ISSN: 2093-5374 / E-ISSN: 2093-5382

Improved Interpretation Model for Heart Disease Diagnosis Using Artificial Neural Networks

Dr.F. Syed Anwar Hussainy[1], J. Jayapradha[2], Dr. Mardeni Roslee[3*], Dr.T. Senthil Kumar[4], Dr. Chilakala Sudhamani[5], Azmi Ismail[6], Anwar Faizd Osman[7], Dr. Fatimah Zaharah Ali[8], and Idris Olalekan Adeoye[9]

[1]Department of Computer Science and Engineering (AI & ML), Faculty of Engineering and Technology, Jain (Deemed to be University), Bangalore, India. syedanwar.f@jainuniversity.ac.in, https://orcid.org/0000-0003-2837-3856

[2]Assistant Professor, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Tamil Nadu, India. jayapraj@srmist.edu.in, https://orcid.org/0000-0002-2548-9135

[3*]Professor, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. mardeni.roslee@mmu.edu.my, https://orcid.org/0000-0001-8250-4031

[4]Associate Professor, SRM Institute of Science and Technology, SRM Nagar, Kattankulathur, Tamil Nadu, India; Postdoctoral Research Fellow, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. senthilt2@srmist.edu.in, https://orcid.org/0000-0002-2200-3339

[5]Postdoctoral Research Fellow, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. sudhamanich@gmail.com, https://orcid.org/0000-0002-8823-9053

[6]Centre of Excellence for Intelligent Network, Telekom Malaysia Research & Development Cyberjaya, Malaysia. azmi@tmrnd.com.my, https://orcid.org/0000-0001-5792-9601

[7]Spectre Solution Sdn Bhd, Malaysia, anwarfaizd@spectresolution.com, https://orcid.org/0000-0001-9563-9138

[8]Universiti Teknologi MARA, Faculty of Electrical Engineering, Shah Alam, Selangor, Malaysia. fatimah_zaharah@uitm.edu.my, https://orcid.org/0000-0002-5467-4049

[9]Centre for Wireless Technology, Faculty of AI & Engineering, Multimedia University Cyberjaya, Malaysia. idreezleks@gmail.com, https://orcid.org/0009-0008-3450-7519

Received: February 17, 2025; Revised: April 04, 2025; Accepted: May 09, 2025; Published: June 30, 2025

Abstract

In the present scenario, health care is a predestined process to be considered in human life. While heart diseases are concerned, cardiovascular disease (CVD) is a wide class of diseases that damages blood vessels and the heart. In the medical field, huge health data are available to study and process; hence, machine learning methods are required for appropriate decision-making, specifically in terms of heart disease prediction and diagnosis. For enhancing the appropriation rate of decision-making in CVD diagnosis, this paper proposes an Improved Interpretation Model for Heart Disease Diagnosis (IIM-HDD) using Artificial neural networks. The model incorporates data acquisition, pre-processing, feature selection, training, and testing for diagnosis. For training and validation, the

*Corresponding author: Professor, Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia.

data from benchmark datasets are combined and used. Moreover, feature selection is computed with a relief-based selection process. The ANN model is trained to produce the output, corresponding to their input features. The results computations are processed with metrics that include classification, accuracy, precision rate, error rate, specificity, sensitivity, F1 score, and other comparisons are also provided for proving the proposed model. The results show that the work outperforms the other compared models in respective metrics.

Keywords: Artificial Neural Network, Cardiovascular Disease, Classification Accuracy, Disease Diagnosis, Feature Selection.

1 Introduction

Cardiovascular Disease (CVD) is the term denoting the state that affects the blood vessels of the heart. This can also refer to artery damage in other organs of the human body, such as the eyes, brain, and kidneys (Baad et al., 2019). Heart diseases are the major reason for human death in the recent scenario in many countries around the world. The disease can be factually prevented by a healthy lifestyle. In clinical definitions, there are four types of cardiovascular diseases, as given in Figure 1. The first type is coronary heart disease caused by blood clots in the heart muscle, which may lead to heart failure. The second type causes blood clots in the brain, termed strokes and Transient Ischaemic Attack (TIA). Peripheral Arterial Disease is the next category; it happens due to blood blockage in limbs. The final category is the damage caused at the aorta, which is the largest blood vessel, and the disease is termed as, which may cause immediate death. Considering the severity of CVD, the disease should be diagnosed as early as possible, and the recovery process should be initiated to save the patient's life.

Because of various constraints in the manual diagnosis of heart diseases, medical practitioners and researchers have moved to advanced diagnostic methods, such as Data Mining, Artificial Intelligence, Machine Learning and Deep Learning methods (Alic et al., 2017; Mezzatesta et al., 2019; Latha & Chandran, 2025). The methodologies are expected to provide better decision-making in disease diagnosis from the massive data collected in healthcare applications and services (Al-rammahi & Alshimaysawe, 2025; Sa, 2013). Moreover, the disease can be appropriately diagnosed using a sequence of lab examinations and image processing. Nevertheless, the main part of detection is based on the family and the medical history of the patients. From the statistical data, the determinations and predictions are coordinated with the disease presence based on the observations. Further, the automated learning procedure can facilitate doctors to make conversant decisions (Morris & McNicholas, 2016; Elyasigomari et al., 2015).

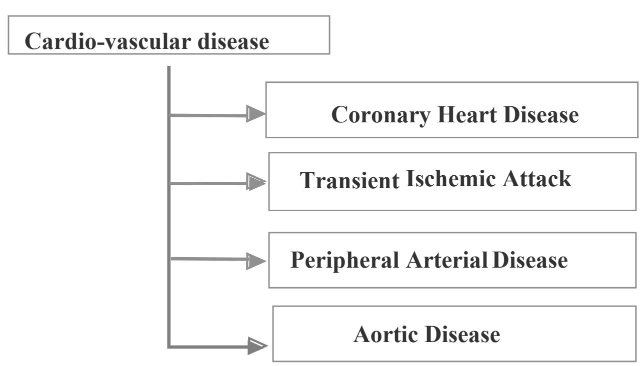

Figure 1: CVD Types

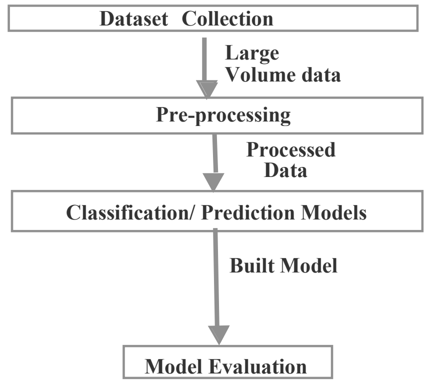

Figure 2: Workflow of Disease Prediction Model

In order to design a machine learning model for predicting heart disease, benchmark patient datasets are used. For doing so, this paper uses the dataset from the UCI Machine Learning Repository, denoted as the Cleveland Heart Disease dataset. An automated disease prediction model can create a unit platform from which the processed data can be obtained and appropriate treatment and care can be given to the patients. Figure 2 presents the general workflow of the detection model.

