A Comparative Analysis of 2D Deep CNN Models for Arrhythmia Detection Using STFT-Based Long Duration ECG Spectrogram

Tabassum Islam Toma and Sunwoong Choi
Department of Electronics Engineering, Kookmin University, Seoul 02707, Korea
Email: tabassum2485@kookmin.ac.kr, schoi@kookmin.ac.kr

***Abstract*—Cardiac arrhythmia detection exploiting electrocardiogram (ECG) signal has gradually become mature in the recent decade. Recent advancement in the deep learning (DL) area has accelerated its extensive application in this area of healthcare research. Especially, convolutional neural network (CNN) based architecture has become more popular and has drawn a lot of research attention for the precise detection of cardiac arrhythmia utilizing single or few QRS complexes or beats of ECG signals. In this article, we have investigated the performance of six CNN based classifiers for cardiac arrhythmia (15 classes) detection based on the spectrogram of the long duration ECG signals. Six pre-trained state-of-the-art CNN based models, i.e. VGG-16, ResNet-50, Inception, MobileNet, DenseNet, and EfficientNet have been used as feature extractors. At the beginning of this investigation, short term fourier transform has been used to extract time-frequency domain information's from the long duration ECG signals as well as generate two dimensional spectrogram images to be fed into these CNN based architectures as input. However, the overall performances of the six ECG rhythm classifiers in terms of accuracy, precision, recall, and F1 score, are also evaluated and investigated explicitly to demonstrate the comparative analysis. In addition, the experiment has been conducted varying the learning rate during training of the classifiers to study the impact of the learning rates on discriminative feature learning.**

***Index Terms*—Cardiac arrhythmia, electrocardiogram (ECG), CNN, STFT, transfer learning.**

I. Introduction

Cardiac arrhythmia has been considered as one of the important and outrageous manifestation of cardiovascular diseases (CVD) which causes majority of cardiac arrests and sudden deaths to human across the world [1]. It refers to heart rhythm disorders which obstruct the origin and physiological diffusion of the electrical stimulus of the heart. Because of adopting the sedentary lifestyle, the occurrence and mortality rate of the SVD is still growing among large set of population specially in developing countries [2]. As a result, it has become very important to monitor heart rhythm regularly in order to manage and prevent the CVDs. Exploiting electrocardiogram (ECG) signal is very useful for the diagnosis of arrhythmia since it is a non-invasive and easy-to-apply method to measure the cardiac activity [3].

ECG (electrocardiogram) is an important and the most widely disseminated medical tool that records the can provide useful information on cardiac excitability, transmission, and recovery. Therefore, automatic diagnosis of cardiovascular disease significantly relies on the precise detection of the irregular heart rhythms from ECG signals [4]. The conventional method of ECG signals rhythm classification includes morphological features extraction of single or few QRS complexes or beat to detect rhythm. But, the current method may fail to achieve satisfactory diagnostic performance during when trained on multi-class ECG data [5]. Hence, diagnosing CVD using long-duration ECG signal fragments is an alternative to the conventional method though it is very challenging.

In the past decade, numerous machine learning (ML) techniques, such as neural network (NN) [6], SVM [7], decision tree [8], logistic regression [9], linear discriminant [10], neuro-fuzzy system [11], K-nearest neighbors (KNN) classification method [12], have been utilized in ECG signal classification. But the performance of ML based classifier largely relies on noise reduction during data pre-processing and discriminative spatiotemporal feature extraction [13]. In recent years, deep learning (DL) algorithms have proven their efficacy in wide range of applications, such as image processing, pattern recognition, computer vision and thus inspired researchers to replace conventional ML methods. In addition, because of the capability of learning discriminative features automatically, DL based model can reduce the complexity in various applications [14]. Recently, many researchers applied several deep learning techniques to the study of ECG classification. Studies manifest that the widely exploited DL based model is convolutional neural network (CNN). Kiranyaz proposed 1D convolutional neural networks for real-time patient-specific ECG classification [15]. The proposed network can effectively classify five typical beats from the ECG records. In order to realize the classification of 5 typical types of arrhythmia signals, i.e., normal, left bundle branch block, right bundle branch block, atrial premature contraction and ventricular premature contraction, Li also proposed 1D-CNN based method in [16]. An ECG monitoring system integrating the Impulse Radio Ultra Wideband (IR-UWB) radar with the CNN have been studied in [17]. Jun proposed an effective ECG arrhythmia classification method using a deep two-dimensional (2D) convolutional neural network considering ECG signal segment as 2D image. In order to identify and classify four types of ECG

TABLE I
ILLUSTRATION OF THE ECG DATA USED IN THIS EXPERIMENT.

