Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 12 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i12.41631

PAPER

Arrhythmia Detection Based on New Multi-Model Technique for ECG Inter-Patient Classification

Zahraa Ch. Oleiwi[1,2](✉), **Ebtesam N. AlShemmary**[3], **Salam Al-Augby**[1]

[1]Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq

[2]College of Computer Science and Information Technology, University Al-Qadisiyah, Qadisiyah, Iraq

[3]IT Research and Development Center, University of Kufa, Najaf, Iraq

zahraa.chaffat@qu.edu.iq; zahraac.albadri@student.uokufa.edu.iq

ABSTRACT

This paper presents a novel model for arrhythmia detection based on a cascading technique that utilizes a combination of the One-Sided Selection (OSS) method, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) algorithms, this model denoted by (OWSK) model to classify four types of electrocardiogram (ECG) heartbeats following inter-patient scheme. The OWSK model consists of three stages. The first stage involves resampling using the One-Sided Selection (OSS) method to solve the imbalance problem and reduce data by removing noisy, borderline, and redundant samples. The second stage involves using Wavelet Transformation (WT) and Power Spectral Density (PSD) to extract the most relevant frequency domain features. The third stage involves a cascading process by constructing the classifier from SVM trained on the whole dataset to classify normal and abnormal beats. Then, KNN (K-Nearest Neighbors) is trained on only the three irregular minority classes to classify the three types of arrhythmias for the detection of ventricular ectopic beats, supraventricular ectopic beats, and fusion beats (V, S, and F). The performance of the proposed model is evaluated in terms of different metrics, including accuracy, recall, precision, and F1 score. The results show the superiority of the proposed model in medical diagnosis compared to the latest works, where it achieves 90%, 90%, 93%, and 91% for accuracy, recall, precision, and F1 score under the inter-patient paradigm and 98%, 98%, 98%, and 98% under the intra-patient paradigm.

KEYWORDS

arrhythmia detection, ECG, inter- and intra-patients paradigm, wavelet transform, One-Sided Selection (OSS) method

1 INTRODUCTION

Heart disease is one of the most severe diseases, the late detection of which may threaten human life. Among these diseases is arrhythmia, as these diseases are identified by analyzing the electrocardiogram (ECG), which is one of the most popular methods used to check the electrical activity and rhythm of the heart [1, 2]. Based on the analysis of the ECG to extract relevant features, the automatic classification models

Oleiwi, Z.C., AlShemmary, E.N., Al-Augby, S. (2023). Arrhythmia Detection Based on New Multi-Model Technique for ECG Inter-Patient Classification. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(12), pp. 78–98. https://doi.org/10.3991/ijoe.v19i12.41631

Article submitted 2023-05-22. Resubmitted 2023-06-26. Final acceptance 2023-06-26. Final version published as submitted by the authors.

of heartbeats are constructed to identify different types of heart diseases, including arrhythmia [3]. This paper proposes a novel model based on machine learning models that aim for the early, accurate detection of heart diseases [4]. Many problems have been faced by these models associated with the ECG datasets as the imbalance between the types of arrhythmia classes where the normal subjects always construct the majority class that contains several samples much more than other abnormal classes [5]. In addition, there is another problem that concerns the scheme of splitting the heartbeats of patient ECG records into training and testing datasets, where there are two schemes; the first is the inter-patient scheme in which there are no heartbeats in both the training and testing datasets for the same patient record [6].

In contrast, in the second intra-patient scheme, there are heartbeats for the same patient in both the training and testing datasets; this scheme causes biased classification results due to a higher correlation between the beats of the same ECG patient record [5]. Many researchers implemented their learning model on an intra-patient scheme and achieved high classification accuracy [7]. However, these results led to a need to generalize the model performance on the practical clinical samples. As for the inter-patient scheme, obtaining high accuracy is not easy, as the model is subject to evaluation using samples that it has never been trained on, especially when this problem combines with the problem of imbalance [5, 8].

A few studies focused on the inter-patient scheme. Most relied on deep learning, characterized by its high computational complexity regarding training time, computational cost, number of parameters needed to construct the model, and model architecture complexity [9, 10].

The main objective of this work is to address the problem of early heart disease detection and diagnosis by developing a more efficient classification model that can produce good results. A classification model based on two machine learning methods was proposed to achieve this objective, which classifies ECG beats using inter and intra-patient schemes. The significant contributions of this work are:

- The imbalance problem was addressed using a down sampling method, the One-Sided Selection (OSS).
- The WT was used as feature extraction at the fourth level of decomposition of the ECG signals.
- Based on two machine learning methods: SVM and KNN were used in a two-cascaded stages process to produce a model denoted by OWSK-model (OSS-WT-SVM-KNN model) with an analysis study of the efficiency of this proposed model.

This paper presents a novel cascading multi-model technique for arrhythmia detection based on inter and intra-patient ECG Heartbeats classification. The proposed model utilizes a combination of the One-Sided Selection (OSS) method, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) algorithms to address the imbalance problem in the MIT-BIH Arrhythmia Database.

2 RELATED WORKS

Several researchers produced different classification models of heartbeats for arrhythmia detection addressing the inter-patient problem, where the inter-patient scheme is most realistic clinically. Most of the research is based on deep learning networks to construct their models.

An architecture of a deep learning model based on the densely connected convolutional neural network (DenseNet) and gated recurrent unit network (GRU) was implemented and evaluated using two standard MIT-BIH databases (arrhythmia and supraventricular databases) in terms of accuracy and sensitivity. The results show that this model achieved 93.61% classification accuracy and 62.7% sensitivity for supraventricular (SVEB) and 93.71% classification accuracy, and 91.25% sensitivity for ventricular (VEB) [11].

A lightweight two-dimensional convolutional neural network (2D-CNN) was designed in [12] to improve the accuracy of inter-patient ECG heartbeats classification. This work uses a computer vision approach, where a plotting algorithm was built to plot the ESG signal as a low-sized image. Then, reliable VGG network concepts were applied to the constructed ECG images. This proposed model achieved the highest classification accuracy of about 98.5% and prediction accuracy of about 98.4% for SVEB heartbeats and VEB heartbeats, respectively.

Another automatic model utilizing deep convolutional neural networks (CNN) was proposed in [13], focusing on inter-patent ECG heartbeats classification. This model was evaluated on the MIT-BIH arrhythmia database regarding sensitivity, positive predictivity, and false positive rate. As a result, this model obtained 97% positive predictivity, 92% sensitivity, and 23% false positive rate for normal heartbeats (N), 56% positive predictivity, 62% sensitivity, and a 2% false positive rate for supraventricular ectopic beat (SVEB), and 51% positive predictivity, 89% sensitivity, and 6% false positive rate for ventricular ectopic beat (VEB).

In [14], a novel deep convolutional neural network was proposed for inter-patient heartbeat classification. This work fixed the unbalancing problem by suggesting a batch-weighted loss function; no preprocessing steps were used. Five classes of MITDB arrhythmia were classified using this model, which achieved 88.34% classification accuracy, 48.25% positive productivity, 90.90% sensitivity, and 88.51% specificity.

A new method for ECG classification using deep learning has been introduced. It solves the issue of insufficient labelled data, improves classification accuracy, and handles varying data distributions. The method comprises three modules: multi-scale feature extraction, domain discrimination, and classification. Experiments show the approach to be effective, with a classification accuracy of 92.3% [15].