The results from the disease prediction model should be more precise to avoid the risks. With that note, this work develops an Improved Interpretation Model for Heart Disease Diagnosis (IIM-HDD). The classification accuracy of the proposed model is enhanced by incorporating ANN-based training. Further, the contributions of the proposed model are listed as follows:

- The medical data is acquired from three datasets for training the model with a reliable dataset.
- At preprocessing, duplicates are removed from the obtained data.
- Feature selection is done with Relief-based Feature Selection (RFS) to improve result accuracy.
- ANN is used for further model training and testing
- The results are analyzed based on various evaluation metrics, and comparisons are provided for justifying model efficacy in terms of predicting NORMAL and ABNORMAL data.

The remained of this paper is framed as follows: Section 2 describes related works that have been done in disease prediction using machine learning methods and artificial intelligence (Armas et al., 2025). Section 3 explains the design and working procedure of the proposed model. The results and discussions are portrayed in the Section 4. Lastly, a brief conclusion and pointers for future enhancement are presented in Section 5.

2 Related Works

Myriad learning methods have been used in medical data analysis and processing to leverage the health state of humans. Some works of health data processing, involve processing diabetes diagnosis using fuzzy-based soft computing models (El-Sappagh et al., 2018; El-Sappagh et al., 2018). There are various other applications that concentrate on heart disease prediction using machine learning and deep learning methods, due to their increasing accuracy and decision-making (Mishra & Tarar, 2020; Khan et al., 2021). The significance of this research is for the prospects to design and select models with more precision and efficacy (Abdeldjouad et al., 2020).

Hybrid learning methods that combine different machine learning models with significant features are an efficient model for disease diagnosis (Tarawneh & Embarak, 2019). In (Latha & Jeeva, 2019), many benchmark datasets are applied and studied. Moreover, the model used ensemble classification techniques to attain better prediction accuracy. The paper stated that the bagging and boosting model provided weak classification results while efficient performances were attained with the Naive Bayes (NB) model, the Multi-Layer Perceptron (MLP) and the Random Forest (RF).

In (Kumar & Sikamani, 2020), different machine learning models are used for diagnosing heart disease. Machine learning methods and traditional methods like Support Vector Machine (SVM) RF and other models were evaluated with the UCI-based machine repository (Javid et al., 2020). The model stated the voting-based classification model improves the result accuracy more than the traditional models. The authors (Ashraf et al., 2020) analyzed both the learning techniques and ensemble classifiers for medical data processing. Amongst, the J48 classification model attained 70.77%. A Multi-task Recurrent Neural Network (MT) has been developed by (Andreotti et al., 2020) for CVD. In a different manner, the Mean Fisher Score Feature Selection Algorithm (MFSFSA) for selecting appropriate features for efficient classification using the Support Vector Machine (SVM) (Nandhinieswari & Indumathi, 2024; Saqlain et al., 2019). In (Mienye et al., 2020), a heart disease prediction model that uses the mean splitting method has been proposed. Additionally, classification and regression tree methods were used for random partitioning of the dataset. The work done on (Roy et al., 2021; Khan et al., 2021) elevates that the availed data has to be securely loaded to the server environment at a higher potential. A two-tier ensemble classification model has been proposed for coronary disease detection in (Tama et al., 2020). The model used three different ensemble techniques, such as,

i. Random Forest (RF)
ii. Gradient Boosting (GB)
iii. Extreme GB

Further, the model computed 98%, 97% and 99% accurate results in disease prediction, respectively. In (Mohan et al., 2019), a new disease diagnosis model has been designed using ANN-based classifications, considering 13 different features (Arora, 2024). An Improved Random Survival Forest (iRSF) has been developed in (Miao et al., 2018), for detecting heart failure mortality. The model differentiated between survivors and non-survivors, using a novel split rule and stop process. The data mining model for detecting CVD has been used in (Raju et al., 2018). The results of algorithms such as Decision Tree (DT), Naive Bayes (NB), K-Nearest Neighbour (KNN) and Support Vector Machine (SVM) and concluded that SVM provided better results in case of accuracy.

The work provided in (Chicco & Jurman, 2020) discusses several machine-learning mechanisms in heart disease prediction (MHM et al., 2025). The features correlated to the important factors are ranked and compared with the traditional bio-evaluations and machine learning models. Congestive Heart Failure (CHF) has been predicted by measuring the Heart Rate Variability (HRV) with the help of Deep Neural Networks (DNN) to provide the gap in corresponding features. The work (Gupta et al., 2019) used Factor Analysis of Mixed Data (FAMD) and Random Forest-based Machine Learning for developing the model. The work stated that the model produced 93% of classification accuracy in disease diagnosis.

In (Balakrishnan et al., 2021), the Gradient Boosting Algorithm has been used. A hybrid model has been designed, and various other methods of CVD diagnosis have been discussed in (Sadeque et al., 2019; Alizadehsani et al., 2013). Further, an experimental evaluation was carried out in (Tan et al., 2009) with 303 datasets obtained from the Cleveland Heart Disease Dataset using a hybrid model with accuracy and precision analysis. Many methods have been used to diagnose CVD in patients. The data pre-processing is done to handle the missing data and data duplications. Further, features are selected using the Relief based method and ANN is used for classification. The evaluation process majorly involves in calculating the detection accuracy and error values for computing the best features.

3 Proposed Model

The proposed model develops an Improved Interpretation Model for Heart Disease Diagnosis (IIM-HDD**)** for accurate CVD diagnosis, using the machine learning algorithm. The training process is made with the data obtained from the UCI machine learning repository, described in (Section 4). Attributes, descriptions and ranges for the obtained data are provided in Table 1. The obtained medical data are divided into two, for training and testing. In the phase, the three layers are provided as follows:

i. Observation Layer:

The data are acquired from the dataset, based on the described features and those data are further submitted to the object layer for processing.

ii. Object Layer:

The layer is responsible for data preprocessing including noise removal, duplication removal and processing. The data are further submitted for feature selection and classification at the next layer

iii. Application Layer:

In this layer, feature extraction and selection operations are processed and further, given for classification, using the ANN model.

In the second phase, the testing process is carried out with factors such as detection accuracy, precision rate, sensitivity, specificity and error rates. The testing process is determined for analyzing the presence and the absence of cardiovascular disease from the obtained patient data. The workflow of the proposed model is given in Figure 3.

