

Received June 8, 2021, accepted July 11, 2021, date of publication July 15, 2021, date of current version July 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3097614

# ECG Heartbeat Classification Using Multimodal Fusion

**ZEESHAN AHMAD<sup>1</sup>**, (Graduate Student Member, IEEE), **ANIKA TABASSUM<sup>2</sup>**,  
**LING GUAN<sup>1</sup>**, (Fellow, IEEE), AND **NAIMUL MEFRAZ KHAN<sup>1</sup>**, (Senior Member, IEEE)

<sup>1</sup>Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada

<sup>2</sup>Data Science and Analytics Program, Ryerson University, Toronto, ON M5B 2K3, Canada

Corresponding author: Zeeshan Ahmad (zlahmad@ryerson.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Dapasoft Inc., to conduct the research is highly appreciated, under Grant CRDPJ529677-18.

**ABSTRACT** Electrocardiogram (ECG) is an authoritative source to diagnose and counter critical cardiovascular syndromes such as arrhythmia and myocardial infarction (MI). Current machine learning techniques either depend on manually extracted features or large and complex deep learning networks which merely utilize the 1D ECG signal directly. Since intelligent multimodal fusion can perform at the state-of-the-art level with an efficient deep network, therefore, in this paper, we propose two computationally efficient multimodal fusion frameworks for ECG heart beat classification called Multimodal Image Fusion (MIF) and Multimodal Feature Fusion (MFF). At the input of these frameworks, we convert the raw ECG data into three different images using Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF). In MIF, we first perform image fusion by combining three imaging modalities to create a single image modality which serves as input to the Convolutional Neural Network (CNN). In MFF, we extracted features from penultimate layer of CNNs and fused them to get unique and interdependent information necessary for better performance of classifier. These informational features are finally used to train a Support Vector Machine (SVM) classifier for ECG heart-beat classification. We demonstrate the superiority of the proposed fusion models by performing experiments on PhysioNet's MIT-BIH dataset for five distinct conditions of arrhythmias which are consistent with the AAMI EC57 protocols and on PTB diagnostics dataset for Myocardial Infarction (MI) classification. We achieved classification accuracy of 99.7% and 99.2% on arrhythmia and MI classification, respectively. Source code at <https://github.com/zaamad/ECG-Heartbeat-Classification-Using-Multimodal-Fusion>

**INDEX TERMS** Convolutional neural network, deep learning, ECG, image fusion, multimodal fusion.

## I. INTRODUCTION

Electrocardiogram is a reliable, effective and non-invasive diagnostic tool and is the best representation of electrophysiological pattern of depolarization and repolarization of the heart muscles during each heartbeat. Heart beat classification based on ECG provides conclusive information to the cardiologists about chronic cardiovascular diseases [1]. An intelligent system for diagnosing cardiovascular diseases is highly desirable because they are the leading source of death around the globe [2].

Arrhythmia is a heart rhythmic problem which occurs when electrical pulses that coordinate hearbeats cause heart to

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

beat irregularly, i.e., either too slow or too fast. Arrhythmias can be caused by coronary artery disease, high blood pressure, changes in the heart muscle (cardiomyopathy), valve disorders etc.

Myocardial Infarction, also known as heart attack, is caused due to the blockage of blood supply to the coronary arteries and in general to the myocardium. This blockage stops the supply of oxygen-rich blood to the heart muscle which can be life-threatening for the patient [3].

ECG beat-by-beat examination is vital for early diagnosis of cardiovascular conditions. However, differences of recording environment, variations of disease patterns among the subjects during testing, complex, non-stationary and noisy nature of ECG signal [4] make heartbeat classification a challenging and laborious exercise for cardiologists [5]. Thus,

computer based novel practices are useful for automatic and autonomous detection of abnormalities in heartbeat ECG classification.

Conventional methods for heartbeat classification using ECG signal rely mostly on hand-crafted or manually extracted features using signal processing techniques such as digital filter-based methods [6], mixture of experts methods [7], threshold-based methods [8], Principal Component Analysis (PCA) [9], Fourier Transform [10] and wavelet transform [11]. Some of the classifiers used with these extracted features are Support Vector Machines (SVM) [12], Hidden Markov Models (HMM) [13] and Neural Networks [14]. The first disadvantage with these conventional methods is the separation of feature extraction part and pattern classification part. Furthermore, these methods need expert knowledge about the input data and selected features [15]. Moreover, extracting features using subject experts is a time consuming process and features may not invariant to noise, scaling and translations and thus can fail to generalize well on unseen data.

Exemplary performance of deep neural networks (DNNs) on ECG [16] and especially the performance of CNN using 1D convolution [17] and 2D convolution [18] has recently attracted attention of many researchers. Deep learning models are capable of automatically learning invariant and hierarchical features directly from the data and employ end-to-end learning mechanism that takes data as input and class prediction as output. Recent deep learning models use 1D ECG signal or 2D representation of ECG by transforming ECG signal to images or some matrix form. For 1D ECG classification, commonly used deep learning models are deep belief networks, restricted Boltzmann machines, auto encoders, CNN [19] and recurrent neural network (RNN) [20]. For 2D ECG classification, CNNs are used and the input ECG data is transformed to images or some other 2D representation. It is experimentally proved in [21] that 2D representation of ECG provides more accurate heartbeat classification compared to 1D. In our previous work [22], univariate ECG signal is transformed to images by segmenting ECG signal between successive R-R intervals and then stacking these R-R intervals row wise to form images. Finally, multidomain multimodal fusion is performed to improve the stress assessment. Experimental results proved that multidomain multimodal fusion achieved highest performance as compared to single ECG modality.

Existing deep learning methods deprived of providing robust fusion framework and rely mostly on concatenation [23] and decision level fusion [24].

In this manuscript, we deal with the shortcomings of existing deep learning models for ECG heartbeat classification by proposing two fusion frameworks that have the capacity of extracting and fusing complementary and discriminative features while reducing dimensionality as well.

The proposed work has following significant contributions:

- 1) Two multimodal fusion frameworks for ECG heartbeat classification called Multimodal Image Fusion (MIF) and Multimodal Feature Fusion (MFF), are proposed. At the input of these frameworks, we convert the heartbeats of raw ECG data into three types of two-dimensional (2D) images using Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF). Proposed fusion frameworks are computationally efficient as they keep the size of the combined features similar to the size of individual input modality features.
- 2) We transform heartbeats of ECG signal to images using Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF) to conserve the spatial domain correlated information among the data samples. These transformations result in an improvement in classification performance in contrast to the existing approaches of transforming ECG to images using spectrograms or methods involving time-frequency analysis (Short time Fourier transform or wavelet transform).

## II. RELATED WORK

Deep Learning models especially CNN has been used over the years for ECG heartbeat classification for the detection of cardiovascular diseases such as arrhythmia and MI. These models include both 1D and 2D CNNs.

