GFDLECG: PAC Classification for ECG Signals Using Gradient Features and Deep Learning

Hashim Abu-gellban, Long Nguyen, Fang Jin
Department of Computer Science, Texas Tech University
{hashim.gellban, long.nguyen, fang.jin}@ttu.edu

***Abstract*—ECG signal classification is a popular topic in healthcare for arrhythmia detection. Recently, ECG signal analysis using supervised learning has been investigated with the goal to help physicians to automatically identify the Premature Atrial Complex (PAC) heartbeats. PAC may be a sign of underlying heart conditions, which may change to supraventricular tachycardia that increases the possibility of sudden death. In this paper, we propose a data-driven approach, GFDLECG, which is based on ECG behavior to detect abnormal beats. We extract further features from ECG using the gradient feature generation algorithm. We also build the classification model by utilizing the Gated Recurrent Unit (GRU) and the residual fully convolutional networks with GRU to learn long-short term patterns of ECG behaviors.**

***Index Terms*—Bioinformatics, Artificial Neural Networks/ Deep Learning, ECG Classification, Feature Generation/ Transformation, Premature Atrial Complex (PAC).**

I. Introduction

Electrocardiogram (ECG) presents the electrical mobility of the heart. The ECG signal is essential for the arrhythmia disease identification. The disease is an irregular heartbeat that is too fast or too slow. The Premature Atrial Complex (PAC) is a kind of heart arrhythmia which is also called the supraventricular premature beat (S type) [1]. PAC is the activation of atria over a pathway other than the sinus node, that is fairly common pervasive in adults who have or do not have heart disease [2], [3]. The common symptoms of PAC are palpitations and missing beats. PAC may be triggered by caffeine, alcohol, abnormal levels of magnesium in the blood, and stress. Furthermore, PAC may be a sign of the underlying heart conditions. In severe conditions, it may change to atrial fibrillation or supraventricular tachycardia that can cause sudden cardiac death.

Using adhesive ECG electrodes is common in care and patient monitoring [4]–[8]. Recently, considerable research focused on providing anomaly detection techniques to assist cardiologists with diagnosing ECG signals. Therefore, the objectives of PAC classification algorithms are: developing an automated detection approach with high performance to increase the productivity of healthcare professionals and to save more lives by early PAC detection, especially for lonely elderly people who stay at home without help. Whereas, the automatic abnormal heart classification from ECG sequences is a challenging task because of imbalance data and other reasons related to the ECG signals (e.g., biomedical contamination, external noise, time-varying dynamics, morphological characteristics) [9].

The performance of PAC classification using the ECG dataset [6] has been a critical issue in earlier research as a result of these challenges in the ECG signals. Wang et al. [7] extracted Shapelet features from multivariate time series (MTS) and using the boosting algorithm to combine weak classifiers into a single classifier, which gave 81% accuracy. Karim et al. [10] developed multivariate Long Short-Term Memory (LSTM)/ Attention LSTM (ALSTM) and Fully Convolutional Networks (FCN) to learn long-short term behavior from the raw ECG data without extracting any new feature, resulting in 86% F1-score. Furthermore, Schäfer et al. [11] created a new algorithm called (WEASEL + MUSE) to extract features from MTS. After that, the Logistic Regression (LR) classifier was executed, which produced 88% accuracy. However, some important features may not be produced by this algorithm because of the complex multi-phase of filtering and selections flows. The previous works did not take advantage of combining good feature extraction techniques and proper neural network architectures for the given problem to mitigate the impact of the ECG signals' issues to foster the performance of their approaches. In general, deep learning approaches are more appropriate than the traditional machine learning classifiers (e.g, LR) in solving high dimensional time series detection problems.

