valvular heart disease screening method. Salau et al. (2024) suggested a more effective methodology for detecting heart problems. The research improved the performance of the medical decision support system to detect heart disease more accurately. Recently, heart disease has gained attention as one of the globe's most common and pervasive diseases. Using related symptoms to identify heart disease early on is an exciting task. Overall, using SVM and sequential feature selection, a state-of-the-art model was developed in this current study for identifying heart disease with a 98.60% predictive accuracy.

Khan et al. (2024) focused on the accurate prediction of CVDs, considering patients' health and socio-economic conditions while mitigating the challenges presented by imbalanced data. In this study, two DL models for accurate CVD prediction and classification, i.e., BICVDD-Net and EnsCVDD-Net, were recommended. The EnsCVDD-Net, with 88% accuracy, 88% F1-score, 91% precision, 85% recall, and 777 s execution time, outperformed all base models, according to the results. Similarly, BICVDD-Net outperformed the state-of-the-art DL models with 91% accuracy, 91% F1-score, 96% precision, 86% recall, and 247 s execution time. In order to simultaneously predict heart disease using binary and multimodal classifications, Yuan et al. (2021) introduced a prediction model based on ML. A Fuzzy-GBDT method was developed which integrates gradient boosting decision tree (GBDT) and fuzzy logic to maximize the generalization of binary classification prediction while minimizing data complexity. According to the evaluation results, when it addressed binary and multiple class prediction, the Bagging-Fuzzy-GBDT demonstrated excellent accuracy and stability.

Abdellatif et al. (2022) improved patient survival and the capacity to identify the presence of cardiac disease by developing a new approach to these issues. In order to predict cardiac diseases, improved weighted random forest (IWRF) and supervised infinite feature selection (Inf-FS_s) were utilized to choose the most important features, and the new IWRF hyperparameters were adjusted through Bayesian optimization. Regarding F-measure and accuracy

on both datasets, according to the results, the proposed Inf-FS_s-IWRF performed better than other models. In order to increase the signal-to-noise ratio (SNR) level among noisy recordings, Ali et al. (2023) proposed a method to effectively decrease background noise with just 1.32 M parameters in both artificially made noisy heart sound recordings and unseen real-world data from low-resource institutions and underserved populations by an average of 5.575 dB. The proposed deep learning-based phonocardiogram (PCG) denoising method greatly enhanced computer-aided auscultation systems' capacity to identify cardiac problems. Tao et al. (2019) created an automated technique for the rapid and accurate localization and diagnosis of ischemic heart disease (IHD). The T wave was divided using averaged magnetocardiography (MCG) data, and 164 features were then retrieved. Several ML classifiers were compared, such as SVM, eXtreme gradient boosting (XGBoost), KNN, and decision tree (DT), and the SVM-XGBoost model performed better than the other models in detecting IHD with an accuracy of 94.03%. The suggested ML approach increased the utilization of MCG data in clinics by giving clinicians a rapid and accurate diagnostic tool to evaluate the data.

Wang et al. (2020) created a two-level stacking model, with the base level as Level 1 and the meta level as Level 2. Predictions of the base-level classifier were used as input. First, the maximum information coefficient and the Pearson correlation coefficient were computed to identify the classifier with the lowest correlation. According to experimental results, by using the suggested model to detect CHD, clinicians differentiated between patients with CHD and those with normal coronary arteries with 95.43% accuracy, 95.84% sensitivity, and 94.44% specificity.

