CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings

SAMIUL BASED SHUVO[1], **SHAMS NAFISA ALI**[1], **(Student Member, IEEE),**
SOHAM IRTIZA SWAPNIL[1], **MABROOK S. AL-RAKHAMI**[2], **(Senior Member, IEEE),**
AND ABDU GUMAEI[2,3]

[1]Department of Biomedical Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[2]Research Chair of Pervasive and Mobile Computing, Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[3]Department of Computer Science, Faculty of Applied Science, Taiz University, Taiz, Yemen

Corresponding author: Mabrook S. Al-Rakhami (malrakhami@ksu.edu.sa)

ABSTRACT The alarmingly high mortality rate and increasing global prevalence of cardiovascular diseases (CVDs) signify the crucial need for early detection schemes. Phonocardiogram (PCG) signals have been historically applied in this domain owing to its simplicity and cost-effectiveness. In this article, we propose CardioXNet, a novel lightweight end-to-end CRNN architecture for automatic detection of five classes of cardiac auscultation namely normal, aortic stenosis, mitral stenosis, mitral regurgitation and mitral valve prolapse using raw PCG signal. The process has been automated by the involvement of two learning phases namely, representation learning and sequence residual learning. Three parallel CNN pathways have been implemented in the representation learning phase to learn the coarse and fine-grained features from the PCG and to explore the salient features from variable receptive fields involving 2D-CNN based squeeze-expansion. Thus, in the representation learning phase, the network extracts efficient time-invariant features and converges with great rapidity. In the sequential residual learning phase, because of the bidirectional-LSTMs and the skip connection, the network can proficiently extract temporal features without performing any feature extraction on the signal. The obtained results demonstrate that the proposed end-to-end architecture yields outstanding performance in all the evaluation metrics compared to the previous state-of-the-art methods with up to 99.60% accuracy, 99.56% precision, 99.52% recall and 99.68% F1- score on an average while being computationally comparable. This model outperforms any previous works using the same database by a considerable margin. Moreover, the proposed model was tested on PhysioNet/CinC 2016 challenge dataset achieving an accuracy of 86.57%. Finally the model was evaluated on a merged dataset of Github PCG dataset and PhysioNet dataset achieving excellent accuracy of 88.09%. The high accuracy metrics on both primary and secondary dataset combined with a significantly low number of parameters and end-to-end prediction approach makes the proposed network especially suitable for point of care CVD screening in low resource setups using memory constraint mobile devices.

INDEX TERMS Phonocardiogram analysis, unsegmented heart sound, cardiovascular disease, lightweight CRNN architecture, deep learning, SqueezeNet.

I. INTRODUCTION

Cardiovascular diseases (CVDs), taking away millions of human lives every year, are inciting major concerns in the global healthcare landscape. In 2016 alone, 17.9 million people, almost 31% of entire global deaths occurred from CVDs [1]. In the same vein, lately an alarmingly upward trend of CVD-orinted deaths has been observed in the developing and under-developed countries [2]. Therefore, with a view to slackening the current pace of death toll and minimizing

The associate editor coordinating the review of this manuscript and approving it for publication was Eunil Park.

the adverse impacts of these diseases, early diagnosis and treatment of CVDs are of paramount importance.

Heart, one of the most important organs of the human body, produces distinct sounds during its course of mechanical activity. Since the occurrence of a specific type of disease alters the heart functionality in a definite manner, the auscultations also change accordingly and thus, they have been historically utilized for screening CVDs [3]. Although Lub (S1) and Dub (S2) play pivotal role for the detection of cardiac anomalies [3], sometimes the irregular variants, third heart sound (S3), fourth heart sound (S4) and murmurs are also taken into account while making pathological inference using the stethoscope [4]. The simple, non-invasive nature of this auscultation based diagnosis strategy have made it the most popular and attractive choice of the cardiologists for performing the initial investigation [5]. Besides auscultation, multiple other cardiac signals involving a wide range of advanced methods such as, electrocardiogram (ECG), angiography, echocardiography, myocardial perfusion imaging (MPI), cardiac computed tomography (CCT), cardiovascular magnetic resonance (CMR), carotid pulse graph, apex cardiogram etc. are being utilized as modern diagnostic tools for effective screening of CVDs as they vividly reflect the overall transthoracic physiological conditions of the cardiovascular system [6], [7]. However, in situations where these advanced methods are scarce, lung auscultation stands out as a simple and reliable medium for detecting CVDs. Nevertheless, even for an expert physician, it is quite challenging to readily detect CVDs just by listening to the auscultation. This situation is further exacerbated owing to subjective dissimilar interpretation of the same auscultation by the physicians [8]. The disproportionate number of skilled medical professionals with respect to the total population adds up to the situation and further retard the initial diagnostic speed. In this scenario, artificial intelligence-empowered automated cardiac screening systems on the basis of PCG classification can play a pivotal role to assist the physicians in their decision-making process.

Since phonocardiogram (PCG), the visualization of heart sound on graphical waveform, can be easily processed to extract essential discriminatory features for the identification of cardiac anomalies [9], monitoring heart condition via PCG is becoming an increasingly popular clinical practice. With the amalgamation of machine learning (ML), state-of-the-art networks and advanced audio processing techniques, the unpropitious manual screening can be replaced with automated classification frameworks for prompt large-scale predictions. Therefore, numerous research works have explored the domain of automated PCG classification over the years [10]–[27]. In the literature, a distinct trend of performing 2-step PCG classification can be conspicuously observed. A predominantly trend of performing a segmentation step (localizing lub (S1) and dub(S2) from cardiac auscultation) prior to conducting the classification of the recordings is seen in these works. For materializing the latter step, initially the researchers in the

past have utilized statistical analysis and several ML-based algorithms like, Support Vector Machines (SVM) [3], [13], k-Nearest Neighbor (k-NN) [18], Random Forest [23], Hidden Markov Models [25], Multilayer Perceptron (MLP) [3], [16] etc. in conjunction with many signal processing and feature extraction techniques based on time-domain, frequency-domain and time-frequency domain. Among the feature extraction techniques, fast Fourier transform (FFT), short Fourier transform (STFT), discrete wavelet transform (DWT), continuous wavelet transform (CWT), Q wavelet transform (TQWT), chirplet transform (CT), S-transform, Hilbert–Huang transform (HHT), TF decomposition, Empirical mode decomposition (EMD), Mel-Frequency Cepstral Coefficients (MFCCs), Variance Fractal Dimension (VFD) are observed in many works [9]. Despite being quite promising, most of these ML-based algorithm rely on hand-engineered features in the classifier and thus, are not generalized [19]. Advent of deep learning (DL) architectures namely Convolutional Neural Networks (CNNs), Unsupervised Pre-trained networks (UPNs), Recurrent and Recursive neural networks (RNNs) and their application in the the domain of PCG classification have intuitively resolved both the generalization and the accuracy concern by utilizing the inherent self-learning competency of these networks [14], [15]. However, these deep networks attain the automated feature extraction capability only after going through a computationally complex extensive training phase with a significantly large dataset [28]. The resource-intense requirement of DL-based frameworks makes them unbefitting to be deployed in the low-resource point-of-care locations of the developing and under-developed countries. Therefore, to implement in mobile platforms and wearable devices, reduction of network parameters along with optimized memory footprint is of paramount importance. A handful of cross-domain studies have investigated the issue and introduced a few advanced strategies such as lightweight networks [29], [30], weight quantization [31] and low precision computation techniques [32]. Recently, some studies have successfully implemented these concepts in various biomedical applications including ECG classification [33], respiratory disease classification [28] and most importantly, PCG classification [15].

