ISSN: 2182-2069 / E-ISSN: 2182-2077

Modelling A Novel Approach for Heart Disease Prediction Due to Covid Using Modern Data Mining Approaches

S.R. Senthilkumar[1*], and Dr.P. Amudhavalli[2]

[1*]Research Scholer, B.S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India. mastersenthil87@gmail.com, https://orcid.org/0009-0004-9825-3635

[2]Assistant Professor (Senior Grade), B.S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India. amudhavalli@crescent.education, https://orcid.org/0000-0001-6979-4847

Received: October 01, 2025; Revised: November 10, 2025; Accepted: December 31, 2025; Published: February 27, 2026

Abstract

The detection and treatment of heart disorders caused by the impact of COVID-19 are improved by intensive work. Mining and recording the data in the medical field provides potential development in maintaining patients' details. This is possible with the convergence of improved technology and medical diagnostics models. It is essential to analyze the interconnection of threat factors in the patient's clinical history to achieve the respective heart disease diagnosis. A meticulous analysis of multiple mechanisms in patient data is used to predict heart disease before COVID-19 infection. The main required attributes for detecting cardiac disorders due to COVID-19 are acquired by applying the feature selection model. The critical patient history details such as age, smoking habits, physical activity, stress levels, gender, previous chest pain occurrences, diabetes, electrocardiogram (ECG) readings, dietary patterns, chest pain type, and troponin levels are considered for predicting heart disease. Different AI technologies, such as deep neural networks with SVM ($d - SVM$), were used, and the results were compared between two datasets from the heart disease database. These methodologies are employed to select features from the database and are also employed on all features of the data repository. The enhanced accuracy rate of 95% is acquired through our proposed model, which uses selected features as input. Early heart disease prediction is achieved through our proposed technique's assisting structure. Successful deployment of an AI model as computationally intensive as this, from the laboratory setting to real-time actual clinical practice, however, relies on an architecture for scalable and fault-tolerant deployment. To this end, we introduce a Cloud Computing and Service Deployment paradigm that is specifically tailored for executing the d-SVM model. It leverages the cloud's elasticity as well as high availability to deal with fluctuating diagnosis loads in healthcare networks. By converting the d-SVM into an Internet-based Service through an Application Programming Interface (API), this work maximizes the potential of the model for large-scale, sustainable, and cost-saving clinical utility.

Keywords: COVID-19, DNN, Data Mining, SVM, Heart Disease, Cloud Computing, Service Deployment.

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 1 (February-2026), pp. 494-517.
DOI: *10.58346/JISIS.2026.I1.028*

*Corresponding author: Research Scholer, B.S. Abdur Rahman Crescent Institute of Science and Technology, Vandalur, Chennai, Tamil Nadu, India.

1 Introduction

Heart disorder is a significant problem due to conditions that disturb the blood capillaries and other internal parts of the body, resulting in irregular cardiac function. The amount of oxygen in the lungs is reduced by the virus SARS-CoV-2, which affects the respiratory system of humans (Mijwil et al., 2021; Abdullah & Ahmed, 2021). This virus can cause some negative functioning of the heart and may also lead to heart failure. The growth of abnormal metabolism in lipids, build-up of lipids, athermanous plaques, and some other substances inside the path of the circulation track are all symptoms of heart disorder (Madjid et al., 2020). A stricture or bowel obstruction can cause oxygen deprivation, tightness in the chest, chest discomfort due to necrosis, myocardial infarction, and myocardial ischemia. These symptoms can cause heart disorders that are considered a significant factor for death in the present world. Approximately 12 million deaths occur due to heart disease across the world each year (Le Glaz et al., 2021) based on the information provided by the World Health Organisation (WHO). Nearly 8.91 million patients lost their lives as a result of 10.6 million infected patients of cardiac heart infection in 2017. About 126.5 million patients were infected by coronary heart disease (Dai et al., 2022) in 2017. The clinical expenses are predicted to rise from \$126.2 billion to \$177.5 billion within 2040, estimated as 41% higher.

The healthcare industry developments mainly rely on machine learning (ML) applications (Olsen et al., 2020). The novel professionals' tasks (Taleb et al., 2021) using artificial intelligence machine learning applications can provide improved precision and efficiency. The issues like congested healthcare systems and clinician shortages are handled by utilizing machine learning applications across many countries. ML plays a vital role in medicinal field as it is capable to predict the patterns in large-scale datasets and assist in identifying disease-related diagnostic features and health risk factors. Medical image evaluation (Allugunti, 2022), tumour or cancer cell diagnosis, and language processing are some of the medical applications that make use of ML models. The machine learning method facilitates clinical management and allows technicians to explore the perfect technique for treatment. ML categorization methods are integrated into the clinical treatment process, which provides information about heart disorder detection and the required treatment to enhance the patient's health condition. Machine learning methods like K-nearest neighbour (K-NN), Naive Bayes (NB), AdaBoost (AB), Support Vector Machine (SVM), Artificial Neural Network (ANN), Fuzzy Logic (FL), Logistic Regression (LR) (Manhas et al., 2022) are extensively used in prior detection of heart disorder. These methods facilitate effective clinical decision techniques and minimize the heart disorder death rate.

The results of different techniques explain that other factors can impact the research conclusion. The influencing factors of the proposed technique are data gathering, the technique used to filter the obtained data, the attributes used, generalization, and randomization of the attributes. The interconnection between the input parameters of the collected datasets is analyzed and generalized by the researchers to find how these factors affect the efficiency of cardiac disorder prediction. This research used many machines learning (ML) techniques like multilayer perceptron (MLP), logistic regression, K-nearest neighbour (K-NN), random forest, Naïve Bayes, Support Vector Machine (SVM), and decision tree methods for predicting heart disease. The extensive description of information in different areas of clinical research (Allugunti, 2022) is provided by logistic regression methods and provides an exact interpretation of the dataset. The logistic regression model is used to identify the attributes and suggest whether an enhancement or no improvement will happen after the intervention, exploring the effects and links between the predictors, presence or absence of infection using various factors identifying appropriate predictors to use and finding whether newly analyzed variables can be added to the predicted value.

Naive Bayes can be employed in health to analyze predictive techniques for various diseases like asthma, breast cancer, brain, and prostate. The supervised machine learning method K-NN is mainly in employment for classifying and predicting heart disease. Liver disease data samples are classified using K-NN as defined. The Random Forest technique identifies the most reliable predictors, facilitating prediction performance in a clinical environment (Jindal et al., 2021). The infection diagnosis of invasive breast carcinoma, basal cell carcinoma and various other diseases are identified using support vector machine model. SVM provides more standardization capability even in small-scale data sample classification. Other fields include handwritten character identification, image categorization, text classification, and recognition.

Early recognition of cancer is possible with the decision tree model which facilitates predicting stroke results, recognizing cardiac arrhythmias and chronic disorder management assistance. MLP is also utilized in different areas of the medical environment, like disease forecasting, clinical image identification and gene selection (Yang & Garibaldi, 2015). This research evaluates all the classification models using full and selected features to check their accuracy and reliability. This research assists in identifying the most feasible classification technique for structuring high-level intelligent models for predicting heart disease. So far, works have not utilized full features and selected feature datasets for heart disease prediction (Nazir et al., 2018). The research picks only one appropriate classification technique based on its performance and accuracy. The suitable $d-SVM$ classification models are chosen based on factors like F1-score, accuracy, precision, receiver operating characteristics, area under the curve score metrics, and recall. The proposed model will help physicians predict whether or not a person is infected with heart disease due to COVID-19. In this research, the selected dataset features performed better than the full features dataset (Fairfax & Sørensen, 2024).

While AI model creation with high accuracy is the top priority, model creation is not enough to solve a real-world clinical problem. For it to be useful in-patient care, the d-SVM model needs to be available in real-time to clinicians via an accepted, high-throughput, and scalable platform. This involves bridging the gap between clinical practice and algorithmic research.

On-premise traditional computing is not able to handle the quantity and variability of diagnostic requests in a sizeable healthcare network. Hence, the secret to realizing maximum utility with our AI is to use a Cloud Computing model (Aravind et al., 2023). The cloud allows for the dynamic resource provisioning and fault tolerance needed to support the d-SVM model as a mission-critical Internet Service, and thus this deployment model is a central theme to this paper.

This research work contains various sections such as section 2 provides detail analysis of different methods, section 3 explains the methodology utilized for this research work, section 4 provides the numerical results and finally section 5 concludes the outcomes of our research.

2 Related Works

The author in (Rajdhan et al., 2023) suggested machine learning (ML) models like K-NN and LR to detect and organize patients with heart disorder (Verma & Kapoor, 2021). This investigation proved that the K-NN technique works better with 88% accurate result. Sahoo et al. recommended various multiple classification models like SVM, logistic regression, K-NN, and Naive Bayes to evaluate the coronary heart disorder dataset collected from the UCI database, which poses 13 essential characteristics (Jindal et al., 2021). The best accuracy is provided by the SVM classification technique, which is 85% accuracy, was concluded by this research. To identify heart disorders, Uyar et al. proposed a Genetic Algorithm (GA) based on a recurrent fuzzy neural network (RFNN) trained dataset, and this study utilized the dataset from the California University Irvin machine learning database (Sahoo & Jeripothula, 2020). Of 297 data samples, 45 were chosen for testing, and 257 were selected for training. The result obtained shows a 97.79% accuracy rate in the testing dataset.