All the mentioned works depend on deep learning methods in constructing their model, and none use machine learning methods. Although models based on deep learning achieve high classification accuracy, they require large amounts of data for training, longer training time, and a specialized GPU (graphics processing unit). Hence, they are more complex than models based on machine learning. So, in this work, the proposed model is based on the machine learning method to ensure efficiency in using simple methods and a small number of training data while achieving promising results in the classification of minority classes of the used dataset.

3 MATERIAL AND METHODS

3.1 Database

The standard MIT/BIH arrhythmia database contains 48 half-hour excerpts of two-channel ambulatory ECG recording files obtained from 47 patients (25 of the patients were men aged 32 to 89 and 22 were women aged 23 to 89, two records numbered 201 and 202 for the same patient), is freely available at (https://physionet.org/content/mitdb/1.0.0/). The recordings were digitized with a sampling frequency of 360 Hz and acquired with 11-bit resolution over a ten-mV range [16]. In this work, 44 files were used from all 48 database records. According to the Association for the Advancement of Medical Instrumentation (AAMI) recommendations, four records (102, 104, 107, 217) with paced beats are excluded. AAMI recommendations categorize the 18 types of heartbeats into five groups: Nontectonic beat (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beats (F), and unknown beat type (Q). The unknown beat (Q) group is excluded due to its few positive samples compared to other classes; it has only 15 samples [5].

Two typical schemes of separating the entire database into training and testing sets are intra-patient and inter-patient. In the intra-patient scheme, the beats of all ECG records are collected together and then split into training and testing datasets with the specific thumb-of-rule ratio. This method ensures that the heartbeat samples of the same patient are found in both the training and testing datasets. As a result, the classification accuracy of the model evaluated on this scheme reached up to 99% due to solid dependence and correlation between heartbeats of the same patient. The result of classification testing beats of the patient which is reset in the training set, was bias result [6].

The inter-patient scheme involved separating the patients' records to the training and testing datasets without intersecting so that the model predicted the class label for patient ECG samples, which were not available in training set during the training process. Generally, the 44 ECG records of MITDB are split as follows: 22 records numbered (101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230) for the training set and denoted by DS1, and the rest 22 records numbered as (100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234) for testing set and denoted by DS2, which are used for testing and evaluation of the model [12].

3.2 One-Sided Selection (OSS) method

An imbalanced database means that the number of examples in each class is unequally distributed. An imbalanced dataset is one of the most common problems that machine learning methods face. This problem affects the performance of the classification model and causes overfitting, which leads to less generalization and high misclassification [17].

Many algorithms were proposed to overcome the limitation and drawbacks of the under-sampling technique, such as One-Sided Selection (OSS), which combines two other under-sampling methods: Tomek Links and the Condensed Nearest Neighbor (CNN) Rule. The OSS benefits from both methods, where noisy and borderline examples are removed by implementing Tomek Links. These redundant examples are far from the decision boundary removed by CNN. These methods ensure no loss of information and balance [18].

3.3 Support Vector Machine (SVM)

SVM is one of the supervised machine learning methods which is used chiefly for binary. The essential working principle (framework) of SVM is the construction of a hyperplane that splits the two classes with high classification accuracy and as large a margin as possible. Where the margin is a perpendicular distance between the two lines on the closest class points, these closest class points are denoted as support vectors. SVM is characterized by significant classification results for the linear and nonlinear separable dataset using kernel trick, which transfers the data to the high dimensional space using a nonlinear function where the data can be separated linearly [19].

SVM performs linear classification for the training set, which represents as $\{(x_i, y_i)\}$ *for* $i = 1, 2, 3 \ldots\ldots N$ where x_i is the n-dimensional feature vector in the original data space, N number of samples in the dataset, y_i is a target label that is either +1 or –1 indicating the class including x_i. Linear SVM aims to find the optimal hyperplane that segregates between x_i points belonging to class +1 and x_i points belonging to class –1 with the maximum gap between the hyperplane and nearest points x_i belonging to either class; this hyperplane is a set of points that satisfy the equation as follows [20]:

$$w^T x + b = 0 \tag{1}$$

Where w is the adjustable weight normal vector to the hyperplane and b is the bias. Then the points belong to class +1 on or above the line as follows:

$$w^T x + b = +1 \tag{2}$$

The points belong to class –1 on or below the line:

$$w^T x + b = -1 \tag{3}$$

The margin is the distance between these two lines, defined geometrically as $\frac{2}{\|w\|}$, so the term $\|w\|$ should be minimized to maximize the margin.

The optimization function which should be solved to find the optimal hyperplane with maximum margin is formulated as:

Minimise $\frac{\|w^2\|}{2}$ (objective function)

Subject to constrain $y_i(w.x + b) \geq 1$, $i = 1, 2, 3, \ldots\ldots N$

If the data has a nonlinear separable pattern, then the optimization function is formulated as follows:

Minimize the objective function $\frac{\|w^2\|}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)$

Subject to constrain $y_i(w.x + b) \geq 1 - \xi_i$, $\xi_i \geq 0$, $i = 1, 2, 3, \ldots\ldots N$

Where ξ_i is slack variables, and C is a regularization parameter that weights two terms in the objective function to obtain a soft margin. So, when the value of C is large, the misclassification examples are penalized and the margin will be hard (small gap). The misclassification examples are given a low penalty for a small value of C, and the margin will be soft (large gap), leading to high generalization performance for new examples prediction with more accuracy and overcoming the overfitting problem [21].

The nonlinear decision function is more sufficiently satisfactory to be considered practically where the nonlinear SVM was extended using inner product kernel notation and the points of original data are transferred to a high dimensional Euclidean space H in which the data can be separated by linear decision boundary, using a nonlinear function as [22]:

$\varnothing: R^n \rightarrow H$ where R^n is the original space of feature vectors.

In the training process, the kernel function K is defined as related to $\varnothing(x_i)$ as:

$$K(x_i, x_j) = \varnothing(x_i).\varnothing(x_j) \tag{4}$$

Hence the w in transferred space will be:

$$w = \sum_{i=1}^{N} \alpha_i y_i \varnothing(x_i) \tag{5}$$

Then the function of hyperplane (decision boundary function) can be calculated by dot product with w as follows:

$$f(x) = w.\varnothing(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right) \tag{6}$$

Where α_i for $i = 1, 2, 3, \ldots N$ are Lagrange multipliers.

There are many kernel functions. The most common one is the Gaussian radial basis kernel denoted by (RBF) which is used in this work, defined as:

$$K(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)} \tag{7}$$

Where σ is the kernel's width, it has a significant effect on the SVM classification accuracy, as with regularisation parameter C effect, which also has an essential role in controlling the classification accuracy, as we will show in the results of classification two inter and intra-patient schemes [19].

3.4 K-Nearest Neighbor (KNN)

KNN is one of the simplest and non-parametric machine learning algorithms for supervised learning, used for classification and regression. The main framework of KNN is the similarity between the new test sample and the training dataset and classifying it to the class with the most similarity. According to that, it considers a non-parametric algorithm since it does not make any assumption on underlying data. In addition, KNN does not model or learn from the training dataset during the training process; instead, it just stores the dataset, so it is called a lazy learning algorithm [23]. While during the testing or prediction phase, when there is a new sample, it computes the similarity between this new sample and all the stored training data.

