Meta Structural Learning Algorithm with Interpretable Convolutional Neural Networks for Arrhythmia Detection of Multi-Session ECG

Maytham N. Meqdad¹, Fardin Abdali-Mohammadi¹ and Seifedine Kadry², (Senior Member, IEEE),

¹Department of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran
²Department of Applied Data Science, Noroff University College, Kristiansand, Norway

Corresponding author: Fardin Abdali-Mohammadi (e-mail: fardin.abdali@razi.ac.ir).

ABSTRACT Detection of arrhythmia of electrocardiogram (ECG) signals recorded within several sessions for each person is a challenging issue, which has not been properly investigated in the past. This arrhythmia detection is challenging since a classification model that is constructed and tested using ECG signals maintains generalization when dealing with unseen samples. This article has proposed a new interpretable meta structural learning algorithm for this challenging detection. Therefore, a compound loss function was suggested including the structural feature extraction fault and space label fault with GUMBEL-SOFTMAX distribution in the convolutional neural network (CNN) models. The collaboration between models was carried out to create learning to learn features in models by transferring the knowledge among them when confronted by unseen samples. One of the deficiencies of a meta-learning algorithm is the non-interpretability of its models. Therefore, to create an interpretability feature for CNN models, they are encoded as the evolutionary trees of the genetic programming (GP) algorithms in this article. These trees learn the process of extracting deep structural features in the course of the evolution in the GP algorithm. The experimental results suggested that the proposed detection model enjoys an accuracy of 98% regarding the classification of 7 types of arrhythmia in the samples of the Chapman ECG dataset recorded from 10646 patients in different sessions. Finally, the comparisons demonstrated the competitive performance of the proposed model concerning the other models based on the big deep models.

INDEX TERMS Arrhythmia Detection, Convolutional Neural Network, Deep Features Representation, Meta Learning, Multi-Session ECG

I. INTRODUCTION

Computer-aided detection (CADe), also known as Computer-aided diagnosis (CADx), are systems that help physicians to interpret medical images and signals. ECG recording techniques, X-ray, MRI, and ultrasound systems create a lot of information and a professional radiologist or physician has to analyze and assess them in a quite short period of time [1]. CAD systems process the digital signals and images and highlight the suspicious parts such as possible diseases to facilitate the physician’s decision-making process. These assistants can help physicians to reduce the human error caused due to fatigue [2]. One of these diagnosis systems, which is quite developed is the system for automatic diagnosis of cardiac diseases using ECG signals. The electrical signals generated by various activities of upper and lower heart muscles are measured with electrocardiograms (ECGs). However, sometimes the performance of each muscle is affected by external factors such as excessive blood fat, which can cause arrhythmia in the electrical performance of the heart. For instance, a heart problem on the walls of the ventricles can be considered a cause of arrhythmia [3]. Moreover, the supraventricular on the upper walls of the ventricles in a region called atria is another complication leading to arrhythmia in the electrical signals of the heart. According to the statistics, heart complications account for one-third of the mortality rate in the adults aged 35–90 years old [4]. Hence, researchers have used artificial computer systems to propose efficient methods for the early diagnosis and classification of arrhythmia through ECG signals. Deep learning models...
such as convolutional neural networks (CNNs) [5, 6], AlexNet [7], VGGNet [8], and GoogleNet [9] have proven greatly efficient in recent years. The relevant analyses have achieved high rates of accuracy in the diagnosis of arrhythmia through ECG signals [8, 10, 11].

From a medical standpoint, every lead observes the heart from its own perspective in multi-lead ECG signals. These signals are categorized as two general classes, the first of which includes the chest signals that observe the heart from various angles of the axial plane. The second class contains the body signals that view the heart from various angles of the coronal plane. Each of these signals highlights a part of the general reality of the heart [12]. Although the arrhythmia of a signal is stable during each ECG signal recording session, the range of this arrhythmia might experience some changes. In general, deep learning methods proposed for the diagnosis of arrhythmia handle ECG signals recorded in a single session. The outstanding characteristic of deep learning models developed through a single-session ECG signal is their high accuracy. For instance, Zheng et al. [13] proposed a CNN model-based method yielding the accuracy of 97% regarding the classification of seven types of arrhythmia in the sample dataset of 10646 patients and 35565 ECG signals, all of which were recorded in a single session. Big-ECG [14] is one of the well-known methods. This method first, diagnosed 45 patients with QRS Complex, and then, classified the signals using the Random Trees model. It obtained an accuracy of 95.6%. In addition, Faust et al. [15] employed a fine-tuned method along with transfer learning in an ImageNet model. They achieved 97% of accuracy in the classification of seven types of arrhythmia in the samples of a dataset including 10646 patients and 35565 ECG signals, all of which were recorded in a single session. Since these models were developed through single-session ECG signals, their main deficiency is their low generalizability while dealing with unseen samples.

As discussed previously, generalization is essential, for each patient’s range of arrhythmia might undergo some changes in different sessions. Moreover, Da Silva et al. [16] proposed a dataset of ECG signals recorded in different sessions called CYBHi. They indicated that the only accurate method for the development and evaluation of each model of ECG signal detection could be achieved when the model was developed and tested through the signals recorded in separate sessions. Hence, the advantage of developing an arrhythmia detection system from the multi-session ECG signals is the comprehensive performance analysis of the arrhythmia diagnosis method through the ECG signals in the disjoint acquisition session. As opposed to the previous techniques, this method will be tested on different states of humans within several days of experimentation. Therefore, the proposed method proved to be generalizable and succeeded in preventing the learning model overfitting while dealing with new and unseen samples. This is regarded as an extension of the permanence analysis. However, due to generalization defects, there is a massive literature gap in the use of deep learning models for the signals recorded in different sessions.

Meta-learning is defined as an emerging method for developing the ability to deal with new conditions in machine learning models [17]. Meta-learning emphasizes the preparation of guidance for machine learning models so that they can make the best decision in new conditions including unseen samples. Moreover, meta-learning seeks a way towards learning to learn. Thus, it plays a key role in resolving the problem of arrhythmia classification in multi-session ECG signals [18]. In recent years, numerous methods have been proposed to develop a model for the demonstration of the same behavior while handling the signals recorded in various sessions, numerous algorithms are introduced for meta-learning including reinforcement learning, transfer learning, and active learning. The following two subjects should be examined to employ meta-learning for multi-session signals: 1. ECG signals are structural signals consisting of QRS Complexes. Therefore, the ordinary meta-learning algorithms are not appropriate for this signal, and an approach that is a combination of meta-learning and structural learning should be used. 2. Interpretability in meta-learning is a crucial principle since every model that seeks to obtain a higher level of generalization must be able to interpret it easily, plus it should not be similar to common neural networks of the black box [19, 20].

The challenge addressed in this article is the arrhythmia classification of multi-session ECG signals as a new problem in this field. Therefore, this article proposed a structural meta-learning method that can be described in CNN to resolve the problems of arrhythmia classification in multi-session ECG signals. The following are the three main keywords in the proposed method: 1. Meta-learning, 2. Structural Learning, 3. The Describablity. The solutions to present each of these parts will be explained in detail in this article. First, a compound loss function is designed in CNN, which is capable of implementing the learning to learn feature in the meta-learning. In this loss function, guidance is provided for the central model to enable it to self-educate. Second, to add describability feature to CNN, this model was encoded by evolutionary trees in the genetic programming. These trees are CNN models that are learned in the course of the evolution process. Third, in order to add the structural learning feature to the proposed meta-learning algorithm of trees in GP, the CNN trees carry out the deep feature learning process using the morphological structural operators. The experimental results section of the number of parameters and computational complexity of the proposed model for classification of the samples of the Chapman ECG dataset including 10646 patients are evaluated [13]. The experimental results suggested that the
student model achieved an average of 97% accuracy through Lead III input signal for the classification of 7 types of arrhythmia in the ChapmanECG dataset. The main innovations of this article are as follows:

1. Providing a new method for arrhythmia classification of multi-session ECG signals for each person (i.e., training and testing the model in two separate sessions)
2. Introducing a new describable meta-learning algorithm via encoding the CNN model in the evolutionary trees of the GP algorithm
3. Adding the structural learning feature to the proposed meta-learning algorithm via encoding the CNN model in the evolutionary trees of the GP algorithm

Then, section II of the article reviews the studies on arrhythmia classification of ECG signals. Section III of the article, describes the proposed structural meta-learning algorithm. Section IV comprises the experimental results. Section V provides a conclusion for the article.

