



## G2-ResNeXt: A novel model for ECG signal classification

Shengnan Haoi, Hang Xui, Hongyu Ji, Zhiwu Wang, Li Zhao, Zhanlin Ji, IVAN GANCHEV

### Publication date

06-04-2023

### Published in

IEEE Access, 2023, vol. 11, pp. 34808-34820

### Licence

This work is made available under the [CC BY-NC-SA 4.0](#) licence and should only be used in accordance with that licence. For more information on the specific terms, consult the repository record for this item.

### Document Version

1

### Citation for this work (HarvardUL)

Haoi, S., Xui, H., Ji, H., Wang, Z., Zhao, L., Ji, Z. and GANCHEV, I. (2023) 'G2-ResNeXt: A novel model for ECG signal classification', available: <https://doi.org/10.34961/researchrepository-ul.22961876.v1>.

This work was downloaded from the University of Limerick research repository.

For more information on this work, the University of Limerick research repository or to report an issue, you can contact the repository administrators at [ir@ul.ie](mailto:ir@ul.ie). If you feel that this work breaches copyright, please provide details and we will remove access to the work immediately while we investigate your claim.

Received 6 March 2023, accepted 3 April 2023, date of publication 6 April 2023, date of current version 12 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265305

## RESEARCH ARTICLE

# G2-ResNeXt: A Novel Model for ECG Signal Classification

SHENGNAN HAO<sup>1</sup>, HANG XU<sup>1</sup>, HONGYU JI<sup>2</sup>, ZHIWU WANG<sup>3</sup>, LI ZHAO<sup>4</sup>,  
ZHANLIN JI<sup>1,5</sup>, (Member, IEEE), AND IVAN GANCHEV<sup>5,6,7</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Artificial Intelligence, North China University of Science and Technology, Tangshan 063009, China

<sup>2</sup>School of Biosciences, University of Sheffield, S10 2TN Sheffield, U.K.

<sup>3</sup>Department of Chemoradiotherapy, Tangshan People's Hospital, Hebei, Tangshan 063009, China

<sup>4</sup>Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China

<sup>5</sup>Telecommunications Research Centre (TRC), University of Limerick, Limerick, V94 T9PX Ireland

<sup>6</sup>Department of Computer Systems, University of Plovdiv "Paisii Hilendarski," 4000 Plovdiv, Bulgaria

<sup>7</sup>Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Corresponding authors: Zhanlin Ji (zhanlin.ji@gmail.com) and Ivan Ganchev (ivan.ganchev@ul.ie)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0135700; in part by the Tsinghua Precision Medicine Foundation under Grant 2022TS003; in part by the Bulgarian National Science Fund (BNSF) under Grant No. КП-06-ИП-КИТАЙ/1 (KP-06-IP-CHINA/1); and in part by the Telecommunications Research Centre (TRC), University of Limerick, Ireland.

**ABSTRACT** Electrocardiograms (ECG) are the primary basis for the diagnosis of cardiovascular diseases. However, due to the large volume of patients' ECG data, manual diagnosis is time-consuming and laborious. Therefore, intelligent automatic ECG signal classification is an important technique for overcoming the shortage of medical resources. This paper proposes a novel model for inter-patient heartbeat classification, named G2-ResNeXt, which adds a two-fold grouping convolution (G2) to the original ResNeXt structure, as to achieve better automatic feature extraction and classification of ECG signals. Experiments, conducted on the MIT-BIH arrhythmia database, confirm that the proposed model outperforms all state-of-the-art models considered (except the GRNN model for one of the heartbeat classes), by achieving *overall accuracy* of 96.16%, and *sensitivity* and *precision* of 97.09% and 95.90%, respectively, for the ventricular ectopic heartbeats (VEB), and of 80.59% and 82.26%, respectively, for the supraventricular ectopic heartbeats (SVEB).

**INDEX TERMS** Cardiovascular disease (CVD), ECG signal classification, ResNeXt, convolutional block attention module (CBAM), MIT-BIH.

## I. INTRODUCTION

Cardiovascular disease (CVD) is a chronic disease of aging with a high mortality rate. As reported in the "Top Ten Causes of Death" issued by the World Health Organization (WHO) [1], CVD is the No. 1 killer in the world as more people die from it annually than from other causes. By 2030, 23.6 million people are expected to die from CVD. Electrocardiogram (ECG) is a non-invasive diagnostic tool for cardiac pathology, which plays an important role in the classification of CVDs. However, the timely and accurate

detection of abnormal heartbeat signals in patients' ECG has become a major difficulty in the medical field.

As a main method of diagnosis of arrhythmia, ECG could objectively reflect the physiological and working conditions of all parts of the heart, which is of great significance for the detection of CVDs. The typical basic waveform of ECG mainly includes a P wave, a QRS wave, and a T wave, as shown on Figure 1. The P wave represents the depolarization process of two atria, the QRS wave represents the depolarization process of two ventricles, and the T wave represents the repolarization process of two ventricles [2].

Usually, arrhythmia is divided into two types, i.e., fatal arrhythmia, and non-fatal arrhythmia. The fatal arrhythmia needs to be immediately treated, otherwise it can be

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei<sup>1</sup>.

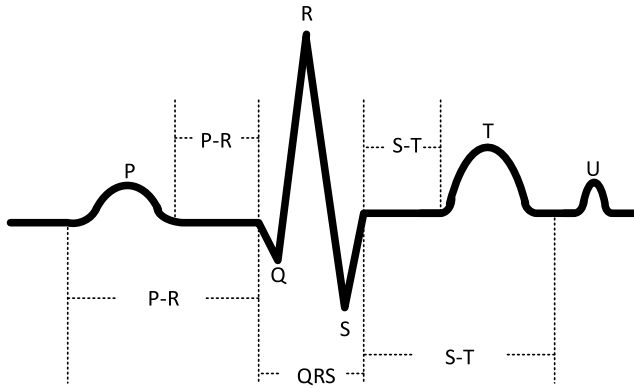


FIGURE 1. A single heartbeat waveform.

life-threatening. For the non-fatal arrhythmia, relevant examination and treatment are also required, and ECGs should be regularly conducted. Basically, early detection and treatment of arrhythmia can tackle the problem of sudden death [3]. Arrhythmia is usually caused by an irregular heartbeat, which can be found from the interval and amplitude of ECG signals, meaning that the shapes and other morphological characteristics of ECG signals determine the type of arrhythmia [4]. In the early clinical practice, doctors analyzed ECGs by visual evaluation and, based on their experience, identified the characteristics of ECG signals and provided diagnostic results. However, the strong non-linearity, non-stationarity, and randomness of ECG signals make the classification of arrhythmia in ECG signals a very difficult task.

This paper proposes an improved version of the ResNeXt model [5] for ECG signal classification, called G2-ResNeXt, which adds a two-fold grouping convolution (G2) to the original ResNeXt structure, along with a modified convolutional block attention module (CBAM), allowing the model to focus more on the changes in the characteristics of the ECG signals and, as a result, to achieve better automatic feature extraction and classification of ECG signals. According to the results obtained by experiments, conducted on the MIT-BIH arrhythmia database, the proposed model outperforms all state-of-the-art models considered, according to all evaluation metrics used, except the GRNN model which achieves better *sensitivity* and *precision* in classifying heartbeats of class S of the Association for the Advancement of Medical Instrumentation (AAMI) standard [6].

The remainder of the paper is structured as follows. Sections II and III present the necessary background of ECG signals and relevant neural networks, respectively. Section IV presents the related work done in the field of applying an artificial intelligence for ECG signal detection and classification. Section V describes the proposed G2-ResNeXt model. Section VI presents experiments, conducted for performance evaluation of the proposed model compared to other state-of-the-art models, and corresponding results obtained. Finally, Section VII concludes the paper.

