A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model

Tamilarasi Suresh[1], Tsehay Admassu Assegie[2], Subhashni Rajkumar[3], Napa Komal Kumar[4]
[1]Department of Information Technology, St. Peter's Institute of Higher Education and Research, Chennai, India
[2]Department of Computer Science, College of Natural and Computational Science, Injibara University, Injibara, Ethiopia
[3]Department of Computer Science and Applications, St. Peter's Institute of Higher Education and Research, Chennai, India
[4]Department of Computer Science and Engineering, St. Peter's Institute of Higher Education and Research, Chennai, India

Article Info

ABSTRACT

Heart disease is one of the most widely spreading and deadliest diseases across the world. In this study, we have proposed hybrid model for heart disease prediction by employing random forest and support vector machine. With random forest, iterative feature elimination is carried out to select heart disease features that improves predictive outcome of support vector machine for heart disease prediction. Experiment is conducted on the proposed model using test set and the experimental result evidently appears to prove that the performance of the proposed hybrid model is better as compared to an individual random forest and support vector machine. Overall, we have developed more accurate and computationally efficient model for heart disease prediction with accuracy of 98.3%. Moreover, experiment is conducted to analyze the effect of regularization parameter (C) and gamma on the performance of support vector machine. The experimental result evidently reveals that support vector machine is very sensitive to C and gamma.

Corresponding Author:

Tsehay Admassu Assegie
Department of Computer Science, Injibara University
P.O.B, 40, Injibara, Amhara, Ethiopia
Email: tsehayadmassu2006@gmail.com

1. INTRODUCTION

In recent years, heart disease have become one of the foremost reason of chronic disease related deaths thought the world population [1]–[5]. Moreover, heart disease is among the most frequently occurring diseases in the world affecting 26 million of the world population [1], [2]. Heart disease cases are widely spreading and the number of cases of heart disease patient is annually growing at a rate of 2%. Thus, identification and diagnosis of heart disease is crucial to save human life and reduce the wide mortality rate caused by heart disease through automated intelligent model to assist medical practitioner on clinical decision making during the diagnosis of heart disease patient [3]. In addition, automated or intelligent model provides more accurate and timely result overcoming the problems caused due to human error [4]. Hence, in order for the survival chance of heart disease patient to be increased, an accurate and timely identification of heart disease through intelligent model is critical for better decision making when diagnosing heart disease patient.

Heart disease prediction involves precise classification of a given sample as heart disease positive or heart disease negative class based on the symptoms or features of a given sample or instance. Many researchers have developed intelligent model with focus on improving the performance of heart disease prediction model and there exist model for heart disease prediction in literature. However, there is still larger scope for improving the performance of the existing model for heart disease prediction [5]. Thus, in this study an effort has been made in designing and implementing an effective and more accurate model that classifies a given sample in

heart disease dataset as heart disease positive (patient) or heart disease negative (not patient) class based on the features of a given sample by using the previous experience gained during training. Iterative feature elimination is employed to remove less informative or irrelevant features that does not have effect on the predictive outcome of the implemented model. In addition, the implemented model is optimized through feature selection with iterative model based feature selection by employing random forest and parameter tuning with grid search to find out optimal parameters for model training. Random forest is employed to select relevant features from the dataset and support vector machine is trained on optimal input feature subset. Overall, the objective of this study is discussed as follows: i) to provide comprehensive summary of the existing work on heart disease prediction using machine-learning approach. ii) to design and implement computationally efficient and effective hybrid model for heart disease prediction by using support vector machine and random forest; and iii) to study the effect of radial basis function regularization parameter (C) and gamma on the performance of support vector machine using 3-fold cross validation.

In addition, experiment is conducted on the proposed hybrid model, the performance of the implemented model is validated and compared with existing state of the art, and most recently published heart disease prediction model. The rest of this study is organized as follows: in section 2, a brief summary of the existing work is presented; section 3 describes the methodology used to conduct the study, heart disease dataset used for experimentation and recursive feature elimination method. Section 4, presents experimental results and discusses the findings of the study comparing with the existing work. Finally, section 5 concludes the study.

