Progress in Artificial Intelligence (2021) 10:333–347
https://doi.org/10.1007/s13748-021-00243-5

REGULAR PAPER

ECG beat classification using neural classifier based on deep autoencoder and decomposition techniques

Roguia Siouda[1] · Mohamed Nemissi[1] · Hamid Seridi[1]

Received: 17 July 2020 / Accepted: 26 March 2021 / Published online: 12 April 2021

Abstract

This paper proposes an ECG beat classification system based on deep autoencoder as feature extractor and a system of multiple neural networks as classifier. The objectives are as follows: First is simplifying the feature extraction step by applying the deep autoencoder, which permits defining high level features without neither pre-processing stage nor expert intervention. Second is enhancing the classification performance by decomposing the original multi-class problem into simpler binary subproblems and solving them using independent classifiers. Third is overcoming the problem of imbalanced data, by applying an oversampling method after the decomposition of the original problem. This allows adding synthetic samples according the number of training instances in each subproblem. To evaluate the proposed system, we conduct experiments on MIT-BIH arrhythmia dataset and we consider the recommendations of the Association for the Advancement of Medical Instrumentation, which defines five classes of interest. Furthermore, we perform two types of tests, i.e. intra- and inter-patient, and compare the obtained results with some of the state-of-the-art methods. We show that solving each subproblem independently can enhance the accuracy, sensitivity and specificity.

Keywords Multi-class · Multilayer perceptron · Stacked sparse autoencoders · ECG arrhythmia · Decomposition strategies

1 Introduction

Despite the recent advances, cardiovascular diseases remain one of the leading causes of death all over the world. An appropriate examination of the ECG constitutes an important tool for detecting cardiac arrhythmias especially in long-time recordings. The computer-aided systems (CAD) provide significant solutions that can help cardiologists in the diagnosis. However, the nonlinear and non-stationary nature of the ECG signals and the noise affecting them complicate their manipulation and require using sophisticated methods. These methods generally include three main steps: pre-processing, feature extraction and classification [1].

✉ Roguia Siouda
siouda.roguia@univ-guelma.dz

Mohamed Nemissi
nemissi.mohamed@univ-guelma.dz

Hamid Seridi
seridi.hamid@univ-guelma.dz

1 LabSTIC Laboratory, University of 8 Mai 1945 Guelma, POB 401, 24000 Guelma, Algeria

Artificial neural networks (ANN) are among the most common and frequently used models in medical classification [2] and ECG diagnosis. In the recent years, many interesting researches based on neural networks have been proposed. For example, the authors in [3] proposed an ECG denoising method based on a feed-forward neural network with three hidden layers. First, Geoffrey Hinton's method is applied to learn the neural network weights following initialization using a stack of restricted Boltzmann machines. Second, backpropagation algorithm is used to fine tune the weights. The authors in [4] proposed a robust method for identifying various cardiovascular diseases by using convolutional neural network (CNN) and multilayer perceptron (MLP). The MLP algorithm is used with four hidden layers and the CNN with four convolution layers. In [5], the authors developed a new method to detect arrhythmia using neural network (NN). First, short time Fourier transform (STFT) and the wavelet transform are used to extract efficient features. Second, neural network is used to distinguish the abnormal beats. Wavelet-based features had shown an improvement of accuracy over STFT features in classifying arrhythmia. In [6], the authors proposed an optimized model for arrhythmia classification using artificial neural network and grey wolf optimization (GWO). The weights of the neural network are optimized using GWO algorithm.

Deep learning (DL) is a current, rapidly growing area in machine learning [7,8]. It has been applied to a wide variety of fields, including image recognition, medical diagnostics and bioinformatics. DL approaches can be divided into two categories: supervised and unsupervised. Supervised models include multilayer perceptron's (MLP), convolutional neural networks (CNNs) and recurrent neural networks (RNNs)). Unsupervised models include self-organizing maps (SOMs), Boltzmann machines and autoencoders (AEs) [9,10].

Deep learning approaches provide the advantage of performing both feature extraction and classification within a single network. This remarkably overcomes the problems of extracting features. In ECG classification, DL models have been abundantly applied. Among them, the sparse autoencoders (SAE) have been efficiently used in several works. For example, Ozal Yildirim et al. [11] proposed a deep approach for the recognition of arrhythmic heartbeats. First, the convolutional autoencoder (CAE) is used to compress the ECG beats and obtain low-dimensional signals for each beat. Second, a long-short term memory (LSTM) network model is constructed to classify the coded signals. Jianli Yang et al. [12] presented an ECG arrhythmia classification method based on stacked sparse autoencoders (SSAEs) and SoftMax regression (SF) model. The SSAEs are employed to hierarchically extract high level features from huge amount of ECG data. The SF is then trained to serve as a classifier for discriminating six different types of arrhythmia heartbeats. M.M. Al Rahhal et al. [13] proposed an approach based on deep learning for the active classification of ECG signals. This approach includes two phases: First, a suitable feature representation from raw ECG data is automatically learned using stacked denoising autoencoders (SDAEs). Then, SoftMax regression layer is added on the top of the resulting hidden representation layer yielding the so-called deep neural network (DNN). Finally, an active learning (AL) criterion for selecting the most valuable ECG beats is applied to update the DNN weights. Ozal Yildirim et al. [14] designed a deep network structure that consists of 27 layers, including coders and decoders, to compress the ECG signals. The deep convolutional autoencoder (CAE) provides a representation of the low and high levels of signals in the hidden layers of the system. Therefore, the original signal can be reconstructed with minimal loss. Siti Nurmaini et al. [15] proposed a deep learning model for multi-class classification of arrhythmia. Throughout the pre-training process, denoising autoencoders (DAEs) and autoencoders (AEs) are staked to produce good feature representation; in the fine-tuning phase, deep neural networks (DNNs) are implemented as a classifier.

Other DL approaches for ECG classification include: deep bidirectional long-short term memory network (LSTM) based on wavelet sequences (DBLSTM-WS) for the classification of electrocardiogram (ECG) signals [16]; faster regions with a convolution neural network algorithm (faster R-CNN) [17].

Most of the above deep models include two parts: automatic feature extraction and classification. In this work, we also propose a classification model with two parts. In the first part, we use deep SAE to extract high level features. In the second part, i.e. classification, we use a system of multiple classifiers based on the decomposition strategies for multi-class classification problems. Indeed, the decomposition strategies have been successfully used in several real-world domains. They divide the original problem into many binary subproblems in order to reduce the complexity [18,19]. Among these methods, On-against-All (OAA) and One-against-One (OAO) strategies are the most commonly used [20–22]. In this work, we adopt both approaches (OAA and OAO) and we use MLPs as base classifiers. We adapt each MLP according to its corresponding subproblem. The training method and the number of the hidden neurons are then set according to the complexity of the subproblem. That is to say, in our systems, every subproblem in OAA method consists in classifying a type of heart beat against all the other types and, in OAO approach, every subproblem consists in classifying a type of heart beat against another type. Furthermore, to overcome the problem of imbalanced data, which highly occurs in most medical problems, we utilize the synthetic minority over-sampling technique (SMOTE) [23]. Therefore, by applying the decomposition strategies, the process of adding synthetic samples is impeccably performed according to the number of instances in each subproblem.

For the evaluation, we use the well-known MIT-BIH arrhythmia dataset and we apply the recommendations of the Association for the Advancement of Medical Instrumentation (AAMI), which defines five classes of interest: normal (N), ventricular (V), supraventricular (S), fusion of normal and ventricular (F) and unknown beats (Q). The AAMI also recommends the adoption of inter-patient test, i.e. the training and test beats should be taken from different patients [24,25]. We carry out both inter-patient and intra-patient tests and compare our results with the state-of-the-art methods.

