



## Classification of Imbalanced Cardiac Arrhythmia Data

Cansu Ecemiş<sup>1\*</sup>, Neslihan Avcu<sup>2</sup>, Zekeriya Sarı<sup>3</sup>

<sup>1\*</sup> Dokuz Eylül University, Faculty of Engineering, Department of Electrical and Electronics Engineering, İzmir, Turkey, (ORCID: **0000-0001-5685-3895**), [cansu.ecemis@ogr.deu.edu.tr](mailto:cansu.ecemis@ogr.deu.edu.tr)

<sup>2</sup> Dokuz Eylül University, Faculty of Engineering, Department of Electrical and Electronics Engineering, İzmir, Turkey, (ORCID: 0000-0001-8481-2863), [neslihan.avcu@deu.edu.tr](mailto:neslihan.avcu@deu.edu.tr)

<sup>3</sup> Dokuz Eylül University, Faculty of Engineering, Department of Electrical and Electronics Engineering, İzmir, Turkey, (ORCID: 0000-0003-4070-9666) [zekeriya.sari@deu.edu.tr](mailto:zekeriya.sari@deu.edu.tr)

(2nd International Conference on Applied Engineering and Natural Sciences ICAENS 2022, March 10-13, 2022)

(DOI: 10.31590/ejosat.1083423)

**ATIF/REFERENCE:** Ecemiş, C., Avcu, N. & Sarı, Z. (2022). Classification of Imbalanced Cardiac Arrhythmia Data. *European Journal of Science and Technology*, (34), 546-552.

### Abstract

Arrhythmias are irregularities in the heartbeat and can be life-threatening. Early diagnosis of Cardiac Arrhythmia is quite crucial for saving patient lives. In this study, the main goal is to detect the presence of cardiac arrhythmia and classify it into 16 groups from the ECG recordings. The arrhythmia dataset in the UCI databank is used to apply different network structures for classification. The number of sample of each class are not the same in the dataset. The dataset has a very immoderate class distribution, and moreover, some classes don't exist. The imbalance condition between the classes causes a decrement in the performance of the classifier such as low classification accuracy. Also, in the cross-validation steps, the data is divided into groups each of which includes the same number of samples from the classes to overcome this difficulty in the classification. The samples of each class are divided into five groups to satisfy that condition. The training and test datasets are obtained as a combination of these groups. To deal with the imbalance condition in the dataset, first, some typical classification algorithms as Multilayer Perceptron (MLP), Support Vector Machine (SVM), Radial Basis Function (RBF), and Random Forest (RF) are used to classify the data. According to the precision and accuracy performance measurements of the classifiers for each data class, the nested classifier structures are constructed to improve the overall accuracy. The different structures are tried to obtain a better classifier performance. The results of classical and proposed four new ensemble networks are presented to compare their performance. The result shows that the random forest classifier has the best performance in terms of accuracy and, even with the ensemble network having the highest accuracy can be obtained almost the same performance results. For this reason, it is planned to increase the dataset and apply the different network structures for the enhancement of classifier performance as to future work.

**Keywords:** Classification, Cardiac Arrhythmia, Imbalanced Data, MLP, SVM, RBF, RF, Ensemble Learning.

## Dengesiz Kardiyak Aritmi Verilerinin Sınıflandırılması

### Öz

Aritmiler kalp atışındaki düzensizliklerdir ve hayati tehlike yaratabilir. Kardiyak Aritminin erken teşhisi, hastaların hayatlarını kurtarmak için oldukça önemlidir. Bu çalışmada, asıl amaç EKG kayıtlarından kardiyak aritmi varlığının saptanması ve 16 gruba sınıflandırılmasıdır. UCI veri bankasındaki aritmi veri seti, sınıflandırma amacıyla farklı ağ yapılarını uygulamak için kullanılmıştır. Veri setinde her sınıfın örnek sayısı aynı değildir. Veri seti oldukça düzensiz bir sınıf dağılımına sahiptir. Ayrıca, veri setinde bazı sınıflar mevcut değildir. Sınıflar arasındaki dengesiz dağılım, sınıflandırıcının performansında düşük sınıflandırma doğruluğu gibi azalmalara neden olur. Ayrıca çapraz geçerlilik sınaması (cross-validation) adımlarında, sınıflandırmadaki zorluğun üstesinden gelmek için veriler, her bir sınıftan aynı sayıda örnek içeren gruplara ayrılmıştır. Her sınıfın örnekleri, bu koşulu sağlamak için beş gruba bölünmüştür. Eğitim ve test veri setleri bu grupların bir kombinasyonu olarak elde edilmiştir. Veri kümesindeki dengesizlik durumuyla başa çıkmak