Shuvo et al. (2021) used CardioXNet, a unique lightweight end-to-end convolutional recurrent neural network (CRNN) architecture, and raw PCG data to autonomously detect aortic stenosis, mitral stenosis, mitral regurgitation, mitral valve prolapse, and normal cardiac auscultation. Three parallel convolutional neural network (CNN) routes were created during the representation learning phase in order to learn the coarse and fine-grained features from PCG and to use the 2D CNN-based squeeze-expansion to examine the prominent features from different receptive fields. Even though the results are computationally comparable, the proposed end-to-end architecture achieves remarkable performance in all evaluation parameters, with an average of 99.60% accuracy, 99.56% precision, 99.52% recall, and 99.68% F1-score compared to the prior state-of-the-art methods. Boukhatem et al. (2022) determined the type of real-time (less than 30 ms) electrocardiogram (ECG) collected and provided a new method to extract information related to ECG. In order to train models with out-of-sample F1-scores between 0.93 and 0.99, the XGBoost algorithm, a popular ML technique, was used. This study was distinguished by its capacity to maintain high predictive accuracy across multiple hospitals, diverse recording protocols, and varying international standards—an achievement not previously reported in existing literature, to the authors' knowledge.

El-Sofany (2024) provided an accurate ML approach, which uses several feature selection strategies, to detect heart disease early. Three different techniques were used for feature selection: mutual information (MI), chi-square, and analysis of variance (ANOVA). With 98% AUC, 97.57% accuracy, 92.68% F1-score, 96.61% sensitivity, 90.48% specificity, and 95.00% precision, the results of the experiment demonstrated that the SF-2 feature subset and combined datasets were the most effective for the XGBoost classifier. With its cheap cost and minimal time, in terms of early-stage heart disease prediction, the proposed approach showed a lot of promise for the healthcare sector. Karhade et al. (2022) coupled the time-frequency-domain deep learning (TFDDL) architecture with PCG signals to automatically detect HVDs. The frequency-domain polynomial chirplet transform (FDPCT) and the time-domain polynomial chirplet transform (TDPCT) in the time-frequency (TF) domain were used to analyze PCG signals. The suggested method for HVD detection using TF pictures of PCG signals based on TDPCT and FDPCT achieved overall accuracy values of 99.48% and 99%, respectively. The proposed deep CNN model offers superior overall accuracy values for HVD identification compared to visual geometry group network (VGGNet)-16 and residual network (ResNet)-50.

There are a number of limitations to the existing studies on heart disease prediction. Most studies rely on tiny and unbalanced datasets, leading to poor generalization results. The selection of overfitting and inconsistent features lowers model dependability. Many methods rely on uninterpretable black-box models (like neural networks), which makes clinical adoption challenging. Accuracy is further impacted by data quality problems, including missing values and measurement discrepancies. Furthermore, rather than real-time patient monitoring, the majority of models concentrate on static datasets. Future studies should focus on bigger datasets, explainable artificial intelligence (AI), real-time monitoring, and customized risk assessment for improved clinical integration in order to enhance predictions.

3. CSS for Heart Disease Prediction

Using ML to predict heart disease entails examining medical data to find trends and risk factors related to CVDs. The Cleveland Heart Disease Dataset, which includes patient data, including age, sex, blood pressure, cholesterol, heart rate, and other medical indications, is a frequently used dataset for this purpose. The dataset includes 303 patient records with 14 attributes and is a component of the University of California, Irvine (UCI) ML Repository. This dataset is frequently used to train and assess ML algorithms for CVD prediction.

The features (attributes) of the dataset are as follows:

- **Age:** Patient's age in years.
- **Sex:** Gender, with 1 indicating male and 0 representing female.
- **Chest pain type (*cp*):** Four categories, with 0, 1, 2, and 3 indicating typical angina, atypical angina, non-anginal pain, and asymptomatic, respectively.
- **Resting blood pressure (*restbps*):** Blood pressure at rest (mm Hg).
- **Serum cholesterol (*chol*):** Cholesterol level in mg/dL.
- **Fasting blood sugar (*fbs*):** Two categories, with 1 indicating >120 mg/dL and 0 representing ≤120 mg/dL.
- **Resting electrocardiographic results (*restecg*):** 0, 1, and 2 indicate normal, ST-T wave abnormality, and left ventricular hypertrophy, respectively.
- **Maximum heart rate achieved (*thalach*):** Measured during exercise.
- **Exercise-induced angina (*exang*):** 1 indicates yes, and 0 indicates no.
- **ST depression (*oldpeak*):** Depression induced by exercise relative to rest.
- **Slope of ST segment (*slope*):** 0, 1, and 2 indicate upsloping, flat, and downsloping, respectively.
- **Number of major vessels colored by fluoroscopy (*ca*):** 0–3.
- **Thalassemia (*thal*):** 1, 2, and 3 indicate normal, fixed defect, and reversible defect, respectively.
- **Target (presence of heart disease):** 0 indicates no disease, and 1, 2, 3, and 4 indicate different severity levels.