In this article, we propose CardioXNet, a novel lightweight CRNN framework with two learning scheme i.e., representation learning and sequence residual learning for automatic detection of CVDs from PCG signal as shown in Figure 1. Representation learning includes adaptive feature extractor, frequency feature extractor and pattern extractor which combinedly extract time-invariant salient features from the raw audio of the Github PCG database [18]. Sequence residual learning contains bidirectional connections and is used for extracting temporal features. The use of these schemes enables the network to classify efficiently and converge rapidly on the raw PCG signals. The major contribution of this work is the automatic end to end classification of valvular heart disease from PCG signals using a lightweight CRNN
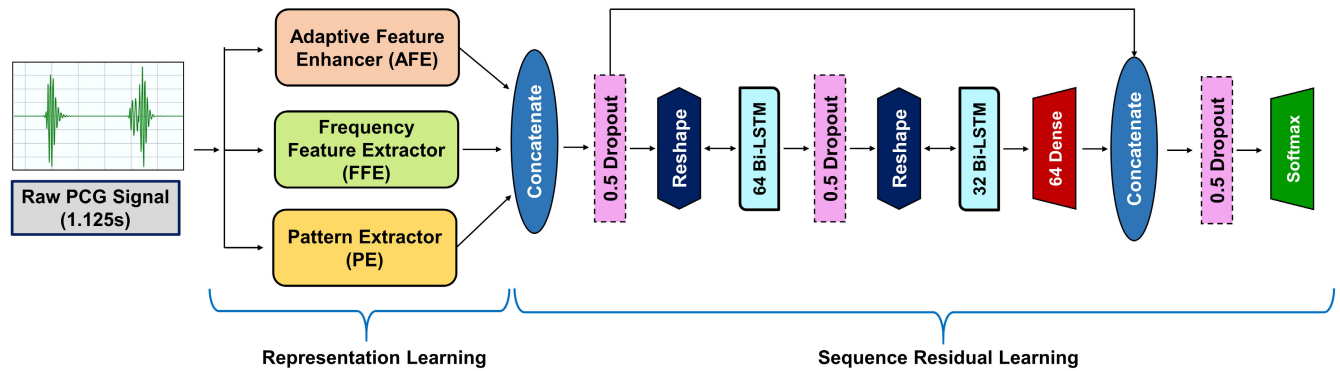
FIGURE 1. An overview architecture of the proposed CardioXNet consisting of 2 phase learning i.e., representation learning and sequence residual learning. The microstructure of the internal modules (AFE, FE and PE) under representation learning phase are shown in figures 5, 6.

network with no manual feature extraction or preprocessing steps like segmentation, augmentation or replication.The capability of the proposed deep lightweight CRNN network to extract salient features directly from PCG signals with minimal training parameters and memory is the main feature of this work. Another highlight is the high performance shown in all matrices. The marked improvement of performance is testament to the effectiveness of feature extraction and classification capability of the proposed combination of 1D squeeze excitation and CRNN networks.

The rest of this article is organized as follows: In Section II, we have discussed the physiological origin of cardiac ausculation and presented an overview of the previous studies on detection of CVDs using PCG siganl. Deep learning network fundamentals are discussed in Section IV; a detailed description of our proposed architecture, along with the intuitions behind designing it is presented in Section V; the implementation settings, evaluation criteria and experimental results are discussed with reference to the literature in Section VI; the future prospects of the work along with the existing limitations are mentioned in Section VII and finally, we conclude our study in Section VIII.

II. BACKGROUND

In the literature, numerous works have been reported on multi-faceted PCG classification. Among them, a vast majority of the existing works have focused on performing normal vs. pathological (abnormal) prediction of the PCG [3], [5], [12]–[15]. Again, some have explored the domain of multi-class (normal, artifact, extrasystole, extrahls, murmurs) cardiac anomaly prediction [6], [10], [11] while some have done multi-class (innocent, systolic, diastolic, pansystolic, ejection) classification of murmurs, the most efficient indicator of CVD incidence [26], [27]. Being motivated by the unparalleled success of the aforementioned schemes, several research works from all over the world are being geared towards direct classification of cardiovascular diseases utilizing cardiac ausculation. By elucidating PCG data, to date

researchers have successfully classified many CVDs such as, coronary artery disease (CAD), atrial fibrillation (AF), aortic stenosis (AS), dilated cardiomyopathy (DC), mitral stenosis (MS), papillary muscle dysfunction (PMD), mitral regurgitation (MR), atrial septal defect (ASD), ventricular septal defect (VSD), mitral valve prolapse (MVP) etc. [16]–[24].

A. PHYSIOLOGICAL ORIGIN OF CARDIAC AUSCULTATION

Cardiac auscultation is a complex sound signal generated by the mechanical activities of the heart. It contains the impact of vibration of the valvular opening and closing, systolic and diastolic movement of muscles and tendons and the internal blood perfusion [34]. However, there are two types of fundamental heart sounds (FHSs). At the onset of the systolic phase, when the atrioventricular (AV) mitral and tricuspid valves close, the first FHS, "lub" (S1) can be heard [35]. The definitive abnormality in S1 might indicate the occurrence of MS, AF and ASD [36]. S1 is followed by the second FHS, "dub" (S2) at the beginning of the diastolic phase from the closure of the pulmonic and aortic valves. This relatively high-pitched sound demonstrates higher frequency [37] in case hypertension and AS during an event of cardiac abnormality.

In normal cardiac condition, distinct pattern of systolic and diastolic period can be observed. Nevertheless, 3rd and 4th heart sound, respectively S3 and S4, murmurs, gallops, clicks might occur because of CVD incidence [34]. While S3 mostly indicates reduced systolic function and congestive heart failure (except for pregnant ladies, athletes and young people); S4 is an clear indicator of diastolic dysfunction. The audible whooshing sounds i.e., murmurs might be generated because of the stiffening of the heart valves, narrowing of the blood vessels and turbulent flow [37]. Murmurs are very significant for classifying CVDs, as their diverse pitch and frequency pattern represent many CVDs including MR, AR, AS, MS etc. [36]. However, innocent murmurs are sometimes produced depending on the physiological condition of the individual.

B. LITERATURE REVIEW

As mentioned earlier, the automatic detection of CVDs and heart valve disorders (HVDs) have been explored in multiple previous works. Some of these works have utilized publicly available data while some implemented the task with in-house datasets. In this subsection, we will provide a detailed overview of some of the prominent works in this domain.

Employing divergence analysis on 50 wavelet features with MLP classifier, Dokur and Ölmez [16] has achieved up to 99% accuracy in determining discriminatory features of VSD, MR, AR, AS, MS, diastolic rumble, summation gallop, aortic insufficiency and some other cardiac abnormalities using a self-constructed multi-source data. In [17], septal defects were detected along with other heart defects like valvular defects, ventricular hypertrophy, constrictive pericarditis etc from publicly a available PCG database using TQWT based sub-bands decomposed from segmented heart disease. The classification was performed using least squares support vector machine (LS-SVM) algorithm and achieved overall accuracy of 99.03%. Using the Github PCG database [18], a multivariate comprehensive analysis for detecting 5 classes, i.e., MS, AS, MR, MVP, N has been carried out in [18]. Discrete Wavelets Transform (DWT) and Mel Frequency Cepstral Coefficient (MFCCs) have been employed as feature extraction technique while SVM, deep neural network (DNN) and centroid displacement based K-NN classifier have been adopted as classifier. The features and classifiers are attempted both single-handily and in an ensemble manner through rigorous experimentation. Merging the DWT and MFCC features, centroid displacement based K-NN, DNN and SVM respectively yields an accuracy of 97.4%, 92.1% and 97.9% [18]. In the same vein, using the same dataset, a recent study has proposed WaveNet, a novel 1D CNN network having residual blocks with dilated 1D CNN and multiple skip connections [19]. Upon 10-fold cross-validation, they have achieved training accuracy of 97% and approximately 90% validation accuracy. Along with sensitivity 92.5% and specificity 98.1%, the model exhibits satisfactory performance. A comprehensive basis for CAD diagnosis has been provided in [20] employing a novel multi-domain feature fusion framework. For the in-house dataset used, the concerted feeding of the selected MFCC features with the deep features into a MLP classifier has obtained a promising accuracy of 90.43%.

Several studies have utilized PCG signals for classifying heart valve disorders (HVDs). In [21], an in-house dataset of 120 PCG signals was used for HVD classification using wavelet packet(WP) decomposition and SVM. The classification task was performed on 2 classes (Normal and HVDs) achieving specificity of 96.67%. Another work [22] involving in-house dataset, has classified normal, pulmonary and mitral stenosis heart valve diseases utilized discrete fourier transform (DFT) and Burg autoregressive (AR) spectrum analysis for feature extraction.

Using Principal Component Analysis (PCA) for dimensionality reduction and ANN for classification, the work has achieved maximum accuracy score of 95%. Maintaining the legacy, Ghosh *et. al* [23] has classified the HVD classes i.e., MR, AR, AS, N using magnitude and phase features, wavelet synchrosqueezing transform (WSST) based time-frequency matrix with a Random Forest classifier on the Github PCG database [18] dropping the MVP class. The proposed paradigm has demonstrated the mean accuracy to be 95.13%. In a later work by the same author [24], using Chirplet transform (CT), local energy (LEN) and local entropy (LENT) features in conjunction with a composite classifier, an overall accuracy of 98.33% was obtained on the same dataset.