Bellow steps illustrate the KNN algorithm:

Step1: Store the training data

Step 2: Choose the initial value for K number of neighbours.

Step 3: For each sample in testing:

1. Calculate the distance to each sample in the training dataset
2. Sort the distances and indices in ascending order and choose the K few distances to determine the K nearest training samples (K training samples close to new test samples).
3. Identify the class labels of these K points and determine the most frequent class label, hence classifying the new sample to this class and assigning it as the predicted label to that new sample.

The above steps are easy to implement and have no training time, but it needs high memory and prediction time. This project aims to address the issue at hand, to enhance the effectiveness of KNN while leveraging its straightforwardness [24].

4 GENERAL PROPOSED FRAMEWORK

4.1 Data preprocessing and perpetration

ECG database contains patients' records that the segmentation process should prepare. Each record has been segmented using a window of –300 ms to 400 ms around the R-wave into the number of beats, each beat of 252 samples using tools of (WFDB) software.

After the segmentation process, the ECG heartbeats for each patient were saved in CSV files individually, i.e., there were 44 CSV files for 44 patients. Then these files are separated into training and testing sets according to inter and intra-patient schemes, as in Table 1.

There are general preprocessing steps that all databases have in common, such as processing the missing, smoothing the outlier data, data normalization, and data reduction. These steps were implemented on the arrhythmia database used in this work as follows:

1. Deleting the records that contain a missing value.
2. Using the wavelet transform, which ensures the removal of noise and data reduction.
3. Using the Z-score normalization method: data normalization reduces DC offset and magnitude variant scaling among different files.
4. On the other hand, there are specific preprocessing steps that depend on the characteristics of the data, the distribution of its samples, how to read it, convert it, and prepare it into data that can be worked on according to the programming language used, as well as make it appropriate for the classification model used.

MIT-BIH arrhythmia database is characterized by a high imbalance problem, as in Table 1, which illustrates the number of samples belonging to each class. Also, it contains high borderline and overlap examples. Therefore, these problems were addressed and fixed using one of the resampling methods known as the One-Sided Selection (OSS) method, as in Table 2.

Table 1. The separation of the dataset into training and testing datasets following inter and intra-patient scheme

Scheme Type	Dataset	Class Label Types				# ECG Beats
		N	S	V	F	
Inter-patient	Training dataset DS1	45827	944	3786	414	50971
	Testing dataset DS2	44217	1837	3219	388	49661
	Total	90044	2781	7005	802	100632
Intra-patient	Training dataset	63030	1947	4904	561	70442
	Testing dataset	27014	834	2101	241	30190
	Total	90044	2781	7005	802	100632

Table 2. Data distribution in each class for training with the binary class label following inter and intra-patient

Scheme Type	Training Dataset	Training Dataset with the Binary Class Label		Total Number of Beats
		Normal(N)	Abnormal	
Inter-patient	DS1 before OSS	45827	5144	50971
	DS1 after OSS	20232	5144	25376
Intra-patient	Before OSS	63030	7412	70442
	After OSS	43226	7412	50638

4.2 Feature extraction

This work extracts two feature categories based on Discrete Wavelet transforms and Power Spectral Density (PSD). The first stage of the feature extraction phase in this work is calculating the detail coefficients of input heartbeats using wavelet transformation at the fourth decompositions level.

The wavelet transforms mathematical formulation expressed by [25]:

$$W_x^{\psi}(t,s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(\tau)\psi * \left(\frac{\tau - t}{s}\right) d\tau \tag{8}$$

Where: s is scale (dilation), y(t) is mother wavelet, τ is a time shift, and ψ^* is y complex conjugate. There is more than one example of a mother wavelet; the Haar wavelet was used in this work [26].

The second stage of the feature extraction phase is calculating the power spectral density (PSD) of WT. Approximate coefficients at the fourth level of the non-parametric periodogram method, which is the Fourier transform of the autocorrelation signal defined, as follows [27]:

$$\hat{P}(F) = \frac{\Delta t}{N} \left| \sum_{n=0}^{N-1} x_n e^{-i2\pi f \Delta t n} \right|^2, \quad -1/2\Delta t < f < 1/2\Delta t \tag{9}$$

Where x_n is sampled at f_s samples per unit time, Δt is the sampling interval. For a one-sided periodogram, the values at all frequencies except 0 and the Nyquist, $1/2\Delta t$, are multiplied by two so that the total power is conserved.

This procedure extracts important discriminate features with less dimension than the original data.

4.3 Proposed model construction

The proposed OWSK_Model is designed and constructed from four basic steps based on the idea of handling data imbalance:

1. The first step includes redistributing the database into two forms. The first distribution includes forming a binary database consisting of normal and abnormal examples by changing the class labels from multi-label to binary labels, i.e.,

we denote the normal label by 0 and any label of any abnormal type with the symbol 1. The second distribution involved the isolation of minority samples from the original database and the construction of a database consisting of three types of abnormal heartbeats S, V, F.

2. The second step is to use the downsample One-Sided Selection (OSS) method for the first binary distribution, which includes removing redundant, noise, outlier, overlap, and borderline samples from the normal class N as in Table 2.
3. The third step is to extract the characteristics from the ECG signals, which include the detail coefficients of the fourth decomposition level of the discrete wavelet transform (DWT), and concatenate it with the PSD coefficients obtained for the approximate coefficients of the fourth decomposition level of DWT. Block 2 in Figure 1, which is decomposed as separated Figure 2, illustrates this step.
4. The fourth step is related to the use of the machine learning-based cascading multi-model technique for classification, which includes the use of the support vector machine (SVM) method and training it on the binary database distribution to classify the data into normal and abnormal heartbeats and then using and training the K-Nearest Neighbors (KNN) method on second minority multiclass database distribution to classify the minority samples into three types of abnormal heartbeats S, V, F. The trained models are then used to predict test samples as follows:

Predict normal and abnormal heartbeats using SVM, then take only the abnormal predicted samples and enter them into the second KNN model to predict their type. Finally, the results of predictions from the two methods are combined to evaluate the OWSK model, Block 5 in Figure 1, which is decomposed after being separated. Figure 3 illustrates the prediction step. Figure 1 illustrates the main components of the proposed OWSK model.