The rest of this paper is organized as follows: Section 2 gives some background on the stacked sparse autoencoders and decomposition strategies. Section 3 describes the proposed method. Section 4 presents and analyses the experimental results conducted on MIT-BIH dataset. Finally, Sect. 5 concludes this paper.

2 Theoretical background

2.1 Sparse autoencoder (SAE)

An autoencoder is a neural network trained through an unsupervised learning algorithm to give the target values equal to the inputs ($y(i) = x(i)$). It aims to learn a function $h(x) \approx x$; therefore, the output $\widehat{x}$ is similar to the input x. Interesting information can thus be found for a limited number of hidden neurons. Autoencoder can be considered as a special type of deep learning used to reduce the inputs into a smaller representation.

A sparse autoencoder (SAE) aims to learn sparse features by adding a sparse penalty term inspired by the sparse coding [26]. This term is added to the cost function in such a way the learned features are not a just repetition of the inputs. The sparse penalty aims to minimize the number of 'active' hidden neurons. Generally, when a neuron's output value is '$1'$' it is active, and when its output value is '$0'$', the neuron is inactive. Suppose $a_j(x)$ denotes the activated neurons in the j^{th} hidden layer. In the process of feature learning, the activation value of a hidden-layer neuron is typically represented as $a = F(wx + b)$ where w is the matrix of weight and b is the vector of deviation. Therefore, the average value of the activation of this neuron is given by:

$$\rho_j = \frac{1}{n}\sum_{i=1}^{n}[a_j(x(i))] \tag{1}$$

Where n is the dimension of the feature space.

The sparsity can be performed by adding a regularization term denoting the difference between the average activation value, $\widehat{\rho_j}$, and a sparsity target value, ρ. This term can be done by the Kullback–Leibler divergence as follows:

$$\Omega_{\text{spar}} = KL(\rho \parallel \widehat{\rho_j}) = \rho\log\frac{\rho}{\widehat{\rho_j}} + (1-\rho)\log\frac{1-\rho}{1-\widehat{\rho_j}} \tag{2}$$

The cost function can be given as follows:

$$f = \text{MSE}(X - \widehat{X}) + a\Omega_{\text{spar}} + \beta\Omega_w \tag{3}$$

Where, $\text{MSE}(X - \widehat{X})$ is the mean squared error and Ω_w is the sum squared of all network weights.

Stacked sparse autoencoder(SSAE)

A system of stacked autoencoders is a deep learning neural network composed of several layers of sparse autoencoders, in which the outputs of each autoencoder are connected to the inputs of the next one. SSAE learning is based on a greedy layer-wise unsupervised training that trains each autoencoder independently [27].

Figure 1 gives an example of a stacked sparse autoencoder neural network. In the encoder part of this model, the signals are reduced to small dimensional vectors, and in the decoder part, they are reconstructed.

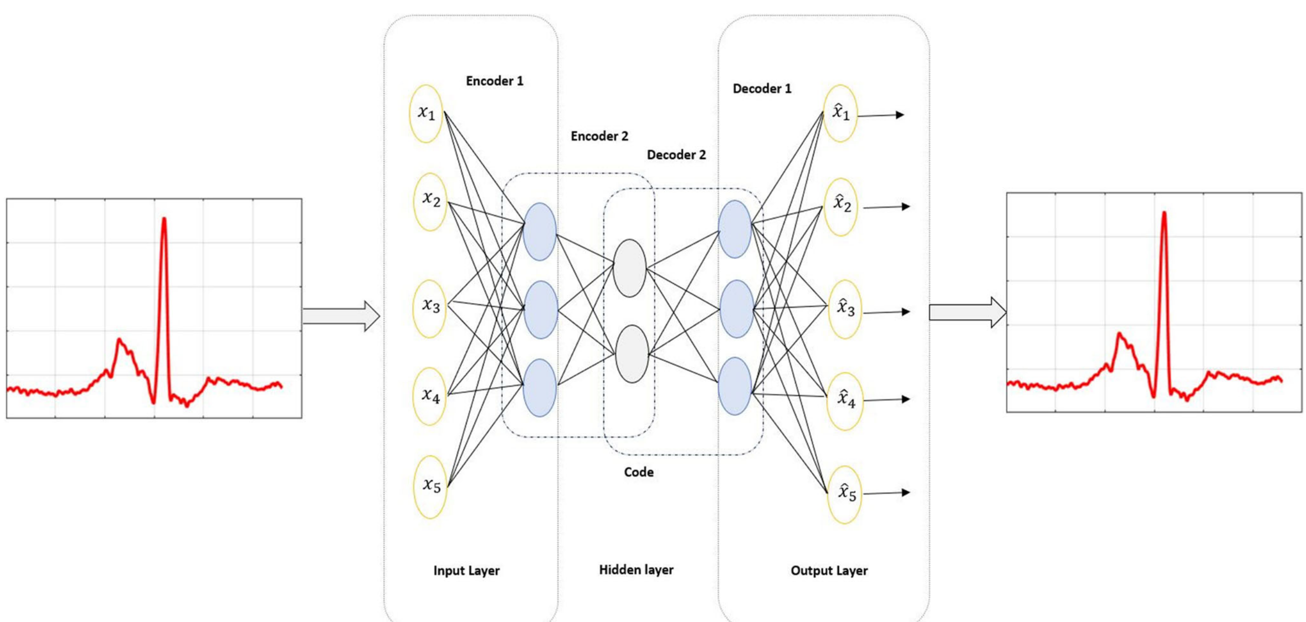


Fig. 1 Structure of a stacked sparse autoencoder network

2.2 Decomposition strategies

Most of the real-world applications include multi-class classification problems, which are more difficult than binary classification problems. An effective way to handle multi-class problems is to divide them into a set of simpler binary subproblems. After decomposition, the original problem is solved using an ensemble of classifiers, in which each classifier is independently built for one subproblem. To classify a new sample, it is presented to all classifiers and the final decision is made by combining the classifiers outputs. The importance of the decomposition methods has been abundantly discussed in the specialized literatures [28–36]. The benefits of these methods can be summarized as follows. First, they simplify the overall problem because binary problems as generally simpler than multi-class problems. Second, most of the existing algorithms were essentially designed for binary classification. Third, each classifier of the ensemble has its own architecture, parameters, learning algorithm and set of features. This generally provides better performances than using one multi-class classifier for the entire problem.

The decomposition can be performed using a variety of strategies; among them One Against All and One Against One are the most commonly used.

2.2.1 The one against all (OAA) decomposition strategy

The OAA decomposition method converts a multi-class classification problem into a set of subproblems that every one of them aims at classifying one class against all the other classes. Consequently, the number of subproblems is equal to the number of classes. The whole training instances are used in all subproblems.

For the aggregation, the Maximum confidence strategy is the most common and simple used method. The output class is simply taken from the classifier with the greatest response [29]:

$$\mathrm{Class}(X) = \arg \max_{i=1\ldots K} Z_i \tag{4}$$

where Z_i is the output of the classifier corresponding to the subproblem: class i against all the other classes.

2.2.2 The one against one (OAO) decomposition strategy

The OAO decomposition method is based on classifying each class against every one of the other classes. It transforms a K-class problem into $K(K-1)/2$ subproblems. Therefore, the number of binary problems in this method is larger than in the OAA method, but each subproblem involves less training data as only the samples of two classes are considered.