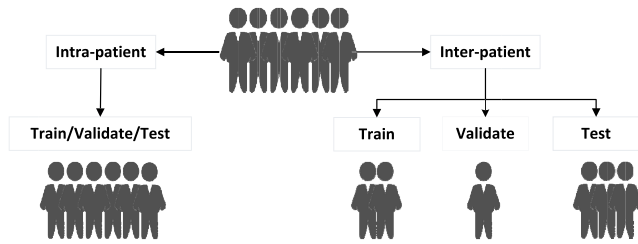
TABLE 1. Mapping of the MIT-BIH heartbeat types to the AAMI heartbeat classes.

AAMI heartbeat classes	MIT-BIH heartbeat types
Normal (N)	Normal beat
	Left bundle branch block beat
	Right bundle branch block beat
	Atrial escape beat
	Junctional escape beat
Supraventricular ectopic beat (S)	Junctional premature beat
	Atrial premature beat
	Aberrated atrial premature beat
	Supraventricular premature beat
	Premature ventricular contractions
Ventricular ectopic beat (V)	Ventricular escape beat
Fusion beat (F)	Fusion of non-ectopic and ventricular beat
	Unclassified beat
Unknown beat (Q)	Paced beat
	Fusion of paced and normal beat

## II. BACKGROUND OF ECG SIGNALS

### A. MIT-BIH ARRHYTHMIA DATABASE

The MIT-BIH arrhythmia database [7], provided by the Massachusetts Institute of Technology – Boston’s Beth Israel Hospital, is among the most internationally-renowned and commonly used databases as a source of clinical ECG signals, along with the AHA database [8], provided by the American Heart Association, and the European ST-T ECG database [9]. The MIT-BIH arrhythmia database contains 48 half-hour two-channel ambulatory ECG recordings obtained from 47 subjects with a resolution of 11 bits and range of 10 mV. The ECG recordings of 25 male subjects (aged 32-89) and 22 female subjects (aged 23-89) are included in the database, 60% of which are inpatients [7]. MIT-BIH is composed of three files – a header file, a data file, and an annotation file. The latter was produced by two experienced cardiologists, whereby around 110,000 computer-readable reference annotations were made by them. In most recordings, the first lead is of type MLII (obtained by placing the electrode on the chest) and the second lead is usually V1 (occasionally V2 or V5, and in one instance V4). The MLII leads were chosen for use in the experiments, presented further in this paper, because these were the most numerous. 15 heartbeat types are mapped to the 5 main classes of AAMI standard, as shown in Table 1. As unknown beats (class Q) are difficult for identification, no data of Class Q are distinguished in the existing mainstream models. Similarly, in this paper, only data of classes N, S, V, and F are distinguished, with class-Q data neglected.



**FIGURE 2.** An illustration of the *intra*-patient and *inter*-patient heartbeat classification methods.

**TABLE 2.** The training set, validation set, and test set, used in the conducted experiments.

Data sets	MIT-BIH patient record numbers
Training set	101, 106, 108, 109, 112, 116, 118, 122, 124, 201, 203, 207, 208, 209, 215, 220, 230
Validation set	114, 115, 119, 205, 223
Test set	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234

### B. INTRA-PATIENT VS. INTER-PATIENT CLASSIFICATION

The classification of heartbeat signals has been subject to many studies. Generally, there are two classification methods that can be used, namely the *intra*-patient method and the *inter*-patient method [10], as depicted on Figure 2.

The *inter*-patient classification method treats all ECG data of each patient as a whole. When splitting the dataset into a training set, validation set, and test set, it ensures that the records of the same patient do not appear in more than one set. The resulting model has a strong generalization ability and can be applied to real-life situations with practical implications. The *intra*-patient classification method mixes all the data in the dataset together and then splits these into a training set, validation set, and test set [11], [12]. This results in training the model on part of the ECG data of a patient and then testing it on the rest of the ECG data of the same patient to achieve a higher performance, which is very unreasonable. Moreover, using the ECG data of the same patient for both model training and testing may lead to biased results: if the classification model has already seen ECG data of a patient during training, then the model is likely to have acquired the characteristics of that patient. Therefore, as being more realistic and meaningful, the *inter*-patient method was used in the experiments described further in this paper. More specifically, the MIT-BIH records of 44 patients were used as follows: the records of 17 patients were used for the training set, the records of 5 patients for the validation set, and the records of 22 patients for the test set (Table 2).

### C. ECG SIGNAL SLICING

At present, most of research on heartbeat classification relatively depends on the accuracy of QRS wave detection. As the characteristics of the human body and the form and parameters of a single signal are variable, certain differ-

**TABLE 3.** Sample classification of ECG signal slices.

Slice	Slice characteristics	Conclusion about slice class	Slice class
[N,N,N,N,N]	Only normal heartbeats	The slice is of normal class	N
[N,N,N,N,F]	Normal heartbeats with only one abnormal heartbeat	The class of the slice corresponds to the class of the abnormal heartbeat	F
[N,V,V,V,F]	Multiple numbers of different abnormal heartbeats	The class of the slice corresponds to the class of the most represented abnormal class	V
[N,F,F,V,V]	Multiple numbers of different abnormal heartbeats with the same highest quantity	The class of the slice corresponds to the class of the first appearing abnormal heartbeat	F

ence may exist in the positioning results of different QRS wave detection algorithms under the same accuracy [13]. For example, some QRS algorithms provide position slightly to the left, while others provide position slightly to the right, resulting in an obvious deviation in the final positioning results even if the same data are processed. As a result, the waveform performance of the acquired heartbeats can vary, leading to dependence on the QRS localization algorithm for the learning process of the neural network later on, making the network somewhat coupled to the QRS wave detection algorithm.

Instead of using a QRS wave detection, an ECG signal slicing method is used in the research, presented in this paper, whereby the location and connotation of useful information are learned by the neural network itself, freeing the network from coupling with the QRS wave detection algorithm, and making the signal diagnosis a simpler and more general process that can be used more widely in reality. In the experiments, reported further in this paper, 1080-length slices were obtained by capturing three seconds of an ECG signal at a sampling rate of 360 Hz. For the classification of slices, the following rules (illustrated in Table 3) were applied:

- 1) When only normal heartbeats are present in a slice, the slice is defined as belonging to the normal class (N).
- 2) When only one abnormal heartbeat is present in a slice, and all other heartbeats are normal, the slice is defined as belonging to the class of the abnormal heartbeat (i.e., V, S, or F).
- 3) When multiple numbers of different abnormal heartbeats exist in a slice, the most represented abnormal class (i.e., with the largest quantity) defines the class of the slice.
- 4) When multiple numbers of different abnormal heartbeats with the same highest quantity exist in a slice, the

class of the first appearing abnormal heartbeat defines the class of the slice.

In addition, a slice overlapping method is used in the research, presented in this paper, to alleviate the class imbalance problem of AAMI heartbeat data contained in the MIT-BIH arrhythmia database, so that the obtained new samples and original samples be different (to some degree) and of better quality (theoretically). This imbalance problem may have a severe effect on the model training process, thus likely invalidating the neural network learning. To alleviate further this problem, a stacking of slices was utilized in the experiments, described further in this paper, allowing to collect more samples by using more overlapping between adjacent slices of a class with a small sample size. Compared with simple oversampling used to obtain completely consistent new samples, this slice-and-stack approach is more effective, because simple oversampling generates new data samples for fewer sample classes to participate in training, but may result in identical new samples, which is prone to model overfitting problems.