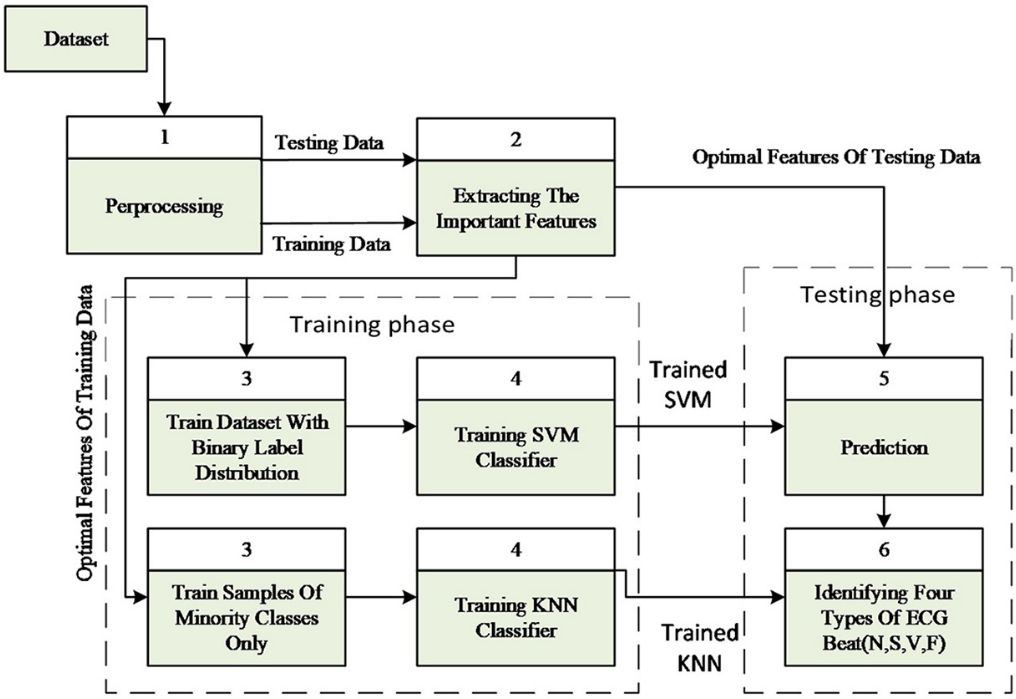

Fig. 1. The main components of the proposed framework

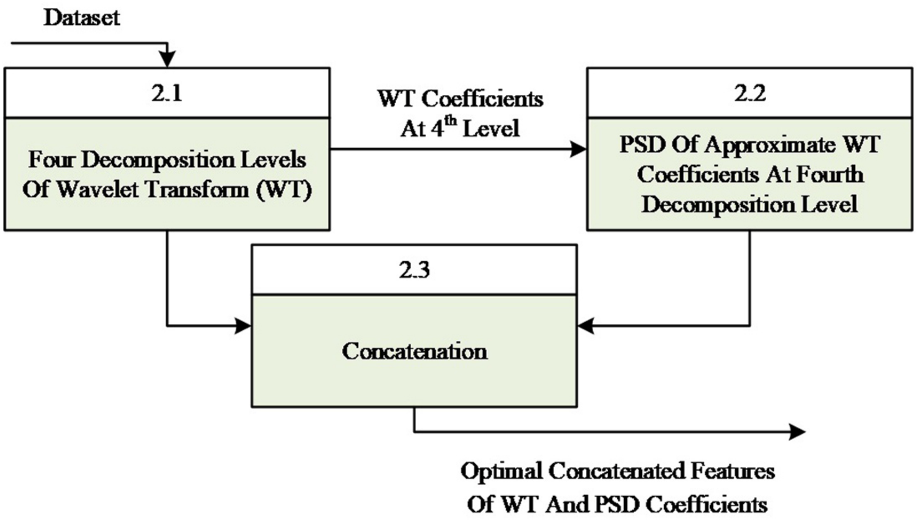

Fig. 2. The features extraction step of the proposed framework

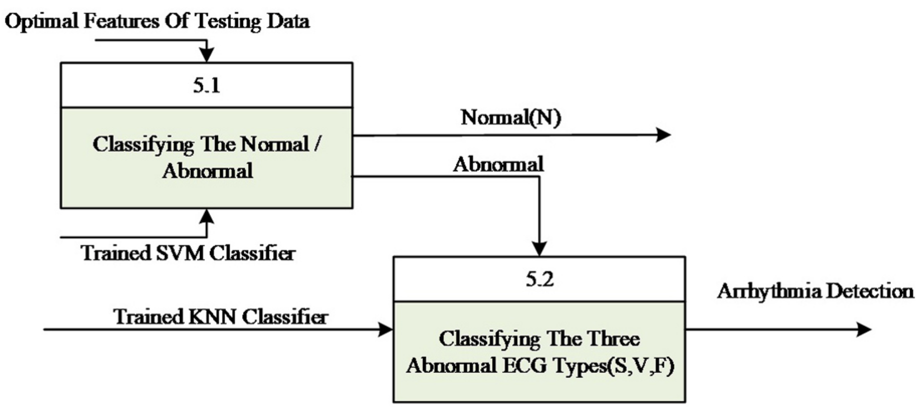

Fig. 3. The prediction step of the proposed framework

5 ANALYSIS OF COMPUTATIONAL COMPLEXITY

The training time complexity of the SVM method is $O(N^3)$, and the space complexity is $O(N^2)$. The testing (prediction) time is $O(kd)$, where N is the number of samples in the dataset (dataset size), d is the dimension of the dataset, and k is the number of support vectors [28].

As well the testing (prediction) time complexity of the KNN method is $O(k_1 dN)$, the space complexity is $O(dN)$, where the N is the number of samples in the dataset (dataset size), d is the dimension of the dataset, and k_1 is the number of Nearest Neighbors [29].

In light of the above **Big O Notation, the training time complexity of the proposed classifier is** $O(N^3)$, and the space complexity is $O(N^2 + dN_1)$. The testing (prediction) time is $O(kd + k_1 dN_1)$, where N_1 is the number of samples in the minority classes (since the KNN in the proposed model is trained on minority classes only).

The proposed model has become more complicated, but using the Wavelet as feature selection and feature reduction and OSS as an under-sampling technique improves the efficiency of the proposed model since the computation complexity of the proposed model depends on the size a dimension of the dataset as follows:

A. Before using WT and OSS

The size of the training set is (50970) ECG heartbeats, and the number of features (dimension) is (252) before using WT and OSS. Then the training time complexity is $O(50970^3)$, the space complexity is $O(50970^2 + (252 * 5144))$, and testing (prediction) time is $O((k * 252) + k(252 * 5144))$.

B. After using WT and OSS

The size of the training set is (25376) ECG heartbeats, and the number of features (dimension) is (25) before using WT and OSS. Then the training time complexity is $O(25376^3)$, the space complexity is $O(25376^2 + (25 * 5144))$, and testing (prediction) time is $O((k * 25) + k(25 * 5144))$.

According to the algorithm of the proposed model, if the predicated sample is normal, the predicting time is $O(25k)$, and if the predicated sample is abnormal, then $O((k * 25) + k (25 * 5144))$.

As a result, the proposed OWSK model training time complexity is reduced by 88%, space complexity is reduced by 75%, and prediction time is reduced by 90% from the original value.

6 RESULTS AND DISCUSSION

The proposed model was trained and tested on the MIT-BIH arrhythmia database according to AAMI recommendations, and two inter and intra- patient schemes were considered.

The performance of the proposed model is evaluated in terms of several evaluation metrics; the most important are four main metrics: accuracy (ACC), recall or sensitivity (true positive ratio (TPR)), precision (positive prediction value (PPV)), and f1-score.

The rest metrics are false positive ratio (FPR), true negative ratio (TNR), false negative ratio (FNR), negative prediction value (NPV), and false prediction ratio (FDR).

In order to highlight the effectiveness of the proposed model, experiments were conducted following the AAMI guidelines and the inter-patient and intra-patient schemes.

6.1 Performance evaluation of proposed model following intra-patient scheme

The results are presented for performance evaluation of the proposed classification model, where the training and test sets were selected randomly as a 70:30 thumb rule.

The proposed model was trained, and its performance was evaluated after the fourth DWT level of decomposition, where the dimension of the data entering the model is (16+9), where 16 equals the number of wavelet detail coefficients at the fourth level, and 9 are the nine features extracted from PSD of approximate WT coefficients at the fourth decomposition level in the aim of increasing the relevant features. These features were concatenated to construct the final feature vector of 25 lengths. The results of the performance evaluation of the second stage of the proposed model based on KNN, which depends on the resultant prediction vector from the first stage, are illustrated in Table 3.