To address the issues related to the nature of ECG signals and imbalanced datasets, we proposed GFDLECG (Gradient Feature and Deep Learning approach for ECG classification). GFDLECG employs a gradient filter and a deep learning based model to detect abnormal (PAC) heartbeats of a human subject from their ECG signal behaviors with high performance. It applies an efficient feature generation method called the Gradient Feature Generation (GFG) algorithm, to extract further features from ECG. Our developed approach also applies a novel neural network architecture called Multivariate Gated Recurrent Unit and Residual Fully Convolutional GRU Networks (MGRU-ResFCNGRU) to identify PAC ECG heartbeat from the original ECG multivariate sequences as well as from the gradient ECG signals. More specifically, our main contributions in this paper are:

- We are one of the first in employing the Gated Recurrent Unit (GRU) and the Residual Fully Convolutional Networks with GRU (ResFCNGRU) in a multivariate time series. First, we utilize fast training in GRU while

keeping the ability to learn temporal behaviors as in LSTM (Long Short-Term Memory). Then, the FCN component is employed as a latent feature extractor for our proposed classification model. We have also added GRU and residual components to FCN, to enhance the overall model performance.
- We proposed the gradient feature generation (GFG) algorithm for given ECG signals in our framework to generate additional important features based on the gradient algorithm. Extracting the features by computing the slopes of electrodes is effective to increase the performance of the model, as the gradient of signals provides the neural networks with the amounts of the signals' changes within a fixed time interval.
- Extensive experiments were conducted to present the capabilities of our approach in discovering PAC from ECG sequences with high performance. The results show that the ECG pattern based on the normal/ abnormal identification with GFDLECG is promising and outperforms the previous approaches. Several other neural networks and feature generation methods were applied, to show the effectiveness of our proposed approach.

II. Problem Formulation

Here, we present our multivariate time series classification problem as follows:

Multivariate time series (MTS): MTS of M attributes are sequences of simultaneous observations. Let $X = [X^1, X^2, ..., X^M]$ be the set of the variables representing M time series and $X^m \in \mathbb{R}^T$. Each time series is called univariate time series (UTS) denoted by $X^m = [x_1, x_2, ..., x_T]$, where $x_t \in \mathbb{R}$ is the t^{th} element in the time series.

The input: The input MTS^i of the model consists of M variables performed by C class labels, where $i \in \mathbb{Z}_n$ and n is the number of examples. We can denote this as pairs: (X, y_i), where $y_i \in \{0, 1\}$ representing a class label. Zero is the normal class label and one is the abnormal class label.

Problem definition: Given the input MTS^i which consists of M features $[X^1, X^2, ..., X^M]$, we have to find a function f that classifies the input as normal or PAC heartbeat, as follows:

$$y = f(X^1, X^2, ..., X^M) \tag{1}$$

III. Related Work

In this Section, we focus on the previous work of the PAC feature generation/ selection and classification.

*a) **Feature Generation and Selection from Multivariate Time Series of ECG:*** Werth et al. [12] employed Lomb-Scargle algorithm [13] to generate the frequency spectrum. The WEASEL + MUSE algorithm was employed [11] to extract features from ECG. [14] applied the Daubechies wavelet 6 filters algorithm [15] to remove noise in the ECG time series, and detected the R-peak features using Pan-Tompkins algorithm [16]. Moreover, Li et al. [17] proposed using different techniques to extract features, such as: generating statistical features, independent component analysis, Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Continuous Wavelet Transform (CWT). Gutiérrez-Gnecchi et al. [18] preprocessed the raw data and extracted both the P and T ECG waves by the Mallat filter bank, the wavelet transformer, and wave detection algorithms. However, discovering P-wave is a complex task [19]. [20] preprocessed ECG using the Wavelet method to decrease the noise, and extracted features of median beat and 8 characteristics points using R-peaks and RandomWalk algorithms. Finally, Walinjkar et al. [21] used and extracted the human subject's age and instantaneous heartbeat, the ECG's amplitude/ WABP, and RR interval from the ECG dataset.

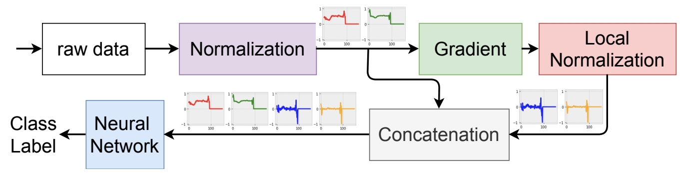

Fig. 1. Data processing pipeline for MTS classification.