The author in (Arabasadi et al., 2017) defined a specific hybrid technique for predicting coronary heart infection. The utilized method can improve the performance of the neural network by roughly 10%, and it upgrades the actual weight using the genetic algorithm technique that recommends enhanced neural network performance. The above technique was employed to the Z-Alizadeh Sni data samples, the results generated provide sensitivity, accuracy, and specificity rates of 97.5%, 93.85%, and 92.5%, respectively. MLP neural-network-based detection of heart disorders has been developed by (Chowdary et al., 2024; Mahdizadeh & Zamanzade, 2019). This model's neural network input is collected from 13 medical features and employs the steepest-descent method to predict the subject's cardiac disorder. This model results in an enhanced accuracy of 98%. Different data mining models are expressed to evaluate heart disease prediction (Reza et al., 2021). The generated output shows that the neural networks utilizing 15 attributes outperform the data mining techniques. The decision tree method with a genetic algorithm and selected attribute technique can achieve high accuracy is the conclusion provided in another research.

A knowledge extraction model is used to invent a detection technique based on the physicians' instruction was proposed by (Seo et al., 2022). A skating algorithm was used to develop the exactness and efficacy of the system. Skating is an ensemble model comparable to the boosting and bagging method. They ensemble four varieties of classification techniques like K-NN, skating, Naive Bayes, and decision tree, they experimented with the fact that skating provides more efficient results than the other classifiers (Mishra et al., 2021; Sethy et al., 2022; Rustam & Angie, 2021). The author utilized the ensemble of five classification models: neural network, SVM, classification through regression, probabilistic classifier, and gradient boosting. By applying these classifiers, the classification accuracy of 73.18% is achieved. These classification algorithms are used to construct the ensemble data extracting approach applying two standard datasets, namely Hungarian and Cleveland, which are extracted from the UCI machine learning database (Yekkala & Dixit, 2018). The regression method had the lowest accuracy, but the random forest improved accuracy by 98.136% (Routray et al., 2018; Le et al., 2019; Afolayan et al., 2022).

Healthcare AI literature downplays the move from an accurate model to a low-latency, world-wide service. Cloud Computing has proven to be the de facto choice for productizing large-scale clinical AI (Verma & Kapoor, 2021). The greatest advantages are scalability and resource elasticity, which are of primary concern for medical diagnostic services with unpredictable, bursty demand (e.g., for a pandemic outbreak). Serverless (FaaS) or PaaS cloud infrastructure makes it possible to execute high-performance models such as d-SVM as Microservices or APIs with no hardware maintenance (Mokhtari et al., 2025). This makes it easier to manage the Service Deployment pipeline (referred to as MLOps often) and provides HA. This discussion forms the basis to suggest an architecture where the accessibility, maintainability, and availability of our diagnostic d-SVM Internet Service are given top priority on its agenda.

The performance of the classification algorithm can be enhanced using a decision tree with the genetic algorithm, as proposed by (Santos et al., 2022). The other two models, such as Naive Bayes (Qiu et al., 2021) and cluster-based classification, provide results different from the previous model. The proposed model is identified to provide 99.2% accuracy. The author in (Mishra et al., 2020) verified the efficiency of the logistic regression and random forest for predicting cardiovascular patients' risk exposure (Liu et al., 2022). The random forest classification technique provides less performance accuracy than the logistic regression, resulting from the above method. The performance of the LR is 89% whereas the gradient boosting provides an accuracy of 88%. The author in (Chaudhuri et al., 2021) compared the standard classification trees and regression trees to find the best accuracy, resulting in the techniques are standard logistic regression results in better recognition for predicting the presence of heart disease (Lai & Deng, 2018; Adler et al., 2020). Table 1 depicts the comparison of various approaches.

Table 1: Comparison of various existing approaches

Reference	Dataset	Method	Approach	Objective
Feeny et al., 2019	Github		NB and LR	To give optimal prediction outcomes
Dalal et al., 2023	US health system		LR	To give superior prediction accuracy
Diwakar et al., 2021	Reported cases		CNN	To reduce error rate
Edeh et al., 2022	Chest X-ray		k-NN	To predict COVID-19 with higher speed based on features
Waris & Koteeswaran, 2021	152 COVID-19 samples	Supervised learning	CNN	To provide better clinical prediction value
Ghouali et al., 2022	5000 COVID cases		Regression model	To make prediction faster and measure the chances of future cases
Jan et al., 2018	7000 samples		ANN	To provide higher efficiency and precision value
Khajehali et al., 2023	320 samples		LR	To provide sensitive prediction value
Kim et al., 2022	370 variables		NB	To concentrate of feature samples to analyse the cause of the disease
Kondababu et al., 2021	290 samples		LR	To provide optimal outcome

3 Methodology

There are various reasons why heart disease is becoming more common. Prior identification of the heart disease is required for the proper treatment. Many machine learning techniques are used in this research to achieve the early detection of heart disease due to COVID-19. These proposed techniques enable each person to become aware of their risk in the prior stage.

Prediction Model

The main objective of the proposed research is to develop a suitable technique that can automatically and exactly predict heart disorders. The goal of this study is achieved through data gathering, multiple machine learning models, pre-processing, characteristics mining and selection, and evaluation of performance. The initial procedure starts with collecting the required dataset, and after that, pre-processing of the gathered data samples is done to fill in the missing attributes, remove indefinite and duplicate datasets, and finally, generalize the dataset. Attribute and extraction methods have been employed to pick the most required features and ensemble features to produce a new reduced feature dataset. The two types of datasets are obtained for the proposed technique, one with full features and other with selected features. Diverse artificial intelligence techniques are employed on full-featured and selected features to identify the perfect dataset for predicting heart disease. Comparative evaluation of various machine learning models is executed to construct a highly efficient and reliable heart disorder prediction technique. The identified prediction technique can categorize heart disorders strictly based on multiple performance factors obtained from the two types of datasets. Figure 1 represents the basic steps in employing the prediction technique that must be tracked to employ AI methods for cardiac disorder prediction by means of suitable confidence.

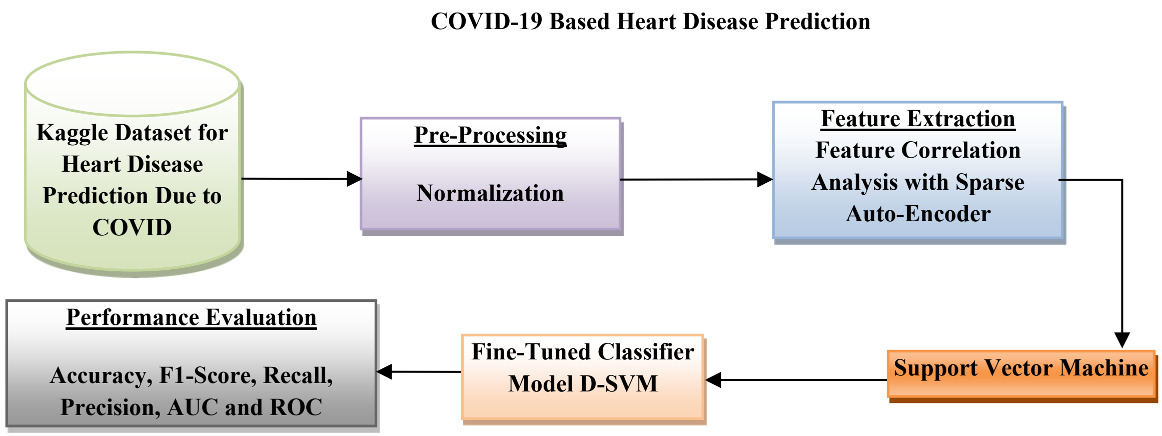

Figure 1: Block diagram of the model

Pre-Processing

The dataset for analysis is collected from diagnostic labs, hospitals and medical centres (Muzammal et al., 2020). Large numbers of datasets are gathered from the hospital's in-patients.

Table 2: Dataset description

Attributes	Type	Value
Age	Nominal	30 – 90
Gender		Female and male
Profession		Service holder, house wife, farmer, worker, businessman, driver, unemployed
Family_details		Yes and No
Smoke		
Adiposity		
Diet_condition		Abnormal and normal
Physical_exercise		Yes and No
Pressure		Abnormal and normal
Type of Cardiac pain		Atypical, typical and non-angina pain
Chest pain history		Yes and No
Oedema		Yes and No
Systolic_pressure		80 – 180
Diastolic_pressure		60 – 120
Cardiac rate		Yes and No
Blood_sugar		Negative and positive
Cardiac troponin		Abnormal and normal
Electrocardiogram		Class 1 and class 0
Result		

The questions are asked to the patients, test outcomes are observed, required feature attributes are recorded. The required dataset contains test outcomes of 59 patients and reactions to different queries. Table 1 illustrates the specifications of 19 features (one dependent target attribute, 18 independent attributes) in the dataset. The attribute ranges may vary; hence, it is necessary to standardize the dataset. The equation obtains normalization.

$$X - normalized = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Where, X-normalized represents the restructured normalized variable. '0' is the least value of every dataset attribute, and '1' indicates the maximum value for the feature attribute. All the other values are calculated using numerals between '0' and '1' (Table 2).

Feature Representation

Feature extraction reduces a large sum of data to a reduced number of related bits. Attribute selection is the procedure of extracting a subset of required variables for the machine learning structure and statistical techniques. The generalization process reduces the data to make it simple for the ML models so that the evaluation process takes less time and increases the speed. All the collected data are not equally important for the prediction process. So, our proposed model utilized the correlation-based attribute subset selection method and the most excellent first search to acquire the required attributes from all 19 heart patients' data records variables. This evaluation technique is correlated with the appropriate association measure between the datasets and the heuristic explore model in the CFS algorithm. When the correlation among the outside attribute and the components improves, the association among the outside attribute and the composite attribute also enhanced. The components have reduced inter-correlation if the association between the complex variable and the outside variable is more robust.