This dataset is important because it offers an actual medical dataset for predictive model training. The target variable aids in multi-class severity prediction and binary classification (presence or absence of heart disease). In addition, the dataset is useful for ML applications because it contains a variety of cardiovascular risk variables. Heart disease can be predicted using a variety of ML methods, such as neural networks, DT, random forests, logistic regression, SVM, and KNN. In order to determine if a person has heart disease or not, these models can be used to analyze patient data.

The key steps in prediction are as follows:

Step 1: Data preprocessing, which involves encoding categorical variables, handling missing data, and normalizing features.

Step 2: Feature selection, which involves finding the most important variables influencing heart disease, such as blood pressure and cholesterol levels.

Step 3: Model training, which involves training algorithms using previous patient data through supervised learning techniques.

Step 4: Evaluation, which involves using metrics of accuracy, precision, recall, and F1-score to evaluate model performance.

By lowering the chance of serious complications, this predictive technique helps physicians with early diagnosis and individualized treatment planning. By offering data-driven insights, ML improves decision-making and outperforms conventional techniques in the prediction of cardiac disease.

SVM and KNN work together to maximize their respective advantages and minimize their disadvantages. SVM struggles with overlapping classes but excels at identifying the best decision boundaries, especially in high-dimensional data. Although KNN, a non-parametric technique, is good at capturing local data patterns, it is computationally costly for large datasets and susceptible to noise. The accuracy, resilience, and adaptability of the hybrid model are enhanced by combining SVM for global classification with KNN for local refinement. This combination is especially helpful in image identification, anomaly detection, and medical diagnosis, where accurate classification is essential for making decisions. By using KNN for local modifications and SVM for global decision boundaries, the SVM+KNN hybrid model improves classification. KNN improves predictions by taking into account neighbouring instances, whereas SVM converts data into a higher-dimensional space for improved separation. Misclassification is decreased by this integration, particularly in datasets with intricate patterns or overlapping classes. Additionally, by reducing the number of neighbours to evaluate, SVM lowers dimensionality, increasing the efficiency of KNN. In fields like fraud detection, picture recognition, and heart disease prediction, where both global structure and local variability are essential for enhancing predictive accuracy and model reliability, this hybrid approach is frequently employed.

The block diagram of the performance evaluation of CSS for heart disease prediction using ML is presented in Figure 1. Predicting cardiac disease is more accurate and efficient when the maximum number of data samples is used. Effective results can be obtained by collecting data samples, storing essential features, and analyzing them. These information tests were separated by sex, such as male and female, and age groups (40–60 years, 60 years, 20–40 years, and over 10–20 years). The machine was initially trained to store a large amount of data about heart disease for future recognizable proof.

Both the proper training and testing of the ML classifier and the efficient representation of the data depend on data preprocessing. Samples of collected data were then preprocessed utilizing efficient methods such as image thresholding to turn an image grayscale and contrast enhancement filtering to remove the background. De-noising and increasing the image, slanting the element inside it, reducing or thickening its phase, and other tasks can be

finished at the preparation stage. Following preprocessing, any of the bundling algorithms can be used to identify the thing in the image through establishment deduction.

