WILEY | Hindawi

*Research Article*

# An Accurate Heart Disease Prognosis Using Machine Intelligence and IoMT

**Jamshid Pirgazi [ID],[1] Ali Ghanbari Sorkhi [ID],[1] and Majid Iranpour Mobarkeh [ID][2]**

[1]*Department of Computer Engineering, University of Science and Technology of Mazandaran, Behshahr, Iran*
[2]*Department of Computer Engineering and IT, Payam Noor University, Tehran, Iran*

Correspondence should be addressed to Jamshid Pirgazi; j.pirgazi@mazust.ac.ir

In recent years, Internet of Medical Things (IoMT) and machine learning (ML) have played a major role in the healthcare industry and prediction of in time diagnosis of diseases. Heart disease has long been considered one of the most common and lethal causes of death. Accordingly, in this paper, a multiple-step method using IoMT and ML has been proposed for diagnosis of heart disease based on image and numerical resources. In the first step, transfer learning based on convolutional neural network (CNN) is used for feature extraction. In the second step, three methods of distributed stochastic neighbor embedding (t-SNE), F-score, and correlation-based feature selection (CFS) are utilized to select the best features. In the end, a combination of outputs of three classifiers including Gaussian Bayes (GB), support vector machine (SVM), and random forest (RF) according to the majority voting is employed for diagnosis of the conditions of heart disease patients. The results were evaluated on the two UCI datasets. The results indicate the improvement of performance compared to other methods.

## 1. Introduction

According to the World Health Organization (WHO) statistics, cardiovascular disease is one of the leading causes of death worldwide, accounting for 17.9 million deaths each year [1]. The main causes of heart disease are various unhealthy activities such as high cholesterol, obesity, an increase in triglyceride levels, and high blood pressure, among others. Sleep problems, irregular heartbeat, swollen legs, and, in some cases, weight gain of 1 to 2 kilograms per day all increase the risk of heart disease [2, 3]. All these symptoms are common within various diseases leading to death in the near future; therefore, the correct diagnosis is difficult.

Smart healthcare presents healthcare platforms which make use of tools such as IoT, wearable appliances, and wireless Internet connection for signing in health evidences and resource connection, organizations, and individuals. IoT, artificial intelligence (AI), big data, cloud networks, 5G, and advanced biotechnology are some of the smart healthcare networks used in disease screening and diagnosis and medical research [4].

As previously mentioned, IoT and IoMT play a great part in the healthcare in prediction of time and chronic illness diagnosis. The volume of information required by the healthcare, security factors, power of processing, and accuracy of information is very important in terms of diagnostic prediction for many illnesses. To tackle these challenges, AI algorithms in previous researches are used to increase the precision of patients' data [5].

IoMT refers to disease diagnosis without human intervention through the development of intelligent sensors, smart devices, and advanced lightweight communication protocols. IoMT-based healthcare, swallowable sensor tracking, mobile health, smart hospitals, and improved treatment of chronic diseases have been shown in [6].

IoMT is a new network-based technique for connecting medical devices and their applications to healthcare information technology systems. In [7], in addition to providing treatment to orthopedic patients, the IoMT approach examines the possibilities of facing with COVID-19 pandemic.

In the recent years, ML is widely utilized in healthcare industry to analyze big data for initial prediction of diseases

leading to the improvement of the quality of healthcare [8, 9]. ML can be used to solve complex health issues and give accurate results. Healthcare industry is one of the largest industries in which ML has shown to be functional. Creating accurate and multidimensional datasets are very important and play a critical role in the functionality of ML algorithms. IoMT enables medical facilities and healthcare products to share real-time data to create a great volume of data for ML [10].

Lately, large amount of research data and patients' cases have become accessible. There are many open sources for gaining access to patients' records, and research can be done to be able to use computer technologies for patient identification and accurate disease diagnosis in order to prevent the lethality of these illnesses. Today, ML and AI are well recognized to play major roles in healthcare industry, and various models of ML and deep learning (DL) can be employed to classify and diagnose diseases or to predict results. Complete analysis of genome data can easily be done using different models of ML [11–13].

Several studies have utilized different models of ML for classification and diagnosis of heart diseases. CART automatic classifier based on classification and regression of congestive heart failure [14], using deep neural network for best feature selection and ECG performance improvement [15], proposing a clinical decision support system for diagnosis of heart failures and its prevention during initial stages of the disease [16], and also rule-based natural language processing (NLP) [17] are among these researches.

In today's digital age, healthcare generates a large amount of patient data. For physicians, manual control of these data is difficult, whereas IoT can manage the produced data very efficiently. IoT records large amounts of data and is capable of diagnosing diseases using machine algorithms with the purpose of applying different methods of ML on the produced data. A ML approach is proposed for initial heart disease prediction in relation to IoT [10].

Cardiac image processing approaches which are obtained from DL manage and supervise large medical data gathered by the IoT. Deep IoMT is a common DL and IoT platform that is in charge of extracting precise cardiac image data of usual instruments and devices. Energy depletion, finite battery life, and high PLR (packet loss ratio) are critical issues that must be addressed in universal medical care. Wearable devices must be stable (i.e., have a longer battery life), energy efficient, and valid in order to improve an affordable and inclusive healthcare environment. In this regard, a new efficient approach based on the consciously enhanced efficient-aware approach (EEA) of self-adaptive power control to decrease energy utilization while increasing validity and battery life is proposed in [18]. For remote cardiac imaging of elderly patients, a new common DL-IoMT framework (DL-based layered architecture for IoMT) has also been proposed.

Medical image classification is critical in the prediction and early detection of critical illnesses. Medical imaging is the most essential record of patient's health which helps to control and cure illnesses, which is one of the important applications of IoMT. In [19], an improved classification of optimal DL for the lung cancer classification, brain imaging, and Alzheimer's disease is introduced. The researches show that medical image classification is based on optimal feature selection using the DL by combining preprocessing, feature selection, and classification. The primary goal of model extraction is to select an effective feature for medical image classification. The opposition-based crow search (OCS) approach is recommended to enhance the efficiency of the DL classifier. In addition, multitextured, gray-level features are chosen for analysis. Finally, it is claimed that the optimal features made better the result of classification.

This study presents a method based on data collected by IoT. In this regard, a general method is presented for numerical and image data. At first, the proposed method examines the type of data resource. If input data were from image resources, in the first step, features are extracted from this type of resource using transfer learning. CNN-based deep network is used for this purpose. Fully connected layer has been utilized for feature extraction, whereas if the input data were from numerical sources, the first step is ignored. The proposed method's next steps include feature selection and classification phases, which are independent of the input resource. In the feature selection step, three methods of distributed stochastic neighbor embedding (t-SNE), F-score, and correlation-based feature selection (CFS) have been used. An individual classifier has been trained for each method of feature selection. In this paper, three classifiers of SVM, GB, and RF have been employed. In the end, voting is used for final label selection. The results demonstrate that the proposed method performs well.

