Hybrid optimal feature selection approach for internet of things based medical data analysis for prognosis

Felcia Bel, Sabeen
Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, India

Article Info

Article history:

Received Jul 8, 2023
Revised Oct 15, 2023
Accepted Nov 14, 2023

Keywords:

Feature selection
Internet of things
Pre-processing
Recursive feature elimination
machine learning

ABSTRACT

Healthcare is very important application domain in internet of things (IoT). The aim is to provide a novel combined feature selection (FS) methods like univariate (UV) with tree-based methods (TB), recursive feature elimination (RFE) with least absolute shrinkage selection operator (LASSO), mutual information (MI) with genetic algorithm (GA) and embedded methods (EM) with univariate has been applied to internet of medical things (IoMT)based heart disease dataset. The well-suited machine learning algorithms for IoT medical data are logistic regression (LR) and support vector machine (SVM). Each combined method has been applied to the machine learning algorithms to find the best classifier for prognosis. The various performance metrices has been calculated for all the combined feature selection methods for logistic regression and support vector machine and found that for precise classification could be done using recursive elimination feature selection method with LASSO applied to logistic regression achieved a better performance than all other combined methods with high accuracy, sensitivity and high area under curve. Decision has been taken by data analytics that RFE+LASSO using LR feature selection method will provide an overall better performance for IoT based medical heart disease dataset after comparing all other combined methods with LR and SVM classifiers.

Corresponding Author:

Sabeen
Department of Computer Science, Faculty of Science and Humanities
SRM Institute of Science and Technology
Kattankulathur, Chennai-603203, Tamilnadu, India
Email: sabeens@srmist.edu.in

1. INTRODUCTION

In humans, there are various chronic diseases like heart failures, kidney failures, cancer, etc. Among all these, this research mainly focused on iot based heart disease. To monitor heart and to take decisions in a fast and rapid manner, we have implemented internet of things (IoT) sensors for acquiring data. The patient's health is more important to survive. This IoT based monitoring of heart take a major role to sense human heart parameters in a better way for earlier prediction to get diagnosed. Acquiring data and analyzing data is the foremost and crucial role in machine learning. Hence, this research paper mainly focuses on the preprocessing stage of this prognosis process. Various preprocessing techniques are commonly used in IoT medical sensor datasets. They are data cleaning, data normalization, feature selection, feature extraction, Dimensionality reduction, data augmentation and data mixing and combinations. Among all these kinds of preprocessing techniques, IoT based heart disease medical dataset can be effectively pre-processed mainly using data cleaning, data normalization and feature selection processes. The IoT based heart disease dataset has been collected from University of California Irvine (UCI) machine learning repository. The tool used for implementing this dataset is python.

2. RELATED WORKS OR BACKGROUND

Li *et al*. [1] proposed a novel fast conditional mutual information feature selection algorithm which can select the suitable features which can improve the classification accuracy and reduce the computation time. They suggested that the above said novel method mainly works well with support vector classifier. Khan and Algarni, [2] mentioned that modified optimization algorithm has been presented in their work to increase the prognosis precision during classification. The proposed algorithm improves the feature search by Levy flight algorithm. This proposed algorithm is used to optimize the parameters of heart disease for better prediction. Ghosh *et al*. [3] proposed a feature selection technique with the combination of relief and least absolute shrinkage selection operator (LASSO) technique which could be used to select important features for classification using bagging and boosting techniques. Zhong *et al*. [4] proposes a new pre-processing methodology based on subspace similarity detection. This technology can handle similar kind of sensing data to provide precise accuracy for classification. Mittal *et al.* [5] have studied and found that these preprocessing techniques like imputation, smoothing, feature extraction and data reduction has been done in sequential order. Also, presented that horizontal data reduction is not suitable for sensor datasets and vertical data reduction method is more suitable for data preprocessing. Krishnamurthi *et al*. [6] data processing techniques such as data denoising, data outlier detection, missing data imputation and data aggregation has been considered for their study. Moreover, understood the necessity of data fusion and various data fusion methods such as direct fusion, associated feature extraction, and identity declaration data fusion. Cofre-Martel *et al.* [7] presented a preprocessing pipeline followed for big systems, their discussion is based on data selection and label generation. To create clean datasets, two case studies has been discussed. Jane and Arockiam [8] reviewed about IoT preprocessing techniques for sensor data and studied about cleaning, transformation, reduction and integration in a detailed manner. Fraser *et al.* [9] evaluated that the high risk of heart patients could be identified easily. Machine learning algorithms are easy to identify diseases when it has been applied on proper suitable datasets proposed by [10]. Sivakumar and Pramod [11] presented minimum of fourteen attributes are needed for better classification and suitable hybrid methods have been used to develop accurate models. Features included in various studies related to IoT based heart disease are given in [3], [12]–[15].