Deep Network Model

It is feasible to characterize a feature vector in a minimized form because, in a predictable auto-encoder, the hidden space has fewer neurons than the input and output layers. In Sparse Auto-Encoder (SAE), the input and output layers have fewer neurons than the latent space. The usage of the neurons in the layers is limited in the $L1$-regularization term, which forces the network to use fewer neurons each time. This kind of network can improve the number of features, which can be analyzed from a diverse viewpoint. The standard structure of SAE is shown in Figure 2. The decoder element in the neural network is detached once the SAE is trained for the input restructuring procedure. This represents that only the encoder part enters into the latent space. The initial N attributes are detached to create M attributes with$N < M$, which is possible through the encoder method. This technique permits the maintenance of all the datasets of the actual information and enhances the extra hidden features. The input to the CNN is two-dimensional data that includes sample images, and such networks can extract the composite features from the dataset repository. During the training, the filter or kernel masses are balanced to perform an accurate feature map of every class. In the convolutional neural network framework, the initial phase is the pooling layer, and the next layer is the convolution layer. The pooling layer reduces the issues of over-fitting and the count of network parameters to minimize the computation complexity. Max pooling is the most general and widely applied method that selects each window's higher value. Dense layers are the final layers of the convolutional networks used to classify kernel-extracted attributes. Figure 2 represents the feature-based CNN, which contains a hidden layer that is an extra layer present between the inputs and the outputs.

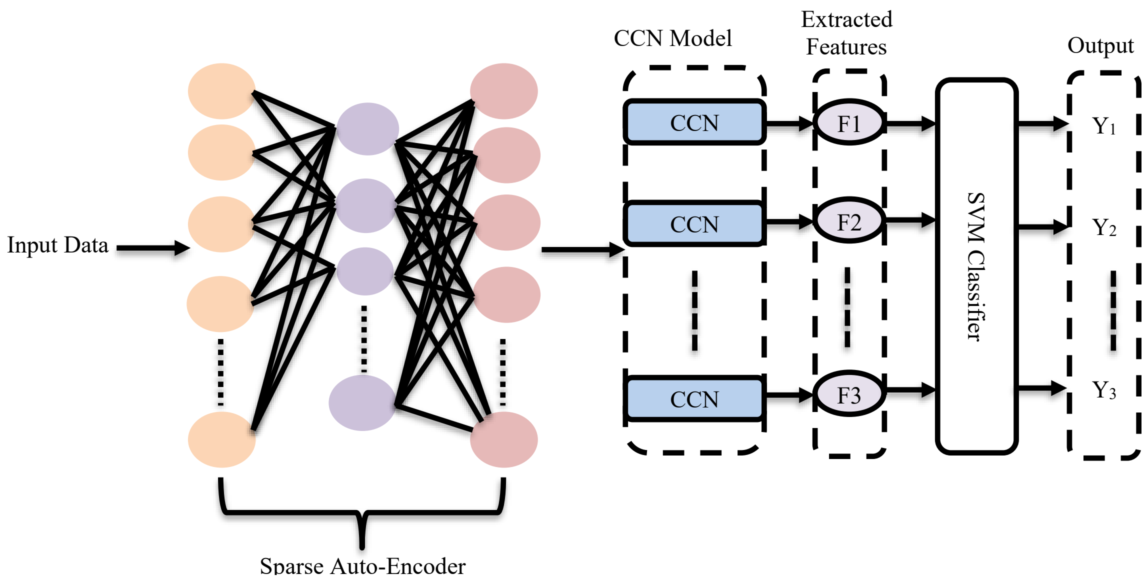

Figure 2: Proposed model

SVM Based Classification

Vapnik (Reza et al., 2021) generated the most common and well-structured machine learning technique, SVM. The boundary among the positive and negative data points is increased nearer to the N-dimensional decision hyper plane in a space. Numerous classes to distinguish are present in N-dimensional space, and the efficiency of the support vector machine (SVM) is considerably affected by distinguishable nonlinear datasets. The difficulty is solved by transmitting the data points from the input data space to the high-dimensional space utilizing single available function of the filter. This kernel aims to detect the most favourable decision plane. To predict heart disease SVM is the commonly used model.

The prognosis of cardiac disorder is considered an SVM classification issue that allocates the feature vector of diseased subject$\vec{x} = [x_1, x_2, \dots, x_n]$, to a class $y_j \in Y = \{y_1, y_2, \dots, y_{|Y|}\}$ or not, where Y represents class set. By assuming, there are N training datasets$\{(\vec{x}_1, \vec{y}_1), (\vec{x}_2, \vec{y}_2), \dots, (\vec{x}_N, \vec{y}_N)\}$,$\vec{x}_i \in R_d$, $y_i \in (\pm 1)$ where $y_i, y_i, \dots, y_N$ represents the class label feature vectors $\{\vec{x}_1, \vec{x}_2, \dots \vec{x}_N,\}$, correspondingly. In linearly separable attributes, the line $\vec{\omega} . \vec{x}_i + b = 0$ indicates the decision boundary among the two classes, positive class and negative class, where $\vec{\omega}$ indicates a vector weight, bias is defined as$b$, and input data is represented as$\vec{x}_i$. The purpose of the support vector machine is to detect the best parameters of $\vec{\omega}$ and b that build the planesH_1 and H_2, where $H_1 \rightarrow \vec{\omega}^T . \vec{x}_i + b \geq +1$ for positive class and $H_1 \rightarrow \vec{\omega}^T . \vec{x}_i + b \leq -1$ for the negative class, as illustrated in Figure 2.

In general, SVM increases the positive and negative boundary space of the data points nearest to the hyperplane. The combination of hyperplanes H_1 and H_2 can be represented as follows, $yi\ ((\vec{\omega} . \vec{x}_i + b) - 1 \geq 0 \forall i = 1, 2, \dots, N$, where $y_i \in \{\pm 1\}$. The formulated hyperplanes are considered as an optimization difficulty in the support vector machine by utilizing Eq. (2) to differentiate the negative and positive classes, which indicates the vector machine margin.

$$\min \left(\frac{1}{2} \vec{\omega}^T . \vec{\omega} \right)$$

$$y_i\ (\vec{\omega}^T . \ \vec{x}_i + b) - 1 \geq 0 \quad for\ all\ \ i = 1,2, \dots, N \qquad (2)$$

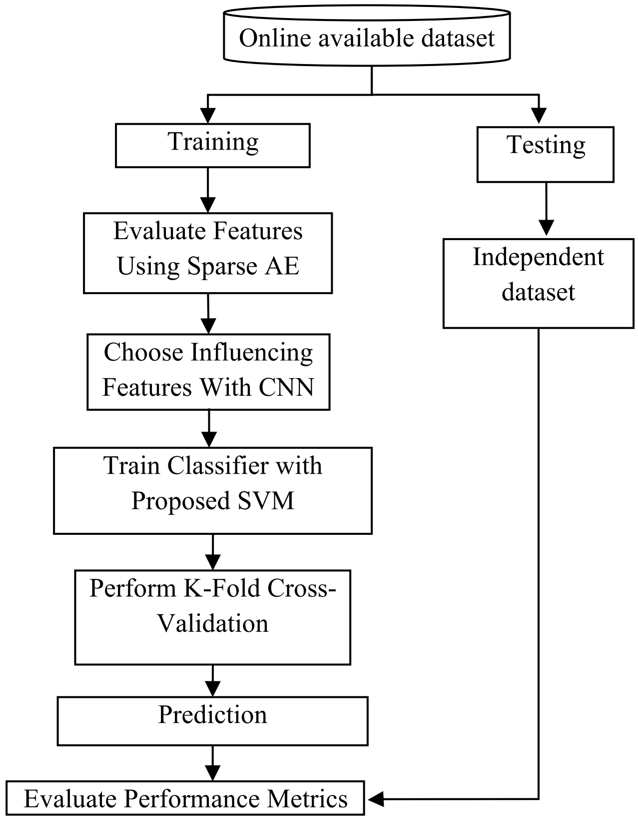

Figure 3: Flow of prediction model

Figure 3, The standard support vector machine cannot classify the new data samples into a perfect class for nonlinearly separable data. To overcome this difficulty, the support vector machine presents a kernel function (ψ), which maps the training data into a high-dimensional space that helps to prevent improper classification of the dataset. Eq. (3) indicates the formulation of the primary function of support vector machine.

$$\min\left(\frac{1}{2}\,\vec{\omega}^T.\vec{\omega}\right) + C\sum_{i=1}^{N} \epsilon_i$$

$$y_i\,(\vec{\omega}^T.\ \vec{x}_i + b) - 1 \geq 0 \quad for\ all\ \ i = 1,2,\ldots,N \tag{3}$$

Here, C indicates a penalty factor between $\in i$ and the length of the margin, and $\in i$ indicates a slack variable. Using nonlinear support vector machine classifier, the feature vector $\vec{x}_i$ is labelled as i^* if the main function f_i produces the maximum value of i^* as follows:

$$i^* = \arg\max f_i(\vec{x}_i) = argmax\ f_i\left((\vec{\omega}^T.\ \psi(\vec{x}_i) + b_i)\right) for\ all\ i = 1,2,\ldots,N \tag{4}$$

The outcomes of i^{*th} objective function may be positive or negative as represented in Eq. (5):

$$f_{i=i^*}\,(\vec{x}_i) > 0,\ \ f_{i\neq i^*}\,(\vec{x}_i) < 0\} \tag{5}$$

At the time of classification, the feature vector $\vec{x}_i$ which will not satisfy Eq. (5) is not classified and declared as an indefinite case as:

$$for\ all\ \vec{x}_i\ \neq \{\vec{x}_i | f_{i=i^*}\,(\vec{x}_i) > 0,\ \ f_{i\neq i^*}(\vec{x}_i) < 0\} \tag{6}$$