The rest of this paper is organized as follows. Section 2 discusses previous research in this area. Section 3 examines the proposed method and its details. Section 4 compares the performance of the proposed method to some of the successful models in this field, and Section 5 concludes the paper.

## 2. Literature

With the recent advances in medical data processing and machine learning, many researchers have been consistently active in this field. One of the most challenging medical data is data related to heart diseases which have drawn many researchers' attention. In [20, 21], multiple machine learning methods were examined for the prediction of heart diseases in which recursive neural network (RNN) and decision tree (DT) were reported to have gained the best results.

In [22], deep neural network (DNN) with the name of Heart Evaluation for Algorithmic Risk-reduction and Optimization five (HEARO-5) was proposed. This method which is consisted of regularization has shown positive results on UCI dataset. In [23], for classifying imbalanced clinical data, a neural network with a convolution layer was used. This study takes advantage of a two-step approach feature weight based on least absolute shrinkage and selection operator (LASSO) and then identification of critical features based on majority voting for achieving more accuracy in classified imbalanced data.

In [24], to increase the performance of the classifier, feature selection approaches based on fast correlation-based feature selection (FCBF) were used to choose efficient features. In this method, classification is done using K-nearest neighbor (KNN), SVM, Naive Bayes (NB), RF, and multilayer perceptron (MLP) optimized using particle swarm optimization (PSO) with ant colony optimization (ACO) [25]. NB, SVM, and RF methods were employed for extraction and classification of the most relevant features in [26, 27].

A k-means method with particle swamp was proposed in [28] for detecting hazard factors in coronary heart disease treatment (CAD). The extracted data are classified using MLP, multinomial logistic regression (MLR), and algorithms of phase rule, as well as C4.5. It was claimed that the results demonstrated the appropriate accuracy of the proposed method on the datasets presented by medical college in India. In [29], heart disease prediction has been done using methods of data mining, ML, and DL, and neural network method was claimed to be more functional than other methods. In [30], genetic algorithms and neural networks were employed for diagnosis of heart disease.

## 3. Proposed Method

The general procedure of the proposed method is shown in Figure 1. As it can be seen, this method is made up of three major steps. In the first step, two different approaches with respect to the input resource are used. If data are numerical, only feature vector gets used for the next step; however, if data are image, the feature vector must be extracted. For the purpose of extracting features from images, transfer learning based on CNN has been used. In this stage, fully connected layer is utilized after convolution layers for feature extraction. The second step of the proposed method is made up of feature selection. This step is independent of the input resource. Three methods of t-SNE, F-score, and CFS have been put to use for feature selection. In the third step of the proposed method, for each feature vector of the previous step, three different classifiers of SVM, GB, and RF are used. In the end, majority voting has been used for selection of the favorable output. Labels of the three classifiers used in the last step are the input of the current step. Eventually, the final input label is selected. In the following, different sections of the proposed method will be described.

*3.1. Feature Extraction Based on Image Resource.* The extraction of features is a critical issue in classification [31]. As illustrated in Figure 1, one of the main steps of the proposed method is feature extraction. In the step of feature extraction, if the resource is image, it must turn into a feature vector. Methods based on DL are among the most successful methods for feature extraction; however, unfortunately, the numbers of images related to heart diseases are very low; therefore, in this step, transfer learning has been utilized for feature extraction (Figure 2). A pretrained CNN network is used in this step as well. This network is merely used for feature extraction that the output of fully connected layer is selected as the feature vector.

Transfer learning is an issue of great significance which focuses on knowledge retention of problem-solving and its usage to solve a different but related problem. Since datasets are not sufficiently available, CNN network is not initially trained; thus, pretrained network weights aid to solve more issues concerning feature extraction or configuration. Very deep networks are costly to be trained. More complex models require more time for training using hundreds of systems with expensive CPUs.

Transfer learning maps a model that has already been trained in specific areas to a new model in new domains; thus, the time required for training by using this method is reduced [32]. Furthermore, in complex models, transfer learning decrease the need for a large number of training samples. Because the number of images available in the field of heart disease is limited, this method is used to compute the initial weights from the well-known ImageNet dataset. The ResNet, AlexNet, VGG-16, and VGG-19 architectures trained on ImageNet are evaluated based on a set of validations. VGG-16 architecture has shown the best performance due to experimental results. As shown in Figure 2, this paper uses CNN-based transfer learning to extract features.

*3.2. Feature Selection.* As it is shown in Figure 1, in this section, the feature vector extracted from the previous step is used as the input for feature selection. In this step, three methods of feature selection including t-SNE, F-score, and CFS are used which are further elaborated in the following.

*3.3. Correlation-Based Feature Selection (CFS).* As a filter method, CFS classifies and evaluates feature subsets based on subsets that are highly correlated with the class but unrelated to one another [33]. Irrelevant features should be ignored if they have a low correlation with the class. Aside from that, the duplicated features can be identified because they are closely related to the remaining ones. The feature can be accepted if it predicts the label that no other features predict. The evaluation function of CFS' feature subset is as follows:

$$M_s = \frac{\overline{kr_{cf}}}{\sqrt{k + k(k-1)r_{ff}}}.$$ (1)

In this equation, $M_s$ shows the heuristic "merit" of a feature subset $S$ including $k$ features, and also, $\overrightarrow{r_{cf}}$ and $\overline{r_{ff}}$ represent the mean feature-class correlation ($f \in S$) and the average feature intercorrelation, respectively. The calculation from this equation has the usage to predict not only the feature subsets but also the redundant ones [34].