3. WORK FLOW AND PROPOSED METHODOLOGY

3.1. Workflow

UCI heart disease dataset can be pre-processed using feature selection techniques and the pre-processed data could be used to train the model using machine learning algorithms. The performance measures are evaluated to do better prognosis. The overall workflow of our research has been given in Figure 1.

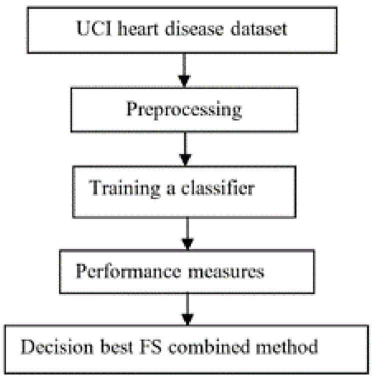

Figure 1. Overall workflow

3.2. The proposed methodology

The dataset has been taken from UCI heart disease repository. This dataset has been preprocessed by replacing missing values, normalization etc. In this preprocessing stage, the four combined feature selection methods are being used. Combined feature selection methods discussed here are univariate with tree-based method (UV+TB) [16], recursive feature elimination (RFE) with LASSO method (RFE+LASSO) [17], mutual information with genetic algorithm (MI+GA) [18] and embedded method with univariate (EM+UV) [19]. These methods are used to select the relevant features from original dataset. These selected features of various combined methods are used to train the machine learning model using logistic regression (LR) and support vector machine (SVM) which are more suitable for IoT based medical dataset. The various suitable machine learning (ML) performance measures have been calculated to identify the abnormal patients to treat them earlier from the risk of danger. The proposed method workflow has been shown in Figure 2.

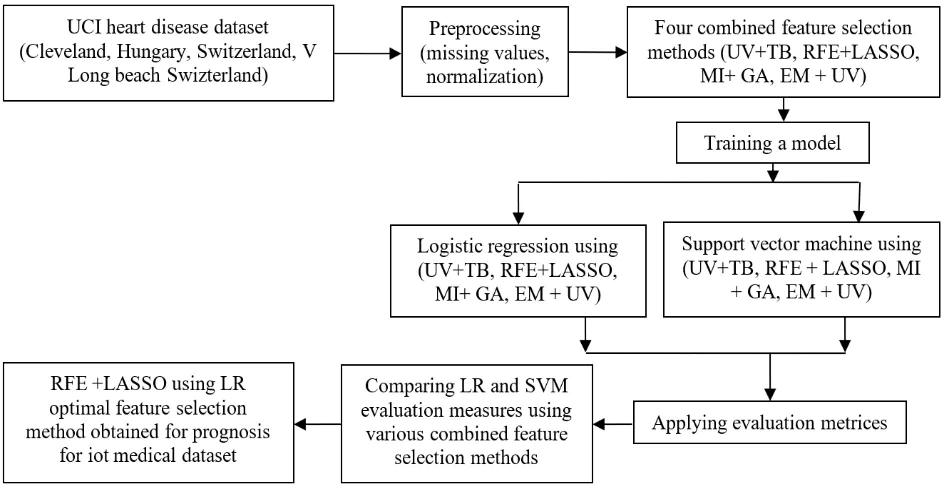

Figure 2. Workflow of the proposed model

3.2.1. Dataset

Data plays a major role in any research analysis. To understand data well, to take better decision, we have considered the possible application domain dataset from secondary source or from real time data. For our research work, the data has been gathered from UCI machine learning repository. The datasets used in our research are from cleveland, hungary, Switzerland, V long beach [20] and statlog heart disease dataset [21]. The samples collected are more than 1,190 cases from these datasets. 13 attributes with one target attribute have been selected for our research work.

