Human-centric Computing and Information Sciences

December 2022 | Volume 12

www.hcisjournal.com

KIPS Korea Information Processing Society

KIPS CSWRG Korea Information Processing Society Computer Software Research Group

HCIS | Human-centric Computing and Information Sciences

RESEARCH(Research Manuscript)Open Access

Human-centric Computing and Information Sciences (2022) 12:57
DOI: https://doi.org/10.22967/HCIS.2022.12.057
Received: December 26, 2021; Accepted: March 24, 2022; Published: December 15, 2022

Predicting the Risk of Heart Failure Based on Clinical Data

Jasminder Kaur Sandhu[1], Umesh Kumar Lilhore[2], Poongodi M[3], Navpreet Kaur[4], Shahab S. Band[5,*]
Mounir Hamdi[3], Celestine Iwendi[6], Sarita Simaiya[1], M.M. Kamruzzaman[7], and Amir H. Mosavi[8,9,*]

Abstract

The disorder that directly impacts the heart and the blood vessels inside the body is cardiovascular disease (CVD). According to the World Health Organization reports, CVDs are the leading cause of mortality worldwide, claiming the human life of nearly 23.6 million people annually. The categorization of diseases in CVD includes coronary heart disease, strokes, and transient ischemic attacks (TIA), peripheral arterial disease, aortic disease. Most CVD fatalities are caused by strokes and heart attacks, with an estimated one-third of these deaths currently happening before 60. The standard medical organization "New York Heart Association" (NYHA) categorize the various stages of heart failure as Class I (with no symptoms), Class II (mild symptoms), Class III (comfortable only when in resting position), Class IV (severe condition or patient is bed-bound), and Class V (unable to determine the class). Machine learning-based methods play an essential role in clinical data analysis. This research presents the importance of various essential attributes related to heart disease based on a hybrid machine learning model. The proposed hybrid model SVM-GA is based on a support vector machine and the genetic algorithm. This research analyzed an online dataset obtainable at the UCI Machine Learning Repository with the medical data of 299 patients who suffered from heart failures and are classified as Class III or IV as per the standard NYHA. This dataset was collected through patients' available follow-up and checkup duration and involved thirteen clinical characteristics. The proposed machine learning models were used to calculate feature importance in this research. The proposed model and existing well-known machine learning based-models, i.e., Bayesian generalized linear model, ANN, Bagged CART, Bag Earth, and SVM, are implemented using Python and various performance measuring parameters, i.e., accuracy, processing time, precision, recall, F-measures are calculated. Experimental analysis shows the proposed SVM-GA model strengthens in terms of better accuracy, processing time, precision, recall, F-measures over existing methods.

Keywords

Heart Failure, Machine Learning, Computing, Healthcare, Biomedical Diagnosis, Hybrid SVM-GA

***Corresponding Author:**Amir Mosavi(amir.mosavi@nik.uni-obuda.hu), Shahab S. Band (shamshirbands@yuntech.edu.tw)
[1]Department of Computer Science and Engineering, Chandigarh University, Mohali, Punjab, India
[2]KIET Group of Institutions, Delhi-NCR, Ghaziabad (UP), India
[3]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
[4]Chitkara School of Health Sciences, Chitkara University, Punjab, India
[5]Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, Yunlin, Taiwan, ROC
[6]School of Creative Technologies, University of Bolton, Bolton, UK
[7]Department of Computer and Information Science, Jouf University, Sakaka, Al-Jouf, Kingdom of Saudi Arabia
[8]German Research Center for Artificial Intelligence, 26129 Oldenburg, Germany
[9]Institute of the Information Society, University of Public Service, 1083 Budapest, Hungary

1. Introduction

These days, one of the leading causes of death is heart disease. In the context of biomedical data analysis, predicting heart disease is a significant challenge. Researchers in heart disease data analysis widely use machine learning (ML)-based models. These models help facilitate the decision-making and prediction of vast amounts of data generated by the medical field. Cardiovascular disease (CVD) has become a common illness spread throughout the world. It can cause death due to the slow-down process of blood pumping. The blood pumping process is dropped or slows down due to a viscous substance accumulating in the blood vessels. CVD can be cured at early stages, but it is challenging to identify the heart diseases in clinical data analysis [1]. As per NYHA (New York Heart Association), heart failure is categorized into five stages. When the condition worsens, the class I stage shifts to the class II stage because of less blood supply to all body parts. Due to the changing state of heart failure can cause, medications, lifestyle, and cardiac devices will be changed. I show no symptoms in class, but it is considered a pre-heart failure. A family history of heart failure is the reason for class I, and it can be treated with regular exercise in day-to-day life [2]. It can be prevented by changing its lifestyle. Class II comprises mild symptoms and is considered in the pre-heart failure category. Class III can be diagnosed by the previous and current symptoms. Shortness of breathing, swollen feet, and abdomen are the common symptoms in Class III. Class IV shows the worst condition, which does not improve with medications. It requires bed rest, and sometimes heart surgery is the treatment. Class V is the category in the class style of heart failure is not the diagnosis [3].

There is a requirement to predict CVD in health monitoring. Due to the advanced growth of heart illness, researchers focus on the current status of heart failure, people's awareness for CVD, prediction of CVD control factors, and significance of the ML approach to predict heart failures. These efforts improved the Quality of Service in the health area [4]. There are several approaches applied in the past decade to predict the accuracy of CVD. There are three ML methods: supervised ML, unsupervised ML, and reinforcement ML. In this research paper, five supervised ML methods are implemented on the dataset, namely as Bayesian generalized linear model (BGLM), artificial neural network (ANN), Bagged CART, Bag Earth, and support vector machine (SVM). These ML models calculate the significance of the features [5].

The cases of heart failure increase day by day across the world due to the lifestyle of the people. CVD becomes life-threatening when it is not cured on time. Some symptoms of CVD are not visible and diagnosed for many years. So, there is a requirement to identify heart failure and its stage during early stages [6]. The author, in state of the art, proposes distinct models. However, these approaches are limited to too few features of CVD. Five supervised learning ML approaches on 13 clinical features best classify and predict heart failures [7]. Coronary disease is one of the most complex and deadly diseases. In any situation when the heart fails to pump enough blood to all body parts to start the brain's normal functions. As a result, heart failure can take place and cause death.

This research presents the importance of various essential heart disease attributes based on a hybrid ML model. The proposed hybrid model SVM-GA is based on a SVM and genetic algorithm (GA). The contributions of this complete research are as follows:

- The proposed model is based on SVM-GA ML methods.
- We carried a fair comparison among the most widely used ML models that enable researchers to decide which classifiers to use.
- This paper mainly emphasizes the heart illness of health science. Several ML techniques are used in state-of-the-art to classify and predict CVD. Supervised learning is applied to provide the best results for predicting CVD in the early stages.
- Five ML models are applied to the dataset having 13 features.
- The proposed model and existing well-known machine learning based-models, i.e., BGLM, ANN, Bagged CART, Bag Earth, and SVM, are implemented using Python and various performance

measuring parameters, i.e., accuracy, processing time, precision, recall, F-measures are calculated [8].

The complete research paper is organized as follows: Section 1 discussed introduction, Section 2 discussed the related work in heart disease analysis using ML-based model analysis of heart disease datasets for the last decade. Section 3 discussed materials and methods, proposed hybrid model working. Section 4 discussed experimental results and discussion, and the conclusion and future work related to the research article.

2. Related Work

Heart failure (HF), also referred to as congestive heart failure, is among the most lethal common diseases, and accurate HF problems and risks are vital for HF treatment and prevention. ANN-based decision support systems that estimate HF risks often presume that HF variables contribute equally to the HF diagnosis. The risk contributions of the attributes, on the other hand, would be different. As a result, the equal risk assumption idea linked with conventional ANN approaches will misrepresent the diagnosis condition of HF patients [9].

Lilhore et al. [10] integrated ANN with a fuzzy analytic hierarchy process (Fuzzy_AHP) to compare the effectiveness of the suggested expert system to that of the conventional ANN method. An experienced cardiac clinician assessed thirteen HF features using an online clinical dataset of 297 HF patients, and their contributions were determined. The experimental findings showed that, in contrast to ANN, the proposed system could obtain a mean success rate of 91.10%, which is 4.40% higher. There have been numerous efforts to create models that foresee 30-day re-hospitalization in patients with HF, but only a few studies have enough discriminatory power to be used in clinical practice. Bag et al. [11] constructed ML-based models to predict all-cause readmission 30 days following discharge from an HF hospitalization and compared the effectiveness of ML models to models developed using traditional statistically-based methods.