II. RELATED LITERATURE

The ECG signals include various arrhythmia, among which 11 types were mentioned in the previous section. The methods are divided into different classes with respect to the number of arrhythmias classified in various investigations. Besides, it should be noted that all investigations introduced for diagnosing arrhythmia were assumed to be carried out based on single-session ECG signals. Among all methods introduced up to now, deep learning is one of the top algorithms in the field of the classification of ECG signals. Employing more layers in deep learning enables us to have a deeper model and when the ECG signals pass through these layers, more distinguishing features can be extracted to facilitate the detection of the pattern of arrhythmia in them. The total process of extracting the features in the deep learning models is carried out automatically and without the need for hand-craft feature extraction. In recent years, a broad spectrum of ECG dataset-based deep learning methods are introduced such as the MIT-BIH dataset and several other private datasets. The methods introduced [21], [22], and [23] are CNN-based deep learning models, which have classified various arrhythmia in ECG signals. The majority of these methods use 1D filters in CNN since the ECG signal in the MITBIH dataset is 1D as well. Here several CNN-based methods are explained. In [24], a CNN model with 8 convolution layers, 4 pooling layers, and one fully connected layer is used, which obtained the accuracy of 93.1% for 22 patients and 109449 accurate pulses. In addition, in [21], a CNN model enjoying 8 convolution layers bearing 1D filters, 4 pooling layers, and one fully connected layer is proposed for the classification of 5 types of arrhythmia (non-ectopic, supraventricular ectopic, ventricular ectopic, fusion, and unknown beats). This method obtained an accuracy of 94%. In [25], a CNN model comprising 8 convolution layers bearing 1D filters, 4 pooling layers, and two fully connected layers is used for the classification of 12 types of arrhythmia. In this article, the proposed model was examined on a dataset, which is not publicly accessible including 91232 ECG records from 23500 male and 2004 female patients. The results suggested that this method calculated accuracy of 98.5%. In [26], a deep learning method enjoying the U-Net architectural standard, which comprised thirty-two 1D residual layers, was introduced to classify 5 types of arrhythmia. The results suggested that this method calculated accuracy of 98.5%. In [22], a CNN model comprising 12 convolution layers was used for the classification of 44 records. In [27], CNN model consisting of 8 convolution layers bearing 1D filters, 4 pooling layers, and two fully connected layers enjoying the End-to-End feature were suggested for the classification of 17 and 15 arrhythmias in ECG signals, respectively. Per the results, this method calculated the accuracy of 98.5% and 94%, respectively.

Besides these methods, some models evaluated their methods on 12-lead ECG signals. In [28], an LSTM model was used on 38,899 patients and for 12 different arrhythmias. The results manifested 90% accuracy. Moreover, in , a CNN model was used on an unknown number of patients and for 8 different arrhythmias. The results manifested an F1-Score amounting to 81%. Despite the advantages of the aforesaid models such as accuracy of higher than 90%, as well as the End-to-End feature, the dataset used in these models including MIT-BIH were collected about 40 years ago and has problems such as the imbalance class and a small number of patients. Therefore, the results of these studies cannot be approved for practical use. Zheng et al. [13] collected a dataset called ChapmanECG from 12-Lead ECG signals from more than 10,000 people. The interesting point about this dataset is that the signals are recorded from the patients within several different days and during different sessions. A primary evaluation based on the gradient boosting tree model was conducted on the classification of this dataset. This model obtained the accuracy of nearly 97% for each class of arrhythmia, separately. Other than the evaluations conducted in [24], a CNN model comprising three 1D convolutions was proposed for the classification of 12-lead signals of the Chapman ECG dataset. It yielded an average AUC of 79.60 (AUC=79.60%). Faust et al. [29] presented a new and more integrated method on this dataset. In this article, a transfer learning-based fine-tuning method was implemented in the Image-Net deep learning model. Following the evaluations of this article, an accuracy of 92.24% was obtained for the classification of 7 types of arrhythmia and 96.13% accuracy was obtained for the classification of 4 types of arrhythmia. The great deficiency of this method is using the folding method for dividing the data between training and testing. This method results in using repetitive data from a special session in training and testing.
I. PROPOSED METHOD

The ECG dataset used in this article was collected by Chapman University and Shaoxing People’s Hospital (Chapman ECG in short) [13]. Table I shows the numerical details of the dataset. The ECG signals for each person were recorded within several days and during different sessions. This enables us to assess the proposed method of this article on multi-session ECG signals. In this dataset, the 12-lead ECG signals were recorded from 10646 people with a frequency higher than 500Hz. Each ECG signal in 12-lead is a 10-second strip. In addition, an initial pre-processing was applied to this dataset to smooth the ECG signals using the Butterworth filter and the Non-Local Means technique.

A. Meta Structural Learning Model

Figure 1 shows a general schematic of the meta-learning-based model proposed in this article. The input of this model is based on a dataset in $D = \{x_i, y\}_{i=1}^{N}$, in which $x_i$ signifies one of the $N$ ECG signals in the Chapman dataset. In addition, $y$ indicates the no arrhythmia label, which includes one of SB, SR, AFIB, ST, A, SI, SVT, AT, AVNRT, AVRT, SAAWR. This model includes two phases for one task, i.e., meta training and meta testing. In the meta training, the dataset of training $D_{meta-train} = \{x_i, y\}_{i=1}^{N_{train}}$ is used for training the classifications, in which $N_{train}$ shows the number of training samples.

The meta testing aims at detecting the labels of a query sample $D_{query} = \{x\}_{j=1}^{N_{query}}$, in which $N_{query}$ indicates the number of queries. Taking into account that ECG signals in the Chapman dataset were recorded during different sessions (i.e., $session_{x_i} \neq session_{x_j}$, $\forall x_i, x_j \in D_{meta-train}$), the meta-learning model must include the following two characteristics: 1. Non-linear mapping in the artificial neural networks must enjoy the generalization feature for unseen samples in various sessions, 2. Mapping must preserve the relationship between the classes of the unseen samples in $D_{query}$. Thus, this article seeks to transfer the knowledge related to ECG signals in various sessions between classification models. In light of that, per Fig. 1, a two-phase communication network is proposed for this problem. First, the model for extraction of transferable features is meta learned on the training samples using the CNN trees. Then, the features extracted from the query model will be extracted based on the same model, and they will be added to the model to calculate their distance. Then, a relationship between training samples and the query will be acquired using the labels space.

In general, the loss related to the model that is manifested in the Figure will be formalized using an equation.

$$L_{meta-learning} = \log(\alpha \times \sum_{n}^{N_{model}} J_{Feat-Sim}^{n} + (1 - \alpha) \times \sum_{n}^{N_{model}} J_{BL-Sim}^{n})$$

(1)

In this equation $J_{Feat-Sim}^{n}$ is the distance between feature maps extracted by CNN tree for two input signals of $x_i$ and $x_j$, which is calculated as equation (2). In this equation, $f_n(x_i)$ shows the interpretable representation learning section for the $x_i$ input. It will be elaborated on in section 3.3.

$$J_{Feat-Sim}^{n} = \sum_{(i,j)} \frac{1}{|x_i + x_j|} e^{-\beta(\cos(f_n(x_i), f_n(x_j)))-\gamma}$$

(2)
In addition, \( \gamma_{\text{LBL-Sim}} \) is the distance between the estimated labels by the \( n \) model for two \( x_i \) and \( x_j \) input signals that will be calculated through Gumbel-Softmax distribution.

\[
\gamma_{\text{LBL-Sim}}(y_1, \ldots, y_M) = \Gamma(N_{\text{class}})\tau^{N_{\text{class}}-1} \left( \sum_{i=1}^{N_{\text{class}}} \pi_i \right)^{-N_{\text{class}}-1} \prod_{i=1}^{N_{\text{class}}} \left( \frac{\pi_i}{y_i^{T+1}} \right)
\]

(3)

In this equation, \( y_i \) are the soft labels that are used in the Gumbel-Softmax distribution. These labels are calculated using the following equation.

\[
y_i = e^{\frac{\log \pi_i}{\tau}} \sum_{t=1}^{N_{\text{class}}} e^{\frac{\log \pi_t}{\tau}}
\]

(4)

In this equation, \( \pi_i \) signifies the calculated probability by \( l \) model, plus \( y_i \) is the soft label of this model. By the hard label coded as one-hot, merely the corresponding information of classes can be transferred among the models, and it will be impossible to train other information such as the relationship between various arrhythmia classes. Therefore, in equation 4, a soft label and a hard label are used together for training models. In equation 4, the \( \tau \) parameter indicates the extent of the softness of the labels. In the normal classification applications, \( T \) is regarded as equal to 1. In the Gumbel-Softmax distribution, a higher amount is considered for \( \tau \) to obtain a softer probable distribution in the classes that include more information about One-Hot coding. Upon the increase of \( \tau \), the \( y_i \) inclines towards a uniform distribution [30, 31].