3.2.2. Preprocessing of dataset

This preprocessing involves noticing and removing any data samples that are significantly different from the rest of the data to reduce errors while building a model. For example, removing data points that are outside the normal range of heart rate or blood pressure. Transforming the data to a common scale or range is very essential for comparisons and analysis. For example, normalizing the heart rate values to a range between 0 and 1. The datasets collected from secondary sources may contain missing values and noise removal [22]. The missing values can be examined by calculating the mean for the missing values and by arranging the values in increasing order. Then the middle value is calculated. After that the data is to be standardized using 0 and 1.

$$\mathrm{Pm} = \beta 0 + \beta \mathrm{k}\, \mathrm{Q} + \in \mathrm{k}, \text{for } \mathrm{k} = 1, 2 \ldots \mathrm{n} \quad (1)$$

The above equation can be used by the random samples present in a dataset and the regression model is given as,

$$\mathrm{Pm} = \beta 0 + \beta \mathrm{k}\, \mathrm{Q} + * \in \mathrm{k} \quad (2)$$

The equal variance and error are expressed as ∈k. The least squares values are represented by β0 and βk respectively. The sample data and the average values are calculated using the deviation. In (3), the average value is estimated as follows [22].

$$\mu = \frac{\sum_{i=1}^{n} Q_i}{N} \quad (3)$$

Where N represents the frequency of data and Qi is the input data. The formula for data normalization is given in (4) and (5),

$$n_r = \frac{\epsilon_i^*}{\sigma_i^*} \quad (4)$$

$$n_r = \frac{Q_i - \mu_i^*}{\sigma_i^*} \quad (5)$$

where $\in_i^*$ - Residual value, σ_i^* - Variance.

3.2.3. The proposed Feature selection combined methods

This proposed technique involves selecting a subset of relevant features from the original data. We can achieve more robust and accurate results after using many techniques together while using various combined feature selection methods. It helps to remove redundant, irrelevant, or noisy features that may not contribute to the accuracy of the model [23].

a. Univariate with tree-based methods (UV+TBM)

This method comprises of selecting the features having highest correlation or statistical importance to target variable. Univariate feature selection method like Chi-squared test or analysis of variance (ANOVA) has been used to find the most relevant features based on statistical test. After performing the above-mentioned test, the tree-based methods like random forest or gradient boosting are used to find the selected features importance. This type of combined feature selection method leverages the statistical importance effectively,

b. Recursive feature elimination with LASSO (RFE+LASSO)

This technique involves iteratively removing the least important features from the dataset until the desired number of features has been obtained. Using a machine learning model, rank the importance of the features and removing the least important feature at each iteration. After obtaining a reduced feature set using the previous step, apply L1 regularization (Lasso) technique to further shrink the less important feature coefficients to zero. This method of combination can provide a sparse set of relevant features while considering both the model's performance and feature importance.

c. Mutual Information with genetic algorithms (MI+GA)

Mutual Information is a measure of the dependency between variables. It quantifies the amount of information that can be obtained about one variable by observing another variable. Calculate mutual information between each feature and the target variable to identify relevant features. Mutual Information can be used as a filter method to rank features based on their relevance to the target variable. GA is a metaheuristic optimization algorithm that uses an evolutionary approach to select features based on their performance, to iteratively refine the feature set. This combinational method allows for capturing both linear and non-linear relationships among features for overall performance of the selected features.

d. Embedded method with univariate (EM + UV)

Embedded feature selection methods, such as L1 regularization with logistic regression or decision trees, which automatically evaluate feature importance during the model training process. Then, combine it with a univariate feature selection approach to further refine the feature set based on statistical tests. This combination takes advantage of the intrinsic feature selection capabilities of the model and the broader statistical analysis**.**