*b) **Traditional Classification Algorithms using Logistic Regression, Decision Tree, and KNN:*** Schäfer et al. [11] used the logistic regression classifier to identify the binary class label. Li et al. [17] applied a random forest algorithm to classify heartbeats. Walinjkar et al. [21] applied the bagged tree and the weighted KNN algorithms to detect 4 class labels including PAC, after dropping normal ECG examples from the training data.

*c) **Deep Learning for Time Series Classification:*** Kachuee et al. [5] applied 5 residual convolutional blocks followed by 2 dense layers to classify the ECG lead II dataset [8]. Werth et al. [12] used ResNet and ResNext architectures employing GRU and BiGRU to learn complex non-linear patterns in ECG sequences. Acharya et al. [14] applied augmentation and CNN to identify five classes (N, S, V, F, Q). Karim et al. [10] employed (MLSTM-FCN and MALSTM-FCN) to classify PAC from ECG. Probabilistic backpropagation Neural Network (PNN) was employed to detect 8 different ECG class labels, where PAC had a low performance with 76.82% accuracy [18]. Xia et al. [20] concatenated 3 deep learning network architectures which consist of different layers (e.g., CNN, BiRNN, and Dense) and attention with context blocks, resulting in F1-measure 89.7% for the PAC class label.

IV. ECG Binary Classification Framework

A. Dataset Preprocessing

We normalize each example to be treated fairly during building the classification model and the evaluation process. The values of the raw data are normalized between [0, 1] by scaling each time series, using the following formula:

$$x_t^{'} = \frac{x_t - min_m}{max_m - min_m} \tag{2}$$

where the maximum and minimum of the m^{th} feature are max_m and min_m for all time series in the dataset. After that, we shuffle the examples in the dataset since shuffling

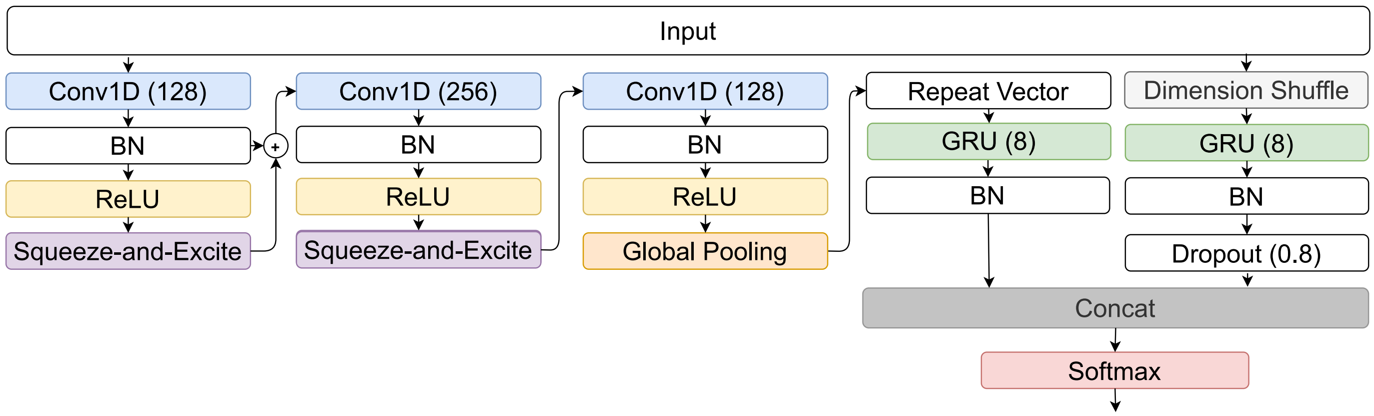

Fig. 2. The architecture of GFDLECG. The width of layer (number of nodes) and percentages of dropout are shown between parenthesis.

is an essential process for the training and testing datasets to represent the entire dataset.

