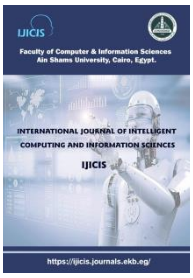
IJICIS, Vol.22, No.2, 44-62
DOI: 10.21608/ijicis.2022.105574.1137

International Journal of Intelligent Computing and Information Sciences
https://ijicis.journals.ekb.eg/



DEEP LEARNING APPROACH BASED ON TRANSFER LEARNING WITH DIFFERENT CLASSIFIERS FOR ECG DIAGNOSIS

Mahmoud. M. Bassiouni*
Computer Science,
Faculty of Computer and Information Technology, Egyptian E-learning University,
El-Giza, Egypt
mbassiouni@eelu.edu.eg

Islam Hegazy
Computer Science departement,
Faculty of Computer and Information Science, Ain shams University, Abbassia
Cairo, Egypt
islheg@cis.asu.edu.eg

Nouhad Rizk
Director of Undergraduate Studies,
Computer science,
Houston Unversity,
Houston, USA
njrizk@central.uh.edu

Sayed. A. El-Dahshan
Physics department,
Faculty of Science,
Ain shams University, Abbassia
Cario, Egypt
seldahshan@sci.asu.edu.eg

Abdelbadeeh M. Salem
Computer Science departement,
Faculty of Computer and Information Science, Ain shams University, Abbassia
Cairo, Egypt
absalam@cis.asu.edu.eg

Received 2021-11-11; Revised 2022-03-04; Accepted 2022-03-09

Abstract: *Heart diseases are one of the main reasons that cause human death. The early-stage detection of heart diseases can prevent irreversible heart muscle damage or heart failure. Electrocardiogram (ECG) is one of the main heart signals that can be useful in early diagnosis because of its obvious peaks and segments. This paper focuses on using a methodology depending on deep learning for the diagnosis of the electrocardiogram records into normal (N), Supraventricular arrhythmia (SV), ST-segment changes (ST), and myocardial infarction (MYC) conditions. The continuous wavelet transform (CWT) converts the ECG signals to the time-frequency domain to compute the scalogram of the ECG signals and for the conversion of ECG signal from one dimension signal to a two-dimension image. In addition to this, a pertained model using transfer learning is applied based on Resnet50. Moreover, three main classifiers are verified to estimate the accuracy of the proposed system which are based on the Softmax, Random Forest (RF), and XGBoost classifier. An experiment is applied for the diagnosis of four main kinds of ECG records. Finally, the results based on the class-oriented schema achieved an accuracy of 98.3% based on Resnet50 with the XGBoost classifier. The comparison with the related previous work presented the excellent performance of the proposed methodology as it can be applied as a clinical application.*

Keywords: *Cardiovascular diseases (CVD), Electrocardiogram (ECG), Continuous wavelet transform (CWT), Convolution neural network, XGBoost Classifier*

* Corresponding author: Mahmoud. M. Bassiouni
Computer Science, Faculty of Computer and Information Technology, Egyptian E-learning University, El-Giza, Egypt
E-mail address: mbassiouni@eelu.edu.eg

1. Introduction

Heart diseases are one of the main reasons for death worldwide and they are sometimes called cardiovascular diseases (CVD). Various people suffer and die from heart diseases annually based on recent research and survey studies. In 2022 [1], it is estimated that about 17.9 million people died from CVD, and this represents about 32% of the global death, and about 85% of these people have died from heart attack and stroke. Moreover, CVD was responsible for 38% of all premature deaths (under the age of 70) due to non-communicable diseases. About 3 quarters of the deaths caused by CVD occur in the low-and middle-income countries. Arrhythmia is one of the salient groups of CVDs. They represent the abnormal electrical conduction or impulse origin in the heart. Most of the arrhythmias are non-life-threatening, while some of them can cause many cardiovascular complications and sudden death. The early diagnosis of arrhythmia can assist in preventing sudden death and help in treating many different cardiovascular diseases. Physicians, experts, and doctors detect arrhythmias based on electrocardiograms (ECG) signals. The ECG measures the variations in the electrical potential in one cycle of the heartbeat. A single ECG signal consists of a group of peaks defined by P, Q, R, S, and T. Moreover, various types of arrhythmia do not appear in a short time and may require a large amount of ECG heartbeats. As a result, a diagnosis method automated should be investigated for the identification of different ECG records and this is the main focus of the proposed methodology.
Several methodologies based on machine learning have been built for extracting features and classifying ECG records. On one hand, extracting features from ECG signals is essential before the classification process because it provides a great impact on the results of the classification. P-QRS-T segment and RR interval were used in almost every research [2]. In addition to this, there are other conventional features extracted from the ECG based on morphological features, wavelet transform features, higher-order statistics, random projection, and wavelet packet entropy. These methodologies require providing a hand-crafted feature before applying any conventional classifier. There are several disadvantages in these processes of feature extraction which are depthless, large time-consuming and they lack any implicit knowledge. On the other hand, several numbers of classifiers were applied such as a k-nearest neighbor, artificial neural network, support vector machine, random forest, and Gaussian mixture models. When these conventional features are fit to these conventional classifiers they suffer from overfitting obstacles. The main causes for overfitting are as follows: (i) noise and unclean data used for training (ii) high variance and complexity of the model (iii) size of the training set is not enough (iv) learning from a small dataset. Deep learning (DL) is preserved to be part of machine learning. It is known by the word “deep” because the network structure consists of many hidden layers [3]. The main concept in DL is that the low-level features are integrated to obtain high-level features. In DL no hand-craft features are obtained and implicit knowledge can be learned easily. DL has also been used in some of the ECG studies, and it showed excellent classification results in diagnosis. Several DL structures were used such as recurrent neural network, stacked denoising auto-encoder, deep neural network, convolutional neural networks, and restricted Boltzmann machine. Finally, based on the advantages of the DL the proposed methodology used the merits of the DL and delivered the following contributions.
The contributions stemming from this paper are three-main folds:

1. The proposed DL is used to diagnose four main ECG records based on balanced datasets of records.

2. In addition to this, hand-craft features are extracted from the continuous wavelet transform (CWT) and combined with automatic features that are obtained with the CNN to reach an efficient performance in the classification.
3. Investigation of the XGBoost for the classification process to achieve an accurate diagnosis performance.

The manuscript is summarized based on different sections. Related works are presented in section 2, while the proposed methodology in terms of capturing ECG records, filtering the ECG signals, extracting features, and classifying the ECG records is presented in section 3. Moreover, the main results and the discussion are illustrated in sections 4 and 5 respectively. Finally, section 6 manifests the conclusion and the future directions.

2. Related Work

There exist several studies for the diagnosis of the ECG signals. Some studies used traditional and classical methods in extracting features and classification, while other studies applied deep learning methods for the classification of the ECG with the aim of diagnosis. V. K. Sudarshan et al. (2017) [4] provided an automatic system for the prediction of CHF, and they used two datasets which are BIDMC, MIT-BIH Normal Sinus rhyme (NSR), fantasia databases. The technique used for feature extraction was dual-tree complex wavelets transform (DTCWT). The DTCWT was applied on ECG segments for 2 seconds aiming in obtaining six levels of coefficients. The classification was performed using k-nearest neighbor (KNN) and decision trees (DT). R. Mahajan, et al. (2017) [5] concentrated on implementing techniques for diagnosing cognitive heart failure (CHF) for avoiding events that can threaten the life of the individual. A model based on a probabilistic symbol pattern recognition to detect the individuals in CHF by discretizing each continuous R-R interval onto a symbolic representation, and the classification process was based on bagging decision trees. A. D Dolatabadi, et al. (2017) [6] applied a method for determining the abnormal and normal individuals based on ECG signals. This study used very limited data for training, and the data applied was long-term ST. The features were extracted from heart rate variability based on non-linear and time-frequency domains. Then, the PCG is applied for reducing the dimensions of the features. Finally, the classification process was based on SVM. N. Henzel et al. (2017) [7] obtained data from the MIT-BIH arrhythmia dataset, and the features were extracted using the RR ECG intervals, and then a proposed linear model is designed for classification. K. H. Boon, et al. (2018) [8] designed an algorithm based on the optimization techniques aiming for the diagnosis of a kind of atrial fibrillation (AF) known as paroxysmal (PAF). The data were collected from the atrial fibrillation prediction database (AFPDB). Their methodology depends on three main phases which are HRV pre-preprocessing, feature extraction, and classification. The filtering was based on heart rate correction, interpolation, detrending, while the feature extraction depends on time-domain analysis, non-linear analysis, and frequency domain analysis, and the features produced are optimized using a proposed optimization technique based on NSGA-II, and finally the classification was based on SVM.