2. LITREATURE SURVEY

In the literature, several machine-learning approaches have been implemented for prediction of heart disease using automated intelligent system. Many supervised and unsupervised machine-learning algorithm have been implemented [6]–[26]. Most of the developed models are tested on real world clinical heart disease dataset collected from Kaggle and University of California Irvine (UCI) data repository. Some of the studies are discussed in this section. In [6], unsupervised machine-learning approach namely k-means clustering is applied to heart disease identification. The authors implemented k-means clustering on heart disease dataset collected from University of California Irvine data repository. The study compared the classification accuracy of supervised methods along with k-means clustering and the experimental result shows 84.5% accuracy on heart disease prediction. The authors developed hybrid model, which is based on support vector machine and k-means algorithm for heart disease prediction. The k-means clustering is employed for dataset visualization and support vector machine is used to implement the heart disease prediction model.

In addition, Naive Bayes and random forest based heart disease prediction model with varying number of feature is developed to discover heart disease pattern using the varying input feature [27]. The authors highlighted that random forest has better performance on heart disease prediction as compared to nave Bayes. The authors conducted experiment on nave Bayes and random forest and the experimental result shows highest accuracy achieved using random forest is 86.81%. The authors claim that the accuracy improves when all of the 13 features are used in training as compared with smaller number of features such as only 10 features from all the feature set characterizing heart disease dataset samples.

Another comparative study [7] analyzed the performance of random forest and neural network for heart disease prediction. The comparative result shows that neural network performs better than random forest for heart disease prediction. The experimental result also reveals that the highest accuracy achieved by the developed model is 85.03% using neural network classifier and 79.93% using random forest. Machine learning algorithms are widely applied to predict heart disease surgery procedure [21]. The researchers developed predictive model for heart disease surgery procedure using logistic regression model. The study conducted experiment to evaluate the performance of the developed model using heart disease dataset and the experimental result appears to prove that the highest predictive accuracy of 87.3% is achieved using logistic regression. In [18], hybrid intelligent framework is developed for heart disease prediction with k-nearest neighbor (KNN), support vector machine (SVM), artificial neural network (ANN) and decision tree (DT). The comparative result on the predictive accuracy of KNN, SVM, ANN, and DT shows that support vector machine outperforms as compared to KNN, ANN and DT. The highest predictive accuracy archived using the implemented hybrid framework is 86% using support vector machine with optimal feature subset selected by sequential feature selection approach.

3. RESEARCH METHOD

The dataset for this study is obtained from real world Indian Pima heart disease dataset available to the scientific community from UCI machine learning data repository. The dataset consists of 1,025 samples of

which 526 samples are heart disease patient and 499 are not heart disease patient. Thirteen input features as shown in Table 1 characterize each of the sample in the heart disease dataset. Each sample in the heart disease dataset is labeled with a target or class label and a sample belongs to either heart disease positive or heart disease negative class. The proposed framework is developed by employing support vector machine. We have employed grid search for parameter tuning. Moreover, iterative feature elimination approach is implemented to remove irrelevant features from the total 13 features charactering the heart disease dataset, which are illustrated in Table 1.

Table 1. Heart disease dataset feature description

Feature	Description
Age	Age in years
Sex	Gender of the patient (1=male, 0=female)
Resting blood pressure (restbps)	Blood pressure in mmHg
Cholesterol (chol)	Serum cholesterol, continuous value in mm/dl
Fasting blood pressure (fbs)	Fasting blood sugar >120 mg/dl (1=yes, 0=no)
Thalassemia (thalach)	Maximum heart rate achieved in mmHg
Thallium scan(thal)	Nominal (3=normal, 6=fixed defect, 7=reversible defect)
Exercise induced angina(exang)	Nominal (presence of exercise induce angina, 1=present, 0=absent)
Slope (slope)	Nominal (1=up slopping, 2=flat, 3=down slopping)
Status of fluoroscopy (cad)	Nominal (number of vessels colored through fluoroscopy (0 to 3)
Chest pain (cp)	Nominal (Having chest pain=1, no chest pain=0)
oldpeak	ST depression induced by exercise relative to rest nominal (0 to 6)
Resting electrocardiography results (restecg)	Nominal (0=normal, 1=having ST-T, 2=hypertrophy)
Target	Nominal (class label, 1=heart disease patient, 0=not-patient)

3.1. Heart disease dataset description

Heart disease dataset employed in this study for experimental test and model development is described in Table 1. The heart disease dataset consists of 13 features and 2 classes, namely heart disease positive (patient=labeled 1) and heart disease negative (not patient=labeled 0) classes. The proposed model uses these 13 input features to predict patterns among the whole heart disease dataset and to describe a given sample as positive or negative class based on the experience gained during training. The target feature or class label is not used as input feature. The dataset is divided into two, training set (70% of the dataset) and testing set (30% of the dataset).