B. Gradient Feature Generation (GFG)

We calculate the gradient for each time series (Lead0, Lead1) using the second-order finite central differences of the interior elements of the time series (X^m) [22]–[24]. We also use the second-order forward and backward differences for the first and last points, respectively. The length of the calculated gradient is the same as the length of the original time series. To compute the gradient, we assume that X^m has 3 continuous derivatives or more ($X^m \in C^3$). Let h_s and h_d be a heterogeneous step size. We need to minimize the consistency error (η_t). η_t is the differences between the gradient's actual value and estimate value from the adjacent points.

$$\eta_t = X_t^{m(1)} - [\alpha x_t + \beta x_{t+h_d} + \gamma x_{t-h_s}] \tag{3}$$

We compute the following linear system by using Taylor series expansion instead of x_{t+h_d} and x_{t-h_s}:

$$\begin{cases} \alpha + \beta + \gamma = 0 \\ \beta h_d - \gamma h_s = 1 \\ \beta h_d^2 + \gamma h_s^2 = 0 \end{cases} \tag{4}$$

Therefore, the estimation of $X_t^{m(1)}$ is calculated using the following formula:

$$\hat{X}_t^{m(1)} = \frac{h_s^2 x_{t+h_d} + (h_d^2 - h_s^2)x_t - h_d^2 x_{t-h_s}}{h_d h_s (h_d + h_s)} + \mathcal{O}\left(\frac{h_d h_s^2 + h_d^2 h_s}{h_d + h_s}\right) \tag{5}$$

In our experiments, we have used a homogeneous step size ($h_d = h_s$). Therefore, the second-order of $\hat{X}_t^{m(1)}$ is computed as the following:

$$\hat{X}_t^{m(1)} = \frac{x_{t+h} - x_{t-h}}{2h} + \mathcal{O}(h^2) \tag{6}$$

We perform a local normalization for every generated gradient lead feature, which has enhanced the performance of our classification model better than the Z-normalization and the Min-Max normalization. We scale each time series individually where the values can be at most 1 and at least -1.

Algorithm 1 describes the steps in more detail. $| TS' |$ is the absolute value of each element in TS'. $\oplus$ is the operator to concatenate the original MTS which contains lead0 and lead1 signals with their gradients. GFG_MTS contains 4 sequences that are sent to NN to identify the class label. Fig. 1 illustrates the steps of preprocessing from raw data to the input data of NN to detect the PAC class label.

C. The Proposed Neural Network

Fig. 2 shows our proposed classification architecture using neural networks. We first explain the GRU layer, and then the residual FCN with GRU. Finally, we combine all neural networks.

1) Gated Recurrent Units Network (GRU): LSTM suffers from the vanishing gradient problem [25], [26]. GRU solves the problem by using the update gate which learns from the previous long short-term pattern, to anticipate the future sequence. We also employ GRU since it is faster to learn than LSTM and it discovers reliances of numerous time scales to enhance the performance of the model [27], [28]. The update formulas of GRU are:

$$\begin{aligned} z_t^j &= \sigma(W_z x_t + U_z h_{t-1}^j) \\ r_t^j &= \sigma(W_r x_t + U_r h_{t-1}^j) \\ \tilde{h}_t^j &= tanh(W x_t + U(r_t^j \odot h_{t-1}^j)) \\ h_t^j &= (1 - z_t^j)h_{t-1}^j + z_t^j \tilde{h}_t^j \end{aligned} \tag{7}$$

where W and U are the weight matrices. z_t^j, r_t^j, $\tilde{h}_t^j$, and h_t^j present the update gate, the reset gate, the candidate activation, and the activation units at time t, respectively.

2) Residual Fully Convolutional Networks (ResFCN) with GRU: We use three convolution layers to build the Fully Convolutional Networks (FCN), since FCN is famous NN in

Algorithm 1: The gradient feature generation algorithm.

function gradient_feature_generation (MTS)

Input : $MTS = X \in \mathbb{R}^{M\times T}$.

Output: $GFG\_MTS = X^{'} \in \mathbb{R}^{M'\times T}$, where $M' = 2 * M$.