Among the aggregation methods used in OAO, the Majority voting and the Weighted voting are the most commonly used.

Majority voting

Every binary classifier gives the predicted class as one vote. Then, the votes received by each class are counted and the class with the largest number of votes is considered as the final decision. Formally, the decision rule can be described as follows [29]:

$$\mathrm{Class}(X) = \arg \max_{i=1\ldots K} \left(\sum_{1 \le j \ne i \le K} I(Z_{ij} > Z_{ji}) \right) \tag{5}$$

With $I(.)$ the standard indicator function which evaluates to one when its argument is true and to zero otherwise.

Where Z_{ij} is the output of the classifier corresponding to the subproblem: class i against class j.

Weighted voting strategy

In this method, every binary classifier vote on both classes. The weight for the vote is determined by the confidence of the classifier predicting the class. The resulting class is the class with the largest sum value. Hence, the decision rule is [29,30]:

$$\mathrm{Class(X)} = \arg \max_{\mathrm{i=1\ldots K}} \sum_{\mathrm{1 \le j \ne i \le K}} \mathrm{Z_{ij}} \tag{6}$$

Other methods include decision directed acyclic graph [31], learning valued preference for classifiers [32], non-dominance criterion [33], binary tree of classifiers [34], probability estimates [35] and Nesting OAO [36].

3 The proposed method

3.1 Main idea and motivations

The main contribution of this paper is twofold: (i) to use deep stacked autoencoder for automatic extraction of the features and (ii) to divide the original multi-class problem into simpler subproblems and solving them using a system of multiple classifiers. The motivations are then the followings:

We make use of the capacity of the stacked sparse autoencoders in representing data by high level features. In addition, these systems permit avoiding pre-processing and feature selection steps.

We make use of the advantages of the decomposition strategies, like simplifying the original problem and enhanc-

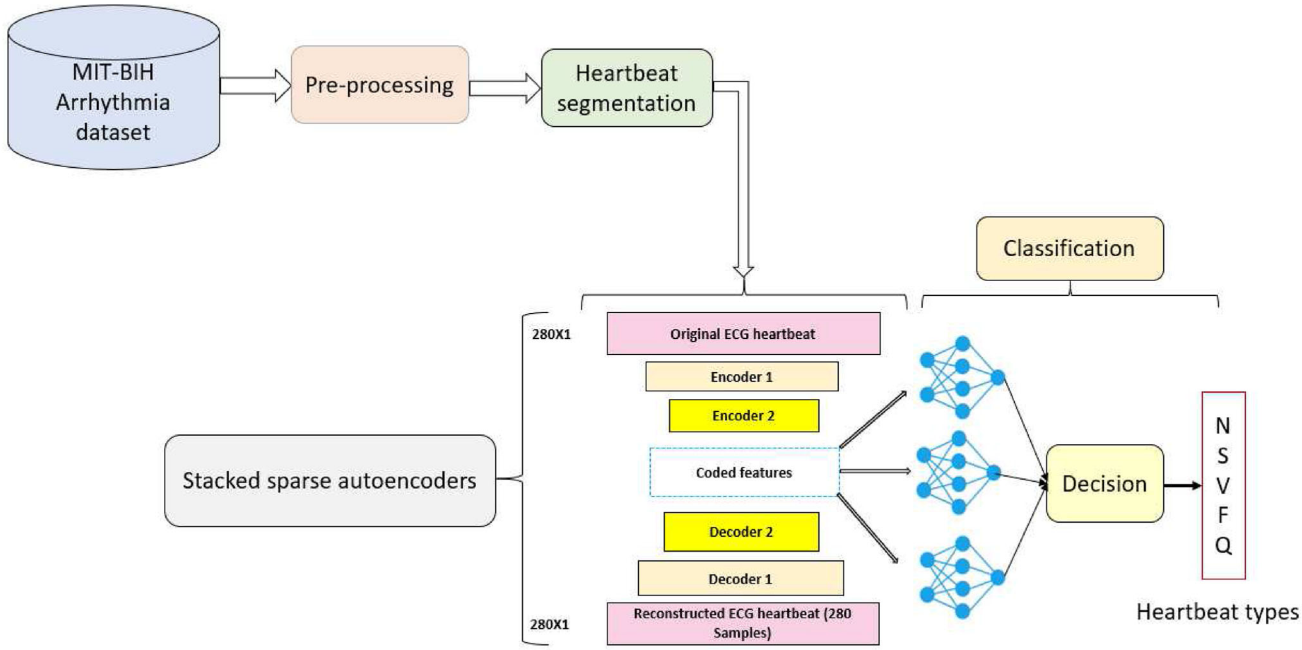


Fig. 2 Principal scheme of the proposed system

ing the performance by using an ensemble of different classifiers.

We apply the process of adding synthetic samples after the decomposition. In this way, the number of added instances is performed according to the number of training instances in each subproblem. Indeed, in the ECG classification problems, the data are highly imbalanced and the application of oversampling methods to each subproblem independently permits optimizing the number of training samples.

According to the most commonly used decomposition strategies for multi-class problems, i.e. OAA and OAO, we proposed two systems. In this works, we use MLPs as base classifiers.

3.2 Architecture

Figure 2 shows the principal scheme of the proposed system. First, the SSAEs are used to extract high-level features from the raw ECG signals (280 features in each heart beat) and then the coded features are used as inputs to the ensemble of MLPs. Every new sample (ECG beat) is presented to all MLPs, and the decision is made according to their outputs.

3.3 Training process

After the beats segmentation, the training of the proposed system is performed in two steps:

(i) Training the autoencoders, one at a time, using unsupervised data.

(ii) Decomposition of the multi-class problem and training the MLPs using supervised data.

3.3.1 Training the autoencoders

The autoencoders are powerful tools for re-encoding data, in which the input is encoded by the network to concentrate exclusively on the most important features. In this work, we use SSAEs to extract features form raw ECG data.

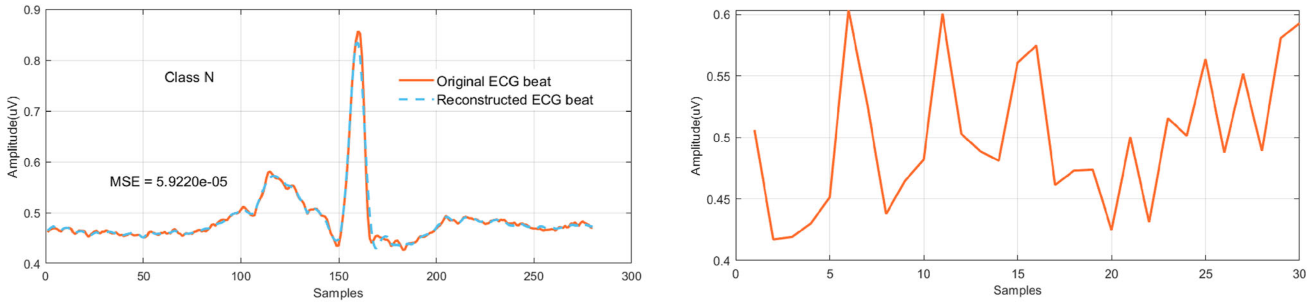
Figure 3 displays an example of ECG beat (from class N) and its corresponding coded features obtained at the encoding stage. In this example, the original signal is coded using 30 features. We can note that these features permit good reconstruction of the original beat.

Figure 4 illustrates some examples (from all classes) of the original and reconstructed ECG signals using the SSAEs. We can note that the original and reconstructed signals are almost equals. This shows the efficiency of the SSAEs coding performances on different beat types.