J. H. Tan.et al. (2018) [9] implemented a way for the verification of the ECG signals diseases, especially the CAD. The data were collected from an online dataset known as Fantasia and the institute of cardiology technics arrhythmia. They applied an RNN model based on LSTM and CNN as the classifier, and they applied these techniques on the physio net database obtaining an accuracy of 99.85%. M. Kumar et al. (2018) [10] applied a nonlinear methodology depending on an analytic wavelet transform (AWT), and then the log energy and permutation entropies abbreviated by LEE and ($\boldsymbol{PE_n}$) respectively are applied for further feature extraction on each ECG sub-band. The features of LEE provided high classification results than the features of $\boldsymbol{PE_n}$ inaccuracy performance. The classifiers used were RF and J48 in the diagnosis process. B. Safarbali et al. (2019) [11] worked on using time of life features and these features rely on persistent homology and fractal dimension. The ECG signals are gathered from MIT-BIH arrhythmia and long-term AF datasets. The MLP was used as a classifier in the diagnosis of normal and AF patients.

A model was proposed by L. Yang et al. (2020) [12] depending on a hybrid convolutional neural network (HCNN) and there was the feeding of this network with RR intervals at 3 different positions. Five databases were used for the application of this work: Ventricular contraction beat (VCB), premature contraction beat (APC), left bundle branch block beat (LBBB), right bundle branch block beat (RBBB), and normal sinus beat. This combination showed an improved diagnosis performance. A model proposed by R. R. Janghel et al. (2020) [13] depending on machine learning techniques for the differentiation of abnormal and normal beat, the data were obtained from the MIT-BIH arrhythmia. In terms of classification, there was the employment of various machine learning methods such as SVM, naïve Bayes, Ada-boast, decision tree, KNN, and RF. An approach was applied by Y. Li et al. (2020) [14] to identify the ECG signals depending on F-CNN for extracting features of ECG heartbeats, and M-CNN was employed by the second tool for classification and they are cascaded together as the final identification network. Five public datasets from Physio net were used to conduct the experiments for evaluating the performance of the method proposed. A method was proposed by C. Chen et al. (2020) [15] based on 15-layer deep CNN combined LSTM, and multi-input network and the performance of QRS detection. The data were gathered from 4 main databases which are the MIT-BIH arrhythmia, CinC challenge database, MIT-BIH NSR, and MIT-BIH AF. They aimed to classify six different types of ECG heartbeats which are normal, AF, ventricular bigeminy (B), atrial flutter (AFL), sinus bradycardia (SBR), and pacing rhythm (P) of ECG heartbeats based on the combination proposed.

Another methodology was proposed by E. H. Houssein, et al. (2021) [16] for the diagnosis of various ECG heartbeats. The ECG heartbeats were collected from MIT-BIH arrhythmia. The methodology is based on the combination of feature extraction techniques such as local binary pattern (LBP), higher-order statistical (HOS), wavelets, and morphological information. The classification was based on manta ray foraging optimization (MRFO) for optimizing the SVM classifier. Finally, a methodology was proposed by F. M. Dias et al (2021) [17] for the diagnosis of three ECG heartbeats from the MIT-BIH arrhythmia dataset which is normal, supraventricular ectopic beat (SE), ventricular ectopic beat (V). The features were based on a combination of RR intervals, signal morphology, and HOS. The method was validated using the inter-patient paradigm. Table 1 shows a summary of the machine and deep learning methodologies used for the diagnosis of different heart diseases based on ECG signals in terms of subjects, records, heartbeats, datasets, methodology, and the final accuracy performance.

Table 1 Summary of Machine and Deep Learning Methodology for ECG diagnosis

Authors & Year	S/ R/ HB	Methodology	Databases	Performance
V. K. Sudarshan et al. [4] 2017	15 S from CHF + 18 S from NSR 40 S from Fantasia all normal	Features: Dual tree complex wavelet (DTCWT) + statistical ranked features Classifiers: KNN, DT	BIDMC CHF, MIT-BIH NSR, Fantasia	2 classes diagnosis ACC = 99.86% SEN = 99.78% SPEC = 99.94%
R. Mahajan et al. [5] 2017	107 S 69 S normal + 38 S abnormal from CHF	Features: probabilistic symbol pattern recognition Classifiers: Bagged decision trees	R-R Interval database	2 classes diagnosis ACC = 99.5% SPEC = 100% SEN = 98.57%
A. D. Dolatabadi, et al. [6] 2017	86 R from 80 S abnormal from CAD + 24 R from 54 S normal	Features: Features obtained from HRV in time frequency and non-linear domains Classifiers: SVM	MIT-BIH NSR, Long term SR database	2 classes diagnosis ACC = 99.2% SEN = 98.43% SPEC = 100%
N. Henzel et al. [7] 2017	25 R	RR intervals + generalized linear model	AFDB	2 classes diagnosis ACC=93.28% SEN=90.34% SPEC=95.46%
K. H. Boon et al.[8] 2018	106 data from 53 R pairs	Features: Time domain, spectral, Bispectrum, nonlinear dynamics features Classifiers: SVM	Atrial Fibrillation prediction (AFPDB) Database	2 classes diagnosis ACC = 87.7%
J. H. Tan et al. [9] 2018	Normal: 40 S, 32000 HB + CAD: 7S, 6120 HB	8 layers stacked CNN-LSTM based on a blindfold strategy	Fantasia, St Petersburg Institute of Cardiology Technis 12-leads arrhythmia	2 classes diagnosis Two Scenarios First Scenario: ACC = 99.85% Second Scenario: ACC = 95.76%
M. Kumar et al. [10] 2018	243 R has N rhythms 247 R has AF	Features: Log energy entropy (LEE)+ Classifiers: RF, J48	MIT-BIH AF	2 classes diagnosis For J48: ACC = 95.72% For RF: ACC=96.8% SEN=95.8% SPEC=97.6%
B. Safarbali et al. [11] 2019	AF : 38 S Normal: 20 S	Features: Fractal dimension and persistent homology to obtain (TOF) features Classifiers: ANN	MIT-BIH, Long-Term AF database	2 classes diagnosis ACC=93.0%
L. Yang et al. [12] 2020	2102 S of abnormal patients	HCNN + RR intervals at 3 different positions	MIT-BIH arrhythmia database	5 classes diagnosis For HCNN: ACC = 86.61% SEN =76.92% SPEC = 78.46% For CNN: ACC = 83.30%
R. R. Janghel et al. [13] 2020	47 R 40% of the 47 R records are patients	Naïve Bayes SVM Ada-boost RF, Decision Tree, and KNN	MIT-BIH arrhythmia database	2 classes diagnosis ACC of the Decision Tree = 88.2%
Y. Li et al. [14] 2020	Normal: 78 S ST: 28 S AF: 23 S	Cascaded CNN based on F-CNN and M-CNN	Fantasia, CEDSDB, MIT-BIH NSR,ST,AF	3 classes diagnosis Cascaded ACC = 94.3%
C. Chen et al. [15] 2020	Normal: 144149 HB AF: 17135 HB B: 144 HB, P: 631 HB AFL: 60 HB, SBR: 180 HB	15 deep CNN layers combined with LSTM	MIT-BIH Challenge 2017 AFDB NSRDB	6 classes diagnosis ACC = 97.15% SEN = 97.11% SPEC = 97.06%
E. H. Houssein, et al. [16] 2021	22 R for Train 22R for Test Forming 110,000 HB	LBP + wavelets + HOS + morphological information + MRFO with SVM	MIT-BIH arrhythmia database	4 classes diagnosis ACC: 98.26% SEN: 97.43%
F. M. Dias, et al. [17] 2021	29 R for Train 15 R for Test 48,856 HB Normal: 45,101 HB SE:915HB, V: 3740 HB	RR intervals + signal morphology + higher-order statistics	MIT-BIH arrhythmia database	3 classes diagnosis SEN: 90.43%