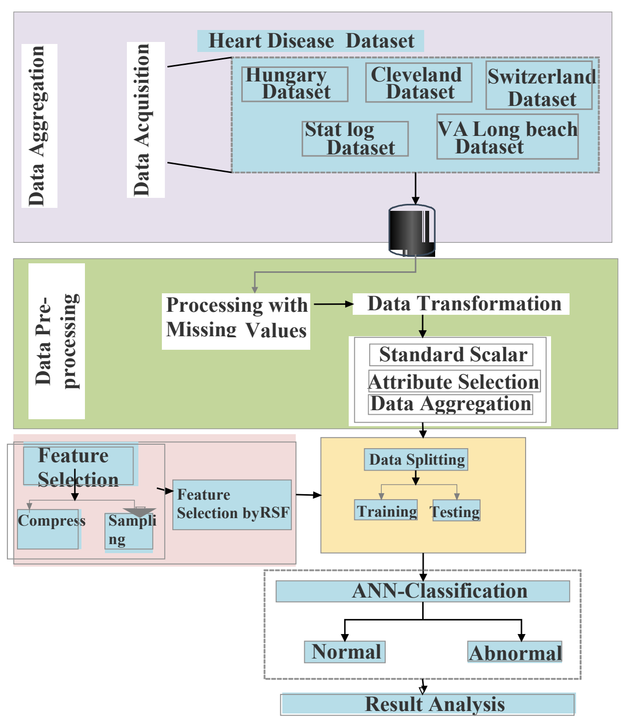

Figure 3: Workflow of Proposed Model

Table 1: Attributes, Descriptions and Ranges

S.No	Attribute	Description	Range	Data Type
1	Age	Patient Age in Years	30-80	Integer
2	Gender	Binary Value, 1-Female, 0-Male	0-1	Integer
3	Chest Pain	0-Non-specific chest pain 1-Atypical chest pain 2-Typical angina	0-2	Integer
4	Height	Patient's Height in cms	NA	Integer
5	Old Peak	Depression Induced by Exercise relative to rest	1-3	Real
6	Weight	Patient's Weight in Kgs	NA	Integer
7	Blood Pressure	Systolic Blood Pressure	100-170	Integer
8	Serum Cholesterol	Cholesterol Level in the blood	40-130	Integer
9	Sugar (fasting)	Blood Sugar level	70-99	Integer
10	ECG (resting)	Electro Cardiography level	60-100	Integer
11	ST depression	ST Segment depression ECG	1-2	Integer
12	Induced Angina Status	Blood Flow	0-2	Integer
13	Slope	Slope of the Peak Exercise ST Segment	1,2, 3	Integer
14	Smoke	Habit of Smoking (Yes/No)	0-1	Integer
15	Alcohol	Habit of Drinking (Yes/No)	0-1	Integer
16	ca	Number of major vessels coloured by fluoroscopy	0-3	Integer
17	thal	Defect Type	3,6,7	Integer
18	CVD Disease	Case History (Presence or Absence)	0-1	Integer

A. Data Pre-Processing

From the data obtained, there are 18 features and one target class rate as its label. The samples included the missing values, which are to be deleted for further evaluation. As machine learning models are used for processing, the feature reduction process is not incorporated. Instead, the proposed model used feature selection that is used for effectively processing with the classification model. It is significant for eliminating outliers and extreme rates to validate the robustness of the developed machine learning model. Moreover, the statistical dispersion of dataset features is determined using the Inter-Quartile Value (IQV) method. Further, there is patient data with some outliers in the 'thal' feature with the value of 9.5, which diverges extremely from the remaining feature values in the dataset. Then, the instance is removed for stands avoiding, producing efficient mean and standard deviation rates. For the extreme rate analysis, it is found that there are 43 patient data samples with extreme rates, comprising 14.5% of sample records. Nevertheless, to mitigate the issue of bias in the process of model development, the impact of removing the extreme rates is to be noted. From the considered records, the 43 instances are categorized under two classes: 23 under the presence of disease and 20 under the absence of disease. The balance between two categories of samples provides a sign that it is effective to have class bias; hence, those records are removed. The feature sets are to be frequently tested to validate that it is homogeneous.

The ranges of the heart feature dataset are noticed and will impact the design of the machine learning algorithm. For instance, the maximal rate for age is taken here as 80 and cholesterol level as 564, respectively. The effectual model is to standardize the numerical data attributes in the dataset with a '0' mean and '1' variance. After performing the data processing steps, it is significant to process feature selection to enhance the model performance under examination.

B. Feature Selection

After removing redundancy and irrelevant features, the feature selection is processed. Careful and serious computations are taken into account for employing the feature selection process on the patient data set with cross-validation to attain a higher rate of accuracy. Some general issues occurred when the entire dataset was used for processing. Hence, in this work, the features are selected based on the data training. Further, each developed classification model is learned on the training dataset, with cross-validation in consequence, followed by model analysis with the test data.

In a machine learning-based classification model, feature selection methods are significant, in which the optimal features are selected for processing. This work uses Relief-based Feature Selection (RFS) and the implementation of the proposed model with the features from the dataset is presented in Figure 4.

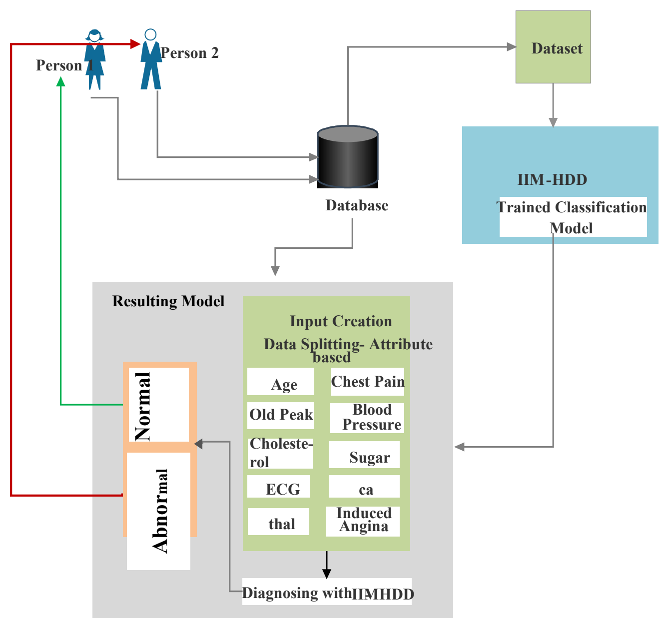

Figure 4: Model Implementation for Disease Diagnosis

RFS is a weight-based feature selection from the obtained model. The major objective is to derive significant features that have larger weights, and the lowest weights are ordered last. The model uses a similar algorithm as KNN for determining weights. In the algorithm, it is considered as the randomly selected instance, and the model finds two nearest neighbors. The nearest neighbor includes the closest hit rate and an opposite hit. The computation for the feature is given, based on, and. When there is a huge difference, the weight is considered to be low. On the other hand, the feature classification arises when there is a larger difference between, causing the weight rate to increase. The process is iterated for ‘n’ times till the process achieves a result.

C. Artificial Neural Network (ANN) Based Classification

The ANN is designed based on the inspiration of biological neural networks, which develop the model based on the human nervous system. The model consists of a neuron processing layer that denotes mathematical computations for the NN. During the supervised learning process, an ANN analysis obtains experience from a pre-defined training sample set. The error minimization operation is learned with the evaluations. In the process of the ANN model, the supervised learning model is utilized to process non-linear mapping in back-propagation-based neural network methods. During the learning phase, the input samples are given to the network, and the actual results are compared with the obtained responses. If the actual results differ from the target results, an error signal is backpropagated to regulate the feature weights. The backpropagation-based ANN classification framework is presented in Figure 5.