The authors of [12] utilized traditional bioinformatics time-dependent frameworks, like as Cox regression and Kaplan-Meier survival graphs, to anticipate death rates and recognize essential factors of 310 patients with HF (age >42 years) to forecast death and recognize essential factors of all these clients. The findings revealed that high BP, renal dysfunction (level of serum creatinine >1.5 mg/dL), advancing age, lower ejection fraction (EF) values, and a higher level of anemia are the main attributes contributing to a higher risk of mortality in HF patients. Later, Guleria et al. [13] analyzed the same dataset to compare the results of two different survival prediction models—each for women and men. Gender as a risk parameter is unreliable in predicting patient survival. Although the results presented by [14] are intriguing, standard biostatistics methods have been used to handle the issue, leaving an opportunity for ML approaches.

Further, the authors of [15] applied different ML classifiers on the number of clinical features of the dataset (299 patients with HF) to predict HF patient survival and rank the attributes according to the most relevant risk factors. The experimental findings reveal that the EF and serum creatinine are two sufficient features for predicting survival in HF patients. Moreover, an analysis that included each patient's follow-up month proved the effectiveness of these two features in predicting patient survival. In [16], patient diagnostic, physician, and hospital department records are all contained in electronic health records (EHRs). In general, EHR time series can extract large amounts of unstructured data. The relationships between diagnostic events can be established and eventually anticipate when a patient will be diagnosed by studying and mining these time-based EHRs. However, because the existing EHR data is scarce and non-standardized, it is difficult to use it directly in the research [17].

Alvarez-Garcia et al. [18] demonstrated the robustness and effectiveness of neural network architecture for HF prediction. The key fundamentals of an extended short-term memory network model are often used to design the specific diagnostic occurrences and forecast HF instances utilizing vector representation and one-hot embedding (both terminologies relate to the direction diagnosing occurrences

were managed). One of the most crucial factors in the care of HF patients is to keep them out of the hospital as much as possible [19].

In [20], investigators searched for the perfect method to predict readmission rates in heart disease patients for decades. Despite specific scores, clinical use of prediction models for HF hospital treatment is limited. Most scores, namely The HF Patient Severity Index (HFPSI) [21], the REDIN score [21], and the CHARM scores [22], were all validated ambulatory outpatients on chronic HF and comprised biochemical, clinical, and often instrumental data. Jin et al. [23] employed ML approaches to analyze 44,886 HF patients and found that these methods may reliably predict results and recognize clinically differentiated subgroups with variable responses to routinely used medications. The cluster analysis used the eight most predictive indicators to show four clinically meaningful HF subgroups with significant differences in 1-year survival [24]. It was determined that advanced analytics on extensive clinical data collection would identify different patient profiles, improve outcome prognostication, and uncover therapy response heterogeneity [25]. Khennou et al. [26] suggested a new model based on SVM and random forest (RF) that achieved an accuracy of 81.34% on the Cleveland cardiovascular disease sample data. Several researchers have explored integrating GAs with selection and prediction approaches, including sampling methods and enabling more accuracy. ANNs are the most favored attribute selection method in these optimization algorithms [27]. The authors of [28] aggregated a GA to SVM for precise gene analysis and heart disease extracted features classification.

2.1 Comparative Analysis of Existing Heart Disease Research

Table 1 represents a comparative analysis of various existing research works based on ML and other models for heart disease analysis [27, 29–35].

Table 1. Comparative analysis of existing research work

Study	Methods	Key findings	Dataset	Future scope
Ahmad et al. [29]	IoMT assisted heart disease diagnostic system using machine learning technique, i.e., SVM, RF, and AdaBoost	The best accuracy was 84.1% for the RF ML method.	Heart disease UCI dataset	More performance measuring parameters can measure.
Gokulnath and Shantharajah [30]	IoT-based diseases prediction	Data-preprocessing and classification of heart disease	Healthcare Kaggle dataset	Accuracy can be improved.
Ketu and Mishra [31]	Naive Bayes, random forest, decision tree	Decision trees show better accuracy 90.5%	Heart disease UCI dataset	More performance measuring parameters can measure.
Kyrimi et al. [32]	SVM, kNN, and ensemble learning methods	SVM method shows better accuracy	Heart disease UCI dataset	Accuracy can be improved by using hybrid models.
Bag et al. [33]	GA, RNN, and DNN	RNN achieves 91.8% accuracy	Heart image dataset	More performance measuring parameters can measure.
Chicco and Jurman [27]	Random oversampling method, adaptive sampling method, and SVM	SVM method shows a better accuracy of 90.4%	Heart disease Kaggle dataset	Accuracy can be improved by using hybrid models.
Kyrimi et al. [34]	Neural network model CNN	CNN achieves more than 90.1% accuracy	Image dataset UCI	More performance measuring parameters can measure.
Rathod and Patil [35]	Genetic algorithm, kNN, and SVM	SVM shows more than 90% accuracy and precision	Online UCI heart disease dataset	Accuracy and precision can be improved.

3. Materials and Methods

3.1 Description of Dataset and Features Analysis

The sample for this research was obtained from the UCI dataset, which includes 299 heart disease patients at the Islamabad Medical College & Hospital and the Allied Hospital in Islamabad (Punjab, Pakistan) [36]. Statistics of the dataset complies with sick people having 105 female and 194 male of the age of 35–90years. Table 2 shows the 13 characteristics/attributes of the heart disease dataset. The dataset contains clinical, body, and style of living details. The disorder was detected by a heart echo preliminary report or written notes by a healthcare professional. The dataset includes the following attributes: patient age, level of serum sodium, details of serum creatinine, patient gender type, and the habit of smoking, level of blood pressure, details of EF, platelets, anemia, level of diabetes, and level ofcreatine phosphokinase (CPK) [37].

In the dataset, anemia is a binary value that indicates a reduction in blood cells and hematocrit. The binary value of hypertension indicates whether or not the person has high blood pressure. The amount of the CPK protease in the body is evaluated in micrograms per liter or micrograms. Diabetes is a binary value that reveals whether the person has symptoms. The percentage of blood that leaves each heart contraction is measured by the EF, expressed in percentages [38]. The person's gender is a Boolean value that indicates whether the person is a male or a female. The count of platelets in the blood is measured in $\times 10^3$ platelets/mL and the concentration of serum creatinine in milligrams per deciliter (mg/dL). The sodium value in the body is calculated in milli-equivalents per liter (mEq/L). Cigarette smoke is a binary variable that indicates whether or not the person is a smoker. The follow-up period is measured in days. The death event parameter is a binary value that indicates whether the person died during the follow-up [39]. The dataset's central tendency measure shows in Table 2.

Table 2. Dataset details (data type, characteristics, and measurement)

S.No.	Characteristic	Measurement or unit	Data type
1	Age	Year	Double
2	Anemia	Boolean	Integer
3	High blood pressure	Boolean	Integer
4	Creatinine phosphokinase	μg/L	Integer
5	Diabetes	Boolean	Integer
6	Ejection fraction	Percentage	Integer
7	Sex	Binary	Integer
8	Platelets	$\times 10^3$ platelets/mL	Double
9	Serum creatinine	mg/dL	Double
10	Serum sodium	mEq/L	Double
11	Smoking	Boolean	Integer
12	Time	Day	Integer
13	Death event	Boolean	Integer

3.2 Statistical Summary of Dataset

The data is limited to the training dataset for the summarization purpose. The dataset consists of 209 observations and 13 features [40]. There are no missing values in the data. Age, platelets, and serum creatinine are all double data types; additionally, in integer data type, anemia, diabetes, CPK, high blood pressure, EF, gender type, serum sodium, smoking habit, time, and death event. The first and third quartile computations show that 25% of respondents of the findings have beliefs of that parameter which are fewer than or higher than the amount mentioned.