S: Subject **R**: Record **HB**: Heartbeat

3. Methodology

The methodology consists of four main stages which are obtaining ECG data, filtering ECG signals, extracting various ECG features, and classifying ECG records. In the data acquisition phase, six main data sets online are downloaded that consist of four different ECG heartbeats. In the second stage, filtering or de-noising is performed on the ECG heartbeats as the ECG signal consists of three main common noises which are line drifting, power interference, and noise based on a high frequency.

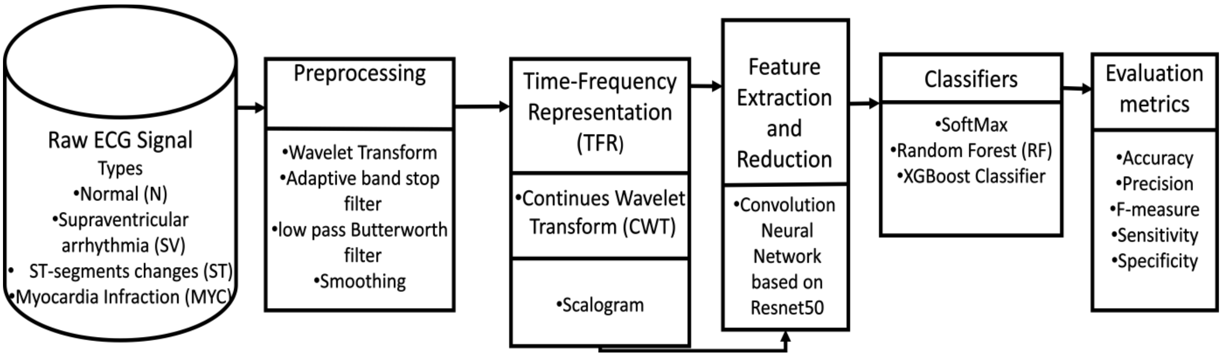


Figure. 1: Proposed Overall Methodology

These distortions are removed using wavelets and a set of filters. The next stage is capturing the features for the ECG heartbeats using a conventional method, and the result of this method act as an input to a pre-trained model which is Resnet50. Finally, in the classification phase, three different classifiers are employed for ECG diagnosis which are Softmax, Random Forest (RF), and XGBoost classifier as shown in Figure. 1.

3.1. Data Acquisition

This stage is one of the most important stages in the methodology proposed. There exist two concepts for capturing the ECG signals. The first concept is the application of a medical device for capturing the ECG heartbeats at different leads, whereas the second way is to download an available ECG signal online. In this methodology, the second way is employed, and the ECG signals are captured from six datasets which are Normal Sinus Rhythm Database (nsrdb) [18], Normal Sinus Rhythm RR Interval Database (nsr2db) [19], MIT-BIH Supraventricular arrhythmia database (svdb) [20], MIT-BIH ST change database (stdb) [21], Long Term ST database (ltsdb) [22] and PTB diagnostic ECG [23], where the number of ECG records in the former datasets are 18, 54, 84, 28, 86, and 549 respectively. Four main different ECG heartbeats were chosen from these datasets. Figure. 2 shows the four ECG types which are one normal and 3 abnormal ECG heartbeats. The ECG heartbeats are in their noising state without any filtration applied to them.

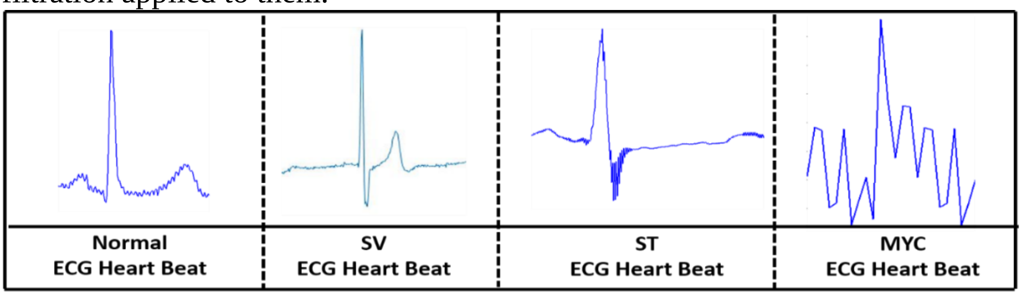


Figure. 2: Four different ECG heartbeats were used for experimentation based on the proposed methodology

3.2. Filteration

ECG signals are always nested in it some distortions and noises that are produced from variant origins. The main three types of noises concentrated in the ECG signals are the drifting in the baseline of ECG signal, interference in the power line, high noise frequency in the main components of the signal, and in some cases, a combination between these types of noises can be found. As a result, a pre-processing chain of filters and wavelets is developed to eliminate these noises by saving the main information of the signal. The pre-processing chain should be summarized in three main tasks which are correcting the drifting in the ECG signal, reducing the interference, selecting low-frequency components, and enhancing the overall signal [24]. The chain contains four main stages based on wavelet drift correction, adaptive bandstop filter, low pass filter, and smoothing. The baseline drift is removed by applying wavelet decomposition with db8 and a decomposition level equal to 9.

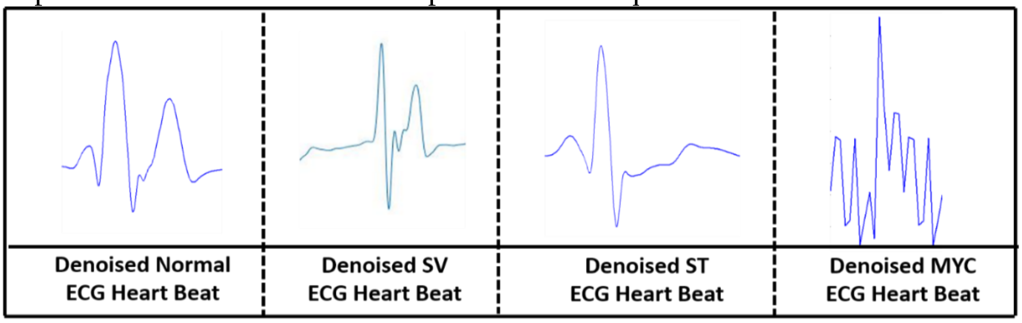


Figure. 3: The Pre-processed ECG heartbeats after applying the preprocessing Chain

Then, the powerline interference is removed using an adaptive bandstop filter with a stopband frequency corner W_s equals to 50Hz. In addition to this, high frequency located in the ECG signals is removed using a low pass Butterworth filter with a passband frequency corner and a stopband frequency corner equal to 40Hz and 60Hz respectively. The values for the passband ripple and stopband ripple attenuation are 0.1dB and 30dB respectively. Finally, a smoothing filter based on Savitzky–Golay (SG) is applied to remove the remaining noise with a smoothing value equal to 5. Figure. 3 presents the filtration results after applying the pre-processing chain on the four ECG heartbeats to express the efficiency of the chain in the removal of noises.