#### D. DATA DENOISING

The ECG signal is a low-frequency and high-impedance weak signal [14], which is easily affected by the in-vivo and in vitro environments during its acquisition. In-vivo effects refer to adverse reactions of the patient's body during the ECG measurement that cause errors in the test results, such as muscle contractions due to the patient's mental stress during the measurement resulting in inotropic interference, or the patient breathing excessively resulting in unstable ECG amplitude. In vitro effects, such as electromagnetic interference in the surrounding environment, have a relatively small impact but nonetheless should be taken into account too. As a result, the acquired ECG signal is usually accompanied with lots of noise, such as baseline drifts, power frequency interference, EMG interference and motion artifact, which may cause wrong classification. To improve the classification precision, in this paper, the wavelet transform is used to eliminate the noise, as follows [15]:

$$WT_x(\alpha, \tau) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x(t) \varphi^*\left(\frac{t-\tau}{\alpha}\right) dt, \quad (1)$$

where  $\alpha$  denotes the scale factor, used to stretch the basic wavelet  $\varphi(t)$  function, and  $\tau$  reflects the displacement. Figure 3 shows the overall denoising process of the ECG signal.

It is very important for signal denoising to select an appropriate decomposition level, because each wavelet in the wavelet transform has its own characteristics. In the research, presented in this paper, the DB8 wavelet was used because it could more effectively reduce the noise in the ECG signals compared with other wavelets, [16]. In response to the large fluctuation of ECG signals, a 9-layer wavelet decomposition was adopted, whereby the wavelet coefficient of each layer was kept for wavelet reconstruction. A soft threshold process



FIGURE 3. The ECG signal denoising process.

was used, as per the following equation:

$$w_{new} = \begin{cases} w_{old}^* f(x), & |w_{old}| > T \\ 0, & |w_{old}| < T, \end{cases} \quad (2)$$

where  $f(x)$  denotes the shrinkage function,  $w_{old}$  denotes the input wavelet values,  $w_{new}$  denotes the output wavelet values, and  $T$  denotes the threshold of the wavelet transform. The denoising effect is shown in Figure 4.

### III. BACKGROUND OF CNNs AND ResNets

#### A. CNN

Compared with traditional neural networks, convolutional neural networks (CNNs) are characterized by weight sharing and local connection, which could greatly improve their feature extraction capability and training efficiency. CNNs can be applied not only in the field of 2D images but also in the field of 1D data, such as natural language processing (NLP) and some physiological signals (e.g., blood pressure signals [17], respiratory signals [18], ECG signals [19], etc.). The basic CNN structure consists of an input layer, a convolutional layer, a pooling layer, a fully-connected layer, and an output layer. Generally, several convolutional layers and pooling layers are used alternately, that is, a convolutional layer is connected to a pooling layer, which in turn is connected to another convolutional layer and so on, as shown in Figure 5. In some studies, Long Short-Term Memory (LSTM)-based models are used for classification. Compared with CNN, the LSTM structure is relatively complex. Each LSTM cell includes four fully-connected layers (MLPs). If a LSTM has a large time span and the network is very deep, the computation complexity is significant, and the training is slow. The parameter sharing nature of the convolutional operations in CNN allows for a significant reduction in the number of parameters to be optimized, increasing the speed of model training. Moreover, since convolutional operations mainly deal with grid-like data, they have significant advantages for the analysis and recognition of time series and image data. In addition, although LSTMs alleviate the long-term dependency problem of RNNs to a certain extent, they are also tricky for use with longer sequence data.

#### B. ResNet

The deeper a CNN, the better its performance. However, with the deepening of the network, problems related to vanishing gradients and exploding gradients occur. In order to solve these problems, He et al. proposed a deep residual network (ResNet) [20], which is more easily optimized than traditional CNNs and could provide a higher precision while increasing the depth. ResNet became the winner



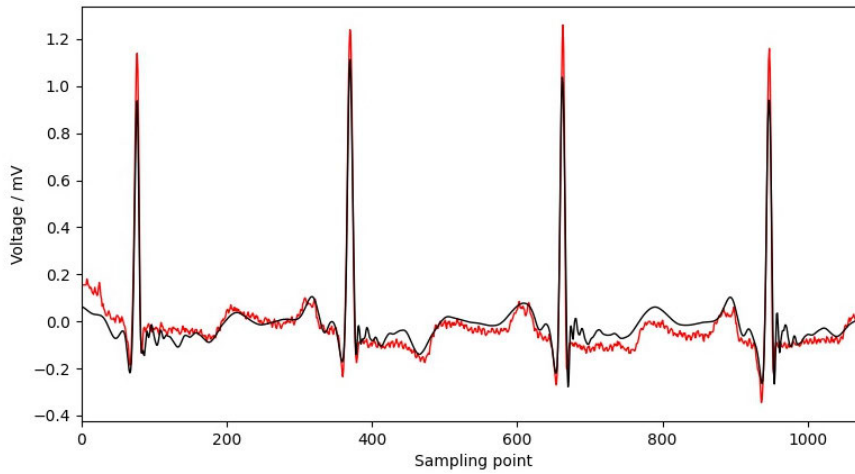


FIGURE 4. A sample ECG signal before denoising (red color) and after denoising (black color).

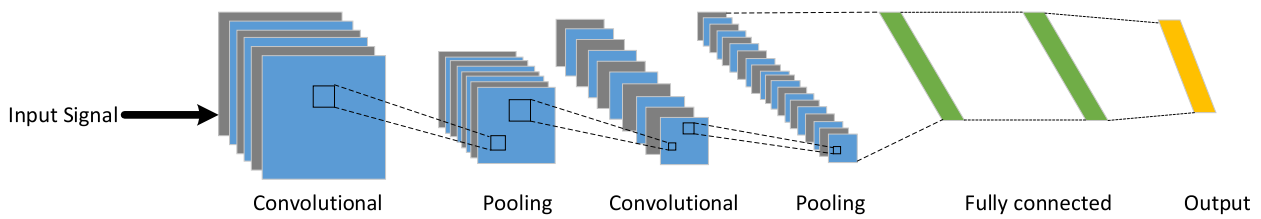


FIGURE 5. The basic structure of a convolutional neural network (CNN).

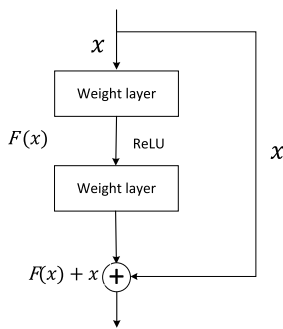


FIGURE 6. The ResNet building block.

of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2015 [21] in image classification, single-object localization, and object detection tasks, as well as the winner of the Microsoft Common Objects in Context (MS COCO) 2015 [22] detection and segmentation tasks (<https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8>).

In order to solve the degradation problem in deep neural networks, a residual structure was proposed in [20] as shown in Figure 6, producing the following output vector:

$$y = F(x, \{W_i\}) + x, \quad (3)$$

where  $x$  denotes the input vector and function  $F(x, \{W_i\})$  expresses the error mapping to be learned. The connections in Figure 6 neither introduce additional parameters nor increase the complexity of computation, but  $x$  and  $F$  must have the same dimensions. If the input/output dimensions need to be changed (e.g., changing the number of channels), the following formula can be used:

$$y = F(x, \{W_i\}) + W_s x, \quad (4)$$

where  $W_i$  denotes the weight matrix of the two-layer network in the ResNet block shown in Figure 6 and  $W_s$  is used to adjust the input dimension to the output dimension so that the matrix summation operations can be performed. For the residual structure, the learned feature is recorded as  $H(x)$ , and this module uses a kind of shortcut connection to allow it to directly learn the residual  $F(x) = H(x) - x$ , so as to output the target of  $F(x) + x$ . When the residual is 0, identify mapping is only made for the accumulation layer, ensuring at least that the network performance will not be degraded and the residual will not be 0; this also enables the accumulation layer to learn new features based on the input features, so as to achieve better performance.