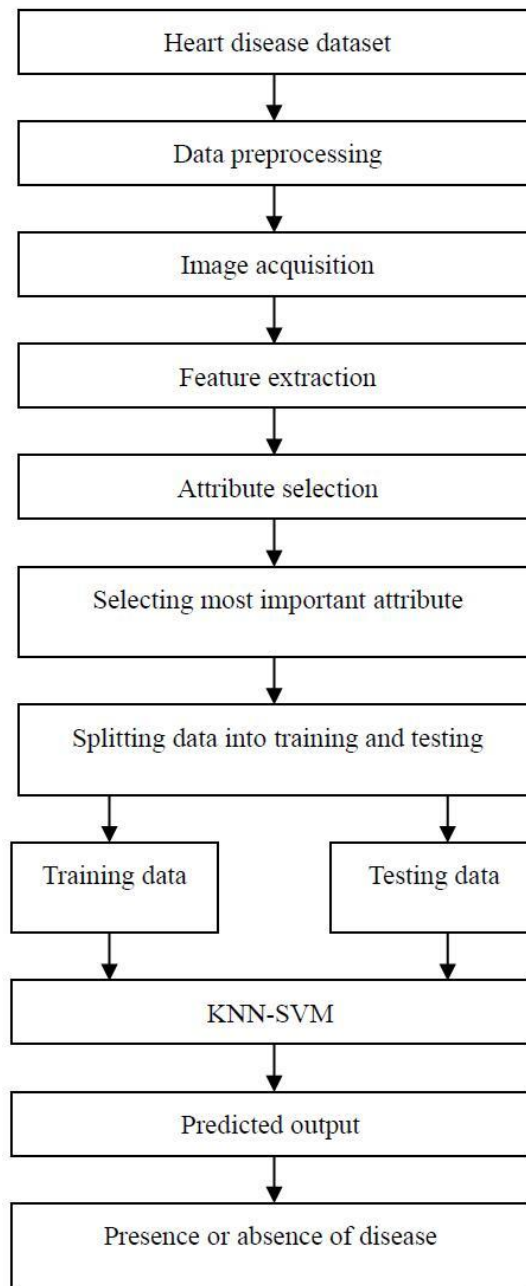


Figure 1. Block diagram of CSS for heart disease prediction

The information was then stored for future identification of a specific cardiac issue once wave structure was extracted from the image using pixel indexing. It provides a clue for choosing the most efficient classifier and feature extraction techniques for effectively predicting heart disease from the specific database images gathered.

Eliminating redundant and irrelevant properties is the main goal of attribute selection. An irrelevant feature reduces the algorithm's accuracy. To choose the most important features in this study, the Boruta attribute selection technique was utilized.

Two phases were used to detect abnormality. The testing phase evaluates the disease's severity, and the training phase gathers and stores data for future heart disease prediction. A new image that has been processed and extracted with the necessary information must be compared to the data from the trained phase. Testing data for heart disease prediction makes up 30% of the dataset, whereas the remaining 70% is thought to represent training data for the model fit and ML methods. Using CSS and ML, this study attempts to predict cardiac disease by using the combined model which includes SVM and KNN classifiers.

In classification, the SVM is one of the most often used supervised learning models. It is described as the finite-dimensional vector spaces in which a feature of a specific sample is characterized by each dimension. The goal of SVM is to identify the optimal hyperplane that separates the two classes with the highest margin. SVM's computational competence on large data sets has demonstrated its efficiency in solving high-dimensional space problems.

The training samples were directly used to categorize the test data using KNN, another supervised learning model. It describes a technique for categorizing objects within the feature space according to the closest training data. Otherwise, several distance metrics which may include a new sample's class are predicted using a simple Euclidean distance. In the practical procedures, after computing k , which is the number of the nearest neighbors, KNN was used to determine and sort the distance between the training data. Following that, the test data were given a class label determined by a majority vote.

A SVM+KNN hybrid classification method combines the benefits of both techniques to increase accuracy and resilience. SVM is a potent supervised learning technique that is excellent at identifying the best decision boundaries and managing high-dimensional data. It finds a hyperplane that best divides classes after converting input data into a higher-dimensional space using kernel functions. When there is a distinct margin of separation between the classes or when dealing with non-linear distributions, SVM performs poorly. In contrast, KNN is an instance-based learning technique that uses the majority class of their closest neighbours to categorize fresh samples. It adapts effectively to complex decision boundaries and does not assume anything about the distribution of data. However, KNN's sensitivity to noise and processing cost for large datasets can hinder its classification effectiveness. The SVM+KNN hybrid strikes a balance between local and global learning by merging these two techniques. In order to effectively reduce dimensionality and filter outliers, SVM was first utilized to determine the best decision boundary. In order to improve predictions based on local neighbourhood information, KNN was then implemented in the converted space.