At this point, the NB classification technique is utilized to classify the ambiguous vectors. The probability that the ambiguous vector $\vec{x}_i$ belongs to a class C_j in naïve bayes classifier is defined using Eq. (7) is:

$$P\left(C_j|\vec{x}_i\right) = \frac{P\left(C_j\right)P(\vec{x}_i|C_j)}{P(\vec{x}_i)}; for\ all\ i = 1,2,\dots,N \tag{7}$$

The i^* is the label for ambiguous vector $\vec{x}_i$, if the conditional probability $P(C_j\,|\vec{x}_i)$ is the maximum for $i\,*$, as specified in Eq. (8):

$$i^* = argmax\left(P\left(C_j|\vec{x}_i\right)\right) for\ all\ \ i = 1,2,\dots,N \tag{8}$$

Algorithm 1:

Input: Dataset = $\left\{\left(x^{(1)},\dots,x^{(n)}\right)\right\}$; input sample size; total amount of samples; prediction class = {Yes/No};

Output: classification accuracy

1: Split training and testing set;

2: Analyze number of features;

3: Analyze selected features;

4: Input processing;

5: Partition (training, testing) = (70:30)

6: Training New = choose (Training set, D)

7: Testing New = choose (Testing set, D)

8: Aggregate deep model with SVM;

For $i = 1\ to\ L$;

$D_i = (Trainingset, D)$;

$h_i = NB\ (D_i)$;

$h_j = SVM\ \left(D_j\right)$;

end

9: Testing:

For all $'x'$ in testing;

End

End

Operationalizing the D-SVM Model Via Cloud Microservices

To deploy the d-SVM prediction model from a research paper to a high-availability Internet Service, a Cloud Computing deployment strategy founded on a microservices approach. The strategy is aimed at meeting both the computational needs of the model as well as the essential need for real-time scalability in healthcare settings. The deployment procedure entails three main steps. Second, the trained d-SVM model and the optimal feature set obtained from the feature selection process are serialized using tools such as Python's joblib or pickle. Subsequently, this serialized model is placed inside a Docker container in a manner that the model's environment, i.e., all the dependencies such as TensorFlow/Keras and Scikit-learn, will be the same and not change irrespective of the cloud environment (e.g., AWS, Azure, or GCP). In phase two, the container application is deployed to a Platform-as-a-Service (PaaS) environment like Kubernetes or an app-managed platform. In this case, the microservice executes as a RESTful API endpoint with an API Gateway as a single point of access for handling traffic, authentication, and routing of requests. This allows patients' important characteristics, such as age and troponin levels, to be uploaded as a simple JSON input to the API, returning the time-calculated heart disease risk score. Lastly, for phase three, elastic scaling and load balancing are implemented. The native load balancer of the cloud directs incoming diagnostic requests between multiple instances of the d-SVM container running. Auto-scaling policies are configured to monitor metrics like CPU utilization or request queue length, allowing the system to automatically scale up or down based on fluctuating demand. This ensures that, even during periods of high demand (such as a COVID-19 surge), the service maintains low latency and high availability (HA) around the clock.

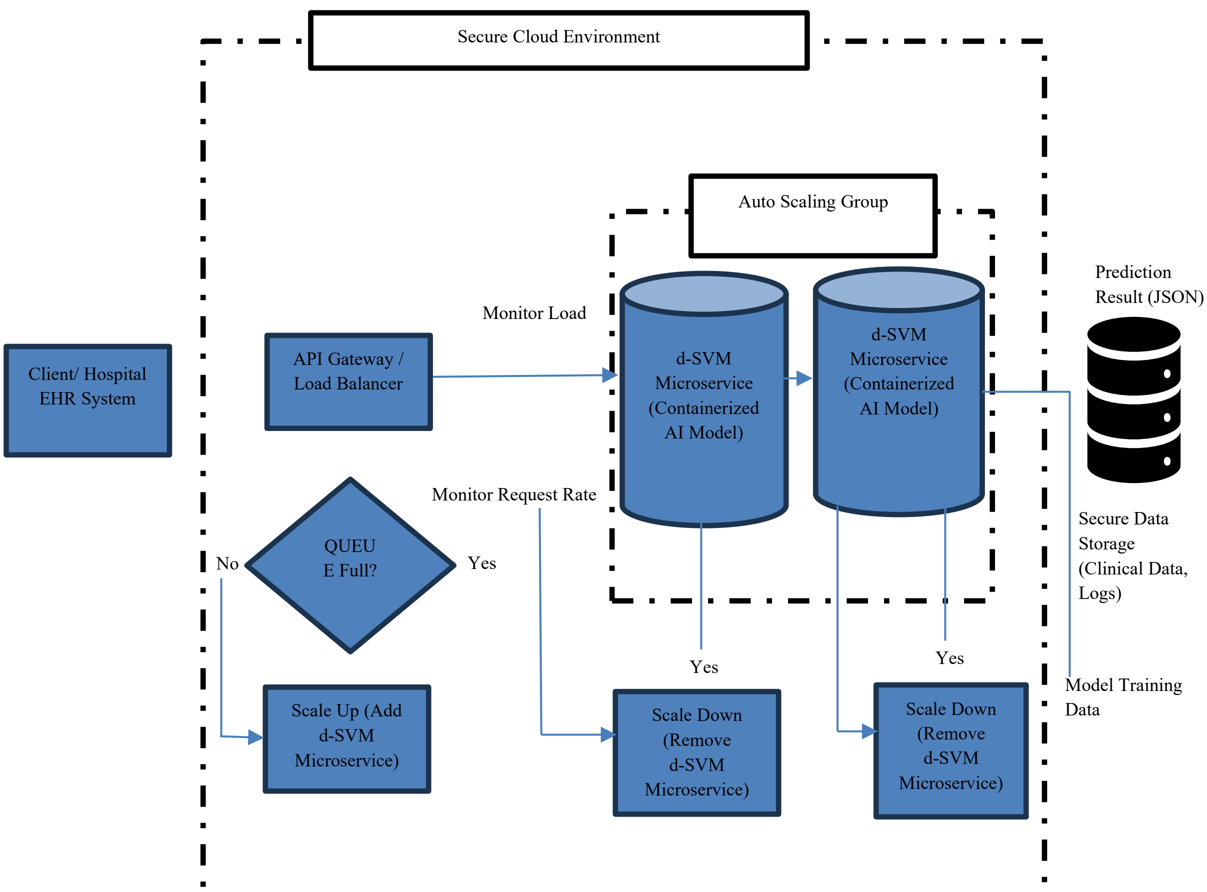

Figure 4: Cloud-based service deployment architecture with flowchart logic for D-SVM prediction model

Figure 4 involves a Cloud-Based Service Deployment Architecture for a d-SVM Prediction Model, showing a secure and scalable method for the distribution of AI-based heart disease diagnosis as an Internet Service. The process starts with a Client/Hospital EHR System transmitting a diagnosis request (JSON data) through a Secure Network (encryption of data in transit). The request initially lands on the

API Gateway/Load Balancer, which offers one secured entry point (TLS/VPN). The greatest innovation lies in the Auto-Scaling Group of the Secure Cloud Environment. There, it applies flowchart logic: when the request queue is full, the system Scales Up automatically by adding a new d-SVM Microservice container to handle the load. When the load is low, the system can Scale Down to save costs. These containerized d-SVM AI Models process the requests and produce two important outcomes: the Prediction Result (JSON) is returned to the client through the API Gateway, and all transactional data are logged into the Secure Data Storage (in addition to the initial Model Training Data), providing both real-time utility as well as strong data stewardship.

4 Numerical Results and Analysis

Table 1 represents the entire dataset partitioned into testing and training datasets. A known output is included in the training dataset, and the method produced on this dataset is used to standardize it later to other data. 70% of the collected data (39 cases from the 59-instance dataset) is used for training. The proposed model performs at an enhanced accuracy level on the training attributes but outperforms the dataset in the testing phase. The performance inaccuracy is prevented by applying the tenfold cross-validation technique. The method is demonstrated with the test data, which is also a part of the gathered dataset. This research used twenty instances from the collected data for the testing phase, which is 30% of the entire dataset. The performance of the cross-validated dataset depends on the classification model used. So, it is necessary to ensure that the classification technique is functioning at its best through the parameter tuning technique (Table 3 and Table 4).

Table 3: Data partitioning

Dataset	Partitioning (%)
Training	70
Testing	30

Table 4: Confusion matrix

Prediction		
	Negative class (0)	Positive class (1)
Actual		
Positive class (1)	FN	TP
Negative class (0)	TN	FP

Performance Analysis

A structured model is used in this research to analyze the efficiency and accuracy of various classification models for predicting heart disease. To evaluate the system's performance metrics, many factors like True Positive, False Positive, True Negative, False negative, precision, recall, ROC-AUC score, and F1-score, along with the confusion matrix, are considered.

Confusion Matrix: The classification models' efficiency and accuracy are determined using the confusion matrix, a straightforward technique. A confusion matrix is a table with two dimensions, "Predicted class" and "Actual class" in all dimensions. The actual classification of heart disease is represented in rows, and the predicted classes are defined in the columns. The dataset contains two classes Cl, Class 0 and Class 1. Table 2 indicates the confusion matrix developed for detecting heart disease.

True Positives (TP): in this case, the actual class is true, and the predicted class of the data points are also true.

True Negatives (TN): in this type of case, the actual data class and the predicted data class are False.

False Positives (FP): in this type, the actual data class is False, but the predicted data class is true.

False Negatives (FN): in this type, the actual class of the data point is true, and the predicted class data point is False.