In all of these works, separate preprocessing steps have been performed for feature extraction. However, in one of the recent studies, a CNN-LSTM model [38] has been used for direct classification using the PCG dataset with 1D wavelet smoothing, achieving 99.32% accuracy. The use of such models without manual preprocessing steps are vital for any end to end mobile health care devices.

III. MATERIALS
A. DATASET DESCRIPTION

In this study, PCG recordings used in the article [18] has been primarily utilized. The recordings were collected from various sources like books and websites and contained a total number of 1000 PCG recordings in.wav format in five different classes i.e., Normal (N), Aortic stenosis (AS), Mitral regurgitation (MR), Mitral stenosis (MS), Mitral valve prolapse (MVP). Each of the classes has 200 recordings for roughly 3s. All the recordings are sampled at 8 kHz. Since the lowest signal length present in this dataset is 1.125s, all the recordings are truncated from the start of the recording up to 1.125s. Figure 2 illustrates the waveform for each of the disease classes.

Additionally, to validate the PCG classification performance of our proposed network, the openly accesible "PhysioNet/CinC Challenge 2016 dataset [39]" is incorporated as a secondary dataset that contains a total number of 3240 PCG recordings within six separately labeled datasets. Seven different research groups have concertedly collected the PCG signals, at either clinical or nonclinical settings. The recordings are originally sampled at 2 kHz and have varying duration (5s-120s). Nevertheless, this dataset does not contain disease-based annotation and contains only 2 classes (normal, abnormal), where the normal annotated 2575 recordings refer to healthy subjects with no valvular defects and the remaining 665 abnormal recordings indicate pathological cases like arrhythmia, coronary heart disease, valvular stenosis, mitral regurgitation etc. For maintaining analogy with the primary dataset and preserving fidelity of the proposed model, only the first 1.125s from each of the PCG recordings of PhysioNet dataset were taken as well.
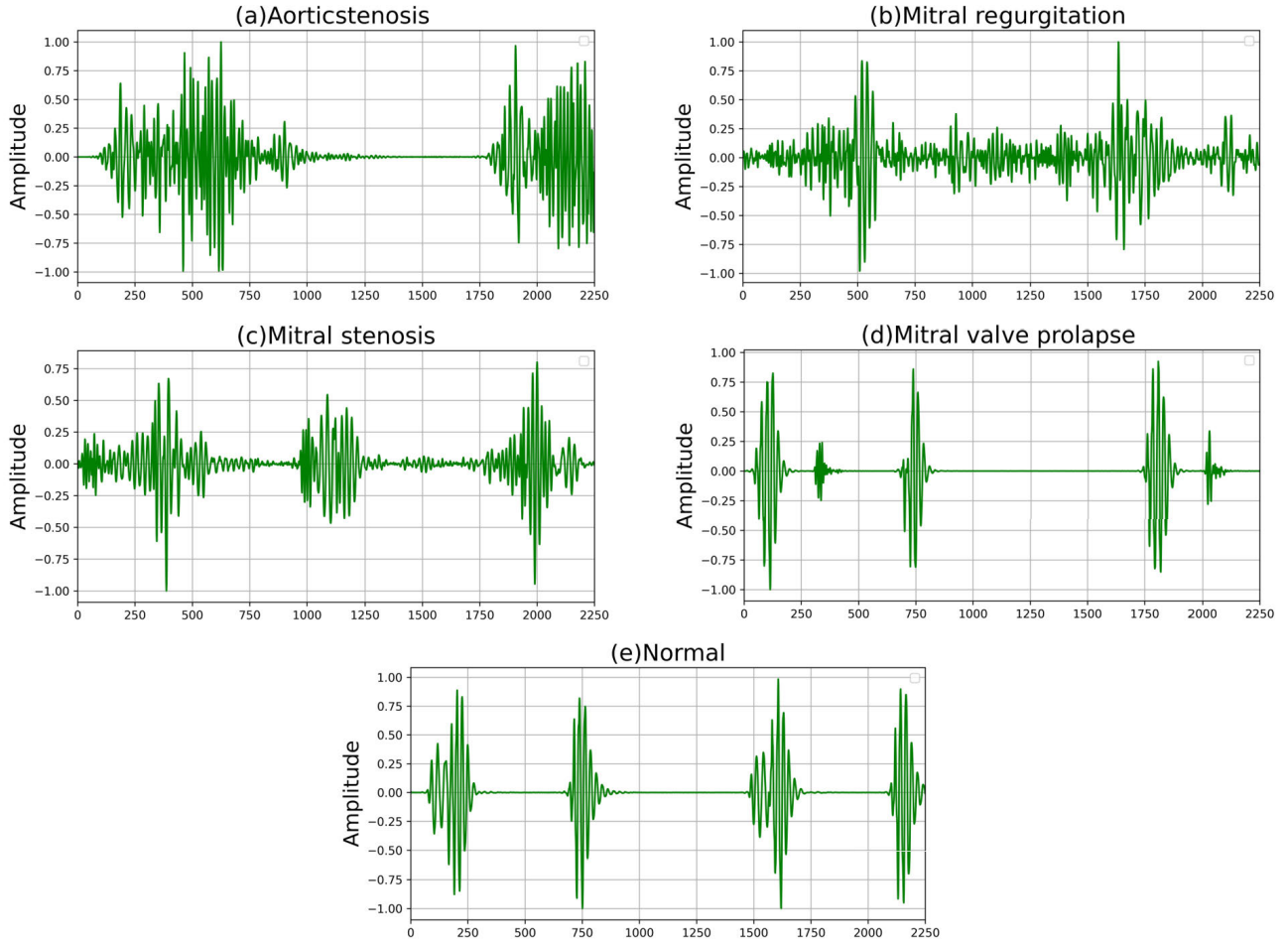
FIGURE 2. Waveform of the existing CVD classes on PCG dataset. (a) Aortic stenosis (AS), (b) Mitral regurgitation (MR), (c) Mitral stenosis (MS), (d) Mitral valve prolapse (MVP) and (e) Normal.

B. PREPROCESSING

All the PCG signals are resampled at 2 kHz that ensures the preservation of important heat sound frequency components as well as lowers the computational cost. The signals are also amplitude normalized to ensure consistency and homogeneity by nullifying the effect of acquisition device/sensor variation. The visible clusters in Figure 3 shows feature distributions corresponding to different CVD class along with normal class.

IV. NETWORK FUNDAMENTALS

A. CONVULATIONAL NEURAL NETWORK (CNN)

Convolutional neural networks, a variant of state-of-the-art artificial neural networks, mainly combine three architectural generalization ideas by ensuring different invariance techniques such as, shared weights, local receptive fields and the subsequent presence of sub-sampling of spatial or temporal channels with a view to reducing the complexity of the network model. Each of the CNN layer consists of a number of neural units and receives a set of extracted features as input from the previous layer. CNNs for images, speech,

and time-series signals have different sets of kernel formation and they map the input into different types of features from each location [40].

1) 1D CNN

In spite of being analogous to the regular neural network (ANN), 1D CNN generally takes raw sequential data i.e., time-series data like PCG, EEG as input. The main advantage of 1D CNN over ANN is that it can show excellent performance on time-series data without engaging any time or resource intensive feature extraction steps.

Assuming the input of 1D CNN is a tensor representing the audio waveform denoted as X, feature extraction of 1D convolution can be defined as,

$$T = F(X|\Theta) = f_L(\ldots f2(f1(X|\Theta_1)|\Theta_2)|\Theta_L) \quad (1)$$

Here, L is the number of hidden layers, Θ is a set of parameters that maps the input to the prediction T.