3.3. Feature Extraction

In this stage, the most important features are obtained using a conventional approach based on wavelets, and the feature or coefficients obtained from the wavelets act as an input to three main pre-trained models for extracting more enhanced features for efficient diagnosis.

3.3.1. Conventional Feature Extraction based on CWT

Wavelets transform are preserved as an expansion of the traditional Fourier transform. The Fourier transform operates on a single frequency or a scale, while wavelets operate on multi-scale frequencies [25]. The analysis based on wavelet is a decomposition of any signal into a different version of shift and scale from the original wavelets. In the proposed methodology the type of wavelet that is considered is the continuous wavelet transform (CWT) [26]. The signal is mapped to a time scale domain and then each time scale is mapped to a specific range of frequencies in the frequency domain [27]. The CWT of a signal $\boldsymbol{E(t)}$ is preserved as a transform based on integral for the $\boldsymbol{E(t)}$ with different families of wavelet functions known as $\boldsymbol{\varphi_{x,y}(t)}$:

$$\boldsymbol{CWT(x,y)} = \frac{\mathbf{1}}{\sqrt{\boldsymbol{x}}} \int_{-\infty}^{\infty} \boldsymbol{E(t)} * \boldsymbol{\varphi_{x,y}}\left(\frac{\boldsymbol{t-x}}{\boldsymbol{y}}\right) \boldsymbol{dt} \qquad (\mathbf{1})$$

$$\boldsymbol{x \in R^{+} - \{0\}}, \qquad \boldsymbol{y \in R}$$

$\boldsymbol{\varphi}$ is known as the mother wavelet, while the $\boldsymbol{\varphi_{x,y}}$ is defined as the daughter wavelets. The former wavelets are called the daughter and they are produced from the shifting and scaling process done on the mother wavelets. $\boldsymbol{x}$ and $\boldsymbol{y}$ are known by the scaling and shifting factors respectively. CWT produces various wavelet coefficients defined by $\boldsymbol{C}$. The pre-processed ECG signal acts as an input to the CWT. The CWT is applied to obtain different CWT wavelet coefficients. It is passed to a set of continuous wavelet filter banks, and the wavelet family used in the filter bank is the analytic Morse wavelet. The result produced is a 2D image that holds the CWT coefficients produced for each corresponding ECG record. Figure. 4 shows the scalogram of the ECG records and this scalogram is calculated based on the absolute of the CWT coefficients.

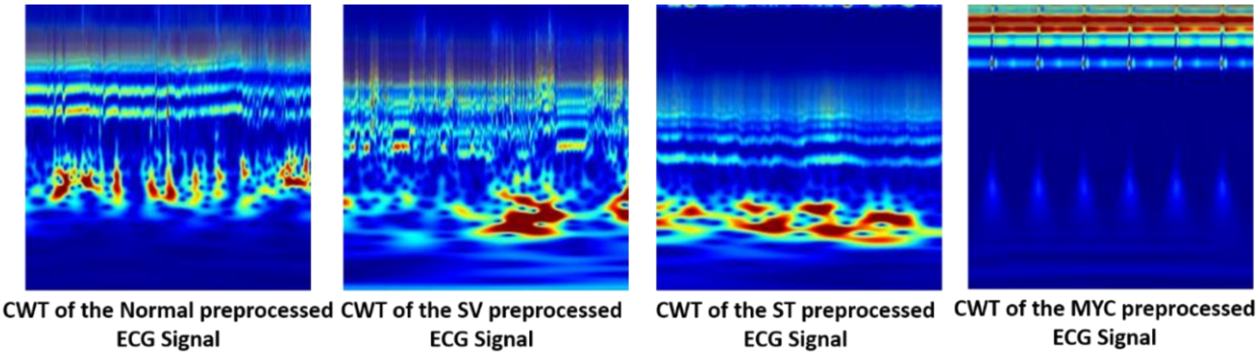


Figure. 4: CWT of the four ECG heartbeats

3.3.2. Automatic Feature Extraction based on Transfer Learning

A series of improvements and enhancements have been achieved with a lot of approaches depending on CNN in different tasks. A real application of CNN is the Alex net model that achieved great success over image net with 10% higher accuracy. After then, a set of equivalent models were proposed for computer vision and signal processing tasks such as SqueezeNet, ShuffleNet, GoogleNet, Resnet18, and more deep and sophisticated models such as VGG16 and VGG19 models. The Resnet50 pre-trained model is applied in the methodology. The output of the CWT is images and they resized with dimensions of 224 x 224 to act as an input for the Resnet50 transfer learning model. Resnet50 is preserved to be one of the essential models in different applications such as computer vision. This pre-trained model is salient in the deep learning world because this framework can train various deep networks. These deep networks consist of several layers and still can result in a high accuracy performance.

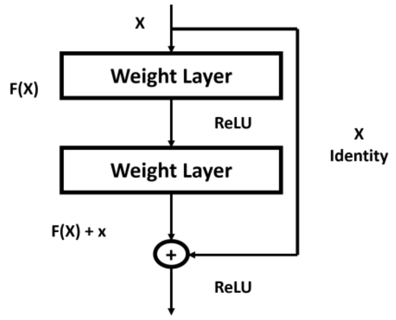


Figure. 5: Residual Mapping for fitting the Resnet50 Layers

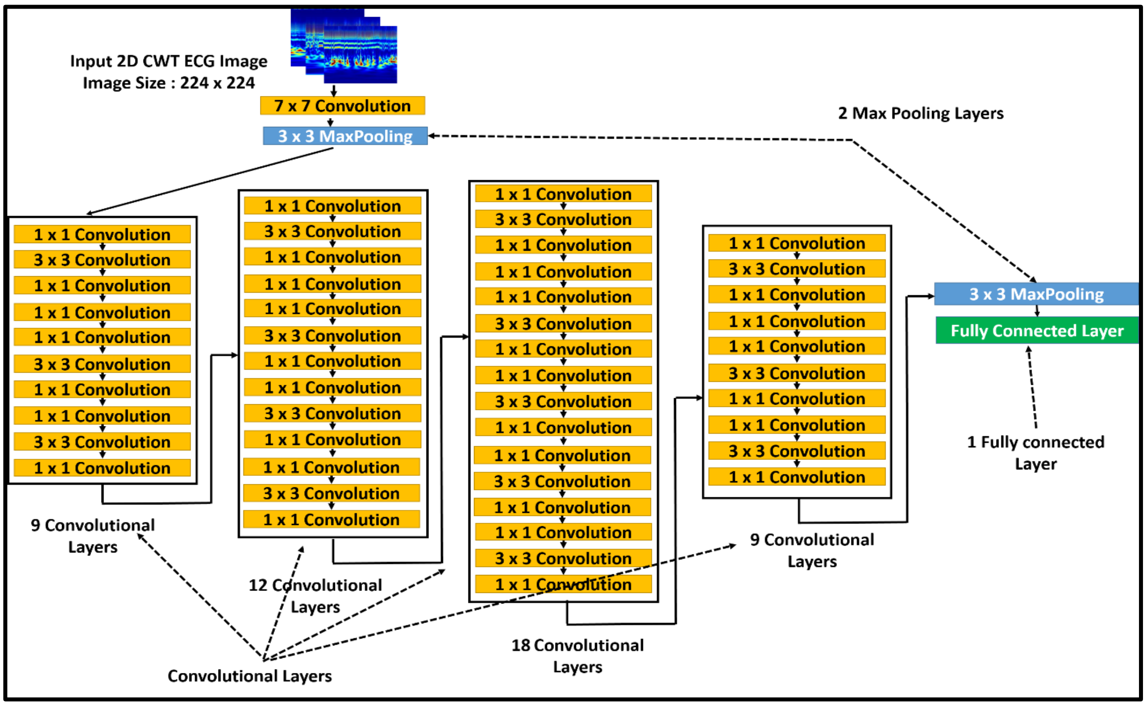


Figure. 6: Resnet50 Architecture

At first, Resnet50 was applied to several image processing tasks, and then it was further used in signal processing tasks. It was proved that it can achieve higher or greater accuracy [28]. Many different researchers claim that adding various layers to the network will always produce higher accuracy performance [29], but this was incorrect because based on experimentation deep learning models with small layers showed a small amount of error than the deeper layers that showed a lot of errors. To address the problem that deep layers or models always produce more error. Several researchers considered this problem by introducing deep learning based on a residual framework. In other words, they provided a connection based on shortcuts that can perform and identify mappings easily.