Accuracy: It is computed using the count of entire exact predictions divided by the total number of datasets. Four classification techniques are used to find the accuracy comparison.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{9}$$

Precision: it displays the ratio of optimistic class predictions that are heart disease positive. The high precision obtains the result of consistent measurement or the repetitive values of the reading. The variation in the measurement is indicated by the low precision.

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

Recall: the test's ability to designate a patient with positive heart disease. Few false negative results are produced with susceptible tests; thus, only a few cases of heart disease are missed. It is referred to as a True Positive Rate (TPR).

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

Specificity: The number of exact pessimistic predictions of cardiac disorder divided by the sum of all negative predictions, and it is also referred to as True Negative Rate (TNR).

$$Specificity = \frac{TP}{TN+FP} \tag{12}$$

F1-score: The F1 score is evaluated as the harmonic average of recall and precision. Maximum F1 score obtains a high accuracy value of recall and precision.

$$F1\ score = 2 * \frac{Recall*Precision}{Recall+Precision} \tag{13}$$

AUC-ROC: The Area Under Curve (AUC) –Receiver Operating Characteristic (ROC) is an efficacy indicator for the issues related to the classification model. The AUC-ROC metric indicates the capacity of the classification models to work in the class division process. The higher AUC value indicates that the model performs better. The curve is generated using the values acquired by plotting the values of True Positive Rate (TPR) and False Positive Rate (FPR) at dissimilar threshold values.

Discussion

This research applies diverse machine learning models to detect heart disease. The efficiency of these ML techniques is analyzed, and a thorough comparison of the results is observed. The following section evaluates the dataset and records the observations for different methods before the performance evaluation process. The graphs also depict varying correlations between the attributes and target values. Tables 3 and 4 illustrate the numerical data correlation. The heatmap of the numerical feature correlation is presented. The highly correlated attributes are diastolic_pressure and systolic_pressure. It is necessary to depreciate either the variables bp_systolic or bp_diastolic. The attribute bp_systolic has been separated and removed from the dataset because bp_diastolic connect higher with the target, according to the association with the parameters and every other target (Table 5 and Table 6).

Table 5: Statistical analysis

	Age	bp_systolic	Bp_diastolic	Heart rate	Target
Count	60	47	46	47	60
Mean	16.76	129	77	83	0.62
Std	16.87	25	15	22	0.50
Min	27	75	50	54	0
25%	50	110	66	70	0
50%	60	130	75	76	1
75%	65	144	90	90	1
Max	90	180	120	165	1

Table 6: Correlation analysis

	Age	bp_systolic	Bp_diastolic	Heart rate	Target
Age	1.0	0.019	0.070	-0.03	0.28
Bp_systolic	0.020	1.0	0.80	0.15	0.13
Bp_diastolic	0.070	0.8	1	0.20	0.20
Heart rate	-0.04	0.15	0.2	1	0.14
Target	0.29	0.13	0.2	0.14	1

In older people's circulatory systems, efficiency declines due to the main factor of age, this increases the possibility of cardiovascular disease. A lot of other age-related issues increase due to cardiovascular diseases like stroke, atherosclerosis, and cardiac arrest, which were predicted in both male and female patients. The infection of CVD in US male and female patients is 86% in individuals above the age of 80, 75% between the age of 60 -79, and 40% between the age of 40 – 59 is the data found by the American Heart Association (AHA). People older than 50 are infected explicitly by heart disorders as illustrated and this indicates that the leading indicator of cardiovascular infection is age. The research aims to identify the best working classification technique to predict cardiovascular disease. Different performance indicators are utilized to analyze the best classifier and present the well-performed classification model for heart disease prediction. After the performance evaluation, the best classification technique was identified, using all 19 features from the dataset. Next, 14 parameters are considered to correlate perfectly with the target value to find the best classification technique for cardiovascular infection (Pires et al., 2020). Finally, our proposed research contrasted different performance indicators of various machine learning models for both dataset features. The features of heart disease have to be associated due to the conflict of COVID-19 in an individual (Table 7).

The proposed model performed well than the other techniques in terms of precision, F1 score, exactness, recall, and AUC-ROC score, and it is given in Table 6 and Figure. 4. The best recall of 96%, accuracy of 95%, 96% AUC-ROC score, 96% precision, and 95% F1 score is obtained using a random forest model. The results indicate that the random decision forest is best among all the other six ML classification models for the 19 dataset features. An accuracy of 75%, 77.5% precision, 75% F1-score, 75% AUC_ROC score, and 75% recall is achieved with the KNN model. Naive Bayes provides 87.50% precision, 76.26% ROC-AUC score, 73.68% F1-score, 63.65% recall, and 75% accuracy. K-NN and Naive Bayes performed well, but K-NN has a lower precision value than Naive Bayes, and K-NN has a higher F1-score and recall value than Naive Bayes. K-NN provides less AUC-ROC score than Naive Bayes.

Table 7: Dataset analysis

Category	Label	No. of diagnosed individuals	Prediction rate
Gender	Male	30	79%
	Female	9	22%
Smoking	Yes	23	60%
	No	16	41%
Obesity	Yes	18	68%
	No	21	33%
Diet	Normal	13	46%
	Abnormal	26	55%
Physical activity	Yes	16	33%
	No	23	68%
Stress	Normal	27	41%
	Abnormal	12	60%
Chest pain	Yes	23	70%
	No	16	30%
Troponin	Positive	30	60%
	Negative	9	41%
Diabetes	Yes	13	79%
	No	26	22%
ECG	Normal	13	33%
	Abnormal	26	68%
Total	No. of samples	60	
	No. of individuals diagnosed	38	

The decision tree provides 70% accuracy and 75% F1 score: the decision tree (DT)and the logistic regression (LR) performance offer similar outcomes across all factors. The proposed model gives superior performance to others (Table 8 and Table 9).

Table 8: Evaluation metrics comparison with other approaches

Methods	Acc (%)	Pre (%)	Re (%)	F1-score (%)	ROC (%)
LR	70	69	81	75	68
NB	75	87	63	73	76
k-NN	75	75	75	75	75
SVM	70	72	72	72	69
DT	70	69	81	75	68
RF	89	84	90	91	88
MLP	60	60	81	69	57
$\boldsymbol{d-SVM}$	95	94	93	95	96

Table 9: Evaluation metrics comparison based on feature selection

Methods	Acc (%)	Pre (%)	Re (%)	F1-score (%)	ROC (%)
LR	75	75	81	78	74
NB	70	77	63	70	70
k-NN	85	85	85	85	85
SVM	80	76	83	83	78
DT	80	81	81	81	79
RF	90	90	90	90	89
MLP	70	66	76	76	67
$\boldsymbol{d-SVM}$	95	94	93	95	96