### A. ONE-DIMENSIONAL CNN APPROACHES

Various models based on 1D CNN has been proposed in the literature for ECG classification. In [25], an active learning model based on 1D CNN is presented for arrhythmia detection using ECG signal. Model performance is improved by using breaking-ties (BT) and modified BT algorithms. Authors in [26] proposed a model for adaptive real time implementation of a patient-specific ECG heartbeat classification based on 1D CNN using end-to-end learning. In [27], a novel algorithm making use of an 11-layer deep CNN is proposed for automatic detection of MI using ECG beats with and without noise. A transfer learning method based on CNN is proposed in [28] where the information learned from arrhythmia classification task is employed as a reference for the training of classifiers. A computationally intelligent method for patient screening and arrhythmia detection using CNN is proposed in [29]. The proposed method is capable of diagnosing arrhythmia conditions without expert domain knowledge and feature selection mechanism. In [30], wavelet transform based on Fourier-Bessel series expansion is proposed for the localization of ECG. The Fourier-Bessel spectrum of the ECG beats is separated into adjacent parts using the fixed order ranges and then multiscale CNN is employed for MI classification of different categories. Multi-Channel Lightweight Convolutional Neural Network (MCL-CNN) which uses squeeze convolution, the depth-wise convolution, and the point-wise convolution is proposed in [31] for MI classification. Two end-to-end deep learning models based on

CNN are proposed in [32]. These models are called two stage hierarchical model. Furthermore, generative adversarial networks (GANs) is used for data augmentation and to reduce the class imbalance. In [33], authors proposed a neural network model for precise classification of heartbeats by following the AAMI inter-patient standards. This model works in two steps. In the first step the signals are preprocessed and then features are extracted from the signals. In the second step, the classification is performed by a two-layer classifier in which each layer consists of two independent fully-connected neural networks. The experiments show that the proposed model precisely detects arrhythmia conditions. In [34], authors proposed a complex deep learning model consists of CNN and LSTM. This model classifies six types of ECG signals by processing ten seconds ECG slices of MIT-BIH arrhythmia dataset. Experimental results proved that the proposed model could be used by cardiologists to detect arrhythmia. In [35], authors presented CNN based model for proper diagnoses of congestive heart failure using ECG. The testing and training of the proposed model was carried out on publicly available ECG datasets. Performance of the proposed model shows the authenticity of model for congestive heart failure detection.

### B. TWO-DIMENSIONAL CNN APPROACHES

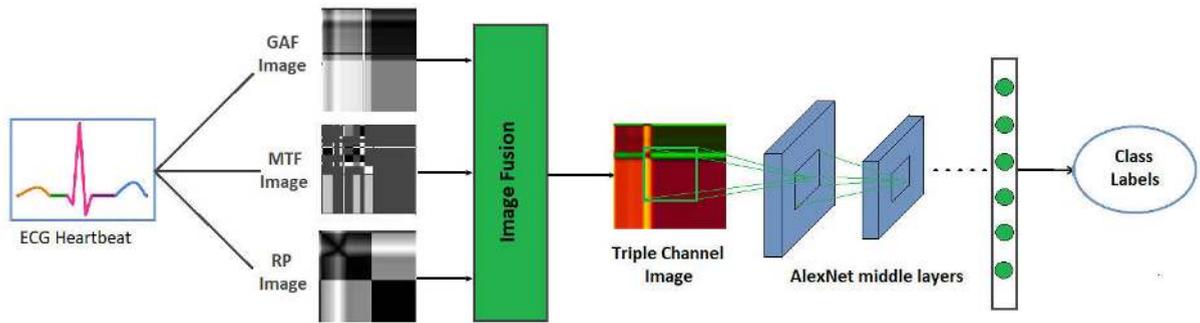
The knock out performance of CNN on 2D data such as images convinced the researchers to convert raw ECG data to images for improved results. In [21], short-time Fourier transform is used to convert ECG signal into time-frequency spectrograms that were used as input to CNN for arrhythmia classification. Experimental results show that 2D-CNN achieved higher classification accuracy than 1D-CNN. In [36], ECG signal is converted into spectro-temporal images that were sent as an input to multiple dense convolutional neural network to capture both beat-to-beat and single-beat information for analysis. Authors in [37] transformed heartbeat time intervals of ECG signals to images using wavelet transform. These images are used to train a six layer CNN for heartbeat classification. In [38], Generative neural network is used to convert the raw 1D ECG signal data into a 2D image. These images are input to DenseNet which produces highly accurate classification, with high sensitivity and specificity using 4 classes of heart beat detection. To distinguish abnormal ECG samples from normal, authors in [39] used pretrained CNNs such as AlexNet, VGG-16 and ResNet-18 on spectrograms obtained from ECG. Using a transfer learning approach, the highest accuracy of 83.82% is achieved by AlexNet. In [40], multi-lead ECG are treated as 2D matrices for input to a novel model called multilead-CNN (ML-CNN) which employs sub two-dimensional (2D) convolutional layers and lead asymmetric pooling (LAP) layers. In [41], authors generated dual beat coupling matrix from the sections of heartbeats. This dual beat coupling matrix was then as 2D input to a CNN classifier. Gray-level co-occurrence matrix (GLCM), obtained from ECG data is employed for features vector description due to its exceptional statistical feature extraction

ability in [42]. In [43], ECG signals were segmented into heartbeats and each of the heartbeats were transformed to 2D grayscale images which were input to CNN. In [44], two second segments of ECG signal are transformed to recurrence plot images to classify arrhythmia in two steps using deep learning model. In the first step the noise and ventricular fibrillation (VF) categories were recognized and in the second step, the atrial fibrillation (AF), normal, premature AF, and premature VF labels were classified. Experimental results show the promising performance of the proposed method.