Arrhythmia type	*Symbol*	*Record*	*Training samples*	*Test samples*
Atrial bigeminy	AB	222	6	2
Atrial fibrillation	AFIB	201-203,210, 217,219,221, 222	632	159
Atrial flutter	AFL	202,203,222	61	16
Ventricular bigeminy	B	106,119,200	196	49
2° heart block	BII	231	55	14
Idioventricular	IVR	124,207	10	3
Normal sinus	N	101-106,108, 109,111-119, 121-124, 205 200-202, 219 207-209, 220 212-215,222 223,228,230, 231,233,234	5048	1263
Nodal (A-V junctional) rhythm	NOD	124,201,222	21	6
Paced rhythm	P	102,104,107, 217	527	132
Pre-excitation	PREX	230	59	15
Sinus bradycardia	SBR	232	144	36
Supraventricular tachyarrhythmia	SVTA	207,209,220, 234	14	4
Ventricular trigeminy	T	106,119,124, 201,208,214, 221,223	86	22
Ventricular flutter	VFL	207	11	3
Ventricular tachycardia	VT	200,203,205, 223,233	13	4

patterns, Salem proposed an ECG arrhythmia classification method in [18] using transfer learning from 2D deep CNN features. However, all these studies focused on developing ECG arrhythmia classifier based on beat levels. A very few studies have been conducted based on rhythm label taking long duration ECG signals under consideration. In this work, we have broadly investigated the comparative performances of five deep 2D CNN models for cardiac arrhythmia (15 classes) detection based on long-duration electrocardiography (ECG) signal analysis. Since, ECG is one-dimensional time domain signal, short-time Fourier transform has been applied to transform into time-frequency spectrograms and use it as 2D input of the deep CNN models at the very beginning. We have utilized VGG-16, ResNet50, InceptionNet, MobileNet, Efficient Net as the backbone of five CNN based models. But, instead of training these models from scratch using ECG data, these model were pre-trained on image database and used as feature extractor in our experiment. since our dataset is not comparatively large, the learning of these models after being pre-trained on huge database associated with image classification and object recognition can be transferred for ECG classification purposes. This is very well known method for image classification purpose and it also known as transfer learning.

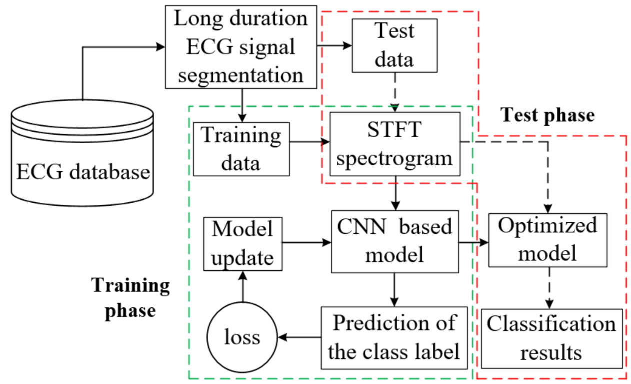

Fig. 1. The architecture of the ECG based arrhythmia detection using STFT based long duration ECG spectrogram.

The rest of this article is organized as follows. In section II, our experimental methodology has been demonstrated broadly. section III illustrates the experimental results with detail discussion. Finally, we draw conclusion of this article in section IV.

II. METHODOLOGY

A. Overview of Method

The architecture of the ECG based arrhythmia detection using STFT based long duration ECG spectrogram is depicted in Fig. 1. This method doesn't require signal filtering, hand-crafted feature extraction from the signal, and feature selection at any stage. The ECG data are obtained from MIT-BIH arrhythmia database [19]. The input ECG signals are segmented into 10 seconds duration recordings and annotated based on the the recordings annotations. Later on, the short time Fourier transform (STFT) is utilized to transform each 10 seconds ECG signal into 2D time-frequency spectrogram and the size of ECG spectrogram images are $256 \times 256 \times 1$. Exploiting these obtained ECG spectrograms, the CNN based classifier learns spatial features at multiple scaling representations and get optimized iteratively. In the Table 1, we have shown the various cardiac arrhythmia diagnostic classes, the associated number of ECG signal fragments collected, and their distributions into training and test sets.