The l-th convolutional layer's operation can be defined as:

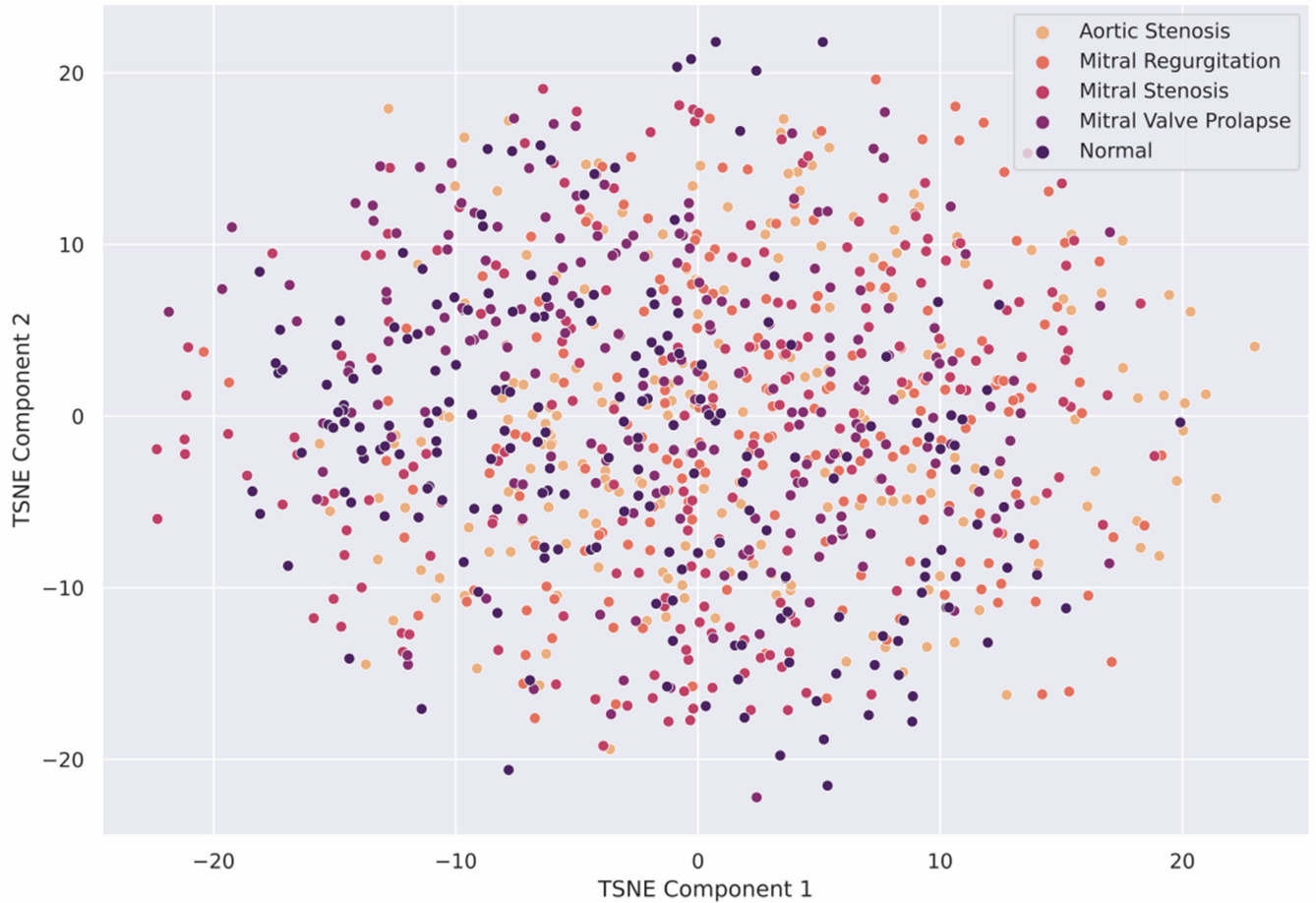$$T_l = f_l(X_l|\Theta_l) = h(W \otimes X_l + b), \Theta_l = [W, b] \quad (2)$$

FIGURE 3. t-Stochastic Neighbour Embedding (t-SNE) visualization of normal class and four CVDs class Aortic stenosis (AS), Mitral regurgitation (MR), Mitral stenosis (MS) Mitral valve prolapse (MVP) heart sound data.

Here, $\otimes$ denotes the convolution operation, X_l is a two-dimensional input matrix of N feature maps, W is a set of N one dimensional kernels (receptive fields) used for extracting a new set of features from the input array, b is the bias vector, and h $(\cdot)$ is the activation function. The shapes of X_l, W and T_l are respectively (N, d), (N, m) and (N, d−m + 1) [41].

2) 2D CNN

2D CNN is generally used for extracting spatial features using its kernel. It has multiple hidden layers, each layer having a number of neural units. The neural unit is basically a function connecting inputs and outputs. The input function can be denoted as [42]:

$$x_k^n = \sum_{n=1}^{N_{n-1}} conv2D(w_{ik}^n, s_i^{n-1}) + b_k^n \qquad (3)$$

here, s_i^{n-1} represents the output of the ith neuron of (n-1)th layer and b_k^n and $(w_{ik}^n$ are the bias and kernel of the kth neuron in the nth layer, accordingly. conv2D represents a two-dimensional regular convolution with valid padding.

The input function of the kth neuron can be denoted as [42]:

$$y_k^n = f\left(\sum_{n=1}^{N_{n-1}} conv2D(w_{ik}^n, s_i^{n-1}) + b_k^n \right) \qquad (4)$$

B. RECURRENT NEURAL NETWORK (RNN)

RNN is a special kind of ANN model where the input samples contain more inter-dependencies. RNN has a significant specialisation at storing the information of the past time steps. In other words, the parameters available at time t_1+1 will be affected by output produced at time t_1 [43]. RNN is mainly used for sequential and dynamic information processing and thus, it is very useful for multiple types of audio signal processing i.e., music and speech processing, pathological auscultation processing etc. The Generalized delta rule [44] and energy minimization function are two of the major algorithms used for RNN training [45].

1) LONG-SHORT TERM MEMORY

The Long-Short Term Memory (LSTM), proposed by Hochreiter and Schmidhuber in 1997 is a primary building
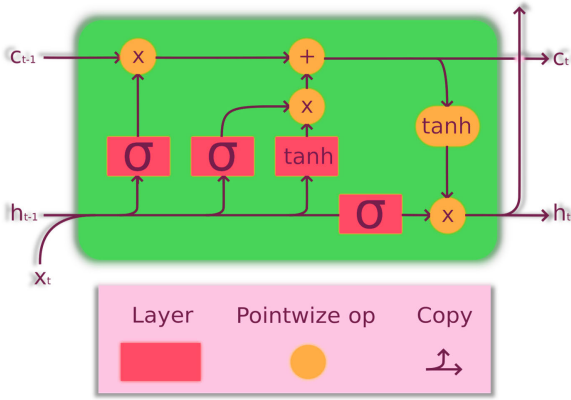
FIGURE 4. The structure of the LSTM cell.

block of RNN. It has been substantially explored for sequence learning. A standard LSTM cell (see Figure 4) consists of three distinct gates and three main cell-memory networks. The respective equations for each of these gates and networks are presented below:

Input gate:

$$i_t = Sigmoid(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + bi) \quad (5)$$

Forget gate:

$$f_t = Sigmoid(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + bf) \quad (6)$$

Output gate:

$$o_t = Sigmoid(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + bo) \quad (7)$$

Input cell:

$$g(t) = tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \quad (8)$$

Memory cell:

$$c(t) = f_t \odot c_{t-1} + i_t \odot g_t \quad (9)$$

Output cell:

$$h(t) = o_t \odot tanh(c_t) \quad (10)$$

Here, sigmoid and tanh are the activation functions which map the non-linearity of the features [46].

2) BIDIRECTIONAL LSTM

The main disadvantage of one-directional LSTM is that it is only encompasses the past information. To overcome this limitation of the one-directional LSTM, a bi-directional LSTM (Bi-LSTM) approach has been introduced by Schuster and Paliwal [47] for efficiently investigating both the past and the future context. The Bi-LSTM is defined as,

$$\overrightarrow{h} = \sigma(x_t U + \overrightarrow{h}_{t-1} W + b_t) \quad (11)$$

$$\overleftarrow{h} = \sigma(x_t U + \overleftarrow{h}_{t-1} W + b_t) \quad (12)$$

Here, $\overrightarrow{h}$ is the forward hidden state and $\overleftarrow{h}$ is the backward hidden state. Concatenation of both of these states formed the hidden state at time t.

C. MICRO-ARCHITECTURE OF FIRE MODULE

The Fire module, composed of two layers i.e., Squeeze and Expansion layer, is a dedicated parameter reduction module introduced in the SqueezeNet architecture [48]. In the Squeeze layer, 3×3 kernels have been replaced with 1×1 kernels which results in 9x reduction in the number of parameters. Then, the squeezed features are fed into an expansion layer. The expansion layer consists of a layer with 1×1 filters and a layer with 3×3 filters and the extracted features of these layers concatenated together in the channel dimension. By maintaining 8x fewer output channels in the squeeze layer compared to the expansion layer, squeeze ratio (SR) of 0.125 has been achieved.