**B. Interpretable Structural Learning \( f_n, \forall n \in \{1, \ldots, N_{\text{model}}\} \)**

This article used the middle layers of CNN models to extract the automatic features. As stated in equation (2), \( f_n, \forall n \in \{1, \ldots, N_{\text{model}}\} \) in model \( n \) is a deep network that uses the convolution layers to extract the deep features from \( x_i \) input signal. Figure 2 shows the architecture used for the CNN model as the \( n^\text{th} \) example. As mentioned in [20], the deep learning methods are the black box models, and what happens inside them is not clear. Thus, every meta-learning method designed based on these models will be uninterpretable. It reduces the possibility of using them for different sciences, as well as for non-professional use. Therefore, this article used a new CNN network that is called evolutionary deep learning. In this approach, the CNN networks are coded by genes in the GP algorithm. In the GP algorithm, genes resemble trees with the ability to make plans, plus they are interpretable like the decision tree and can be written as simple mathematical equations [32].

Figure 3 signifies the general diagram concerning the general stages of constructing the CNN trees as the GP algorithm. At first, the CNN models were encoded as GP trees. At the beginning of training, \( M \) multi-gene samples are constructed to generate the population of the first generation in the GP algorithm. The collection of the initial population can be displayed as a collection of \( \mathcal{P} = \{p_1, \ldots, p_N\} \). In this collection, \( p^m \) is the \( m \) member of this population that is generated as a multi-gene sample.

Moreover, \( p^m \) is a set of mapping functions that are coded as genes and can be expanded as \( p^m = \{f^m_1, \ldots, f^m_n, \ldots, f^m_{N_{\text{model}}}, \forall m \in \{1, \ldots, M\}, \forall n \in \{1, \ldots, N_{\text{model}}\}\}. \)

In coding \(^1\) of One-Hot the amount of real class is considered to be 1 and the rest of the classes are zero.

---

In this equation, \( \pi_i \) signifies the calculated probability by \( l \) model, merely the corresponding information of classes can be transferred among the models, and it will be impossible to train other information such as the relationship between various arrhythmia classes. Therefore, in equation 4, a soft label and a hard label are used together for training models. In equation 4, the \( \tau \) parameter indicates the extent of the softness of the labels. In the normal classification applications, \( T \) is regarded as equal to 1. In the Gumbel-Softmax distribution, a higher amount is considered for \( \tau \) to obtain a softer probable distribution in the classes that include more information about One-Hot coding. Upon the increase of \( \tau \), the \( y_i \) inclines towards a uniform distribution [30, 31].

---

\(^1\) In coding One-Hot the amount of real class is considered to be 1 and the rest of the classes are zero.
FIGURE 2. The CNN architecture used in the proposed method (comprising thirteen convolution layers, and four pooling layers bearing 3x3 filters) to extract deep structural features from ECG signals.

Here, $N_{\text{model}}$ signifies the number of mapping functions or genes that equal the number of classifying models in Fig. 1. Besides, $f_n^m$ is a $d$ dimensional function that is displayed as CNN trees. The process of representing an evolutionary CNN tree as $f_n^m$ genes for the extraction of the structural features will be elaborated in section III.C. In the course of the GP algorithm in each generation, these trees are used to transfer the input to the new space in the form of representation learning. The process of changing the numerical space can be represented as $Q^{(m)} = p^{(m)}(x_i), \forall x_i \in D_{\text{meta-train}}, \forall m \in \{1, \ldots, M\}$, in which $x_i$ is the input ECG signal in $D_{\text{meta-train}}$ set. Furthermore, $Q^{(m)} = \{\mathcal{V}_n^{(m)}\}, \forall n \in \{1, \ldots, N_{\text{model}}\}$ indicates the structural features in the new numerical space. $\mathcal{V}_n^{(m)}$ can be expanded with the following equation.

$$\mathcal{V}_n^{(m)} = p^{(m)}(x_i) = [f_1^m(x_i), \ldots, f_n^m(x_i), \ldots, f^m_{N_{\text{model}}}(x_i)]^T \in \mathcal{R}^{N_{\text{model}}} \quad (5)$$

In which $T$ signifies the transpose operator and $N_{\text{model}}$ is the number of classifying models. When all samples are executed in the current generation of the GP algorithm, a set of deep features will be calculated in the new numerical space using the mapping functions. In the GP algorithm, the accuracy of the deep features obtained in the classification of ECG signals, as well as the complication of its structure is used as fitting functions [33]. To implement this compound fitting function in this article, the Pessimistic Error Estimate (PLE) is used for each gene in the GP algorithm like the following equation.

$$\text{Fitness}(p^m) = \frac{\sum_{n=1}^{N_{\text{model}}} [e(f_n^m) + \Omega(f_n^m)]}{\sum_{n=1}^{N_{\text{model}}} T(f_n^m)}$$

$$= \frac{e(p^m) + \Omega(p^m)}{L} \quad (6)$$

In this equation, the $T(f_n^m)$ equation calculated the number of the tree nodes for each $f_n^m$ gene in the $p^m$ sample. Besides, $e(f_n^m)$ is the classification fault of this gene for the classification of input signals in $D_{\text{meta-train}}$.

In addition to this equation, the $\Omega$ function shows the complexity of gene calculation and $L$ is the set of all nodes of all genes in $p^m$. Taking into account that this article examines the classification of various arrhythmia of ECG signals, the classification error of signals in the $D_{\text{meta-train}}$ dataset is a great parameter to evaluate the CNN trees. Moreover, per Occam’s Razor’s theory, to reduce the total computational overload of the GP algorithm, the $\Omega$ function is used in the fitting function of each tree.
Following equation (6), the fitting of each population is displayed as a set of $\mathcal{F} = \{F_1, \ldots, F_m, \ldots, F_n\}$, in which $F(m) = \text{Fitness}(p^m)$. At the end of the GP algorithm, the best population sample can be selected as follows.

$$p_{\text{best}} = p^{(\arg \max_m F(m))}, \text{ s.t. } F(m) \in \mathcal{F}$$

When the performance of $p_{\text{best}}$ meets the determined amount, or when the number of generations is finished, the stages of GP algorithm execution are completed.

Otherwise, the population of the current population are transferred to the subsequent generation via the following three stages:

1. The best samples of the current population are selected based on their fitting function and the rest are eliminated, which is called the natural selection stage.

2. Afterward, the crossover and mutation genetic operators are applied to the remaining samples to construct a more mature set of samples.

3. The mature samples created in the previous stage construct a new generation together with a set of randomly constructed samples to observe the population size in all generations.