3.3.2 Decomposition of the multi-class problem and training the MLPs

The main contribution of this paper consists in decomposing the entire ECG beats classification problem into simpler subproblems, then solving each subproblem using a different classifier.

In this work, we consider the AAMI recommendations, that defines five classes of heartbeats. Therefore, the classification task is a 5-class problem. According to the most com-



(a) Original and reconstructed ECG beats

(b) The coded features

Fig. 3 An example of ECG signal coding using SSAE (from class N)

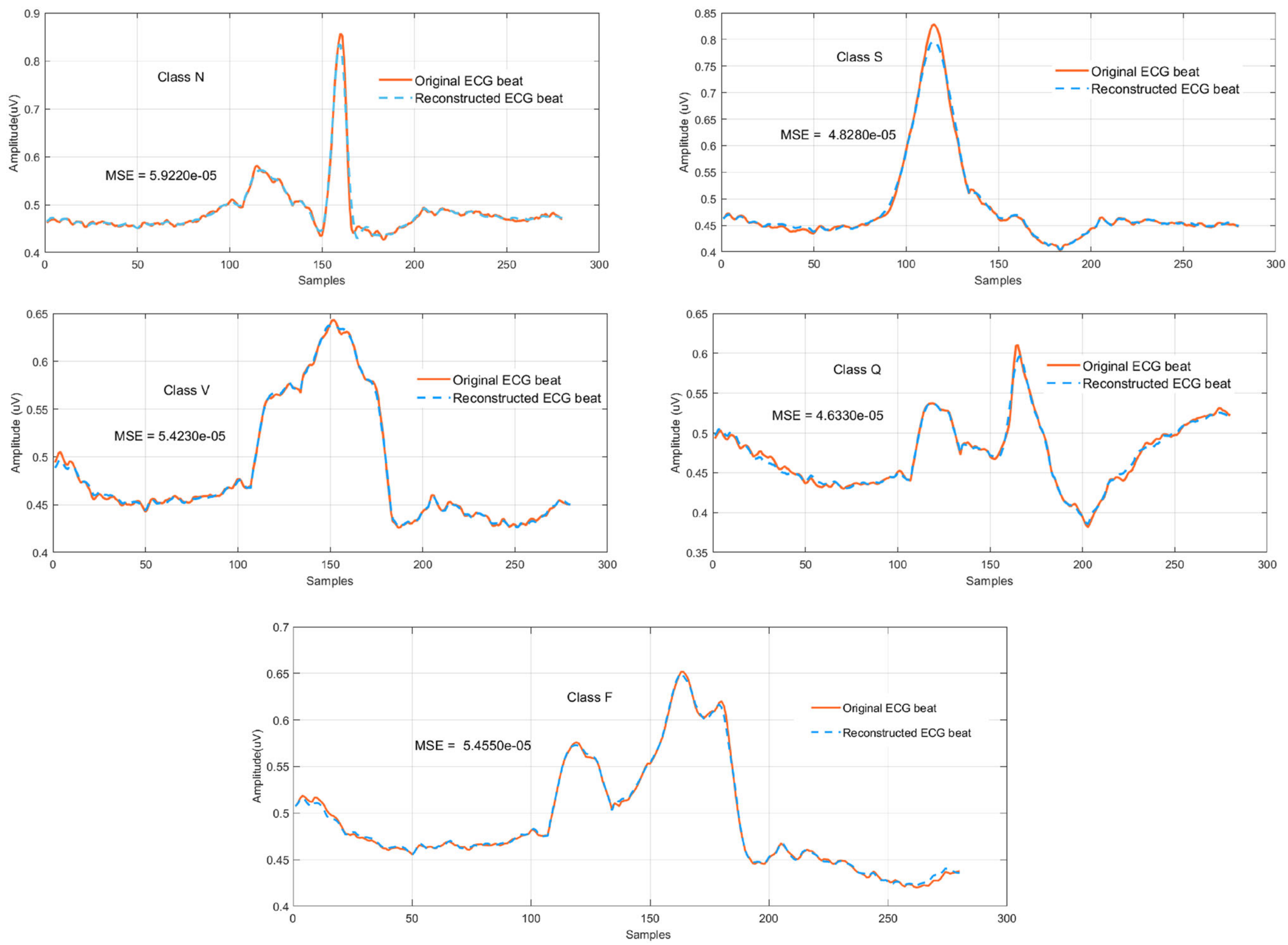


Fig. 4 Some examples of the original and reconstructed signals obtained using SSAEs. The original ECG signals (in red) and the reconstituted ECG signals (in blue) are superimposed

monly used strategies for decomposing multi-class problems, i.e. OAA and OAO, we proposed two models: OAA-MLP and OAO-MLP systems.

First model: OAA -MLP system

This system is based on OAA method, in which every subproblem consists in classifying a class against all the other classes. For a 5-class heartbeat classification problem, the proposed OAA-MLPs system includes five binary neural networks (Net_1, Net_2, ..., Net_5). Each network, Net_i, is trained

Table 1 Types of heartbeats in the MIT-BIH dataset recommended by AAMI

AAMI heartbeats	MIT-BIH heartbeats
Normal (*N*)	Normal beat (*N*)
	Left and right bundle branch block beats (*L*, *R*)
	Atrial escape beat (*e*)
	Nodal (junctional) escape beat (*j*)
Supraventricular ectopic beat (*S*)	Atrial premature beat (*A*)
	Aberrated atrial premature beat (*a*)
	Nodal (junctional) premature beat (*J*)
	Supraventricular premature beat (*S*)
Ventricular ectopic beat (*V*)	Premature ventricular contraction (*V*)
	Ventricular escape beat (*E*)
Fusion (*F*)	Fusion of ventricular and normal beat (*F*)
Unknown beat (*Q*)	Paced beat (/)
	Fusion of paced and normal beat (*f*)
	Unclassifiable beat (*U*)

to classify an ECG beat class (B_i) against all the other classes. Net_i has one output $Z_i(X)$ to indicate whether the presented ECG beat, X, belongs to the class (B_i) or not.

To classify a new beat, it is presented to all the networks of the system. The decision is then given by combining their outputs. We use the max rule Eq. 4, which is the most commonly used. The predicted class is thus given by the network that provides the highest output.

Second model: OAO-MLP system

The proposed OAO-MLP system is based on OAO method, which considers all possible pairs of classes. For a 5-class heartbeat classification problem, the proposed system includes ten binary neural networks: Net_{ij} , $i = 1...5 - 1$, $j = i + 1...5$. Each network, Net_{ij}, is trained to classify the beat class (B_i) against the beat class $\left(B_j\right)$. Net_{ij} has an output $Z_{ij}(X)$ indicating whether the presented beat, X , belongs to the class B_i or B_j. The training of Net_{ij} is based only on beats from classes B_i and B_j.

To classify a new beat, it is presented to all the networks of the system. The decision is then given by combining their outputs. We use the majority voting Eq. 5, which is the most commonly used. Each binary network therefore gives a vote for a predicted class, and the votes received by each class are counted. The class with the largest number of votes is chosen as the output class.

4 Tests and Experiments

4.1 Dataset description

To evaluate the performance of proposed classifier, we carried out tests on PhysioNet MIT-BIH Arrhythmia dataset [37,38]. This dataset involves ECG signals collected at the sampling rate of 360Hz for 48 distinct patients. There are two ECG leads in each record: lead II and lead V1. The lead II is usually utilized in the literature to identify heartbeats. Similarly, we used this lead in our work. This dataset is suggested by the American Association of Medical Instrumentation (AAMI) [39], as it contains the five major classes of arrhythmias outlined in Table 1.