3.2.4. Machine learning classifiers

a. Logistic regression

Logistic regression [24] is a binary classifier and it takes one or more features of x and predicts the response y. The mapping function in logistic regression method is known as sigmoidal function. A sigmoidal function is of ‘S’ shaped function that ranges the values between o and 1. This is known as logit function. This can be represented as follows.

$$\text{logit(x)} = \frac{1}{1 + e^{-x}} \quad (6)$$

Where x is an independent variable and e is a Euler number. The purpose of logit function is to map any real number to zero or 1. The regression coefficients $a_{0,}$ a_1 can be learned and the predictor predicts p(x) directly using the threshold function as,

$$\text{y} = \begin{Bmatrix} 1 \ if \ p(x) > 0.5 \\ 0 \ otccherwise \end{Bmatrix} \quad (7)$$

To obtain the parameters, we need to determine the relationship between the dependent and independent variable. In this model the maximum likelihood function (MLE) using the training data. The idea is to learn the values of parameters of logistic model (a's) by minimizing the error in the probability predicted by the model.

b. Support vector machine

Support vector machine [25] which is very much used when dealing with complex decision boundaries and for high dimensional data. The margin has been maximized to provide an optimal hyperparameter by separating the datapoints of different classes. The hyperplane is defined by a set of weights (coefficients) and bias term. Support vectors are those datapoints that fall on the boundary lines. All training samples that fall on the boundary lines are called support vectors.

4. RESULTS AND DISCUSSION

In this section, various performance measures, comparison of combined feature selection methods with LR and SVM and about the best combined feature selection methods with LR and SVM are projected. The feature selection based combined methods are implemented in classifiers to provide an effective combined method after performing various performance measures. The suitable performance measures used for this IoT based medical data are given as below,

4.1. Performance measures

The effectiveness and quality of model's predictions could be evaluated using performance measures in machine learning. The various performance measures evaluated for this research are accuracy, precision, recall, f-measure, sensitivity, specificity and area under curve [26] which has been given using the formula from (8) to (14). From these measures, we can understand which feature selection method is best suited for predicting iot based medical data. These measures are explained as below.

- Accuracy: The proportion of correctly classified instances.
- Precision: The ratio of true positive (TP) predictions to the total predicted positives. It measures the model's ability to avoid false positives.
- Recall (Sensitivity): The ratio of true positive predictions to the total actual positives. It measures the model's ability to identify positive instances correctly.
- f -measure: The harmonic means of precision and recall, providing a balanced measure of both metrics.
- Sensitivity: Sensitivity measures the capability of a model to correctly classify patients with heart disease or samples from the total number of actual positive instances in the dataset. It calculates the proportion of true positive predictions (correctly predicted positive samples) out of all the actual positive samples.
- Specificity: Specificity is also referred to as the true negative rate (TNR). It calculates the proportion of true negative predictions (correctly predicted negative samples) out of all the actual negative samples.
- Area under the receiver operating characteristic curve (AUC-ROC): Measures the model's ability to discriminate between classes across different classification thresholds.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{9}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{10}$$

$$\text{F measure} = 2 * \frac{Precision * Recall}{Precision+Recall} \tag{11}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{12}$$

$$\text{Specificity} = \frac{FP}{FP+TN} \tag{13}$$

$$\text{Area under curve (AUC)} = \Sigma[(\mathrm{TPR(i)} + \mathrm{TPR(i+1)})/2] * (\mathrm{FPR(i+1)} - \mathrm{FPR(i)}) \tag{14}$$

4.2. Comparison of combined feature selection methods with LR and SVM

The combined feature selection methods explained above has been undergone various performance measures applied to LR and SVM which has been given in Table 1. Three suitable combined feature selection methods have been selected which has obtained high accuracy, better sensitivity and good in auc. The overall combined feature selection methods with its calculated performance measures representation have been given in Figure 3.

4.3. Best combined feature selection methods with LR and SVM

The three suitable combined feature selection methods have been taken from Table 1 and compared among them. The comparison of three combined feature selection methods shows that RFE+LASSO with LR obtained with high accuracy value, high sensitivity and with very high auc value which is greater than 0.5 and less than 1 has been given in Table 2. and the representation also been given in Figure 4. High sensitivity and high auc gives the true positive rate (TPR) which is useful for predicting the patients with heart disease correctly to get treatment before undergoing the conditions worse.