*3.4. F-Score.* F-score by evaluating the difference between two real numbers sets presents a simple feature selection filter method [35] which for feature $i$ is calculated as follows:

$$F - \text{score}(i) = \frac{\sum_{k=1}^{m} \left(\bar{x}_i^k - \bar{x}_i\right)^2}{\sum_{k=1}^{m} \left(1/n^k - 1\right)\sum_{j=1}^{n^k} \left(x_{j,i}^k - \bar{x}_i^k\right)^2}.$$ (2)
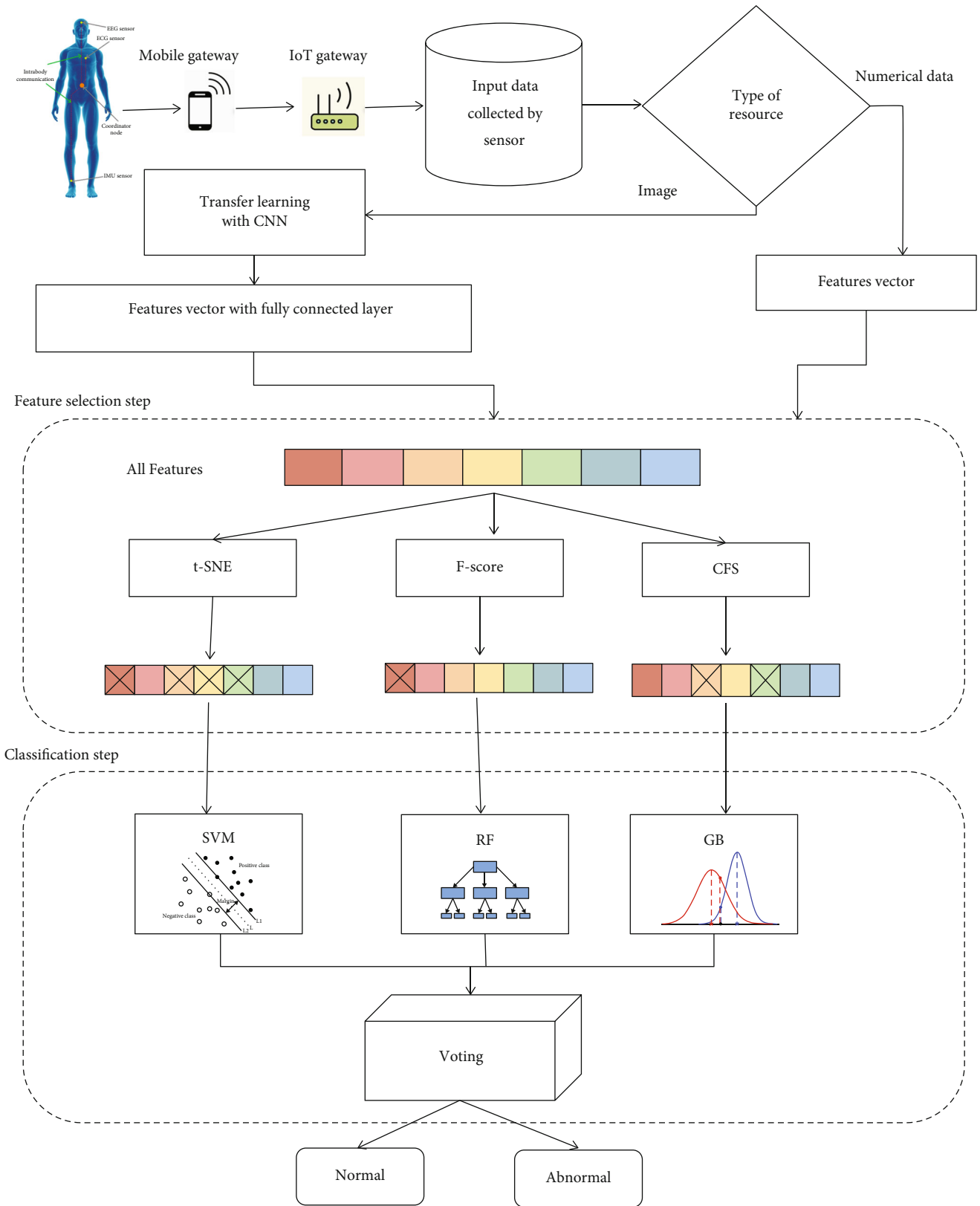
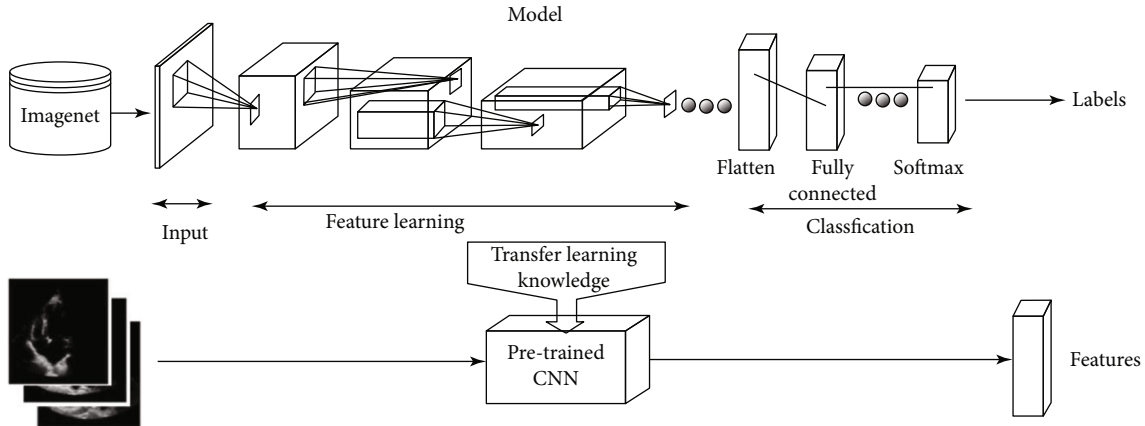FIGURE 1: An overview of the proposed method.

FIGURE 2: Using transfer learning to extract features form image resources.

TABLE 1: Hyperparameters of the basic classifiers.

| Methods | Parameters | Amounts |
| --- | --- | --- |
| SVM | C_SVM | 1 |
| | Kernel_SVM | Radial basis function (RBF) |
| | Degree_SVM | 3 |
| | Gamma_SVM | Scale |
| | Coef0_SVM | 0 |
| GB | Priors_GB | None |
| | Var-smoothing_GB | 1e-08 |
| RF | Min_samples_split_RF | 2 |
| | Min_samples_leaf_RF | 1 |

In the above equation, $m$ refers to the number of classes, $n^k$ shows the samples number of class $k$, $\bar{x}_i$ presents the mean of feature $i$ among data, also $\bar{x}_i^k$ demonstrates the mean of feature $i$ in class $k$, and $x_{j,i}^k$ shows the amount of feature $i$ in the sample $j$ of the class $k$. If F-score related to a feature is high, it shows that the respected feature includes proper information which belongs to classification.

### 3.5. Distributed Stochastic Neighbor Embedding (t-SNE).

This method is an unsupervised nonlinear method which is used for discovery and reduction of data dimensions. In other words, it will provide the user with an understanding of the manner of data organization in a high-dimensional space. This method has been introduced in 2008 by Laurens van der Maatens and Geoffery Hinton [36]. The main difference between this method and principal component analysis (PCA) is that PCA is a method of reducing the linear dimensions which attempts to maximize the variance and preserve the large distance between the pares, while t-SNE preserves PCA in preserving the small distance between pares by using local similarities. t-SNE algorithm computes a similarity measure between the pare of samples in large-dimensional data and low-dimensional space. Then, it attempts to optimize these two similarity measures using a cost function.