Predicted and actual value combinations for normal and abnormal classes using SVM with regularization parameter C=20 under the intra-patient scheme is illustrated in the confusion matrix Figure 4.

Table 3. Evaluation results of the second stage of the proposed model based on KNN for the intra-patient scheme under different values of regularization parameter C

Class Type	TPR	FNR	TNR	PPV	NPV	FDR	ACC
C=0.1							
N	98.5	1.41	65.1	96.0	84.3	3.98	95.0
S	10.4	89.5	99.9	78.3	97.5	21.6	97.4
V	87.4	12.5	98.6	83.1	99.0	16.8	97.8
F	43.1	56.8	99.9	78.7	99.5	21.2	99.4
C=1							
N	99.6	0.30	82.2	97.9	96.9	2.05	97.8
S	58.3	41.6	99.7	89.1	98.8	10.8	98.6
V	92.2	7.71	99.8	97.6	99.4	2.31	99.3
F	57.6	42.3	99.9	84.7	99.6	15.2	99.5
C=20							
N	99.5	0.45	88.665	98.679	95.848	1.32	98.4
S	74.4	25.5	99.7	87.8	99.2	12.1	99.0
V	94.336	5.66	99.8	97.3	99.5	2.60	99.4
F	68.8	31.1	99.9	84.6	99.7	15.3	99.6
C=100							
N	99.4	0.57	89.6	98.7	94.8	1.21	98.3
S	77.9	22.062	99.6	86.4	99.3	13.5	99.0
V	94.1	5.80	99.7	96.6	99.5	3.32	99.3
F	69.2	30.7	99.8	83.0	99.7	16.9	99.6

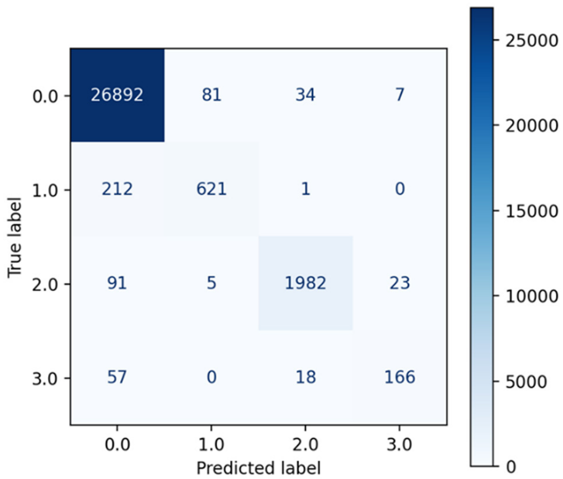

Fig. 4. Confusion matrix for multiclassification under the intra-patient scheme

The results show that the value of C significantly impacts all the performance metrics values, as all metrics increase with the increase of C, and the performance of the proposed model improved with a high value of parameter C for the SVM algorithm.

The impact of parameter C on the performance of the proposed model under the intra-patient scheme in terms of sensitivity was visualized in Figures 5 and 6.

As it is known, the enormous value of C makes the SVM try to minimize margin size (a decision boundary with a smaller margin) by giving a high penalty to misclassified examples to reduce the number of misclassified examples. So, the evaluation results come true with this knowledge due to the natural character of the MIT dataset with an intra-patent scheme. The examples in the testing set correlated to examples in the training set, leading to an overfitting problem in case of a large margin. This small decision boundary suits the intra-patient scheme due to the high correlation between beats in training and testing datasets. Although the OSS method plays a significant role in solving the imbalance problem as it clears acceptable and promising results, the classes with few examples have more misclassification examples than classes with many examples. In contrast, the dataset used in this work needs a vast and significant difference between the examples for each class, especially between N, S, and F classes.

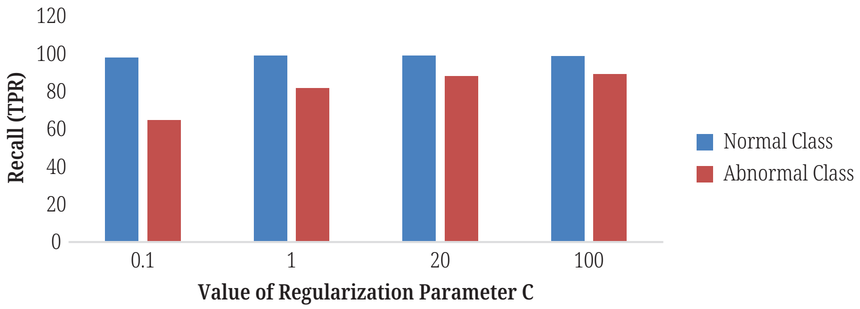

Fig. 5. Sensitivity metric at different values of regularization parameter C for binary classification under the intra-patient scheme

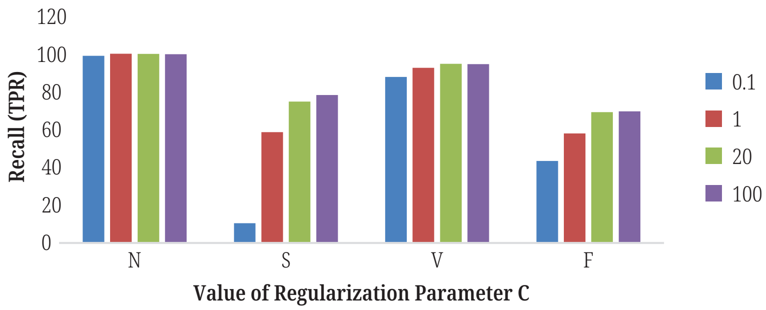

Fig. 6. Sensitivity metric at different values of regularization parameter C for multiclassification under the intra-patient scheme

6.2 Performance evaluation of proposed model following inter-patient scheme

With the AAMI recommendations and the inter-patient scheme, the model's generalization performance was evaluated using DS2, which was trained using DS1.

The researches of medical diagnoses experience showed that the delay of the best and small time of treatment could be caused by minority abnormal ECG heartbeats, which means that the incorrect misclassification of minority beats has more impact than majority beats, according that the most critical metrics are TPR, which refers to the number of correctly classified samples, and FNR, which indicates the ratio of the number of misclassifications.

So, the results in Table 4 are promising since the FNR (misclassification of abnormality samples) of the minority abnormality class is only 19%, which is considered a small ratio concerning the small number of samples in the abnormality class as compared to the number of samples in normal majority classes.

Table 4. Evaluation results of the second stage of the proposed model based on KNN for the inter-patient scheme under different values of regularization parameter C

Class	TPR	FNR	TNR	PPV	NPV	FDR	ACC
C=0.1							
N	91.2	8.78	80.1	97.3	52.9	2.60	90.0
S	72.7	27.2	95.4	37.9	98.9	62.0	94.5
V	84.3	15.6	96.1	60.5	98.8	39.4	95.4
F	43.8	56.1	99.8	69.9	99.5	30.0	99.4
C=1							
N	70.4	29.5	84.9	97.4	26.1	2.57	72.0
S	82.6	17.3	78.1	12.6	99.1	87.3	78.3
V	84.0	15.9	94.8	53.1	98.8	46.8	94.1
F	55.4	44.5	99.1	35.1	99.6	64.8	98.8
C=20							
N	59.6	40.3	70.0	94.1	17.6	5.82	60.7
S	52.3	47.6	70.7	6.42	97.4	93.5	70.0
V	77.2	22.7	92.8	42.8	98.3	57.1	91.8
F	47.9	52.0	98.5	20.8	99.5	79.1	98.1
C=100							
N	65.2	34.7	58.0	92.6	17.0	7.34	64.4
S	36.5	63.4	75.0	5.33	96.8	94.6	73.6
V	69.7	30.2	93.8	43.9	97.8	56.03	92.2
F	22.9	77.06	98.50	10.76	99.38	89.23	97.91

Predicted and actual values combinations for all classes (N, S, V, and F) using SVM with C=0.1 under the inter-patient scheme are illustrated in the confusion matrix in Figure 7.