Table 1. Overall combined feature selection methods and its performance measures

Sl. No.	Feature Selection Methods	Accuracy	Precision	Recall	F-Measure	Sensitivity	Specificity	AUC
1	UV+ TBM-LR	0.85	0.82	0.88	0.85	0.88	0.82	0.91
2	RFE+LASSO-LR	0.87	0.83	0.85	0.84	0.88	0.82	0.92
3	MI + GA-LR	0.85	0.78	0.85	0.81	0.85	0.76	0.81
4	EM+univariate- LR	0.82	0.85	0.80	0.80	0.75	0.85	0.80
5	UV+TBM-SVM	0.84	0.83	0.75	0.79	0.75	0.86	0.87
6	RFE+LASSO-SVM	0.83	0.82	0.70	0.75	0.88	0.80	0.89
7	MI+GA-SVM	0.85	0.85	0.80	0.82	0.80	0.90	0.92
8	EM+univariate-SVM	0.85	0.82	0.70	0.75	0.88	0.80	0.89

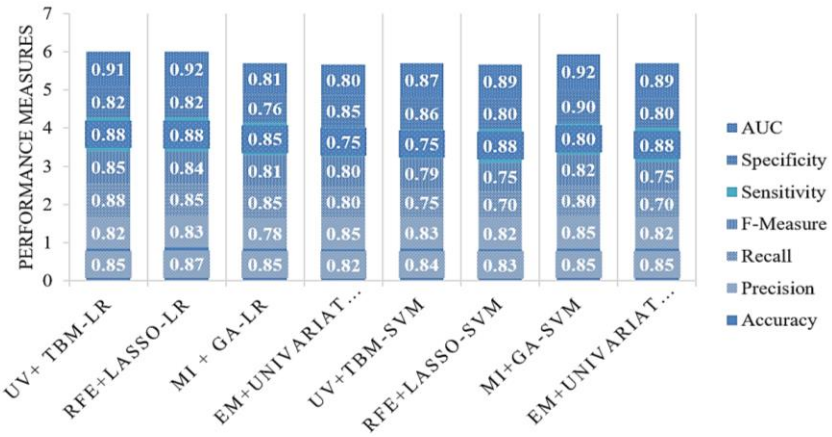

Figure 3. Overall combined feature selection methods and its performance measures

Table 2. Best combined feature selection method and its performance measures

Sl. No.	Feature Selection Methods	Accuracy	Precision	Recall	F-Measure	Sensitivity	Specificity	AUC
1	UV+ TBM-LR	0.85	0.82	0.88	0.85	0.88	0.82	0.91
2	RFE+LASSO-LR	0.87	0.83	0.85	0.84	0.88	0.82	0.92
3	MI+GA-SVM	0.85	0.85	0.80	0.82	0.80	0.90	0.92