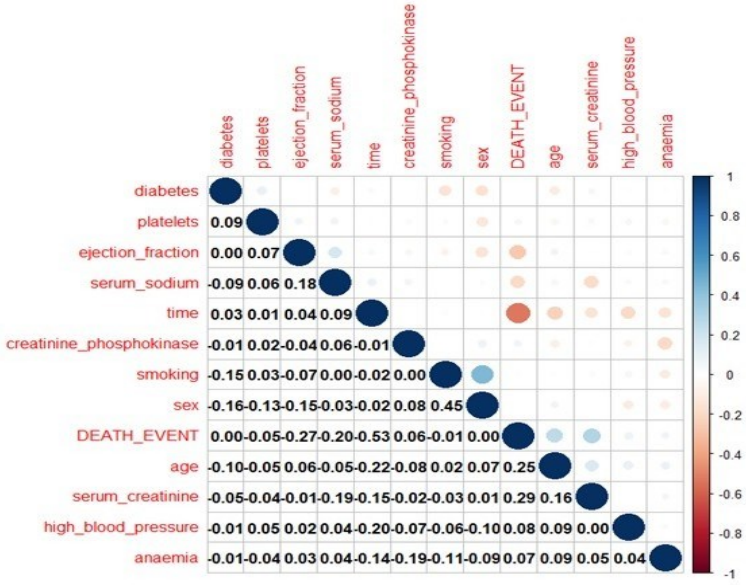

Fig. 1. Visualization of the correlation matrix.

The graphical representation of the correlation matrix with the numerical values is shown in Fig. 1. The degree of correlation can be interpreted with the help of color, shape, and the numerical values in the matrix. A particular feature is perfectly correlated with itself. The color depicts the strength of correlation. The negative correlation among the dataset's characteristics is reflected by the red shades, while the blue shades reflect the positive correlation [41]. Box plots help visualize the skewness of the feature distribution of datasets for particular characteristics of the heart dataset. The box plot of the 12 input features is shown below in Fig. 2(a)–2(j).

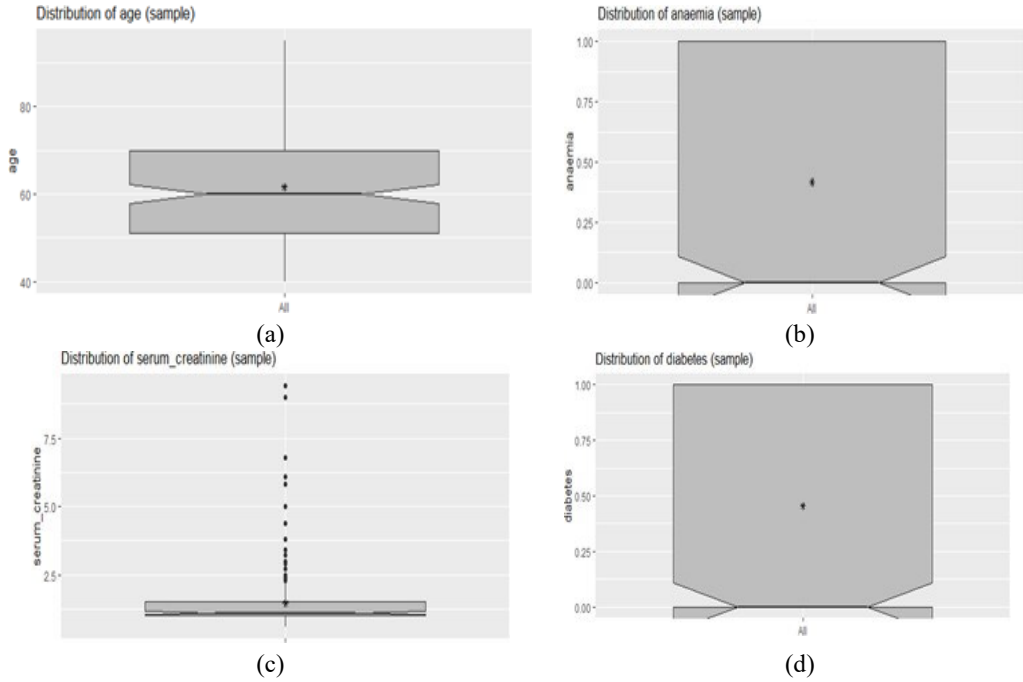

Fig. 2. Continued.

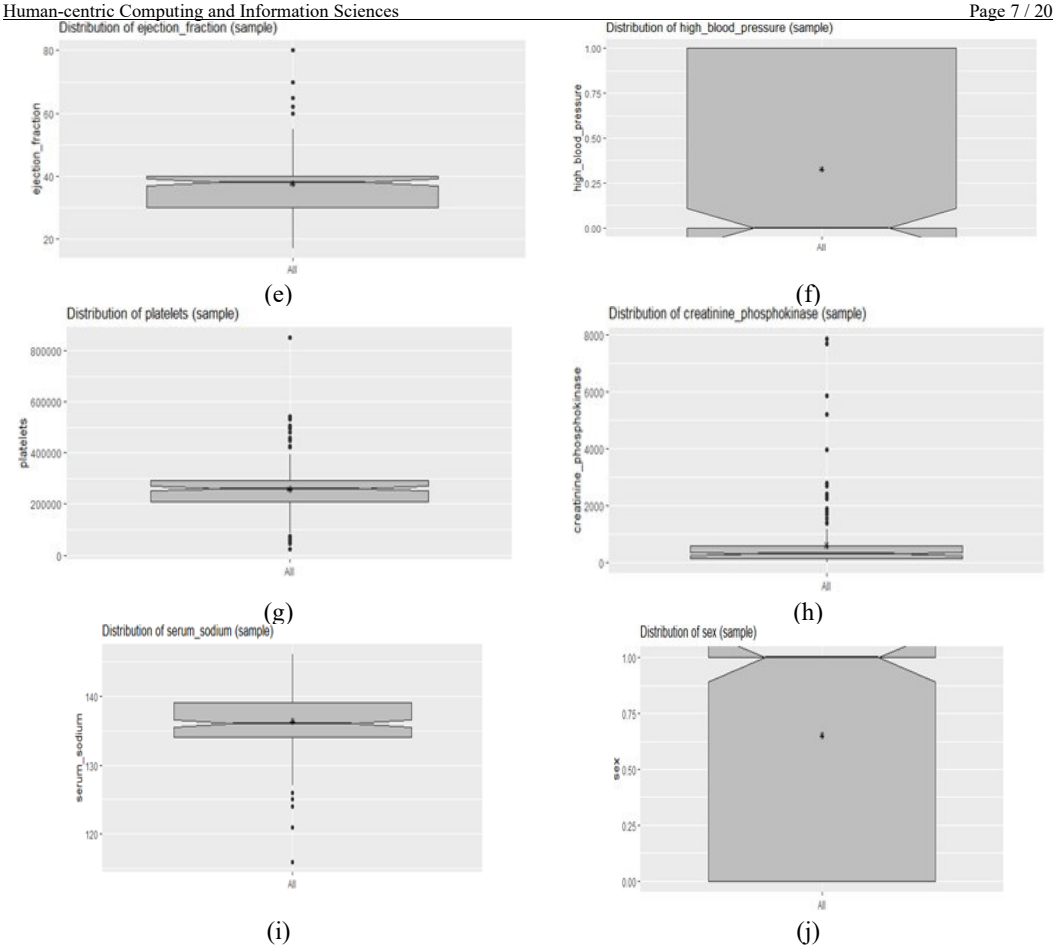

Fig. 2. Box plot of input features: (a) age, (b) anemia, (c) creatinine, (d) diabetes, (e) ejection fractions, (f), blood pressure,(g) platelets, (h) creatinine phosphokinase, (i) serum sodium, and (j) sex.

The data is available on the UCI repository with a detailed description of features mentioned in Table 3. The method and the tuning parameters used to implement these models in R are discussed in Table 3. Table 4 shows the details of ML models and their tuning parameters.

Table 3. Dataset central tendency measure

Feature	Min	1st quartile	Median	Mean	3rd quartile	Max
Patient age	40.00	52.00	60.00	60.76	69	94
Anemia	0	0	0	0.445	1.00	1.00
High blood pressure	0	0	0	0.3589	1.00	1.00
Creatinine phosphokinase	23	119	250	563	582	5861
Diabetes	0	0	0	0.3923	1.00	1.00
Ejection fraction	14.00	30.00	38.00	38.26	48.00	30.00
Sex	0	0	1.00	0.6411	1.00	1.00
Platelets	25100	213000	255000	260007	302000	742000
Serum creatinine	0.500	0.900	1.100	1.438	1.400	9.400
Serum sodium	121.0	134.0	137.0	136.7	139.0	148.0
Smoking	0	0	0	0.3158	1.00	1.00
Time	6.0	74.0	109.0	126.1	198.0	278.0
Death event	0	0	0	0.3014	1.00	1.00

Table 4. Machine learning models and their tuning parameters

S.No.	Machine learning classifier	Method name	Tuning parameters
1	Bayesian generalized linear model	bayesglm()	family = gaussian
2	Artificial neural network	nnet()	linout = TRUE, skip = TRUE, MaxNWts = 10000, trace = FALSE, maxit = 100; size = 10;
3	Bagged CART	bagging()	coob=TRUE
4	Bag Earth	bagEarth()	B = 50
5	Support vector machine	ksvm()	type = "C-svc", kernel = "rbf- dot", prob.model = TRUE

3.3 Existing Machine Learning-Based Models

This section describes the working of various existing ML-based models used in this research for comparison with the proposed hybrid model.