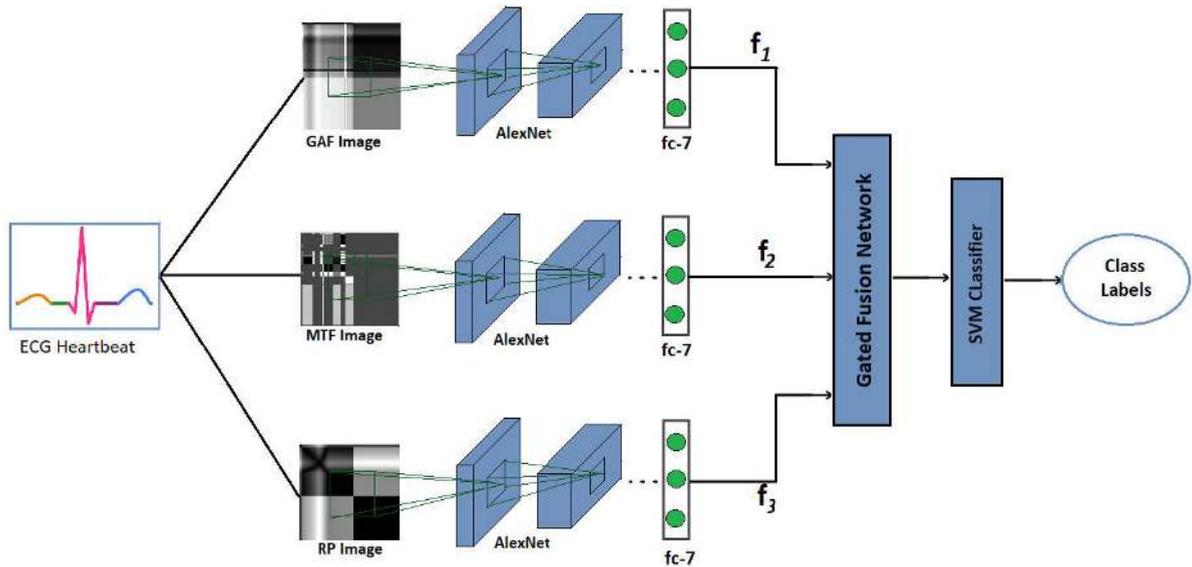
### C. FUSION BASED APPROACHES

Fusing different modalities mitigates the weaknesses of individual modalities both in 1D and 2D forms by integrating complementary information from the modalities to perform the analysis and classification tasks accurately. In [45], a Multi-scale Fusion convolutional neural network (MS-CNN) is proposed for heartbeat classification using ECG signal. The Multi-scale Fusion convolutional neural network is a two stream network consisting of 13 layers. The features obtained from the last convolutional layer are concatenated before classification. Another Deep Multi-scale Fusion CNN (DMSFNet) is proposed in [46] for arrhythmia detection. Proposed model consists of backbone network and two different scale-specific networks. Features obtained from two scale specific networks are fused using a spatial attention module. Patient-specific heartbeat classification network based on a customized CNN is proposed in [47]. CNN contains an important module called multi-receptive field spatial feature extraction (MRF-SFE). The MRF-SFE module is designed for extracting multispatial deep features of the heartbeats using five parallel convolution layers with different receptive fields. These features are concatenated before being sent to the third convolutional layer for further processing. Two stage serial fusion classifier system based on SVM's rejection option is proposed in [48]. SVM's distance outputs are related with confidence measure and then ambiguous samples are rejected with first level SVM classifier. The rejected samples are then forwarded to a second stage Logistic Regression classifier and then late fusion is performed for arrhythmia classification. Authors in [49] presented a unique feature fusion method called parallel graphical feature fusion where all the focus is given to geometric features of data. Original signal was first split into subspaces, then multidimensional features are extracted from these subspaces and then mapped to the points in high-dimensional space. Multi-stage feature fusion framework based on CNN and attention module was proposed in [50] for multiclass arrhythmia detection. Classification is performed by extracting features from different layers of CNN. Combination of CNN and the attention module shows the improved discrimination power of the proposed model for ECG classification.

The shortcoming in the existing fusion methods is that they depend mostly on concatenation fusion. Concatenation leads towards the problem computational complexity, curse of dimensionality and hence the degradation in classification



**FIGURE 1.** Complete overview of the proposed multimodal image fusion (MIF) framework. We fused GAF, RP and MTF image to form a triple-channel (GAF-RP-MTF) compound image containing both static and dynamic features of input images.



**FIGURE 2.** Complete overview of the proposed multimodal feature fusion (MFF) framework. The MFF extracted features from fc-7 layer of AlexNet. These features are then integrated through gated fusion network (GFN) and are finally sent to the classifier.

accuracy [51]. In this paper, we address the imperfections of the existing literature and propose two fusion frameworks called Multimodal Image Fusion (MIF) and Multimodal Feature Fusion (MFF) which extract and fuse the features while reducing dimensionality as well. The proposed fusion frameworks are described in section III.

### III. MATERIALS AND METHODS

This section explains the proposed fusion frameworks called Multimodal Image Fusion (MIF) and Multimodal Feature Fusion (MFF). The common element in both of the proposed fusion framework is ECG signal to image transformation as shown in Figures 1 and 2. Therefore in this section, first we will explain ECG signal to image transformation and then MIF, MFF and the two important elements of MFF, gated fusion network shown in Fig. 3 and architecture of CNN shown in Fig. 4, will be explained.

#### A. ECG SIGNAL TO IMAGE TRANSFORMATION

For each fusion framework, we transform the input heart-beats into three types of images called GAF, RP and MTF images.

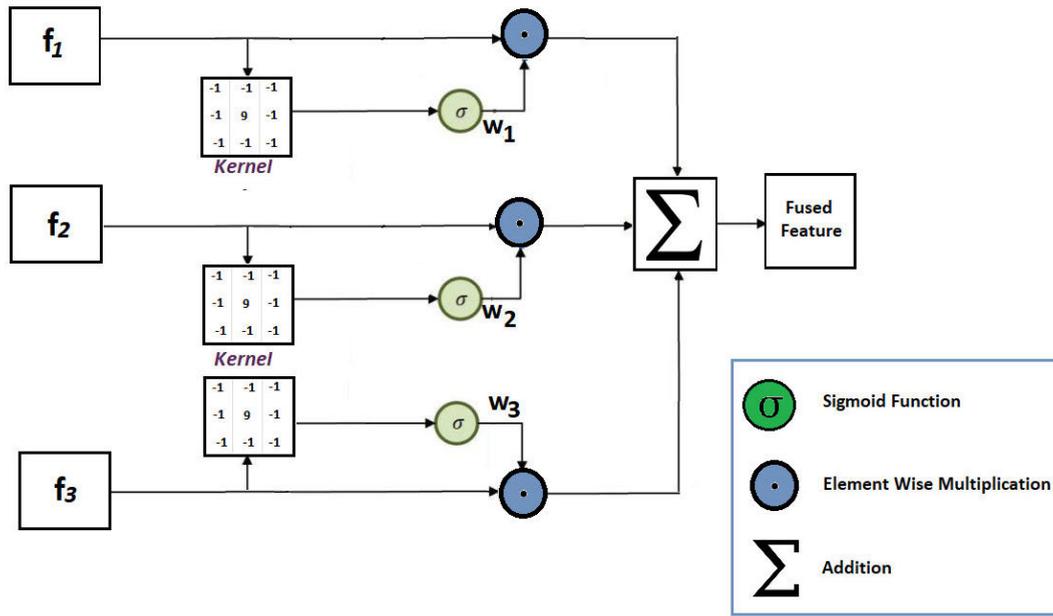
#### 1) FORMATION OF IMAGES BY GRAMIAN ANGULAR FIELD (GAF)

Converting heart-beats of ECG into Gramian Angular Field (GAF) images maps the ECG in an angular coordinate system instead of typical rectangular coordinate system.

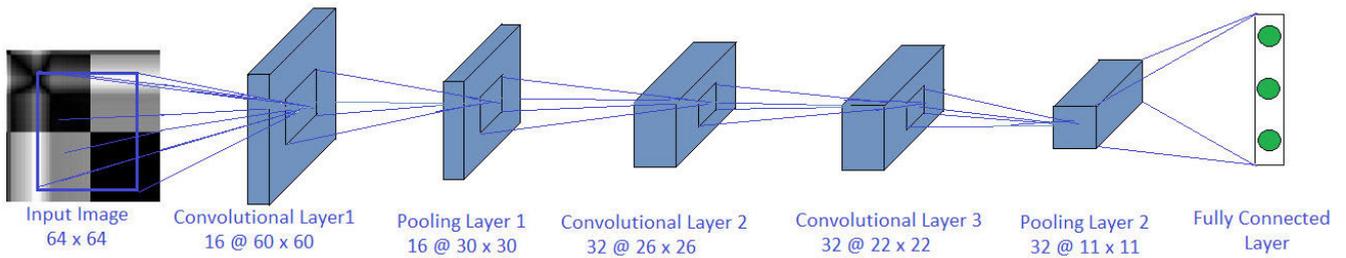
Consider that  $E$  is an ECG signal of  $n$  samples such that  $E = \{s_1, s_2, s_3, \dots, s_k, s_l, \dots, s_n\}$ . We normalized  $E$  between 0 and 1 to get  $\bar{E}$ . Now we map the normalized ECG in angular coordinate system by transforming the value into the angular cosine and the time stamps into the radius. Following equation is used to explain this encoding.

$$\left. \begin{aligned} \beta &= \arccos(s_{k0}) \\ R &= \frac{t_k}{C} \end{aligned} \right\} \quad (1)$$

In the above equation,  $s_{k0}$  is normalized  $k$ th sample of the ECG,  $t_k$  is the time stamp for  $s_{k0}$  and  $C$  is a constant to adjust the spread of the angular coordinate system. This encoding provides two benefits. It is bijective and it conserves the spatial domain affiliations through the  $R$  [52]. Since the image location with respect to the ECG heart beat samples is consistent along the principal diagonal, therefore, the original



**FIGURE 3.** Structure of the proposed gated fusion network. Input feature  $f_1$ ,  $f_2$  and  $f_3$  from modalities are convolved with high boost kernel and then gated values  $w_1$ ,  $w_2$  and  $w_3$  are generated using sigmoid function. Finally, these gated values are multiplied element-wise with input features to perform fusion.