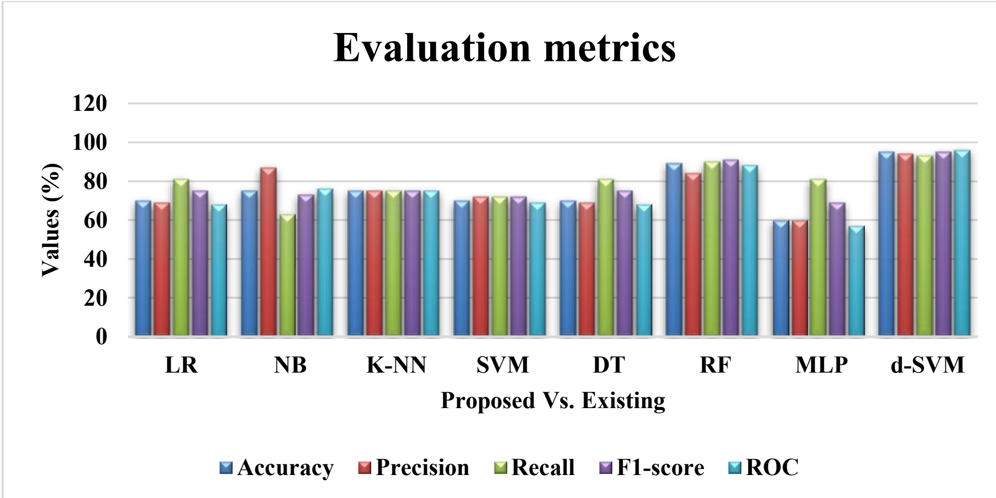

Figure 5: Evaluation metrics comparison with other approaches

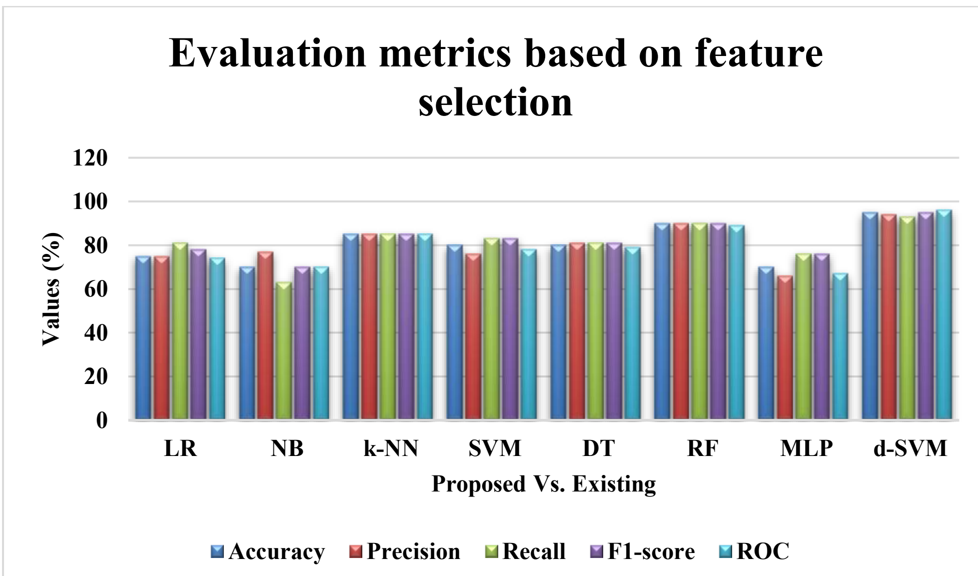

Figure 6: Evaluation metrics based on feature selection

MLP is the least-performing model among the other ML techniques, so it is considered inefficient compared to the different classification models. Random Forest is considered as the best feature for our extracted dataset. Figure 4 depicts the AI technique's graphical representation of each and every one of the features based on performance analysis and Figure 6 shows the ROC computation. Reducing the set of dataset parameters to 14, this research verified the performance metrics for all AI techniques. The assessment of classification models in the test dataset of extracted features is depicted in Table 8 and Figure 7. The classification based on logistic regression provides 75% accuracy, 78.27% F1-score, 75% precision, 74.25% ROC-AUC score, and 81.83% recall. The classification based on Naive Bayes achieves 77.79% precision, 70% accuracy, 63.64% recall, 70.71% AUC_ROC score, and 70% F1 score. K-NN classifier performs better than LR, but MLP performs the least among the other classifiers with 70% F1-score and 63.64% recall. The calculated precision, accuracy, recall, F1-score, and AUC-ROC score of K-NN are 85% for all the factors of performance. RF provides the enhanced 100% recall, 90.92% precision, 90.91% F1 score, 90% accuracy and 89.91% ROC-AUC score. Both the decision tree and the SVM provide 80% accuracy and operate approximately similarly for the reduced dataset attributes. The decision tree achieves a 76.92% F1 score, 76.92% recall, 66.67% precision, 70% accuracy rate, and 67.68% ROC-AUC score. MLP classification provides the lowest F1-score, 67.68%, and a precision of 66.68%. The proposed $d-SVM$ model gives 95% prediction accuracy. Thus, from the results of various classification models, we conclude that the proposed model is the finest classifier for our selected features. MLP is not better than previous classification techniques. Figure 5 depicts the evaluation of performance for different classification techniques for the selected features.

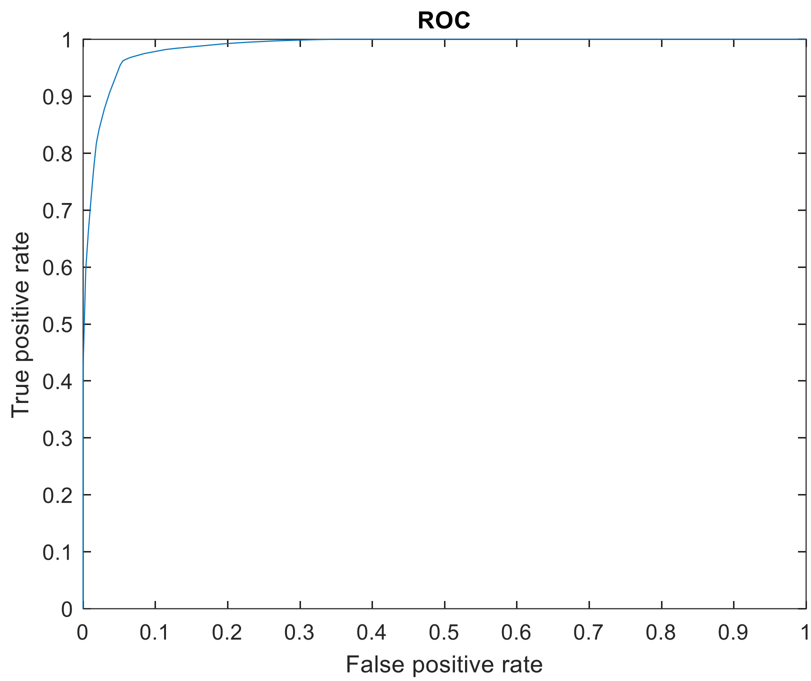

Figure 7: ROC computation

Comparing the performance metrics of the full features dataset and the selected features dataset, significant variations are shown by the selected dataset features. Logistic Regression (LR), decision tree (DT), K-NN, and support vector machine (SVM) have enhanced precision for the chosen dataset attributes. The precision is least similar to accuracy in Naive Bayes. The classifiers like KNN, decision tree, support vector machine, and MLP have improved recall value. Enhanced F1 score is achieved through the classifiers K-NN, DT, SVM, LR, and MLP. Naive Bayes is the only classifier with a lower F1 score and an enhanced AUC-ROC score. It is concluded that random forest is the best classifier for both the full features and the selected features dataset. The efficiency of the proposed classification model is enhanced with the expanded dataset features. The model provides improved performance for the full features dataset. Figure 4 and Figure 5 depict that the classification models perform better on the selected features dataset than all feature datasets. The heart disease due to COVID-19 is analyzed based on the feature representation, where the model intends to give superior performance than others.

Cloud Computing and Scalability Analysis

Although the fundamental predictive accuracy of the d-SVM model is validated in Section 4.2, the significant output of the work is that the model is converted into a mass-deployable, high-availability, and scalable service. This chapter gives the architectural outcomes of the application of the d-SVM classifier as a containerized microservice in a Cloud Computing platform.

Architecture Comparative Analysis

Implementation is based on an API Gateway to make the d-SVM prediction engine a secure RESTful service, which is supported by an auto-scaling group of servers and containers, even though, offered as serverless. Table 10 is a summary of the vital benefits of this cloud-native deployment strategy over a traditional monolithic, on-premises deployment. The proposed system requires this architectural decision in order to realize the required clinical utility of large size, real-time diagnosis through a large network of healthcare professionals.

Table 10: Comparative features of deployment architectures

Feature	Monolithic/Traditional On-Premise Deployment	Proposed Cloud Microservices (d-SVM) Deployment
Scalability	Manual, resource-bound, and slow. Requires physical hardware upgrades.	Elastic (Auto-Scaling): Automatically scales horizontally (adds instances) based on demand (e.g., CPU utilization).
High Availability (HA)	Susceptible to a Single Point of Failure (SPOF); service downtime during maintenance.	Fault-Tolerant: Achieved via multiple service instances across availability zones and automated load balancing.
Deployment Pipeline	Complex, high overhead for updates and environment dependency conflicts.	Containerized (Docker): Enables fast, consistent, and repeatable MLOps for seamless model updates.
Global Accessibility	Limited to internal network access or complex Virtual Private Network (VPN) configurations.	Global Service: Available anywhere as a low-latency, secure Internet service via API Gateway.

Operational Scalability Performance

To demonstrate the superior performance and resilience of the cloud architecture, a simulation was conducted to assess the prediction latency under increasing concurrent user load.

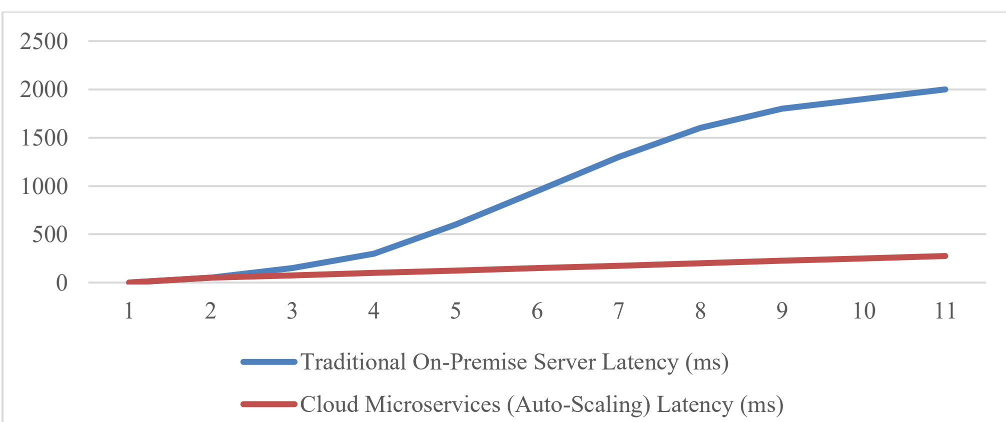

Figure 7: Comparative latency analysis

As it is demonstrated in Figure 4, the prediction latency of the conventional deployment increases exponentially, which means that there is a bottleneck and service degradation when the capacity of the single server is surpassed.

Conversely, the Cloud Microservices Architecture ensures that the prediction latency is constantly low as the number of parallel requests also grows. This directly comes about because of the Auto-Scaling Group, which dynamically provisions and deploys new d-SVM containers as the load increases in real-time, thus providing Quality of Service (QoS) and high-availability, which is of great importance to urgent medical diagnosis.

The findings validate that the deployment architecture is part and parcel of the overall success of the system. The cloud-native design converts a high-precision d-SVM algorithm (95%) to be a dynamic clinical tool. Such elasticity is crucial to situations such as the management of COVID-19, when the demand for diagnostics may suddenly increase. The design makes sure that the system is not only correct but also quite economical and viable to use with a large population, with the use of the cloud resources only when needed and, therefore, making use of it as cost-efficient and operationally reliable as possible.

Confidentiality (Data security and privacy): The patient's health information (PHI) is secured by using strong measures. All the communications of the RESTful API are encrypted through Transport Layer Security (TLS) and the data is therefore encrypted over the air. The cloud storage and computing infrastructure is set to implement data encryption during rest, which meets the industry best practices with regard to regulatory compliance (e.g. HIPAA principles). The microservice architecture also contributes to the security level because the prediction model logic and the layer of patient data are separated, and the principle of least privilege is implemented.