3.3.1 Bayesian generalized linear model

BGLM method is a linear regression approach for establishing associations. It is used to solve the problem of overfitting and fit a good dataset into a reasonable size [42–47]. Based on preliminary data, it calculates the prior distribution. The sample data is then combined with the prior data to get the posterior distribution. Because it integrates expert opinions with sample data, the information produced by the posterior distribution is closer to genuine information. The R programming languages' arm package has been used to implement BGLM in this work. Bayesian techniques are used for modeling complex research problems [48].

This article uses the bayesglm() function for generalized Bayesian modeling using the Gaussian family. The vital building blocks of GLM are the error distribution of the predicted dependent variable, the linking function on which independent variable effects are collected additively, and the collection of terminologies used for the linear independent variable. The main advantage of this model is that it uses external information to enhance the calculation of the linear model coefficients [49]. Further, this model is expressed as primary parameters w representing weights and the hyperparameters q. The set of observations is denoted by S, the likelihood will be represented by P(S|w), and P(w) denotes the Bayesian prior distribution [50]. Equation (1) represents the BGLM method function.

$$P(\omega|S)\ \alpha\ Qi\ \theta i*(Ui)\ ,\ \ Ui = \psi T\phi i \tag{1}$$

Equation (1), known as the Bayesian posterior, is a linear function with the scalar value and is a scalar function with non-negative values are log-concave values. This linear model uses Gaussian inference. Bayesian inference can be computed analytically [51].

Algorithm 1. Bayesian generalized linear model

Input: Selection of training and testing dataset partition.

Output: Calculate the prediction accuracy from the computations.

Select an optimum value for Value=Family

while the stop criteria are not fulfilled

do Execute the training procedure of BGLM using the bayesglm() function

Set the default tuning parameters to end while

Return the prediction accuracy as output

It is observed that the prediction output is comparative to the linear model, the generalized linear model. Also, the prediction output converges to the same values with an increase in the sample size.

3.3.2 Artificial neural network

ANN exhibits the properties of the human brain. ANN learns from the training data and provides the classification category as the output. ANN is a non-linear statistical technique that helps understand a complex relationship between explanatory and response variables and enables the discovery of a novel pattern in the data [52]. The structure of an ANN is demonstrated in Fig. 3. The first layer of ANN, which receives the input information from the dataset, is the input layer. The center layer of ANN, known as the hidden layer, performs various mathematical computations on the data to discover the novel pattern. There can be single or multiple hidden layers in an ANN. After these computations, the output yielding the prediction outcome of classification is provided to the output layer [53].

Also, the prediction outcome depends on specific parameters, namely, batch size, weights, biases, rate of learning. In the ANN structure, every node has some assigned weight. The weighted summation of input data and the bias is computed with the help of the transfer function. The activation function is then applied to realize the result. The input neurons are denoted by, and the output neurons are denoted by ψ. The hidden layers use the activation function for the binary classification problem, calculated as [54].

$$L = F\left(\Sigma nWTi * ki + \psi\right) \tag{2}$$

Equation (2) represents the ANN function for the hidden layer. Here are the connected input() and output neuron (). A represents the bias. This article applies the nnet() function for the feed-forward ANN with a single hidden layer [55].

3.3.3 Bagged CART

The main idea of using the ensemble learning technique is to consider multiple models for calculating the prediction accuracy. This approach has the advantage of reducing the high bias and high variance because multiple averaging models are carried out. Bagging Classification and Regression Trees or the Bagged CART model is implemented in two main steps: creating bootstrapping samples and applying the bagging method [56]. Bootstrapping is a statistics-based technique used to create multiple samples from the dataset by not disturbing the properties of the existing dataset. Every individual sample is termed a bootstrap sample and tends to replicate the complexity of original data [57]. The bagging process is represented with the help of Equation (3):

$$Pbag = P1 * (X) + P2 * (X) + P3 * (X) \ldots\ldots\ldots + Pb * (X) \tag{3}$$

The term Pbag represents the bagged prediction, and the terms (P1*(X) + P2*(X) +….. Pb*(X)) denotes the individual learners based on the input features of the dataset. For the classification problem in this article, the concept of hard-voting is used where the most voted class is considered [3]. In Fig. 4, data points are selected randomly and with replacement properties for equal size samples. In the bagging technique, all the models are constructed parallelly. Bagging is also known as the bootstrap aggregation technique, as all the bootstrap samples are treated equally. The final prediction output is based on the fundamentals of majority voting, as shown in Fig. 3.

Algorithm 2. Artificial neural network

Input: Selection of training and testing dataset partition.
Output: Calculate the prediction accuracy from the computations.
Stop Criteria: Maximum number of iterations as 100

Select an optimum value for Wti and A in the ANN classifier
while the stop criteria are not fulfilled
do Execute the training procedure of ANN using the nnet() function

Set the tuning parameters of nnet(): Number of units in the hidden layer <-10
Set the parameter to linear output units
Add skip-layer connections from input to output
Maximum allowable weights! 100000
end while
Return the prediction accuracy as output

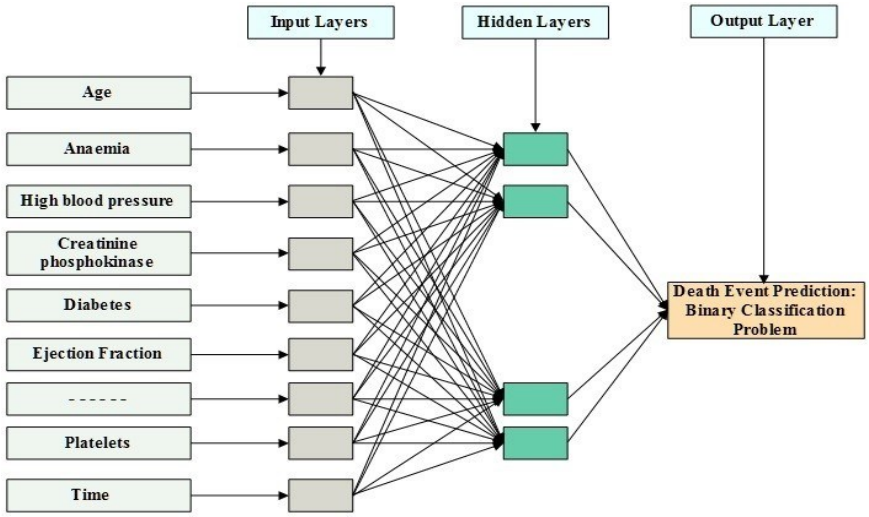

Fig. 3. Structure of feed-forward ANN with single-hidden layer.

3.3.4 Bag Earth

This is a bagging wrapper function for multivariate adaptive regression splines (MARS) using the earth function. It takes the matrix of input variables and the target outcomes for training the model [58]. The main advantage of this model is that it improves Accuracy and reduces variance, hence eliminating the overfitting issue. This further helps in increasing the stability of models. The Bag Earth model is based on the fundamentals of the MARS model. The model is constructed using Equation (4).

$$\mathrm{Fn(X)} = \Sigma \mathrm{N\ Ci} * \mathrm{Bi\ (X)} \quad (4)$$

This Bag Earth model is the weighted summation of the primary function Bi(X), which represents the constant coefficient. The primary function can be a constant value, a hinge function, or a product of multiple hinge functions [59]. The main steps involved in the bagging procedure are the original dataset is divided into multiple sub-sets with the same number of tuple records, and the records are selected with replacement, models based on these base observations are created which is known as the base models, these models are trained in parallel and executed independently, the final prediction is calculated based on these intermediate results and thus is more accurate [60]. Fig. 4 shows the structure or Bagged CART method.