**FIGURE 4.** Architecture of CNN for signal image of size  $64 \times 64$ .

heart beat samples of ECG can be restored from angular coordinates [53].

The angular viewpoint of the encoded image can be exploited by taking into account the sum/difference between each sample to indicate the correlation among various time stamps. The summation method, used in this article is explained by the following set of equations.

$$Grammian\ field = \cos(\beta_k + \beta_l) \quad (2)$$

$$Grammian\ field = \bar{E}^T \cdot \bar{E} - \sqrt{I - \bar{E}^2}^T \cdot \sqrt{I - \bar{E}^2} \quad (3)$$

$I$  is the unit row vector in equation 3

GAF Images of five different categories for MIT-BIH dataset are shown in Fig 5.

## 2) FORMATION OF IMAGES BY RECURRENCE PLOT (RP)

ECG is a non-stationary signal, therefore to visualize the recurrent behavior and to observe the recurrence pattern of ECG signal [54], we encode ECG heartbeats into RP images. An RP image obtained from a heartbeat of ECG represents spacing between time points [55].

For ECG signal  $E$  defined in section III-A1, the recurrence plot is given by

$$R\text{-plot} = \alpha(\lambda - ||s(k) - s(l)||) \quad (4)$$

where  $\lambda$  is threshold and  $\alpha$  is the heaviside function.

RP Images of five different categories for MIT-BIH dataset are shown in Fig 5.

## 3) ECG TO MARKOV TRANSITION FIELD (MTF) IMAGE CONVERSION

For ECG heartbeats to MTF image encoding, we used the same approach explained in [56]. Let  $E$  is the ECG signal defined in section III-A1, then the foremost step is to define its  $B$  bins based on quantiles and assign every  $s_k$  to the related bins  $b_j(j \in [1, B])$ . Second step is the construction of  $B \times B$  weighted adjacency matrix  $W$  by computing transformations within quantile bins like a first-order Markov chain on the time axis. Weighted adjacency matrix in the normalized form is called Markov transition matrix and is non-relative to the spatial domain characteristics, resulting in information loss. For handling the loss of

Categories	N	S	V	F	Q
Heartbeats					
GAF Images					
RP Images					
MTF Images					

FIGURE 5. GAF, RP and MTF images of MIT-BIH dataset according to the five different heartbeats defined in Table 2.

information, Markov transition matrix is transformed to Markov transition field matrix (MTF) by stretching the transition likelihoods corresponding to the spatial domain locations. The MTF matrix is denoted by  $M$  and is shown below

$$M = \begin{bmatrix} w_{lk|s_1 \in b_l, s_1 \in b_k} & \dots & w_{lk|s_1 \in b_l, s_n \in b_k} \\ w_{lk|s_2 \in b_l, s_1 \in b_k} & \dots & w_{lk|s_2 \in b_l, s_n \in b_k} \\ \vdots & \ddots & \vdots \\ w_{lk|s_n \in b_l, s_1 \in b_k} & \dots & w_{lk|s_n \in b_l, s_n \in b_k} \end{bmatrix} \quad (5)$$

where  $w_{lk}$  is the frequency of transition of a point between two quantiles. Since the formation of transformed matrix depends upon the chances of moving element, the MTF cannot be restored to original ECG signal.

Bins are the quantiles where the probability distribution is same. Any number of bins can be selected for ECG to MTF images. We decided to take 10 bins as the data is normalized between 0 and 1. These bins are defined during the formation of Weighted adjacency matrix which is the first step for creating MTF matrix shown in equation 5.

MTF Images of five different categories for MIT-BIH dataset are shown in Fig 5.

For ECG to image transformation using GAT, RP and MTF methods, we are using the full length of heartbeats to transform 1D information to 2D. Therefore, ECG signal of any length can be transformed to images and then can be resized using interpolation.

We can see from Fig. 5, that for each kind of image (GAF, RP and MTF), the gray scale images are more interpretable. These images show different patterns for each of the five categories of MIT-BIH dataset. The x-y values of the 2D images are just pixel values of the GAF, RP, and MTF images.

### B. MULTIMODAL IMAGE FUSION FRAMEWORK

Multimodal Image Fusion (MIF) framework is shown in Fig. 1. At the input, we transform the heartbeats of raw ECG signal into three types of images as described in section III-A and shown in Fig. 5. The motivation of choosing GAF, MTF and RP is that they are three different statistical methods of transforming ECG to images. During transformation they preserve the temporal information and hence they are lossless transformations. We combine these three gray scale images to form a triple channel image (GAF-RP-MTF). A triple channel image is a colored image in which GAF, RP and MTF images are considered as three orthogonal channels like three different colors in RGB image space. However, this three-channel image is not conventional way of converting a gray scale image to RGB, rather in this paper all three gray scale images are formed from raw ECG data with different statistical methods. Thus, a three-channel image in the presented work carries statistical dynamics of the ECG and therefore, is more informative. Furthermore, three-channel image can be easily utilized with off-the-shelf CNNs like AlexNet.

We use AlexNet, (CNN based model) [57] for feature extraction and classification tasks and thus employ end-to-end deep learning where feature extraction and classification parts are embedded in a single network as shown in Fig. 1.

### C. MULTIMODAL FEATURE FUSION FRAMEWORK

At the input of MFF, we transform ECG heartbeats into images as shown in Fig. 2. AlexNets are employed to learn features from input imaging modality. We extract these learned features from (fc-7) of each AlexNet and are then fused by an efficient Gated Fusion Network (GFN), backbone of the proposed MFF, which fuses the features effectively

by taking care of their dimensionalities as well. These fused features are input of the SVM classifier as shown in Fig. 2.

### 1) GATED FUSION NETWORK

The architecture of our proposed gated fusion network (GFN) is shown in Fig. 3. We have adapted this network from our previous work in [58]. The input to the GFN are the features extracted from the second last fully connected layer (fc-7) of each AlexNet as shown in Fig. 2.

Let  $f_1, f_2$  and  $f_3$  be the features from each imaging modality respectively. These feature are then convolved with high boost kernel  $K$  as shown in Fig. 3.

We used high boost filter for convolution with features since this filter precisely recognize important information of feature and accredits boosted value to every element of features according to its importance [59]. High boost filter is the difference between scaled version and low-pass version of the input image as shown below in equation 6.

$$f_{hb}(m, n) = cf(m, n) - f_{lp}(m, n) \quad (6)$$

where  $cf(m, n)$  and  $f_{lp}(m, n)$  are respectively the scaled version and low pass version of image  $f(m, n)$

In general, high boost filter is given by

$$K = \begin{bmatrix} -1 & -1 & -1 \\ -1 & c + 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (7)$$

where  $c$  is the amplification factor that assigns the weights to the feature during convolution.

The best filter performance is obtained for  $c = 1$ . Other values of  $c$  produces less amplification.

Thus, following high boost kernel is selected empirically that highlights the important characteristics.

$$K = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad (8)$$

High boost filter highlights the high frequency components while conserving the low frequency components.