3.2. Statistical summary of numeric feature

Descriptive statistics such as maximum, minimum and standard deviation is employed to analyze the numerical features of heart disease dataset. The statistical summary of numerical input features such as age cholesterol, total resting blood pressure, fasting blood pressure and maximum heart rate achieved is summarized in Table 2. As we observe summary statistics from Table 2, the maximum and minimum values of age is 77 and 29 respectively.

Table 2. Descriptive statistics for numerical heart disease features

Feature	Min	Max	STD
Age	29	77	9.07
Resting blood pressure	94	200	17.51
Cholesterol	126	564	51.59
Fasting blood pressure	0	1	0.35
Maximum heart rate achieved	71	202	23.00

3.3. Iterative feature elimination

We have employed iterative feature elimination for removing irrelevant features of heart disease dataset that does not have effect on the predictive outcome of the developed model to obtain good outcome on heart disease prediction. Iterative feature elimination removes irrelevant features that mislead the model's predictive capability and ultimately reduce the performance of classification model [20], [23]. Moreover, with reduced feature, the computational time required for model training and storage space requirement is optimized [22], [24]. The input features in the heart disease dataset employed in in training are shown in Table 1. The goal of feature selection is to choose feature subset of *X_subset*, from complete set of input features X = X1, X2 . . . XN, so that the subset *X_subset* predicts the output feature Y with accuracy comparable to the performance of the complete input feature set X, and with great decrease to the computational time [8], [26].

Iterative feature elimination is model based feature selection method [9], [13]. RFE fits a model and removes feature that does not have effect on predictive outcome of model [11]. In addition, iterative feature elimination removes dependencies and collinearity between features [10].

Algorithm 1. Iterative feature elimination

```
Input: Y=y1, y2 . . . yd, d – dimensional feature set.
Output: Xk=xj | j=1, 2, 3 . . . Ik, xj ∈ Y where k=(0, 1, 2 . . . d)
Initialize: number of feature: T
Initialize classification model: M
for I = 0 to N
do
fit M on N [I] and record the predictive accuracy: P [I]
repeat the procedure for I = 1, 2, 3 . . . N
compare N [I] with P [I] If N [I] < P [I] drop the feature N[I],
 else consider the feature as relevant
 end if
end do
```

3.4. Performance metric

To develop heart disease prediction model, support vector machine is trained on heart disease training set. The selected sample in the heart disease dataset is classified into either heart disease patient (positive) or not heart disease patient (negative) class. To evaluate the effectiveness of the developed model on predicting a given sample into the positive or negative class, we have employed different performance measures called performance metric for machine learning model. The performance metrics employed for evaluation of the developed model include accuracy and confusion. Accuracy and confusion matrix are among the most widely used performance measures used by researchers for evaluation of machine learning model on classification task [12], [14], [15]. Accuracy is defined as the ration of true predictions to total samples predicted by a given model [16], [17], [19]. Mathematically, accuracy is defined by the formula given in (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

where TP=true positive (number of correctly classified heart disease patient samples), FP=False Positive (number of incorrectly classified heart disease patient samples which were not heart disease patient), TN=true negative (number of correctly classified not heart disease patient samples), FN=false negative (number of incorrectly classified not heart disease patient samples which were heart disease patient).

4.   RESULT AND DISCUSSION

This section discusses the experimental setup and the number of heart disease dataset samples used in experiment to test the performance of the developed hybrid heart disease prediction framework. Confusion matrix, the effect of gamma and regularization parameter on the performance of SVM is presented. Finally, the experimental result of the developed model is compared to existing recently published researchers work on heart disease prediction using supervised machine learning algorithm. In the experiment, we have used 308 samples (30% of original heart disease dataset) consisting of 141 heart disease patient samples and 167 heart disease negative (not patient) samples. The system has predicted 141 of the positive samples correctly and miss-classified 4 samples of the negative class as demonstrated in confusion matrix in Figure 1.