Algorithm 3. Bag Earth

Input: Selection of training and testing dataset partition.
Output: Calculate the prediction accuracy from the computations.

Select an optimum value for the number of bootstrap samples as B=50 while the stop criteria are not fulfilled
do Execute the training procedure of Bagged Earth using the bayes_Earth() function end while

Return the prediction accuracy as output

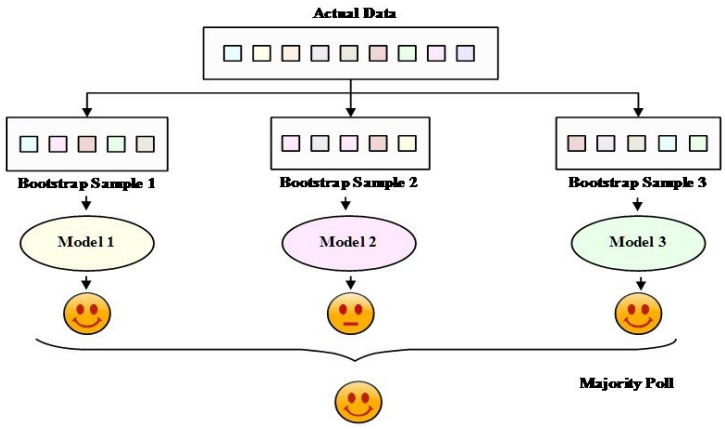

Fig. 4. Structure of Bagged CART.

3.3.5 Support vector machine

It is mainly a dual-use classifier that can be implemented on regression and classification data. However, it is more predominantly preferred for classification data [61]. The primary function of SVM is to provide distinct classification to the data points with the help of a hyperplane in K-dimensional space, where K represents the number of features [62]. SVM is a ubiquitous, effective, and robust supervised ML algorithm commonly used to solve prediction-related problems. Using non-linear kernels, it extracts the various data points and separates them into the feature space of n-dimensions. In this work, the hyperplane partition the feature space into severity classes using a tagged training dataset. A new category is assigned to tagged classes [63]. SVM's operation is based on two primary phases. It first determines the decision limits that accurately identify the training dataset followed by the boundary selection with the most significant distance from the nearest data points from among those boundaries.

The primary goal of this classifier is to find the best hyperplane to partition the class. It contains various variables that require adjusting, such as P and γ. The former variable controls how the accurate prediction and smooth decision boundaries of training data points interact; however, the latter represents the impact of a single training. If the P variable has a significant value, a complex curve boundary is produced that fits all the data points to gather more training data points accurately. Different values of P are necessary for the dataset to avoid the problem of overfitting and create a fully stable curve. Higher and lower values of variables imply that each data point has a close and extensive reach, respectively. Algorithm 1 depicts the steps followed for the execution of the SVM classifier. In this article, SVM is used for the classification scenario. Unlike support vectors, an optimum hyperplane is selected to calculate the highest margin with the closest data points and function. The optimization process for the input is a quadratic equation when training the supervised model. The hyperplane construction is defined by Equation (5).

$$Wt * (X + i) = 0 \tag{5}$$

where W denotes the weight vector, t represents stages, and i represent the bias. To maximize the boundary inside the hyperplane and their closest point denoted the support vector. The kernel trick used in this article is the radial basis kernel or the rbfdot.

Algorithm 4. Support vector machine

Input: Selection of training and prediction dataset.

Output: Calculate the prediction accuracy from the computations.

Select an optimum value for P and γ in the SVM classifier

while the stop criteria are not fulfilled, do

Execute the training procedure of SVM for each training data point Predict function is executed for predicting data points using SVM end while

Return the prediction accuracy as output

3.3.6 Genetic algorithm

A GA is a type of search heuristic method dependent on Charles Darwin's concept of natural development. This method mimics natural law, in which the fittest participants are recruited for propagation to generate the next gen's descendants [64]. Fig. 5 shows the complete working of the GA method.

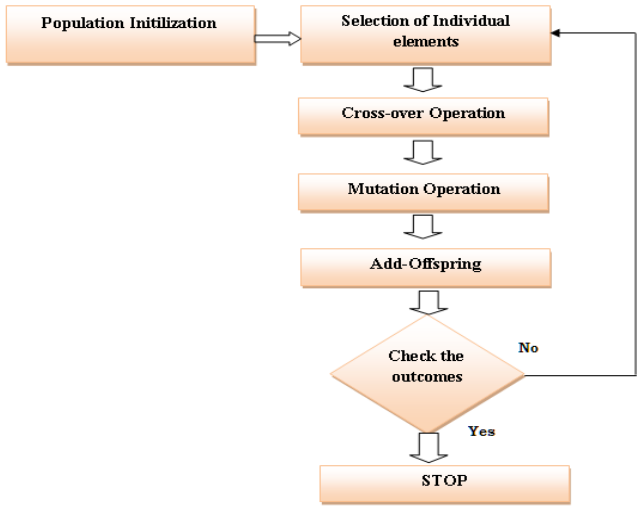

Fig. 5. Working of GA method.

The working of GA method is as follows.

- Initialization of populations: The strength of evolutionary methods relies on their parallel processing searching space expansion. It can occur owing to several alternatives, each examining a neighborhood of the search space. The genetic population is a complete set of solutions that have been initialized randomly.
- Selection of population: The selection algorithms should better decide which participants can propagate, produce and who cannot.
- Crossover operation: Once all the participants are correctly decided and selected, a crossover operation is performed. This operation merged the participants to generate a better population in the next phase. The fundamental assumption behind the crossover operation is that if two participants who are well-adapted to the ecosystem are chosen, their offspring have genetic data from parents (mother and father).
- Mutation operation: The number of a participant's genes and nodes, generally only one, alter arbitrarily due to the mutation operation.

3.3.7 Proposed hybrid SVM-GA model

The proposed hybrid model is based on the SVM and GA method. The hybrid technique suggested in this article is focused on building a hybrid solution that integrates GA and SVM to perform classification before choosing the smallest number of meaningful variables. The SVM is highly advantageous while

attempting to perform dichotomous categorization, that is, when distinguishing between two different classes, as explained in the earlier sections. However, a similar principle can be expanded to determine the n classes that a given data originally belonged to the new dataset.

We have selected an SVM-GA procedure in this research work. A SVM is formed by assuming that n types (t_1, t_2,..., t_n) can be described for a feasible t_i group inside the input sequence of the entire database. It will aim to identify whether a variable fits the specified t_i type of the sampling portion. Finally, to evaluate which particular subset each data point originally belonged to, we developed a set for all the definite SVM entries and implemented them. In choosing the best output, a better degree of data is considered in this approach. We adjusted the GA's conventional function to produce a subset of the population of varied lengths to develop the proposed hybrid model. Fig. 6 shows the working of the proposed hybrid model, and Algorithm 5 shows the complete functioning of the proposed method.

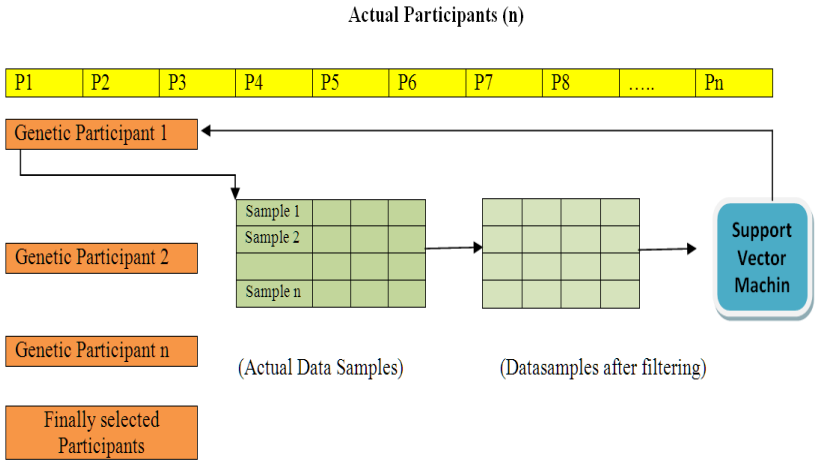

Fig. 6. Working of proposed hybrid (SVM-GA) model.