This process is undertaken through three main steps. They are as follows:

(1) In the first step, the interpoint similarity in high-dimensional space is measured. To better understand this, suppose a set of scattered data points in a two-dimensional space. For each data point of $x_i$, the Gaussian distribution is spread around that point by the user. Then, the density of all $x_i$ points will be computed based on that Gaussian distribution. Then, renormalization is applied to all data points. This will result in a set of $P_{ij}$ probabilities for all data points. These probabilities are proportional to their similarities. This actually means that if $x_1$ and $x_2$ data points possess a similar value under the Gaussian circle, their proportions and similarities will be equal consequently; hence, the local similarities will hold true in the structure of high-dimensional space

(2) The second step is quite similar to the first; but conversely, Student's $T$-distribution with one level of freedom is used instead of Gaussian distribution which is also known as the Cauchy distribution. This will result in a second set of $Q_{ij}$ probabilities in a low-dimensional space

(3) The last step is associated with the reflection of high-dimensional space probabilities $P_{ij}$ through low-dimensional space probabilities $Q_{ij}$ in the best possible manner. The basic requirement here is the similarity of the two mappings. The difference between two-dimensional space probability distributions is computed through the Kullback-Leibler (KL) divergence criteria. This study does not elaborate upon KL. The only point to be considered is that it is an asymmetrical approach in which the effective comparison of $P_{ij}$ and $Q_{ij}$ values does not suffice. Eventually, the optimal value of the KL cost function is found using gradient descent

### 3.6. Classification.
An ensemble classifier is used on the reduced feature vector. In these types of classifications,

TABLE 2: Description of Cleveland dataset [39].

| No. | Name of attribute | Description |
| --- | --- | --- |
| 1 | Age | Age in years |
| 2 | Sex | Male is equal 1 and female is equal 0 |
| 3 | CP | Type of chest pain |
| 4 | Trestbps | A criterion which shows resting blood pressure |
| 5 | Chol | A criterion which shows serum cholesterol |
| 6 | FBS | A variable which is boolean, when fasting blood sugar > 120 mg/dl is true otherwise it is false |
| 7 | Restecg | A criterion which shows resting electrocardiographic results |
| 8 | Thalach | A criterion which shows maximum heart rate |
| 9 | Exang | A binary variable which shows exercise-induced angina |
| 10 | Oldpeak | A criterion which shows ST depression |
| 11 | Slope | A criterion which shows the peak exercise ST segment |
| 12 | CA | A criterion which shows major vessel number |
| 13 | Thal | A criterion which shows heart rate |
| 14 | NUM | A criterion which shows heart disease status |
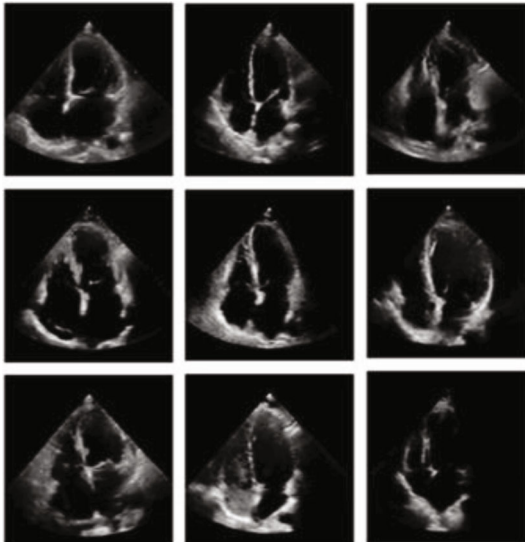


FIGURE 3: Sample image from dataset.

combination of a number of basic classifiers creates an accurate and robust classification. One of the most common ways to combine classifiers is majority voting. As shown in Figure 1, since the diversity of the consisting classifiers gives rise to the power of an ensemble classifier, the SVM, BG, and RF are suggested as basic classifiers. Therefore, it is expected that the sample data to be covered in the maximum range and the generalizability of the classification to be increased. It is better not to use the classification with the similar results in group classification. In order to reduce the classification error, it is important to choose the appropriate classifier and combination strategy.

Support vectors in the SVM model are the most important component of the model, which is obtained through convex optimization. In this model, the classification margin creates the maximum distance within classes. The main assumption in Bayesian classifier is statistical independence between features and in most cases maximizes the performance of the acquisition. In this classifier, model parameters are estimated with a small set of training data. Random forest is a simple machine learning technique that usually produces outstanding results even when its hyperparameters are not adjusted. This technique is one of the most extensively used machine learning algorithms for both regression and classification because of its simplicity and usability [37, 38]. This method works based on building a large number of decision trees. In the proposed method, the classifications are combined by voting according to label repetitions. The main reason for choosing three different classifiers, SVM, BG, and RF, as the basic classifier which is the main component in constructing ensemble classifiers is "diversity." All of these classifiers are trained differently leading to the increase of the level of classification diversity and ensemble generalization.

## 4. Experimental Results

This section summarizes the results of experiments conducted to evaluate the suggested method's performance. It should be noted that all the presented methods and analysis of their results are done on same datasets and similar hardware. All the implementation is done on a computer with Core (TM) i7 M620 CPU, 4GB memory card, and T4 graphic card with Python as programming language as well as Keras framework. It also should be mentioned that Scikit-learn-0.22.0 toolbox has been used for classification and all the parameters in this toolbox also have been utilized by default. For instance, SVC employs the "one vs. one" approach for ensemble classification. Table 1 shows the main classifier parameters.