This two-step procedure improves classification accuracy, especially when KNN by itself would be too sensitive to noise or when SVM has trouble with overlapping classes. Applications where high accuracy and versatility are crucial, such as text categorization, image recognition, and medical diagnosis, frequently employ this hybrid technique. A more dependable and efficient classification framework for complex datasets is offered by the SVM+KNN model, which combines the global learning power of SVM with the local adaptability of KNN. AUC, F1-score, accuracy, recall, precision, and some more measurements can be computed to evaluate the predictive model's performance.

4. Result Analysis

Tests of input data were separated into different age groups and sexes, such as male and female. Seventy percent of the dataset was designed to provide training input for ML algorithms and model fit; the remaining 30% was used to assess information for heart disease prediction. Measurements like F1-score, accuracy, recall, AUC, and precision were used to evaluate the model's performance for the heart disease dataset. There are four possible outcomes for each prediction.

- **True negatives (TN):** The case is incorrectly predicted as negative and is actually negative.
- **True positives (TP):** The case is incorrectly predicted as positive, but it is actually positive.
- **False negatives (FN):** The data is actually positive despite being predicted as negative.
- **False positives (FP):** Despite being expected to be negative, the data is actually positive.

For classification problems, AUC is a performance indicator. The true positive rate (TRP), also known as sensitivity or recall, is analyzed against the false positive rate (FPR) to obtain the AUC. The more accurately the model predicts cardiac disease, the higher the AUC.

Making a division between the total number of samples and accurate predictions, classification accuracy is a commonly used metric to assess classification algorithms.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The ratio of TP to all positive predictions is called precision, or the positive predictive value (PPV).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

The ability of a model to recognize positive examples in a dataset is measured by a metric called recall in ML. The percentage of accurately categorized true positive instances is known as recall. It is sometimes referred to as the chance of detection, sensitivity, or TPR.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score: The combination of precision and recall is known as the F1-score. It is computed by calculating the precision and recall harmonic means.

$$F1-Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Table 1 represents the comparative analysis of the performance evaluation of CSS for heart disease prediction using SVM+KNN, DT and naïve bayes (NB), taking into consideration all relevant features.

Table 1. Comparative analysis

Parameters	Prediction of Heart Disease		
	SVM+KNN	NB	DT
AUC	97.1	84.5	86.7
Accuracy	97.6	86.1	85.6
Precision	97.1	85.2	86.2
Recall	96.8	85.4	85.6
F1-score	98.2	84.9	86.4

Figure 2 represents the AUC comparative analysis of CSS for heart disease prediction using SVM+KNN, NB and DT. It is observed that AUC of the proposed hybrid model is high compared to other models.

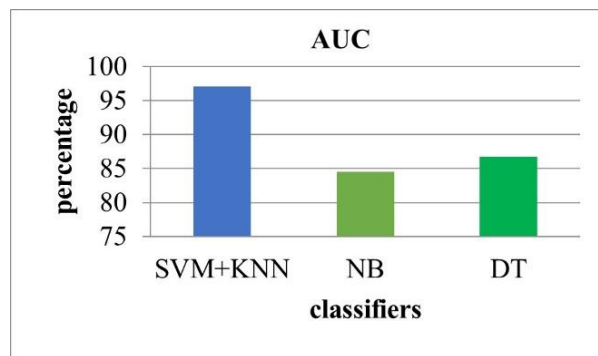


Figure 2. Comparative analysis of the AUC parameter

Figure 3 presents a comparison of the accuracy and precision metrics of CSS for heart disease prediction using SVM+KNN, NB, and DT. From the figure, it is observed that accuracy and precision values of the proposed model (SVM+KNN) are high and effective to other models in terms of heart disease prediction.