\*Corresponding Author: [cansu.ecemis@ogr.deu.edu.tr](mailto:cansu.ecemis@ogr.deu.edu.tr)

amacıyla, ilk olarak, verileri sınıflandırmak için Çok Katmanlı Algılayıcı (MLP), Destek Vektör Makinesi (SVM), Radyal Temel Fonksiyon (RBF) ve Rastgele Orman (RF) gibi bazı tipik sınıflandırma algoritmaları kullanılmıştır. Sınıflandırıcıların kesinlik ve doğruluk performans ölçümlerine göre, genel doğruluğu artırmak için, her bir veri sınıfı için iç içe sınıflandırıcı yapıları oluşturulmuştur. Daha iyi bir sınıflandırıcı performansı elde etmek için farklı yapılar denenmiştir. Klasik ve önerilen dört yeni topluluk ağının (ensemble network) performansları karşılaştırma için sunulmuştur. Karşılaştırma sonuçları, rastgele orman sınıflandırıcısının doğruluk açısından en iyi performansa sahip olduğunu ve en yüksek doğruluğa sahip topluluk ağıyla bile hemen hemen aynı performans sonuçlarının alınabileceğini göstermiştir. Bu nedenle ilerideki çalışmalarda sınıflandırıcı performansının iyileştirilmesi için veri setinin artırılması ve farklı ağ yapılarının uygulanması planlanmaktadır.

**Anahtar Kelimeler:** Sınıflandırma, Kardiyak Aritmi, Dengesiz Veri, MLP, SVM, RBF, RF, Topluluk Öğrenmesi..

## 1. Introduction

An early detection and accurate medical assistance is quite important for heart disease patients and can save human lives for heart diseases that can be life-threatening causing sudden death. The most widely used tool for diagnosing the heart disease is the electrocardiogram (ECG) [1]. However, sometimes it may be difficult for a doctor to look at these long duration ECG recordings and detect little irregularities in the wave. Therefore, using neural networks and machine learning algorithms for automating arrhythmia diagnosis can be very beneficial for doctors [2]. For this reason, the main aim is to detect the presence of cardiac arrhythmia and classify it into 16 groups from the ECG recordings. The name of 16 groups can be found in the Table (1) with class codes and number of instances.

## 2. Material and Method

### 2.1. Dataset

For this study, the dataset has been taken from the UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/Arrhythmia>. The data set contains 452 instances representing the medical record of different patients and 279 features representing age, weight, height, and patient's ECG-related data. The last column represents the label of the classes. There are 16 different classes. Class 01 refers to the absence of disease, ECG classes from 02 to 15 refer to different arrhythmia classes. Class 16 refers to the rest of the unclassified ones. In the dataset, about 0.33% of the feature values are missing and missing attribute values are distinguished with a question mark (?) [3]. In this study, the missing attribute values are exchanged with '0'. In Table (1), the class names of the dataset and corresponding class numbers can be found, and the distribution of the classes for the number of samples is in the Figure (1). This figure shows that the data is distributed in imbalanced way, and samples of classes 11, 12, and 13 are absent in the dataset. It can be a massive problem for multiclass classification. Furthermore, ECG data belonging to some patients have the properties of more than one arrhythmia class [3]. For this reason, it makes the classification harder. A stratified K-Folds cross-validation was used to deal with imbalanced data. The data was split into five chunks to be used in train and test by preserving the same percentage of samples for each class. Overall accuracy and confusion matrices were calculated by using the average of these chunks. In this paper, the data is normalized to obtain better performance at results and reduce the computational cost.