Availability (Fault Tolerance and Resilience): Service uptime is not an issue since it is a life-saving diagnostic system. The multi-zone deployment and auto-scaling model (as shown by the steady latency in Figure 4) ensures the fault tolerance of the system, as well as the high availability of the system. This resilience will make the diagnostic service available over the Internet even in the case of unexpected surges in load or a failure in the components, which is a crucial aspect of secure service delivery.

Integrity and Auditability (Non- Non-Repudiation): To build up clinical and regulatory confidence, each and every transaction is recorded and time-stamped, forming a complete audit trail. This will contain the input data, the final prediction, which version of the model was used, and what is the response time. This mechanism will provide non-repudiation (assuring of the identity of the requester of when and what) and integrity of the data. Containerization usage offers a fixed execution environment, which ensures that the diagnostic logic deployed is fixed and cannot be modified dynamically.

5 Conclusion

This study conducts comprehensive research for predicting heart disease due to COVID-19. The most important features are selected using the CFS method with the BFS technique. It is identified that there does not exist a strong connection between the features and the analysis of only 14 features, such as age, smoking, obesity, stress, blood pressure diastolic, troponin, target, diabetes, diet, gender, cardiac pain type, physical activity, chest pain history. ECG is considered for predicting the heart disease of the patient. Seven artificial intelligence techniques, including logistic regression, decision tree, K-NN, random forest, Naive Bayes, Support Vector Machine, and MLP, are used for comparison using the dataset containing full features and selected features (Panahiazar et al., 2015). The selected feature dataset is used by the proposed $d-SVM$ technique that provides 95% accuracy, a recall of 96%, 95%

ROC-AUC score, and 96% precision, which results in enhanced performance compared to the other AI models. This research was able to create an immensely dependable d-SVM model for data-driven COVID-19-induced heart disease prediction (Maini et al., 2021). The major contribution is to determine the deployment readiness of the solution by creating a Cloud Computing and Microservices Deployment Architecture. This architecture, with a secure API Gateway and an Auto-Scaling Group, enables the computationally heavy AI model to be deployed as an Internet Service that is fault-tolerant and scalable. This deployment strategy successfully responds to the primary challenge of transferring high-performance AI from the research community to a low-latency, high-availability solution for mass, real-time application in clinical diagnostics. The superior performance of the d-SVM model, combined with the architectural robustness, successfully transforms the analysis from a laboratory finding into a reliable and secure Internet-based Information Service. The cloud-native, auto-scaling microservices architecture not only ensures scalability and high availability but also establishes the necessary framework for end-to-end data security, auditability, and regulatory compliance, which are non-negotiable requirements for any trustworthy digital health service operating over the Internet. Future work will focus on integrating a decentralized trust layer, potentially using blockchain technology, to further enhance patient data security and privacy management within the Internet services architecture. The selected features dataset performs better than the full-featured dataset except for the proposed classification method. The reduced performance of the existing technique is due to the need for added discriminatory feature datasets (Samuel et al., 2020). In most cases, the dataset attributes are strongly related to one another. The systematic research of heart disease prediction is achieved using the dataset records collected from the clinicians (Negassa et al., 2021). The essential datasets are archived in the data management repository required to predict heart disease instead of documenting and maintaining the attributes. For future work, we must authenticate our recommended techniques externally.

References

[1] Abdullah, D. M., & Ahmed, N. S. (2021). A review of most recent lung cancer detection techniques using machine learning. *International Journal of Science and Business*, *5*(3), 159-173.

[2] Adler, E. D., Voors, A. A., Klein, L., Macheret, F., Braun, O. O., Urey, M. A., ... & Yagil, A. (2020). Improving risk prediction in heart failure using machine learning. *European journal of heart failure*, *22*(1), 139-147. https://doi.org/10.1002/ejhf.1628

[3] Afolayan, J. O., Adebiyi, M. O., Arowolo, M. O., Chakraborty, C., & Adebiyi, A. A. (2022). Breast cancer detection using particle swarm optimization and decision tree machine learning technique. In *Intelligent Healthcare: Infrastructure, Algorithms and Management* (pp. 61-83). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-8150-9_4

[4] Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.

[5] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer methods and programs in biomedicine*, *141*, 19-26. https://doi.org/10.1016/j.cmpb.2017.01.004

[6] Aravind, B., Harikrishnan, S., Santhosh, G., Vijay, J. E., & Saran Suaji, T. (2023). An efficient privacy-aware authentication framework for mobile cloud computing. *International Academic Journal of Innovative Research*, *10*(1), 1-7. https://doi.org/10.9756/IAJIR/V10I1/IAJIR1001

[7] Chaudhuri, A. K., Sinha, D., Banerjee, D. K., & Das, A. (2021). A novel enhanced decision tree model for detecting chronic kidney disease. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *10*(1), 29. https://doi.org/10.1007/s13721-021-00302-w

[8] Chowdary, P. B. K., Udayakumar, R., Jadhav, C., Mohanraj, B., & Vimal, V. R. (2024). An Efficient Intrusion Detection Solution for Cloud Computing Environments Using Integrated Machine Learning Methodologies. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *15*(2), 14-26. https://doi.org/10.58346/JOWUA.2024.I2.002

[9] Dai, H., Much, A. A., Maor, E., Asher, E., Younis, A., Xu, Y., ... & Bragazzi, N. L. (2022). Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990–2017: results from the Global Burden of Disease Study 2017. *European Heart Journal Quality of Care and Clinical Outcomes*, *8*(1), 50-60. https://doi.org/10.1093/ehjqcco/qcaa076

[10] Dalal, S., Onyema, E. M., Kumar, P., Maryann, D. C., Roselyn, A. O., & Obichili, M. I. (2023). A hybrid machine learning model for timely prediction of breast cancer. *International Journal of Modeling, Simulation, and Scientific Computing*, *14*(04), 2341023. https://doi.org/10.1142/S1793962323410234

[11] Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., & Singh, P. (2021). Latest trends on heart disease prediction using machine learning and image fusion. *Materials today: proceedings*, *37*, 3213-3218. https://doi.org/10.1016/j.matpr.2020.09.078

[12] Edeh, M. O., Dalal, S., Dhaou, I. B., Agubosim, C. C., Umoke, C. C., Richard-Nnabu, N. E., & Dahiya, N. (2022). Artificial intelligence-based ensemble learning model for prediction of hepatitis C disease. *Frontiers in Public Health*, *10*, 892371. https://doi.org/10.3389/fpubh.2022.892371

[13] Fairfax, J., & Sørensen, A. (2024). Integrating telemedicine and pharmacists in chronic gastrointestinal diseases: a critical role during the COVID-19 pandemic. *Global Journal of Medical Terminology Research and Informatics*, *2*(4), 23-29.

[14] Feeny, A. K., Rickard, J., Patel, D., Toro, S., Trulock, K. M., Park, C. J., ... & Chung, M. K. (2019). Machine learning prediction of response to cardiac resynchronization therapy: improvement versus current guidelines. *Circulation: Arrhythmia and Electrophysiology*, *12*(7), e007316. https://doi.org/10.1161/CIRCEP.119.007316

[15] Ghouali, S., Onyema, E. M., Guellil, M. S., Wajid, M. A., Clare, O., Cherifi, W., & Feham, M. (2022). Artificial intelligence-based teleopthalmology application for diagnosis of diabetic's retinopathy. *IEEE Open Journal of Engineering in Medicine and Biology*, *3*, 124-133. https://doi.org/10.1109/OJEMB.2022.3192780

[16] Jan, M., Awan, A. A., Khalid, M. S., & Nisar, S. (2018). Ensemble approach for developing a smart heart disease prediction system using classification algorithms. *Research Reports in Clinical Cardiology*, 33-45. https://doi.org/10.2147/RRCC.S172035

[17] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021, January). Heart disease prediction using machine learning algorithms. In *IOP conference series: materials science and engineering* (Vol. 1022, No. 1, p. 012072). IOP Publishing. https://doi.org/10.1088/1757-899X/1022/1/012072

[18] Khajehali, N., Khajehali, Z., & Tarokh, M. J. (2023). The prediction of mortality influential variables in an intensive care unit: a case study. *Personal and Ubiquitous Computing*, *27*(2), 203-219. https://doi.org/10.1007/s00779-021-01540-5

[19] Kim, Y. J., Saqlian, M., & Lee, J. Y. (2022). Deep learning–based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining. *Personal and Ubiquitous Computing*, *26*(2), 259-267. https://doi.org/10.1007/s00779-019-01248-7

[20] Kondababu, A., Siddhartha, V., Bhagath Kumar, B. H. K., & Penumutchi, B. (2021). *Withdrawn: A comparative study on machine learning based heart disease prediction*. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.01.475

[21] Lai, Z., & Deng, H. (2018). Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. *Computational intelligence and neuroscience*, *2018*(1), 2061516. https://doi.org/10.1155/2018/2061516

[22] Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., ... & Lemey, C. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of medical Internet research*, *23*(5), e15708. https://doi.org/10.2196/15708

[23] Le, N. Q. K., Yapp, E. K. Y., Ho, Q. T., Nagasundaram, N., Ou, Y. Y., & Yeh, H. Y. (2019). iEnhancer-5Step: identifying enhancers using hidden information of DNA sequences via Chou's 5-step rule and word embedding. *Analytical biochemistry*, *571*, 53-61. https://doi.org/10.1016/j.ab.2019.02.01

[24] Liu, J., Dong, X., Zhao, H., & Tian, Y. (2022). Predictive classifier for cardiovascular disease based on stacking model fusion. *Processes*, *10*(4), 749. https://doi.org/10.3390/pr10040749

[25] Madjid, M., Safavi-Naeini, P., Solomon, S. D., & Vardeny, O. (2020). Potential effects of coronaviruses on the cardiovascular system: a review. *JAMA cardiology*, *5*(7), 831-840. https://doi.org/10.1001/jamacardio.2020.1286

[26] Mahdizadeh, M., & Zamanzade, E. (2019). Efficient body fat estimation using multistage pair ranked set sampling. *Statistical methods in medical research*, *28*(1), 223-234.