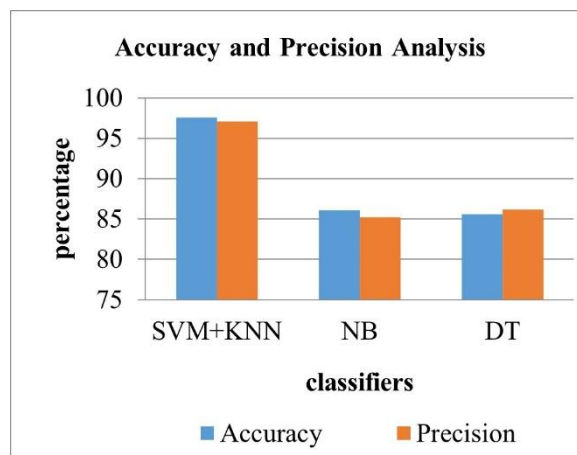


Figure 3. Comparative analysis of in terms of accuracy and precision parameters

The recall and F1-score metrics of CSS for heart disease prediction using SVM+KNN, NB, and DT were compared in Figure 4 below. It can be seen that the proposed hybrid model is more efficient and achieves high percentage in predicting heart disease compared to other models.

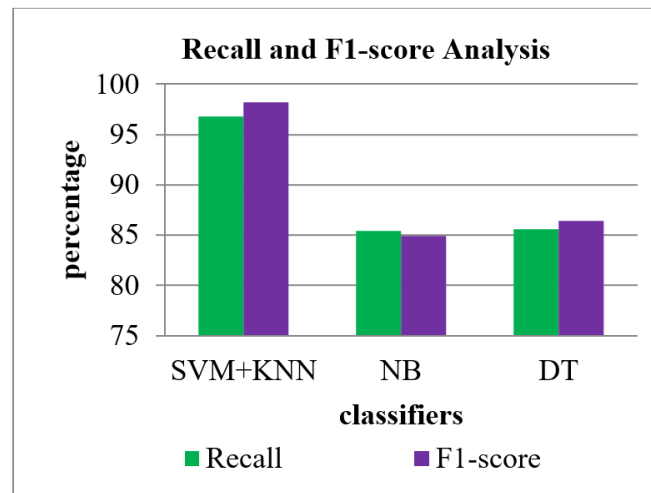


Figure 4. Comparative analysis of in terms of recall and F1-score parameters

Overall, the results suggest that the performance evaluation of CSS for heart disease prediction using SVM+KNN is effective. The described model's accuracy, recall, precision, AUC, and F1-score parameters achieved results of 97.2%, 97.6%, 96.8%, 97.1%, and 98.2%, respectively.

There are a number of restrictions on how well CSS for ML-based cardiac disease prediction performs. Model accuracy may be lowered by problems with data quality, such as missing values and inconsistencies. Biased forecasts that favor non-disease cases are frequently the result of class imbalance. Overfitting limits a model's ability to be applied to new patient data. Clinical adoption of sophisticated models (like deep learning) is difficult due to their lack of interpretability. Furthermore, the majority of systems lack tailored risk assessment and real-time monitoring and instead depend on static statistics. Future studies should concentrate on real-world testing, explainable AI, robust validation, and interaction with electronic health records (EHRs) in order to enhance performance.

5. Conclusion

This study evaluated the performance of CSS for heart disease prediction using ML. Given that one of the greatest risks to life is heart disease, in clinical data analysis, it is now very challenging to predict heart disease early on in order to reduce the number of deaths. More ML techniques are being applied to develop pattern-recognition and classification screening systems. Using CSS and ML, this study attempted to predict cardiac disease. The combined model which includes SVM and KNN classifiers was used for prediction. Tests of input data were separated into different age groups and sexes, such as male and female. The remaining 30% of the dataset was regarded as testing input, and the remaining 70% was assumed to be training input. The model achieved 97.2% accuracy, 97.6% recall, 96.8% AUC, 97.1% precision, and 98.2% F1-score, respectively.