### C. ResNeXt

In 2017, Xie et al. improved the ResNet model by proposing a combination of the ResNet and Inception models, called ResNeXt, which compares favorably with ResNet and is

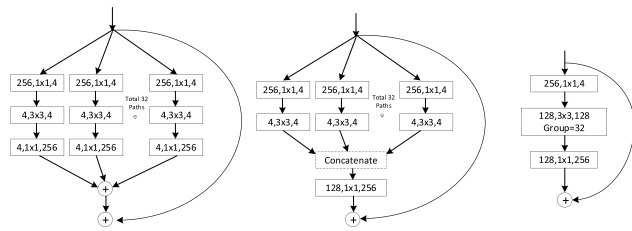


FIGURE 7. The ResNeXt equivalent building blocks (adapted from [5]).



FIGURE 8. The structure of ResNeXt (simplified view).

much simpler than all Inception models, as it requires considerably fewer hyperparameters to be set manually. ResNeXt is essentially grouped convolution, where the number of groups is controlled by the number of variable bases. Figure 7 shows the three equivalent building blocks of ResNeXt.

Figure 8 shows a simplified view of the ResNeXt structure, where the input information passes through a convolution, a batch normalization (BN), and an activation layer before entering a group convolution module, consisting of 32 branches with the same topology, whose outputs are combined and finally normalized, activated, and summed with the shortcut connection to obtain the output information.

#### IV. RELATED WORK

In 2012, AlexNet [23] reached a milestone by making the following contributions:

- 1) Using a graphics processing unit (GPU) for network training for the first time, which greatly accelerated the model training.
- 2) Using a Rectified Linear Unit (*ReLU*) non-linear activation function [24] instead of the traditional *Sigmoid* and *Tanh* activation functions.
- 3) Using a local response normalization (LRN).
- 4) Using dropout random deactivation neuron operations in the first two layers of the fully-connected layer to reduce overfitting.

This was the first use of deep learning (DL) in the form of a CNN, which paved the way for a new generation of machine learning (ML) models and techniques. Since then, ML has experienced revolutionary changes. DL models, such as CNNs, LSTMs [25] and their variants, almost fully dominate over the other ML models in all application fields. In the medical field, for instance, researchers use DL or ML models for classification of arrhythmia, in order to make its detection and prediction an easier and more reliable process [26], [27].

In terms of data processing, a method, where the ECG signals are treated as one-dimensional data, is increasingly well

accepted. In [28], a one-dimensional adaptive CNN model is proposed, in which two ECG modules are integrated into a single learning system for extraction and classification of features. In [29], a 9-layer deep CNN is proposed to classify one-dimensional ECG signals. In [30], following the stage of feature learning, a *SoftMax* layer is added to the top of the hidden layer obtained to form a deep neural network (DNN). After each iteration, the most relevant and uncertain heartbeats in the test records are marked and used to update the DNN weight, which significantly improves the detection and classification accuracy. A model, based on a one-dimensional CNN, is proposed in [31], where the data of the MIT-BIH database is divided into normal and abnormal heart activity, and a grid search method is used to find the hyperparameters of the CNN.

Recently, many powerful DL models have been produced to effectively classify ECG signals. In [32], a single-lead ECG signal extraction method based on wavelet transform is proposed, especially for positioning the end point of the T wave. In [33], a new automatic heartbeat classification method is proposed to effectively diagnose arrhythmia in a unsupervised medical environment. In [34], a 34-layer DNN based algorithm is developed to detect various cardiac rhythm disturbances with the single-lead ECG data generated by a sensing/monitoring equipment. The algorithm provides a higher accuracy rate than that of ordinary cardiologists, possibly due to the powerful feature learning ability of the DNN used. In [35], a new multi-scale wavelet CNN is proposed to automatically identify various cardiac rhythm disturbances, combined with an one-dimensional CNN and stationary wavelet transform. Distinguishable features are extracted from the ECG signals, along with their wavelet subbands, which greatly improves the feature learning process of the model at different scales. In [36], a parallel general regression neural network (GRNN) is used to classify heartbeats, and an online learning module is added to the GRNN to construct a personalized automatic classification model for patients. In [37], a multi-model DL integration model is proposed that can extract useful information from the input of multiple ECG waves. In [38], a symbolization model is proposed, specifically designed for ECG signals, based on a multi-perspective CNN (MPCNN). In [39], a new model for deep bidirectional LSTM network-based wavelet sequences, called DBLSTM-WS, is proposed for ECG signals classification. In [40], a 16-layer 1D CNN is designed for efficient and automatic classification of ECG signals, including AF. In [41], an end-to-end ECG signals classification algorithm, based on a novel segmentation strategy and 1D CNN, is proposed. In [42], a novel method is proposed, based on genetic algorithm back propagation neural network (GA-BPNN) for ECG signals classification with feature extraction using wavelet packet decomposition (WPD).

There are also many non-DL models that have achieved good results as well. In [43], a wavelet packet entropy (WPE) and random forest (RF) based ECG signal classification

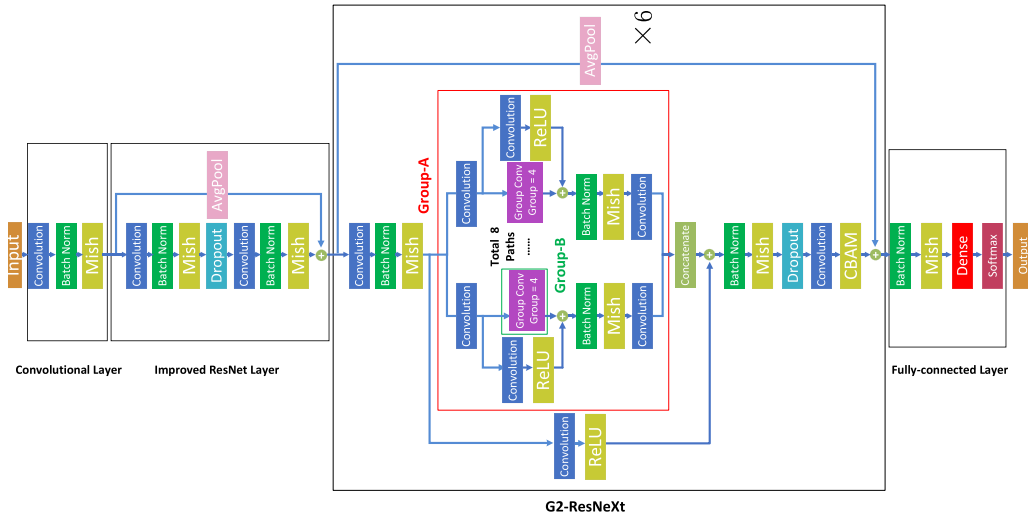


FIGURE 9. The structure of the proposed G2-ResNeXt model.

model is proposed. A RF model is used in [44] to select features using a filter method based on a mutual information ranking criterion, by reducing the number of features while improving the performance. An automatic ECG signals classification model, based on the combination of multiple support vector machines (SVM), is proposed in [45], showing a very good performance. In [46], a novel ensemble multi-label classification model is proposed, which combines multiple multi-label classification methods to build a high-performance classifier.

## V. PROPOSED MODEL: G2-ResNeXt

The classic ResNeXt structure is usually used for classification of two-dimensional images. As such, it is not quite suitable for ECG signals classification considered in this paper, so it has to be improved. The improvement proposed in this paper, called G2-ResNeXt, adopts a one-dimensional convolution kernel of size 32, while the original ResNeXt uses  $3 \times 3$  small convolution kernels. Such difference in sizes is mainly due to the different nature of the ECG signals and the image data. The resolution of the image input to the network is generally low and a  $3 \times 3$  area may also contain significant information even with a small receptive field. However, for low-frequency and low-sampling-rate ECG signals, the area of 3 sampling points at any position is difficult to constitute a waveform change with specific significance and is easily interfered by noise, thus badly effecting the feature learning. For solving this problem, the proposed G2-ResNeXt model uses a large convolution kernel.