In this study, two types of tests were considered: inter-patient and intra-patient. In the intra-patient paradigm, we used the holdout approach to evaluate the generalization capacities. The dataset is randomly divided into two parts: the training part, 80% of data, and the test part that consists of the remaining 20%. In this test, the heartbeat samples of the same patient may be in the training and test sets.

In the inter-patient paradigm, the training and test sets are built from different patients. In this work, we adopt the protocol proposed by de Chazal et al. [25], which has been widely adopted in the literature [24,40]. According to this protocol, the heartbeats from the MIT-BIH dataset (44 records in accordance with AAMI) are split into two sets of records: Dataset1 = 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223,230 and Dataset2 = 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234.

Dataset1 is used to train the classification systems and Dataset2 is used for the test. Within this separation, there is no concern about including the heartbeats from the same patient in both training and test sets.

Table 2 illustrates the number of beats in each class in both Intra- and inter-patient tests.

4.2 Evaluation metrics

The metrics recommended by AAMI for arrhythmia classification methods are: Sensitivity (Se), Positive predictive (+P), False positive rate (FPR), Specificity (Spe) and Overall accuracy (Acc). In this work, we use the Overall accuracy, Sensitivity (Se), Specificity (Spe) and Positive predictive (+P). These metrics are given as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \quad (7)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \quad (9)$$

$$\text{Positive predictive} = \frac{TP}{TP + FP} \times 100\% \quad (10)$$

where TP: true positive, FN: false negative, TN: true negative, and FP: false positive.

4.3 Intra-patient test

4.3.1 Results using OAA-MLP system

As mentioned above, we used an oversampling method, i.e. SMOTE, to overcome the problem of imbalanced data. The application of SMOTE in OAA strategy is performed by adding samples to the minor class in each subproblem. Since each subproblem consists in classifying one class against the others, this class is labelled as positive class, and all the others are considered as one class and labelled as negative. The process of adding samples is applied to the minor class whether it is the positive or the negative one. For example, in the subproblem: N against All, N is the positive class and all the others constitute the negative class. Although the negative class contains several classes, but it remains the minor class and SMOTE method is applied to increase the number of its instances.

To evaluate the proposed model, we first analysed the effect of the number of hidden layers in the stacked autoencoder and the number of features. Therefore, we performed several tests with different structures. Figure 5 shows the obtained results. We notice that using two layers and thirty features provided the best performances.

Table 3 displays the performances (accuracy, sensitivity and specificity) and the number of hidden neurons corresponding to each classifier in the OAA-MLP system. We

Table 2 The number of samples in each class

Heartbeat class	Number of heartbeats intra-patient	inter-patient Dataset1	Dataset2
N	90502	45798	44198
S	2777	941	1836
V	7226	3782	3217
F	802	414	388
Q	8031	7	7
Total	109338	50915	94646

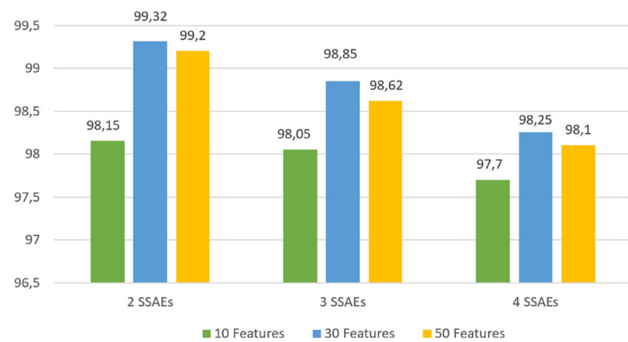


Fig. 5 Classification accuracies of OAA-MLP system obtained using different numbers of features and autoencoders

independently trained each classifier using different parameters and structures. We note that the most difficult subproblem is the $1st$ one that correspond to: N vs. All. Its corresponding network was trained with larger number of hidden neurons. On the other hand, the other networks provided good results with less hidden neurons and fewer iterations.

The results obtained on the test data using the full OAA-MLP system and a single MLP are shown on Table 4. The sensitivity (Se) and the Positive predictive (+P) are detailed for each class. We note that the system of multiple MLPs provided an accuracy of 99.32% while the single MLP provided only 97.83%. This indicates that using a system of multiple MLPs considerably enhances the accuracy. The enhancement can also be noted in terms of sensitivity.

Indeed, in Table 4, when reading the sensitivities of different classes obtained using the single MLP, it is clear that they are proportional to the number of instances in each class. Class N has the highest sensitivity (99.2%) and classes S and F have the smallest (84.4% and 82.3%, respectively). The application of the system of multiple MLPs improved the sensitivity of all classes.

4.3.2 Results using OAO-MLP system

The application of SMOTE in OAO strategy is performed by adding samples to the minor class in each subproblem, i.e. for each pair of classes the number of instances in the minor class

Table 3 The classification results of each classifier in OAA-MLP system on intra-patient paradigm

# Classifiers	Subproblems	# Hidden neurons	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	N vs. All	40	98.65	99.30	95.50
2	S vs. All	24	99.80	92.06	99.98
3	V vs. All	23	99.60	97.25	99.88
4	F vs. All	26	99.72	95.68	99.86
5	Q vs. All	25	99.86	98.66	99.96

Table 4 Performance of OAA-MLP system and single MLP on intra-patient paradigm

Classifiers	Accuracy (%)	Class *N*		Class *S*		Class *V*		Class *F*		Class *Q*	
		Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)
Single MLP	97.83	99.2	98.7	84.4	87.3	93.5	95.1	82.3	91.6	96.9	96.6
OAA-MLP system	99.32	99.73	99.52	94	95.93	97.74	97.84	95.42	96.52	98.73	99.24

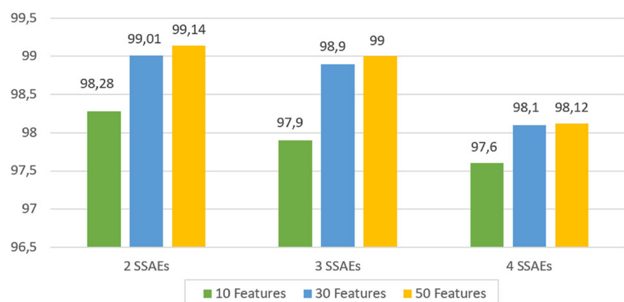


Fig. 6 Classification accuracies of OAO-MLP system obtained over MIT-BIH Arrhythmia using different numbers of features and autoencoders

is increased to reach the major class. Therefore, the process of oversampling is applied in an appropriate manner, especially to subproblems with small number of instances in each class.

Figure 6 shows the results obtained using OAO-MLP system and a SSAE with different structures and different numbers of features. We note that the best performance was also provided by two hidden layers. With regards to the number of features, we note that fifty features gave slightly better results than thirty.

Table 5 displays the performances (accuracy, sensitivity and specificity) and the number of the hidden neurons corresponding to each classifier in the OAO-MLP system. Each classifier was independently trained with different parameters and structures. The most difficult subproblems are the $5th$,$6th$ and $8th$ that correspond, respectively, to S vs. V, S vs. F and V vs. F. The networks corresponding to these subproblems were trained with larger number of hidden neurons and more iterations compared to the other networks.

Table 6 illustrates the classification results obtained on the test data using an OAO-MLP system and a single MLP. The system of multiple MLPs provided an accuracy of 99.14%, whereas the single MLP provided only 97.93%.