After convolution of features with the high boost filter, sigmoid function is used for generating proper gated weights  $w_1, w_2$  and  $w_3$  respectively as shown in Fig. 3. Finally, we obtained point-wise product of the weights  $w_1, w_2$  and  $w_3$  and the features  $f_1, f_2$  and  $f_3$  respectively, to perform feature fusion and to generate fused features. The working of GFN can be understood by the following equations.

$$w_1 = \sigma(f_1 \otimes K) \quad (9)$$

$$w_2 = \sigma(f_2 \otimes K) \quad (10)$$

$$w_3 = \sigma(f_3 \otimes K) \quad (11)$$

$$F_f(j) = w_1 \odot f_1(j) + w_2 \odot f_2(j) + w_3 \odot f_3(j) \quad (12)$$

where,

$\sigma(x) \triangleq \frac{1}{1+e^{-x}}$ : Sigmoid Function.

$a \otimes b$ : Convolution

$a \odot b$ : Point Wise Multiplication

$F_i(k)$ :  $k$ th feature of  $i$ th modality

$F_f(k)$ :  $k$ th Fused feature

### 2) CNN ARCHITECTURE

Architecture of CNN used in proposed MFF is shown in Fig. 4. It consists of three convolutional layers, two pooling layers, and a fully connected layer. The first convolutional layer has 16 kernels of size  $5 \times 5$ , followed by pooling layer of size  $2 \times 2$  and stride 2. Second and third convolutional layers have 32 kernels of size  $5 \times 5$  followed by  $2 \times 2$  pooling layer with stride 2.

### D. CLASSIFICATION TASK AND CLASSIFIER

The classification task of the proposed methods is ECG heart beat classification for arrhythmia and MI detection.

The classification metrics used for classification are accuracy, precision and recall as shown in Tables 5, 6, 7 and 8. The accuracies, precisions and recalls are calculated using following equations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Recall = \frac{TP}{TP + FN} \quad (15)$$

where,

$TP$  = True positive

$TN$  = True negative

$FP$  = False positive

$FN$  = False negative

We used Softmax classifier in proposed MIF and Support Vector Machines (SVM) classifier in proposed MFF for classification task.

Softmax classifier is a multiclass classifier or regressor used in the fields of machine learning. Score function for softmax classifier computes the class specific probabilities whose sum is 1.

The mathematical representation of score function for softmax classifier is shown below.

$$f(y) = \frac{e^{y_j}}{\sum_k e^{y_k}} \quad (16)$$

where  $y$  is the input vector and the score function maps the exponent domain to the probabilities.

In simplest form, the score function for SVM is the mapping of the input vector to the scores and is a simple matrix operation as shown in Equation 17.

$$f = Wx + b \quad (17)$$

where  $x$  is the input vector,  $W$  is the weight determined by input vector and the number of classes and  $b$  is the bias vector.

### E. TRAINING AND OPTIMIZATION

We resize images to  $227 \times 227$  to perform experiments with AlexNet. We also perform experiments with smaller but

**TABLE 1.** Training parameters for AlexNet and CNN.

Training Parameters	Values
Momentum	0.9
Initial Learn Rate	0.005
Learn Rate Drop Factor	0.5
Learn Rate Drop Period	10
$L_2$ Regularization	0.004
MiniBatchSize	128

computationally efficient CNN, whose architecture is shown in Fig. 4, to show that proposed frameworks can achieve comparable performance even with the smaller CNN. The comparison in terms of computational cost between both CNN models is provided in Table 11. We fine tune Alexnet by reducing the size of second last fully connected layer 'fc7' from 4096 to 512 and the size of last fully connected layer 'fc8' from 1000 to size equal to the number of classes in our datasets. The size of "fc7" layer of AlexNet is 4096 which is according to size of classification layer which is 1000. For our MIT-BIH dataset and PTB dataset, we need the size of classification layer equal to 5 and 2 respectively due to number of classes in these datasets. Thus to make 'fc7' compatible with classification layer, we reduce its size to 512. The training parameters for AlexNet and CCN are shown in Table 1.

For optimization of the deep networks, we used Stochastic Gradient Descent with Momentum (SGDM) algorithm. SGDM is a method which helps accelerate gradients vectors in the right directions, thus leading to faster converging. It is one of the most popular optimization algorithms and many state-of-the-art models are trained using it.

## IV. EXPERIMENTAL RESULTS

### A. ECG DATABASES

Experiments are performed with PhysioNet MIT-BIH Arrhythmia dataset [60], [61] for heartbeat classification and PTB Diagnostic ECG dataset [62] for MI classification using both proposed fusion frameworks. For experiments, ECG lead-II re-sampled data at sampling frequency of 125Hz is used as the input.

We used the standardized form of both datasets provided in [63]. These datasets are already denoised and the training and testing parts are provided in the form of standard ECG heartbeats. Furthermore, five classes of arrhythmia and MI localization has already been done and provided in terms of standard ECG heart-beats. Our study focused on ECG to image transformation and to the design of proposed multimodal fusion frameworks. The main focus is increasing the overall performance of classification of heartbeats. We did not attempt at modeling or solving for a specific type of noise.

We conduct our experiments on Matlab R2020a on a desktop computer with NVIDIA GTX-1070 GPU.

The experimental results are discussed in detail in section V.

#### 1) PhysioNet MIT-BIH ARRHYTHMIA DATASET

Forty seven subjects were involved during the collection of ECG signals for the dataset. The data was collected at

**TABLE 2.** Mapping between annotations and AAMI EC57 [64] categories.

Categories	Annotations
N	Normal
	Left/Right bundle branch block
	Atrial escape
	Nodal escape
S	Atrial Premature
	Aberrant atrial premature
	Nodal premature
	Supra-ventricular premature
V	Premature ventricular contraction
	Ventricular escape
F	Fusion of ventricular and normal
	Paced
Q	Fusion of paced and normal
	Unclassifiable

**TABLE 3.** Information about number of heartbeats before and after SMOTE for training component of MIT-BIH dataset.

Dataset	Classes	Original heartbeats	Number of heartbeats after SMOTE
MIT-BIH	N	72471	72471
	S	2223	30000
	V	5788	20000
	F	641	20000
	Q	6431	10000

**TABLE 4.** Training and testing samples of datasets.

Dataset	Training Samples	Testing Samples
MIT-BIH	152471	21892
PTB	11641	2911

the sampling rate of 360Hz and each beat is annotated by at least two experts. Using these annotations, five different beat categories are created in accordance with Association for the Advancement of Medical Instrumentation (AAMI) EC57 standard [64] as shown in Table 2.

For training on CNN, we need large number of samples. We use the same testing and training segments provided in [63] to train on CNNs. Since there is a class-imbalanced in the training part of the dataset as apparent from the numbers, we applied SMOTE [65] to upsample the minority classes (classes other than N) and finally settled on the numbers shown in the right column of Table 3.

SMOTE is a data augmented technique which is used to reduce overfitting during training and is helpful to reduce the biasness of classifier.

We perform experiments using both proposed fusion frameworks on MIT-BIH dataset with the training and testing samples shown in Table 4 and with the training parameters shown in Tables 1. The experimental results are shown in Tables 5 and 6.