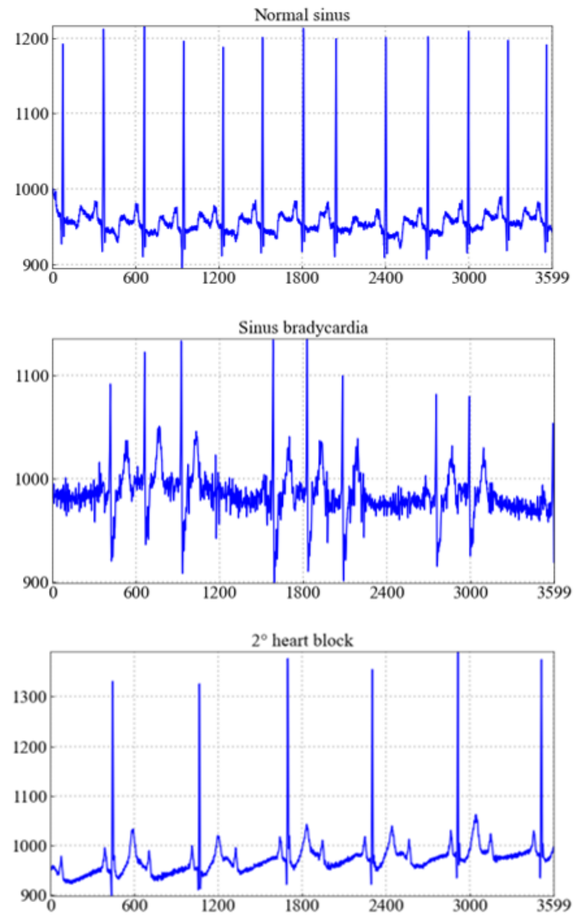

Fig. 2. Typical long duration ECG samples of different classes.

B. Data Acquisition and Selection

This subsection provides comprehensive demonstration of ECG data acquisition and selection methods. The original MIT-BIH Arrhythmia Database hosted by PhysioNet contains 48 records studied by the BIH Arrhythmia Laboratory and each of them is slightly over 30 minutes long [19]. The sampling rate of each ECG recordings are set to 360 Hz. Two cardiologists independently worked to annotate them in 15 rhythm labels. since, we are focused on long duration ECG signal in this study, for each type of rhythm we segmented 10 second long ECG record which contains 3600 samples. Fig. 2 depicts the typical ECG signal samples of class normal sinus, sinus bradycardia, 2° heart block. The number records used in this experiment for each class are given Table I. Besides, the number of training and testing samples used in this experiment are also shown in the table.

C. ECG Data Pre-processing

As our classifier is 2D CNN based model, the input of the classifier must be an 2D image. Therefore, both the training and testing ECG data samples represented in time domain are transformed into 2D time-frequency spectrograms using STFT.

TABLE II
PERFORMANCE MEASURES COMPARISON OF THE USED PRE-TRAINED 2D CNN MODEL.

Pre-trained CNN model	*Parameters*	*Top-5 accuracy*	*Time (ms) per inference step (GPU)*
VGG-16	138.4M	90.1%	4.2
ResNet50	25.6M	92.1%	4.6
Inception	23.9M	93.7%	6.9
MobileNet	4.3M	90.1%	3.4
DenseNet	8.1M	92.3%	5.4
EfficientNet	5.3M	93.3%	4.9

Before applying STFT transformation, the data samples are normalized between 0 to 1.

However, learning feature variation from a ECG signal becomes very difficult without analyzing the frequency domain properties. To extract frequency domain behavior of a signal, classical Fourier transform (FT) was used in many previous studies. But, the FT method assumes the signal as stationary, i.e. the signal has no time domain properties [20]. But, it observed from the experiment that the ECG signal is a non-stationary signal and its frequency also varies according to the time. The short time Fourier transform (STFT) is a well-known and widely used linear operator which was developed to overcomes this drawback by considering an analysis window that has a specific time-frequency resolution property. Thu, the STFT can explore both the instantaneous frequency behavior as well as the instantaneous amplitude behavior of a signal at the same time [21].

In the STFT, a window function is utilized for extracting time domain information. The window function has a given interval and the value of this window function outside the interval is zero. This window function is shifted over the whole non-stationary signal sequentially and every time it is multiplied with the signal to calculate frequency domain information. However, for a discretized digital signal, the time-frequency spectrogram of a non-stationary signal can be calculated as

$$STFT\left\{x[n]\right\} = \sum_{-\infty}^{\infty} x[n]g[n-\tau]e^{-jwn} \tag{1}$$

where $x[n]$ represents the discretized digital signal and in our experiment $x[n]$ the ECG signal which sampling rate was 360 Hz. $g[n]$ is the window function and τ is the shifting parameter. There are several windows available for STFT operation but we used "Hann" window in generating 2D spectrogram image. Besides, the window length is chosen as 128 and the value of τ is 14.