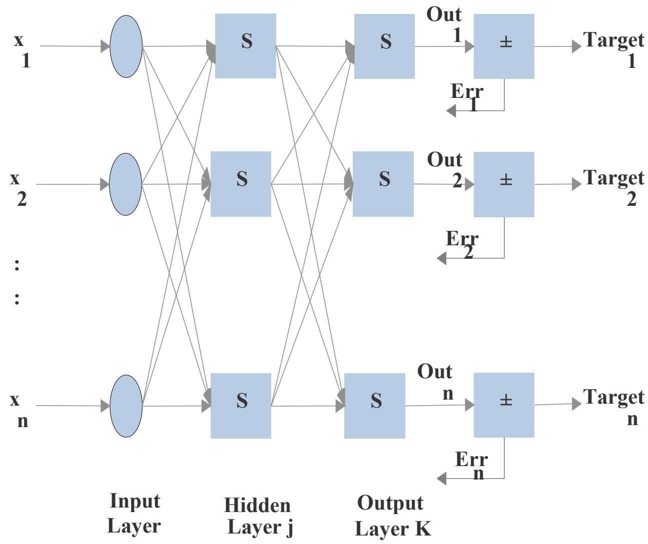

Figure 5: Model for Classification

The directional flows of backpropagation NN are given a,

i. Forward propagation of operational signals
ii. Back-propagation of error signals.

For the first functional flow, computations of actual output and error rates are carried out. The backpropagation function involves reducing the error rates, which is calculated as, mean-square error (MSE), given as,

$$MSE = \frac{1}{2}(target\ result - actual\ result)^2 \tag{1}$$

The learning factor computations (L_r), in the opposite direction makes the maximal reduction in error function. And, the weight rate (WR) is computed as,

$$WR_{new} = WR_{old} - L_r \frac{\sigma E}{\sigma WR_{old}} \tag{2}$$

The net weight rate for backpropagation is computed, based on,

iii. Compute WR for all Ws of the output layer
iv. Compute WR for all Ws from all the hidden layers to the first layer

After processing of training the model with datasets, the real-time patient data are given for testing and the evaluations are carried out based on the evaluation metrics such as sensitivity, specificity, accuracy, precision, and error rates.

4 Results and Discussions

A. Dataset Description

For training the machine learning model, dataset acquisition is the important step considered here for attaining accurate classification results. The utilized dataset is obtained from an eminent data repository called UCI Machine Learning Repository, comprising five datasets: Hungary, Cleveland, Switzerland, Statlog Heart Disease dataset, and VA Long Beach. The data is combined and provided for the machine learning model to produce accurate results. The data from 1190 patient datasets is obtained with 14 significant features as presented in Table 1, and 'num' is considered as the main factor. The predicted rate is '0' indicating the absence of heart disease, and the other value denotes the stages of heart disease.

B. Evaluation Metrics

The evaluation metrics for model analysis in disease diagnosis are presented in this section. The computations are performed based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). Here, the ROC graph is plotted between the TP and FP rates. The formulas for measuring classification accuracy, precision rate, sensitivity, specificity, error rate, and F1-score are provided below.

A. Classification Accuracy:

The accuracy rate is defined as the rate of correctly detected data to the total number of data given for analysis. The formula is given as

$$classification\ accuracy = \frac{No.of\ correctly\ classified\ instances}{No.of\ instance\ \ processed} \times 100\% \quad (3)$$

B. Sensitivity:

The sensitivity rate is computed based on the TP rate from the classifications, and the equation is given as

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

C. Specificity:

Specificity rate can also be termed as, the rate of true negative rate and the computation is given as,

$$Specificity = \frac{FP}{TN+FP} \quad (5)$$

D. Error Rate:

The error rate is a significant factor for computing the model efficacy. Here, K-fold validation is analyzed from the classification results and the average error rate is calculated as,

$$Error\ Rate = \frac{1}{n}\sum_{i=0}^{n} \frac{TN_k+FN_k}{Total\ Instances} \quad (6)$$

E. Precision Value:

The precision rate is computed based on the classification of data under appropriate class, as, Abnormal (presence of CVD). And, the formula is given as,

$$Precision = \left(\frac{TP}{TP+TN}\right) \times 100\% \quad (7)$$

F. F1-Score:

The F1 score is computed from the values obtained for precision and recall, where recall is the specificity rate obtained from (5). And, the equation is given as,

$$F_1 - Score = \frac{2*Precision*Recall}{Precision+Recall} \quad (8)$$

C. Analysis and Comparisons

The results are evaluated and compared with the existing models, such as the Naive Bayes (NB) model, Multi-Layer Perceptron (MLP), and Support Vector Machine (SVM). The received data file is converted to Attribute Relation File Format (ARFF) for adapting into the WEKA environment for processing. Figure 6 and Figure 7 present the results for sensitivity and specificity rate analysis and the obtained values are presented in Table 2 and Table 3, respectively. It is observed from both the figures that the proposed model achieves better results of sensitivity and specificity than the compared methods, analyzing with varying numbers of datasets obtained. The efficient incorporation of RSF for feature selection makes the classification process more appropriate.

Table 2: Results for Sensitivity Rate Analysis

MODEL	10 Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	51.9	71.9	75.6	78.5	85.4	78.7
NB	62.4	76	74.1	85.8	82.7	75.7
MLP	78.8	87.4	90.6	87.8	90.9	88.6
IIM-HDD	89.8	92.4	97.3	96.8	95.8	95

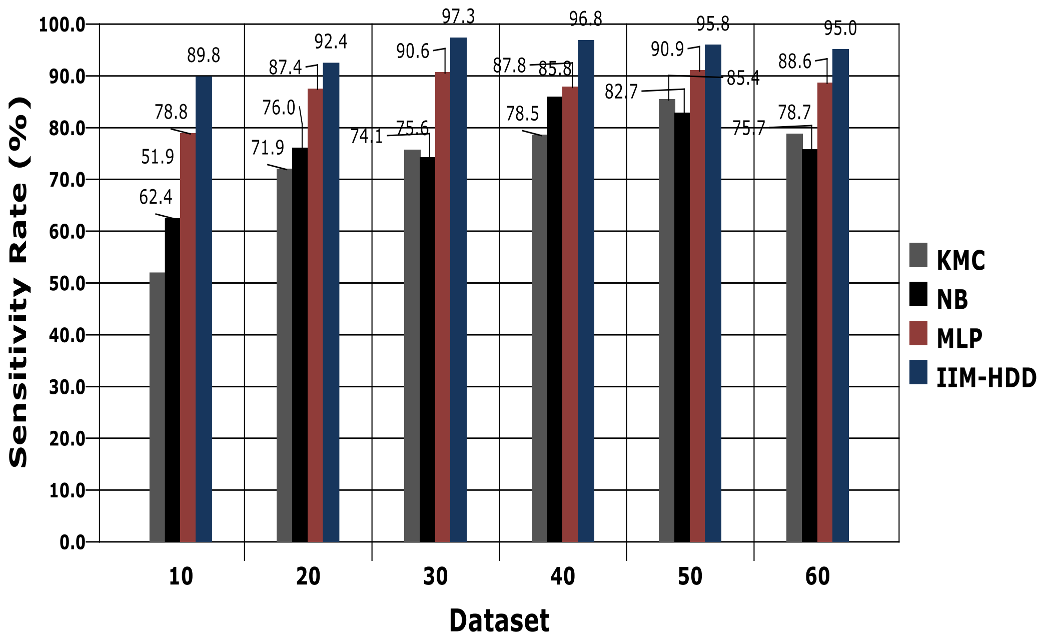

Figure 6: Sensitivity Rate Comparisons

Table 3: Results For Specificity Rate Analysis

MODEL	10 Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	70.0	84.8	71.7	82.2	78.9	71.8
NB	76.5	68.3	75.5	83.3	73.2	66.7
MLP	79.3	83.9	83.3	87.8	86.2	81.2
IIM-HDD	96.6	98.4	97.2	97.6	93.7	95.5