$GFG\_MTS \leftarrow copy(MTS)$

foreach $TS \in MTS$ **do**

    /* Compute TS′ (the slope of TS) */

    $TS' \leftarrow copy(TS)$

    foreach $x'_t \in TS'$ **do**

        /* $x_{t+h}, x_t, x_{t-h} \in TS$ */

        if $t = 1$ **then**

            $x'_t \leftarrow \frac{x_{t+h}-x_t}{h}$

        else

            if $t = T$ **then**

                $x'_t \leftarrow \frac{x_t-x_{t-h}}{h}$

            else

                $x'_t \leftarrow \frac{x_{t+h}-x_{t-h}}{2h}$

            end

        end

    end

    /* Scale TS′ */

    $max_{TS'} \leftarrow max(|\, TS' \,|)$

    foreach $x'_t \in TS'$ **do**

        $x'_t \leftarrow \frac{x'_t}{max_{TS'}}$

    end

    $GFG\_MTS \leftarrow GFG\_MTS \oplus TS'$

end

return GFG_MTS

finding the semantic segmentation of images [29]. The basic block of convolution:

$$
\begin{aligned}
y &= W \otimes x + b \\
s &= BN(y) \\
h &= ReLU(s)
\end{aligned}
\tag{8}
$$

where $\otimes$ is the convolution operator. BN is important to decrease the building time and provide the generalization to NN. ReLU is the Rectified Linear Unit. Squeeze-and-excite is used to extract interdependencies between feature channels [30]. The residual operation (h^{res}) is a shortcut connection to combine the outputs (h^s and h^{se}) before and after the first squeeze-and-excite block, where "+" is the residual operator. In our experiments, we have found that this shortcut improves the accuracy of the PAC classification model.

$$h^{res} = h^s + h^{se} \tag{9}$$

The repeat vector layer is applied to convert the dimensionality of the data from 2D to 3D, to be adequate for GRU layer.

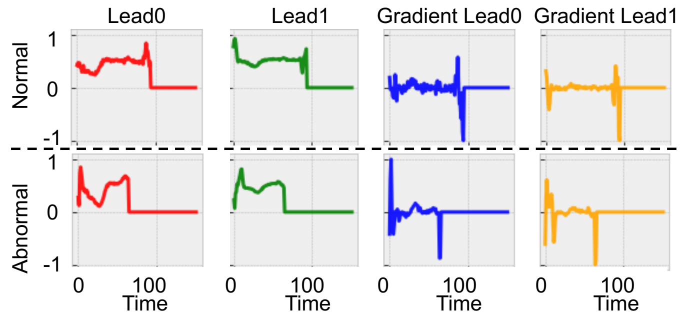

Fig. 3. Examples of normal and abnormal heartbeats sequences including the new gradient lead features.

GRU is also employed after the global pooling (h^{gp}) to learn the long short-term pattern, which increases the efficiency of our proposed model.

$$h^r = repeat\_vector(h^{gp}) \tag{10}$$

3) Combining GRU and ResFCNGRU: Fig. 2 shows the concatenation between ResFCNGRU and GRU (after the dropout) neural networks, as follows:

$$
\begin{aligned}
h^c &= h^b \oplus h^d \\
y &= softmax(h^c)
\end{aligned}
\tag{11}
$$

where $\oplus$ is the concatenation operator, and h^c is the output of concatenating the left ResFCNGRU neural network (h^b) and the rightmost GRU with the dropout layer (h^d). The class label is predicted by the $softmax$ activation function.

V. Experiments and Results

We start by describing the dataset as in Section V-A. In all experiments, we set $h = 1$ for the GFG algorithm. Section V-B shows that the performance of our methodology outperforms the results of other previous approaches. Sections V-C and V-D brief the advantages of using GRU layer in MGRU-resFCNGRU architecture. At the end, we show the importance of adding the gradient features in the preprocessing phase of ECG signals in enhancing the performance.

A. Dataset Description

The ECG dataset has 200 examples (133 normal and 67 abnormal) [6]. Each example is a heartbeat and contains two heart electrodes (Lead0, Lead1) sequences. The abnormal class label represents cardiac pathology (supraventricular premature beat) known as PAC or S type. Time series is a sequence of values ordered by time (t) (equally gaps ECG records). ECG signals have variable lengths between 39 and 152. We unify the length for all time series to be the maximum length ($T = 152$), where empty values are replaced by zeroes. We split the dataset approximately into training (50%) and testing (50%) per a class label.