C. CNN Tree Structure

It was stated in the previous section that $f_m^m$ gene, in the $p^m$ sample is a mapping function in the form of a CNN tree, which is used to extract the structural features from $x_t$ signal. The representation of the CNN tree encoded in $f_m^m$ gene, as well as its functions and terminals, will be explained here. The representation of the CNN model in the form of GP trees is addressed in [34]. This article used a modified version of it per [20], which utilizes the morphological convolutional functions in this tree and it is appropriate for QRS Complex analysis in ECG signal. Figure 4 shows an instance of the CNN trees. As displayed in this figure, the surface of this tree comprises a variety of layers. There is an input layer in the leaf nodes of this tree that receives the $x_t$ signal as the input. Afterward, there is a morphological convolutional layer, which is normally followed by a pooling layer. In this tree, a node can be a layer of the combination of morphological convolutional / pooling. Before the root layer, there is a concatenation layer. Finally, there is the root layer, which is the output layer. In this three, the convolution layer carries out the operation of extracting the structural features using the morphological operators, which is fully explained in the next section. The pooling layer is used after the convolution layer to reduce the output dimensions of the convolution. Besides, the concatenation layer connects two input layers together. The output layer is the same as the Flatten layer in the CNN, which forms the feature diagram.
D. Set of Functions

Table II shows the set of functions used in the CNN tree. First, the functions used to change the numerical space of the features include Conv, SQRT, ADD, ReLU, Sub, and Abs, among which Conv is the most important function. This function is responsible for applying the morphological operators to the input ECG signals. On account of their special geometrical characteristics, the morphological functions such as dilation and erosion are capable of perfectly analyzing the complexities of ECG signals including QRS Complex. The morphological operators such as erosion and dilation are beneficial for analyzing the shape-oriented signals due to their theoretical framework and lower computational complexity [35, 36]. The morphological operators, i.e. erosion and dilation, are the restricted form of counter-harmonic mean morphology [37].

\[
\lim \limits_{q \to +\infty} s_q^q x_i(o) = \sup_{g \in E(u)} x_i(g) = \delta_u x_i(o) \tag{8}
\]

\[
\lim \limits_{q \to -\infty} s_q^q x_i(o) = \inf_{g \in E(u)} x_i(g) = \epsilon_u d(x) \tag{9}
\]

If \(x_i(o)\) is the input signal (\(x_i: \Omega \subset \mathbb{Z}^2 \to \mathbb{R}\), \(o \in \Omega\) shows the coordinates of an entry inside it. \(s_q^q x_i(o)\) in equations 2 and 3 is the counter-harmonic mean morphology filter that is expanded as follows.

\[
s_q^q x_i(o) = \frac{(x_i^{q+1} * u)(o)}{(x_i^q * u)(o)} \tag{10}
\]

The * indicates the convolution operations and \(u(o)\) shows the kernel function. The convolution operations are formulated in the Conv function as follows using the counter-harmonic mean morphology filter.

\[
\text{Conv}(x_i, u, q) = \frac{(x_i^{q+1} * u)(o)}{(x_i^q * u)(o)} \tag{11}
\]

The type of the operations of this function is determined with respect to the amount considered for \(q\), (\(q = 0\) linear, \(q < 0\) pseudoerosion, \(q > 0\) pseudodilation).

Sub and Add functions carry out the weight addition and subtraction operations of two signals based on the weights of \(n_1\) and \(n_2\). Two input ECG signals might have different measurements. Thus, the ECG signal will be cut using the aforementioned function to obtain two ECG signals of the same size. The Sqrt, ReLU, and Abs functions are used to change the amount of ECG signal and change the numerical space of the respective sample. In the new networks, the ReLU is preferred over the activating Sigmoid function for hidden layers for two reasons. First, it is simple and easy to use. Second, it does not cause a local minimum problem. In this function, in case the input amount is less than zero, the output will be the same as the input, and in case the input is less than or equal to zero, then the output will be zero.

The ReLU function has a fixed derivative for all inputs greater than zero. This fixed derivative accelerates the network’s learning. The Concat1, Concat2, Concat3, and Concat4 functions are used in the concatenation layer, which receives several ECG signals as input and displays them as a diagram in the output. The MaxP function performs a downsampling operation on the ECG input signal. This function can reduce the dimensions of the received ECG signal.

E. Set of Terminals

Table III shows a set of terminals and amounts authorized for them to be used in the program. The terminals include \(x_i\), \(\text{filter}^3_{x3}\), \(\text{filter}^5_{x5}\), \(\text{filter}^7_{x7}\), \(n_1\), \(n_2\), \(k_1\), and \(k_2\). \(x_i\) shows the input ECG signal. The \(\text{filter}^3_{x3}\), \(\text{filter}^5_{x5}\), and \(\text{filter}^7_{x7}\) terminals show the filter kernel that is added to this function as the second input of Conv. Considering that the convolution function performs the morphological operator, each of these filters is a diamond, disk, line, rectangle, and square-shaped matrix. \(n_1\) and \(n_2\) are added to Sub and Add functions as input and their amounts range between 0.000 to 1.000. \(k_1\) and \(k_2\) terminals are the same size as the kernel of MaxP functions. Their amounts are created randomly based on the initial range, and they are evolved in the course of evolutionary learning.

Table III shows the list of the terminals and the range of the authorized amounts used for them in the nodes of the leaves of the CNN tree. The ECG input signal is displayed through the measurement of \(n \times 1\). \(\text{filter}^3_{x3}\), \(\text{filter}^5_{x5}\), and...
filter_{3x3} terminals show the convolution filter and the second input of Conv. Taking into account that the used morphological filter is a matrix in the shape of a diamond, disk, line, rectangle, and square. $n_1$ and $n_2$ are added to Sub and Add functions as input and their amounts range between 0.000 to 1.000. $k_1$ and $k_2$ are the same size as MaxP kernel functions. Their amounts will be evolved randomly and in the course of execution of the GP algorithm.

### TABLE III
**TERMINALS EMPLOYED TO LEARN FEATURES IN CNN TRESS.**

| Terminal | VALUE RANGE | DESCRIPTION |
|----------|-------------|-------------|
| $x_1$    | [-1,1]      | The input matrix includes functional relativity information |
| filter_{3x3} | {0,1}      | Filter at the size of $3 \times 3$ in MConv function |
| filter_{5x5} | {0,1}      | Filter at the size of $5 \times 5$ in MConv function |
| filter_{7x7} | {0,1}      | Filter at the size of $7 \times 7$ in MConv function |
| $n_1, n_2$ | [0.000,1.000] | Random numbers that are the inputs of Add and Sub functions. |
| $k_1, k_2$ | {2,4}      | The kernel measurement of MaxP function |

## IV. Experimental Results

The experimental results section evaluates the performance of the proposed model for the classification of samples in the Chapman ECG dataset. It is clear that a sufficient number of samples are required as inputs to construct a classification model. However, in the Chapman dataset, some of the 11 classes have a very small number of samples to train the deep learning model. A number of these classes enjoy Atrial Tachycardia=121, Atrioventricular Node Reentrant Tachycardia=16, Atrioventricular Reentrant Tachycardia=8, Sinus Atrium to Atrial Wandering Rhythm=7 signal samples, which is quite small. Therefore, in these tests merely seven classes of arrhythmia with a sufficient number of samples are used. As reported earlier [38], the only method for the comprehensive and accurate evaluation of a model for classification of ECG signals is a model in which the signals of the training and testing parts are recorded during different sessions and within several days. This is the only acceptable modality to test the model. Consequently, all evaluations carried out in this section are performed on the basis of configuration.

All implementations and evaluations were performed on a PC with Microsoft Windows 10 OS, Intel i9-9750H processor, 32GB of RAM, and Nvidia GTX 1060 GDDR4 SLI GPU with 4GB memory. Finally, in equation (1), the value of 0.8 is considered for $\alpha$ parameter. Besides, in equation (2), the value of 0.6 is considered for $\beta$ parameters and 0.3 for $\gamma$ parameters.

### A. RESULT ANALYSIS METHODOLOGY

To analyze the results first the confusion matrix is calculated for test samples. Table IV shows the method of formation of a confusion matrix for tumors with the label of true and predicted $c_{predicted label, true label}$. The combination of 8 labels of true and predicted results in an 8*8 confusion matrix. Four evaluations criteria, i.e., ACC\textsubscript{Fold}, ACC\textsubscript{cl}, SEN\textsubscript{cl}, and SPE\textsubscript{cl}, created based on the confusion matrix are employed for evaluation of the classification models. The total accuracy of ACC\textsubscript{Fold} demonstrates the performance of classification model for all sub-classes in a special fold. This criterion is calculated by dividing the number of classified accurate samples into the total samples. This criterion can be easily obtained in the confusion matrix through adding the principal diameter with the total samples of the matrix as follows.

\[
TP\textsubscript{cl} = c_{ct,ct} \quad (12)
\]

\[
TN\textsubscript{cl} = \left( \sum_{l \in class \text{ set}} c_{cl} \right) - c_{ct,cl} \quad (13)
\]

\[
FP\textsubscript{cl} = \left( \sum_{l \in class \text{ set}} c_{cl} \right) - c_{ct,cl} \quad (14)
\]

\[
FN\textsubscript{cl} = \left( \sum_{l \in class \text{ set}} c_{cl} \right) - c_{ct,cl} \quad (15)
\]

\[
ACC\textsubscript{cl} = \frac{TP\textsubscript{cl} + TN\textsubscript{cl}}{TP\textsubscript{cl} + TN\textsubscript{cl} + FP\textsubscript{cl} + FN\textsubscript{cl}} \quad (16)
\]

\[
PRE\textsubscript{cl} = \frac{TP\textsubscript{cl} + FP\textsubscript{cl}}{TP\textsubscript{cl}} \quad (17)
\]

\[
SEN\textsubscript{cl} = \frac{TP\textsubscript{cl}}{TP\textsubscript{cl} + FN\textsubscript{cl}} \quad (18)
\]

\[
SPE\textsubscript{cl} = \frac{TN\textsubscript{cl}}{TN\textsubscript{cl} + FP\textsubscript{cl}} \quad (19)
\]