Table 1. Class table of dataset [3].

Class Code	Class	Number of Instances
01	Normal	245
02	Ischemic changes (Coronary Artery Disease)	44
03	Old Anterior Myocardial Infarction	15
04	Old Inferior Myocardial Infarction	15
05	Sinus Tachycardia	13
06	Sinus Bradycardia	25
07	Ventricular Premature Contraction (PVC)	3
08	Supraventricular Premature Contraction	2
09	Left Bundle Branch Block	9
10	Right Bundle Branch Block	50
11	First Degree Atrio Ventricular Block	0
12	Second Degree AV Block	0
13	Third Degree AV Block	0
14	Left Ventricle Hypertrophy	4
15	Atrial Fibrillation or Flutter	5
16	Others	22

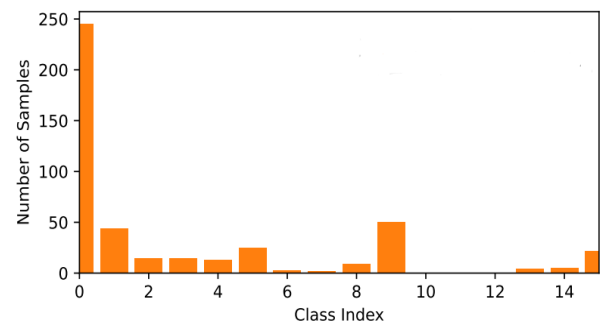


Figure 1. Distribution of dataset classes

## 3. Results and Discussion

In this section, the classification algorithms used and performance results of the algorithms will be discussed.

### 3.1. Multilayer Perceptron (MLP)

MLP is a kind of artificial neural network used as a classifier in this study. MLP was modelled as four hidden layers with 128, 32, 32 and 16 neurons by using the rectifier linear unit (ReLU) activation function. In the output layer, there are 16 neurons with

softmax activation function in order to obtain 16 classes. Categorical cross-entropy was used as a loss function. Furthermore, Adam optimizer was used with a learning rate of 0.001. By using this method, accuracy was obtained as 62%. The confusion matrix is as shown in Figure (2). It is obvious that MLP succeeded at determining classes of 0, 1, 2, 3, 8 and 9.

### 3.2. Support Vector Machine (SVM)

SVM is a supervised machine learning method used for classification and regression. In this study, SVM is used to classify the class of arrhythmias with normalization and the stratified K-Folds cross-validation method to increase the classifier's performance. The type of SVM kernel used is 'linear'. It can be seen in the Figure (3) that 0, 2, 3, 8, 9, and 14 classes can be determined correctly by using SVM. The accuracy is obtained as 70% in that case.

### 3.3. Radial Basis Function (RBF)

Radial Basis Function (RBF) is a special case of SVM with a 'RBF' kernel. The same methods used in SVM are applied to RBF to obtain better results. The accuracy in that case is 60%. The successful classes can be seen in confusion matrix in Figure (4) as 0, 2, and 8.

### 3.4. Random Forest (RF)

Random forests (RF) classifier is one of the most accomplished ensemble learning techniques that have been proven to be very popular and powerful techniques in machine learning for multi-dimensional classification [4]. By using the RF method, the accuracy is increased up to 73%. Furthermore, the overall confusion matrix in Figure (5) depicts that the RF method was successful in detecting the classes of 0, 1, 2, 3, 8, and 9.

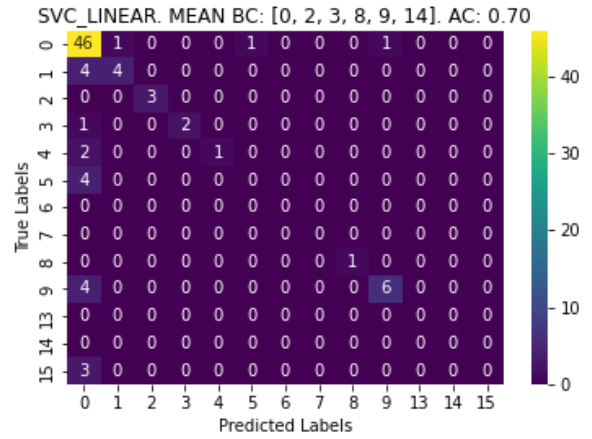


Figure 3. Overall Confusion Matrix of Linear SVM.