#### 2) PTB DIAGNOSTIC ECG DATASET

Two hundred and ninety (290) subjects took part during collection of ECG records for PTB Diagnostics dataset. 148 of them are diagnosed as MI, 52 healthy control, and the rest are diagnosed with 7 different diseases. Frequency of 100Hz is

**TABLE 5.** Experimental results of MIT-BIH dataset using AlexNet.

Modalities	Accuracies%	Precision%	Recall%
GAF Images only	97.3	85	91
RP Images only	97.2	82	93
MTF Images only	91.5	86	89
Concatenation Fusion	97	82	91
Average Fusion	98.5	95	93.1
Proposed MIF	98.6	93	92
Proposed MFF	99.7	98	98

**TABLE 6.** Experimental results of MIT-BIH dataset using simpler CNN of Fig. 4.

Modalities	Accuracies%	Precision%	Recall%
GAF Images(gray scale)	94.2	74.2	91
RP Images(gray scale)	96.3	80	90
MTF Images(gray scale)	94	72	86
Concatenation Fusion	94.6	80.4	84
Average Fusion	97.6	87	92
Proposed MFF	98.3	90.5	93

**TABLE 7.** Experimental results of PTB dataset using AlexNet.

Modalities	Accuracies%	Precision%	Recall%
GAF Images only	98.4	98	96
RP Images only	98	98	94
MTF Images only	95.3	94	89
Concatenation Fusion	97.4	95	95
Average Fusion	98.5	97	98
Proposed MIF	98.4	98	94
Proposed MFF	99.2	98	98

**TABLE 8.** Experimental results of PTB dataset using simpler CNN of Fig. 4.

Modalities	Accuracies%	Precision%	Recall%
GAF Images (gray scale)	94.7	91	90
RP Images (gray scale)	95.1	95	87
MTF Images(gray scale)	86.6	80	69
Concatenation Fusion	92.2	88	84
Average Fusion	96.3	91	94
Proposed MFF	96.5	94	93

used for each ECG record from 12 leads. However, for our experiments, we used lead II ECG recordings and worked with healthy control and MI categories.

We perform experiments using both proposed fusion frameworks on PTB dataset with training and testing samples shown in Table 4 and with training parameters shown in Tables 1. Training and testing parts of the dataset are provided in [63] to train CNN models. The experimental results are shown in Tables 7 and 8

## V. DISCUSSION

We present the comparative results of the proposed frameworks with the state-of-the-art methods in Tables 9 and 10. As we can see, our proposed frameworks considerably outperform the existing methods in terms of accuracy, precision, and recall.

To justify the importance of the proposed fusion frameworks, we assess the performance of different components of the proposed framework with both datasets by concatenation and average fusion methods. We performed average fusion by accrediting the unity value to all the weights, i.e.,  $w_1 = 1$ ,  $w_2 = 1$  and  $w_3 = 1$  in the gated fusion network. Since we have three modalities, therefore, by taking simple average, we get

**TABLE 9.** Comparison of heart beat classification results of MITBIH dataset with previous methods.

Previous Methods	Accuracies%	Precision%	Recall%
Izci et al. [43]	97.96	-	-
Dang et al. [23]	95.48	96.53	87.74
Li et al. [47]	99.5	97.3	98.1
Zhao et al. [49]	98.25	-	-
Oliveria et al. [37]	95.3	-	-
Huang et al. [21]	99	-	-
Shaker et al. [32]	98	90	97.7
Kachuee et al. [28]	93.4	-	-
Xu et al. [66]	95.9	-	-
He et al. [67]	98.3	-	-
Qiao et al. [68]	99.3	-	-
<b>Proposed MIF</b>	<b>98.6</b>	<b>93</b>	<b>92</b>
<b>Proposed MFF</b>	<b>99.7</b>	<b>98</b>	<b>98</b>

**TABLE 10.** Comparison of MI classification results of PTB dataset with previous methods.

Previous Methods	Accuracies%	Precision%	Recall%
Dicker et al. [39]	83.82	82	95
Acharya et al. [27]	95.22	95.49	94.19
Kojuri et al. [69]	95.6	97.9	93.3
Kachuee et al. [28]	95.9	95.2	95.1
Liu et al. [40]	96	97.37	95.4
Sharma et al. [12]	96	99	93
Chen et al. [31]	96.18	97.32	93.67
Cao et al. [70]	96.65	-	-
Ahamed et al. [71]	97.66	-	-
<b>Proposed MIF</b>	<b>98.4</b>	<b>98</b>	<b>94</b>
<b>Proposed MFF</b>	<b>99.2</b>	<b>98</b>	<b>98</b>

**TABLE 11.** Comparison of computational cost of AlexNet and CNN of Fig. 4 using MFF framework on MIT-BIH dataset.

CNN Model	Fusion Framework	Training Parameters
AlexNet	MFF	9259427
AlexNet	MIF	3086475
CNN of Fig. 4	MFF	612069

the equal value of 0.333 for each weight. We also experiment with 0.333 and get the same results. Since weights are equal in average fusion, therefore, to make things simpler, we assign a unity value to every weight. It is possible that better weight can be acquired through trainable weight coefficients. This is something we plan to investigate in future. Tables 5, 6, 7 and 8 reports the results of assessing different fusion methods along with proposed fusion frameworks.

The performance of concatenation fusion is poor as compared to other methods as shown by experimental results. Concatenation fusion creates high dimensional feature vector that leads to the additional computational cost and deterioration of information during classification [72].

We also provide the comparison of both proposed fusion frameworks in terms of inference speed as shown in Table 12. Inference speed is the time consumed by classifier to recognize one test sample. It is expressed in microseconds ( $\mu$ s). It is observed that MFF yields high accuracy, precision and recall for both datasets as compared to MIF, however, MIF is computationally efficient in terms of inference speed.

Since we experiment with two different CNNs, we provide comparison between both CNNs in terms of computational cost as shown in Table 11. Since there is a tradeoff between accuracy and computational cost, we observe from

**TABLE 12. Comparison of inference speed of both proposed fusion frameworks using AlexNet.**

Dataset	Fusion Method	Inference Speed( $\mu$ s)
MIT-BIH	Multimodal Image Fusion	1233
	Multimodal Feature Fusion	1670
PTB	Multimodal Image Fusion	1205
	Multimodal Feature Fusion	1470

Tables 5, 6 and 11 that CNN, shown in Fig. 4, is less accurate than AlexNet but is computationally efficient.

We prefer SVM classifier over softmax classifier since we have experimentally proved in our previous work [73] that SVM performs better than softmax, which is typically built into any CNN framework. Softmax classifier reduces the cross entropy function while SVM employs a margin based function. The more rigorous nature of classification is the reason of better performance of SVM over softmax.

The comparison provided in Tables 9 and 10 is on the basis of datasets and the performance metrics. There are slight changes in the conditions for testing in few of the comparisons, However, it is appropriate to compare the results.

The limitation of the proposed Multimodal Image Fusion (MIF) Framework is that it requires exactly three different statistical gray scale images for creating a triple channel compound image. Since Multimodal Feature Fusion (MFF) Framework is using three separate AlexNet for training on GAF, RP and MTF images, it requires more time for training and inference.

## VI. CONCLUSION

We proposed two computationally efficient multimodal fusion frameworks for ECG heart beat classification called Multimodal Image Fusion (MIF) and Multimodal Feature Fusion (MFF). At the input of these frameworks, we convert ECG signal into three types of images using Gramian Angular Field (GAF), Recurrence Plot (RP) and Markov Transition Field (MTF). In MIF, we first perform image fusion by combining three input images to create a three channel single image which used as input to the CNN. In MFF, highly informative cues are pulled out from penultimate layer of CNN and they are fused and used as input for the SVM classifier. We demonstrate the superiority of the proposed fusion frameworks by performing experiments on PhysionNet's MIT-BIH for five different arrhythmias and on PTB diagnostics dataset for MI classification. Experimental results prove that we beat the previous state-of-the-art in terms of classification accuracy, precision and recall. The important finding of this study is that the multimodal fusion of modalities increases the performance of the machine learning task as compare to use the modalities individually.