Algorithm 5. Proposed hybrid SVM-GA method

Input: Dataset D = (P_1, P_2, P_3, P_4 . . ., P_n, C), $Size_p$: Population size, $Size_c$: length of the Chromosome, P_{cross}: Probability of Crossover, P_{mut}: Probability of Mutation, I: Iterations

Where: features/ variables P_1, P_2, P_3, P_4 . . ., P_n, Class C,

Output: Best data samples will get

Step1: Initialize all the parameters

Step 2: Select the dataset D

2.1 Normalization (D); // call normalization function

Step3: Genrates the initial population

3.1 Genrate_Population(D);

3.2 Assign the population

Population(i,j) = Random_Population (D);

Step 4: creates subgroups

4.1 Repeat step till n

4.2 Sub_set(D, n);

Step 5: Calculate the fitness value by SVM

5.1 Fitness_value(D_i) = fitness[n];

5.2 Arrange the values based on fitness

D_n= Sort(fitness[n]);

Step6: Apply cross over operation

6.1 Chromosom(I,Population)= {Chromosom (I,$Population_{crossover}$)/ $Croos_{Result}$}

Step7: Apply mutaion operation

Step 8: Find out the best particpants

8.1 Optimize ($Best_{individual}$(Dn)n fitness[n]);

25: return optimum and best fitness

To achieve the above, we have built an initial procedure that will be accountable for generating the length of each participant in the GA population randomly or by specific size and, subsequently, initializing the values of each gene variable. We have chosen a fraction of arbitrary size from the complete collection of attributes; for this reason, so even if the characteristics created vary, the fraction will be allocated to the individuals. Alternatively, it generates a unique random value until the status stipulated is encountered. This operation is continued until all GA participants have been allocated to a subgroup.

4. Results and Discussion

The proposed model and existing well-known ML based-models, i.e., BGLM, ANN, Bagged CART, Bag Earth, and SVM methods are implemented using Python and various performance measuring parameters, i.e., accuracy, processing time, precision, recall, F-measures and confusion matrix are calculated. Equations (6) to (9) represent the performance measuring parameters. Fig. 7 shows the confusion matrix to compare the actual and predicted results.

		Predicted value	
		P	N
True value	P	TP	FN
	N	FP	TN

Fig. 7. Confusion matrix.

$$\text{Accuracy} = \{(TP + TN) /[(TP+TN+FN+FP)]\} \quad (6)$$

$$\text{Precision} = \{[TP/ (FP+TP)]\} \quad (7)$$

$$\text{Recall} = \{[TP/ (FN+TP)]\} \quad (8)$$

$$\text{F-Measure} = \{2*[(Recall*Precision) / Recall + Precision)]\} \quad (9)$$

where TP true positive, FP false positive, TN true negative, and FN false negative.

The sample for this research was obtained from the UCI dataset, which includes 299 heart disease patients at the Islamabad Medical College & Hospital and the Allied Hospital in Islamabad (Punjab, Pakistan) [36]. The dataset consists of 13 clinical features, including the target death event and 12 input features. The UCI dataset was initially divided into the training and testing samples using the stratification 10-fold cross-validation method. The approach subdivides into 10-folds randomly while keeping the number of data for each class training and test data fold. It guarantees that all folds have the same data distribution as their source sample. This training set is now being used to train the proposed SVM-GA. The method is then combined with feature selection methods to improve the classification performance. An analysis of prediction results of five existing machine learning models and proposed SVM-GA is carried out on the training-testing partition size of 70% and 30%. To predict the patients' survival, we

used several methodologies. Each method was tested on 50 instances, and the average outcome scores are presented in Table 5. The accuracy of the models is assessed, and the moment required to developing them. At last, K-fold cross-validation is used to test the better classifier model's reliability. The ML models used for this work include BGLM, ANN, Bagged CART, Bag Earth, Random Forest, Decision Tree, and SVM [65–67].

Table 5. Experimental results mean of 50 executions

S.No.	Machine learning classifier	Accuracy (%)	Precision (%)	Recall (%)	F-measure (%)	Time taken (ms)
1	Bayesian generalized linear model	88.33	80.51	82.6	84.35	117.04
2	Artificial neural network (ANN)	87.89	88.98	86.7	83.98	112.61
3	Bagged CART	86.67	87.58	87.2	85.6	114.08
4	Bag Earth	85.1	86.57	88.9	85.9	117.47
5	Support vector machine (SVM)	83.33	84.96	86.6	84.65	116.54
6	Random Forests	87.09	85.91	84.95	85.98	111.25
7	Decision tree	85.14	84.09	83.56	84.97	114.56
8	Proposed SVM-GA method	91.49	94.25	93.6	90.89	105.67

Table 5 and Fig. 8 demonstrates the experimental results of the existing seven ML techniques and the proposed SVM-GA model using HCV datasets. It is observed that BGLM and ANN models obtain higher Accuracy of 88.33% because BGLM uses external information to enhance the calculation of the linear model coefficients, and ANN facilitates understanding a complex relationship between explanatory and response variable thereby enabling discovery of a novel pattern in the data. However, the BGLM model outperforms the ANN in terms of the execution time for the models to execute. Also, it is noticed that the models Bagged CART, Bag Earth, and SVM exhibit good accuracy and the considerably same average time for execution. Class imbalance is overcome by the proposed SVM method. The proposed SVM-GA method shows 91.49% accuracy, which is better than other ML methods, and it also takes less processing time of 105.67 ms. Subsequently, the proposed model shows precision 94.25%, recall 93.6%, F-measures 90.89%, which is better than the existing ML model.

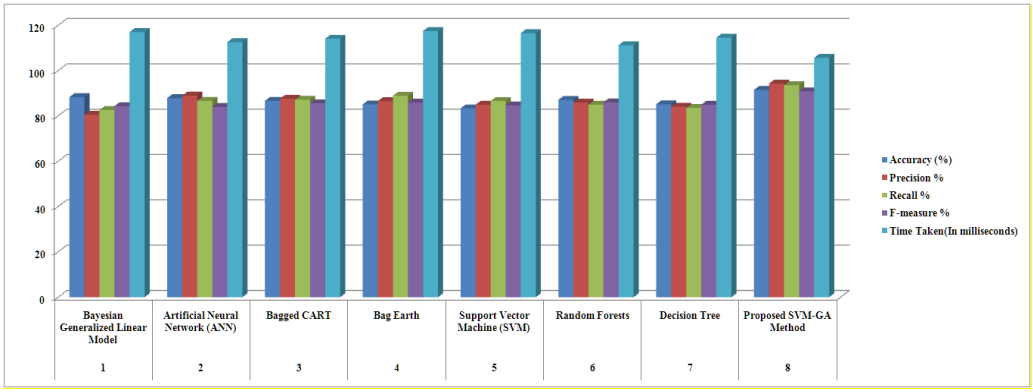

Fig. 8. Comparisons graph for proposed and existing ML method mean of 50 executions.

5. Conclusion

This article considers the CVD Category III and IV of NYHA to predict the death event rate in patients who suffered from HFs. The proposed hybrid model is based on the SVM and GA method. The hybrid technique suggested in this article is focused on building a hybrid solution that integrates GA and SVMs

to perform classification before choosing the smallest number of meaningful variables. The proposed method is compared with existing ML models BGLM, ANN, Bagged CART, Bag Earth, and SVM. Both the model's BGLM, ANN outperforms the other models with 88.33% classification accuracy. These prediction results have enormous potential to influence the clinical records and act as a supporting tool for physicians while predicting the death event in case of HF. The most crucial features for predicting the death event are serum creatinine and ejection fraction. Class imbalance is overcome by the proposed SVM method. The findings show that the proposed framework is influential in determining the risk of HF. The proposed model shows accuracy of 91.49%, precision 94.25%, recall 93.6%, F-measures 90.89%, which is better than the existing ML model.

As mentioned in the previous section, SVM is robust and helpful when performing forecasting statistics with an error margin. This assessment process has the undeniable benefit of the reproducibility of outcomes. Consequently, each group of factors will often have a relatively similar forecasting model linked with about the same fitness level, utterly contrary to how it would occur in the particular instance of someone using ANN as a benchmark of perfection for genetic participants. In the future, ensemble modeling can be carried out further to enhance the prediction accuracy of the proposed architecture.

Author's Contributions

Conceptualization, JKS, UKL, AM, PM, SS, CI, SSB, NP. Funding acquisition, AM, SSB. Investigation and methodology, PM, UKL. Writing of the original draft, JKS, PM, UKL. Writing of the review and editing, UKL, SSB, AM. Validation, JKS, UKL, AM, PM, SS, CI, SSB, NP. Formal analysis, JKS, UKL, AM, PM, SS, CI, SSB, NP. Data curation, JKS, UKL, AM, PM, SS, CI, SSB, NP. Visualization, UKL, PM, SS.