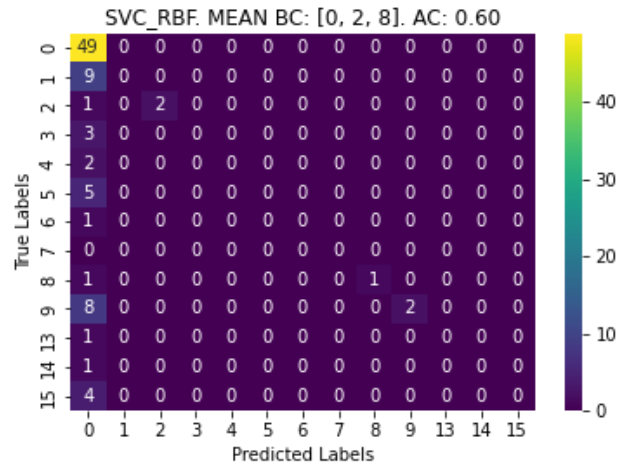


Figure 4. Overall Confusion Matrix of RBF

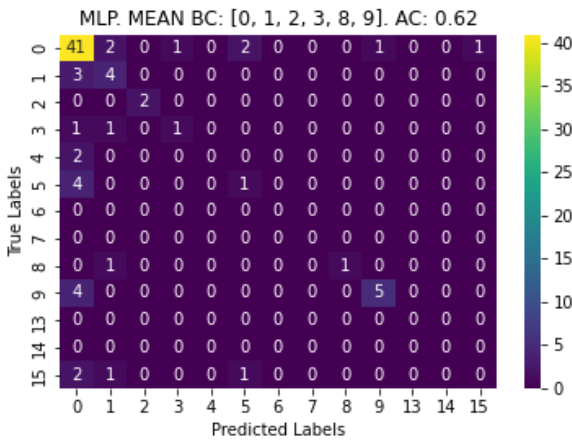


Figure 2. Overall Confusion Matrix of MLP

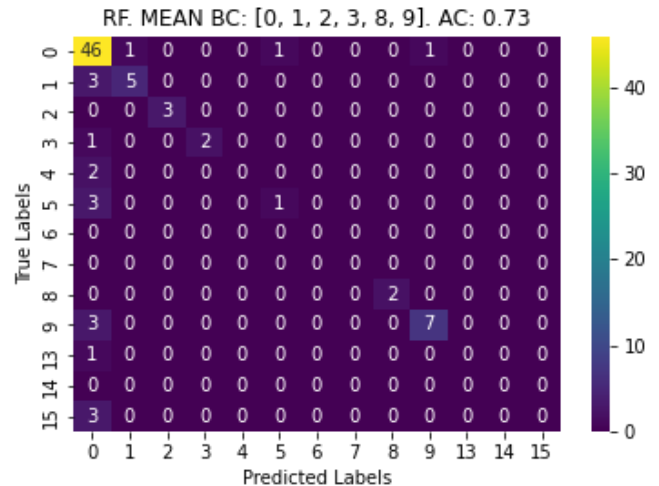


Figure 5. Overall Confusion Matrix of RBF

### 3.5. Ensemble Networks

Because of the unfair distribution of classes and inclusion of more than one arrhythmia type for some patient's ECG data, some ensemble network types are proposed. In Figure (6), (7), (8) and (9), proposed ensemble networks are illustrated.