In the conducted experiments, the ECG signals were sliced up with the sampling rate of 360 Hz, whereby the length of each slice was 1080 sampling points, resulting in a 3-sec slice. However, as the slice length of 1080 is reduced to 135 after three down sampling operations and cannot be down sampled again, all slices were re-sampled to 1024 sampling points in

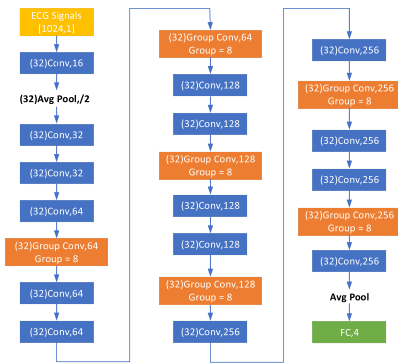


FIGURE 10. Layers' parameters of the proposed G2-ResNeXt model.

order to allow more down sampling operations. The minimum number of features was set to 16 and this value increases with passing through each convolutional layer.

The structure of the proposed G2-ResNeXt model is depicted on Figure 9, whereas Figure 10 shows the specific dimensions and values used in each layer of it.

For measuring the computational complexity of the proposed model, the number of parameters (Params) and floating-point operations (FLOPs) can be used, calculated as follows:

$$FLOPs = (2 * C_{in} * K^2 - 1) * H * W * C_{out}; \quad (5)$$

$$Params = C_{out} * (K^2 * C_{in} + 1), \quad (6)$$

where  $C_{in}$  denotes the number of input channels,  $C_{out}$  denotes the number of output channels,  $K$  denotes the size of the convolution kernel,  $H$  denotes the height of input features, and  $W$  denotes the width of input features. The number of FLOPs used by different G2-ResNeXt nodes is shown in



**TABLE 4.** The number of floating-point operations (FLOPs) used by G2-ResNeXt nodes (M – millions of FLOPs, K – thousands of FLOPs).

Node name	FLOPs
Square	8.72 M
Sum	8.72 M
Mul	4.37 M
Sqrt	3.04 K
Sub	1.52 K
Add	357

Table 4. The number of G2-ResNeXt parameters (Params) equals 4.37 million.

The data are first convoluted once and then passed through a BN layer for batch normalization, in order to solve the problem of changing the data distribution in the middle layer during training, prevent the gradient from disappearing or exploding, and speed up the model training. The normalized data are then fed into an activation layer that uses a *Mish* activation function, which proved to have more stable performance and higher precision than *ReLU* [47], and, in addition, can keep small negative values to stabilize the network gradient current. The data are then grouped and convoluted, which increases the accuracy without increasing the complexity, while also reducing the number of hyperparameters.

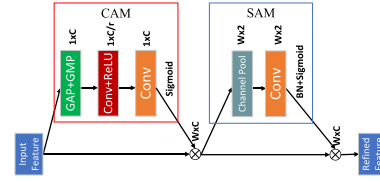
Unlike ResNeXt, the proposed G2-ResNeXt model utilizes a novel two-group structure, consisting of Group-A (the red box in Figure 9) and Group-B (the green box in Figure 9). Group-A consists of eight branches (with identical topologies). In each branch, the data pass through a convolution first and then through a Group-B and a parallel shortcut, which is down sampled using a combination of convolutional module and *ReLU*. Then, the output of the shortcut is summed with the output of Group-B, before undergoing normalization, activation, and another convolution at the end of the branch. The outputs of all eight branches of Group-A are fused at the end with features according to the following *concat* operation:

$$R_{concat} = \sum_{i=1}^z X_i * K_i + \sum_{i=1}^z Y_i * K_{i+z}, \quad (7)$$

where  $X_i$  and  $Y_i$  denote the two input channels,  $K_i$  represents parameters of the first set of input convolution kernels, and  $K_{i+z}$  represents parameters of the second set of input convolution kernels.

Given  $C_0$  input channels,  $C_1$  output channels, and a convolution kernel of size  $K$ , for conventional convolution, the computational effort (measured in FLOPs) can be calculated as:

$$FLOPs = K^2 * C_0 * C_1. \quad (8)$$

**FIGURE 11.** The modified CBAM used by the proposed G2-ResNeXt model.

For Group-A, the computational effort (measured in FLOPs) is:

$$FLOPs = K^2 * \frac{C_0}{G} * \frac{C_1}{G} * G = K^2 * \frac{C_0 * C_1}{G}, \quad (9)$$

where  $G$  denotes the number of branches in Group-A. From (9), it is clear that the computational effort in Group-A is reduced  $G$  times compared to the conventional convolution. This allows to speed up the model training.

Group-B is a group convolution with four identical topologies, with only one convolution operation in each branch. The outputs of all four branches are fused with the features.

During the model training, if there are too many parameters, the trained model will be easily overfitted. In order to reduce the number of parameters in the grid along with lowering the model training time, a dropout is added in the convolutional layer.

To further improve the model performance, a modified convolutional block attention module (CBAM) [48] is added at the end. A typical CBAM combines a channel attention module (CAM) and a spatial attention module (SAM). CAM utilizes different feature information, whereby each channel of the features represents a special detector. After adding the features together, an activation function gets their weights and figures out what features are meaningful. SAM is spliced with the channel descriptions, obtained through average pooling and maximum pooling, and processed by the convolutional layer to obtain the weight coefficients. CBAM is mostly applied to two-dimensional images, whereas the ECG data is processed as a one-dimensional time-series signal. Therefore, some modifications were made to the typical CBAM structure. More specifically, in CAM, both global mean pooling (GAP) and global maximum pooling (GMP) are performed as one-dimensional pooling. In SAM, the two-dimensional convolution kernel of size  $7 \times 7$  is replaced by a one-dimensional convolution kernel of size 7. The resultant modified CBAM (shown in Figure 11) is used in the proposed G2-ResNeXt model.

After the feature map is inputted into the modified CBAM, a channel attention is firstly performed, based on the width of the feature map for GAP and GMP. Then, the attention weight of the channel is obtained through multi-layer perceptron (MLP), followed by obtaining the normalized attention weight through a *Sigmoid* function. Finally, the channel attention is re-calibrated to the original features by multiplying the channel-by-channel weighting to the original input feature map and completing the channel attention to the original

features as follows:

$$\begin{aligned} M_c(F) &= \text{Sigmoid}(\text{MLP}(\text{AvgPool}(F))) \\ &\quad + \text{MLP}(\text{MaxPool}(F)) \\ &= \text{Sigmoid}\left(W_1\left(W_0\left(F_{avg}^c\right)\right) + W_1\left(W_0\left(F_{max}^c\right)\right)\right) \end{aligned} \quad (10)$$

where  $F_{avg}^c$  and  $F_{max}^c$  denote the average pooling feature and the maximum pooling feature, respectively. In order to obtain the attentional features in the spatial dimension, the feature map output of the channel attention is also globally max-pooled and globally averaged based on the width of the feature map, followed by reducing the feature map dimension after convolution with a convolution kernel of size 7 and a *ReLU* activation function, then raising it to the original dimension after a convolution, and finally merging the feature map normalized by a *Sigmoid* activation function with the feature map output of the channel attention. Finally, the *Sigmoid* activation function normalized feature map is combined with the channel attention output feature map to complete the rescaling of the feature map in both spatial and channel dimensions as follows:

$$\begin{aligned} M_s &= \text{Sigmoid}\left(f^7([\text{AvgPool}(F); \text{MaxPool}(F)])\right) \\ &= \text{Sigmoid}\left(f^7\left(\begin{bmatrix} F_{avg}^s \\ F_{max}^s \end{bmatrix}\right)\right). \end{aligned} \quad (11)$$

The proposed G2-ResNeXt model consists of four main parts, namely a convolutional layer, an improved ResNet layer, a G2-ResNeXt layer, and a fully-connected layer. The convolutional layer is mainly used to extract the features of the input data prepared for the next layer.