Despite high unbalancing between classes and uncorrelated between training and testing examples due to the inter-patient scheme, this approach has potential for clinical use given the recall (sensitivity) metric in identifying four ECG heartbeat classes for arrhythmia detection, where this metric is considered the most reliable in medical diagnosis.

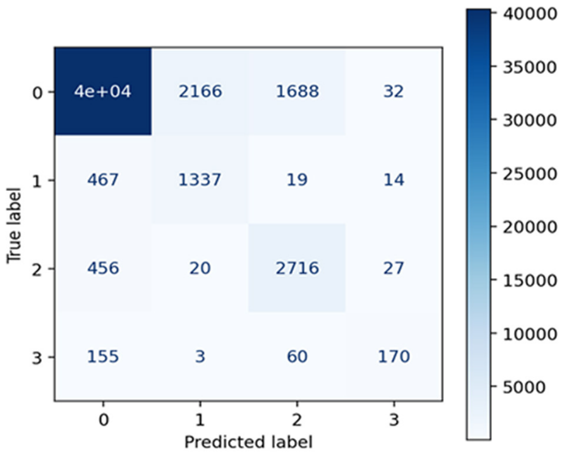

Fig. 7. Confusion matrix for multiclassification under the inter-patient scheme

The presented results in Table 4 show that only class F has a low recall value (high number of misclassification examples). The reason beyond these results is the small number of examples in this class.

In this scheme, only one value of C was considered, equal to 0.1, where a value of more than 0.1 causes noticeable degradation in the model's performance in terms of all metrics. These results are due to independence and uncorrelated between the samples in the testing and training sets. So, the SVM was required to be trained with a small value of C, which led to less penalty of misclassification example and a large distance of decision boundary, i.e., a large margin to overcome the overfitting caused by the inter-patient scheme and unbalancing problem. Where the small margin due to the large value of C produces a hard separated boundary between classes, which causes overfitting in the case of inter-patient (unseen data during the training process), then this problem needs to maximize the margin to increase the generalization.

As a result of a metric, the choice of the C parameter depends on the dataset's characteristics, and it should be tuned after deep database analyses.

The impact of parameter C on the performance of the proposed model under the inter-patient scheme in terms of sensitivity was visualized in Figures 8 and 9.

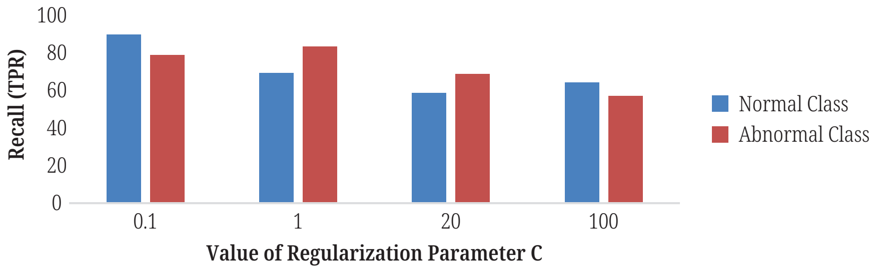

Fig. 8. Sensitivity metric at different values of under different values of regularization parameter C for binary classification under the inter-patient scheme

Fig. 9. Sensitivity metric at different values of under different values of regularization parameter C for multiclassification under the inter-patient scheme

6.3 Comparative study

The comparative study between our proposed classification model and previous related works of ECG beats classification to predict arrhythmia disease under intra and inter-patient schemes has been produced and explained in terms of the classifier and its architecture, resultant classification accuracy, and number of extracted features as in Tables 5 and 6.

Based on the intra-patient scheme, the proposed classification technique exhibits exceptional accuracy and efficiency performance compared to other works' classifiers. Furthermore, it achieves higher accuracy with fewer coefficients and less computational complexity.

Compared to the classification technique discussed in [30], the proposed technique in this paper may have lower accuracy but is more robust against unbalancing problems. The other technique in [30] only balanced 14,878 ECG beats across all classes, which limits its reliability in handling unbalancing problems. Our technique, on the other hand, excels in its generality and overcomes the unbalancing problem more effectively. Therefore, our proposed technique is superior in terms of overall performance.

The proposed work utilizes WT to extract only 25 coefficients after the fourth decomposition level. However, in [31], the extracted features (input dimensions) were 62 and 356, respectively. As explained in Section 5, this makes the proposed technique less complex than the technique in [31]. Furthermore, the proposed work achieves a higher average recall value than in [31].

In [32], a total of 44 features were extracted, which were a combination of heartbeat interval, temporal, wavelet transform, empirical mode decomposition (EMD), and variational mode decomposition (VMD) features. It should be noted that the classifiers in that reference performed better compared to the proposed ones used in this article. However, the work in [32] utilized only 10274 training and 2203 testing samples, and the datasets were balanced. Additionally, like [30], this article should have addressed the issue of unbalanced datasets, which our proposed technique overcomes, making it more versatile and generic.

Table 5. The comparison between the proposed classification model under the intra-patient scheme and previous works on the MIT-BIH arrhythmia database regarding the classifier, dataset, input size, and evaluation metrics

REF.	Classifier	Number of Extracted Features	Evaluation Metrics		
			Accuracy(%)	R(%)	P(%)
[32]	DNN	Two heartbeat interval features, 4 temporals, 20 WT, 6 EMD and 12 VMD features	99	98	99
	RF		99	98	99
[31]	RF	8 levels of DWT and PCA (356 features)	98	60.00	98.2
[30]	SVM	12 Approximate coefficients at the sixth level	99	96	99
The proposed work	SVM+KNN	25(16 detail coefficients at the fourth level+9 PSD coefficients)	98	85	91

Table 6 illustrates how the efficiency of the OWSK model under the inter-patient scheme can be attributed to the input signal length and the classifier's complexity. Compared to other works that utilized the entire length of the ECG beat, the proposed model had a shorter input feature vector, directly impacting the computational complexity during both the training and testing phases, as discussed in section (5). On the other hand, all mentioned previous work used all training samples without any data reduction, affecting the complexity. In the proposed model, the OSS model reduces the dataset samples to half the original sample.

Even though the study in [15] showed slightly better results for the N, S, and V classes, our proposed work outperformed it in class F. Additionally, our proposed work used fewer features, resulting in lower computational complexity than the study in [15]. Therefore, our proposed work has the advantage of being simpler. Based on general knowledge, the models based on deep learning networks are greedy for data. The reduction of data degrades the performance of these models, and the extensive data increases their complexity simultaneously.

In addition, the architecture and computation of deep learning models are generally more than machine learning models. Hence, the proposed model is superior to the previous work in complexity, efficiency, and performance evaluation.