Fig. 3 shows an example for each class label with different original time series lengths. The gradient sequences represent the new features extracted from Lead0 and Lead1 using GFG. We can see that the two class labels have clear differences for

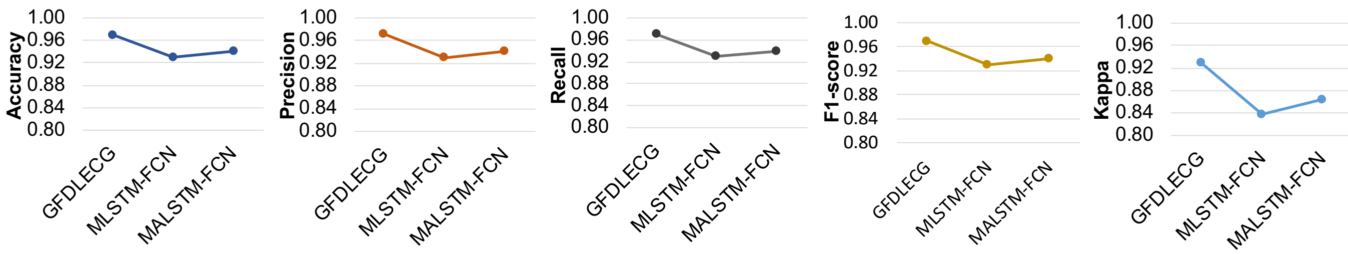

Fig. 4. Performance metrics for GFDLECG (i.e., using the MGRU-ResFCNGRU architecture) and other neural networks using the same gradient method employed in GFG. The y axes are the metrics and x axes are the neural networks for building the models.

the gradient leads. The gradient lead0 of the normal heartbeat has at the end of the original time series (i.e., before adding zeros to have a fixed length) a short crest and a large trough, while the PAC heartbeat contains one high crest at the start of time series and a large trough at the end. Additionally, the gradient lead1 of the normal heartbeat consists of a small crest and a large trough at the end of the original signal. Whereas, the gradient lead1 has a deep trough without any crest at the end of the original abnormal heartbeat.

TABLE I
PERFORMANCE SUMMARY OF OUR PROPOSED APPROACH COMPARED WITH OTHER ALGORITHMS FROM PREVIOUS RESEARCH USING THE SAME DATASET. THE BOLDFACE REPRESENTS THE BEST PERFORMANCE.

Algorithm	Accuracy
Boosting [7]	0.81
WEASEL + MUSE [11]	0.88
MLSTM-FCN [10]	0.86
MALSTM-FCN [10]	0.86
GFDLECG	**0.97**

B. Overall Performance Summary

Table I shows the comparative results of our proposed algorithm against the other algorithms [7], [10], [11] from the previous research using the same dataset [6]. Our approach using the gradient feature generation method and the MGRU-ResFCNGRU architecture outperforms the previous methods with (97% accuracy and 97% F1-score). Furthermore, the F1-score of the MDDNN method [31] is 88% while our approach performs better with 9% more.

C. The Effect of GRU in GFDLECG

We ran the preprocessed dataset including the gradient leads features on the same neural network architecture MGRU-ResFCNGRU (GFDLECG) and we also replaced GRU with different layers (BiLSTM [32], LSTM [33], ALSTM [10]). The purpose of this experiment was to manifest the effectiveness of GRU in our proposed architecture. GRU was the best layer with F1-score (97%) as shown in Fig. 5. We can see that BiLSTM (96% F1-score) performed better than LSTM (94% F1-score) and ALSTM (95% F1-score).

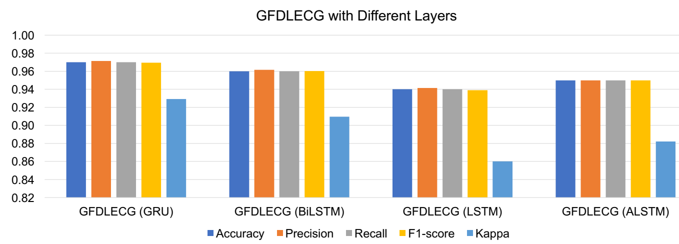

Fig. 5. Performance metrics by replacing GRU in GFDLECG with other layers (BiLSTM, LSTM, MALSTM). The best layer for GFDLECG is GRU which is employed in our neural network architecture for GFDLECG.