The proposed hybrid model offers a lot of promise for practical use, especially in the prediction of heart disease. It can help physicians by offering real-time risk evaluations based on patient data when incorporated with EHR systems. Local adaptation (KNN) and global categorization (SVM) were used in the model to improve diagnostic accuracy and lower FP and FN. For high-risk patients, this integration can facilitate remote monitoring, individualized treatment programs, and early identification. In summary, the model is a useful tool for contemporary medical decision-making since it can enhance patient outcomes, streamline healthcare processes, and lower hospital readmission rates.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- Abdellatif, A., Abdellatef, H., Kanesan, J., Chow, C. O., Chuah, J. H., & Gheni, H. M. (2022). Improving the heart disease detection and patients' survival using supervised infinite feature selection and improved weighted random forest. *IEEE Access.*, *10*, 67363-67372. <https://doi.org/10.1109/ACCESS.2022.3185129>.
- Ahmad, G. N., Fatima, H., Ullah, S., & Saidi, A. S. (2022). Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *IEEE Access.*, *10*, 80151-80173. <https://doi.org/10.1109/ACCESS.2022.3165792>.
- Ali, S. N., Shuvo, S. B., Al-Manzo, M. I. S., Hasan, A., & Hasan, T. (2023). An end-to-end deep learning framework for real-time denoising of heart sounds for cardiac disease detection in unseen noise. *IEEE Access.*, *11*, 87887-87901. <https://doi.org/10.1109/ACCESS.2023.3292551>.
- Bertsimas, D., Mingardi, L., & Stellato, B. (2021). Machine learning for real-time heart disease prediction. *IEEE J. Biomed. Health Inform.*, *25*(9), 3627-3637. <https://doi.org/10.1109/JBHI.2021.3066347>.
- Boukhatem, C., Youssef, H. Y., & Nassif, A. B. (2022). Heart disease prediction using machine learning. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)*. (pp. 1-6). <https://doi.org/10.1109/ASET53988.2022.9734880>.
- Cenitta, D., Arjunan, R. V., & Prema, K. V. (2022). Ischemic heart disease prediction using optimized squirrel search feature selection algorithm. *IEEE Access.*, *10*, 122995-123006. <https://doi.org/10.1109/ACCESS.2022.3223429>.
- El-Sofany, H. F. (2024). Predicting heart diseases using machine learning and different data classification techniques. *IEEE Access.*, *12*, 106146-106160. <https://doi.org/10.1109/ACCESS.2024.3437181>.
- Jain, S. K. & Bhaumik, B. (2017). An energy efficient ECG signal processor detecting cardiovascular diseases on smartphone. *IEEE Trans. Biomed. Circuits Syst.*, *11*(2), 314-323. <https://doi.org/10.1109/tbcas.2016.2592382>.
- Karhade, J., Dash, S., Ghosh, S. K., Dash, D. K., & Tripathy, R. K. (2022). Time-frequency-domain deep learning framework for the automated detection of heart valve disorders using PCG signals. *IEEE Trans. Instrum. Meas.*, *71*, 1-11. <https://doi.org/10.1109/TIM.2022.3163156>.
- Khan, H., Javid, N., Bashir, T., Akbar, M., Alrajeh, N., & Aslam, S. (2024). Heart disease prediction using novel ensemble and blending based cardiovascular disease detection networks: EnsCVDD-Net and BICVDD-Net. *IEEE Access.*, *12*, 109230-109254. <https://doi.org/10.1109/ACCESS.2024.3421241>.
- Kumar, A., Singh, K. U., & Kumar, M. (2023). A clinical data analysis based diagnostic systems for heart disease prediction using ensemble method. *Big Data Min. Analyt.*, *6*(4), 513-525. <https://doi.org/10.26599/BDMA.2022.9020052>.
- Li, H., Ren, G., Yu, X., Wang, D., & Wu, S. (2020). Discrimination of the diastolic murmurs in coronary heart disease and in valvular disease. *IEEE Access.*, *8*, 160407-160413. <https://doi.org/10.1109/ACCESS.2020.3021093>.
- Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access.*, *7*, 81542-81554. <https://doi.org/10.1109/ACCESS.2019.2923707>.
- Rahim, A., Rasheed, Y., Azam, F., Anwar, M. W., Rahim, M. A., & Muzaffar, A. W. (2021). An integrated machine learning framework for effective prediction of cardiovascular diseases. *IEEE Access.*, *9*, 106575-106588. <https://doi.org/10.1109/ACCESS.2021.3098688>.
- Rezaeieh, S. A., Bialkowski, K. S., & Abbosh, A. M. (2014). Microwave system for the early stage detection of congestive heart failure. *IEEE Access.*, *2*, 921-929. <https://doi.org/10.1109/ACCESS.2014.2352614>.
- Salau, A. O., Assegie, T. A., Chhabra, G., Kaushik, K., & Braide, S. L. (2024). Heart disease detection model using support vector machine with feature selection. In *2024 2nd International Conference on Advancement*