In terms of training time, the deep learning model needs very high time since it depends on the number of samples in the training dataset, the length of each sample (which is equal to the input layer size), the number of layers, the number of neurons in each layer, number and size of the kernel in case of CNN, and the number of epochs, and the more accurate results required a large number of layers, a more significant number of kernels, and more epoch. In contrast, the proposed training time and the space complexity depend only on the dataset size, which is reduced using OSS methods.

However, the KNN needs a high prediction time since it depends on the number of samples in the dataset; the proposed model solved this problem by training KNN on only minority classes, which contain only 5144 samples which certainly does not exceeds the prediction time of the deep learning network.

So, using a machine learning-based model with reduced data and the same performance overcomes this problem and increases the model efficiency as in the proposed model.

The further comparison in terms of precision (P), recall (R), and false positive rate (FPR) for each class is produced in Table 7.

Generally, the proposed model achieved the best results for all classes compared to the results obtained by previous works.

Table 6. The comparison between the proposed classification model under the inter-patient scheme and previous works on the MIT-BIH arrhythmia database regarding the classifier, dataset, input size, and evaluation metrics

Ref.	Classifier	Length of an Input Feature Vector	Evaluation Metrics		
			Accuracy(%)	R(%)	P(%)
[11]	densely connected convolutional neural network (DenseNet) and gated recurrent unit network (GRU)	8 data points of ECG signal	–	–	–
[12]	2-D convolution neural network	Image of 150*150 pixels	93.7	–	–
[13]	Deep convolutional neural networks (CNN)	Entire ECG with a length of 130 datapoint	90	–	–
[14]	deep convolutional neural network	Entire ECG with a length of 170 datapoint	88	69	55
[15]	adversarial domain adaptation based deep learning network	Entire ECG with a length of 411 datapoint	92	73	68
Proposed work	SVM+KNN	25(16 detail coefficients at the fourth level+9 PSD coefficients)	90	71	68

Table 7. The comparison between the proposed classification model and previous works in terms of evaluation metrics for each class in the dataset

Ref.	N			S			V			F		
	P	R	FPR	P	R	FPR	P	R	FPR	P	R	FPR
[11]	–	–	–	61	63	–	88	91	–	–	–	–
[12]	96	97	30	73	54	0.8	84	84	1.1	0.2	0.3	1.2
[13]	97	92	23	56	62	2	51	89	6	2	0.5	0.15
[14]	98	88	9	30	82	7	72	92	2.4	26	68	1.5
[15]	97	93	–	73	76	–	57	85	–	44	38	–
proposed _OWSK	97	91	19	38	73	4	61	84	3	70	44	0.14

7 CONCLUSION

This paper proposes a novel ECG beats classification model based on a cascading technique for arrhythmia detection to address the imbalance and inter-patient ECG beats arrhythmia classification problems. The OWSK model was designed from three stages: preprocessing the data and under-sampling; it is based on a one-sided selection method, feature extraction and reduction using Wavelet transform (WT), and a two-cascaded stages-based classifier is constructed from SVM and KNN machine learning algorithms. The proposed model was implemented according to two schemes, inter-patient and intra-patient, and the comparative study for the results following these two schemes were produced to prove the generalization performance of the proposed model. In conclusion, the value of parameter C in SVM significantly impacts the results where its value was different in each scheme because of the variable distribution of data in the two paradigms and the degree of correlation between the training and testing data following each paradigm. Generally, the proposed model achieved promising results in accuracy and recall metrics for each class of arrhythmia. The proposed model was the best compared to other implemented models and those offered by related work in improving the inter-patient problem and efficiency with less computation complexity since it depends on a simple machine learning classifier and fewer data samples and dimensions.

There are some limitations with insufficient data belonging to a minority type of arrhythmia; in the future, we plan to solve the imbalance problem using a more efficient technique and apply our proposed model to another dataset and another application.

8 REFERENCES

[1] A. A. Ahmed, W. Ali, T. A. Abdullah, and S. J. Malebary, "Classifying cardiac arrhythmia from ECG signal using 1D CNN deep learning model," *Mathematics,* vol. 11, no. 3, pp. 562–577, 2023, https://doi.org/10.3390/math11030562

[2] A. A. Fadhel and H. M. Hasan, "Reducing delay and packets loss in IoT-Cloud based ECG monitoring by Gaussian modeling," *International Journal of Online & Biomedical Engineering,* vol. 19, no. 6, 2023, https://doi.org/10.3991/ijoe.v19i06.38581

[3] V. A. Ardeti, V. R. Kolluru, G. T. Varghese, and R. K. Patjoshi, "An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to AI-based approaches," *Expert Systems with Applications,* p. 119561, 2023, https://doi.org/10.1016/j.eswa.2023.119561

[4] A.-D. Abdou, N. F. Ngom, and O. Niang, "Arrhythmias prediction using an hybrid model based on convolutional neural network and nonlinear regression," *International Journal of Computational Intelligence and Applications,* vol. 19, no. 03, p. 2050024, 2020, https://doi.org/10.1142/S1469026820500248

[5] J. Jiang, H. Zhang, D. Pi, and C. Dai, "A novel multi-module neural network system for imbalanced heartbeats classification," *Expert Systems with Applications: X,* vol. 1, p. 100003, 2019, https://doi.org/10.1016/j.eswax.2019.100003

[6] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, and P. Zhang, "Inter-patient ECG arrhythmia heartbeat classification based on unsupervised domain adaptation," Neurocomputing, vol. 454, pp. 339–349, 2021, https://doi.org/10.1016/j.neucom.2021.04.104

[7] J. Rahul, M. Sora, L. D. Sharma, and V. K. Bohat, "An improved cardiac arrhythmia classification using an RR interval-based approach," *Biocybernetics and Biomedical Engineering,* vol. 41, no. 2, pp. 656–666, 2021, https://doi.org/10.1016/j.bbe.2021.04.004

[8] Z. C. Oleiwi, E. N. AlShemmary, and S. Al-augby, "Efficient ECG beats classification techniques for the cardiac arrhythmia detection based on wavelet transformation," *International Journal of Intelligent Engineering & Systems,* vol. 16, no. 2, 2023, https://doi.org/10.22266/ijies2023.0430.16

[9] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," *IEEE Access,* vol. 9, pp. 103452–103464, 2021, https://doi.org/10.1109/ACCESS.2021.3098986

[10] S. M. Mangj, P. H. Hussan, and W. M. R. Shakir, "Efficient deep learning approach for detection of brain tumor disease," *International Journal of Online & Biomedical Engineering,* vol. 19, no. 6, 2023, https://doi.org/10.3991/ijoe.v19i06.40277

[11] L. Guo, G. Sim, and B. Matuszewski, "Inter-patient ECG classification with convolutional and recurrent neural networks," *Biocybernetics and Biomedical Engineering,* vol. 39, no. 3, pp. 868–879, 2019, https://doi.org/10.1016/j.bbe.2019.06.001

[12] K. Ye, "Inter-patient electrocardiogram heartbeat classification with 2-D convolutional neural network," Master's Thesis, University of Victoria, 2021, [Online]. Available: http://hdl.handle.net/1828/12586

[13] J. Takalo-Mattila, J. Kiljander, and J.-P. Soininen, "Inter-patient ECG classification using deep convolutional neural networks," in *2018 21st Euromicro Conference on Digital System Design (DSD),* 2018: IEEE, pp. 421–425, https://doi.org/10.1109/DSD.2018.00077