D. Architecture of the CNN based ECG Arrhythmia Classifier

In this subsection, the detail of the CNN based ECG arrhythmia classifier is described. We have used state-of-the-art pre-trained CNN model to extract discriminative features from the input 2D spectrogram. When a 2D spectrogram

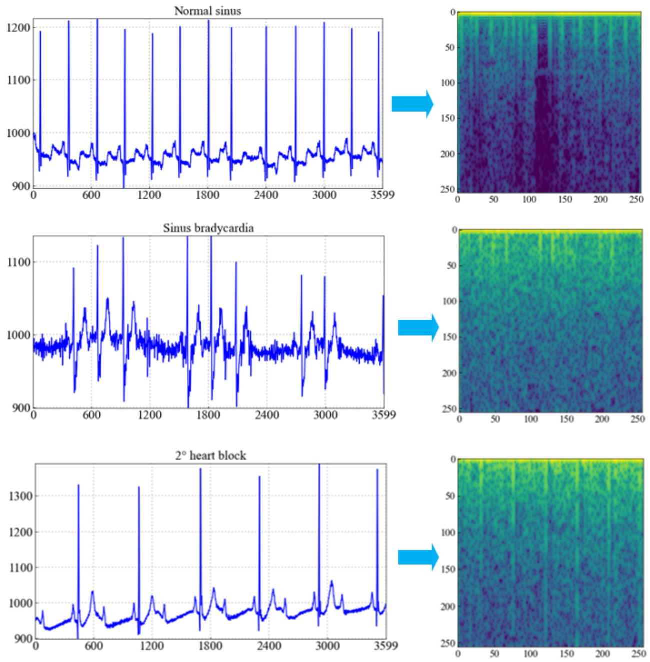

Fig. 3. Spectrograms of long duration ECG samples belonging to different classes.

TABLE III
HYPER-PARAMETERS OF THE CNN BASED CLASSIFIERS TRAINING.

Name of the hyper-parameter	*Values*
Shape of the training data	(6889, 256, 256, 1)
Shape of the testing data	(1723, 256, 256, 1)
Epoch	60
Optimizer	Adam
Batch size	64
Loss	Categorical cross-entropy
Learning rate	.001, .005, .0001

image is given as input, a Conv2D layer is used between the input and the pre-trained CNN model. Because the pre-trained CNN model can only take 3 channel input, but the given input channel is 1. Therefore to produce 3 channel input, three convolutional kernels of size (3×3) are applied in this first convolutional layer. Afterwards, pre-trained CNN model extracts features from the out of the first convolutional layer. In our experiment, we have utilized five state-of-the-art CNN models, i.e. VGG16, ResNet50, Inception, MobileNet, DenseNet, EfficientNet, which was previously trained on a large database "imagenet" in order to extract spatial discriminative features and we named the classifiers as Model I, Model

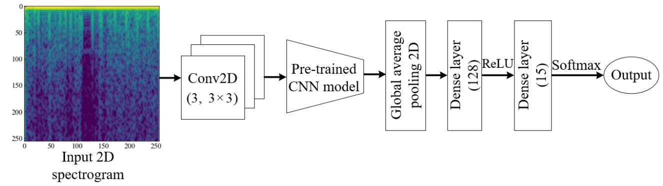

Fig. 4. The architecture of the 2D CNN based ECG arrhythmia classifier.

II, Model III, Model IV, Model V, and Model VI, respectively. Table II presents the overview of the five CNN models. However, after extraction of the features, global average pooling layer is used to down sample the detection of features in feature maps. Then two dense layers are applied along with two activation function ReLU, softmax, respectively.

III. EXPERIMENTAL RESULTS

A. Model Training and Optimization

For this experiment, we trained and optimized six CNN based classifiers architecture. This subsection elucidates the overall training and optimization details of six CNN based classifiers. Based on the basic hyper-parameters shown in Table III, the six ECG heart rhythm classifier are trained and optimized. However, the raw ECG data are segmented into 10s signals and afterwards, the spectrograms of size (256×256) are produced using the STFT. The whole dataset is divided into training and testing sets by the ratios of 80% and 20%, respectively. The number of training and testing samples of our experiment are 6889 and 1723. "Categorical cross-entropy" function is used to measure th loss between actual label and predicted label. To optimize the CNN based classifier, "Adam" optimizer is applied with the different learning rates .001, .005, and .0001. We have applied 60 iterations to converge the classifier models. In addition, the batch size is set to 64. All the training and testing programs have been performed in anaconda python 3.7 on a system equipped with 3.80 GHz CPU, 256 GB RAM, and a single NVIDIA Quadro RTX 6000 GPU.