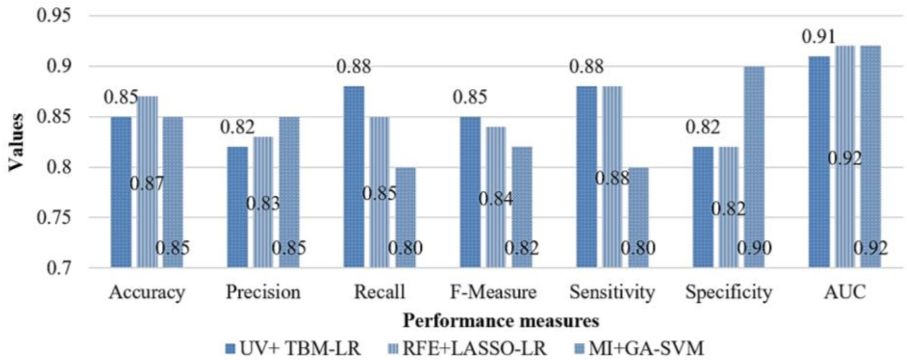

Figure 4. Best combined feature selection method and its performance measures

5. CONCLUSION

Human lives are more valuable and must be saved by providing various technological solutions in health sector. To acquire better decision among IoT based heart data, and to understand the patient condition more precisely, the machine learning based classification provide a better solution to health sector. Various preprocessing techniques like missing values replacement, data normalization has been applied to the above UCI heart dataset. Feature selection techniques will provide better results in classifier performance. In this research, the four combined feature selection methods like univariate with tree based, recursive feature elimination with LASSO, mutual information with genetic algorithm, embedded with univariate methods are implemented. These FS methods are implemented in two well-known machine learning classifiers like logistic regression and support vector machine. The classifier performance has been precisely calculated using various performance measures. Implementing all the combined feature elimination with LASSO method using LR has obtained high sensitivity and high AUC with 0.91 which can perform well for this UCI heart dataset. We conclude that, this FS method can correctly indicate the positive patients and make them undergo treatment earlier before going to the dangerous situation and to mitigate the disease to provide the life longer for the patients. In future as for as the need of the patients, the feature selection techniques could be enhanced to develop an opt model for prediction system in healthcare.

REFERENCES

[1] 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, doi: 10.1109/ACCESS.2020.3001149.

[2] M. A. Khan and F. Algarni, "A healthcare monitoring system for the diagnosis of heart disease in the IoMT cloud environment using MSSO-ANFIS," *IEEE Access*, vol. 8, pp. 122259–122269, 2020, doi: 10.1109/ACCESS.2020.3006424.

[3] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.

[4] Y. Zhong, S. Fong, S. Hu, R. Wong, and W. Lin, "A novel sensor data pre-processing methodology for the internet of things using anomaly detection and transfer-by-subspace-similarity transformation," *Sensors (Switzerland)*, vol. 19, no. 20, 2019, doi: 10.3390/s19204536.

[5] S. Mittal, K. Gopal, and S. L. Maskara, "Preprocessing methods for context extraction from multivariate wireless sensors data - An evaluation," *2013 Annual IEEE India Conference, INDICON 2013*, 2013, doi: 10.1109/INDCON.2013.6725984.

[6] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of iot sensor data processing, fusion, and analysis techniques," *Sensors (Switzerland)*, vol. 20, no. 21, pp. 1–23, 2020, doi: 10.3390/s20216076.

[7] S. Cofre-Martel, E. L. Droguett, and M. Modarres, "Big machinery data preprocessing methodology for data-driven models in prognostics and health management," *Sensors*, vol. 21, no. 20, 2021, doi: 10.3390/s21206841.

[8] V. Jane and L. Arockiam, "Survey on IoT data preprocessing," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 9, pp. 238–244, 2021, doi: 10.17762/turcomat.v12i9.3001.

[9] E. Anwaitu Fraser, Okonkwo, and Obikwelu R, "Artificial neural networks for medical diagnosis: a review of recent trends," *International Journal of Computer Science & Engineering Survey*, vol. 11, no. 3, pp. 1–11, 2020, doi: 10.5121/ijcses.2020.11301.

[10] R. Rabiei, S. M. Ayyoubzadeh, S. Sohrabei, M. Esmaeili, and A. Atashi, "Prediction of breast cancer using machine learning approaches," *Journal of Biomedical Physics and Engineering*, vol. 12, no. 3, pp. 297–308, 2022, doi: 10.31661/jbpe.v0i0.2109-1403.

[11] S. Sivakumar and T. C. Pramod, "Comprehensive analysis of heart disease prediction: machine learning approach," *2022 IEEE 3rd Global Conference for Advancement in Technology, GCAT 2022*, 2022, doi: 10.1109/GCAT55367.2022.9972035.

[12] F. Andreotti *et al.*, "Prediction of the onset of cardiovascular diseases from electronic health records using multi-task gated recurrent units," 2020, [Online]. Available: http://arxiv.org/abs/2007.08491.

[13] Wiharto, H. Kusnanto, and H. Herianto, "Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis," *International Journal of Electrical and Computer Engineering*, vol. 7, no. 2, pp. 1023–1031, 2017, doi: 10.11591/ijece.v7i2.pp1023-1031.