We can also note from Table 6 that the sensitivities obtained using the single MLP are proportional to the number of instances in each class. Class N had the highest sensitivity (99.3%) and class F had the smallest (82.8%). The application of the system of multiple MLPs improved the sensitivity of all classes. This indicates that the system of multiple MLPs considerably outperformed the single MLP.

We have assessed the effectiveness of SMOTE on both original and divided multi-class problem. In both cases, we have compared the results obtained with and without this method. Table 7 illustrates the classification results obtained using single MLP, OAA-MLP and OAO-MLP systems; SMOTE enhanced the performances of all classifiers. We can also conclude that the performances can first be enhanced with the decomposition and then be further enhanced using SMOTE.

It is worth mentioning that we have found that the decomposition using OAA gave better results than OAO. This can be explained by the fact that OAA can better capture the relation between all classes. In fact, in OAA, each subproblem classifies a class against all the other classes and uses all the training samples from all classes.

4.3.3 Comparison with other works

Table 8 compares the proposed models with other works applied on the same dataset (MIT-BIH arrhythmia) with intra-patient test. These works include a variety of machine learning and deep learning techniques, i.e. CNN, neural networks, support vector machine (SVM), ELM, etc.

We note that our models outperformed all these works except the work of S. Mousavi et al. [24], in which the authors

Table 5 The classification results of every classifier in OAO-MLP system on intra-patient paradigm

# Classifiers	Subproblems	# Hidden neurons	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	N vs. S	26	99.79	99.98	91.37
2	N vs. V	25	99.77	99.86	98.50
3	N vs. F	26	99.73	96.28	99.84
4	N vs. Q	24	99.83	99.95	98.43
5	S vs. V	40	99.11	98.09	99.58
6	S vs. F	34	99.46	100	98.78
7	S vs. Q	22	99.70	99.83	99.31
8	V vs. F	38	99.22	99.53	99.13
9	V vs. Q	23	99.78	99.76	99.80
10	F vs. Q	26	99.89	99.87	99.90

Table 6 Performance of OAO-MLP system and single MLP on intra-patient paradigm

Classifiers	Accuracy (%)	Class *N*		Class *S*		Class *V*		Class *F*		Class *Q*	
		Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)
Single MLP	97.93	99.3	98.6	88.7	93.0	92.2	94.1	82.8	91.9	95.4	97.5
OAO-MLP system	99.14	99.63	99.52	89.64	97.53	97.64	96.94	95.52	95.61	98.74	99.12

Table 7 The classification results with and without SMOTE

Models		Accuracy (%)	Sensitivity (%)	Specificity (%)
Single MLP	Without SMOTE	96.07	39.49	96.48
	With SMOTE	97.83	90.13	97.72
OAA-MLP	Without SMOTE	97.92	72.19	98.12
	With SMOTE	99.32	96.03	99.45
OAO-MLP	Without SMOTE	97.85	72.83	98.03
	With SMOTE	99.14	96.50	99.23

used a CNN with three consecutive one-dimensional convolutional layers. It is worth mentioning that our models are simpler than the compared works as our models are based on a SSAE with only two layers.

4.4 Inter-patient test

In this test, we consider only three classes (N, S and V). Indeed, the classes F and Q are represented by only few beats, and they have been ignored in several works [47,48].

4.4.1 Results using OAA-MLP system

In the same manner as in intra-patient test, we first analysed the effect of the number of features in SSAE. Table 9 illustrates the accuracy, sensitivity and specificity obtained using a system of OAA-MLP and a two-layer stacked autoencoder with different numbers of features. We found that using fifty features provided the best results.

Table 10 illustrates the accuracy, sensitivity, specificity and the number of the hidden neurons corresponding to each classifier of OAA-MLP system. As in the intra-patient test, the most difficult subproblem is the 1^{st} one that correspond to: N vs. All. The corresponding network gave only 93.75% even it had been trained using more hidden neurons and for more iterations than the other networks.

Table 11 compares the full OAA-MLP system with a single MLP. The OAA-MLP system gave an accuracy of 94.69% on the test data, whereas the single MLP gave only 90.12%. Table 11 indicates that using a system of multiple MLPs considerably enhanced the overall accuracy and the sensitivity and positive predictive as well.

4.4.2 Results using OAO-MLP system

Table 12 illustrates results of an OAA-MLP system and a two-layer stacked autoencoder with different numbers of features. Again, fifty features provided the best classification results in terms of accuracy, sensitivity and specificity .

Table 8 Comparison of the classification results on intra-patient paradigm

Authors	# of classes	Extraction features method	Classifier	Performance (%)	Cross-validation (CV)
M. Zubair et al. [41]	5	Raw data	Deep CNN	Acc: 92.7	-
S. Shadmand et al. [42]	5	Hermit function coefficient and temporal	BBNN	Acc: 97.00	-
UR. Acharya et al. [43]	5	Raw data	9 layers CNN	Acc: 94.03; Sen: 96.71; Spe: 91.45	[10-CV]
M.Kachuee et al. [44]	5	Raw data	Deep residual CNN	Acc: 93.4	-
W.Yang et al. [45]	5	PCANet	Linear SVM	Acc: 97.77	[10-CV]
O.Yildirim et al. [11]	5	Raw data	LSTM	Acc: 99.23	70% -30%
Y.Ji et al. [17]	5	CNN	Faster R-CNN	Acc: 99.21; Sen: 98.06; Spe: 99.45	-
S.Mousavi et al. [24]	4	CNN	CNN+encoder-decoder network	Acc: 99.92; Sen: 98.66; Spe: 99.70	80% - 20%
Z. Shuren [46]	4	CNN	ELM	Acc: 98.77; Sen: 94.41 ;Spe: 98.45	[10-CV]
OAA-MLP system	5	SSAEs	MLPs	Acc: 99.32; Sen: 96.03 ;Spe: 99.45	80% - 20%
OAO-MLP system	5	SSAEs	MLPs	Acc: 99.14; Sen: 96.50 ;Spe: 99.23	80% - 20%

BBNN Block-based neural network; *RR* the time between the R peaks of two heartbeats; *wRR* weighted RR; *PCA* Principal component analysis

Table 9 Results obtained using OAA-MLP system with different numbers of features on inter-patient paradigm

# of features	Accuracy (%)	Sensitivity (%)	Specificity (%)
10	93.88	71.30	89.47
30	94.42	70.20	89.95
50	94.69	71	89.53
80	93.95	69.15	90.66

Table 10 The classification results of each classifier in OAA-MLP system on inter-patient paradigm

# Classifiers	Subproblems	# Hidden neurons	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	N vs. All	45	93.75	73.16	95.58
2	S vs. All	24	95.78	98.65	32.23
3	V vs. All	30	97.79	98.68	86.69

Table 11 Performance of OAA-MLP system and single MLP on inter-patient paradigm

Classifiers	Accuracy (%)	Class N Se(%)	+P(%)	Class S Se(%)	+P(%)	Class V Se(%)	+P(%)
Single MLP	90.12	94.71	94.8	12.9	11.2	81.5	82.1
OAA-MLP system	94.69	98.21	96.64	23.98	47.70	86.53	81.65

Table 13 illustrates the accuracy, sensitivity, specificity and the number of the hidden neurons in each classifier of OAO-MLP system. Here, the most difficult subproblem is the 3^{rd} that correspond to: S vs. V. Its corresponding network gave only 78.55%, while the other networks provided better results with less hidden neurons and few iterations.