V. NETWORK ARCHITECTURE

The proposed network namely CardioXNet is constructed in a modular way, employing two major learning paradigms i.e., representation learning and sequence residual learning. This architecture is designed for effectively classifying short unsegmented PCG recordings. Figure 1 portrays the high-level network architecture.

A. REPRESENTATION LEARNING

This part is trained with a view to extracting time-invariant features from each of the raw PCG signals. It consists of three parallel CNN pathways namely, Frequency Feature Extractor (FFE), Pattern Extractor (PE) and Adaptive Feature Enhancer (AFE) [43], [48].

1) FREQUENCY FEATURE EXTRACTOR (FFE)

FFE consists of four 1D convolutional layers and two max-pooling layers with the primary filter size of sampling frequency (Fs) $\times$ 4 and stride size set to Fs/2 for the 1D convolutional (conv1) layer to capture the frequency components. The subsequent filter and stride sizes are chosen by rigorous hyperparameter tuning (see Figure 5).
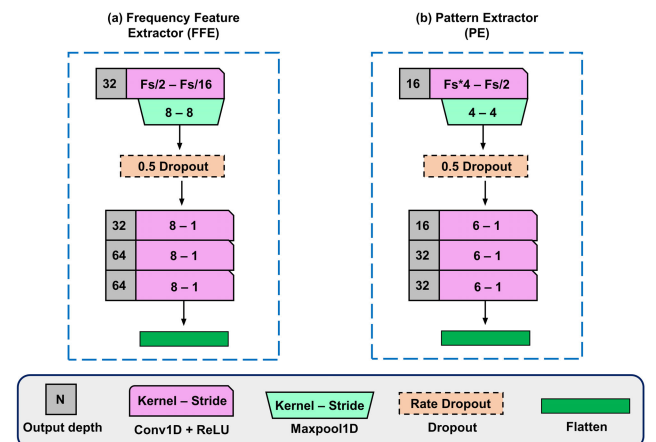

FIGURE 5. Micro structure of (a) Frequency Feature Extractor (FFE); (b) Pattern Extractor (PE). Fs represents sampling frequency.

2) PATTERN EXTRACTOR (PE)

Similar to FFE, PE also consists of four 1D convolutional layers and two max-pooling layers. However, fine-grained convolution with filter and stride size set to Fs/2, and Fs/16 respectively are used which helps to recognise the appearance of various PCG pattern. With the help of hyper-parameter tuning, the filter and stride sizes are chosen (see Figure 5).

3) ADAPTIVE FEATURE ENHANCER (AFE)

The input sequences are reshaped into a 2D tensor and fed to AFE. AFE consists of 2D convolutional layers, batch-normalization, max-pooling layers and squeeze-expansion layers, inspired by the Fire module of SqueezeNet architecture [48]. Batch normalization layers helps to stabilize and speed up the training process [49] while abstract feature maps generated from the initial convolutional layers are then passed to the squeeze-expansion layers which results in lowering of parameter count by manifolds(see Figure 6). Apart from efficient feature extraction, it also ensures faster convergence in the training phase, a highly appreciated task from the perspective of training deep networks in mobile platforms.
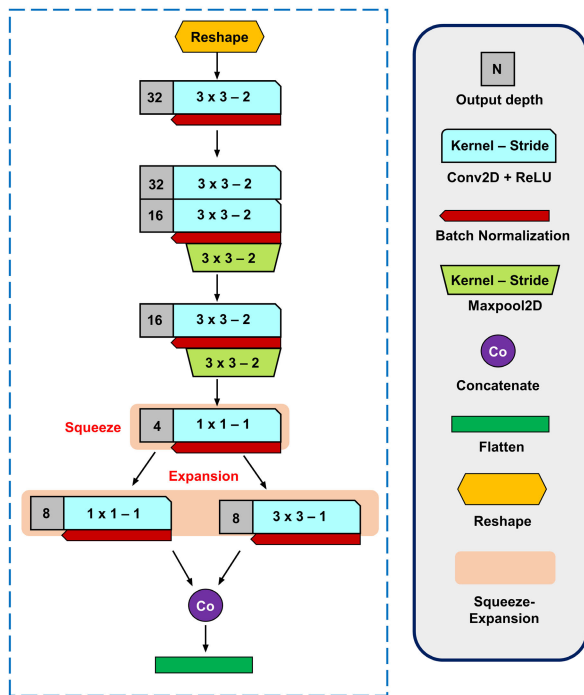

FIGURE 6. Micro structure of Adaptive Feature Enhancer (AFE).

Outputs from three CNN paths are concatenated together and forwarded to the sequence residual learning part.

$$X_{\text{concat}} = X_{\text{AFE}} \oplus X_{\text{FFE}} \oplus X_{\text{PE}} \qquad (13)$$

Here, X_{concat} is the concatenated feature vector and X_{AFE}, X_{FFE}, X_{PE} are the extracted feature vectors from three CNN paths.

B. SEQUENCE RESIDUAL LEARNING

Sequence residual learning is trained to extract the temporal information from the sequence of extracted features in the representation learning part. Two layers of bi-LSTMs have been employed to learn temporal information which enables the encoding of both past and future information by processing with two independent LSTMs. Since PCG is a time-series data, the application of Bi-LSTM is best suited for capturing the sequential information embedded in them.

A skip connection has been employed to implement residual functionality and enabling the addition of temporal information and previously extracted features from the CNNs.The concatenated feature vector is fed into a prediction layer with probability nodes, calculated by the softmax function.

In all the convolutional layers and LSTM layers, Rectified Linear Unit (ReLU) [50] activation function has been used, which is defined as:

$$f(x) = max(0, x), \qquad (14)$$

It has been employed to diminish the vanishing gradient issue and also contributes to the faster convergence. For minimizing the over-fitting phenomenon, dropout layers have been employed.

VI. EXPERIMENTS

A. EVALUATION METRICS

The dataset was divided into training, validation and testing parts with 70% of the PCG signal used for training the model and 10% of the signal used for validation and remaining 20% data used for testing. 10-fold cross validation was performed randomly on the dataset for generailization of the obtained results. Well-known and important performance evaluation metrics such as, Accuracy, Precision, Recall and F1- score are chosen to evaluate the proposed framework quantitatively.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \qquad (15)$$

$$Precision = \frac{TP}{TP + FP} \qquad (16)$$

$$Recall = \frac{TP}{TP + FN} \qquad (17)$$

$$F1\ score = \frac{2 * TP}{2 * TP + FP + FN} \qquad (18)$$

B. EXPERIMENTAL SETUP

The CardioXNet is designed using Keras and TensorFlow backend and training work has been performed using NVidia K80 GPUs provided by Kaggle notebooks.

Sparse categorical crossentropy loss and adaptive learning rate optimizer (Adam) with the learning rate of 0.00001 have been engaged in the training process and a batch size of 16 is selected for training and validation purpose.

C. PERFORMANCE OF THE PROPOSED FRAMEWORK

In this work, the proposed lightweight model has been evaluated on the Github PCG datst, PhysioNet/CinC challenge

TABLE 1. Results of 10-fold cross-validation.

Fold iteration	Accuracy	Recall	Precision	F1-score
1	0.995	0.995	0.997	0.996
2	0.99	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00
4	0.995	0.996	0.995	0.995
5	0.990	0.984	0.984	0.991
6	0.995	0.997	0.995	1.00
7	0.995	0.996	0.995	0.995
8	1.00	1.00	1.00	1.00
9	0.995	0.994	0.994	0.995
10	0.995	0.996	0.996	0.996
Mean	0.996	0.9952	0.9956	0.9968
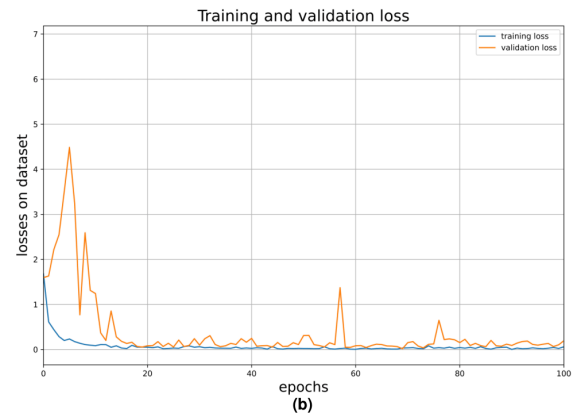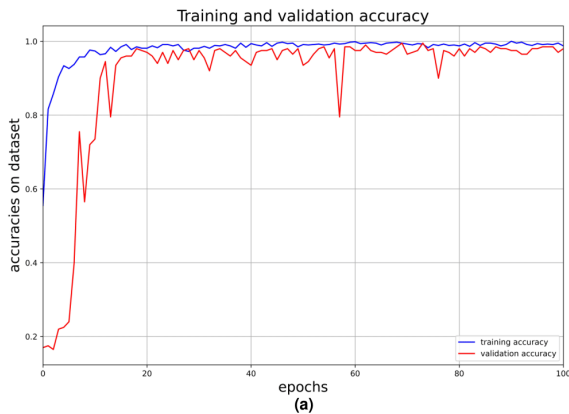
FIGURE 7. (a) Epoch vs accuracy of the CRNN during training and validation (b) Epoch vs loss of the CRNN during training and validation for 10th cross validation step.