4.1. Database. The Cleveland dataset from UCI is used to evaluate the proposed method. This dataset is available at http://archive.ics.uci.edu/ml/datasets.php. Cleveland dataset

TABLE 3: Descriptions of echocardiogram dataset [4].

| Name of attribute | Description |
| --- | --- |
| Survival | This variable indicates the number of months the patient survives |
| Still-alive | A variable which is binary, still-alive is shown by 1 and dead by 0 |
| Age at heart attack | Age of heart attack occurrence (in years) |
| Pericardial effusion | A variable which is binary. Fluid around the heart is shown by 1 and no fluid by 0 |
| Fractional shortening | A criterion which measures contractility around the heart |
| Epss | Another criterion which measures contractility (E-point septal separation) |
| Lvdd | A criterion which measures the size of the heart (left ventricular end-diastolic dimension) |
| Wall motion score | A criterion which measures the movement of the left ventricle segments |
| Wall motion index | This criterion depends on number of segments seen that can be used instead of the wall motion score |
| Mult | An ignorable var which is derivative |
| Name | Patient's name |
| Group | Meaningless |
| Alive at 1 | A variable which is boolean, patient was dead after one year is shown by 0 and patient was alive at one year by 1 |
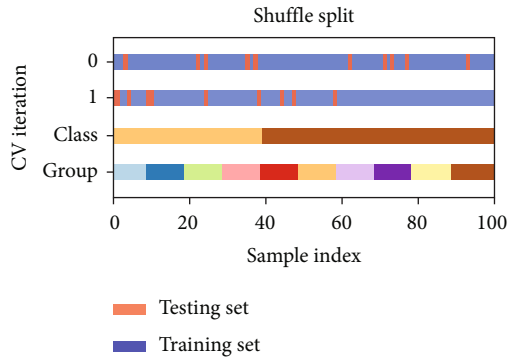


FIGURE 4: Stages of sample selection in testing and training sets.

owns 76 attributes and 303 samples. Nonetheless, only 14 attributes of Cleveland dataset were put to use for training and testing. These features are further elaborated in Table 2. These types of data have been used as numerical resources in the present paper.

In the following, echocardiogram images have been employed as image resources. Figure 3 shows some examples of these images. The suitable attributes are described in Table 3. UCI database was used for echocardiography image retrieval using 66 normal images from 30 participants and 66 abnormal images from 30 subjects [4]. When the variables of "survival" and "still-alive" are combined together, it shows whether the patient has stayed alive at least one year after the heart attack or not.

In the experiments performed to evaluate the proposed method, 10-fold cross-validation was used. The steps for building a training and test set are described in Figure 4. Accordingly, in each repetition, 10% of the data were used as a test set and the rest as a training set. In addition, 10% of the training image sets have been used to create the validation set.

*4.2. Evaluation Criteria.* Several quantitative criteria including specificity (Spe), accuracy (ACC), recall (sensitivity)

(RE), precision (PR), and F1 are used to show the performance of the proposed method [40].

Generally, accuracy (ACC) refers to a model's ability to accurately predict the output label. Equation (3) depicts the accuracy criterion. It also should be mentioned that variance and mean in 10 numbers of repetitions are considered to calculate accuracy for 10-fold cross validation. This criterion examines the training level and functionality of the model, although it has no further information regarding the model accurate functionality.

$$Accuracy = \frac{TP + TN}{total\ examples}. \tag{3}$$

In equation (4), precision criterion is shown that is appropriate for amounts with high false positive.

$$PR = \frac{TP}{TP + FP}. \tag{4}$$

In equation (5), recall (sensitivity) criterion is shown that is appropriate for amounts with high false negative.

$$RE = \frac{TP}{TP + FN}. \tag{5}$$

In equation (6), specificity criterion is shown.

$$Spe = \frac{TN}{TP + FN}. \tag{6}$$

F1 criterion is shown in equation (7). This criterion also contains accuracy and recall (sensitivity) criteria. F1 approaches 0 and 1, respectively, in its worst and best cases.

$$F_1 = \frac{2 * RE * PR}{PR + RE}. \tag{7}$$

In the aforementioned equations, TP presents the number of images which is correctly allocated to $C_i$ class by
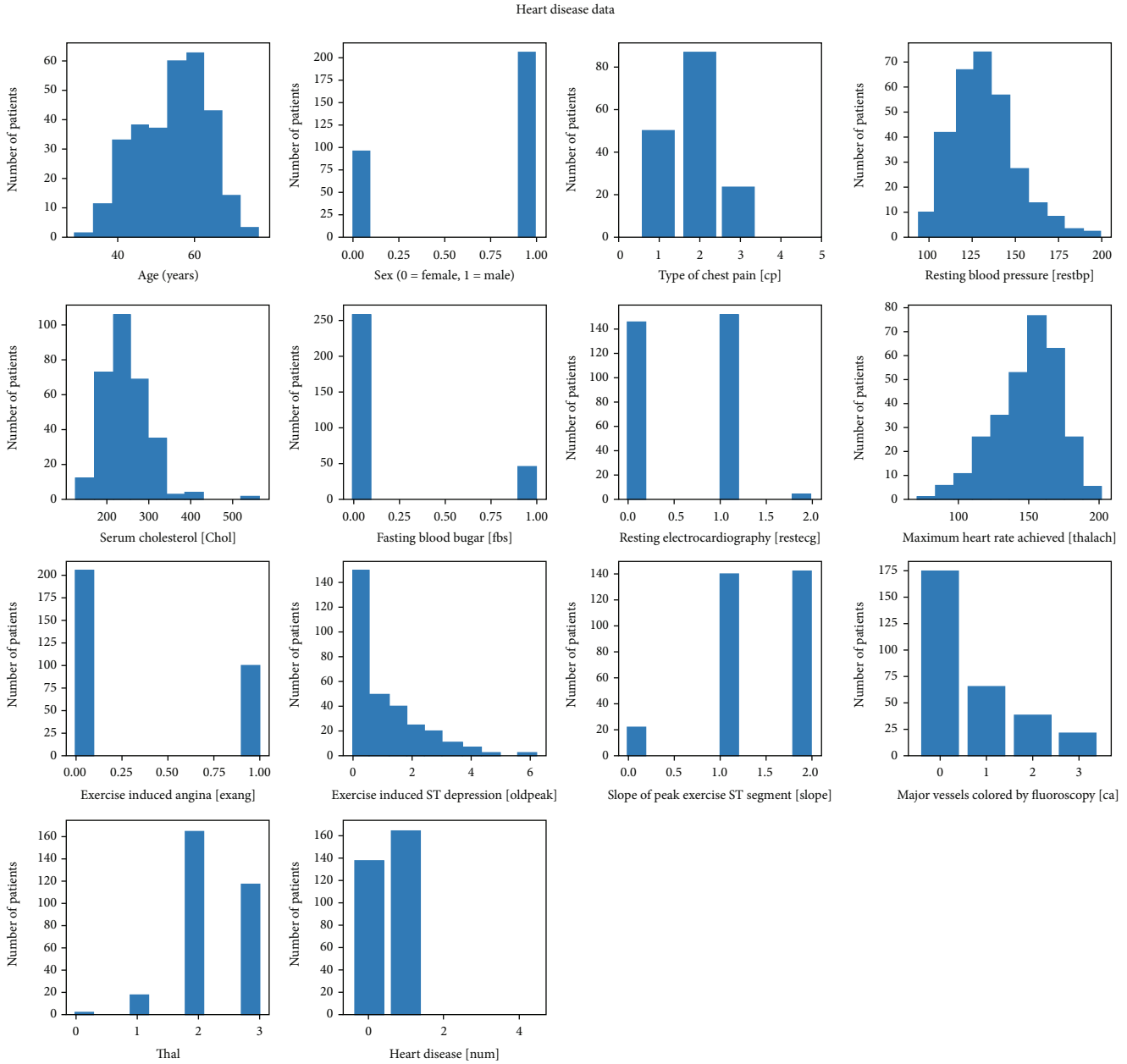
FIGURE 5: Histogram of the number of patients based on different attributes of Cleveland dataset.

classifier and FN presents the number of images from class $C_i$ which are wrongly allocated to other classes using classifier. FP presents the number of images belonging to class $C_i$ which are allocated to other classes. TN criterion is the number of images which do not belong to class $C_i$ nor allocated to this class using classifier.