In the second layer, the improved ResNet, the data is convoluted twice, whereby a *Mish* activation function and a dropout are added between the two convolutions. In addition, just before the convolution, the data are inputted into an average-pool module, the samples are divided into feature regions, and the mean value in the area is used as the region representative, which simplifies the computation and reduces the number of parameters. Finally, the data outputted from the average pooling are added to the data (with the same dimension) produced by the *Mish* activation function, aiming to inherit the optimization effect of the previous step, so that the model could be further converged.

The next layer, G2-ResNeXt, is used to increase the speed of the model convergence and maintain stability. After passing through the G2-ResNeXt structure six times, the data are inputted to the fully-connected layer, where these are mapped to a one-dimensional vector, regressed by a *SoftMax* function [49]. Finally, all outputs are added together and normalized in order to show the multi-classification results in the form of probability.

The loss function used is the focal loss function, which allows to handle various data imbalances present in the data

set [50], as per the following formula:

$$\begin{aligned} FL &= -\alpha_1 * (1 - p_1)^\gamma * y_1 * \log(p_1) \\ &\quad - \alpha_2 * (1 - p_2)^\gamma * y_2 * \log(p_2), \dots, \\ &\quad - \alpha_n * (1 - p_n)^\gamma * y_n * \log(p_n), \end{aligned} \quad (12)$$

where  $\gamma$  is a parameter with a value within the range of [5, 0] (when  $\gamma = 0$ ,  $FL$  is the same as the common cross entropy loss function),  $\alpha$  and  $y$  denote the input data and labels, respectively, and  $p_n$  represents the probability value of the  $n^{\text{th}}$  class of the *SoftMax* output. The  $(1 - p_n)^\gamma$  parts reduce the loss contribution of the easy samples and increase the loss proportion of the hard samples. Because the label is in the form of One-Hot encoding, the value in the label of a certain sample type is 1 in the corresponding position, and the rest are 0. The optimized formula is shown below:

$$FL = -\alpha_c(1 - p_c)^\gamma \log(p_c), \quad (13)$$

where  $\alpha_c$  denotes the weight of the class- $c$  sample and  $p_c$  denotes the probability value of the class- $c$  output produced by *SoftMax*.

In order to speed up the convergence and limit the overfitting phenomenon, a L2 regularization is added to the convolutional layer, as follows [51]:

$$\text{Loss} = \|W^T x - p\|^2 + \alpha \|W\|^2, \quad (14)$$

where  $\alpha$  denotes the regular term coefficient,  $W$  denotes the network weight,  $p$  denotes the predicted value of the heartbeat category,  $x$  denotes the heartbeat feature, and  $T$  denotes the number of weighted items.

## VI. EXPERIMENTS AND RESULTS

### A. EVALUATION METRICS

Typical evaluation metrics, used in multi-class classification problems, were utilized for comparing the performance of different models, namely *sensitivity* ( $Se$ ), *precision* ( $P+$ ), and *overall accuracy* ( $Acc$ ).

*Sensitivity* (also called *recall*) is used to measure the proportion of the positive samples that are correctly identified as such by a classification model to the total number of positive samples in a dataset, as follows:

$$Se = \frac{TP}{TP + FN} \times 100\%, \quad (15)$$

where  $TP$  (true positive count) denotes the number of positive samples that are correctly identified as positive, and  $FN$  (false negative count) denotes the number of positive samples that are incorrectly identified as negative. In our case, *sensitivity* expresses the percentage of the actual heartbeats correctly classified by a model, which reflects its ability to discover heartbeats of the classes considered.

*Precision* is used to measure the proportion of the positive samples that are correctly identified as such by a classification model, to the total number of samples identified as positive, as follows:

$$P+ = \frac{TP}{TP + FP} \times 100\%, \quad (16)$$

**TABLE 5.** Comparison of G2-ResNeXt results achieved in five different training sessions.

Training Session's Model	Acc (%)	AAMI heartbeat class N		AAMI heartbeat class V		AAMI heartbeat class S		AAMI heartbeat class F	
		Se (%)	P+ (%)	Se (%)	P+ (%)	Se (%)	P+ (%)	Se (%)	P+ (%)
Training_1	96.16	98.34	97.24	97.09	95.90	80.59	82.26	4.05	24.27
Training_2	95.83	98.39	96.91	95.33	95.69	78.60	81.67	2.27	13.33
Training_3	96.22	98.43	97.26	97.63	95.40	78.63	89.71	4.86	11.11
Training_4	95.75	97.85	97.27	96.99	95.34	80.30	79.01	5.19	16.84
Training_5	96.29	98.60	97.19	96.62	95.16	79.89	89.52	6.16	20.54

**TABLE 6.** Performance comparison of optimizers used in the experiments.

Optimizer	Acc (%)	AAMI heartbeat class N		AAMI heartbeat class V		AAMI heartbeat class S		AAMI heartbeat class F	
		Se (%)	P+ (%)	Se (%)	P+ (%)	Se (%)	P+ (%)	Se (%)	P+ (%)
G2-ResNeXt + Adagrad	95.87	98.02	97.47	96.34	95.16	80.33	82.81	11.35	19.07
G2-ResNeXt + Adam	90.86	91.93	97.25	93.60	83.66	84.07	52.52	14.26	21.57
G2-ResNeXt + Adadelta	94.80	96.76	97.61	96.48	93.00	80.18	84.69	4.21	3.07
G2-ResNeXt + Adamax	89.43	96.75	91.40	85.72	95.52	5.69	11.22	1.78	4.45
G2-ResNeXt + SGD	96.16	98.34	97.24	97.09	95.90	80.59	82.26	4.05	24.27

**TABLE 7.** The multi-class confusion matrix of the proposed G2-ResNeXt model (Training\_1).

Actual number of AAMI heartbeats per class		Predicted numbers of AAMI heartbeats per class			
		N	V	S	F
N	42097	41401	241	436	19
V	7453	127	7236	33	57
S	2705	489	34	2180	2
F	617	557	34	1	25

where *FP* (false positive count) denotes the number of negative samples that are incorrectly identified as positive. Generally, *precision* expresses the positive predictive value (PPV) of a model. In our case, it is more important to improve (as much as possible) *sensitivity* than *precision* of the classification model, because diagnostics of a possible disease is more important than not discovering it at all.

In addition to *sensitivity* and *precision*, another evaluation metric, used in the conducted experiments, is the *overall accuracy* (*Acc*) which indicates the overall classification accuracy of a model. It is calculated by dividing the number of all correctly classified heartbeats to the total number of heartbeats present, as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (17)$$

where *TN* (true negative count) denotes the number of negative samples that are correctly identified as negative.

## B. EXPERIMENTS

In the experiments, the initial learning rate was set to 0.1 and then dynamically adjusted during the training process. When the loss has been no longer reduced, the learning rate was reduced to half of the original, which allowed the network to converge faster in the right direction. As for the batch size, we first chose a larger batch size to fill up the memory, then observed the convergence of the loss, and reduced the batch size if it did not converge, or if the convergence was not good. Finally we chose a batch size of 128.

The initialization of the weights is also a very important parameter. Correct weight initialization can promote fast convergence of the model. As the combination of 'ReLU + Conv' is used in many places in the proposed G2-ResNeXt model, the He initialization method was utilized as it works well for ReLU [52].

Conducted experiments showed that the use of the stochastic gradient descent (SGD) optimizer for tuning the parameters of the G2-ResNeXt model produces the best results compared to other optimizers used, i.e., Adagrad, Adam, Adadelta, and Adamax. This is reported in Table 6.