4.1. Performance of the developed model

In addition to predictive accuracy and confusion matrix, Mathew's correlation coefficient is employed for evaluation of the developed model. Confusion matrix shows class-wise performance of the implemented model, quantifying the classification result on the positive and negative class. As demonstrated in Figure 1, the model miss-classified 4 samples as false positive and perfectly classified the all of 141 true negative class samples.

Figure 1 shows the classification accuracy for the proposed hybrid heart disease prediction model. We observe from Figure 1, that the model miss-classified 4 samples among 308 samples of test set. The experimental result evidently appears to prove that the accuracy of the proposed model over all classes in the dataset is 98.70%. Furthermore, the model does not predicate any of the 308 samples as false negative, which is highly risky outcome in medical dataset classification. Because, if the model predicts more samples as false negative, then heart disease patient is not recommended for further test and this causes death or series health problem and complication at latter stage. We observe in Figure 1 that class-wise accuracy on the true positive

class is 100% accurate. Thus, the model is more effective on predicting heart disease patient as compared to the predictive accuracy on heart disease negative class, with accuracy of 97%. Hence, the approach is effective and recommended for medical decision support during heart disease identification.
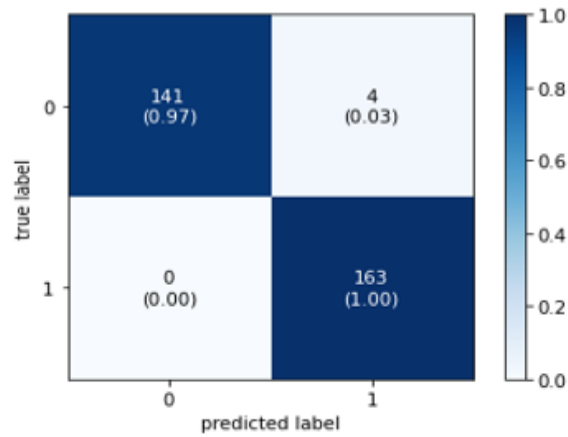

Figure 1. Confusion matrix for the proposed approach

4.2. Effect of hyper-parameters on performance

The effect of gamma and C on the performance of SVM model is demonstrated in Figure 2. Moreover, the classification accuracy for different parameter setting is summarized in Table 3. We observe from Figure 2, that the proposed SVM mode performs well when lower value of gamma and regularization (C) value is used for training the model.
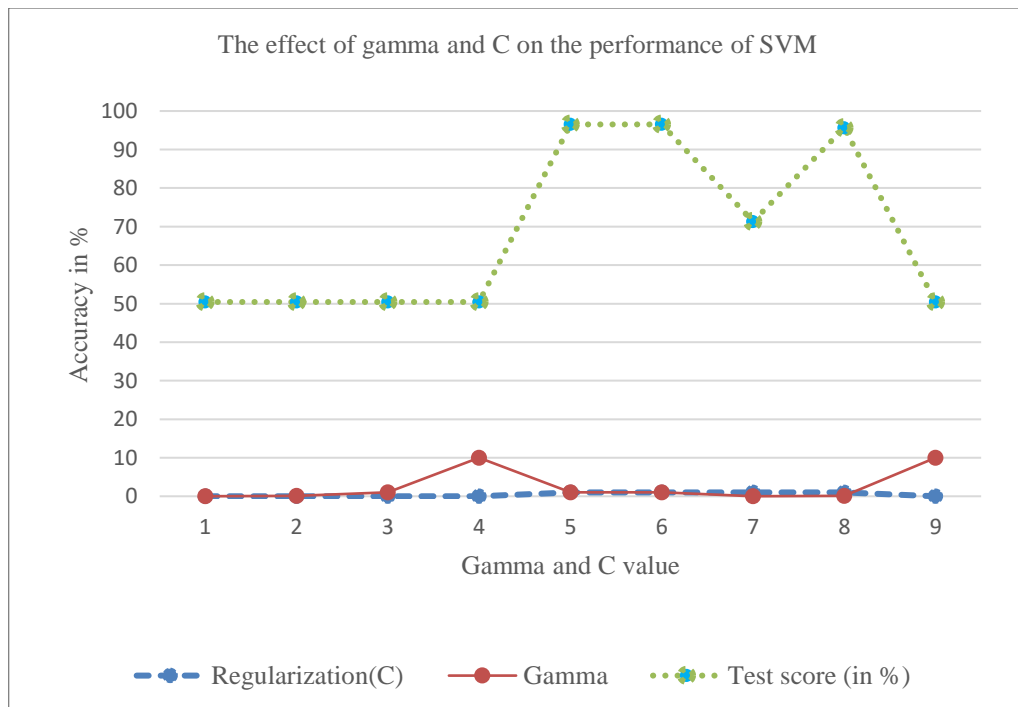