The layer that represents the residual mapping is defined by $\boldsymbol{H(x)}$, while the layers that represent nonlinear mappings are known as $\boldsymbol{F(x) \coloneqq H(x) - x}$ so the original mapping become $\boldsymbol{H(x) \coloneqq F(x) + x}$ [30]. This is shown clearly in Figure. 5.

The architecture of Resnet50 contains the following element or structure: A convolutional layer with 64 various kernels and each kernel is of size 7 x 7 and a stride of size 2. The next layer is a max-pooling with the same stride as the convolutional layer. Then, further convolutional layers are 64 kernels with a

size of 1 x1 followed by 64 kernels with size 3 x 3, and finally 256 kernels with size 1 x 1. These convolutional layers are repeated three times producing nine layers in this step only [31].
Then, it can be seen a further convolution is performed based on 64 kernels with a size of 1 x 1 followed by 128 kernels with size 3 x 3, and finally 512 kernels with size 1 x 1. These convolutional layers are repeated four times producing 12 layers. In addition to this, other convolutional layers are applied based on 256 kernels of size 1 x 1, 256 kernels of size 3 x 3, and 1024 kernels of size 1 x 1 forming three convolutional layers. The former convolutional layers are represented six times to obtain eighteen layers. Again, more convolutional layers are added with 512 kernels with a size of 1 x1 followed by two 512 kernels with a size 3 x 3, and finally 2048 kernels with a size 1 x 1. These convolutional layers are repeated three times producing nine layers. Finally, an average pooling and fully connected layers are added to end the architecture of the Resnet50. Figure. 6 presents the whole architecture of the Renet50 and its deep residual mapping.

3.4. Classification

This step is the final step of the proposed methodology in which the result of the diagnosis will be determined with average accuracy. Three main classifiers are determined to examine the performance of the methodology, and these classifiers are Softmax, Random forest (RF), and XGBoost classifier.

3.4.1 Softmax Classifier

Softmax is known as a multinomial logistic regression and it is well accepted in statistical mathematics as it is applied to classify a categorical class placement. Softmax gives a more intuitive classification output and probabilistic interpretation [32]. For instance, let us assume that 4 classes are presented, the Softmax classifier will have 4 main nodes generated and defined by $\boldsymbol{P_i}$, where i = 1,2,3,4. These probabilities depend on a discrete target function, and these probabilities are input to the Softmax classifier in the form of the following equation:

$$\boldsymbol{S_i = \sum_k y_k t_{ki}} \tag{2}$$

Where y is the activation produced from the nodes found in the last layers, and the $\boldsymbol{t}$ is the weight that joins the last layers of nodes to the last layers in the DL model [64].
The probability of the $\boldsymbol{S_i}$ will be defined using the following equation:

$$\boldsymbol{P_i = \frac{exp(S_i)}{\sum_j^2 exp(S_i)}} \tag{3}$$

The predicated class $\boldsymbol{i}$ will be obtained by the following equation:

$$\boldsymbol{i = argmax\ (P_i)} \tag{4}$$

3.4.2 Random Forest (RF) Classifier

Random Forests (RF) are one of the powerful ensemble learning methods. RF was developed to overcome the drawbacks of decision trees (DT). The major disadvantage of the DT is the high variance. In order words, it is not natural that a small variance in the training data can lead to a major change in the structure of a decision tree. This makes the decision trees as a classifier largely unstable in comparison to other decision predictors. Also, if an error happens in a node that is near the root, it propagates to the leaves of the tree. This leads to different and worse classification results. Therefore, the classifier of the random forest is invented by Breiman [33]. RF is built based on the combination of various decision trees. It integrates the output obtained from each separate decision tree to generate the

final result. In addition to that, RF relies on uncorrelated decision trees. In other words, if similar decision trees are used in the forest, then the overall result will not vary so much and it will be equivalent to the result of a single decision tree. To achieve the concept of uncorrelated decision trees in RF features randomness and bootstrapping are applied. Random forests work considering a learning set known by $\boldsymbol{L} = ((\boldsymbol{X_1}, \boldsymbol{Y_1}), \ldots\ldots, (\boldsymbol{X_i}, \boldsymbol{Y_i}))$ designed with $\boldsymbol{i}$ vector. Where $\boldsymbol{X}$ is a set of features and samples and the $\boldsymbol{Y}$ is the set of labels. In the classification problems, RF maps $\boldsymbol{X}$ to $\boldsymbol{Y}$ and new input features are recognized by each tree of the forest. Then, each tree produces a specific classification result and the decision forest selects the classification based on the most votes obtained over all the trees in the forest.
The training of the RF is achieved relying on the result obtained from each decision tree. The training data is distributed randomly based on drawing $\boldsymbol{N}$ examples with a special kind of replacement in which the $\boldsymbol{N}$ is considered the original size of the training data. The learning method produces a classifier obtained from various trials and then the classifiers are gathered together to form the final classifier. In the classification stage, each classifier starts to record a vote for the class to which it belongs and the feature is drawn to the class with the highest votes.

3.4.2 XGBoost Classifier

XGBoost classifier is an ensemble of trees in which various classification and regression trees are fused [34-35]. Suppose a given data set is defined with $\boldsymbol{y}$ classes and $\boldsymbol{x}$ features. The mathematical formula that represents the tree ensemble structure is known by:

$$\hat{\boldsymbol{y}}_{\boldsymbol{\iota}} = \sum_{\boldsymbol{k}=1}^{\boldsymbol{k}} \boldsymbol{h_k}(\boldsymbol{x_i}), \boldsymbol{h_k}\ \boldsymbol{h_k} \in \boldsymbol{R} \tag{5}$$

The number of trees is defined by $\boldsymbol{k}$, $\boldsymbol{h}$ is defined as a method in the functional space known as $\boldsymbol{R}$. The former space is the number of regression trees and classifiers possible. Then, the XGBoost objective equation is defined as follows:

$$\boldsymbol{F_1}(\boldsymbol{\theta}) = \sum_{i}^{\boldsymbol{n}} \boldsymbol{h}(\boldsymbol{y_i}, \hat{\boldsymbol{y}}_{\boldsymbol{\iota}}) + \sum_{\boldsymbol{k}=1}^{\boldsymbol{k}} \boldsymbol{\Omega}(\boldsymbol{h_k}) \tag{6}$$

Where $\sum_{i}^{\boldsymbol{n}} \boldsymbol{h}(\boldsymbol{y_i}, \hat{\boldsymbol{y}}_{\boldsymbol{\iota}})$ is defined as the training loss function, and also $\boldsymbol{\Omega}(\boldsymbol{h_k})$ is known as the regularization function or sometimes called regularization complexity and it is represented by:

$$\boldsymbol{\Omega}(\boldsymbol{h_k}) = \boldsymbol{\gamma L} + \frac{1}{2}\boldsymbol{\lambda} \sum_{\boldsymbol{j}=\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{w_j^2} \tag{7}$$

Where $\boldsymbol{\gamma}$ and $\boldsymbol{L}$ are known as the gamma and the number of leaves, $\boldsymbol{\lambda}$, and $\boldsymbol{w}$ are the regularization coefficient on weights and the vector of scores resulting from the leaves respectively. The XGBoost classifier aims to decrease $\boldsymbol{F_1}(\boldsymbol{\theta})$ as much as possible. Finally, $\boldsymbol{h}$ is defined as the loss function represents the variance between and the target $\boldsymbol{y_i}$ and the forecasted $\hat{\boldsymbol{y}}_{\boldsymbol{\iota}}$.