The results obtained using the full OAO-MLP system and a single MLP are shown in Table 14. We note that the system of multiple MLPs provided an accuracy of 94.65% and the single MLP provided only 90.12%. The overall accuracy was then significantly enhanced using the system of multiple MLPs. The improvement can also be noted in terms of sensitivity and positive predictive, especially for classes N and S.

4.4.3 Comparison with other works

Table 15 compares the proposed models with other works applied on the same dataset, MIT-BIH arrhythmia, with inter-patient test. These works include several feature extracting methods and classification models. We note that our models outperformed all these works except the work of J.Niu et al. [51], in which the authors concatenated raw ECG signals and RR interval into a symbolic representations, and they used a multi-perspective convolutional neural networks for the classification. It can be also noted from this table that combining RR intervals with wavelets-based features or raw data provided good performances. Our models provided promising results as they are based only on the features extracted using SSAE from raw data.

Table 12 Results obtained using OAO-MLP system with different numbers of features on inter-patient paradigm

# of features	Accuracy (%)	Sensitivity (%)	Specificity (%)
10	94.12	65.47	85.50
30	94.48	64.95	86.05
50	94.65	66.78	87.96
80	94.45	66.66	89.75

5 Conclusion

We have proposed an ECG beat classification system that consists of two parts. In the first part, a deep autoencoder is used for defining high level features. In the second part, a system of multiple neural networks, based on the decomposition of the multi-class problem, is used for the classification. According to the most commonly used strategies for the decomposition of the multi-class problems, i.e. OAA and OAO, we have proposed two models. To deal with the problem of imbalanced data, we have proposed using an oversampling method, i.e. SMOTE, after the decomposition of the original problem. Adding synthetic samples is then performed according to the number of the training samples in each subproblem.

To evaluate the proposed models, we conducted experiments on the well-known MIT-BIH arrhythmia dataset and we performed two types of test: intra- and inter-patient. For each test, we first analysed the effect of the number of stacked autoencoders and the number of features. Then, we compared the proposed models with single MLP. Finally,

Table 13 The classification results of each classifier in OAO-MLP system on inter-patient paradigm

# Classifiers	Subproblems	# Hidden neurons	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	N vs.S	24	95.94	49.13	93.92
2	N vs.V	16	97.82	85.91	99.51
3	S vs.V	41	78.55	18.85	98.99

Table 14 Performance of OAO-MLP system and single MLP on inter-patient paradigm

Classifiers	Accuracy (%)	Class N		Class S		Class V	
		Se(%)	+P(%)	Se(%)	+P(%)	Se(%)	+P(%)
Single MLP	90.12	94.71	94.8	12.9	11.2	81.5	82.1
OAA-MLP system	94.65	98.66	96.20	18.92	67.68	82.77	78.09

Table 15 Comparaison of the classification results on inter-patient paradigm

Authors	# of classes	Extraction features method	Classifier	Performance (%)
T. Li et al. [49]	5	Wavelet packet Entropy + RR	Random forests	Acc: 94.61
G.Garcia et al. [47]	3	Complex network	SVM	Acc: 92.4
S. Chen et al. [48]	3	Projections +wRR	SVM	Acc: 93.1
Luo et al. [50]	4	Three layers (SDAE) + multi-layer DNN	DNN	Acc: 89.3; Sen: 42.9
V.Mondéjar-Guerra et al. [40]	4	Wavelet, HOS, RR interval, morphological	Ensemble of SVM	Acc: 94.5
J.Niu et al. [51]	2	SBCX+RR	MPCNN	Acc: 96.4
S. Haotian [52]	3	Raw data+RR	MIDNN	Acc: 94.2
H. Wang [53]	3	CNN	CNN	Acc: 93.4
OAA-MLP system	3	SSAEs	MLPs	Acc: 94.69; Sen: 71; Spe: 89.53
OAO-MLP system	3	SSAEs	MLPs	Acc: 94.65; Sen: 66.78; Spe: 87.96

RR the time between the R peaks of two heartbeats; *MIDNN* Multiple input layers deep neural network; *wRR* weighted RR; *SBCX* Symbolic baseline corrected approXimation; *MPCNN* Multi-perspective convolutional neural network

we compared the obtained results with some state-of-the-art methods. These experiments allowed us to conclude the following:

First, the decomposition of the multi-class problem provides better performances. The results obtained using multi-MLPs are clearly higher than single MLP in terms of accuracy, sensitivity and specificity.

Second, SMOTE considerably enhances the performances of the proposed system of multiple MLPs. In fact, the performances were first enhanced with the decomposition; afterwards, they were further enhanced using SMOTE. The decomposition and SMOTE can therefore be used in a complementary manner.

Third, the proposed models show promising performances compared to the state-of-the-art methods. This confirms the efficiency of the proposed approach, which consists in extracting numerical features using SSAE, then using these features as inputs to multiple simple classifiers. This approach can therefore replace the direct processing of the time series data using specialized recurrent neural networks.

Finally, it is worth mentioning that since the proposed system is based on decomposing the multi-class problem into independent binary problems, one can use any type of binary classifiers. This allows: (i) using a large variety of binary classifiers; (ii) using classifiers with different structures for each subproblem; (iii) using different features for each subproblem. The performances can then be further enhanced by adding other features to the difficult subproblems, like classifying the beats from class S and F.

References

1. Luz, E.J.D.S., Schwartz, W.R., Cámara-Chávez, G., Menotti, D.: ECG-based heartbeat classification for arrhythmia detection: a survey. Comput. Methods Program. Biomed. **127**, 144–164 (2016)
2. Roguia, S., Mohamed, N.: An optimized RBF-neural network for breast cancer classification. Int. J. Inform. Appl. Math. **1**(1), 24–34 (2019)
3. Rodrigues, R., Couto, P.: A neural network approach to ECG denoising. arXiv preprint arXiv:1212.5217 (2012)