Funding

The European Union's Horizon 2020 Research and Innovation Programme under the Programme SASPRO 2 COFUND Marie Sklodowska-Curie grant agreement No. 945478.

Competing Interests

The authors declare that they have no competing interests.

References

[1] K. Divya, A. Sirohi, S. Pande, and R. Malik, "An IoMT assisted heart disease diagnostic system using machine learning techniques," in *Cognitive Internet of Medical Things for Smart Healthcare*. Cham, Switzerland: Springer, 2021, pp. 145-161.

[2] V. Sharma, S. Yadav, and M. Gupta, "Heart disease prediction using machine learning techniques," in *Proceedings of 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Greater Noida, India, 2020, pp. 177-181.

[3] I. Raeesi Vanani and M. Amirhosseini, "IoT-based diseases prediction and diagnosis system for healthcare," in *Internet of Things for Healthcare Technologies*. Singapore: Springer, 2021, pp. 21-48.

[4] S. Barik, S. Mohanty, D. Rout, S. Mohanty, A. K. Patra, and A. K. Mishra, "Heart disease prediction using machine learning techniques," in *Advances in Electrical Control and Signal Systems*. Singapore: Springer, 2020, pp. 879-888.

[5] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart disease prediction using hybrid machine learning model," in *Proceedings of 2021 6th International Conference on Inventive Computation Technologies (ICICT)*, Coimbatore, India, 2021, pp. 1329-1333.

[6] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive approach for heart disease prediction using machine learning," in *Proceedings of 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, 2020, pp. 1-5.

[7] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020.

[8] F. Ali, S. El-Sappagh, S. R. Islam, D. Kwak, A. Ali, M. Imran, and K. S. Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208-222, 2020.

[9] I. M. El-Hasnony, O. M., Elzeki, A. Alshehri, and H. Salem, "Multi-label active learning-based machine learning model for heart disease prediction," *Sensors*, vol. 22, no. 3, article no. 1184, 2022. https://doi.org/10.3390/s22031184

[10] U. K. Lilhore, S. Simaiya, D. Prasad, and D. K. Verma, "Hybrid weighted random forests method for prediction & classification of online buying customers," *Journal of Information Technology Management*, vol. 13, no. 2, pp. 245-259, 2021.

[11] S. Bag, S. Gupta, T. M. Choi, and A. Kumar, "Roles of innovation leadership on using big data analytics to establish resilient healthcare supply chains to combat the COVID-19 pandemic: a multimethodological study," *IEEE Transactions on Engineering Management*, 2021. https://doi.org/10.1109/TEM.2021.3101590

[12] K. Siau and Z. Shen, "Mobile healthcare informatics,' *Medical informatics and the Internet in Medicine*, vol. 31, no. 2, pp. 89-99, 2006.

[13] K. Guleria, A. Sharma, U. K. Lilhore, and D. Prasad, "Breast cancer prediction and classification using supervised learning techniques," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 6, pp. 2519-2522, 2020.

[14] K. Harimoorthy and M. Thangavelu, "Multi-disease prediction model using improved SVM-radial bias technique in healthcare monitoring system," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 3, pp. 3715-3723, 2021.

[15] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "COVID-19 pandemic: role of machine learning & deep learning methods in diagnosis," *International Journal of Current Research and Review*, vol. 13, no. 06, pp. 150-156, 2021.

[16] M. Poongodi and S. Bose, "Stochastic model: reCAPTCHA controller based co-variance matrix analysis on frequency distribution using trust evaluation and re-eval by Aumann agreement theorem against DDoS attack in MANET," *Cluster Computing*, vol. 18, no. 4, pp. 1549-1559, 2015.

[17] M. Poongodi and S. Bose, "A novel intrusion detection system based on trust evaluation to defend against DDoS attack in MANET," *Arabian Journal for Science and Engineering*, vol. 40, no. 12, pp. 3583-3594, 2015.

[18] J. Alvarez-Garcia, A. Ferrero-Gregori, T. Puig, R. Vazquez, J. Delgado, D. Pascual-Figal, et al., "A simple validated method for predicting the risk of hospitalization for worsening of heart failure in ambulatory patients: the Redin-SCORE," *European Journal of Heart Failure*, vol. 17, no. 8, pp. 818-827, 2015.

[19] D. Misra, V. Avula, D. M. Wolk, H. A. Farag, J. Li, Y. B. Mehta, et al., "Early detection of septic shock onset using interpretable machine learners," *Journal of Clinical Medicine*, vol. 10, no. 2, article no. 301, 2021. https://doi.org/10.3390/jcm10020301

[20] I. Ahmad, I. Ullah, W. U. Khan, A. Ur Rehman, M. S. Adrees, M. Q. Saleem, Q. Cheikhrouhou, H. Hamam, and M. Shafiq, "Efficient algorithms for E-healthcare to solve multiobject fuse detection problem," *Journal of Healthcare Engineering*, vol. 2021, article no. 9500304, 2021. https://doi.org/10.1155/2021/9500304

[21] F. M. Zahid, S. Ramzan, S. Faisal, and I. Hussain, "Gender based survival prediction models for heart failure patients: a case study in Pakistan," *PLoS One*, vol. 14, no. 2, article no. e0210602, 2019. https://doi.org/10.1371/journal.pone.0210602

[22] S. Simaiya, U. K. Lilhore, D. Prasad, and D. K. Verma, "MRI brain tumour detection & image segmentation by hybrid hierarchical K-means clustering with FCM based machine learning model," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 1, pp. 88-94, 2021.

[23] B. Jin, C. Che, Z. Liu, S. Zhang, X. Yin, and X. Wei, "Predicting the risk of heart failure with EHR sequential data modeling," *IEEE Access*, vol. 6, pp. 9256-9261, 2018.

[24] A. Kishor and C. Chakraborty, "Artificial intelligence and Internet of Things based healthcare 4.0 monitoring system," *Wireless Personal Communications*, 2021. https://doi.org/10.1007/s11277-021-08708-5

[25] M. Poongodi, S. Bose, and N. Ganeshkumar, "The effective intrusion detection system using optimal feature selection algorithm," *International Journal of Enterprise Network Management*, vol. 6, no. 4, pp. 263-274, 2015.

[26] F. Khennou, C. Fahim, H. Chaoui, and N. E. H. Chaoui, "A machine learning approach: using predictive analytics to identify and analyze high risks patients with heart disease," *International Journal of Machine Learning and Computing*, vol. 9, no. 6, pp. 762-767, 2019.

[27] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1-16, 2020.

[28] R. Sujitha and B. Paramasivan, "Distributed healthcare framework using MMSM-SVM and P-SVM classification," *Computers, Materials and Continua*, vol. 70, no. 1, pp. 1557-1572, 2021.

[29] T. Ahmad, L. H. Lund, P. Rao, R. Ghosh, P. Warier, B. Vaccaro, et al., "Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients," *Journal of the American Heart Association*, vol. 7, no. 8, article no. e008081, 2018. https://doi.org/10.1161/JAHA.117.008081

[30] C. B. Gokulnath and S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease," *Cluster Computing*, vol. 22, no. 6, pp. 14777-14787, 2019.

[31] S. Ketu and P. K. Mishra, "Scalable kernel-based SVM classification algorithm on imbalance air quality data for proficient healthcare," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2597-2615, 2021.

[32] E. Kyrimi, S. McLachlan, K. Dube, M. R. Neves, A. Fahmi, and N. Fenton, "A comprehensive scoping review of Bayesian networks in healthcare: past, present and future," *Artificial Intelligence in Medicine*, vol. 117, article no. 102108, 2021. https://doi.org/10.1016/j.artmed.2021.102108

[33] R. Bag, M. Ghosh, B. Biswas, and M. Chatterjee, "Understanding the spatio-temporal pattern of COVID-19 outbreak in India using GIS and India's response in managing the pandemic," *Regional Science Policy & Practice*, vol. 12, no. 6, pp. 1063-1103, 2020.