## REFERENCES

[1] L. Sun, Y. Lu, K. Yang, and S. Li, "ECG analysis using multiple instance learning for myocardial infarction detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3348–3356, Dec. 2012.

[2] Y. Xia, X. Liu, D. Wu, H. Xiong, L. Ren, L. Xu, W. Wu, and H. Zhang, "Influence of beat-to-beat blood pressure variability on vascular elasticity in hypertensive population," *Sci. Rep.*, vol. 7, no. 1, pp. 1–8, Dec. 2017.

[3] R. U. Acharya, N. Kannathal, L. M. Hua, and L. M. Yi, "Study of heart rate variability signals at sitting and lying postures," *J. Bodywork Movement Therapies*, vol. 9, no. 2, pp. 134–141, Apr. 2005.

[4] U. R. Acharya, Y. Hagiwara, J. E. W. Koh, S. L. Oh, J. H. Tan, M. Adam, and R. S. Tan, "Entropies for automated detection of coronary artery disease using ECG signals: A review," *Biocybern. Biomed. Eng.*, vol. 38, no. 2, pp. 373–384, 2018.

[5] Z. Zhang, J. Dong, X. Luo, K.-S. Choi, and X. Wu, "Heartbeat classification using disease-specific feature selection," *Comput. Biol. Med.*, vol. 46, pp. 79–89, Mar. 2014.

[6] E. Pasolli and F. Melgani, "Active learning methods for electrocardiographic signal classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 6, pp. 1405–1416, Nov. 2010.

[7] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ECG beat classifier using a mixture of experts approach," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 9, pp. 891–900, Sep. 1997.

[8] V. Chouhan and S. Mehta, "Threshold-based detection of P and T-wave in ECG using new feature signal," *Int. J. Comput. Sci. Netw. Secur.*, vol. 8, no. 2, pp. 144–153, 2008.

[9] N. A. Bhaskar, "Performance analysis of support vector machine and neural networks in detection of myocardial infarction," *Procedia Comput. Sci.*, vol. 46, pp. 20–30, Jan. 2015.

[10] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 179–185, Feb. 1999.

[11] H. Khorrami and M. Moavenian, "A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5751–5757, Aug. 2010.

[12] L. N. Sharma, R. K. Tripathy, and S. Dandapat, "Multiscale energy and eigenspace approach to detection and localization of myocardial infarction," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1827–1837, Jul. 2015.

[13] P.-C. Chang, J.-J. Lin, J.-C. Hsieh, and J. Weng, "Myocardial infarction classification with multi-lead ECG using hidden Markov models and Gaussian mixture models," *Appl. Soft Comput.*, vol. 12, no. 10, pp. 3165–3175, Oct. 2012.

[14] H. L. Lu, K. Ong, and P. Chia, "An automated ECG classification system based on a neuro-fuzzy system," in *Proc. Comput. Cardiol.*, vol. 27, 2000, pp. 387–390.

[15] K. A. Sidek, I. Khalil, and H. F. Jelinek, "ECG biometric with abnormal cardiac conditions in remote monitoring system," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 44, no. 11, pp. 1498–1509, Nov. 2014.

[16] V. Krasteva, S. Ménétré, J.-P. Didon, and I. Jekova, "Fully convolutional deep neural networks with optimized hyperparameters for detection of shockable and non-shockable rhythms," *Sensors*, vol. 20, no. 10, p. 2875, May 2020.

[17] I.-C. Tanoh and P. Napoletano, "A novel 1-D CCANet for ECG classification," *Appl. Sci.*, vol. 11, no. 6, p. 2758, Mar. 2021.

[18] M. Wasimuddin, K. Elleithy, A. Abuzneid, M. Faezipour, and O. Abuzaghle, "Multiclass ECG signal analysis using global average-based 2-D convolutional neural network modeling," *Electronics*, vol. 10, no. 2, p. 170, Jan. 2021.

[19] M. Långkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.

[20] R. Salloum and C.-C.-J. Kuo, "ECG-based biometrics using recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 2062–2066.

[21] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.

[22] Z. Ahmad and N. M. Khan, "Multi-level stress assessment using multi-domain fusion of ECG signal," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2020, pp. 4518–4521.

[23] H. Dang, M. Sun, G. Zhang, X. Zhou, Q. Chang, and X. Xu, "A novel deep convolutional neural network for arrhythmia classification," in *Proc. Int. Conf. Adv. Mech. Syst. (ICAMEchS)*, Aug. 2019, pp. 7–11.