In the experiments, the proposed G2-ResNeXt model was trained according to Algorithm 1. After setting the learning rate, number of epochs, and batch size, the model is initialized with the weights and then the epoch's loop starts updating the weights, based on the input signal data (signalsData) and signal labels (signalsTags), calculating the loss value, and updating the value of the learning rate when the next round's loss is greater than the previous round's loss.

**TABLE 8.** Performance comparison of the proposed G2-ResNeXt model (Training\_1) to other state-of-the-art models, in terms of *overall accuracy (Acc)*, and *sensitivity (Se)* and *precision (P+)* in classifying AAMI heartbeats of class V or S.

Model	Acc (%)	AAMI heartbeat class V		AAMI heartbeat class S	
		Se (%)	P+ (%)	Se (%)	P+ (%)
WPE and RF [43]	94.61	94.20	89.78	20.00	0.16
RF [44]	96.14	87.36	94.62	78.55	68.29
GRNN [36]	94.81	92.13	88.16	88.58	88.78
Multi-Model [37]	95.81	93.91	94.55	65.51	66.19
Ensemble SVM [45]	94.47	94.75	93.79	78.10	49.75
MPCNN [38]	95.36	87.89	83.63	48.12	75.30
<b>G2-ResNeXt</b>	<b>96.16</b>	<b>97.09</b>	<b>95.90</b>	<b>80.59</b>	<b>82.26</b>

**Algorithm 1** G2-ResNeXt Training Process**Input:** preprocessed ECG signals;**Output:** classified heartbeats

```

1: Initialize the learning rate,  $\alpha$  total number of max iteration
    $epoch$ , previous round error  $e_p$ .  $Total\ Batch$ 
2: Generate random weights  $\theta$  of the G2-ResNeXt
3: Init G2-ResNeXt model ( $\theta$ )
4: while  $iteration < epoch$  do
5:   if  $iteration > 1$  then
6:      $e_p = error$ 
7:   end if
8:   Initialize  $error = 0$ 
9:   for  $batch = 1$  to  $Total\ Batch$  do
10:     $\theta = \text{G2-ResNeXt model.train}(\text{signalsData}, \text{signalsTags})$ 
11:    Update  $\theta$ 
12:     $error = error + J(\theta)$ 
13:   end for
14:   if  $e_p < error$  then
15:      $\alpha = \alpha * f$ 
16:   end if
17:    $iteration = iteration + 1$ 
18: end while

```

In order to probe stability of the proposed model, it was trained five different times, one after another, with 80 iterations in each training session. The *overall accuracy*, and *sensitivity* and *precision*, achieved by the proposed model in each training session (Training\_1–Training\_5) for each AAMI heartbeats class, are shown in Table 5. As can be seen from the table, the *overall accuracy* fluctuates slightly (within 1% only) between different training sessions, and the mean square error (MSE) is equal to 0.2434. The model with the most balanced performance (corresponding to the median value of the *overall accuracy* achieved in the five training sessions), named Training\_1, was chosen for participation in the performance comparison with the state-of-the-art models, presented in the next subsection.

**C. RESULTS**

Table 7 shows the multi-class confusion matrix of the proposed G2-ResNeXt model (Training\_1) applied to the MIT-BIH arrhythmia database.

Table 8 presents the results of the performance comparison of the proposed model to the state-of-the-art models considered,<sup>1</sup> as regards the *overall accuracy*, and *sensitivity* and *precision* achieved in classifying AAMI heartbeats (results for classes V and S are presented only, as these contain most arrhythmias). The presented results clearly demonstrate the superiority of the proposed G2-ResNeXt model, according to all evaluation metrics used, except the GRNN model which achieves better *sensitivity* and *precision* in classifying AAMI heartbeats of class S. More specifically, the G2-ResNeXt superiority over the other models compared ranges: from 0.02% to 1.69% based on *overall accuracy*, from 2.34% to 9.73% and from 1.28% to 12.27% based on *sensitivity* and *precision*, respectively achieved in classifying AAMI heartbeats of class V, and from 2.04% to 60.59% and from 6.96% to 82.10% according to *sensitivity* and *precision*, respectively achieved in classifying AAMI heartbeats of class S.

**VII. CONCLUSION**

This paper has proposed a novel model, called G2-ResNeXt, for the classification of inter-patient ECG signals with *overall accuracy* of 96.16%. A slice-and-stack method is used to process the MIT-BIH data set by achieving an effective data preprocessing. The presented experimental results have clearly demonstrated that the proposed G2-ResNeXt model could effectively identify arrhythmia, by achieving *sensitivity* and *precision* values of 97.09% and 95.90% for the ventricular ectopic heartbeats (VEB), and 80.59% and 82.26% for the supraventricular ectopic heartbeats (SVEB), respectively, thus surpassing all state-of-the-art models used for performance comparison, based on all evaluation metrics used.

<sup>1</sup>The results of the state-of-the-art models considered are taken from the corresponding literature sources, as the code of these models is not open source and therefore the results cannot be reproduced.



Therefore, the proposed model has great clinical application prospect.

Although the overall accuracy of the model is high, there is still room for further improvements, aiming at improving the recognition rate of class-F heartbeats, and reducing the increased computational complexity and execution time (due to the introduction of a modified CBAM module) as to be competitive in that sense to the original ResNeXt model.