[34] E. Kyrimi, K. Dube, N. Fenton, A. Fahmi, M. R. Neves, W. Marsh, and S. McLachlan, "Bayesian networks in healthcare: what is preventing their adoption?," *Artificial Intelligence in Medicine*, vol. 116, article no. 102079, 2021. https://doi.org/10.1016/j.artmed.2021.102079

[35] S. R. Rathod and C. Y. Patil, "Performance assessment of ensemble learning model for prediction of cardiac disease among smokers based on HRV features," *International Journal of Biomedical and Clinical Engineering (IJBCE)*, vol. 10, no. 1, pp. 19-34, 2021.

[36] UCI Machine Learning Repository, "Heart Disease Data Set," 2021 [Online]. Available: https://archive.ics.uci.edu/ml/datasets/heart+disease.

[37] K. Sahu, U. K. Lilhore, and N. Agarwal, "An improved data reduction technique based on KNN & NB with hybrid selection method for effective software bugs triage," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 3, no. 5, pp. 633-639, 2018. https://ijsrcseit.com/CSEIT1835146

[38] K. Nakamura, R. Kojima, E. Uchino, K. Ono, M. Yanagita, K. Murashita, K. Itoh, S. Nakaji, and Y. Okuno, "Health improvement framework for actionable treatment planning using a surrogate Bayesian model," *Nature Communications*, vol. 12, article no. 3088, 2021. https://doi.org/10.1038/s41467-021-23319-1

[39] S. L. Hummel, H. H. Ghalib, D. Ratz, and T. M. Koelling, "Risk stratification for death and all-cause hospitalization in heart failure clinic outpatients," *American Heart Journal*, vol. 166, no. 5, pp. 895-903, 2013.

[40] T. Ahmad, A. Munir, S. H. Bhatti, M. Aftab, and M. A. Raza, "Survival analysis of heart failure patients: a case study," *PLoS One*, vol. 12, no. 7, article no. e0181001, 2017. https://doi.org/10.1371/journal.pone.0181001

[41] A. Kumari, N. Agrawal, and U. Lilhore, "Clustering malicious spam in email systems using mass mailing," in *Proceedings of 2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, 2018, pp. 870-875.

[42] S. J. Pocock, D. Wang, M. A. Pfeffer, S. Yusuf, J. J. McMurray, K. B. Swedberg, et al., "Predictors of mortality and morbidity in patients with chronic heart failure," *European Heart Journal*, vol. 27, no. 1, pp. 65-75, 2006.

[43] M. Poongodi, V. Vijayakumar, F. Al-Turjman, M. Hamdi, and M. Ma, "Intrusion prevention system for DDoS attack on VANET with reCAPTCHA controller using information based metrics," *IEEE Access*, vol. 7, pp. 158481-158491, 2019.

[44] M. Poongodi, T. N. Nguyen, M. Hamdi, and K. Cengiz, "Global cryptocurrency trend prediction using social media," *Information Processing & Management*, vol. 58, no. 6, article no. 102708, 2021. https://doi.org/10.1016/j.ipm.2021.102708

[45] K. Arumugam, J. Srimathi, S. Maurya, S. Joseph, A. Asokan, M. Poongodi, A. A. Algethami, M. Hamdi, and H. T. Rauf, "Federated transfer learning for authentication and privacy preservation using novel supportive twin delayed DDPG (S-TD3) algorithm for IIoT," *Sensors*, vol. 21, no. 23, article no. 7793, 2021. https://doi.org/10.3390/s21237793

[46] S. K. Sahoo, N. Mudligiriyappa, A. A. Algethami, P. Manoharan, M. Hamdi, and K. Raahemifar, "Intelligent trust-based utility and reusability model: enhanced security using unmanned aerial vehicles on sensor nodes," *Applied Sciences*, vol. 12, no. 3, article no. 1317, 2022. https://doi.org/10.3390/app12031317

[47] M. Poongodi, T. N. Nguyen, M. Hamdi, and K. Cengiz, "Global cryptocurrency trend prediction using social media," *Information Processing & Management*, vol. 58, no. 6, article no. 102708, 2021. https://doi.org/10.1016/j.ipm.2021.102708

[48] M. Varun and C. Annadurai, "PALM-CSS: a high accuracy and intelligent machine learning based cooperative spectrum sensing methodology in cognitive health care networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4631-4642, 2021.

[49] J. D. Frizzell, L. Liang, P. J. Schulte, C. W. Yancy, P. A. Heidenreich, A. F. Hernandez, D. L. Bhatt, G. C. Fonarow, and W. K. Laskey, "Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches," *JAMA Cardiology*, vol. 2, no. 2, pp. 204-209, 2017.

[50] U. K. Lilhore, S. Simaiya, D. Prasad, and K. Guleria, "A hybrid tumour detection and classification based on machine learning," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 6, pp. 2539-2544, 2020.

[51] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," *Informatics in Medicine Unlocked*, vol. 26, article no. 100655, 2021. https://doi.org/10.1016/j.imu.2021.100655

[52] U. K. Lilhore, S. Simaiya, K. Guleria, and D. Prasad, "An efficient load balancing method by using machine learning-based VM distribution and dynamic resource mapping," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 6, pp. 2545-2551, 2020.

[53] A. Hassan, D. Prasad, M. Khurana, U. K. Lilhore, and S. Simaiya, "Integration of internet of things (IoT) in health care industry: an overview of benefits, challenges, and applications," in *Data Science and Innovations for Intelligent Systems*. Boca Raton, FL: CRC Press, 2021, pp. 165-180.

[54] N. K. Trivedi, S. Simaiya, U. K. Lilhore, and S. K. Sharma, "An efficient credit card fraud detection model based on machine learning methods," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3414-3424, 2020.

[55] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 376, article no. 113609, 2021. https://doi.org/10.1016/j.cma.2020.113609

[56] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, "Reptile search algorithm (RSA): a nature-inspired meta-heuristic optimizer," *Expert Systems with Applications*, vol. 191, article no. 116158, 2022. https://doi.org/10.1016/j.eswa.2021.116158

[57] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-Qaness, and A. H. Gandomi, "Aquila optimizer: a novel meta-heuristic optimization algorithm," *Computers & Industrial Engineering*, vol. 157, article no. 107250, 2021. https://doi.org/10.1016/j.cie.2021.107250

[58] D. Mpanya, T. Celik, E. Klug, and H. Ntsinjana, "Predicting mortality and hospitalization in heart failure using machine learning: a systematic literature review," *IJC Heart & Vasculature*, vol. 34, article no. 100773, 2021. https://doi.org/10.1016/j.ijcha.2021.100773

[59] Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, and A. A. Yarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm," *Computer Methods and Programs in Biomedicine*, vol. 141, pp. 19-26, 2017.

[60] D. A. Anggoro and N. D. Kurnia, “Comparison of accuracy level of support vector machine (SVM) and K-nearest neighbors (KNN) algorithms in predicting heart disease,” *International Journal*, vol. 8, no. 5, pp. 1689-1694, 2020.

[61] S. Ekiz and P. Erdogmus, “Comparative study of heart disease classification,” in *Proceedings of 2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, Istanbul, Turkey, 2017, pp. 1-4.

[62] J. Singh, A. Kamra, and H. Singh, “Prediction of heart diseases using associative classification,” in *Proceedings of 2016 5th International Conference on Wireless Networks and Embedded Systems (WECON)*, Rajpura, India, 2016, pp. 1-7.

[63] S. P. Siddique Ibrahim and M. Sivabalakrishnan, “An evolutionary memetic weighted associative classification algorithm for heart disease prediction,” in *Recent Advances on Memetic Algorithms and its Applications in Image Processing*. Singapore: Springer, 2020, pp. 183-199.

[64] S. P. Shaji, “Prediction and diagnosis of heart disease patients using data mining technique,” in *Proceedings of 2019 international conference on communication and signal processing (ICCSP)*, Chennai, India, 2019, pp. 0848-0852.

[65] S. Ghosh, G. Samanta, and M. De la Sen, “Bayesian analysis for cardiovascular risk factors in ischemic heart disease,” *Processes*, vol. 9, no. 7, article no. 1242, 2021. https://doi.org/10.3390/pr9071242

[66] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Classification of heart disease using artificial neural network and feature subset selection,” *Global Journal of Computer Science and Technology Neural & Artificial Intelligence*, vol. 13, no. 3, pp. 5-14, 2013.

[67] I. K. A. Enriko, “Comparative study of heart disease diagnosis using top ten data mining classification algorithms,” in *Proceedings of the 5th International Conference on Frontiers of Educational Technologies*, Beijing, China, 2019, pp. 159-164.