In Figure (6), ensemble network consists of RF, SVM, MLP, and RBF. After obtaining the overall confusion matrices of MLP, SVM, RBF, and RF separately, the prediction results of classes are

compared. According to the performance of predictions on confusion matrices of classical network structures, the structure of ensemble network was designed. For each class, the classifier having the best performance for it was chosen and the priority of the classifier of each class was defined according to the number of the samples of the class. For instance, class 0 has its best performance on RBF, with the number of predicted true positive classes as 49 in the confusion matrix. The rest of the class outputs were decided by looking at the best value of true positive corresponding classes. In the nested structure, class 9 has the second priority level. In Appendix 1, the algorithm of the network can be found in detail.

The ensemble network in Figure (7) has the same logic as the network previously, but this network is constituted from only MLP and RF.

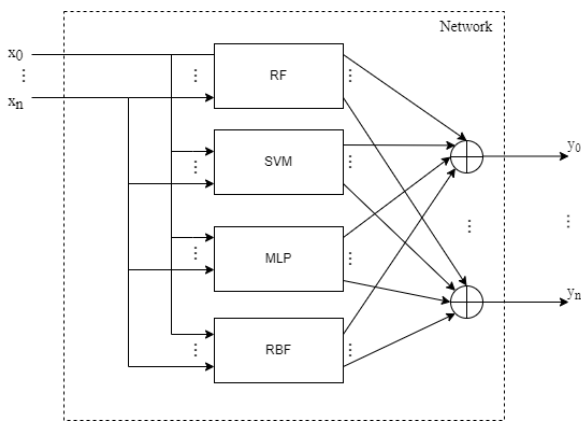


Figure 6. Proposed Ensemble Network Block Diagram of MLP, SVM, RBF and RF

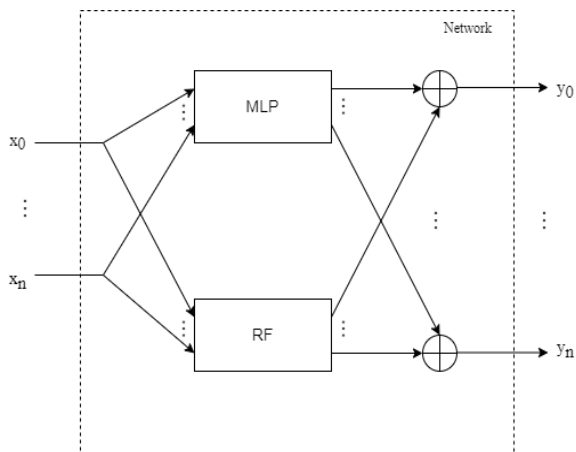


Figure 7. Proposed Ensemble Network Block Diagram of RF and MLP

In the previous confusion matrices, class 0 was predicted a lot instead of true ones. That means methods are not good at distinguishing between class 0 and others. In that case, an RF classifier is used to decide whether the class is 0 or not. If not, the class will be decided by another classifier block. In Figure (8), the other classifier block is RF, and an ensemble of MLP, SVM, RBF, and RF in Figure (9). In the second step, the classifiers are trained without class 0. The algorithm of the network in Figure (8) can be found in Appendix 2.