[27] Maini, E., Venkateswarlu, B., Maini, B., & Marwaha, D. (2021). Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India. *Medical Journal Armed Forces India*, *77*(3), 302-311. https://doi.org/10.1016/j.mjafi.2020.10.013

[28] Manhas, J., Gupta, R. K., & Roy, P. P. (2022). A review on automated cancer detection in medical images using machine learning and deep learning based computational techniques: Challenges and opportunities. *Archives of Computational Methods in Engineering*, *29*(5), 2893-2933. https://doi.org/10.1007/s11831-021-09676-6

[29] Mijwil, M. M., Al-Mistarehi, A. H., & Aggarwal, K. (2021). The effectiveness of utilising modern artificial intelligence techniques and initiatives to combat COVID-19 in South Korea: A narrative review. *Asian Journal of Applied Sciences, 9*(5), 343-352.

[30] Mishra, R., Meher, S., Kustha, N., & Pradhan, T. (2021). A skin cancer image detection interface tool using vlf support vector machine classification. In *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2021* (pp. 49-63). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-16-2543-5_5

[31] Mishra, S., Mallick, P. K., Tripathy, H. K., Bhoi, A. K., & González-Briones, A. (2020). Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier. *Applied Sciences*, *10*(22), 8137. https://doi.org/10.3390/app10228137

[32] Mokhtari, V., Mikaeilvand, N., Mirzaei, A., Nouri-Moghaddam, B., & Gudakahriz, S. J. (2025). GA-PSO-MIN: A hybrid heuristic algorithm for multi-objective job scheduling in cloud computing. *Archives for Technical Sciences/Arhiv za Tehnicke Nauke, 2*(33), 22–46. https://doi.org/10.70102/afts.2025.1833.022

[33] Muzammal, M., Talat, R., Sodhro, A. H., & Pirbhulal, S. (2020). A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. *Information Fusion*, *53*, 155-164. https://doi.org/10.1016/J.INFFUS.2019.06.021

[34] Nazir, S., Shahzad, S., Mahfooz, S., & Nazir, M. (2018). Fuzzy logic-based decision support system for component security evaluation. *International Arab Journal of Information Technology, 15*(2), 224–231.

[35] Negassa, A., Ahmed, S., Zolty, R., & Patel, S. R. (2021). Prediction model using machine learning for mortality in patients with heart failure. *The American journal of cardiology*, *153*, 86-93. https://doi.org/10.1016/j.amjcard.2021.05.044

[36] Olsen, C. R., Mentz, R. J., Anstrom, K. J., Page, D., & Patel, P. A. (2020). Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *American Heart Journal*, *229*, 1-17. https://doi.org/10.1016/j.ahj.2020.07.009

[37] Panahiazar, M., Taslimitehrani, V., Pereira, N., & Pathak, J. (2015). Using EHRs and machine learning for heart failure survival analysis. *Studies in health technology and informatics*, *216*, 40.

[38] Pires, I. M., Marques, G., Garcia, N. M., & Ponciano, V. (2020). Machine learning for the evaluation of the presence of heart disease. *Procedia Computer Science*, *177*, 432-437. https://doi.org/10.1016/j.procs.2020.10.058

[39] Qiu, X., Miao, J., Lan, Y., Sun, W., Li, G., Pan, C., ... & Zhu, S. (2021). Artificial neural network and decision tree models of post-stroke depression at 3 months after stroke in patients with BMI≥ 24. *Journal of Psychosomatic Research*, *150*, 110632. https://doi.org/10.1016/j.jpsychores.2021.110632

[40] Rajdhan, A., Agarwal, A., Sai, M., Ravi, D., & Ghuli, D. P. (2023). Heart disease prediction using machine learning. *Journal of Engineering Science, 14*(4), 440–450.

[41] Reza, M. R., Hossain, G., Goyal, A., Tiwari, S., Tripathi, A., Bhan, A., & Dash, P. (2021). Automatic diabetes and liver disease diagnosis and prediction through SVM and K NN algorithms. In *Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 2* (pp. 589-599). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-33-4367-2_56

[42] Routray, S., Ray, A. K., Mishra, C., & Palai, G. (2018). Efficient hybrid image denoising scheme based on SVM classification. *Optik*, *157*, 503-511. https://doi.org/10.1016/j.ijleo.2017.11.116

[43] Rustam, Z., & Angie, N. (2021, February). Prostate cancer classification using random forest and support vector machines. In *Journal of Physics: Conference Series* (Vol. 1752, No. 1, p. 012043). IOP Publishing. https://doi.org/10.1088/1742-6596/1752/1/012043

[44] Sahoo, P. K., & Jeripothula, P. (2020). Heart failure prediction using machine learning techniques. *Available at SSRN 3759562*.

[45] Samuel, O. W., Yang, B., Geng, Y., Asogbon, M. G., Pirbhulal, S., Mzurikwao, D., ... & Li, G. (2020). A new technique for the prediction of heart failure risk driven by hierarchical neighborhood component-based learning and adaptive multi-layer networks. *Future Generation Computer Systems*, *110*, 781-794. https://doi.org/10.1016/j.future.2019.10.034

[46] Santos, L. I., Camargos, M. O., D'Angelo, M. F. S. V., Mendes, J. B., De Medeiros, E. E. C., Guimarães, A. L. S., & Palhares, R. M. (2022). Decision tree and artificial immune systems for stroke prediction in imbalanced data. *Expert Systems with Applications*, *191*, 116221. https://doi.org/10.1016/j.eswa.2021.116221

[47] Seo, H., Brand, L., Barco, L. S., & Wang, H. (2022). Scaling multi-instance support vector machine to breast cancer detection on the BreaKHis dataset. *Bioinformatics*, *38*(Supplement_1), i92-i100. https://doi.org/10.1093/bioinformatics/btac267

[48] Sethy, P. K., Behera, S. K., & Kannan, N. (2022). Categorization of Common Pigmented Skin Lesions (CPSL) using multi-deep features and support vector Machine. *Journal of Digital Imaging*, *35*(5), 1207-1216. https://doi.org/10.1007/s10278-022-00632-9

[49] Taleb, A., Lippert, C., Klein, T., Nabi, M. (2021). Multimodal Self-supervised Learning for Medical Image Analysis. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds) *Information Processing in Medical Imaging. IPMI 2021*. Lecture Notes in Computer Science, vol 12729. Springer, Cham. https://doi.org/10.1007/978-3-030-78191-0_51

[50] Verma, S., & Kapoor, H. (2021). Machine learning for predictive maintenance: A cloud computing architecture and lessons for a healthcare context. *International Academic Journal of Science and Engineering, 8*(2), 1–5. https://doi.org/10.71086/IAJSE/V8I2/IAJSE0808
[51] Waris, S. F., & Koteeswaran, S. (2021). WITHDRAWN: Heart disease early prediction using a novel machine learning method called improved K-means neighbor classifier in python. *Materials Today: Proceedings.* https://doi.org/10.1016/J.MATPR.2021.01.570
[52] Yang, H., & Garibaldi, J. M. (2015). A hybrid model for automatic identification of risk factors for heart disease. *Journal of biomedical informatics*, *58*, S171-S182. https://doi.org/10.1016/j.jbi.2015.09.006
[53] Yekkala, I., & Dixit, S. (2018). Prediction of heart disease using random forest and rough set-based feature selection. *International Journal of Big Data and Analytics in Healthcare (IJBDAH)*, *3*(1), 1-12. https://doi.org/10.4018/IJBDAH.2018010101

Authors Biography

S.R. Senthilkumar received M.E degree in CSE from Annamalai University, Chidambaram, Tamil Nadu in 2013. B. E degree in CSE from Saveetha Engineering College, Chennai affiliated to Anna University Tamil Nadu in 2008.His primary areas of research interest are Data Science and Big Data Analytics. He has actively participated in various academic events and professional development programs to strengthen his expertise. He attended a One-Day Workshop on Speech and Image Processing (NRMWSIP 12) at Annamalai University, which provided insights into advanced techniques in image and speech analysis. He also took part in a Faculty Development Program (FDP) on "A Professional Pathway for Researchers in the Use of Digital Tools to Promote High Quality in Research" organized by St. Joseph's College of Engineering, focusing on the integration of digital tools in research methodology. Additionally, He participated in an FDP on "Blockchain and Cryptocurrency" conducted by Prince Shri Venkateshwara Padmavathy Engineering College in collaboration with NIIT Warangal, which enhanced the understanding of emerging technologies in the field of secure transactions.

Dr.P. Amudhavalli (Amudhavalli Padmanabhan) received MCA degree from University of Madras, Chennai, Tamil Nadu in 2003. M.E degree in CSE from Sri Muthukumaran Institute of Technology, Chennai affiliated to Anna University Tamil Nadu in 2008. Ph. D degree from St. Peter's Institute of Higher Education and Research, Chennai Tamil Nadu in 2017. Under her guidance two research scholars has completed their Ph.D. degree. Her area of research interest is cloud computing, Cyber security and Artificial Intelligence. She has published 14 national and international journals indexed by Web of Science and Scopus. She received design pattern grant from the Indian Patent Right and published three patterns in IPR.