- in Computation & Computer Technologies (InCACCT)*. (pp. 204-207).
<https://doi.org/10.1109/InCACCT61598.2024.10550988>.
- Sarkar, S. & Koehler, J. (2012). A dynamic risk score to identify increased risk for heart failure decompensation. *IEEE Trans. Biomed. Eng.*, 60(1), 147-150. <https://doi.org/10.1109/TBME.2012.2209646>.
- Schmidt, S. E., Holst-Hansen, C., Hansen, J., Toft, E., & Struijk, J. J. (2015). Acoustic features for the identification of coronary artery disease. *IEEE Trans. Biomed. Eng.*, 62(11), 2611-2619. <https://doi.org/10.1109/TBME.2015.2432129>.
- Shokouhmand, A., Wen, H. R., Khan, S., Puma, J. A., Patel, A., Green, P., Ayazi, F., & Ebadi, N. (2023). Diagnosis of coexisting valvular heart diseases using image-to-sequence translation of contact microphone recordings. *IEEE Trans. Biomed. Eng.*, 70(9), 2540-2551. <https://doi.org/10.1109/TBME.2023.3253381>.
- Shuvo, S. B., Ali, S. N., Swapnil, S. I., Al-Rakhami, M. S., & Gumaei, A. (2021). CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings. *IEEE Access.*, 9, 36955-36967. <https://doi.org/10.1109/ACCESS.2021.3063129>.
- Su, Y. S., Ding, T. J., & Chen, M. Y. (2021). Deep learning methods in internet of medical things for valvular heart disease screening system. *IEEE Internet Things J.*, 8(23), 16921-16932. <https://doi.org/10.1109/JIOT.2021.3053420>.
- Tao, R., Zhang, S. L., Huang, X., Tao, M. F., Ma, J., Ma, S. X., Zhang, C. X., Zhang, T. X., Tang, F. K., Lu, J. P., Shen, C. X., & Xie, X. M. (2019). Magnetocardiography-based ischemic heart disease detection and localization using machine learning methods. *IEEE Trans. Biomed. Eng.*, 66(6), 1658-1667. <https://doi.org/10.1109/tbme.2018.2877649>.
- Wang, J., Huang, X., Tang, S. -Y., Shi, G. M., Ma, X., & Guo, J. (2019). Blood triglyceride monitoring with smartphone as electrochemical analyzer for cardiovascular disease prevention. *IEEE J. Biomed. Health Inform.*, 23(1), 66-71. <https://doi.org/10.1109/jbhi.2018.2845860>.
- Wang, J., Liu, C., Li, L., Li, W., Yao, L., Li, H., & Zhang, H. (2020). A stacking-based model for non-invasive detection of coronary heart disease. *IEEE Access.*, 8, 37124-37133. <https://doi.org/10.1109/ACCESS.2020.2975377>.
- Yuan, X., Chen, J., Zhang, K., Wu, Y., & Yang, T. (2021). A stable AI-based binary and multiple class heart disease prediction model for IoMT. *IEEE Trans. Ind. Informat.*, 18(3), 2032-2040. <https://doi.org/10.1109/TII.2021.3098306>.