4. Savalia, S., Emamian, V.: Cardiac arrhythmia classification by multi-layer perceptron and convolution neural networks. Bioengineering **5**(2), 35 (2018)
5. Mohapatra, S.K., Palo, H.K., Mohanty, M.N.: Detection of arrhythmia using neural network. Ann. Comput. Sci. Inform. Syst. **14**, 97–100 (2017)
6. Mohapatra, S. K., Sahoo, S., Mohanty, M. N.: Classification of arrhythmia using artificial neural network with grey wolf optimization. In international conference on biologically inspired techniques in many-criteria decision making (pp. 3-10). Springer, Cham (2019, December)
7. Bengio, Y.: Learning deep architectures for AI. Now Publishers Inc (2009)
8. Kuwajima, H., Tanaka, M., Okutomi, M.: Improving transparency of deep neural inference process. Prog. Artif. Intell. **8**(2), 273–285 (2019)
9. Dhillon, A., Verma, G.K.: Convolutional neural network: a review of models, methodologies and applications to object detection. Progress in Artificial Intelligence **9**(2), 85–112 (2020)
10. Lotfollahi, M., Siavoshani, M.J., Zade, R.S.H., Saberian, M.: Deep packet: a novel approach for encrypted traffic classification using deep learning. Soft Comput. **24**(3), 1999–2012 (2020)
11. Yildirim, O., Baloglu, U.B., Tan, R.S., Ciaccio, E.J., Acharya, U.R.: A new approach for arrhythmia classification using deep coded features and LSTM networks. Comput. Methods Program. Biomed. **176**, 121–133 (2019)
12. Yang, J., Bai, Y., Lin, F., Liu, M., Hou, Z., Liu, X.: A novel electrocardiogram arrhythmia classification method based on stacked sparse auto-encoders and softmax regression. Int. J. Mach. Learn. Cybern. **9**(10), 1733–1740 (2018)
13. Al Rahhal, M.M., Bazi, Y., AlHichri, H., Alajlan, N., Melgani, F., Yager, R.R.: Deep learning approach for active classification of electrocardiogram signals. Inform. Sci. **345**, 340–354 (2016)
14. Yildirim, O., San Tan, R., Acharya, U.R.: An efficient compression of ECG signals using deep convolutional autoencoders. Cognit. Syst. Res. **52**, 198–211 (2018)
15. Nurmaini, S., Darmawahyuni, A., Sakti Mukti, A.N., Rachmatullah, M.N., Firdaus, F., Tutuko, B.: Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification. Electronics **9**(1), 135 (2020)
16. Yildirim, Ö.: A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in biology and medicine **96**, 189–202 (2018)
17. Ji, Y., Zhang, S., Xiao, W.: Electrocardiogram classification based on faster regions with convolutional neural network. Sensors **19**(11), 2558 (2019)
18. Guan, X., Liang, J., Qian, Y., Pang, J.: A multi-view OVA model based on decision tree for multi-classification tasks. Knowledge-Based Syst. **138**, 208–219 (2017)
19. Zhou, L., Wang, Q., Fujita, H.: One versus one multi-class classification fusion using optimizing decision directed acyclic graph for predicting listing status of companies. Inform. Fus. **36**, 80–89 (2017)
20. Zhang, Z.L., Luo, X.G., García, S., Tang, J.F., Herrera, F.: Exploring the effectiveness of dynamic ensemble selection in the one-versus-one scheme. Knowledge-Based Syst. **125**, 53–63 (2017)
21. Nemissi, M., Seridi, H., Akdag, H.: One-against-all and one-against-one based neuro-fuzzy classifiers. J. Intell. Fuzzy Syst. **26**(6), 2661–2670 (2014)
22. Nemissi, M., Seridi, H., Akdag, H.: The labeled systems of multiple neural networks. Int. J. Neural Syst. **18**(04), 321–330 (2008)
23. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
24. Mousavi, S.,Afghah, F.: Inter-and intra-patient ecg heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach. In ICASSP 2019-2019 IEEE International conference on acoustics, speech and signal processing (ICASSP) (pp. 1308-1312) (2019, May)
25. De Chazal, P., O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. IEEE Trans. Biomed. Eng. **51**(7), 1196–1206 (2004)
26. Ng, A.: Sparse autoencoder. CS294A Lecture notes, 72, 1-19 (2011)
27. Oliveira, T. P., Barbar, J. S.,Soares, A. S.: Multilayer perceptron and stacked autoencoder for Internet traffic prediction. In IFIP International conference on network and parallel computing (pp. 61-71). Springer, Berlin, Heidelberg (2014, September)
28. Sáez, J.A., Galar, M., Krawczyk, B.: Addressing the overlapping data problem in classification using the one-vs-one decomposition strategy. IEEE Access **7**, 83396–83411 (2019)
29. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. Pattern Recognit. **44**(8), 1761–1776 (2011)
30. Hüllermeier, E., Vanderlooy, S.: Combining predictions in pairwise classification: an optimal adaptive voting strategy and its relation to weighted voting. Pattern Recognit. **43**(1), 128–142 (2010)
31. Platt, J., Cristianini, N.,Shawe-Taylor, J.: Large margin DAGs for multiclass classification. In Advances in neural information processing systems (pp. 547-553) (2000)
32. Hühn, J., Hüllermeier, E.: FR3: a fuzzy rule learner for inducing reliable classifiers. IEEE Trans. Fuzzy Syst. **17**(1), 138–149 (2009)
33. Fernández, A., Calderón, M., Barrenechea, E., Bustince, H., Herrera, F.: Solving multi-class problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations. Fuzzy Sets Syst. **161**(23), 3064–3080 (2010)
34. Fei, B., Liu, J.: Binary tree of SVM: a new fastmulticlass training and classification algorithm. IEEE Trans. Neural Netw. **17**(3), 696–704 (2006)
35. Wu, T., Lin, C., Weng, R.: Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. **5**, 975–1005 (2004)
36. Liu, B., Hao, Z., Tsang, E.: Nesting one-against-one algorithm based on SVMs for pattern classification. IEEE Trans. Neural Netw. **19**(12), 2044–2052 (2008)
37. PhysioNet: Physionet MIT-BIH arrhythmia database https://archive.physionet.org/cgi-bin/atm/ATM. Accessed 2020. (2001)
38. Moody, George B., Mark, Roger G.: The impact of the mit-bih arrhythmia database. IEEE Eng. Med. Biol. Mag. **20**(3), 45–50 (2001)
39. ANSI-AAMI: Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms, American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI), ANSI/AAMI/ISO (1998-2008)
40. Mondéjar-Guerra, V., Novo, J., Rouco, J., Penedo, M.G., Ortega, M.: Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers. Biomed. Signal Process. Control **47**, 41–48 (2019)
41. Zubair, M., Kim, J., Yoon, C.: An automated ECG beat classification system using convolutional neural networks. In 2016 6th international conference on IT convergence and security (ICITCS) (pp. 1-5). IEEE (2016, September)
42. Shadmand, S., Mashoufi, B.: A new personalized ECG signal classification algorithm using block-based neural network and particle swarm optimization. Biomed. Signal Process. Control **25**, 12–23 (2016)
43. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San Tan, R.: A deep convolutional neural network model to classify heartbeats. Comput. Biology Med. **89**, 389–396 (2017)

44. Kachuee, M., Fazeli, S., Sarrafzadeh, M.: Ecg heartbeat classification: a deep transferable representation. In 2018 IEEE International conference on healthcare informatics (ICHI) (pp. 443-444) (2018, June)
45. Yang, W., Si, Y., Wang, D., Guo, B.: Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine. Comput. Biol. Med. **101**, 22–32 (2018)
46. Zhou, S., Tan, B.: Electrocardiogram soft computing using hybrid deep learning CNN-ELM. Appl. Soft Comput. **86**, 105778 (2020)
47. Garcia, G., Moreira, G., Menotti, D., Luz, E.: Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO. Sci. Rep. **7**(1), 1–11 (2017)
48. Chen, S., Hua, W., Li, Z., Li, J., Gao, X.: Heartbeat classification using projected and dynamic features of ECG signal. Biomed. Signal Process. Control **31**, 165–173 (2017)
49. Li, T., Zhou, M.: ECG classification using wavelet packet entropy and random forests. Entropy **18**(8), 285 (2016)
50. Luo, K., Li, J., Wang, Z., Cuschieri, A.: Patient-specific deep architectural model for ECG classification. J. Healthc. Eng. (2017)
51. Niu, J., Tang, Y., Sun, Z., Zhang, W.: Inter-Patient ECG classification with symbolic representations and multi-perspective convolutional neural networks. IEEE J. Biomed. Health Inform. **24**(5), 1321–1332 (2019)
52. Shi, H., Qin, C., Xiao, D., Zhao, L., Liu, C.: Automated heartbeat classification based on deep neural network with multiple input layers. Knowledge-Based Syst. **188**, 105036 (2020)
53. Wang, H., Shi, H., Lin, K., Qin, C., Zhao, L., Huang, Y., Liu, C.: A high-precision arrhythmia classification method based on dual fully connected neural network. Biomed. Signal Process. Control **58**, 101874 (2020)