Figure 2. The effect of gamma and regularization parameter (C) on the performance of SVM

As demonstrated in Table 3, predictive performance of support vector machine varies for varied gamma and C values. Table 3 summaries the variation of accuracy using 3-fold cross validation score on heart

disease test set is given in Table 3. We observe from Table 4 that the accuracy varies between 95.65% and 50.43%. Therefore, we conclude that shows that the accuracy of support vector machine is improved up to 46.22% for heart disease prediction by tuning the parameters and using optimal parameters.

4.3. Performance of the proposed framework on optimal input feature

The optimal feature subset selected after removing less informative features using iterative feature elimination consists of feature index (0, 2, 4, 6, 7, 9, 11, 12). The proposed model performed with 98.3% accuracy on heart disease prediction using the optimal feature. The optimal feature subsets selected by RFE, which produced highest possible accuracy on heart disease prediction, is described in Table 4. From Table 4, we observe that the optimal features selected by the iterative feature elimination approach are 5 features among the 13 features in the heart disease dataset. The classification accuracy with the iterative feature elimination is 98.3%.

Table 3. Classification accuracy for different parameter setting and folds

Mean-fit time	Std_score_time	Parameters	split0_test_score	split1_test_score	split2_test_score
0.0342	0.0034	{'C': 0.001, 'gamma': 0.01}	0.5086	0.5043	0.5043
0.0596	0.0196	{'C': 0.001, 'gamma': 0.1}	0.5086	0.5043	0.5043
0.0351	0.0030	{'C': 0.001, 'gamma': 1}	0.5086	0.5043	0.5043
0.0306	0.0027	{'C': 0.001, 'gamma': 10}	0.5086	0.504	0.5043
0.0424	0.0022	'C': 1.0, 'gamma': 1.0}	0.8879	0.8956	0.9652
0.0385	0.0013	{'C': 1.0, 'gamma': 1.0}	0.8879	0.8956	0.9652
0.0144	0.0017	{'C': 1.0, 'gamma': 0.001}	0.7758	0.7217	0.7130
0.0800	0.0036	{'C': 1.0, 'gamma': 0.1}	0.8793	0.8956	0.9565
0.0402	0.0033	{'C': 0.001, 'gamma': 10.0}	0.5086	0.5043	0.5043

Table 4. Optimal features selected by iterative feature elimination method

No.	Feature Index	Feature Name
1	0	Age
2	2	Chest pain
3	4	Cholesterol
4	6	Exercise induced angina relative to rest
5	8	Exercise induced angina

4.4. Comparative study

The developed hybrid heart disease prediction model is compared with existing recently published heart disease prediction models implemented using supervised machine learning algorithm. In comparing the developed model with existing work, accuracy is used as performance measure. Table 5 summarizes the comparative study on the performance of proposed model and existing works. As demonstrated in Table 5, the proposed model outperforms compared to the existing work.