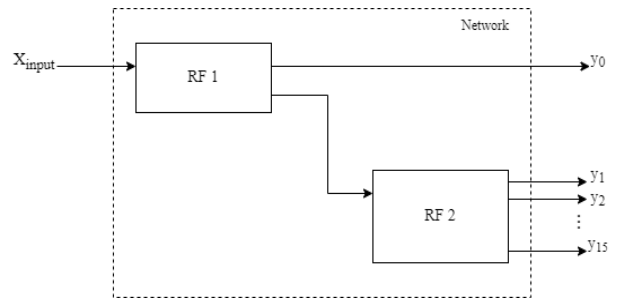


Figure 8. Proposed Ensemble Network Block Diagram of RF 1 and RF 2

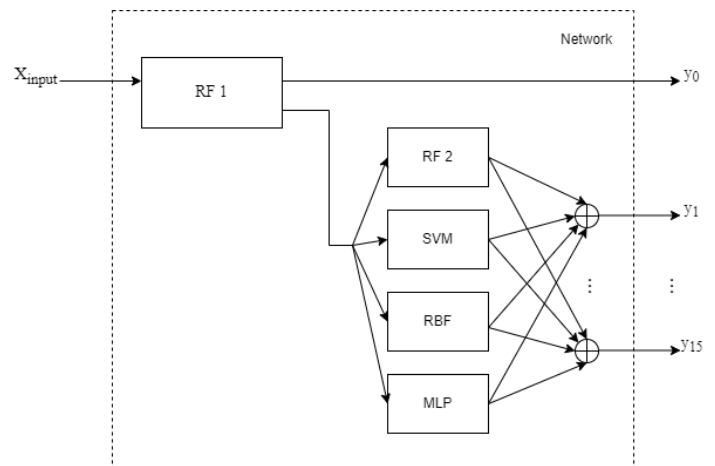


Figure 9. Proposed Ensemble Network Block Diagram of RF 1, RF 2, SVM, RBF and MLP

### 3.6. Performance Measures

In this subsection, the performance metrics have been used to analyse the results of classification will be discussed.

- Confusion Matrix: Confusion matrix represents how many elements were correctly predicted and how many were wrongly classified. For this scope, scientists invented several confusion matrix rates in the past [6] and they are given below.
- Accuracy:

$$\frac{TP + TN}{FP + FN + TP + TN}$$

- Precision:

$$\frac{TP}{TP + FP}$$

where true positive (TP) is the correct classification of positive class, False-negative (FN) is the incorrect prediction of the positive case, True negative (TN) is the correct classification of the samples in the negative class, and False-positive (FP) is the incorrect prediction of the negative case [5].

### 3.7. Performance Results

In this subsection, performance results of classification methods are given. In Table (2), the methods and corresponding

accuracy results are presented for MLP, SVM, RBF, and RF cases. In this case, the data was used without normalization. Training and testing data was split to 80% - 20%, and 5-folds cross-validation was applied. The best and worse results were obtained in the RF case as 72% and in the MLP case as 57.8%, respectively.

The methods and corresponding accuracies with normalized dataset are given in Table (3). Apart from MLP, SVM, RBF, and RF, proposed ensemble networks are investigated. In Table (3), all datasets were normalized, and stratified 5-folds cross-validation was applied in order to suppress the effects of imbalanced data. It seems that the methods used in the dataset have answered the purpose. The accuracies were increased for MLP from 57.8% up to 62%, for RF from 72% to 73%. Proposed ensemble networks accuracies also can be found in Table (3). The best result was attained as 72% from the proposed ensemble networks. Overall, the best accuracy result of 73% can be achieved by using only an RF classifier. The ensemble network variations that consist of RF, MLP, RBF, and SVM are not enough to increase the accuracy enough for such problematic data that is used in that study.

not limited to. In future work, other classification techniques can be used in an ensemble network; also, Synthetic Minority Oversampling Technique (SMOTE) can be used to overcome the imbalance of data.

**References**

Mustaqeem, A., Anwar, S. M., & Majid, M. (2018). Multiclass Classification of Cardiac Arrhythmia Using Improved Feature Selection and SVM Invariants. *Computational and Mathematical Methods in Medicine*, 2018, 1–10. <https://doi.org/10.1155/2018/7310496>

Gupta, A., Banerjee, A., Babaria, D., Lotlikar, K., & Raut, H. (2021). Prediction and Classification of Cardiac Arrhythmia. *Advances in Intelligent Systems and Computing*.

Guvener, H., Acar, B., Demiroz, G., & Cekin, A. (1997). A supervised machine learning algorithm for arrhythmia analysis. *Computers in Cardiology 1997*. <https://doi.org/10.1109/cic.1997.647926>

Azar, A. T., Elshazly, H. I., Hassaniien, A. E., & Elkorany, A. M. (2014). A random forest classifier for lymph diseases. *Computer Methods and Programs in Biomedicine*, 113(2), 465–473. <https://doi.org/10.1016/j.cmpb.2013.11.004>