TABLE 2. Classification Report for different classes using CardioXNet.

Class	Precision	Recall	F1-score
AS	0.9733($\pm$0.028)	0.9951($\pm$0.010)	0.9840($\pm$0.014)
MR	0.9730($\pm$0.023)	0.9807($\pm$0.024)	0.9776($\pm$0.023)
MS	0.9899($\pm$0.013)	0.9846($\pm$0.021)	0.9842($\pm$0.016)
MVP	0.9949($\pm$0.011)	0.9696($\pm$0.019)	0.9821($\pm$0.014)
N	1.0000($\pm$0.000)	1.0000($\pm$0.000)	1.0000($\pm$0.000)

dataset and on both of these dataset combined. In the following sections a detailed explanation of these results are discussed.

1) PERFORMANCE ON GitHub PCG DATASET

The PCG dataset has been classified into 5 different classes using CardioXNet. The proposed model achieved near perfect validation accuracy on the given dataset, on 10 fold cross validation, showing very high precision and recall scores for all the classes (see Table. 1 and Table. 2). Furthermore, Figure 7 (a) depicts that the model converged with great

rapidity reaching 100% training accuracy, while validation accuracy reached 99.5% within 100 epochs. Additionally, Figure 7 (b) provides an insight into the minimization of the categorical cross entropy loss. The incorporation of adaptive feature extractor block with the squeeze and expansion layers mainly attributed to this faster convergence of the model.

2) PERFORMANCE ON PhysioNet/CinC 2016 DATASET

To demonstrate the effectiveness of the proposed model, classification between healthy and unhealthy classes were performed on a secondary dataset, namely PhysioNet/CinC 2016 challenge dataset. The confusion matrix of the best performing models for both GitHub PCG dataset and secondary dataset is shown in Figure 8 and 9. From Figure 9, we can see, the proposed framework performs quite well on PhysioNet dataset showing around 86.57%($\pm$2.31) accuracy, 93.28%($\pm$2.10) precision, 90.33%($\pm$1.90) recall and 91.78%($\pm$1.99) F1 Score. This remarkable performance on a secondary dataset proves the excellent capability of the proposed model to differentiate features without any pre-processing steps like segmentation and augmentation.
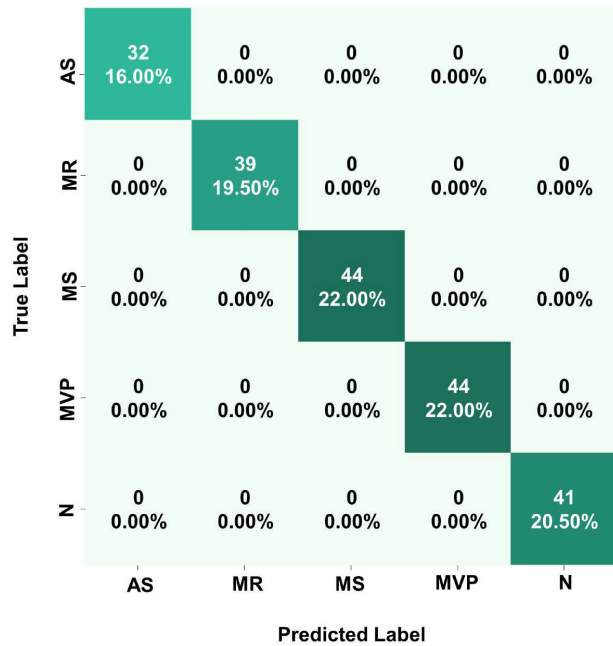
FIGURE 8. Confusion matrix of the GitHub PCG dataset.

FIGURE 10. Confusion matrix of the best model on PhysioNet/CinC 2016 and GirHub PCG mixed dataset.

The four diseased class data from GitHub PCG database was merged with unhealthy data from PhysioNet dataset while Normal class data from GitHub PCG dataset was matched with healthy dataset from PhysioNet dataset. Overall the mixed dataset consists of 4240 (3240 + 1000) PCG recordings in which 2757 (2575+200) recordings are normal and 1465 (665+800) are abnormal. The total recordings available for training are 2968 while 424 and 848 recordings were respectively, used for testing purpose.Two class classification performance on the mixed dataset is shown in Figure 10. CardioXNet showed overall accuracy of 88.09%, 88.08% precision, 87.98% recall and 88.03% F1 Score. This excellent performance shows the superior feature extraction and classification performance of our proposed model.

D. COMPARISON WITH THE EXISTING WORKS

A detailed comparative analysis among the proposed CardioXNet and the previous works have been presented in Table 3. In [18], manually crafted MFCC, DWT and MFCC-DWT features were extracted for feeding into the classifiers i.e., KNN, SVM and DNN. Although multiple analysis were performed adopting several feature extraction and classifier combinations, MFCC-DWT based conjoined features demonstrated the best performance, giving 97.4%, 97.9% and 92.1% accuracy respectively, on KNN, SVM and DNN classifiers. Another work proposing a novel deep WaveNet architecture has achieved a training accuracy of 97% and validation accuracy of 90% for 5 class CVD classification [19].Furthermore, two recent studies on the classification of HVDs have utilized this dataset dropping the MVP class. The first work has achieved the mean accuracy
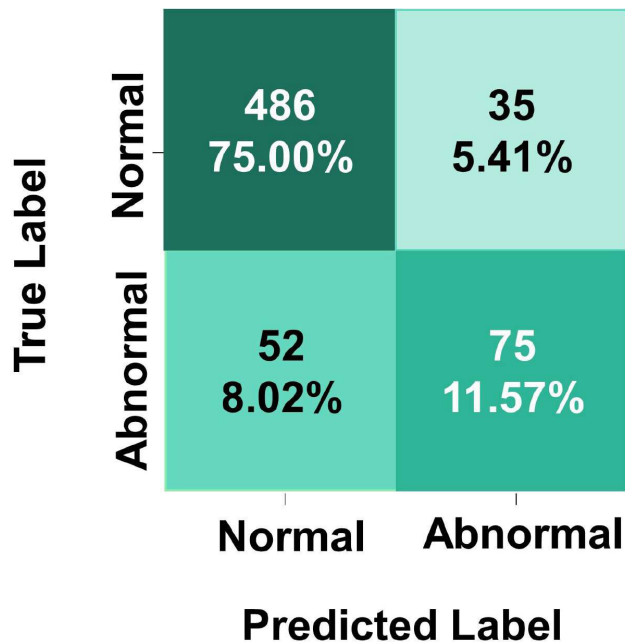

FIGURE 9. Confusion matrix of PhysioNet/CinC 2016 dataset.

The difference in accuracy between the GitHub PCG dataset and PhysioNet dataset might be attributed to the presence of noise and variation in recording devices and variable sources within the PhysioNet dataset.