[14] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Systems with Applications,* vol. 122, pp. 75–84, 2019, https://doi.org/10.1016/j.eswa.2018.12.037

[15] L. Niu, C. Chen, H. Liu, S. Zhou, and M. Shu, "A deep-learning approach to ECG classification based on adversarial domain adaptation," in *Healthcare*, vol. 8, no. 4, p. 437, 2020, https://doi.org/10.3390/healthcare8040437

[16] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine,* vol. 20, no. 3, pp. 45–50, 2001, https://doi.org/10.1109/51.932724

[17] X. Luo, L. Yang, H. Cai, R. Tang, Y. Chen, and W. Li, "Multi-classification of arrhythmias using a HCRNet on imbalanced ECG datasets," *Computer Methods and Programs in Biomedicine,* vol. 208, p. 106258, 2021, https://doi.org/10.1016/j.cmpb.2021.106258

[18] D. Sisodia and D. S. Sisodia, "A hybrid data-level sampling approach in learning from skewed user-click data for click fraud detection in online advertising," *Expert Systems,* vol. 40, no. 2, p. e13147, 2023, https://doi.org/10.1111/exsy.13147

[19] J.-B. Wang, C.-A. Zou, and G.-H. Fu, "AWSMOTE: An SVM-based adaptive weighted SMOTE for class-imbalance learning," *Scientific Programming,* vol. 2021, pp. 1–18, 2021, https://doi.org/10.1155/2021/9947621

[20] E. H. Houssein, I. E. Ibrahim, N. Neggaz, M. Hassaballah, and Y. M. Wazery, "An efficient ECG arrhythmia classification method based on Manta ray foraging optimization," *Expert systems with applications,* vol. 181, p. 115131, 2021, https://doi.org/10.1016/j.eswa.2021.115131

[21] W. Xie, G. Liang, Z. Dong, B. Tan, and B. Zhang, "An improved oversampling algorithm based on the samples' selection strategy for classifying imbalanced data," *Mathematical Problems in Engineering,* vol. 2019, 2019, https://doi.org/10.1155/2019/3526539

[22] M. V. Gowrishankar, K. Sudhakara Pandian, R. Deivasigamani, and S. Chun Kit Ang, "A novel SVM and K-NN classifier based machine learning technique for epileptic seizure detection," *International Journal of Online and Biomedical Engineering (iJOE),* vol. 19, no. 07, pp. 99–124, 2023, https://doi.org/10.3991/ijoe.v19i07.37881

[23] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-Nearest neighbour, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analytics Journal,* vol. 3, p. 100071, 2022, https://doi.org/10.1016/j.dajour.2022.100071

[24] P. Dhivya and A. Bazilabanu, "Deep hyper optimization approach for disease classification using artificial intelligence," *Data & Knowledge Engineering,* vol. 145, p. 102147, 2023, https://doi.org/10.1016/j.datak.2023.102147
[25] A. Aballe, M. Bethencourt, F. Botana, and M. Marcos, "Wavelet transform-based analysis for electrochemical noise," *Electrochemistry communications,* vol. 1, no. 7, pp. 266–270, 1999, https://doi.org/10.1016/S1388-2481(99)00053-3
[26] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen, "Discrete wavelet transform-based time series analysis and mining," *ACM Computing Surveys (CSUR),* vol. 43, no. 2, pp. 1–37, 2011, https://doi.org/10.1145/1883612.1883613
[27] Z. M. Hussain, A. Z. Sadik, and P. O'Shea, "*Digital signal processing: an introduction with MATLAB and applications*," New York: Springer Science & Business Media, 2011.
[28] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (SVM) in LibSVM," *International Journal Computer and Application,* vol. 128, no. 3, pp. 28–34, 2015, https://doi.org/10.5120/ijca2015906480
[29] B. Gawri, A. Kasturi, L. B. M. Neti, and C. Hota, "An efficient approach to KNN algorithm for IoT devices," in *2022 14th International Conference on Communication Systems & Networks (COMSNETS)*, 2022: IEEE, pp. 734–738, https://doi.org/10.1109/COMSNETS53615.2022.9668540
[30] C. K. Jha and M. H. Kolekar, "Cardiac arrhythmia classification using tunable Q-wavelet transform based features and support vector machine classifier," *Biomedical Signal Processing and Control,* vol. 59, p. 101875, 2020, https://doi.org/10.1016/j.bspc.2020.101875
[31] S. Nurmaini, B. Tutuko, M. N. Rachmatullah, A. Darmawahyuni, and F. Masdung, "Machine learning techniques with low-dimensional feature extraction for improving the generalizability of cardiac arrhythmia," *IAENG International Journal of Computer Science,* vol. 48, no. 2, pp. 369–378, 2021.
[32] S. Sahoo, P. Dash, B. Mishra, and S. K. Sabut, "Deep learning-based system to predict cardiac arrhythmia using hybrid features of transform techniques," *Intelligent Systems with Applications,* vol. 16, p. 200127, 2022, https://doi.org/10.1016/j.iswa.2022.200127

9 AUTHORS

Zahraa Ch. Oleiwi received a B.Sc. degree in Computer and Software Engineering from AL–Mustansiriyah University, Iraq, and an M.Sc. degree in Computer Science from the University of Kufa Iraq in 2003 and 2017, respectively. From 2003 to 2014 she lectured on computer and software engineering and worked as an engineer at the University of Al-Qadisiyah, Iraq. In 2017, she was promoted to Associate Lecturer and lectured on signal processing. Oleiwi has many publications on signal processing and image processing. Her interests are in signal processing, programming, computer security, data processing, image processing, machine learning, deep learning, AI, and data mining (email: zahraac.albadri@student.uokufa.edu.iq).

Ebtesam N. Alshemmary, Prof., Ph.D., received her MSc. and Ph.D. from the University of Technology, Baghdad, in 1999 and 2007. She is the IT-RDC manager at the University of Kufa. Her general specialization is computer and control engineering, and her specific specialization is mechatronics engineering (robotics and image processing). She is incredibly well known for her profound contributions to different fields of computer programs, biometrics, and automatic identification systems. AlShemmary's research interests include statistical analyses of health, medical studies, and information technology. She has published research in international and local journals in different fields of computer science. AlShemmary worked as a rapporteur in the computer department of the Faculty of Education for Girls at the University of Kufa in Iraq. She is also a member of many committees at Kufa University, such as scientific and examination committees. AlShemmary is a member of the Editorial and Reviewer Board of the International Journal of Electronics and Computer Science Engineering (IJECSE). She is also a member of the Reviewer Board of the European Journal of Science and Engineering (EJSE), and the Editorial Board of the International Journal of Advanced Computer Science and Information Technology (IJACSIT) (email: dr.alshemmary@uokufa.edu.iq).

Salam Al-Augby received his B.Sc. Degree in Electronic and Electrical Engineering from MEC in 1997 and a Master's degree in Computer Science from the University of Technology, Iraq in 2005. He received his Ph.D. degree in IT in Management from the University of Szczecin in Szczecin, Poland, in 2015. His areas of interest are data analysis, data mining, text mining, behavioral finance, sentiment analysis, IT in management, big data analysis, social media analysis, and natural language processing (email: salam.alaugby@uokufa.edu.iq).