### TABLE IV
**THE CONFUSION MATRIX FOR ARRHYTHMIA CLASSES IN CHAPMAN DATASET.**

| Predicted Label | AF | SB | SVT | ST | SR | AFIB | SI |
|-----------------|----|----|-----|----|----|------|----|
| True Label      |   |    |     |    |    |      |    |
| AF              | c_{AF,AF} | c_{AF,SB} | c_{AF,SVT} | c_{AF,ST} | c_{AF,SR} | c_{AF,AFIB} | c_{AF,SI} |
| SB              | c_{SB,AF}  | c_{SB,SB}  | c_{SB,SVT}  | c_{SB,ST}  | c_{SB,SR}  | c_{SB,AFIB} | c_{SB,SI}  |
| SVT             | c_{SVT,AF} | c_{SVT,SB} | c_{SVT,SVT} | c_{SVT,ST} | c_{SVT,SR} | c_{SVT,AFIB} | c_{SVT,SI} |
| ST              | c_{ST,AF}  | c_{ST,SB}  | c_{ST,SVT}  | c_{ST,ST}  | c_{ST,SR}  | c_{ST,AFIB} | c_{ST,SI}  |
| SR              | c_{SR,AF}  | c_{SR,SB}  | c_{SR,SVT}  | c_{SR,ST}  | c_{SR,SR}  | c_{SR,AFIB} | c_{SR,SI}  |
| AFIB            | c_{AFIB,AF} | c_{AFIB,SB} | c_{AFIB,SVT} | c_{AFIB,ST} | c_{AFIB,SR} | c_{AFIB,AFIB} | c_{AFIB,SI} |
| SI              | c_{SI,AF}  | c_{SI,SB}  | c_{SI,SVT}  | c_{SI,ST}  | c_{SI,SR}  | c_{SI,AFIB} | c_{SI,SI}  |

Besides a criterion that evaluates the samples from a general viewpoint, several criteria are considered for investigating its performance as a class-oriented model. These criteria specifically examine the performance of the model regarding the classification of the samples of a $cl$ class compared to other classes. This criterion transforms classification from 8 classes into a binary classification. The binary classification enjoys two states, in which the models’ operations were correct, plus it has two states, in

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/
which the models’ operation were wrong. The two correct states are called TP and TN. Besides, two wrong states are called FP and FN. The positive signifies that the samples belongs to the current class, cl, and the negative signifies that the samples belongs to the other classes. Equations 12 to 19 show the formula definitions of these criteria. In these equations, \( cl \in \{ \text{Class set} \} \) and \( \text{set} = \{ A, F, PT, TA, DC, LC, MC, FC \} \). The three class-oriented evaluation criteria of the model are formulated as follows through employing these definitions.

B. Evaluating the Performance of the Proposed Method Using Single-Lead

To evaluate the proposed method, first, the best CNN tree model is selected to be classified in the GP algorithm. Then, it is constructed and tested by every single 12-lead ECG signal in the Chapman dataset. Tables V to XI show the results of this test separated per class. In these tests, every classification model is trained with 12-lead including Lead I, Lead II, Lead III, Lead aVR, Lead aVL, Lead aVF, Lead V1, Lead V2, Lead V3, Lead V4, Lead V5, and Lead V6. Tables V to XI show the results of the performance of the proposed classification model per each lead, as well as their mean.

### TABLE V

**RESULTS OF PERFORMANCE OF THE PROPOSED METHOD FOR CLASSIFICATION OF SIGNALS OF CLASS AF CONCERNING VARIOUS 12 LEADS OF ECG**

| Class Name | ECG Leads | PREC<sub>cl</sub> (%) | SEN<sub>cl</sub> (%) | SPEC<sub>cl</sub> (%) | ACC<sub>cl</sub> (%) |
|------------|-----------|------------------------|---------------------|----------------------|-------------------|
| AF
| Lead I | 94.15±0.1 | 94.34±0.2 | 96.01±0.2 | 97.01±0.0 |
| Lead II | 96.19±0.3 | 93.94±1.0 | 96.26±0.4 | 97.92±1.0 |
| Lead III | 95.02±0.0 | 95.56±0.4 | 95.29±0.1 | 97.15±0.8 |
| Lead aVR | 93.54±0.5 | 94.34±1.0 | 96.17±3.3 | 97.01±0.2 |
| Lead aVL | 94.62±0.3 | 95.57±0.2 | 96.65±0.7 | 97.00±0.6 |
| Lead aVF | 95.75±1.0 | 96.83±1.0 | 96.74±0.2 | 98.00±1.1 |
| Lead V1 | 92.18±0.8 | 94.23±0.9 | 95.10±1.0 | 96.00±1.1 |
| Lead V2 | 96.20±0.1 | 94.22±1.0 | 95.12±0.1 | 98.11±0.3 |
| Lead V3 | 95.16±0.4 | 96.96±0.1 | 96.17±5.0 | 97.14±0.9 |
| Lead V4 | 94.23±0.1 | 93.00±0.7 | 95.98±0.9 | 97.01±1.1 |
| Lead V5 | 97.59±1.1 | 96.48±0.6 | 96.01±1.1 | 97.19±1.2 |
| Lead V6 | 96.15±1.1 | 94.01±0.4 | 96.14±0.1 | 97.00±1.1 |
| **Average** | **94.92±0.5** | **94.79±0.6** | **95.70±0.5** | **97.21±0.8** |

### TABLE VI

**RESULTS OF PERFORMANCE OF THE PROPOSED METHOD FOR CLASSIFICATION OF SIGNALS OF CLASS SB CONCERNING VARIOUS 12 LEADS OF ECG**

| Class Name | ECG Leads | PREC<sub>cl</sub> (%) | SEN<sub>cl</sub> (%) | SPEC<sub>cl</sub> (%) | ACC<sub>cl</sub> (%) |
|------------|-----------|------------------------|---------------------|----------------------|-------------------|
| SB
| Lead I | 94.63±0.0 | 93.34±0.5 | 97.68±0.6 | 98.74±0.1 |
| Lead II | 95.28±0.3 | 94.94±1.4 | 94.48±8.8 | 97.67±0.5 |
| Lead III | 95.10±0.1 | 95.56±1.8 | 95.83±1.2 | 98.35±0.8 |
| Lead aVR | 96.67±0.3 | 93.34±1.3 | 95.84±0.9 | 97.58±0.3 |
| Lead aVL | 94.21±0.9 | 93.57±0.8 | 97.69±0.3 | 97.69±0.8 |
| Lead aVF | 94.86±0.1 | 95.83±0.4 | 96.84±0.0 | 96.42±1.0 |
| Lead V1 | 96.39±0.5 | 96.23±0.0 | 95.66±0.0 | 96.69±0.3 |
| Lead V2 | 96.12±0.8 | 94.22±1.2 | 96.00±0.0 | 98.20±0.0 |
| Lead V3 | 95.11±0.4 | 96.96±0.8 | 97.74±0.3 | 97.13±0.4 |
| Lead V4 | 94.19±0.7 | 94.00±1.4 | 95.02±0.1 | 98.10±1.3 |
| Lead V5 | 94.59±0.2 | 96.48±1.0 | 95.20±0.0 | 98.69±0.1 |
| Lead V6 | 95.83±0.5 | 95.01±0.2 | 95.29±0.9 | 96.93±0.0 |
| **Average** | **95.33±0.4** | **94.95±0.9** | **96.10±0.4** | **97.68±0.5** |

For instance, per the results of Table V, the proposed method for the classification of the samples of the AF class enjoyed the mean of PREC<sub>cl</sub>=94.92%, SEN<sub>cl</sub>=94.79%, SPEC<sub>cl</sub>=95.97%, and ACC<sub>cl</sub>=97.21% for 12 leads. This table demonstrated that the proposed method delivered a better performance for lead III (that is from the class of the
chest leads) than other leads. The proposed method for the classification of the samples of the class AF amounted to the mean value of PREC\(_{cl}\)=95.02\%, SEN\(_{cl}\)=95.56\%, SPEC\(_{cl}\)=95.10\%, and ACC\(_{cl}\)=97.15\% for the lead III in the Chapman dataset. The proposed method delivered a proper performance regarding the classification of samples of the class AF with respect to every constructed lead and test. For instance, the proposed method demonstrated a lower performance when trained by lead V1 and obtained the mean value of PREC\(_{cl}\)=92.18\%, SEN\(_{cl}\)=94.23\%, SPEC\(_{cl}\)=95.29\%, and ACC\(_{cl}\)=96.00\% for seven classes of arrhythmia in the Chapman dataset. However, these results were dropped by the mean of PREC\(_{cl}\)=2.84\%, SEN\(_{cl}\)=1.33\%, SPEC\(_{cl}\)=0.19\%, and ACC\(_{cl}\)=1.15\% when trained by lead III. It signifies that regardless of the lead it is trained to, the proposed method preserves its appropriate performance. In general, per Tables V to XI, the proposed method delivered a relatively better performance. It is interesting to know that the lead III enjoyed a better performance in all of these tables.