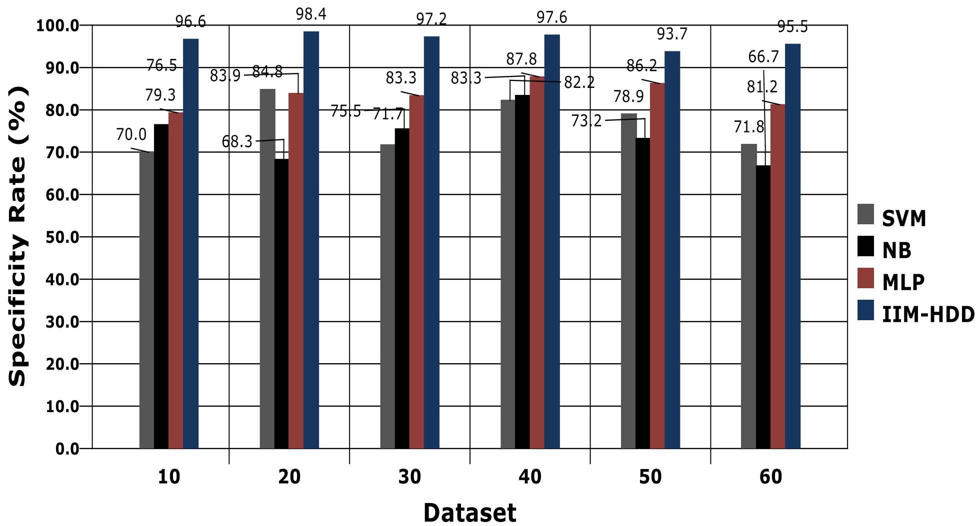

Figure 7: Specificity Rate Comparisons

The classification accuracy and precision analysis are employed, and the results are portrayed in Figure 8 and Figure 9, with the corresponding values provided in Table 4 and Table 5. The classification accuracy analysis shows that the proposed IIM-HDD model achieves a 96.3% average, which is higher than other existing models in disease diagnosis. Further, the precision rate is considered for model efficiency analysis; while concerning, the IIM-HDD model attains a higher rate of precision. Incorporation of ANN for classification, the learning model is trained with the optimal features for disease diagnosis.

Table 4: Results For Classification Accuracy

MODEL	10 Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	69.9	69	58.1	58.7	65.2	69.6
NB	76.9	80.3	64.4	50.7	50.0	66.2
MLP	88.7	91.1	78.6	68	68.1	82.6
IIM-HDD	95.8	97.2	94.7	93.8	96.7	95.0

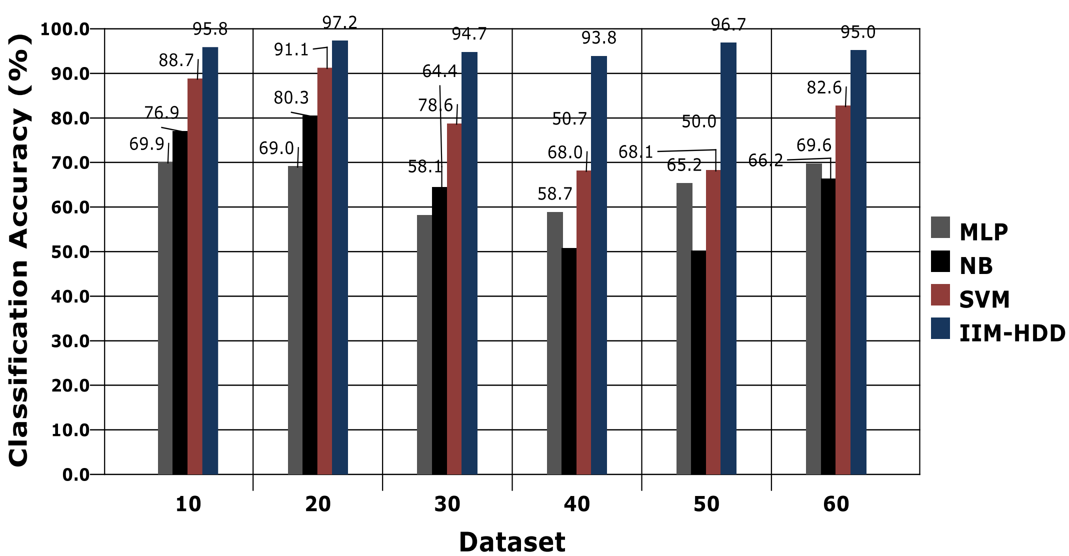

Figure 8: Accuracy Based Comparison

Table 5: Precision Results

MODEL	10Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	90.5	93.1	81.2	73.4	83.6	78.2
NB	95.5	86.2	83.3	77.7	80.8	70.8
MLP	96.1	87.8	93.4	82.7	90.5	88.5
IIM-HDD	97.7	97.6	98.2	92	98.4	96

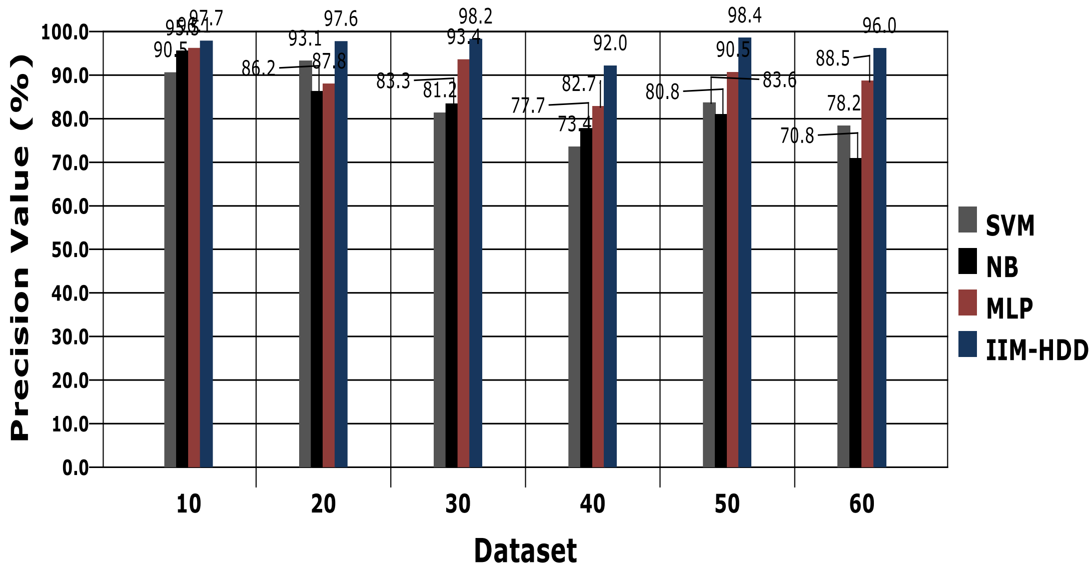

Figure 9: Precision Result

In the proposed model, back-propagation based ANN is trained for disease diagnosis from instances. The error rate is calculated and the results are portrayed in Figure 10 and the obtained rates of analysis are given in Table 6. For the varying range of datasets as {10, 20, 30,40, 50, 60}, the error rates are obtained as, {17.1, 29.4, 35.9. 44.4. 36.2. 44.6}, is less than the other models and validates the efficiency of the proposed work. When medical data processing is considered, time effectiveness is the main factor, for treating the patients without delay. Hence, time-effectiveness based calculations are also carried out and the results are given in Table 7 with the graph comparisons in Figure 11. As the graph denotes, the proposed model achieves better results than existing works in terms of time effectiveness.