*4.3. Results.* In this section, we investigate the proposed method's performance on two datasets with varying input resources. In the first dataset, data are numerical and extracted from Cleveland dataset. As it was previously mentioned, these types of data directly go into the step of feature selection as inputs. In this section, in order to show the influence of each attribute, the attributes of this dataset are examined. Figure 5 illustrates the histogram of the number of

patients per attribute. As it is evident, the amount of most attributes is imbalanced among patients.

Figure 6 shows the frequency of attributes according to the individuals' condition (healthy or sick). With respect to the aforementioned figure, it is certain that amounts of some of the attributes have more significant relationships with the condition of samples and show more separability toward individuals' conditions. This relationship and separability, however, is less noticeable in some of the attributes.

The system's performance can be influenced by choosing the right features. Three feature selection approaches are employed in this case: t-SNE, F-score, and CFS.

As stated in the proposed method, for the three classifiers SVM, RF, and GB, the extracted features based on t-SNE, F-score, and CFS methods have been used, respectively.
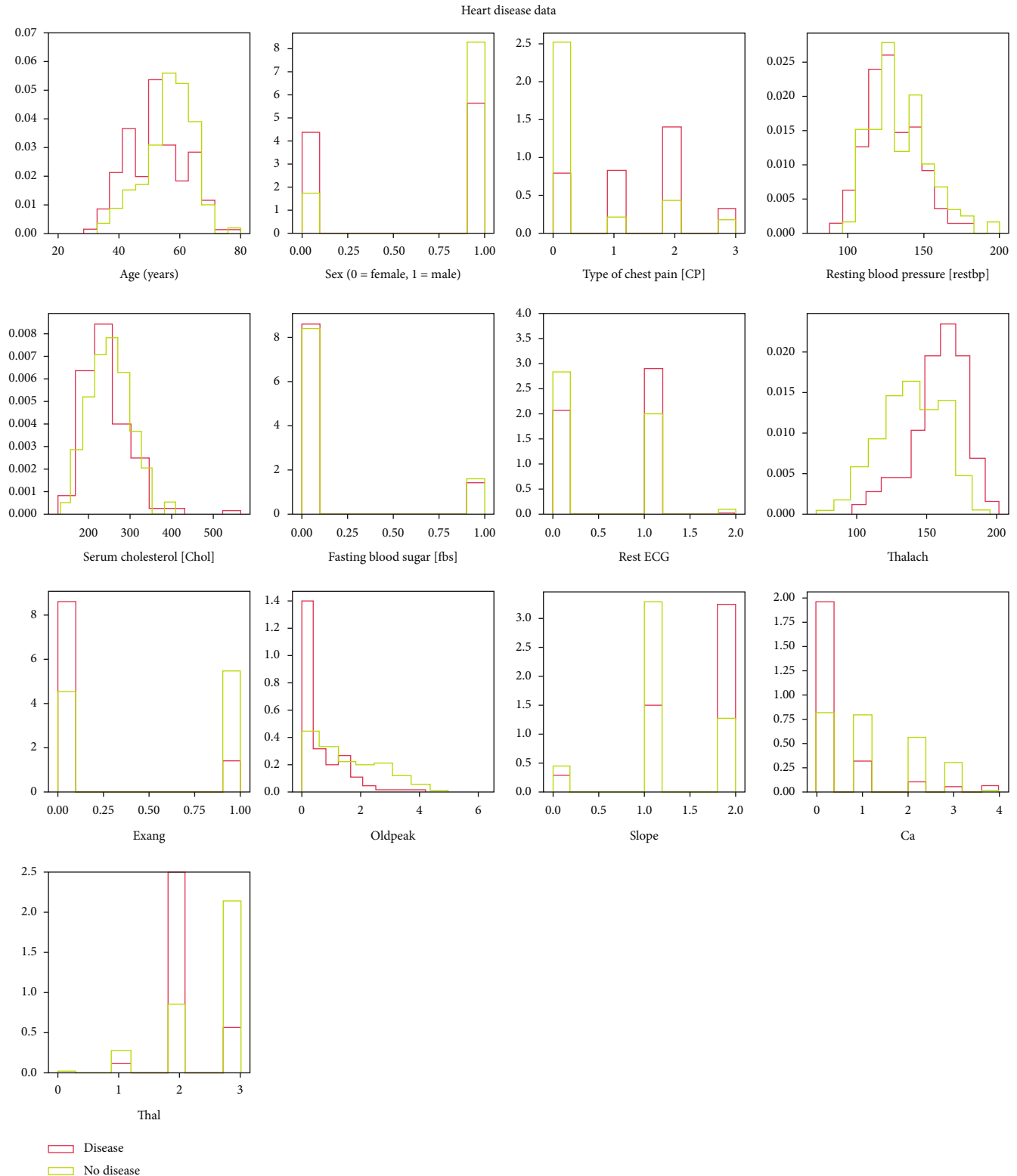
FIGURE 6: Frequency of different attributes in two conditions of the healthy and the sick in Cleveland dataset.

Each classifier's features are chosen using a validation set. Table 4 displays the outcomes of each approach in the validation set. It should be noted that the mean accuracy for 10 iterations is reported in this table. According to the results obtained in both types of input sources (image or numerical), the t-SNE feature has the best performance in the SVM classification, the F-score feature in the RF classification, and the CFS feature in the GB classification, respectively.