[24] P. de Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

- [25] Y. Xia and Y. Xie, "A novel wearable electrocardiogram classification system using convolutional neural networks and active learning," *IEEE Access*, vol. 7, pp. 7989–8001, 2019.
- [26] 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.
- [27] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," *Inf. Sci.*, vol. 415, pp. 190–198, Nov. 2017.
- [28] M. Kachuee, S. Fazeli, and M. Sarrafzadeh, "ECG heartbeat classification: A deep transferable representation," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Jun. 2018, pp. 443–444.
- [29] B. Pourbabaee, M. J. Roshkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2095–2104, Dec. 2018.
- [30] R. K. Tripathy, A. Bhattacharyya, and R. B. Pachori, "Localization of myocardial infarction from multi-lead ECG signals using multiscale analysis and convolutional neural network," *IEEE Sensors J.*, vol. 19, no. 23, pp. 11437–11448, Dec. 2019.
- [31] Y. Chen, H. Chen, Z. He, C. Yang, and Y. Cao, "Multi-channel lightweight convolution neural network for anterior myocardial infarction detection," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Oct. 2018, pp. 572–578.
- [32] A. M. Shaker, M. Tantawi, H. A. Shedeed, and M. F. Tolba, "Generalization of convolutional neural networks for ECG classification using generative adversarial networks," *IEEE Access*, vol. 8, pp. 35592–35605, 2020.
- [33] H. Wang, H. Shi, K. Lin, C. Qin, L. Zhao, Y. Huang, and C. Liu, "A high-precision arrhythmia classification method based on dual fully connected neural network," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101874.
- [34] C. Chen, Z. Hua, R. Zhang, G. Liu, and W. Wen, "Automated arrhythmia classification based on a combination network of CNN and LSTM," *Biomed. Signal Process. Control*, vol. 57, Mar. 2020, Art. no. 101819.
- [35] M. Porumb, E. Iadanza, S. Massaro, and L. Pecchia, "A convolutional neural network approach to detect congestive heart failure," *Biomed. Signal Process. Control*, vol. 55, Jan. 2020, Art. no. 101597.
- [36] C. Hao, S. Wibowo, M. Majmudar, and K. S. Rajput, "Spectro-temporal feature based multi-channel convolutional neural network for ECG beat classification," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5642–5645.
- [37] A. T. Oliveira and E. G. O. Nobrega, "A novel arrhythmia classification method based on convolutional neural networks interpretation of electrocardiogram images," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Feb. 2019, pp. 841–846.
- [38] M. M. Al Rahhal, Y. Bazi, H. Almubarak, N. Alajlan, and M. Al Zuair, "Dense convolutional networks with focal loss and image generation for electrocardiogram classification," *IEEE Access*, vol. 7, pp. 182225–182237, 2019.
- [39] A. Diker, Z. Cömert, E. Avcı, M. To açar, and B. Ergen, "A novel application based on spectrogram and convolutional neural network for ECG classification," in *Proc. 1st Int. Informat. Softw. Eng. Conf. (UBMYK)*, Nov. 2019, pp. 1–6.
- [40] W. Liu, M. Zhang, Y. Zhang, Y. Liao, Q. Huang, S. Chang, H. Wang, and J. He, "Real-time multilead convolutional neural network for myocardial infarction detection," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 5, pp. 1434–1444, Sep. 2018.
- [41] X. Zhai and C. Tin, "Automated ECG classification using dual heartbeat coupling based on convolutional neural network," *IEEE Access*, vol. 6, pp. 27465–27472, 2018.
- [42] W. Sun, N. Zeng, and Y. He, "Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network," *IEEE Access*, vol. 7, pp. 67123–67129, 2019.
- [43] E. Izci, M. A. Ozdemir, M. Degirmenci, and A. Akan, "Cardiac arrhythmia detection from 2D ECG images by using deep learning technique," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, Oct. 2019, pp. 1–4.
- [44] B. M. Mathunjwa, Y.-T. Lin, C.-H. Lin, M. F. Abbod, and J.-S. Shieh, "ECG arrhythmia classification by using a recurrence plot and convolutional neural network," *Biomed. Signal Process. Control*, vol. 64, Feb. 2021, Art. no. 102262.
- [45] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.
- [46] R. Wang, J. Fan, and Y. Li, "Deep multi-scale fusion neural network for multi-class arrhythmia detection," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2461–2472, Sep. 2020.
- [47] F. Li, J. Wu, M. Jia, Z. Chen, and Y. Pu, "Automated heartbeat classification exploiting convolutional neural network with channel-wise attention," *IEEE Access*, vol. 7, pp. 122955–122963, 2019.
- [48] A. Uyar and F. Gurgun, "Arrhythmia classification using serial fusion of support vector machines and logistic regression," in *Proc. 4th IEEE Workshop Intell. Data Acquisition Adv. Comput. Syst., Technol. Appl.*, Sep. 2007, pp. 560–565.
- [49] Y. Zhao, X. Yin, and Y. Xu, "Electrocardiograph (ECG) recognition based on graphical fusion with geometric algebra," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 1482–1486.
- [50] R. Wang, Q. Yao, X. Fan, and Y. Li, "Multi-class arrhythmia detection based on neural network with multi-stage features fusion," in *Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC)*, Oct. 2019, pp. 4082–4087.
- [51] N. Manshor, A. A. Halin, M. Rajeswari, and D. Ramachandram, "Feature selection via dimensionality reduction for object class recognition," in *Proc. 2nd Int. Conf. Instrum., Commun., Inf. Technol., Biomed. Eng.*, Nov. 2011, pp. 223–227.
- [52] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 1–7.
- [53] C.-L. Yang, Z.-X. Chen, and C.-Y. Yang, "Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images," *Sensors*, vol. 20, no. 1, p. 168, Dec. 2019.
- [54] J. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *World Sci. Ser. Nonlinear Sci. A*, vol. 16, pp. 441–446, Sep. 1995.
- [55] *RecuPlots and CNNs for Time-Series Classification*. Accessed: Jul. 14, 2021. [Online]. Available: <https://www.kaggle.com/tigurius/recuplots-and-cnns-for-time-series-classification>
- [56] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–7.
- [57] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [58] Z. Ahmad and N. Khan, "CNN-based multistage gated average fusion (MGAF) for human action recognition using depth and inertial sensors," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3623–3634, Feb. 2021.
- [59] H. B. Mitchell, *Image Fusion: Theories, Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2010.
- [60] A. L. Goldberger, L. A. N. 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," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.
- [61] 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.
- [62] R. Boussetlot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-signaldatenbank CARDIODAT der PTB über das internet," *Biomedizinische Technik/Biomed. Eng.*, vol. 40, no. s1, pp. 317–318, 1995.
- [63] *ECG Heartbeat Categorization Dataset*. Accessed: Jul. 14, 2021. [Online]. Available: <https://www.kaggle.com/shayanfazeli/heartbeat>
- [64] *Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. Standard ANSI/AAMI ECG38, Association for the Advancement of Medical Instrumentation, 1998.
- [65] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.
- [66] X. Xu, S. Jeong, and J. Li, "Interpretation of electrocardiogram (ECG) rhythm by combined CNN and BiLSTM," *IEEE Access*, vol. 8, pp. 125380–125388, 2020.

- [67] R. He, Y. Liu, K. Wang, N. Zhao, Y. Yuan, Q. Li, and H. Zhang, "Automatic detection of QRS complexes using dual channels based on U-Net and bidirectional long short-term memory," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1052–1061, Apr. 2021.
- [68] F. Qiao, B. Li, Y. Zhang, H. Guo, W. Li, and S. Zhou, "A fast and accurate recognition of ECG signals based on ELM-LRF and BLSTM algorithm," *IEEE Access*, vol. 8, pp. 71189–71198, 2020.
- [69] J. Kojuri, R. Boostani, P. Dehghani, F. Nowroozipour, and N. Saki, "Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram," *J. Cardiovascular Disease Res.*, vol. 6, no. 2, pp. 51–59, May 2015.
- [70] Y. Cao, T. Wei, N. Lin, D. Zhang, and J. J. P. C. Rodrigues, "Multi-channel lightweight convolutional neural network for remote myocardial infarction monitoring," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Apr. 2020, pp. 1–6.
- [71] M. A. Ahamed, K. A. Hasan, K. F. Monowar, N. Mashnoor, and M. A. Hossain, "ECG heartbeat classification using ensemble of efficient machine learning approaches on imbalanced datasets," in *Proc. 2nd Int. Conf. Adv. Inf. Commun. Technol. (ICAICT)*, Nov. 2020, pp. 140–145.
- [72] E. Akbas and F. T. Y. Vural, "Automatic image annotation by ensemble of visual descriptors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [73] Z. Ahmad and N. Khan, "Towards improved human action recognition using convolutional neural networks and multimodal fusion of depth and inertial sensor data," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2018, pp. 223–230.



**ZEESHAN AHMAD** (Graduate Student Member, IEEE) received the B.Eng. degree in electrical engineering from the NED University of Engineering and Technology, Karachi, Pakistan, in 2001, the M.Sc. degree in electrical engineering from the National University of Sciences and Technology, Pakistan, in 2005, and the M.Eng. degree in electrical and computer engineering from Ryerson University, Toronto, Canada, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His research interests include machine learning, computer vision, multimodal fusion, and signal and image processing.



**ANIKA TABASSUM** received the B.A. degree in computer science from McGill University, in 2013. She is currently pursuing the M.Sc. degree at the Data Science and Analytics Program, Ryerson University. Previously, she has worked as a software engineer/developer for more than four years.



**LING GUAN** (Fellow, IEEE) is currently a Professor of electrical and computer engineering at Ryerson University, Toronto, Canada, where he was a Tier I Canada Research Chair in Multimedia and Computer Technology, from 2001 to 2015. He has published extensively in multimedia processing and communications, human-centered computing, machine learning, adaptive image and signal processing, and more recently, multimedia computing in immersive environment. He is an Elected Member of the Canadian Academy of Engineering and a Distinguished Lecturer of IEEE Circuits and System Society.



**NAIMUL MEFRAN KHAN** (Senior Member, IEEE) is currently an Assistant Professor of electrical and computer engineering at Ryerson University, where he co-directs the Ryerson Multimedia Research Laboratory (RML). His research focuses on creating user-centric intelligent systems through the combination of novel machine learning and human-computer interaction mechanisms. He is a member of ACM. He was a recipient of the Best Paper Award at the IEEE International Symposium on Multimedia, the OCE TalentEdge Postdoctoral Fellowship, and the Ontario Graduate Scholarship.

...