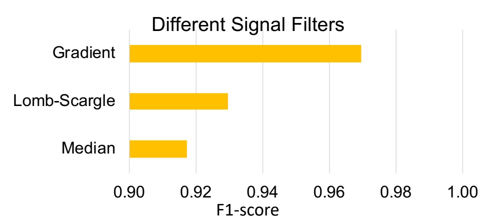

Fig. 6. F1-score of the classification models using different signal filter algorithms. We have also replaced the gradient algorithm in GFG with the Lomb-Scargle and the median algorithms. All preprocessed data have been sent to the MGRU-ResFCNGRU neural network to build the classification model.

D. GFDLECG vs Other Neural Networks Using Gradient Filter

To show the effectiveness of the new gradient features in our neural network architecture (MGRU-ResFCNGRU) and some other neural networks (MLSTM-FCN and MALSTM-FCN [10]). Using GFG, the accuracy of MLSTM-FCN and MALSTM-FCN has been increased from 86% to 93% and 94%, respectively. The results are shown in Fig. 4.

E. The Effect of Gradient Feature

We generated different signal filtering methods (Gradient, Lomb-Scargle, Median) to compare the gradient feature generation with other features. Next, we built our classification model using MGRU-ResFCNGRU for all different filtering signals and raw data. Fig. 6 illustrates the increase of performance using the gradient features.

VI. Conclusion

In this research, we present a new method for ECG classification to discover PAC heartbeats based on the gradient signal processing and the neural networks. We combine the original signals with the new gradient features using our proposed algorithm called the Gradient Feature Generation (GFG). We have trained the preprocessed ECG signals using a new neural network architecture (MGRU-ResFCNGRU) which takes advantage of the GRU layers, the CNN layers, and the residual operation to foster the performance. GFDLECG outperforms other approaches from previous research using the same dataset. Moreover, the results show that the gradient feature and the neural network architecture are more effective than the other methodologies conducted in the experiments. For future work, we will apply this approach to other multivariate time series problems, to study its effectiveness to enhance the performance.

References

[1] A. S. Manolis, H. Calkins, and B. C. Downey, "Supraventricular premature beats." 2015.

[2] "Cardiac health - providing information, risk assessment and advice about your heart health, 2011." http://www.cardiachealth.org/, accessed: February 24, 2020.

[3] Y. Wang, M. M. Scheinman, W. W. Chien, T. J. Cohen, M. D. Lesh, and J. C. Griffin, "Patients with supraventricular tachycardia presenting with aborted sudden death: incidence, mechanism and long-term follow-up," *Journal of the American College of Cardiology*, vol. 18, no. 7, pp. 1711–1719, 1991.

[4] J. Werth, X. Long, E. Zwartkruis-Pelgrim, H. Niemarkt, W. Chen, R. M. Aarts, and P. Andriessen, "Unobtrusive assessment of neonatal sleep state based on heart rate variability retrieved from electrocardiography used for regular patient monitoring." vol. 113, pp. 104–113, 2017.

[5] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation." in *IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2018, pp. 443–444.

[6] "Bobski's World." http://www.cs.cmu.edu/ bobski/, accessed: January 8, 2020.

[7] H. Wang and J. Wu, "Boosting for real-time multivariate time series classification." in *Thirty-First AAAI Conference on Artificial Intelligence*. AAAI, 2017.

[8] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals." vol. 101, no. 23, pp. e215–e220, 2000.

[9] R. Hoekema, G. J. Uijen, and A. V. Oosterom, "Geometrical aspects of the interindividual variability of multilead ECG recordings." in *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 5. IEEE, 2001, pp. 551–559.

[10] F. Karim, S. Majumdar, H. Darabi, and S. Harford, "Multivariate LSTM-FCNs for time series classification." in *Neural Networks 116*. Elsevier, 2019.

[11] P. Schäfer and U. Leser, "Multivariate time series classification with WEASEL+ MUSE." in *arXiv preprint*, vol. 1711, no. 11343. arXiv, 2017.