B. Evaluation Metrics

In this section, we have evaluated and demonstrated elaborately the classification performance of our CNN based ECG rhythm classifier. In order to investigate the performance of recognition, four performance metrics, e.g., accuracy, precision, recall, and F1 score are measured. For a binary class problem, theses metrics of the classifier are defined as,

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$F1 = 2 \times \frac{precision.recall}{precision + recall}$$

TABLE IV
PERFORMANCE OF CNN BASED ARCHITECTURES FOR LONG DURATION ECG RHYTHM CLASSIFICATION.

Models	*learning rate=.001*				*learning rate=.005*				*learning rate=.0001*			
	Accuracy	*Precision*	*Recall*	*F1 score*	*Accuracy*	*Precision*	*Recall*	*F1 score*	*Accuracy*	*Precision*	*Recall*	*F1 score*
Model I	85.4%	.840	.860	.842	83.3%	.843	.833	.836	84.0%	.797	.840	.808
Model II	87.6%	.850	.871	.857	86.4%	.856	.864	.854	86.2%	.826	.862	.831
Model III	79.1%	.713	.792	.733	79.3%	.705	.793	.716	80.3%	.739	.803	.759
Model IV	85.9%	.839	.860	.845	84.0%	.835	.840	.832	85.1%	.816	.852	.817
Model V	83.7%	.805	.835	.809	88.4%	.870	.885	.857	78.1%	.736	.781	.697
Model VI	83.6%	.821	.836	.821	88.2%	.854	.882	.863	81.5%	.742	.815	.757

where, TP, TN, FP, FN represents true positive, true negative, false positive, false negative, respectively. In the multi-class cases, several averaging techniques are used to extend these binary metrics to multi-class. In our experiment, binary classification metrics are measured employing one-vs-rest strategy for each class. Afterwards, we compute the weighted average of individual binary metric since class imbalance exists in our experiment.

C. Performance Analysis

This section reveals the overall performances of the classifiers and also compares the performances of the six pre-trained CNN model based ECG arrhythmia classifiers in terms of accuracy, precision, recall, and F1 score. However, the performance are measures for three learning rates in order to investigate the impact of learning rate on the classification performances.

Table IV exhibits the comparison of overall performances of the six classifiers. When the learning rate is $.001$, the maximum classification accuracy is 87.6% which is achieved by model II that includes ResNet50 and the minimum accuracy is 79.1% achieved by model III which includes InceptionNet. In addition, the precision, recall and F1 score is also higher for the model II. Since the network is very large, residual connection in the model II assists to reduce effect of overfitting and to enhance classification performance. On the other hand, model V that includes DenseNet shows comparatively better performance than the model II when the learning rate is $.005$. At this learning rate, the maximum accuracy is 88.4% which is higher than the previous case as well as than the later case. Besides, the maximum value of precision, recall, and F1 score achieved by model V is $.870$, $.885$, $.857$, respectively at the learning rate $.005$. The model III gain shows lowest performance at this learning rate. Now, when the learning rate becomes $.0001$, model II shows the maximum accuracy along with the precision, recall, and F1 score. But, the model V performance decreases at this case and it attains lowest performance compared with other classifiers.

IV. CONCLUSION

Recognizing patterns of ECG signal rhythm using deep learning based architecture is one of the popular method to classify the type of arrhythmia. In this article, we have evaluated and investigated the performance of six structurally different CNN models for multi-class (15 classes) long duration ECG signal classification task. STFT has been applied to extract both time and frequency domain features from the ECG signals and create 2D spectrogram images. However, six pre-trined CNN based models are used as spatial feature extractor to study the importance of transfer learning in the arrhythmia detection task. Advantages of such methods are also observed in this paper. The overall experimental results depict the efficacy of the pre-trained CNN model based ECG rhythm classifiers with minimal training time due to transfer learning.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (Nos. 2016R1A5A1012966 and 2021R1F1A1062285).