[14] A. K. Paul, P. C. Shill, M. R. I. Rabin, and M. A. H. Akhand, "Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease," *2016 5th International Conference on Informatics, Electronics and Vision, ICIEV 2016*, pp. 145–150, 2016, doi: 10.1109/ICIEV.2016.7759984.

[15] X. Liu *et al.*, "A hybrid classification system for heart disease diagnosis based on the RFRS method," *Computational and Mathematical Methods in Medicine*, vol. 2017, 2017, doi: 10.1155/2017/8272091.

[16] D. P. M. Abellana and D. M. Lao, "A new univariate feature selection algorithm based on the best–worst multi-attribute decision-making method," *Decision Analytics Journal*, vol. 7, 2023, doi: 10.1016/j.dajour.2023.100240.

[17] M. Hamada, J. J. Tanimu, M. Hassan, H. A. Kakudi, and P. Robert, "Evaluation of recursive feature elimination and LASSO regularization-based optimized feature selection approaches for cervical cancer prediction," *Proceedings - 2021 IEEE 14th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip, MCSoC 2021*, pp. 333–339, 2021, doi: 10.1109/MCSoC51149.2021.00056.

[18] Y.-D. Lan, "A hybrid feature selection based on mutual information and genetic algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, no. 1, p. 214, Jul. 2017, doi: 10.11591/ijeecs.v7.i1.pp214-225.

[19] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, pp. 1200–1205, 2015, doi: 10.1109/MIPRO.2015.7160458.

[20] S. Mohamed, M. Malhat, and G. Elhady, "Prediction of cardiovascular disease using machine learning techniques," *IJCI. International Journal of Computers and Information*, vol. 0, no. 0, pp. 0–0, 2022, doi: 10.21608/ijci.2022.129472.1071.

[21] R. Sarmento, "Statlog (Heart). UCI machine learning repository," 2019, doi: 10.24432/C57303.

[22] A. Rairikar, V. Kulkarni, V. Sabale, H. Kale, and A. Lamgunde, "Heart disease prediction using data mining techniques," *Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017*, vol. 2018-January, pp. 1–8, 2018, doi: 10.1109/I2C2.2017.8321771.

[23] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 4, pp. 1060–1073, 2022, doi: 10.1016/j.jksuci.2019.06.012.
[24] K. Sarkar, H. Midi, and S. Rana, "Detection of outliers and influential observations in binary Logistic regression: An empirical study," *Journal of Applied Sciences*, vol. 11, no. 1, pp. 26–35, 2011, doi: 10.3923/jas.2011.26.35.
[25] X.-D. Zhang, "Support vector machines (SVM)," *Gesture*, vol. 23, no. 6, pp. 349–361, 2001.
[26] F. Bel and S. Selvaraj, "Efficient machine learning classifier to detect and monitor COVID-19 cases based on internet of things framework," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 4, pp. 4605–4611, 2023, doi: 10.11591/ijece.v13i4.pp4605-4611.

BIOGRAPHIES OF AUTHORS

Felcia Bel received a B.Sc. degree in physics from Manonmanium Sundaranar University and M.C.A. degree in computer applications from Bharadhidasan University, in 2002 and 2005, respectively, and M.Phil. degree in computer science in 2008. She had been working as an assistant professor in various colleges from 2008 to 2020. At present, she is pursuing research at SRM Institute of Science and Technology, Chennai 603203, Tamil Nadu, India. Her research interests are the Internet of things, data mining, and machine learning. She can be contacted at fh4219@srmist.edu.in.

Sabeen is working at the Department of Computer Science, SRM Institute of Science and Technology. He received his doctoral degree from Anna University in 2012. He received his postgraduate degree M.C.A in the year 2002 and M.Tech. (CSE) in the year 2015. He has 20 years of teaching experience in various engineering colleges: Mohamad Sathak Engineering College, Noorul Islam College of Engineering, Jaya Engineering College and Sidharth Institute of Engineering and Technology, India. He is a life member of ISTE. His area of interests includes data mining, IoT, machine learning. He has published many papers in international journals. He has guided more than 100 projects for the post graduate students. He can be contacted at sabeens@srmist.edu.in.