TABLE 4: Results of different classifiers based on different feature selection methods in validation set.

| Type of data | Method | Accuracy | | |
| | | t-SNE | F-score | CFS |
| --- | --- | --- | --- | --- |
| Numerical resources | SVM | 90.12 (±0.032) | 88.34 (±0.570) | 81.22 (±0.0078) |
| | RF | 85.12 (±0.0322) | 86.43 (±0.120) | 83.11 (±0.056) |
| | GB | 90.21 (±0.0167) | 78.45 (±0.077) | 88.25 (±0.110) |
| Image resources | 92.60(±0.570) | 93.12 (±0.061) | 95.65 (±0.018) | SVM |
| | 94.16(±0.420) | 96.32 (±0.045) | 89.32 (±0.130) | RF |
| | 95.78(±0.220) | 86.25 (±0.190) | 90.74 (±0.470) | GB |

TABLE 5: Results of the proposed method in comparison with other methods based on numerical resources in Cleveland dataset.

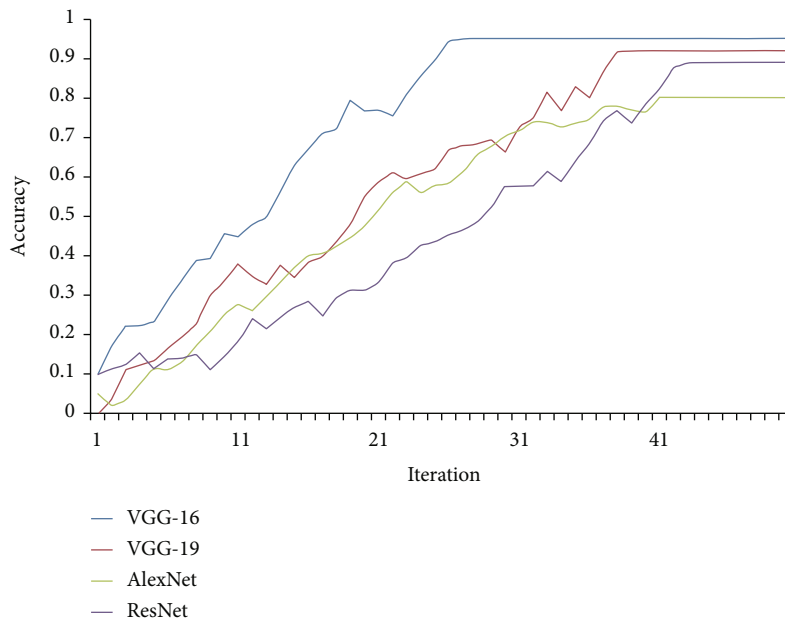| Method | Accuracy | Precision | Recall | Specificity | F-score |
| --- | --- | --- | --- | --- | --- |
| Logistic regression [8] | 83.3 | — | 86.3 | 82.3 | — |
| K-neighbors [8] | 84.8 | — | 85.0 | 77.7 | — |
| SVM [8] | 83.2 | — | 78.2 | 78.7 | — |
| Random forest [8] | 80.3 | — | 78.2 | 78.7 | — |
| Decision tree [8] | 82.3 | — | 78.5 | 78.9 | — |
| DL [8] | 94.2 | — | 82.3 | 83.1 | — |
| K-nearest neighbor [5] | 75.73 | — | — | — | — |
| Decision trees [5] | 72.45 | — | — | — | — |
| Random forest [5] | 75.73 | — | — | — | — |
| Multilayer perceptron [5] | 67.54 | — | — | — | — |
| Naïve Bayes [5] | 76.26 | — | — | — | — |
| Linear support vector machine [5] | 77.73 | — | — | — | — |
| Faster R-CNN with SE-ResNeXt-101 [4] | 98.00 | 96.16 | 98.47 | 96.02 | 97.58 |
| Proposed method | 98.7 | 96.61 | 99.18 | 96.65 | 98.48 |



FIGURE 7: The initial results based on different architectures in image classification.

TABLE 6: Results of the proposed method in comparison with other methods based on image resources.

| Method | Accuracy | Precision | Recall | Specificity | F-score |
| --- | --- | --- | --- | --- | --- |
| VGG-19 [4] | 95.23 | 93.96 | 94.80 | 93.19 | 95.58 |
| ResNeXt-101 [4] | 96.15 | 94.00 | 95.42 | 92.98 | 95.99 |
| Inception-ResNet-v2 [4] | 96.48 | 94.07 | 96.14 | 94.11 | 96.04 |
| SE-ResNet-101 [4] | 97.94 | 95.18 | 97.31 | 95.03 | 98.25 |
| Faster R-CNN with SE-ResNeXt-101 [4] | 99.15 | 98.06 | 98.95 | 96.32 | 99.02 |
| Proposed method | 99.84 | 98.64 | 99.61 | 97.19 | 99.12 |

TABLE 7: Results of the proposed method with the different voting method.

| Type of data | Accuracy | Method |
| --- | --- | --- |
| Numerical resources | 97.32 | Proposed method-weighted majority voting |
| | 98.7 | Proposed method-majority voting |
| Image resources | 98.00 | Proposed method-weighted majority voting |
| | 99.84 | Proposed method-majority voting |

The proposed method's results are shown in Table 5. As is obvious, the proposed method outperformed all of the other methods.

In the following, the performance of the proposed method based on the image resource is examined. It was noted in the proposed method section that the choice of convolutional network design affects the method's performance; hence, four different architectures were investigated: AlexNet, ResNet, VGG-16, and VGG-19. Training occurs solely in the fully connected layers, which is identical to an MLP network used for classification, and the convolutional layers needed to extract the feature are not learned due to the usage of transfer learning. The output layer has the same number of layers as the number of classes and is made up of two layers. The accuracy performance of each type of architecture with 50 repetitions to train fully connected layers is shown in Figure 7. This comparison shows that the VGG-16 architecture performs better, and as a result, this architecture has been used to extract features. The results show that a fully connected neural network (e.g., MLP) reports accuracy of 96.4% for image classification, and this approach can improve performance.

Table 6 shows the results of the proposed method, and as it can be seen, the proposed method has proved to have a suitable performance on these types of data.

In this section, the voting method is evaluated with two different perspectives. In the proposed method, the same weight for each classifier is considered. Table 7 shows the results obtained from the proposed method based on weighted majority voting with different weights for each classifier. As can be seen in the below table, the proposed method has performed better.

## 5. Conclusion

Many researchers have been interested in using ML to diagnose heart diseases in recent years. In this paper, IoMT is used for receiving input data based on numerical and image resources. In this paper, to diagnose the condition of heart disease patients, a hybrid method based on feature extraction from images using transfer learning, feature selection using t-SNE, F-score, and CFS, and classification using the combined output of three classifiers including GB, SVM, and RF using majority voting is used. It was indicated that feature selection or a subset of suitable features is a fundamental part of these types of systems and highly influences the accuracy of their performance.

## Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

[1] World Health Organization, *Cardiovascular Diseases*, WHO, Eneva, Switzerland, 2020.