Table 5. Comparison of the performances of the proposed and existing work

Authors	Year	Algorithm	Accuracy in %
Divya Krishnani [2]	2019	RF, DT, KNN	92.89% with KNN
Assegie T.A [3]	2019	SVM	73.4%
Asraa Abdullah Hussein [4]	2019	K-means	84.74%
Stella Mary [5]	2019	NB, RF	86.81%
Wan Hajarul [6]	2018	DT and RF	82.99% with RF
Amin Ul Haq [8]	2018	SVM, DT, RF, NB, DT	86% with SVM
Kathleen H. Miaoa [11]	2018	Deep neural network	83.67%
Wiharto Wiharto [12]	2019	Ensemble classifier	88.33%
Noor Basha [18]	2019	KNN, NB, SVM, DT	85%, with KNN
Edsel Ing [19]	2019	SVM and LR	82.71% with LR
Márcio Dias [20]	2020	SVM	87.71%
Khaled Mohamad [21]	2020	SVM, NB	84.19% with SVM
Pooja Rani [22]	2021	NB, LR, NB, SVM, RF	84.79% with SVM
Suja Panicker [23]	2020	SVM	90%
G. Magesh [24]	2020	RF	89.30%
Ashir Javeed [25]	2020	Deep neural network	91.83%
G. Saranya [26]	2020	SVM	91.4%
Proposed approach	2021	Hybrid (SVM & RF)	98.3%

5. CONCLUSION

Automated intelligent approaches are crucial for timely prediction of heart disease. Incorrect and false negative outcome on heart disease prediction leads to risky decision during heart disease diagnosis. Thus, we have developed hybrid model with support vector machine and random forest using iterative feature elimination to obtain accurate prediction of heart disease at early stage. The developed model is tested on real world Indian heart disease dataset and result is compared with existing works. Comparative result shows that the proposed model outperforms compared to existing models. Overall, we have developed more efficient and accurate support vector machine and random forest base hybrid model for heart disease prediction with accuracy of 98.3%. The result evidently appears to prove that the developed model is helpful for better decision making for heart disease diagnosis. In the feature work, the authors plan to extend this study by experimentally evaluating the features that are contributing to positive or negative prediction results by applying model explanation approaches in order to make the model trusted and adopted by interpreting the predictive outcome and reasoning why the model have reached on a particular prediction.

REFERENCES

[1] K. Bhagchandani and D. P. Augustine, "IoT based heart monitoring and alerting system with cloud computing and managing the traffic for an ambulance in India," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 6, pp. 5068–5074, Dec. 2019, doi: 10.11591/ijece.v9i6.pp5068-5074.

[2] M. Z. Suboh *et al.*, "Portable heart valve disease screening device using electronic stethoscope," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 1, pp. 122–132, Jul. 2019, doi: 10.11591/ijeecs.v15.i1.pp122-132.

[3] W. B. A. Samad, M. A. Bin Othman, N. B. M. Safri, and M. A. B. A. Razak, "Portable cardiac self-stress test device development based on metabolic equivalent (MET)," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 3, pp. 1223–1231, Sep. 2019, doi: 10.11591/ijeecs.v15.i3.pp1223-1231.

[4] T. A. Assegie, "A support vector machine based heart disease prediction," *Journal of Software Engineering and Intelligent Systems*, vol. 4, no. 3, pp. 111–116, 2019.

[5] A. Anwar, R. Sigit, A. Basuki, and I. P. A. S. Gunawan, "Implementation of optical flow: good feature definition for tracking of heart cavity," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 2, pp. 1057–1065, May 2020, doi: 10.11591/ijeecs.v18.i2.pp1057-1065.

[6] A. A. Hussein, "Improve the performance of k-means by using genetic algorithm for classification heart attack," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 2, pp. 1256–1261, Apr. 2018, doi: 10.11591/ijece.v8i2.pp1256-1261.

[7] A. M. Yusof, N. A. M. Ghani, K. A. M. Ghani, and K. I. M. Ghani, "A predictive model for prediction of heart surgery procedure," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 3, pp. 1615–1620, Sep. 2019, doi: 10.11591/ijeecs.v15.i3.pp1615-1620.

[8] T. A. Assegie and P. S. Nair, "The performance of different machine learning models on diabetes prediction," *International Journal of Scientific and Technology Research*, vol. 9, no. 1, pp. 2491–2494, 2020.

[9] T. A. Assegie, S. J. Sushma, B. G. Bhavya, and S. Padmashree, "Correlation analysis for determining effective data in machine learning: detection of heart failure," *SN Computer Science*, vol. 2, no. 3, Apr. 2021, doi: 10.1007/s42979-021-00617-5.