4. Experimental Results

The experimental result was reached using the proposed model based on three main classifiers which are SoftMax, RF, and XGBoost. The deep learning model was implemented using MATLAB software. An experiment is applied based on the proposed methodology for the diagnosis of four different ECG records. The whole experiment is performed on a computer with Intel (R) Core i7-8565U CPU of 1.99 GHz, 12 GB memory, and NVIDIA graphical card with GM 310M. The total number of records selected from 6 datasets for the four types of ECG heartbeats is 294 records. These records are collected as follows: 72 normal records (NSR) from the first two datasets (18 from nsrdb) and (54 from nsr2db), 74 superventricular arrhythmias (SV) records from the third dataset (svdb), 74 records representing ST-segment changes from the fourth and the fifth datasets (28 from stdb) and (46 from ltsdb), and finally 74 myocardial infractions (MYC) records from the sixth dataset. The experiment was based on dividing the whole ECG records into three different parts training, validation, and test. This division made 177 records used for training and 57 records for validation and 60 for the test. The parameters of the training are adjusted properly to achieve the highest training performance and the lowest loss error.

4.1 Training Parameters Setting

The parameters of the Renet50 model applied for the ECG diagnosis are determined in Table 2. There exist various hyper-parameters that can be set before the training process. The selected parameters are the optimizer, mini-batch size, maximum epochs, the total number of iterations, regularization factor, and the validation frequency.

Table 2 Parameters adjusted for the Resnet50 deep learning model.

Training Parameters				
Optimizer	**Mini Batch Size**	**Maximum Epochs**	**Number of iterations**	**Validation Accuracy (%)**
Stochastic gradient descent Momentum (Sgdm)	8	100	2300	91.75%
	16	100	1100	89.58%
	32	100	500	85.42%
	35	100	500	87.50%
Adaptive Moment estimation (adam)	8	100	2300	91.67%
	16	100	1100	**95.83%**
	32	100	500	93.75%
	35	100	500	93.75%
Root mean square propagation (RMSprop)	8	100	2300	91.67%
	16	100	1100	91.75%
	32	100	500	93.75%
	35	100	500	93.75%

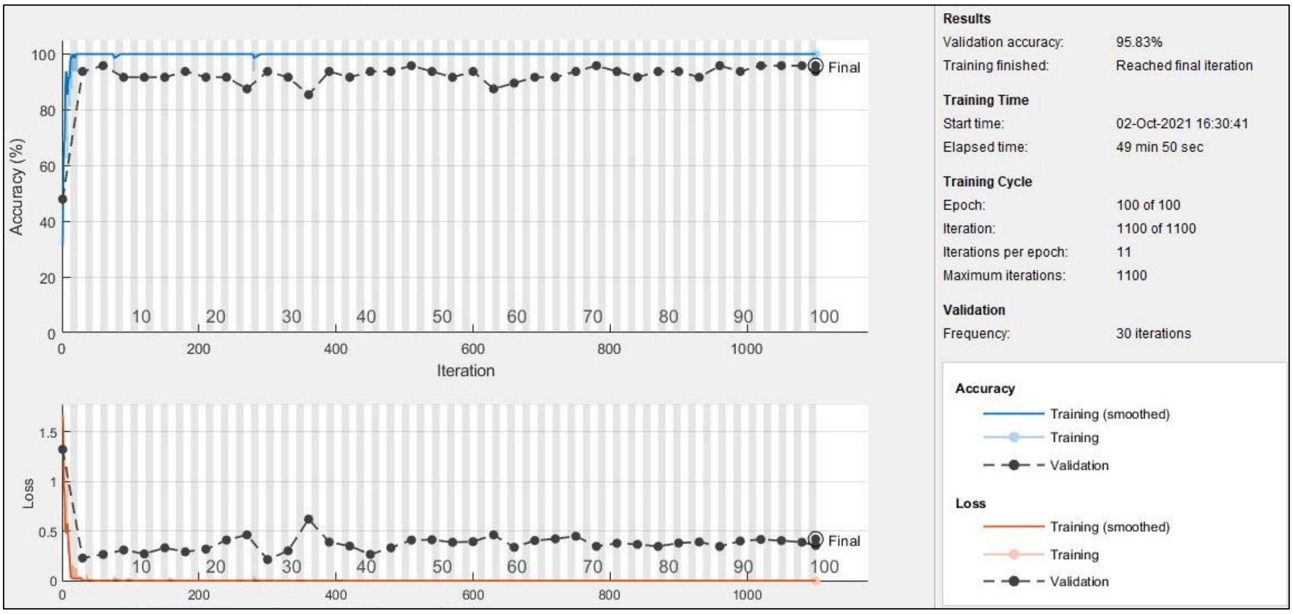


Figure. 7: Training and Loss curve performance of proposed Resnet50 model

The optimal parameters selected for the Resnet50 for the diagnosis of the ECG records are determined experimentally. The optimizers used for training the Resnet50 model are the stochastic gradient descent with momentum (SGDM), adaptive moment estimation (adam), and root mean square propagation (RMSprop). The mini-batch size parameter is applied with different values such as 8, 16, 32, and 35 on the three optimizers. The maximum epochs are 100 and the iterations vary relying on the size of the data and the mini-batch size values. A validation data is input during the training process with a validation frequency equal to 30, and an L2 regularization factor is defined with a value equal to $\mathbf{1} * \mathbf{10^{-4}}$. It can be seen that when the optimizer is set to adam, and the mini-batch size is 16 the performance of the validation data has the highest accuracy. Therefore, the test data are passed to the model with the highest validation accuracy. In the training stage, the accuracy and the loss curves are obtained for the pre-trained model which is the Resnet50. Figure. 7 shows the highest performance achieved based on the validation data. The blue curve presents the training curve, while the red curve presents the loss curve, whereas, the black dashed curve presents the validation accuracy and loss during the training phase.

4.2 Classification Parameters

It is important to determine the hyper-parameters required before training the Resnet50. It is also essential to determine the parameters used on each classifier after applying Resnet50 for feature extraction. As mentioned before, three main classifiers are used to determine the diagnosis performance of the Resnet50 features on the CWT ECG images. The first classifier is the Softmax classifier and its main parameter is the loss function which is defined by the cross-entropy. The next classifier is the random forest and it has a set of parameters such as the number of features extracted, number of trees, and the maximum depth of each tree. Finally, the last classifier applied is the XGBoost classifier and it also has several parameters. These parameters are chosen depending on the kind of classification that XGBoost will perform. In the case of multi-class classification (as in this study) the booster and the evaluation matrix must be defined by gbtree and mlogloss respectively. The rest of the XGBoost parameters are used based on their default values in the library of XGBoost. Table 3 shows the main parameters' values for the classifiers used after applying Resnet50.