### TABLE IX
RESULTS OF PERFORMANCE OF THE PROPOSED METHOD FOR CLASSIFICATION OF SIGNALS OF CLASS AFIB CONCERNING VARIOUS 12 LEADS OF ECG

| Class Name | ECG Leads | PREC\(_{cl}\) (%) | SEN\(_{cl}\) (%) | SPEC\(_{cl}\) (%) | ACC\(_{cl}\) (%) |
|------------|-----------|-------------------|-----------------|-----------------|---------------|
| Lead I     | 93.46±1.4 | 93.42±1.1 | 94.59±1.4 | 97.24±1.4 |
| Lead II    | 94.73±0.8 | 92.58±0.5 | 95.41±0.6 | 94.81±1.0 |
| Lead III   | 93.70±0.2 | 93.82±0.3 | 95.70±0.8 | 95.85±0.3 |
| Lead aVR   | 92.00±1.8 | 94.00±0.1 | 94.00±1.1 | 97.69±0.0 |
| Lead aVL   | 93.47±0.1 | 93.22±0.1 | 94.51±0.0 | 95.20±0.2 |
| Lead aVF   | 93.12±1.1 | 94.62±0.7 | 93.59±1.3 | 96.58±1.9 |
| Lead V1    | 94.58±0.6 | 93.77±0.3 | 95.92±1.9 | 95.80±0.4 |
| Lead V2    | 93.03±0.3 | 93.82±1.8 | 93.70±0.1 | 97.62±0.8 |
| Lead V3    | 92.99±1.8 | 95.04±1.2 | 94.92±0.0 | 96.12±0.0 |
| Lead V4    | 93.44±0.1 | 94.10±1.0 | 95.58±1.4 | 96.85±0.2 |
| Lead V5    | 93.83±0.9 | 94.37±0.3 | 95.03±0.0 | 97.71±1.3 |
| Lead V6    | 93.38±1.0 | 94.72±0.0 | 95.00±0.1 | 97.10±0.0 |
| Average    | 93.47±0.8 | 93.95±0.6 | 94.82±0.7 | 96.54±0.6 |

Following the results of Tables V to XI, the proposed method trained by lead III provides better performance for all arrhythmia classes. Thus, the performance of the classification model constructed with lead III is investigated in more detail via a confusion matrix. In Figure 5, section (a), the confusion matrix that belongs to the lead III-based classification model is provided in the test stage for various arrhythmia classes in the Chapman dataset including Sinus Bradycardia (SB), Sinus Rhythm (SR), Atrial Fibrillation (AFIB), Sinus Tachycardia (ST), Atrial Flutter (AF), Sinus Irregularity (SI), and Supraventricular Tachycardia (SVT). This matrix, the principal diameter is an indication of the correct classes (TP), which is quite crucial for examining a classification model. In general, this table shows that the proposed method enjoyed an appropriate distribution regarding all classes with no emphasis on a specific class with overfitting and no bad performance concerning a specific arrhythmia class. In this matrix, the majority of wrong states occur between Atrial Flutter (AF) and Atrial Fibrillation (AFIB) arrhythmia classes, as well as Sinus Rhythm (SR) and Sinus Irregularity (SI) classes. In this matrix, the average of 1.8% of the AF and AFIB samples, and 3.2% of SR and SI samples are wrong. It should be noted that from the medical viewpoint, classification fault between these two arrhythmias is not crucial.

To manifest the appropriate performance of the classification model constructed concerning lead III, the convergence diagram in Figure 5, Section (b) applied to the dataset will be examined. This Figure indicates how the classification model constructed concerning lead III changes by the increase of the epoch of the accuracy criterion, as well as the epoch value in which it is stabilized. Per diagram, the accuracy increases upon the increase of epoch for various arrhythmia classes. The accuracy of this architecture was increased among the epochs of 50 to 300 and reached 95.18\%. The accuracy was stabilized when the epoch reached the range of (350-500) and no increase has occurred. However, the lead III-based constructed classification model was stabilized in the epoch amounting to 550 and it recorded an accuracy of 95.18\%. This diagram in various arrhythmia classes indicated that the number of convolution layers resulted in an increase in accuracy and architectural performance. In general, the diagram in Figure 6 shows that the classification model constructed concerning lead III enjoys a suitable convergence for the classification of the arrhythmia samples in the Chapman dataset.
C. Evaluating the Performance of the Proposed Method Using Multiple Leads

In this section, the proposed method of the best CNN tree model is selected to be classified in the GP algorithm. Then, it is evaluated by 12 lead ECG signals in the Chapman dataset. Tables XII show the results of this test separated per class. In these tests, the classification model is trained with all 12 leads including Lead I, Lead II, Lead III, Lead aVR, Lead aVL, Lead aVF, Lead V1, Lead V2, Lead V3, Lead V4, Lead V5, Lead V6. Following the results of Table XII, the proposed method enjoys the mean of PRECcl=98.13%, SENcl=98.09%, SPECcl=97.90%, and ACCcl=98.29% regarding seven classes of arrhythmia. Table XIII demonstrates the performance of the proposed method using the Kappa criterion. In this table, separate results are published per different classes of arrhythmia. Using the observed data to calculate the possibilities of each observer, which randomly indicated one of the classes.

In case the classifier is full in the agreement, then, K=1. In case there is no contract between the assessors, despite being accessed randomly, then K=1. Per the Table, the classifier of the proposed method obtained the Kappa criterion near 1 for all classes. This indicates the quality of the proposed method with regard to this criterion. Per the results of this Table, the classification model trained by all 12 leads obtained a higher mean PRECcl=3.11%, SENcl=2.53%, SPECcl=2.80%, and ACCcl=1.14% than the classification model trained by lead III. Therefore, it can be concluded that the construction of a classification method for the detection of arrhythmia of 12-lead ECG signal delivered a better performance. However, it manifested a computational complexity and had a larger number of parameters. Then, Figure 6 sections (a) and (b) show the confusion matrix and convergence diagram of the classification model constructed concerning all 12 leads of ECG in the test phase, respectively. In general, this confusion matrix demonstrates that the proposed method constructed concerning all 12-lead ECG manifested an appropriate distribution regarding all classes, with no emphasis on a specific class with overfitting and no bad performance concerning specific arrhythmia classes. In this matrix, there were errors between the arrhythmia classes that are not crucial from the medical viewpoint. Figure 5, Section (b), shows the convergence diagram of the classification model based on 12-lead ECG signals. The proposed method enjoys a proper convergence regarding various arrhythmia classifications in the Chapman dataset. Besides, even the convergence rate is improved in comparison to Figure 5, Section (b) and it has rapidly obtained convergence.
FIGURE 6. (a) The confusion matrix includes 8 different classes of arrhythmia in the Chapman dataset Constructed Concerning 12-Lead ECG and (b) Its Convergence Rate Curve

![Confusion Matrix and Convergence Rate Curve](image)

**FIGURE 7.** Accuracy of the classification model by increasing the number of samples Constructed Concerning 12-Lead ECG

![Accuracy Curve](image)