## REFERENCES

- [1] World Health Organization. (Dec. 9, 2020). *The Top 10 Causes of Death*. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>
- [2] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A survey on ECG analysis," *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, May 2018.
- [3] E. Izci, M. A. Ozdemir, R. Sadighzadeh, and A. Akan, "Arrhythmia detection on ECG signals by using empirical mode decomposition," in *Proc. Med. Technol. Nat. Congr.*, Nov. 2018, pp. 1–4.
- [4] S. Karimifard and A. Ahmadian, "Morphological heart arrhythmia classification using Hermitian model of higher-order statistics," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 3132–3135.
- [5] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.
- [6] R. Mark, "AAMI-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms," in *Proc. Assoc. Advance. Med. Instrum., Arrhythmia Monit. Subcommittee. AAMI ECAR*, 1987, pp. 1–51.
- [7] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May/Jun. 2001.
- [8] R. E. Hermes, D. B. Geselowitz, and G. C. Oliver, "Development, distribution, and use of the American heart association database for ventricular arrhythmia detector evaluation," in *Proc. Comput. Cardiol.*, 1980, pp. 263–266.
- [9] A. Taddei, G. Distanti, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography," *Eur. Heart J.*, vol. 13, no. 9, pp. 1164–1172, Sep. 1992.
- [10] M. Sraïti, Y. Jabrane, and A. Atlas, "An overview on intra- and inter-patient paradigm for ECG heartbeat arrhythmia classification," in *Proc. Int. Conf. Digit. Age Technol. Adv. Sustain. Develop. (ICDATA)*, Jun. 2021, pp. 1–7.
- [11] C. Han, R. Huang, F. Yu, X. Huang, and L. Cui, "EasiECG: A novel inter-patient arrhythmia classification method using ECG waves," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 1–8.
- [12] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ECG heartbeat classification with temporal VCG optimized by PSO," *Sci. Rep.*, vol. 7, no. 1, pp. 1–11, Sep. 2017.
- [13] M. Mitrokhin, A. Kuzmin, N. Mitrokhina, S. Zakharov, and M. Rovnyagin, "Deep learning approach for QRS wave detection in ECG monitoring," in *Proc. IEEE 11th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Sep. 2017, pp. 1–3.
- [14] Z. Gao, J. Wu, J. Zhou, W. Jiang, and L. Feng, "Design of ECG signal acquisition and processing system," in *Proc. Int. Conf. Biomed. Eng. Biotechnol.*, May 2012, pp. 762–764.
- [15] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.
- [16] A. R. Gómez and Á. Jiménez-Casas, "Analysis of the ECG signal recognizing the QRS complex and P and T waves, using wavelet transform," *Amer. J. Eng. Res.*, vol. 7, pp. 51–59, Jan. 2018.
- [17] S. Shimazaki, H. Kawanaka, H. Ishikawa, K. Inoue, and K. Oguri, "Cuffless blood pressure estimation from only the waveform of photoplethysmography using CNN," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5042–5045.
- [18] S.-H. Kim and G.-T. Han, "1D CNN based human respiration pattern recognition using ultra wideband radar," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIIC)*, Feb. 2019, pp. 411–414.
- [19] Y. Zhao, J. Cheng, P. Zhan, and X. Peng, "ECG classification using deep CNN improved by wavelet transform," *Comput., Mater. Continua*, vol. 64, no. 3, pp. 1615–1628, 2020.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [22] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2014, pp. 740–755.
- [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.
- [24] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [26] S. Nurmaini, R. U. Partan, W. Caesarendra, T. Dewi, M. N. Rahmatullah, A. Darmawahyuni, V. Bhayyu, and F. Firdaus, "An automated ECG beat classification system using deep neural networks with an unsupervised feature extraction technique," *Appl. Sci.*, vol. 9, no. 14, p. 2921, Jul. 2019.
- [27] S. Naik, S. Debnath, and V. Justin, "A review of arrhythmia classification with artificial intelligence techniques: Deep vs machine learning," in *Proc. 2nd Int. Conf. Emerg. Technol. (INCET)*, May 2021, pp. 1–14.
- [28] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2015.
- [29] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Comput. Biol. Med.*, vol. 89, pp. 389–396, Oct. 2017.
- [30] M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, pp. 340–354, Sep. 2016.
- [31] N. Adam, D. Valko, and M. Havrilla, "Using neural networks for ECG classification," in *Proc. IEEE 20th Jubilee World Symp. Appl. Mach. Intell. Informat. (SAMI)*, Mar. 2022, pp. 367–372.
- [32] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna, "A wavelet-based ECG delineator: Evaluation on standard databases," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 4, pp. 570–581, Apr. 2004.
- [33] U. Satija, B. Ramkumar, and M. S. Manikandan, "A new automated signal quality-aware ECG beat classification method for unsupervised ECG diagnosis environments," *IEEE Sensors J.*, vol. 19, no. 1, pp. 277–286, Jan. 2019.
- [34] Y. H. Awni, P. Rajpurkar, M. Haghpahani, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, pp. 65–69, Jan. 2019.
- [35] L. E. Bouny, M. Khalil, and A. Adib, "ECG heartbeat classification based on multi-scale wavelet convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3212–3216.
- [36] P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-S. Zhou, T. Li, and J.-S. Li, "High-performance personalized heartbeat classification model for long-term ECG signal," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 78–86, Jan. 2017.
- [37] E. Essa and X. Xie, "An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification," *IEEE Access*, vol. 9, pp. 103452–103464, 2021.
- [38] J. Niu, Y. Tang, Z. Sun, and W. Zhang, "Inter-patient ECG classification with symbolic representations and multi-perspective convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 5, pp. 1321–1332, May 2020.
- [39] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification," *Comput. Biol. Med.*, vol. 96, pp. 189–202, May 2018.



- [40] Z. Xiong, M. Stiles, and J. Zhao, "Robust ECG signal classification for the detection of atrial fibrillation using novel neural networks," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2017, pp. 1–4.
- [41] X. Hua, J. Han, C. Zhao, H. Tang, Z. He, Q. Chen, S. Tang, J. Tang, and W. Zhou, "A novel method for ECG signal classification via one-dimensional convolutional neural network," *Multimedia Syst.*, vol. 2020, pp. 1–13, Nov. 2020.
- [42] H. Li, D. Yuan, X. Ma, D. Cui, and L. Cao, "Genetic algorithm for the optimization of features and neural networks in ECG signals classification," *Sci. Rep.*, vol. 7, p. 41011, Jan. 2017.
- [43] T. Li and M. Zhou, "ECG classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, pp. 285–1–285–16, Aug. 2016.
- [44] J. F. Saenz-Cogollo and M. Agelli, "Investigating feature selection and random forests for inter-patient heartbeat classification," *Algorithms*, vol. 13, no. 4, p. 75, Mar. 2020.
- [45] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers," *Biomed. Signal Process. Control*, vol. 47, pp. 41–48, Jan. 2019.
- [46] Z. Sun, C. Wang, Y. Zhao, and C. Yan, "Multi-label ECG signal classification based on ensemble classifier," *IEEE Access*, vol. 8, pp. 117986–117996, 2020.
- [47] D. Misra, "Mish: A self regularized non-monotonic neural activation function," 2019, *arXiv:1908.08681*.
- [48] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.
- [49] F. Wang, J. Cheng, W. Liu, and H. Liu, "Additive margin softmax for face verification," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 926–930, Jul. 2018.
- [50] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [51] F. Kamalov and H. H. Leung, "Deep learning regularization in imbalanced data," in *Proc. Int. Conf. Commun., Comput., Cybersecur., Informat. (CCCI)*, Nov. 2020, pp. 1–5.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.



**SHENGNAN HAO** received the B.S. degree from the North China University of Science and Technology, China, in 1996, and the M.S. degree from the Beijing University of Technology, China, in 2009. She joined the North China University of Science and Technology, in 1996, and became an Associate Professor, in 2009. Her current research interests include complex systems, impulsive systems, and stochastic control.



**HANG XU** was born in 1998. He received the B.S. degree from the North China University of Science and Technology, in 2020, where he is currently pursuing the master's degree. His research interests include machine vision and algorithm research.



**HONGYU JI** is currently pursuing the MBiomed-SCI degree with the School of Biosciences, The University of Sheffield. Her research interests include medical dataset analysis, age-related and noise-induced hearing loss, mechanobiology, electronic health records processing, and pharmaceutical analysis.



**ZHIWU WANG** received the Ph.D. degree from Tianjin Medical University, in 2014. He is currently a Chief Physician with the Department of Radiotherapy and Chemotherapy, Tangshan People's Hospital. He is engaged in the comprehensive medical treatment of lung cancer and digestive system tumors. His current research interests include medical data processing and screening tumor immunotherapy effect prediction markers based on artificial intelligence.



**LI ZHAO** received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 1997 and 2002, respectively. He is currently an Associate Professor with the Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University. His current research interests include mobile computing, the Internet of Things (IoT), e-health systems, intelligent transportation systems (ITS), home networking, machine learning, and digital multimedia.



**ZHANLIN JI** (Member, IEEE) received the M.Eng. degree from Dublin City University, in 2006, and the Ph.D. degree from the University of Limerick, Ireland, in 2010. He is currently a Professor with the North China University of Science and Technology, China, and a Researcher with the Telecommunications Research Centre (TRC), University of Limerick. He has authored/coauthored more than 100 research papers in refereed journals and conferences. His research interests include ubiquitous consumer wireless world (UCWW), the Internet of Things (IoT), cloud computing, big data management, and data mining.



**IVAN GANCHEV** (Senior Member, IEEE) received the Engineering and Ph.D. degrees (summa cum laude) from the Saint Petersburg University of Telecommunications, in 1989 and 1995, respectively. He is currently an International Telecommunications Union (ITU-T) Invited Expert and an Institution of Engineering and Technology (IET) Invited Lecturer associated with the University of Limerick, Ireland, the University of Plovdiv "Paisii Hilendarski," Bulgaria, and the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria. He was involved in more than 40 international and national research projects. He has served on the TPC of more than 370 prestigious international conferences/symposia/workshops and has authored/coauthored one monograph, three textbooks, four edited books, and more than 300 research papers in refereed international journals, books, and conference proceedings. He is on the editorial board and has served as the guest editor for multiple prestigious international journals.

...