Sharifrazi, D., Alizadehsani, R., Roshanzamir, M., Joloudari, J. H., Shoeibi, A., Jafari, M., Hussain, S., Sani, Z. A., Hasanzadeh, F., Khozimeh, F., Khosravi, A., Nahavandi, S., Panahiazar, M., Zare, A., Islam, S. M. S., & Acharya, U. R. (2021). Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomedical Signal Processing and Control*, 68, 102622. <https://doi.org/10.1016/j.bspc.2021.102622>

Chicco, D., Tötsch, N., & Jurman, G. (2021b). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, 14(1). <https://doi.org/10.1186/s13040-021-00244-z>

Table 2. Performance of Classification Methods without Cross-validation and Normalization.

Methodology	Accuracy(%)
MLP	57.8
SVM	65.7
RBF	59.8
RF	72

Table 3. Performance of Classification Methods

Methodology	Accuracy(%)
MLP	62
SVM	70
RBF	60
RF	73
Ensemble Network of MLP, SVM, RBF and RF	71
Ensemble Network of MLP, RF	66
Ensemble Network of RF 1 and RF 2	72
Ensemble Network of RF 1, RF 2, MLP, SVM and RBF	71

**4. Conclusions and Recommendations**

This study aimed to increase classification accuracy for the cardiac arrhythmia dataset with an unfair class distribution. The inclusion of more than one arrhythmia type for some patients' ECG data makes the dataset problematic. For this reason, some classification methods such as MLP, RBF, SVM, and RF and the different variations of ensemble networks of these methods have been used to classify cardiac arrhythmia. These classification methods' accuracy was enhanced using stratified K-fold cross-validation and data normalization. The classification methods have been compared, and the best accuracy was attained as 73% with RF. With proposed ensemble networks, the overall accuracy was enhanced to 72%. Still, the different ensemble network variations that consist of RF, MLP, RBF, and SVM are not good enough as RF classification for such problematic data used in that study. These techniques have increased accuracy overall but are



## Appendix 1

The algorithm of the ensemble network in Figure (6) is presented below.

### Inputs:

- The dataset  $\mathbf{X} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^{279}, y_i \in \{0, 1, \dots, 15\}, i = 1, 2, \dots, N\}$

### Initialize:

- Let  $m_0 \geq m_9 \geq m_1 \geq m_5 \geq m_{15} \geq m_2 \geq m_3 \geq m_4 \geq m_8 \geq m_{14} \geq m_{13} \geq m_6 \geq m_7$  where  $m_i$  be the number of samples in  $i^{th}$  class.
- Construct the networks *RF*, *SVM*, *MLP*, *RBF*.
- Train and test *RF*, *SVM*, *MLP*, *RBF*. Define the best classified labels,
  - $y_{RF} \leftarrow \{1, 3, 9\}$
  - $y_{RBF} \leftarrow \{0, 8\}$
  - $y_{SVC} \leftarrow \{2, 4, 5, 13, 14, 15\}$
  - $y_{MLP} \leftarrow \{6, 7\}$

for  $j \in \{1, 2, \dots, 5\}$  do

$\mathbf{X}_j^{test} \leftarrow \{(x_{j,i}, y_{j,i}) \mid (x_{j,i}, x_{j,i}) \in \mathbf{X} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1, 2, \dots, 1/5N\}$

$\mathbf{X}_j^{train} \leftarrow \{(x_{j,i}, y_{j,i}) \mid (x_{j,i}, x_{j,i}) \in \mathbf{X} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1/5N + 1, 1/5N + 2, \dots, N\}$

Using  $\mathbf{X}_j^{train}$ , fit *RF*, *SVM*, *MLP*, *RBF*.