**TABLE XIV**

| Class Name          | Value                  |
|---------------------|------------------------|
| No. of Parameters   | ~104.25 million        |
| GPU memory          | 0.78 giga byte         |
| Computation (GFLOPs)| ~ 2.98                 |
| Training Time       | 25:20:40               |
| Validation Time     | 03:40:20               |
| Testing Time        | 00:59:30               |

**D. COMPARING THE PROPOSED METHOD TO OTHER DEEP LEARNING-BASED METHODS**

In this section, the performance of the classification model constructed concerning lead III is compared to the other state-of-the-art methods based on deep learning. First, in Table XIV the classification model constructed concerning lead III and methods whose input is based on Single-lead ECG are compared. This Table includes the details such as the number of patients, number of ECG records, the number of the diagnosed classes and rhythms, as well as the method used in them. The results of the performance of the previous methods are displayed on the basis of the criterion that is reported in them such as accuracy or F1-Score. Various deep learning models such as LSTM, CNN, and RNN are used in these methods. Traditionally, a broad spectrum of these methods are evaluated on an MIT-BIH Arrhythmia dataset, which belong to PhysioNet. For instance, in [39], a deep CNN model was employed for the classification of 12 rhythms. In this model, various stages such as batch normalization and data augmentation were used. The results of this investigation indicated the value of an F1 value of 83% for 53,549 patients. In [26], a deep learning model, with the standard U-Net architecture was used for the classification of 5 types of arrhythmia
The single-lead-III model demonstrated 0.02% higher accuracy in comparison to the [15], which also used the Chapman ECG dataset. In most of the previous investigations, similar records were used in the train and test dataset, which reduced the integration of the methods and cast doubt on when dealing with unseen data. A subject that was taken into account in the proposed model of this article. For a better representation of the performance of the proposed method, the proposed classification model developed considering all 12 leads of ECG was compared to the previous methods developed with regard to 12-lead ECG in Table XVI. The previous methods listed in this table are based on deep learning models such as CNN and LSTM. In [29], on account of combining CNN+LSTM, in addition to the classification of seven rhythms based on single-lead ECG illustrated in Table XV, a separate section was allocated for the classification based on the 12-lead ECG signals.

Its results for 10,436 patients revealed an Acc of 92.24% at the evaluation section. Moreover, in [15] printed recently, a method based on Detrending+ResNet model was introduced for 10,093 patients in the Chapman ECG dataset. Its results for the classification of seven rhythms manifested the Acc of 92.24% in the evaluation section. Deep learning models such as CNN deliver better performance by increasing the number of inputs. Thus, methods provided in Table XVI are regarded as pioneers and most of them enjoy quite a high accuracy.

| References         | #Subjects | #Records | #Rhythm | Method          | Performance |
|--------------------|-----------|----------|---------|-----------------|-------------|
| Acharya et al. [40]| 47        | 109,449  | 5 Class | CNN             | Acc: 94.03  |
| Xu et al. [39]     | 22        | 50,977   | 5 Class | DNN             | Acc: 93.10  |
| Gao et al. [41]    | -         | 93,371   | 8 Heartbeats | LSTM         | Acc: 90.26  |
| Hamun et al. [21]  | 53,549    | 9,232    | 12 Rhythm | CNN           | F1: 83.00   |
| Yildirim et al. [23]| 45      | 1,000    | 5 Heartbeats | CNN          | Acc: 91.33  |
| Shaker et al. [27] | 44        | 102,098  | 12 Class | CNN            | Acc: 94.30  |
| Oh et al. [26]     | 47        | 16,499   | 5 Heartbeats | UNet          | Acc: 93.10  |
| Xiong et al. [42]  | 12,186    | 12,186   | 4 Class | CNN+RNN        | F1: 82.00   |
| Oh et al. [25]     | 70        | 150,268  | 3 Cardiac Disease | CNN+LSTM   | Acc: 94.51  |
| Mousavi et al. [43]| -         | 750      | 5 Rhythm | CNN+LSTM      | Acc: 93.75  |
| Wu et al. [44]     | -         | 8,528    | 4 Class | Binarized CNN  | F1: 86.00   |
| Fujita et al. [45]| 47        | 109,449  | 4 Class | Normalization+ CNN | Acc: 93.45 |
| Salem et al. [46]  | 22        | 7000     | 4 Class | STFT + CNN     | Acc: 94.23  |
| Xia et al. [47]    | -         | -        | 2 Class | SWT+CNN        | Acc: 95.63  |
| Yildirim et al. [29]| 10,436  | 10,436   | 7 Rhythm | CNN+LSTM      | Acc: 92.24  |
| Proposed Method    | 10,646    | 10,646   | 7 Rhythm | Meta Learning+ CNN Tree | Acc: 96.92 |

| References         | #Subjects | #Records | #Rhythm | Method          | Performance |
|--------------------|-----------|----------|---------|-----------------|-------------|
| Chang et al. [28]  | 38,899    | 65,932   | 12 Rhythm | LSTM           | Acc: 90.00  |
| Yao et al. [48]    | -         | 6,877    | 8 Rhythm | CNN             | F1: 81.00   |
| Li et al. [22]     | 47        | 94,013   | 5 Class | ResNet          | Acc: 99.38  |
| Yildirim et al. [29]| 10,436  | 10,436   | 7 Rhythm | CNN+LSTM      | Acc: 96.13  |
| Faust et al. [15]  | 10,093    | 10,093   | 5 Rhythm | Detrending+ResNet | Acc: 98.55 |
| Proposed Method    | 10,646    | 10,646   | 7 Rhythm | Meta Learning+ CNN Tree | Acc: 98.29 |
V. DISCUSSION
Subsection IV.D demonstrates the superiority of the proposed model over other available methods of diagnosing arrhythmia of ECG signal by manifesting the numerical results and necessary comparisons. This subsection counts several crucial reasons and required justifications for this superiority. Deep learning models are statistical models that are dependent on data distribution. These models perform well merely when faced with data distribution that they are trained with. For instance, in the [42] method, the model obtained an F-measure of almost 99% for training and testing data, however, in experiments carried out in this study it was observed that it enjoys the F-measure of 82% for the Chapman dataset. In light of that, the generalization of a statistical model is directly dependent on the sample distribution. In the proposed method, the statistical model was trained based on data distribution, and this distribution does not necessarily exist in the testing data section. As already mentioned in [19], the latent medical variable cannot be directly achieved in the medical data. Methods such as frequency analysis can express these variables. In the proposed method special frequencies were used in the ECG signal, which helped extract the functional dependency in the ECG leads. It was not examined in methods such as [15] and [29]. The final feature considered for the proposed method is the independence of this method in specifying and extracting QRS Complex. Methods such as Big-ECG [14] of their performance are directly dependent on specifying the stroke. It reduces the flexibility of the model since, in some types of arrhythmia, the QRS peak undergoes changes and is quite difficult to specify. In general, these are several justifications that can be stated to improve the proposed method in comparison to other previous methods.

VI. CONCLUSION
This article proposed an Interpretable Meta Structural Learning algorithm regarding the challenging problems to classify various arrhythmia of ECG signals recorded in several sessions for each person. Therefore, a compound loss function was provided that included a structural feature extraction fault and a space label fault with GUMBEL-SOFTMAX distribution in the CNN models. The collaboration was carried out between models to create the learning to learn feature in these models via transferring the knowledge among them when dealing with unseen samples. This article encoded the models in the form of evolutionary trees of the GP algorithm to create the interpretability feature for CNN models. These trees learn the process of extracting deep structural features in the course of the evolution of the GP algorithm. The experimental results suggested that the proposed classification model enjoys an accuracy of 98% for the classification of 7 types of arrhythmia in the samples of the ChapmanECG dataset on 10646 patients, which were recorded in different sessions. Finally, the comparisons demonstrated the competitive performance of the proposed model through state-of-the-art methods based on the big learning models.