3) PERFORMANCE ON PhysioNet-GitHub MERGED DATASET

The generalization potential of the proposed CardioXNet has been tested using a merged dataset of PhysioNet and Github PCG recordings for training, validation and testing.
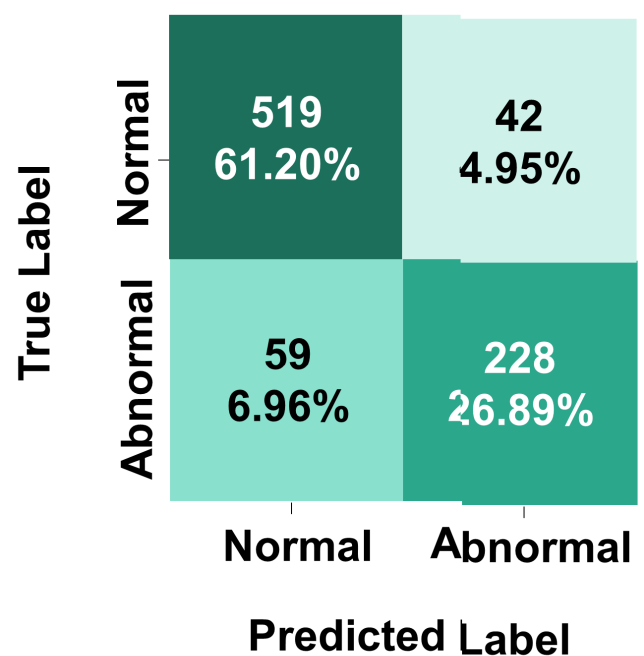
TABLE 3. Comparison between previous works and CardioXNet on CVD classification.

Feature Extraction	Model	Accuracy	F1-score
$MCC + DWT$ [18]	KNN	97.4	99.2
	DNN	97.9	99.7
	SVM	92.1	98.3
-	WaveNet [19]	90.0	-
$WSST$ [23]	Random Forest	95.13	-
$Chirplet\ Transform$ [24]	MCC	98.33	-
$1D\ wavelet\ smoothing,\ z-score\ normalization$ [39]	CNN-LSTM	99.32	98.30
-	**CardioXNet (Proposed)**	**99.6**	**99.4**

of 95.13% using wavelet synchrosqueezing transform function (WSST) for extracting the magnitude and phase features from the PCG and classified the signals with the Random Forest (RF) classifier [23] while the second one contains several complex processing steps. After performing filtering and Shannon energy based cardiac cycle segmentation, local energy and local entropy features were extracted in conjunction with Chirplet transform (CT). Upon the evaluation of the features using multi-class composite classifier (MCC), the framework has obtained overall accuracy of 98.33% [24]. In another recent work CNN-LSTM model is used on PCG recordings from the same dataset with 1D wavelet smoothing and z-score normalization achieving relatively high accuracy and F1 score of 99.32% and 98.30% respectively. Therefore, it is clearly evident that the proposed CardioXNet architecture outperforms or performs on par with the previous works, achieving an accuracy of 99.6%. Individual class-wise evaluation metrics also show similar trend of superior performance in comparison with all the existing works.

The strategies adopted in [18], [23], [24] require several computationally complex and time-consuming hand-engineered feature extraction techniques on the PCG dataset, while the CardioXNet is an end-to-end model which is capable of directly predicting CVDs from the raw PCG signal. Thus, the proposed scheme is well-performing and highly efficient with reduced complexity and latency in terms of making automated predictions. Although the WaveNet model can optimally classify the raw PCG recordings with relatively lower training parameters, it fails to perform as well as the proposed CardioXNet despite having advanced network and deep feature extraction components. Rather, it lagged by around 9.6% accuracy in the validation phase in comparison to our proposed model. This clearly marks the lacking of robustness of the WaveNet model. In addition, the model is ambiguously represented with no clear depiction of kernels and layers; thereby, making the network irreproducible.

The case of [38] is different as in that work the primary focus was the classification of HVD directly from the dataset with minimal processing. However steps such as 1D wavelet

smoothing and z-score normalization was performed while in our proposed work raw PCG signals were used in the model. Moreover the proposed model in [38] is vague and important details such as architecture and number of parameters is absent. Whereas, detailed architecture of CardioXNet and its computational complexity is detailed intricately in this article.

From this exhaustive analysis, it can be inferred that the proposed CardioXNet manifests significant improvement over all the exsting state-of-the-art works using the GitHub PCG database.

E. COMPUTATIONAL EFFICIENCY OF THE PROPOSED MODEL

In this part, a detailed bench-marking is performed on the CardioXNet, considering trainable parameters, floating point operations per second (FLOPS) required end-to-end time and size after training the parameters. The proposed lightweight CNN model has extremely low end to end time of 54.60($\pm$0.06) ms. Since the proposed paradigm can directly classify the raw PCG recordings requiring no preprocessing and hand crafted feature extraction technique which evidently demonstrates the robustness of the model. Moreover, the proposed model has 0.67 M trainable parameters, 26 M FLOPS and a smaller memory requirement of only 7.96 MB. Therefore, it can achieve trade-off between higher accuracy than the considered networks while ensuring the significantly lower necessity of storage space and processing power. This makes the CardioXNet more suitable for real-time wearable and mobile applications [51].

VII. LIMITATIONS AND SCOPES FOR FUTURE WORKS

Although the current framework demonstrates satisfactory performance on two well-known PCG datasets representing real-life scenarios, the work can be further improved if multiple larger PCG datasets are available with diverse CVD annotations. The number of PCG data of HVD is also limited and patient number and independence is not provided within the dataset.

Several advanced network strategies like Generative Adversarial Networks (GAN), Variational Autoencoder (VAE) can be attempted to obtain further optimized performances with more efficiency and non-latency. In future, we aim to integrate our CardioXNet in digital stethoscopes or wearable devices with a cloud server connection for performing automatic PCG classification and predicting on different CVDs using our pre-trained model in real time to assist the clinicians in their diagnostic decision.

VIII. CONCLUSION

In this work, a novel lightweight CRNN network, CardioXNet, has been proposed for automatic detection of different cardiovascular disease classes without performing any preprocessing steps on the PCG signal. The model involves both representation learning and sequence residual learning by utilizing both the CNN and bi-LSTM layers to extract time-invariant and temporal features. The framework demonstrated state-of-the-art performance with an accuracy score of 99.6% outperforming all the previous works on the given PCG dataset by a considerable margin. Besides showing superior results, the use of the proposed end-to-end network with significantly low number of parameters make this model well-qualified for usage in any embedded system applications. We believe that this work can contribute in the advancement of real-time, automated CVD classification from PCG recordings and can have significant real life impacts in the domain of clinical diagnostics.

REFERENCES

[1] *CVD Causes One-Third of Deaths Worldwide: Study Examines Global Burden of CVD From 1990 to 2015—American College of Cardiology.* Accessed: Aug. 26, 2020. [Online]. Available: https://www.acc.org/latest-in-cardiology/articles/2017/05/17/16/02/cvd-causes-one-third-of-deaths-worldwide

[2] A. D. Bowry, J. Lewey, S. B. Dugani, and N. K. Choudhry, "The burden of cardiovascular disease in low-and middle-income countries: Epidemiology and management," *Can. J. Cardiol.*, vol. 31, no. 9, pp. 1151–1159, 2015.

[3] A. Bourouhou, A. Jilbab, C. Nacir, and A. Hammouch, "Heart sounds classification for a medical diagnostic assistance," *Int. J. Online Biomed. Eng.*, vol. 15, no. 11, pp. 88–103, 2019.

[4] V. N. Varghees and K. I. Ramachandran, "A novel heart sound activity detection framework for automated heart sound analysis," *Biomed. Signal Process. Control*, vol. 13, pp. 174–188, Sep. 2014.

[5] H. Alaskar, N. Alzhrani, A. Hussain, and F. Almarshed, "The implementation of pretrained AlexNet on PCG classification," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2019, pp. 784–794.

[6] F. Chakir, A. Jilbab, C. Nacir, and A. Hammouch, "Phonocardiogram signals processing approach for Pascal classifying heart sounds challenge," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1149–1155, Sep. 2018.

[7] Q.-U.-A. Mubarak, M. U. Akram, A. Shaukat, and A. Ramazan, "Quality assessment and classification of heart sounds using PCG signals," in *Applications of Intelligent Technologies in Healthcare.* Cham, Switzerland: Springer, 2019, pp. 1–11.

[8] U. Alam, O. Asghar, S. Q. Khan, S. Hayat, and R. A. Malik, "Cardiac auscultation: An essential clinical skill in decline," *Brit. J. Cardiol.*, vol. 17, no. 1, p. 8, 2010.

[9] S. Ismail, I. Siddiqi, and U. Akram, "Localization and classification of heart beats in phonocardiography signals—a comprehensive review," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, p. 26, 2018.

[10] S.-W. Deng and J.-Q. Han, "Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps," *Future Gener. Comput. Syst.*, vol. 60, pp. 13–21, Jul. 2016.

[11] O. Deperlioglu, U. Kose, D. Gupta, A. Khanna, and A. K. Sangaiah, "Diagnosis of heart diseases by a secure Internet of health things system based on autoencoder deep neural network," *Comput. Commun.*, vol. 162, pp. 31–50, Oct. 2020.