Table 3 Classifiers applied in the methodology and its optimal parameters

Classifiers	**Optimal Parameters**
Softmax	Loss function : "Cross Entropy (CE)"
Random Forest	Number of trees = 100 Max depth of each tree = 0 (zero indicates unlimited) Number of features = (log2(no.of.predictors)+1)
XGBoost Classifier	Booster = 'gbtree', Evaluation matrix = 'logistic' Iteration = 10, Max Depth = 5, Eta = 0.1,Gamma = 0

4.3 Classification Results

The features obtained from the fully connected layer of the Resnet50 model are forward for the three classifiers. The classifiers start to operate on the test data to ensure the performance of the validation accuracy obtained during the training. Table 4 shows various statistical performance measurements such as true positive rate (TPR), precision, false-positive rate (FPR), recall, receiver operating characteristic (ROC), Mathew's correlation coefficient (MCC), and precision-recall characteristic (PRC) value [36]. These measurements are calculated for each classifier on the test data. In addition to this, the confusion matrix is manifested to determine the overall diagnosis performance on the classifiers. The confusion matrix is a figure or a table that is needed to describe the diagnosis performance of the tested data. It is a heat map in which the true value must be known. It gives the chance to visualize the performance of the three applied classifiers on the Resnet50 deep learning model as shown in figure 8 (a, b, and c). It can be manifested that ***the*** XG-Boost classifier has the highest accuracy performance over other classifiers.

Table 4 Classifiers performance using Resnet50 pre-trained model based on different statistical measurements

Classifiers Performance	Performance Measurements (%)								
	TPR	**FPR**	**Precision**	**Recall**	**F-Measure**	**Accuracy**	**MCC**	**ROC Value**	**PRC Value**
Softmax	93.2	2.3	93.9	93.2	93.2	93.2	91.3	98.6	96.3
RF	94.9	1.6	95.0	94.9	94.9	94.9	93.3	96.7	92.1
XGBoost	98.3	0.5	98.4	98.3	98.3	**98.3**	97.8	97.8	98.8

-

Figure 8. (a) The confusion matrix of the SoftMax results based on the Resnet50 (b) The confusion matrix of the ANN results based on the Resnet50 (c) The confusion matrix of the XGBoost results based on the Resnet50

5. Discussion

The paper proposed a methodology for the diagnosis of four main types of ECG heartbeats. The methodology consists of four main phases, and these phases are obtaining ECG data, filtering ECG signals, extracting various ECG features, and classifying ECG records. In the phase of gathering ECG data, the ECG records are obtained from six various online ECG datasets. In the phase of signal filtration, the records are filtered using wavelets and a set of filters based on band stop, low pass, and smoothing filter to eliminate the main common noises in the ECG signals. In the phase of extracting features, the most discernment feature was obtained in the time-frequency domain from the CWT and then forward to one of the transfer learning models known as Resnet50. Finally, the classification is applied based on three main classifiers and these classifiers are Softmax, RF, and XGBoost classifier. To overcome the chances of overfitting in the proposed model, a regularization factor is defined to shrink the learned estimates to zero. In other words, this regularization can tune the loss function by providing a penalty term to the optimizer of the Resnet50 model, and this will encourage smaller weights avoiding excessive changes of the coefficient. Also, the regularization applied to optimize the objective function in the XG-Boost classifier allows it to be trained against overfitting. In addition to this, the number of ECG records in each of the four ECG classes are nearly equal leading to a balanced number of records in each category, dropping the probability of overfitting. Moreover, the ECG records are filtered using a pre-processing chain for reducing common noises that can cause overfitting during the training. Finally, the training accuracy obtained from the model with the highest accuracy validation is 99.5% and the value of the highest test accuracy is 98.3%, the slight difference between the training and the test accuracies shows that the model appropriately fits.

For comparison with others, several algorithms applied different methodologies for ECG diagnosis as shown in Table 5. S. N. Yu et al. [37], the authors approached an accuracy of 96.38% with the bispectrum feature set and SVM as the classifier, and when the authors added the genetic feature selector to the bispectrum and the SVM was used for classification, the accuracy increased to 98.10%. The number of records used was 54 and 29 from each of the normal sinus rhythm (NSR) and cognitive heart failure (CHF) data sets, respectively. K. H. Boon et al. [8] applied a diagnosis method to differentiate between normal and abnormal based on PAF. The features were produced from 106 ECG data collected from 53 ECG recordings. The SVM classifies based on 5 mins heart rate variability (HRV) segment and its distance from the PAF event. If it is at least 45 min distant from the event, the recording is called normal, but if the HVR segment goes before the event the recording is called abnormal. The accuracy achieved was 87.7%. Based on the improvement in the deep learning models in the diagnosis of the ECG heartbeats. H. B. Bae et al. [38] tried to classify normal NSR and abnormal ECG records such as AF, and ventricular fibrillation (VF) and they also focused on balancing the number of records used. The classification was based on Gamma distribution using probability output networks (CPON), and it proved that the performance was higher than KNN, SVM, aiming at an accuracy of 97.33%. Janghel et al. [13] aimed at building automated classification of regular and irregular ECG heartbeats. They applied their system on 47 records and the best results were achieved by using the decision tree, obtaining an accuracy of 88.2%.

Table 5 Proposed DL model compared to other previous work for ECG diagnosis

Authors	Records	Methodology	Classes	Databases	Performance
S. N. Yu et al. [37] 2012	54 R from NSR + 29 R from CHF	Features: Bisecpectrum + genetic feature set Classifier: SVM-	2	MIT-BIH NSR and CHF	Bisecptrum + SVM = 96.38% Bispecturm + genetic feature set + SVM = 98.10%
K. H. Boon et al. [8] 2018	106 data from 53 R pairs	Features: Time domain, spectral, Bispectrum, nonlinear dynamics features Classifiers: SVM	2	Atrial Fibrillation prediction (AFPDB) Database	ACC = 87.7%
H.B. Bae et al. [38] 2019	NSR: 15 R VF: 15 R AF: 15 R	R–R interval + (CPON)	3	MIT-BIH (NSRDB), (VFDB), (AFDB)	ACC = 97.33%
R. R. Janghel et al. [13] 2020	47 R 40% of the 47 R records are patients	Naïve Bayes SVM Ada-boost RF, Decision Tree, and KNN	2	MIT-BIH arrhythmia database	ACC of the Decision Tree = 88.2%
The proposed Method	**294 R 177 R for train 117 R for validation and test**	**CWT + Resnet50 + Softmax, RF, and XGBoost Classifier**	**4**	**6 main datasets**	**Softmax ACC =93.2% RF ACC = 94.9% XGBoost Classifier ACC=98.2%**

The proposed methodology worked on 294 recordings obtained from 4 different ECG heartbeats. The features are obtained from CWT and forward to the Resnet50 pre-trained model. Three main classifiers were applied to reach 93.2% using Softmax, 94.9% using RF, and 98.2% using the XGBoost classifier. The advantages of the proposed model are illustrated in three main points. The first point is the removal of the three common noises related to the ECG signals using a well-defined pre-processing chain. The second point is obtaining robust features from the combination of CWT and Resnet50 transfer learning model. The last point is the superiority of the XB-boost in the classification because it is highly flexible, can be paralleled, supports generalization, and is faster than gradient boosting.

6. Conclusions and Future Work

In this study, a methodology is presented for the diagnosis of the four different types of ECG heartbeats based on continuous wavelet transform combined with a transfer learning model known as Resnet50. The proposed methodology produces better results makes it adaptable for the diagnosis of different ECG records. The data were collected from 6 public available datasets. The ECG records were filtered to drifting in the ECG signals, powerline interference, and the high noise frequencies. The filtering chain is based on wavelets and a set of filters. Then, the ECG records are converted from 1D to 2D images using CWT and the coefficients of the CWT are used to compute the scalogram for each ECG signal. The scalogram is in form of a 2D image representing the main features in the time-frequency domain. The 2D image is passed to the Resnet50 for further features extraction. Finally, the classification is based on Softmax, RF, and XGBoost classifiers achieving an accuracy of 93.2%, 94.9%, and 98.2% respectively. ECG signals have future directions that can contribute and provide assistance in the field of medical informatics. There is a need for a real-time diagnosis application that can verify various types of heart diseases. In addition to this, it was discovered recently that the ECG signals can diagnose COVID patients based on the ECG image reports. It is recommended to develop diagnosis systems that can identify COVID patients from normal and various abnormal heartbeats. It is also suggested to use stratified k-fold cross-validation in future experiments to provide more information about the methodology performance. It is also advised to select the hyper-parameters based on various methods such as grid or random search or various metaheuristic techniques to reach the optimal values on the parameters for the proposed model.