Using  $\mathbf{X}_j^{test}$ , compute predictions *rf*, *svm*, *mlp*, *rbf* of *RF*, *SVM*, *MLP*, *RBF*, respectively.

for  $i \in \{1, 2, \dots, N/5\}$  do

if *rbf*[*i*] = 0 then

ensemble[*i*]  $\leftarrow$  *rbf*[*i*]

else if *rf*[*i*] = 9 then

ensemble[*i*]  $\leftarrow$  *rf*[*i*]

else if *rf*[*i*] = 1 then

ensemble[*i*]  $\leftarrow$  *rf*[*i*]

else if *svc*[*i*] = 5 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *svc*[*i*] = 15 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *svc*[*i*] = 2 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *rf*[*i*] = 3 then

ensemble[*i*]  $\leftarrow$  *rf*[*i*]

else if *svc*[*i*] = 4 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *rbf*[*i*] = 8 then

ensemble[*i*]  $\leftarrow$  *rbf*[*i*]

else if *svc*[*i*] = 14 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *svc*[*i*] = 13 then

ensemble[*i*]  $\leftarrow$  *svc*[*i*]

else if *mlp*[*i*] = 6 then

ensemble[*i*]  $\leftarrow$  *mlp*[*i*]

else if *mlp*[*i*] = 7 then

ensemble[*i*]  $\leftarrow$  *mlp*[*i*]

end if

end for

end for

return ensemble

## Appendix 2

The below algorithm belongs to the ensemble network depicted in Figure (8).

### Inputs:

- The dataset  $\mathbf{X} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^{279}, y_i \in \{0, 1, \dots, 15\}, i = 1, 2, \dots, N\}$

### Initialize:

- Let  $m_i$  be the number of samples in  $i^{th}$  class.
- The dataset  $\tilde{\mathbf{X}} = \{(\tilde{\mathbf{x}}_i, \tilde{y}_i) \mid (\tilde{\mathbf{x}}_i, \tilde{y}_i) \in \mathbf{X}, \tilde{y}_i \in \{1, \dots, 15\}, i = 1, 2, \dots, \tilde{N}\}$  where  $\tilde{N} = N - m_0$
- Construct the networks  $RF$  1,  $RF$  2.
- Train and test  $RF$  1 and  $RF$  2 with  $\mathbf{X}$  and  $\tilde{\mathbf{X}}$ , respectively.

for  $j \in \{1, 2, \dots, 5\}$  do

$\mathbf{X}_j^{test} \leftarrow \{(\mathbf{x}_{j,i}, y_{j,i}) \mid (\mathbf{x}_{j,i}, y_{j,i}) \in \mathbf{X} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1, 2, \dots, 1/5N\}$

$\mathbf{X}_j^{train} \leftarrow \{(\mathbf{x}_{j,i}, y_{j,i}) \mid (\mathbf{x}_{j,i}, y_{j,i}) \in \mathbf{X} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1/5N + 1, 1/5N + 2, \dots, N\}$

$\tilde{\mathbf{X}}_j^{test} \leftarrow \{(\tilde{\mathbf{x}}_{j,i}, \tilde{y}_{j,i}) \mid (\tilde{\mathbf{x}}_{j,i}, \tilde{y}_{j,i}) \in \tilde{\mathbf{X}} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1, 2, \dots, 1/5\tilde{N}\}$

$\tilde{\mathbf{X}}_j^{train} \leftarrow \{(\tilde{\mathbf{x}}_{j,i}, \tilde{y}_{j,i}) \mid (\tilde{\mathbf{x}}_{j,i}, \tilde{y}_{j,i}) \in \tilde{\mathbf{X}} \text{ where } j = 1, 2, \dots, 5 \text{ and } i = 1/5\tilde{N} + 1, 1/5\tilde{N} + 2, \dots, \tilde{N}\}$

Using  $\mathbf{X}_j^{train}$  and  $\mathbf{X}_j^{test}$ , fit  $RF$  1 and compute predictions  $rf1$ , respectively.

Using  $\tilde{\mathbf{X}}_j^{train}$  and  $\tilde{\mathbf{X}}_j^{test}$ , fit  $RF$  2 and compute predictions  $rf2$ , respectively.

for  $i \in \{1, 2, \dots, N/5\}$  do

if  $rf1[i] = 0$  then

$ensemble[i] \leftarrow rf1[i]$

else

$ensemble[i] \leftarrow rf2[i]$

end if

end for

end for

return  $ensemble$