[12] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 8, pp. 2189–2198, Jan. 2020.

[13] W. Zhang, J. Han, and S. Deng, "Heart sound classification based on scaled spectrogram and partial least squares regression," *Biomed. Signal Process. Control*, vol. 32, pp. 20–28, Feb. 2017.

[14] S. Latif, M. Usman, R. Rana, and J. Qadir, "Phonocardiographic sensing using deep learning for abnormal heartbeat detection," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9393–9400, Sep. 2018.

[15] B. Xiao, Y. Xu, X. Bi, J. Zhang, and X. Ma, "Heart sounds classification using a novel 1-D convolutional neural network with extremely low parameter consumption," *Neurocomputing*, vol. 392, pp. 153–159, Jun. 2020.

[16] Z. Dokur and T. Ölmez, "Feature determination for heart sounds based on divergence analysis," *Digit. Signal Process.*, vol. 19, no. 3, pp. 521–531, 2009.

[17] S. Patidar, R. B. Pachori, and N. Garg, "Automatic diagnosis of septal defects based on tunable-Q wavelet transform of cardiac sound signals," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3315–3326, May 2015.

[18] Y. Khan, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, p. 2344, Nov. 2018.

[19] S. L. Oh, V. Jahmunah, C. P. Ooi, R.-S. Tan, E. J. Ciaccio, T. Yamakawa, M. Tanabe, M. Kobayashi, and U. Rajendra Acharya, "Classification of heart sound signals using a novel deep WaveNet model," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105604.

[20] H. Li, X. Wang, C. Liu, Q. Zeng, Y. Zheng, X. Chu, L. Yao, J. Wang, Y. Jiao, and C. Karmakar, "A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103733.

[21] S. Choi, "Detection of valvular heart disorders using wavelet packet decomposition and support vector machine," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1679–1687, Nov. 2008.

[22] H. Uğuz, "A biomedical system based on artificial neural network and principal component analysis for diagnosis of the heart valve diseases," *J. Med. Syst.*, vol. 36, no. 1, pp. 61–72, Feb. 2012.

[23] S. K. Ghosh, R. K. Tripathy, R. N. Ponnalagu, and R. B. Pachori, "Automated detection of heart valve disorders from the PCG signal using time-frequency magnitude and phase features," *IEEE Sensors Lett.*, vol. 3, no. 12, pp. 1–4, Dec. 2019.

[24] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals," *Comput. Biol. Med.*, vol. 118, Mar. 2020, Art. no. 103632.

[25] H. M. Fahad, M. U. Ghani Khan, T. Saba, A. Rehman, and S. Iqbal, "Microscopic abnormality classification of cardiac murmurs using ANFIS and HMM," *Microsc. Res. Techn.*, vol. 81, no. 5, pp. 449–457, May 2018.

[26] P. Wang, C. S. Lim, S. Chauhan, J. Y. A. Foo, and V. Anantharaman, "Phonocardiographic signal analysis method using a modified hidden Markov model," *Ann. Biomed. Eng.*, vol. 35, no. 3, pp. 367–374, 2007.

[27] V. N. Varghees and K. I. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors J.*, vol. 17, no. 12, pp. 3861–3872, Jun. 2017.

[28] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," 2020, *arXiv:2009.04402*. [Online]. Available: http://arxiv.org/abs/2009.04402

[29] M. S. Hussain and M. A. Haque, "SwishNet: A fast convolutional neural network for speech, music and noise classification and segmentation," 2018, *arXiv:1812.00149*. [Online]. Available: http://arxiv.org/abs/1812.00149

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[31] S. Seo and J. Kim, "Efficient weights quantization of convolutional neural networks using kernel density estimation based non-uniform quantizer," *Appl. Sci.*, vol. 9, no. 12, p. 2559, 2019.

[32] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 6869–6898, 2017.

[33] Q. Wu, Y. Sun, H. Yan, and X. Wu, "Ecg signal classification with binarized convolutional neural network," vol. 2020, Art. no. 103800.

[34] A. K. Dwivedi, S. A. Imtiaz, and E. Rodriguez-Villegas, "Algorithms for automatic analysis and classification of heart sounds—A systematic review," *IEEE Access*, vol. 7, pp. 8316–8345, 2019.

[35] D. Mason, *Listening to Heart: A Comprehensive Collection Heart Sounds Murmurs*. Philadelphia, PA, USA: FA Davis Company, 2000.

[36] *Abnormal Heart Sounds and Murmurs—62 Rationale | Medical Council of Canada*. [Online]. Available: https://mcc.ca/objectives/expert/key/62/

[37] S. Barma, B.-W. Chen, K. Lok Man, and J.-F. Wang, "Quantitative measurement of split of the second heart sound (S2)," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 12, no. 4, pp. 851–860, Jul. 2015.

[38] M. Alkhodari and L. Fraiwan, "Convolutional and recurrent neural networks for the detection of valvular heart diseases in phonocardiogram recordings," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105940.

[39] C. Liu *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological Meas.*, vol. 37, no. 12, p. 2181, 2016.

[40] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998, pp. 255–258.

[41] S. Abdoli, P. Cardinal, and A. L. Koerich, "End-to-end environmental sound classification using a 1D convolutional neural network," *Expert Syst. Appl.*, vol. 136, pp. 252–263, Dec. 2019.

[42] Y. Chen, S. Wei, and Y. Zhang, "Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network," *Med. Biol. Eng. Comput.*, vol. 58, no. 9, pp. 2039–2047, Sep. 2020.

[43] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.

[44] M. Huk, "Backpropagation generalized delta rule for the selective attention sigma-if artificial neural network," *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 2, pp. 449–459, 2012.

[45] N. Kalchbrenner and P. Blunsom, "Recurrent continuous translation models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2013, pp. 1700–1709.

[46] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[47] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[48] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with $50\times$ fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[50] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: http://arxiv.org/abs/1505.00853

[51] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, "Real-time human detection as an edge service enabled by a lightweight CNN," in *Proc. IEEE Int. Conf. Edge Comput. (EDGE)*, Jul. 2018, pp. 125–129.

SAMIUL BASED SHUVO is currently pursuing the B.Sc. degree in biomedical engineering with the Bangladesh University of Engineering and Technology, Bangladesh. He is currently working on biomedical signal processing and deep learning-based health informatics projects. His other research interests include biomedical simulation, biomedical instrumentation, and edge device optimization.

SHAMS NAFISA ALI (Student Member, IEEE) is currently pursuing the B.Sc. degree in biomedical engineering with the Bangladesh University of Engineering and Technology, Bangladesh. Her current research interests include biomedical signal and image processing, computer-aided diagnosis, deep learning, medical imaging, smart wearable systems, and the aspects of biocompatible materials through the involvement of artificial intelligence.
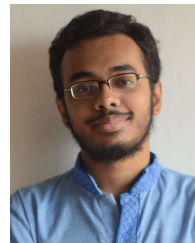
SOHAM IRTIZA SWAPNIL is currently pursuing the B.Sc. degree in biomedical engineering with the Bangladesh University of Engineering and Technology, Bangladesh. He is currently working on using signal processing and machine learning on biomedical signals for use in low resource settings. His other interest includes the world of material sciences and using engineering knowledge for creating better health care solutions.

MABROOK S. AL-RAKHAMI (Senior Member, IEEE) received the master's degree in information systems from King Saud University, Riyadh, Saudi Arabia, where he is currently pursuing the Ph.D. degree with the Department of Information Systems, College of Computer and Information Sciences. He was a Lecturer with King Saud University, Muzahimiyah Branch, and taught many courses, such as programming languages in computer and information science. He has authored several papers in peer-reviewed IEEE/ACM/Springer/Wiley journals and conferences. His research interests include edge intelligence, social networks, cloud computing, the Internet of Things, big data, and health informatics.

ABDU GUMAEI received the Ph.D. degree in computer science from King Saud University, Riyadh, Saudi Arabia, in 2019. He was a Lecturer with the Department of Computer Science, Taiz University, Yemen, and taught many courses, such as programming languages. He is currently an Assistant Professor with the College of Computer and Information Sciences, King Saud University. He has authored and coauthored more than 30 journal and conference papers in well-reputed international journals. He received a patent from the United States Patent and Trademark Office in 2013. His research interests include software engineering, image processing, computer vision, machine learning, networks, and the Internet of Things.

• • •