[2] American Heart Association, *Classes of Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, https://www.heart.org/en/health-topics/heart-failure/what-is-heartfailure/classes-of-heart-failure.

[3] American Heart Association, *Heart Failure*, American Heart Association, Chicago, IL, USA, 2020, https://www.heart.org/en/health-topics/heart-failure.

[4] S. Manimurugan, S. Almutairi, M. M. Aborokbah et al., "Two-stage classification model for the prediction of heart disease using IoMT and artificial intelligence," *Sensors*, vol. 22, no. 2, p. 476, 2022.

[5] A. Kishor and W. Jeberson, "Diagnosis of heart disease using internet of things and machine learning algorithms," in

*Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, pp. 691–702, Springer, Singapore, 2021.

[6] S. Vishnu, S. R. J. Ramson, and R. Jegan, "Internet of medical things (IoMT)-an overview," in *2020 5th international conference on devices, circuits and systems (ICDCS)*, Coimbatore, India, 2020.

[7] R. P. Singh, M. Javaid, A. Haleem, R. Vaishya, and S. Ali, "Internet of medical things (IoMT) for orthopaedic in COVID-19 pandemic: roles, challenges, and applications," *Journal of Clinical Orthopaedics and Trauma*, vol. 11, no. 4, pp. 713–717, 2020.

[8] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 11 pages, 2021.

[9] A. G. Sorkhi, Z. Abbasi, M. I. Mobarakeh, and J. Pirgazi, "Drug–target interaction prediction using unifying of graph regularized nuclear norm with bilinear factorization," *BMC Bioinformatics*, vol. 22, no. 1, pp. 1–23, 2021.

[10] Z. Al-Makhadmeh and A. Tolba, "Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach," *Measurement*, vol. 147, p. 106815, 2019.

[11] S. Shalev-Shwartz and S. Ben-David, "Understanding machine learning," in *From 4eory to Algorithms*, Cambridge University Press, Cambridge, UK, 2020.

[12] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," in *Data Mining, Inference, and Prediction*, Springer, Cham, Switzerland, 2020.

[13] S. Marsland, *Machine Learning*, An Algorithmic Perspective CRC Press, Boca Raton, FL, USA, 2020.

[14] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, "Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 727–733, 2013.

[15] M. M. A. Rahhal, Y. Bazi, H. Alhichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340–354, 2016.

[16] G. Guidi, M. C. Pettenati, P. Melillo, and E. Iadanza, "A machine learning system to improve heart failure patient assistance," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 6, pp. 1750–1756, 2014.

[17] R. Zhang, S. Ma, L. Shanahan, J. Munroe, S. Horn, and S. Speedie, "Automatic methods to extract New York heart association classification from clinical notes," in *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1296–1299, Kansas City, MO, USA, November 2017.

[18] T. Zhang, A. H. Sodhro, Z. Luo et al., "A joint deep learning and internet of medical things driven framework for elderly patients," *IEEE Access*, vol. 8, pp. 75822–75832, 2020.

[19] R. J. S. Raj, S. J. Shobana, I. V. Pustokhina, D. A. Pustokhin, D. Gupta, and K. Shankar, "Optimal feature selection-based medical image classification using deep learning model in internet of medical things," *IEEE Access*, vol. 8, pp. 58006–58017, 2020.

[20] K. Saxena and R. Sharma, "Efficient heart disease prediction system," *Procedia Computer Science*, vol. 85, pp. 962–969, 2016.

[21] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1275–1278, Coimbatore, India, 2018.

[22] N.-S. Tomov and S. Tomov, "On deep neural networks for detecting heart disease," 2018, https://arxiv.org/abs/1808.07168.

[23] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An efficient convolutional neural network for coronary heart disease prediction," *Expert Systems with Applications*, vol. 159, article 113408, 2020.

[24] Y. Khourdifi and M. Bahaj, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 1, pp. 242–252, 2019.

[25] J. Patel, A. A. Khaked, J. Patel, and J. Patel, "Heart disease prediction using machine learning," in *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security*, pp. 653–665, Springer, Singapore, 2021.

[26] A. N. Repaka, S. D. Ravikanti, and R. G. Franklin, "Design and implementing heart disease prediction using Naives Bayesian," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 292–297, Tirunelveli, India, 2019.

[27] S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, "Improving heart disease prediction using feature selection approaches," in *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 619–623, Islamabad, Pakistan, 2019.

[28] L. Verma, S. Srivastava, and P. C. Negi, "A hybrid data mining model to predict coronary artery disease cases using non-invasive clinical data," *Journal of Medical Systems*, vol. 40, no. 7, p. 178, 2016.

[29] H. Sharma and M. A. Rizvi, "Prediction of heart disease using machine learning algorithms: a survey," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 5, no. 8, pp. 99–104, 2017.

[30] K. Uyar and A. İlhan, "Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks," *Procedia Computer Science*, vol. 120, pp. 588–593, 2017.

[31] A. G. Sorkhi, J. Pirgazi, and V. Ghasemi, "A hybrid feature extraction scheme for efficient malonylation site prediction," *Scientific Reports*, vol. 12, no. 1, pp. 1–16, 2022.

[32] S. M. R. Hashemi, H. Hassanpour, E. Kozegar, and T. Tan, "Cystoscopic image classification by unsupervised feature learning and fusion of classifiers," *IEEE Access*, vol. 9, pp. 126610–126622, 2021.

[33] M. A. Hall, *Correlation-Based Feature Selection for Machine Learning*, Citeseer, 1999.

[34] N. Sánchez-Marono, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection–a comparative study," in *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 178–187, Springer, Berlin, Heidelberg, 2007.

[35] S. Ding, "Feature selection based F-score and ACO algorithm in support vector machine," in *2009 Second International Symposium on Knowledge Acquisition and Modeling*, vol. 1, pp. 19–23, Wuhan, China, 2009.

[36] P. Jiang, W. Ning, Y. Shi et al., "FSL-Kla: a few-shot learning-based multi-feature hybrid system for lactylation site prediction," *Computational and Structural Biotechnology Journal*, vol. 19, 2021.

[37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[38] M. Rahimi and M. A. Riahi, "Reservoir facies classification based on random forest and geostatistics methods in an offshore oilfield," *Journal of Applied Geophysics*, vol. 201, article 104640, 2022.

[39] 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.

[40] J. Pirgazi, A. R. Khanteymoori, and M. Jalilkhani, "TIGRNCRN: trustful inference of gene regulatory network using clustering and refining the network," *Journal of Bioinformatics and Computational Biology*, vol. 17, no. 3, article 1950018, 2019.