Table 6: Error Rate Results

MODEL	10Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	33.7	36.8	59.0	74.0	84.1	80.9
NB	44.1	41.8	51.9	75.5	72.7	87.6
MLP	35.4	51.6	64.3	86.5	62.4	66.3
IIM-HDD	17.1	29.4	35.9	44.4	36.2	44.6

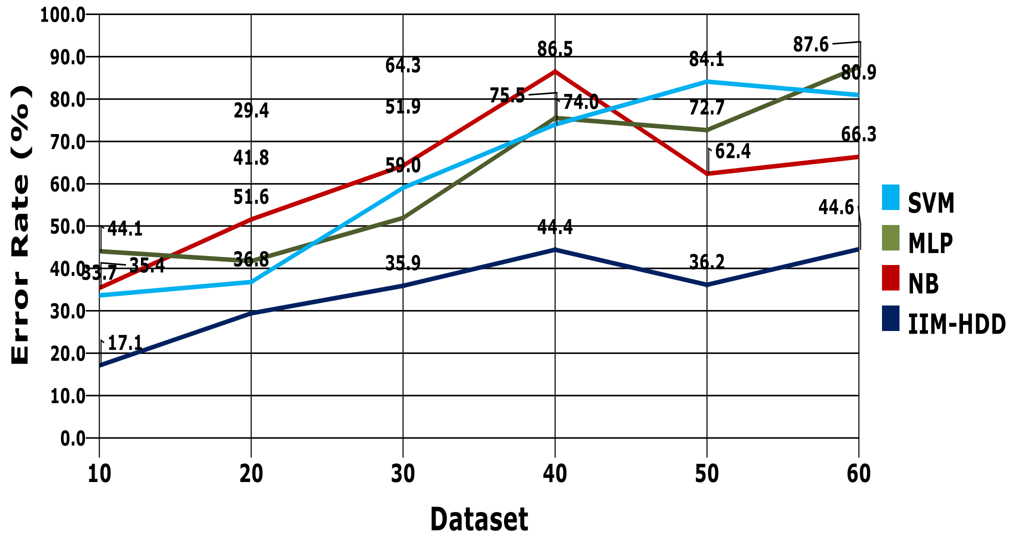

Figure 10: Error Rate Comparison

Table 7: Processing Time Results

MODEL	10Samples	20 Samples	30 Samples	40 Samples	50 Samples	60 Samples
SVM	38.5	47.1	65	70.2	77.1	69.6
NB	52.8	44.6	45.4	64.7	67.2	80.5
MLP	38.7	44.5	66.4	73.4	53.1	60.3
IIM-HDD	13.9	24.5	29.4	34.7	38.9	46.2

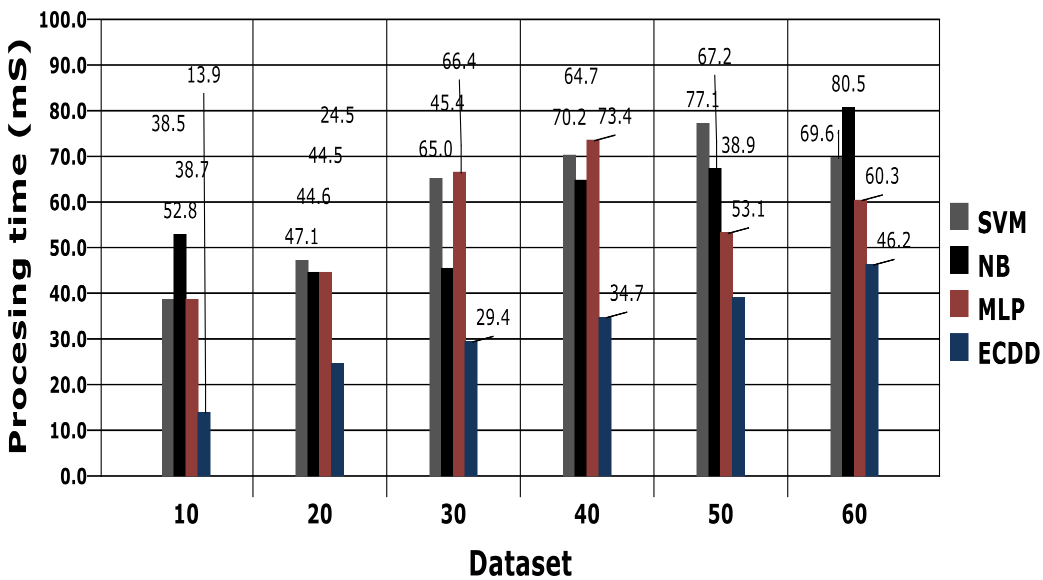

Figure 11: Time Effectiveness

5 Conclusion and Future Work

Identifying the heart disease risks with a higher rate of accuracy can effectively have a greater impact on the long-term death rate of people. The main objective of this model is the earlier and more accurate diagnosis of cardiovascular disease. Hence, the paper proposes the Improved Interpretation Model for Heart Disease Diagnosis (IIM-HDD), which is trained and tested with larger datasets to obtain accurate results. The model used RSF-based feature selection, which involves selecting highly coherent features for training the machine learning model. The proposed model is trained with 14 selected features, which impacts in attaining a higher rate of accuracy and precision, with minimal error and time than other compared models. Here, the proposed model achieves 96.3% accuracy in HDD.

In the future, the model can be enhanced considering the robustness of the model while handling missing data. In the other direction, the work can be focused on developing a novel model for real-time implementation with easy practical settings.

Acknowledgment

This work was supported and funded by a Telekom Malaysia Research & Development (TMR&D) grant, RDTC 241126/ MMUE/240081

References

[1] Abdeldjouad, F. Z., Brahami, M., & Matta, N. (2020). *A hybrid approach for heart disease diagnosis and prediction using machine learning techniques* (Vol. 12157, LNCS). Springer International Publishing.

[2] Alić, B., Gurbeta, L., & Badnjević, A. (2017, June). Machine learning techniques for classification of diabetes and cardiovascular diseases. In *2017 6th mediterranean conference on embedded computing (MECO)* (pp. 1-4). IEEE. 10.1109/MECO.2017.7977152.

[3] Alizadehsani, R., Habibi, J., Sani, Z. A., Mashayekhi, H., Boghrati, R., Ghandeharioun, A., ... & Alizadeh-Sani, F. (2013). Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features. *Research in cardiovascular medicine*, *2*(3), 133-139. 10.5812/cardiovascmed.10888.