[10] A. M. Khairuddin, K. N. F. Ku Azir, and P. Eh Kan, "A general framework for improving electrocardiography monitoring system with machine learning," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 1, pp. 261–268, Mar. 2019, doi: 10.11591/eei.v8i1.1400.

[11] J. P. George and S. M. Gaikwad, "Simulation modeling for heart attack patient by mapping cholesterol level," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 1, pp. 16–23, Apr. 2019, doi: 10.11591/ijeecs.v18.i1.pp16-23.

[12] Z. Rustam, A. Arfiani, and J. Pandelaki, "Cerebral infarction classification using multiple support vector machine with information gain feature selection," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 9, no. 4, pp. 1578–1584, Aug. 2020, doi: 10.11591/eei.v9i4.1997.

[13] F. S. Alotaibi, "Implementation of machine learning model to predict heart failure disease," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, pp. 261–268, 2019, doi: 10.14569/ijacsa.2019.0100637.

[14] S. Sumiati, H. Saragih, T. A. Rahman, and A. Triayudi, "Expert system for heart disease based on electrocardiogram data using certainty factor with multiple rule," *IAES International Journal of Artificial Intelligence (IJAI)*, vol. 10, no. 1, pp. 43–50, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp43-50.

[15] S. Abbas, "An innovative IoT service for medical diagnosis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, pp. 4918–4927, Oct. 2020, doi: 10.11591/ijece.v10i5.pp4918-4927.

[16] R. L. Figueroa, Q. Zeng-Treitler, S. Kandula, and L. H. Ngo, "Predicting sample size required for classification performance," *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, Feb. 2012, doi: 10.1186/1472-6947-12-8.

[17] M. A. Tamal, M. S. Islam, M. J. Ahmmed, M. A. Aziz, P. Miah, and K. M. Rezaul, "Heart disease prediction based on external factors: a machine learning approach," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 446–451, 2019, doi: 10.14569/ijacsa.2019.0101260.

[18] G. T. Reddy and N. Khare, "An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model," *Journal of Circuits, Systems and Computers*, vol. 26, no. 4, Apr. 2017, doi: 10.1142/S021812661750061X.

[19] Y. K. Singh, N. Sinha, and S. K. Singh, "Heart disease prediction system using random forest," in *Communications in Computer and Information Science*, vol. 721, Springer Singapore, 2017, pp. 613–623.

[20] N. Kumar, N. Narayan Das, D. Gupta, K. Gupta, and J. Bindra, "Efficient automated disease diagnosis using machine learning models," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–13, May 2021, doi: 10.1155/2021/9983652.

[21] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, R. Sun, and I. Garciá-Magarinõ, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mobile Information Systems*, vol. 2018, pp. 1–21, Dec. 2018, doi: 10.1155/2018/3860146.

[22] T. A. Assegie, R. L. Tulasi, and N. K. Kumar, "Breast cancer prediction model with decision tree and adaptive boosting," *IAES International Journal of Artificial Intelligence (IJAI)*, vol. 10, no. 1, pp. 184–190, Mar. 2021, doi: 10.11591/ijai.v10.i1.pp184-190.

[23] W. Wiharto, E. Suryani, and V. Cahyawati, "The methods of duo output neural network ensemble for prediction of coronary heart disease," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 7, no. 1, Mar. 2019, doi: 10.52549/ijeei.v7i1.458.

[24] G. Saranya and A. Pravin, "A comprehensive study on disease risk predictions in machine learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, pp. 4217–4225, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4217-4225.

[25] N. Basha, S. P. Ashok Kumar, C. Gopal Krishna, and P. Venkatesh, "Early detection of heart syndrome using machine learning technique," in *4th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques, ICEECCOT 2019*, Dec. 2019, pp. 387–391, doi: 10.1109/ICEECCOT46775.2019.9114651.

[26] P. Rani, R. Kumar, N. M. O. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *Journal of Reliable Intelligent Environments*, vol. 7, no. 3, pp. 263–275, Jan. 2021, doi: 10.1007/s40860-021-00133-6.

[27] T. R. Stella Mary and S. Sebastian, "Predicting heart ailment in patients with varying number of features using data mining techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2675–2681, Aug. 2019, doi: 10.11591/ijece.v9i4.pp2675-2681.