REFERENCES
[1] R. Hu, J. Chen, and L. Zhou, "A transformer-based deep neural network for arrhythmia detection using continuous ECG signals," (in eng), Comput Biol Med, vol. 144, p. 105325, May 2022, doi: 10.1016/j.compbiomed.2022.105325.
[2] A. Ahmadi, M. Kashefi, H. Shahrokhi, and M. A. Nazari, "Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes," Biomedical Signal Processing and Control, vol. 63, p. 102227, 2021/01/1/ 2021, doi: https://doi.org/10.1016/j.bspc.2020.102227.
[3] D. Mozaffarian et al., "Executive summary: heart disease and stroke statistics—2016 update: a report from the American Heart Association," Circulation, vol. 133, no. 4, pp. 447-454, 2016.
[4] F. M. Kusumoto, ECG Interpretation, Springer International Publishing, 2020.
[5] S. Kusurna and K. R. Jothis, "ECG signals-based automated diagnosis of congestive heart failure using Deep CNN and LSTM architecture," Biohybernetics and Biomedical Engineering, 2022.
[6] J. Huang, B. Chen, B. Yao, and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," IEEE Access, vol. 7, pp. 92871-92880, 2019, doi: 10.1109/ACCESS.2019.2928017.
[7] L. V. Kumar and Y. P. Sai, "A new transfer learning approach to detect cardiac arrhythmia from ECG signals," Signal, Image and Video Processing, pp. 1-9, 2022.
[8] M. Sepahvand and F. Abdali-Mohammadi, "A Novel Method for Reducing Arrhythmia Classification from 12-Lead ECG Signals to Single-Lead ECG with Minimal Loss of Accuracy through Teacher-Student Knowledge Distillation," Information Sciences, 2022.
[9] J.-A. Lee and K.-C. Kwak, "Personal Identification Using an Ensemble Approach of 1D-LSTM and 2D-CNN with Electrocardiogram Signals," Applied Sciences, vol. 12, no. 5, p. 2692, 2022.
[10] P. Sodmann, M. Vollmer, N. Nath, and L. Kaderali, "A convolutional neural network for ECG annotation as the basis for classification of cardiac rhythms," Physiological Measurement, vol. 39, no. 10, p. 104005, 2018/10/24 2018, doi: 10.1088/1361-6579/aae304.
[11] C.-H. Hsieh, Y.-S. Li, B.-J. Hwang, and C.-H. Hsiao, "Detection of Atrial Fibrillation Using 1D Convolutional Neural Network," Sensors, vol. 20, no. 7, 2020, doi: 10.3390/s20072136.
[12] H. Liu, D. Chen, and G. Sun, "Detection of Fetal ECG R Wave From Single-Lead Abdominal ECG Using a Combination of RR-Time-Series Smoothing and Template-Matching Approach," IEEE Access, vol. 7, pp. 66633-66643, 2019, doi: 10.1109/ACCESS.2019.2917826.
[13] J. Zheng, J. Zhang, S. Daniokio, H. Yao, H. Guo, and C. Rakovski, "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," Scientific Data, vol. 7, no. 1, p. 48, 2020/02/12 2020, doi: 10.1038/s41597-020-0386-x.
[14] I. Hussain and S. J. Park, "Big-ECG: Cardiographic Predictive Cyber-Physical System for Stroke Management," IEEE Access, vol. 9, pp. 123146-123164, 2021, doi: 10.1109/ACCESS.2021.3109806.
[15] O. Faust and U. R. Acharya, "Automated classification of five arrhythmias and normal sinus rhythm based on RR interval signals," Expert Systems with Applications, vol. 181, p. 115031, 2021/11/01/ 2021, doi: https://doi.org/10.1016/j.eswa.2021.115031.
[31] M. Sepahvand and F. Abdali-Mohammadi, "Overcoming breast cancer histopathological images through a novel decomposed feature-based knowledge distillation method," *Computers in Biology and Medicine*, vol. 145, p. 105413, 2022/06/01/ 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105413.

[32] M. Sepahvand and F. Abdali-Mohammadi, "A novel representation in genetic programming for ensemble classification of human motions based on inertial signals," *Expert Systems with Applications*, vol. 185, p. 115624, 2021/12/15/ 2021, doi: https://doi.org/10.1016/j.eswa.2021.115624.

[33] M. Sepahvand, F. Abdali-Mohammadi, and F. Mardukhi, "Evolutionary Metric-Learning-Based Recognition Algorithm for Online Isolated Persian/Arabic Characters, Reconstructed Using Inertial Pen Signals," *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2872-2884, 2017, doi: 10.1109/TCYB.2016.2633318.

[34] Y. Bi, B. Xue, and M. Zhang, "An Evolutionary Deep Learning Approach Using Genetic Programming with Convolution Operators for Image Classification," in 2019 IEEE Congress on Evolutionary Computation (CEC), 8-13 July 2019 2018, pp. 1-6, doi: 10.1109/CEC.2018.8477933.

[35] A. Y. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, no. 1, pp. 65-69, 2019/01/01 2018, doi: 10.1038/s41591-018-0268-3.

[36] F. Li, J. Wu, M. Jia, Z. Chen, and Y. Pu, "Automated Heartbeat Classification Exploiting Convolutional Neural Network With Channel-Wise Attention," *IEEE Access*, vol. 7, pp. 122955-122963, 2019, doi: 10.1109/ACCESS.2019.2938617.

[37] Ö. Yıldırım, P. Pławiak, R.-S. Tan, and U. R. Acharya, "Arrhythmia detection using deep convolutional neural network with long duration ECG signals," *Computers in Biology and Medicine*, vol. 102, pp. 411-420, 2018/11/01/ 2018, doi: https://doi.org/10.1016/j.compbiomed.2018.09.009.

[38] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664-675, 2015.

[39] S. L. Oh, E. Y. K. Ng, R. S. Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats," *Computers in Biology and Medicine*, vol. 102, pp. 278-287, 2018/11/01/ 2018, doi: https://doi.org/10.1016/j.compbiomed.2018.06.002.

[40] M. Sepahvand and F. Abdali-Mohammadi, "A Deep Learning-Based Compression Algorithm for 9-DOF Inertial Measurement Unit Signals Along With an Error Compensating Mechanism," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 632-640, 2019, doi: 10.1109/JSEN.S.2018.2877360.

[41] A. M. Shaker, M. Tantawi, H. A. Sheedeed, and M. F. Tolba, "Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks," *IEEE Access*, vol. 8, pp. 35592-35605, 2020, doi: 10.1109/ACCESS.2020.2974712.

[42] K.-C. Chang et al., "Usefulness of machine learning-based detection and classification of cardiac arrhythmias with 12-lead electrocardiograms," *Canadian Journal of Cardiology*, vol. 37, no. 1, pp. 94-104, 2021.

[43] O. Yıldırım, M. Talo, E. J. Ciaccio, R. S. Tan, and U. R. Acharya, "Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ECG records," *Computer Methods and Programs in Biomedicine*, vol. 197, p. 105740, 2020/12/01/ 2020, doi: https://doi.org/10.1016/j.cmpbi.2020.105740.

[44] M. Sepahvand, F. Abdali-Mohammadi, and A. Taherkordi, "Teacher-student knowledge distillation based on decomposed deep feature representation for intelligent mobile applications," *Expert Systems with Applications*, vol. 202, p. 117474, 2022/09/15/ 2022, doi: https://doi.org/10.1016/j.eswa.2022.117474.

[45] M. Sepahvand and F. Abdali-Mohammadi, "Overcoming limitation of dissociation between MD and MI classifications of

VOLUME XX, 2017
MAYTHAM N. MEQDAD was born in 1987, in Babylon. He received the B.Sc. degree in computer science from Baghdad College of Economic Sciences University in 2009. In 2012, he obtained his M.Sc. degree in computer science & communication from Arts, Sciences and Technology University in Lebanon. Currently he is working as assistant lecturer in Al-Mustaqbal University College, Babylon, Iraq since 2013. His research interests include deep learning, evolutionary algorithms, and biomedical signal processing.

FARDIN ABDALI-MOHAMMADI was born in KhorramAbad, Lorestan, Iran, in 1979. He received the B.Sc. degree in computer engineering from the Sharif University of Technology in 2001, the M.Sc. degree in software engineering from the Amirkabir University of Technology in 2003, and the Ph.D. degree from Isfahan University in 2012. He is currently an Associate Professor with the Department of Computer Engineering and Information Technology, Razi University, Kermanshah, Iran. His research interests include Medical Informatics, Biomedical signal and image processing, Data mining and Bioinfo.

SEIFEDINE KADRY (Senior Member, IEEE) received the bachelor's degree from Lebanese University, in 1999, the dual M.S. degree from Reims University, France, and EPFL, Lausanne, in 2002, the Ph.D. degree from Blaise Pascal University, France, in 2007, and the H.D.R. degree from Rouen University, in 2017. He is an ABET Program Evaluator for computing and ABET Program Evaluator for engineering tech. His research interests include data science, education using technology, system prognostics, stochastic systems, and applied mathematics.