Various steps can be provided to improve the performance of the proposed model. Instead of computing the scalogram using CWT, the spectrogram can be obtained representing the spectrum of the signal and how it is changed through time. The Resnet50 applied in the proposed method can be replaced with other pre-trained models such as Vgg16, Vgg19, Xception, and others. Finally, the XB-Boost classifier can be replaced with a sparse representation classifier as it is considered a powerful technique for pixel-wise classification of images.

References

1. World Health Organization, Cardiovascular Diseases (CVDs). http://www.who.int/mediacentre/factsheets/fs317/en/, 2022.
2. S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, M. B. Gulmezoglu, A survey on ECG analysis, Biomed Signal Process Control, 43 (2018) 216-235.
3. O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, U. R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, Biomed Comput Meth Prog Bio 161 (2018) 1-13.
4. V. K. Sudarshan, U. R. Acharya, S. L. Oh, M. Adam, J. H. Tan, C. K. Chua, R. San Tan, Automated diagnosis of congestive heart failure using dual tree complex wavelet transform and statistical features extracted from 2 s of ECG signals, Comput. Biol. Med. 83 (2017) 48-58.
5. R. Mahajan, T. Viangteeravat, O. Akbilgic, Improved detection of congestive heart failure via probabilistic symbolic pattern recognition and heart rate variability metrics, Int. J. Med. Inform. 108 (2017) 55-63.
6. A. D. Dolatabadi, S. E. Z. Khadem, B. M. Asl, Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM, Comput. Biol. Med. 138 (2017) 117-126.
7. N. Henzel, J. Wrbel, K. Horoba, Atrial fibrillation episodes detection based 340 on classification of heart rate derived features, InMIXDES-24th Inter-National Conference Mixed Design of Integrated Circuits and Systems, 2017, pp. 571-576.
8. K. H. Boon, M. Khalil-Hani, M. B. Malarvili, Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm, Comput. Biol. Med. 153 (2018) 171-184.
9. J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam and U. R. Acharya, Application of stacked convolutional and long short-term memory network for accurate identification of CAD ECG signals, Comput. Biol. Med. 94 (2018) 19-26,
10. M. Kumar, R.B. Pachori, U.R. Acharya, Automated diagnosis of atrial fibrillation ECG signals using entropy features extracted from flexible analytic wavelet transform, Biocybern Biomed. Eng. 38 (2018) 564-573

11. B. Safarbali, S.M.R.H. Golpayegani, Nonlinear dynamic approaches to identify atrial fibrillation progression based on topological methods, Biomed. Signal Process. 53 (2019) 101563,
12. L. Yang, J. Zhu, T. Yan, Z. Wang, S. Wu, A Modified Convolutional Neural Network for ECG Beat Classification, J Med Imag. Health In. 10 (2020) 654-660.
13. R. R. Janghel, S. K. Pandey, A Classification of ECG Arrhythmia Analysis Based on Performance Factors Using Machine Learning Approach. In Computational Network Application Tools for Performance Management Springer, Singapore, 2020, pp. 65-74.
14. Y. Li, Y. Pang, K. Wang, X. Li, Toward Improving ECG Biometric Identification Using Cascaded Convolutional Neural Networks, Neurocomputing. (2020).
15. C. Chen, Z. Hua, R. Zhang, G. Liu, W. Wen, Automated arrhythmia classification based on a combination network of CNN and LSTM, Biomed. Signal Proces. 57 (2020) 101819.
16. E. H. Houssein, I. E. Ibrahim, N. Neggaz, M. Hassaballah, Y. M.Wazery An efficient ECG arrhythmia classification method based on Manta ray foraging optimization. Expert Syst. Appl. 181 (2021) 115131.
17. F. M. Dias, H. L. Monteiro, T. W. Cabral, R. Naji, M. Kuehni, E. J. D. S. Luz,. Arrhythmia classification from single-lead ECG signals using the inter-patient paradigm. Comput. Methods Programs Biomed. 202 (2021) 105948.
18. MIT-BIH Normal Sinus Rhythm: https://www.physionet.org/content/nsrdb/1.0.0/ , 1999 (accessed 2 October 2021).
19. Normal Sinus Rhythm RR Interval: https://physionet.org/content/nsr2db/1.0.0/ , 2003 (accessed 2 October 2021).
20. MIT-BIH Supraventricular Arrhythmia https://physionet.org/content/svdb/1.0.0/, 1999 (accessed 2 October 2021).
21. MIT-BIH ST Change Database: https://physionet.org/content/stdb/1.0.0/ , 1999 (accessed 2 October 2021).
22. Long Term ST Database: https://physionet.org/content/ltstdb/1.0.0/ ,2000 (accessed 2 October 2021)
23. PTB Diagnostic ECG Database: https://www.physionet.org/content/ptbdb/1.0.0/ , 2004 (accessed 2 October 2021).
24. M. M. Bassiouni, E. S. A. El-Dahshan, W. Khalefa, A. M. Salem, Intelligent hybrid approaches for human ECG signals identification, Signal Image Video Process, 12(5) (2018) 941-949.
25. H. Khorrami, M. Moavenian, A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. Expert Syst. Appl. 37(8) (2010) 5751-5757.
26. F. Bosché, B. Biotteau, Terrestrial laser scanning and continuous wavelet transform for controlling surface flatness in construction–A first investigation. Adv. Eng. Inform. 29(3) (2015) 591-601.
27. E. Maghawry, R. Ismail, T. Gharib, Unsupervised Real-time diagnosis system for ECG streaming DAT. International Journal of Intelligent Computing and Information Sciences, 21(2) (2021) 1-16.
28. D. McNeely-White, J. R. Beveridge, B. A. Draper, Inception and resnet features are (almost) equivalent. Cogn. Syst. Res. 59 (2020) 312-318.
29. S. Dong, P. Wang, K. Abbas, A survey on deep learning and its applications. Comput. Sci. Rev., 40 (2021) 100379.
30. Z. Lu, X. Jiang, A. Kot, Deep coupled resnet for low-resolution face recognition. IEEE Signal Process. Lett. *25*(4) (2018) 526-530.
31. J. Peng, S. Kang, Z. Ning, H. Deng, J. Shen, Y. Xu, L. Liu, Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging, Eur. Radiol. 30(1) (2020) 413-424.
32. A. F. Agarap, Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, (2018).

33. A. T. Azar, H. I. Elshazly, A. E. Hassanien, A. M. Elkorany, A random forest classifier for lymph diseases. Comput. Biol. Med., 113(2) (2014). 465-473.
34. S. Y. El-Bakry, E. S. El-Dahshan, M. Y. El-Bakry, Total cross section prediction of the collisions of positrons and electrons with alkali atoms using Gradient Tree Boosting, Indian J. Phys, *85*(9) (2011) 1405-1415.
35. H. T. Weldegebriel, H. Liu, A. U. Haq, E. Bugingo, D. Zhang, A new hybrid convolutional neural network and eXtreme gradient boosting classifier for recognizing handwritten Ethiopian characters. *IEEE Access,* 8 (2019) 17804-17818.
36. S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. Soft Comput. 13(10) (2009) 959.
37. S. N. Yu, M. Y. Lee, Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability, Comput. Biol. Med., 42(8) (2012) 816-825.
38. H. B. Bae, M. S. Park, R. M. Kil, H. Y. Youn, Classifying heart conditions based on class probability output networks. Neurocomputing, 360 (2019) 198-208.