[12] J. Werth, M. Radha, P. Andriessen, R. M. Aarts, and X. Long, "Deep learning approach for ecg-based automatic sleep state classification in preterm infants," *Biomedical Signal Processing and Control*, vol. 56, p. 101663, 2020.

[13] T. Ruf, "The lomb-scargle periodogram in biological rhythm research: analysis of incomplete and unequally spaced time-series," *Biological Rhythm Research*, vol. 30, no. 2, pp. 178–201, 1999.

[14] 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." in *Computers in biology and medicine*, vol. 89, 2017, pp. 389–396.

[15] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ecg signal denoising," *Digital signal processing*, vol. 16, no. 3, pp. 275–287, 2006.

[16] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[17] T. Li and M. Zhou, "Ecg classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, no. 8, p. 285, 2016.

[18] J. A. Gutiérrez-Gnecchi, R. Morfin-Magana, D. Lorias-Espinoza, A. del Carmen Tellez-Anguiano, E. Reyes-Archundia, A. Méndez-Patiño, and R. Castañeda-Miranda, "Dsp-based arrhythmia classification using wavelet transform and probabilistic neural network," *Biomedical Signal Processing and Control*, vol. 32, pp. 44–56, 2017.

[19] M. Potse, T. A. Lankveld, S. Zeemering, P. C. Dagnelie, C. D. Stehouwer, R. M. Henry, A. C. Linnenbank, N. H. Kuijpers, and U. Schotten, "P-wave complexity in normal subjects and computer models," *Journal of electrocardiology*, vol. 49, no. 4, pp. 545–553, 2016.

[20] Z. Xia, Z. Sang, Y. Guo, W. Ji, C. Han, Y. Chen, S. Yang, and L. Men, "Automatic multi-label classification in 12-lead ECGs using neural networks and characteristic points." in *Machine Learning and Medical Engineering for Cardiovascular Health and Intravascular Imaging and Computer Assisted Stenting*. Springer, Cham, 2019, pp. 80–87.

[21] A. Walinjkar and J. Woods, "ECG classification and prognostic approach towards personalized healthcare." in *2017 International Conference On Social Media, Wearable And Web Analytics (Social Media)*. IEEE, 2017, pp. 1–8.

[22] B. Fornberg, "Generation of finite difference formulas on arbitrarily spaced grids." *Mathematics of computation*, vol. 41, no. 184, 1988.

[23] D. R. Durran, *Numerical methods for wave equations in geophysical fluid dynamics.* Springer Science Business Media, 2013, vol. 32.

[24] A. Quarteroni, R. Sacco, and F. Saleri, "Foundations of matrix analysis," in *Numerical Mathematics*. Springer, 2007, pp. 1–32.

[25] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[26] H. Abu-gellban, L. H. Nguyen, and F. Jin, "LiveDI: An Anti-theft Model Based on Driving Behavior." in *2020 ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'20), https://doi.org/10.1145/3369412.3395069 (in press)*, 2020.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling." in *arXiv preprint arXiv:1412.3555*. arXiv, 2014.

[28] L. H. Nguyen, Z. Pan, O. Openiyi, H. Abu-gellban, M. Moghadasi, and F. Jin, "Self-boosted time-series forecasting with multi-task and multi-view learning." in *arXiv preprint arXiv:1909.08181*, 2019.

[29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2015, pp. 3431–3440.

[30] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks." in *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2015, pp. 7132–7141.

[31] H.-S. Huang, C.-L. Liu, and V. S. Tseng, "Multivariate time series early classification using multi-domain deep neural network." in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2018, pp. 90–98.

[32] Y. Zhang, T. Xu, and Y. Dai, "Research on chatbots for open domain: Using BiLSTM and sequence to sequence." in *the 2019 International Conference on Artificial Intelligence and Computer Science*. ACM, 2019, pp. 145–149.

[33] F. Xiong, K. Zou, Z. Liu, and H. Wang, "Predicting learning status in MOOCs using LSTM." in *the ACM Turing Celebration Conference-China*. ACM, 2019, pp. 1–5.