[4] Al-rammahi, G. G., & Alshimaysawe, U. A. (2025). Diagnosis and Pathogenesis of Fungi F. Oxysporum and R. Solani Associated with Tomato Seed Rot and Seedlings Death in Najaf. *Natural and Engineering Sciences*, *10*(1), 176-187. https://doi.org/10.28978/nesciences.1643488

[5] Andreotti, F., Heldt, F. S., Abu-Jamous, B., Li, M., Javer, A., Carr, O., ... & Dürichen, R. (2020). Prediction of the onset of cardiovascular diseases from electronic health records using multi-task gated recurrent units. *arXiv preprint arXiv:2007.08491*. https://doi.org/10.48550/arXiv.2007.08491

[6] Armas, D. G. A., Toapanta, S. M. T., Díaz, E. Z. G., Guerrero, J. L. J., Arellano, R. M., & Hifóng, M. M. B. (2025). Influence of Social Media and Artificial Intelligence on Cyberbullying for Decision-Making with Legal or Judicial Foundations in Ecuador. Journal of Internet Services and Information Security, 15(1), 32-50. https://doi.org/10.58346/JISIS.2025.I1.003

[7] Arora, G. (2024). Desing of VLSI Architecture for a flexible testbed of Artificial Neural Network for training and testing on FPGA. *Journal of VLSI circuits and systems*, *6*(1), 30-35. https://doi.org/10.31838/jvcs/06.01.05

[8] Ashraf, M., Ahmad, S. M., Ganai, N. A., Shah, R. A., Zaman, M., Khan, S. A., & Shah, A. A. (2020, August). Prediction of cardiovascular disease through cutting-edge deep learning technologies: an empirical study based on TENSORFLOW, PYTORCH and KERAS. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2020, Volume 1* (pp. 239-255). Singapore: Springer Singapore.

[9] Baad, B., Bhatkhande, N., Kalokhe, K., & Raghatwan, V. P. J. (2019). Heart disease prediction and detection. *Int. J. Res. Appl. Sci. Eng. Technol*, *7*(4), 2293-2299.

[10] Balakrishnan, M., Christopher, A. A., Ramprakash, P., & Logeswari, A. (2021, February). Prediction of cardiovascular disease using machine learning. In *Journal of Physics: Conference Series* (Vol. 1767, No. 1, p. 012013). IOP Publishing. 10.1088/1742-6596/1767/1/012013

[11] Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart

failure from serum creatinine and ejection fraction alone. *BMC medical informatics and decision making*, *20*(1), 16. https://doi.org/10.1186/s12911-020-1023-5

[12] El-Sappagh, S., Ali, F., Ali, A., Hendawi, A., Badria, F. A., & Suh, D. Y. (2018). Clinical decision support system for liver fibrosis prediction in hepatitis patients: A case comparison of two soft computing techniques. *IEEE Access*, *6*, 52911-52929. https://doi.org/10.1109/ACCESS.2018.2868802

[13] El-Sappagh, S., Alonso, J. M., Ali, F., Ali, A., Jang, J. H., & Kwak, K. S. (2018). An ontology-based interpretable fuzzy decision support system for diabetes diagnosis. *IEEE Access*, *6*, 37371-37394. doi: 10.1109/ACCESS.2018.2852004.

[14] Elyasigomari, V., Mirjafari, M. S., Screen, H. R., & Shaheed, M. H. (2015). Cancer classification using a novel gene selection approach by means of shuffling based on data clustering with optimization. *Applied Soft Computing*, *35*, 43-51. https://doi.org/10.1016/j.asoc.2015.06.015

[15] Gupta, A., Kumar, R., Arora, H. S., & Raman, B. (2019). MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE access*, *8*, 14659-14674. 10.1109/ACCESS.2019.2962755.

[16] Javid, I., Alsaedi, A. K. Z., & Ghazali, R. (2020). Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method. *International Journal of Advanced Computer Science and Applications*, *11*(3).

[17] Khan, S. A., Shayea, I., Ergen, M., El-Saleh, A. A., & Roslee, M. (2021, December). An improved handover decision algorithm for 5G heterogeneous networks. In *2021 IEEE 15th Malaysia International Conference on Communication (MICC)* (pp. 25-30). IEEE. https://doi.org/10.1109/MICC53484.2021.9642076

[18] Kumar, N., & Sikamani, K. (2020). Prediction of chronic and infectious diseases using machine learning classifiers-A systematic approach. *Int J Intell Eng Syst*, *13*(4), 11-20. doi: 10.22266/IJIES2020.0831.02.

[19] Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, *16*, 100203. https://doi.org/10.1016/j.imu.2019.100203

[20] Latha, R. S., & Chandran, M. (2025). Role of Artificial Intelligence-Enabled Marketing Strategies on Purchase Decisions. Indian Journal of Information Sources and Services, 15(1), 260–266. https://doi.org/10.51983/ijiss-2025.IJISS.15.1.33

[21] Mezzatesta, S., Torino, C., De Meo, P., Fiumara, G., & Vilasi, A. (2019). A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Computer methods and programs in biomedicine*, *177*, 9-15. https://doi.org/10.1016/j.cmpb.2019.05.005.

[22] MHM, N., Deepthi, S., Murugan, S., Farzana, Y., Kabir, M. S., Ahmed, S. U., Doustjalali, S. R., Sabet, N. S., Udayah, M. W., Ying, T. S., Chia, T. Y., Shirin, L., & Subramaniyan, V. (2025). Microbial contamination in aquatic ecosystems: Implications for human health and disease prevention. I*nternational Journal of Aquatic Research and Environmental Studies, 5*(1), 408–430. https://doi.org/10.70102/IJARES/V5I1/5-1-38

[23] Miao, F., Cai, Y. P., Zhang, Y. X., Fan, X. M., & Li, Y. (2018). Predictive modeling of hospital mortality for patients with heart failure by using an improved random survival forest. *Ieee Access*, *6*, 7244-7253. 10.1109/ACCESS.2018.2789898

[24] Mienye, I. D., Sun, Y., & Wang, Z. (2020). An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked*, *20*, 100402. https://doi.org/10.1016/j.imu.2020.100402

[25] Mishra, J., & Tarar, S. (2020). Chronic disease prediction using deep learning BT—advances in computing and data sciences.

[26] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554. 10.1109/ACCESS.2019.2923707.

[27] Morris, K., & McNicholas, P. D. (2016). Clustering, classification, discriminant analysis, and dimension reduction via generalized hyperbolic mixtures. *Computational Statistics & Data Analysis*, *97*, 133-150. https://doi.org/10.1016/j.csda.2015.10.008.

[28] Nandhinieswari, S., & Indumathi, A. (2024). BILEVEL OPTIMIZED RECURSIVE FEATURE ELIMINATOR FOR CERVICAL CANCER FEATURE SELECTION PROCESS. *Archives for Technical Sciences*, *2*(31), 311-328. https://doi.org/10.70102/afts.2024.1631.311

[29] Raju, C., Philipsy, E., Chacko, S., Suresh, L. P., & R., D. (2018, March). *Mining techniques* (pp. 2–3).