REFERENCES

[1] B. Hemmeryckx, Y. Feng, L. Frederix, M. Lox, S. Trenson, R. Vreeken, H. R. Lu, D. Gallacher, Y. Ni, and H. R. Lijnen,"Evaluation of cardiac arrhythmic risks using a rabbit model of left ventricular systolic dysfunction," *Eur. J. Pharmacol.,* vol. 832, pp. 145–155, Aug. 2018.

[2] World Health. (2017). Cardiovascular Diseases (CVDs). Accessed: Apr. 18, 2018. [Online]. Available: *http://www.who.int/mediacentre/ fact-sheets/fs317/en/*.

[3] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith,"A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.,* vol. 50, no. 3, pp. 289–294, Mar. 2003.

[4] R. Hoekema *et al.*,"Geometrical aspects of the interindividual variability of multilead ECG recordings," *IEEE Trans. Biomed. Eng.,* vol. 48, no. 5, pp. 551–559, May 2001.

[5] E. J. da S. Luz, W. R. Schwartz, G. Camara-Chavez, and D. Menotti,"ECG-based heartbeat classification for arrhythmia detection: A survey," *Comput. Methods Programs Biomed.,* vol. 127, pp. 144–164, 2016.

[6] L. B. Marinho, N. D. M. M. Nascimento, J. W. M. Souza, M. V. Gurgel, P. P. R. Filho, and V. H. C. de Albuquerque,"A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification," *Future Gener. Comput. Syst.,* vol. 97, pp. 564–577, Aug. 2019.

[7] A. F. Khalaf, M. I. Owis, and I. A. Yassine,“A novel technique for cardiac arrhythmia classification using spectral correlation and support vector machines,” *Expert Syst. Appl.,* vol. 42, no. 21, pp. 8361–8368, Nov. 2015.
[8] M. Thomas, M. K. Das, and S. Ari,“Automatic ECG arrhythmia classification using dual tree complex wavelet based features,” *AEU-Int. J. Electron. Commun.,* vol. 69, no. 4, pp. 715–721, Apr. 2015.
[9] M. A. Escalona-Morán, M. C. Soriano, I. Fischer, and C. R. Mirasso,“Electrocardiogram classification using reservoir computing with logistic regression,” *IEEE J. Biomed. Health Informat.,* vol. 19, no. 3, pp. 892–898, May 2015.
[10] I. Christov, V. Krasteva, I. Simova, T. Neycheva, and R. Schmid,“Multiparametric analysis for atrial fibrillation classification in the ECG,” in *Proc. Comput. Cardiol. Conf. (CinC),* Rennes, France, Sep. 2017, pp. 1–4.
[11] N. Razmjooy, M. Ramezani, and N. Ghadimi,“Imperialist competitive algorithm-based optimization of neuro-fuzzy system parameters for automatic red-eye removal,” *Int. J. Fuzzy Syst.,* vol. 19, no. 4, pp. 1144–1156, Aug. 2017.
[12] Y. Kaya and H. Pehlivan, “Classification of premature ventricular contraction in ECG,” *Int. J. Adv. Comput. Sci. Appl.,* vol. 6, no. 7, pp. 34–40, Jul. 2015.
[13] R. C. Gonzalez, “Deep convolutional neural networks [Lecture notes],” *IEEE Signal Process. Mag.,* vol. 35, no. 6, pp. 79–87, Nov. 2018.
[14] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature,* vol. 521, no. 7553, pp. 436–444, May 2015.
[15] 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. 2016.
[16] D. Li, J. Zhang, Q. Zhang, and X. Wei, “‘Classification of ECG signals based on 1D convolution neural network,” in *Proc. IEEE 19th Int. Conf. E-Health Netw.,* Oct. 2017, pp. 1–6.
[17] W. Yin, X. Yang, L. Zhang, and E. Oki, “ECG monitoring system integrated with IR-UWB radar based on CNN,” *IEEE Access,* vol. 4, pp. 6344–6351, 2016.
[18] T. J. Jun, H. M. Nguyen, D. Kang, D. Kim, D. Kim, and Y.-H. Kim, “ECG arrhythmia classification using a 2-D convolutional neural network,” 2018 *arXiv:1804.06812.*
[19] 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.
[20] G. Serbes, B. E. Sakar, H. O. Gulcur, N. Aydin, “An emboli detection system based on Dual Tree Complex Wavelet Transform and ensemble learning,” *Applied Soft Computing,* Volume 37, 2015, Pages 87-94.
[21] S. Haykin and B. V. Veen, *Signals and Systems,* Hoboken, NJ, USA: Wiley, 1999.