[30] Roslee, M. B., Abdullah, R. S. A. R., & Shafr, H. Z. (2010). Road pavement density analysis using a new non-destructive ground penetrating radar system. *Progress In Electromagnetics Research B*, *21*, 399-417. http://dx.doi.org/10.2528/PIERB10032202

[31] Roy, S., Tiang, J. J., Roslee, M. B., Ahmed, M. T., & Mahmud, M. P. (2021). A quad-band stacked hybrid ambient RF-solar energy harvester with higher RF-to-DC rectification efficiency. *IEEE Access*, *9*, 39303-39321.

[32] Sa. S., Intelligent heart disease prediction system using data mining techniques. (2013). *Int. J. Healthc. Biomed. Res.,* 1, 94–101.

[33] Sadeque, M. G., Yusoff, Z., Roslee, M., & Hadi, N. S. R. (2019, August). Design of a broadband continuous class-F RF power amplifier for 5G communication system. In *2019 IEEE Regional Symposium on Micro and Nanoelectronics (RSM)* (pp. 145-148). IEEE. https://doi.org/10.1109/RSM46715.2019.8943508

[34] Saqlain, S. M., Sher, M., Shah, F. A., Khan, I., Ashraf, M. U., Awais, M., & Ghani, A. (2019). Fisher score and Matthews correlation coefficient-based feature subset selection for heart disease diagnosis using support vector machines. *Knowledge and Information Systems*, *58*(1), 139-167. https://doi.org/10.1007/s10115-018-1185-y

[35] Tama, B. A., Im, S., & Lee, S. (2020). Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble. *BioMed Research International*, *2020*(1), 9816142. https://doi.org/10.1155/2020/9816142

[36] Tan, K. C., Teoh, E. J., Yu, Q., & Goh, K. C. (2009). A hybrid evolutionary algorithm for attribute selection in data mining. *Expert Systems with Applications*, *36*(4), 8616-8630. https://doi.org/10.1016/j.eswa.2008.10.013

[37] Tarawneh, M., & Embarak, O. (2019, February). Hybrid approach for heart disease prediction using data mining techniques. In International Conference on Emerging Internetworking, Data & Web Technologies (pp. 447-454). *Cham: Springer International Publishing.*

Authors Biography

Dr.F. Syed Anwar Hussainy is an Assistant Professor in the Department of Computer Science and Engineering (AI & ML) at Jain University, Bangalore. He obtained his B.E Degree in Computer Science and Engineering in the year 2004, and M. Tech. Degree in Information Technology in the year 2010.He completed his Ph.D. in Computer Science and Engineering (CSE) in the year 2024 with a specialization in Deep Learning. He has published many research articles and holds a granted patent in machine learning. His research interests include healthcare analytics, artificial intelligence, machine learning, and deep learning.

J. Jayapradha is an Assistant Professor at department of Computing Technologies, SRMIST. She had completed her Ph.D(CSE) in Privacy Preservation. and has published many research articles on Machine Learning, Data Mining, and Privacy. She has 6 patents published and 2 patents granted in the field of ML and data privacy. Her research interests include Machine Learning, Databases, and Privacy.

Prof. Dr. Mardeni Roslee is a Professor and Chairman of Centre for Wireless Technology, Faculty of Engineering, Multimedia University since 2008. His current research interests are 5G/6G telecommunication, D2D, satellite, Internet of Things and radar communication. At International/local level, he was a Chairman of IEEE Malaysia in Comsoc/VTS. He is a registered Chartered Engineer with Engineering Council United Kingdom, and Member with The Institution of Engineering and Technology (IET), UK. He was invited as keynote speaker for 26 International conferences and been invited by IEEE international conference at 16 different countries since 2009. He has hold as General Chair of 11 International conference world wide. He has 220 publications in indexed journals, conferences and books. His contributions to academic and the engineering professional over the years have earned him recognitions nationally and internationally, he has awarded more than 100 International/local awards including IEEE Outstanding Large Chapter Award for 2019 and 2020, 2021 IEEE Malaysia Outstanding Volunteer Award and winner of Top Research Scientist in Malaysia 2020 from Academy of Science Malaysia (ASM).

Dr.T. Senthil Kumar is an Associate Professor, graduated with a bachelor's degree in Electronics and Communication Engineering and a master's degree in computer science and engineering in 2004 and 2006, respectively. He completed his Ph.D. in CSE from SRMIST, Chennai, 2019. His areas of expertise include Wireless Communication, Network Security, Machine Learning, the Internet of Things, and Vehicular Networks.

Dr. Chilakala Sudhamani received her B.Tech degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2005, M.Tech degree in Digital Electronics and Communication Science from Jawaharlal Nehru Technological University, Kakinada, India in 2011 and completed her Ph.D. degree in Wireless Communications and Networking from Acharya Nagarjuna University, Andhra Pradesh, India in 2020. She is currently working as a Postdoctoral Research Fellow in the Center for Wireless Technology Lab, Faculty of Engineering, Multimedia University Malaysia. She published more than 30 research papers in reputed journals and conferences. Her areas of research interest are 5G Networks, wireless channel modeling, spectrum sensing in cognitive radio networks, Cooperative Spectrum sensing, Signal Processing.

Azmi Ismail is an experienced engineer (researcher) with a demonstrated history of working in the telecommunications, semiconductor & electronics industry. Programming skills in Javascript (Back-end), Python, Java/Android, C/C++ and hardware skills in mixed-signal circuit design and data converter. Strong research professional with a Bachelor of Engineering (B.Eng) majored in Electrical and Electronics from Universiti Teknologi Malaysia (UTM).

Anwar Faizd Osman was born in Georgetown, Penang. He has a Master of Science degree in Electrical and Electronic engineering from Universiti Sains Malaysia (USM). His master's thesis is on wideband low noise amplifier design. He is currently the Head of 5G Product Development for Telekom Malaysia. In Telekom Malaysia, he leads the End to End roadmap development for Telekom Malaysia (TM) 5G Core as a Service, which include Network Slicing, 5G Standalone capable solutions and related 5G Value Added Services. Prior to Telekom Malaysia, he has worked in Rohde & Schwarz, Agilent Technologies, Motorola Solutions and Intel Malaysia, either in manufacturing or field application positions. He has published multiple technical papers on LNA, RF switches, and RF filter designs. His research interests include wireless testing for mobile operators and interference hunting. Mr. Osman has been a

Committee Member of IEEE ED/MTT/SSC Penang Chapter, since 2015, currently serves as the Chapter Author.

Dr. Fatimah Zaharah Ali is a senior lecturer at College of Engineering, Universiti Teknologi MARA (UiTM), Malaysia since 2012. She obtained her Bachelor of Engineering (Honours) in Electrical and Electronics from Universiti Teknologi Petronas (UTP), Perak, Malaysia in 2009; MSc. in Telecommunication and Information Engineering from Universiti Teknologi MARA (UiTM), Selangor, Malaysia in 2012; and PhD in Electrical Engineering, UiTM, Selangor, Malaysia in 2023. She also had worked as an Assistant Manager in Telekom Malaysia (TM) Berhad in 2010 for a year before she pursued her master's degree.

Idris Olalekan Adeoye is a senior lecturer at Centre for Wireless Technology, Faculty of AI & Engineering, Multimedia University Cyberjaya, Malaysia.