IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 8, 2021, accepted August 4, 2021, date of publication August 9, 2021, date of current version August 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3103316

Phonocardiogram Signal Based Multi-Class Cardiac Diagnostic Decision Support System

SHAMIK TIWARI[1], ANURAG JAIN[1], AKHILESH KUMAR SHARMA[2], (Member, IEEE), AND KHALED MOHAMAD ALMUSTAFA[3]

[1]Department of Virtualization, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand 248007, India
[2]Department of Information Technology, Manipal University Jaipur, Jaipur, Rajasthan 303007, India
[3]Department of Information Systems, College of Computer and Information Science, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding authors: Anurag Jain (anurag.jain@ddn.upes.ac.in) and Akhilesh Kumar Sharma (akhileshshm@gmail.com)

ABSTRACT A Phonocardiogram (PCG) signal represents murmurs and sounds signals made by vibrations caused for the period of a cardiac cycle. Acoustic wave generated through the beat of the cardiac cycle propagates through the chest wall. It can be easily recorded by a low-cost small handheld digital device called a stethoscope. It provides information like heart rate, intensity, tone, quality, frequency, and location of various components of cardiac sound. Due to these characteristics, phonocardiogram signals can be used to detect heart status at an early stage in a non-invasive manner. In previous studies, the Convolutional Neural Network (ConvNet) is the most studied architecture, which was fed by features, namely Mel Frequency Cepstral (MFC), Chroma Energy Normalized Statistics (CENS), and Constant-Q Transform (CQT). This work has proposed a ConvNet model trained by Hybrid Constant-Q Transform (HCQT) for heart sound beat classification. CQT, Variable-Q Transform (VQT), and HCQT are extracted from each phonocardiogram signal as the acoustic features, including the dominant MFCC features, feed into five-layer regularized ConvNets. After analyzing the literature in the same domain, it can be stated that this is the first time HCQT is being utilized for PCG signals. The findings of the experiments demonstrate that HCQT is more effective than standard CQT and other variants. Also, the accuracies of the system proposed in this work on the validation datasets are 96% in multi-class classification, which outperforms the proposed work relative to other models significantly. The source code is available on the Github repository https://github.com/shamiktiwari/PCG-signal-Classification-using-Hybrid-Constant-Q-Transform to support the research community.

INDEX TERMS Cardiovascular disease, convolutional neural network, decision support system deep learning, multi-class classification, phonocardiogram signal.

I. INTRODUCTION

As per the fact sheet available with WHO, CVD claims the lives of around 17.9 million people each year, and it is 31% of total death in a year, which makes CVD disease the number one cause of death. Most deaths due to CVD occur in middle and low-income countries where medical facilities are either not easily available or very costly [1]. Diagnose at an early stage is the only way to decrease the death rate due to CVD. There are many invasive and non-invasive methods to diagnose CVD. All Invasive techniques are costly, painful, and readily unavailable at all places, especially in remote areas. Usage of a non-invasive method to diagnose CVD at an early stage is less expensive and painless. ECG and PCG are two such non-invasive ways to diagnose CVD. But their analysis requires an expert doctor of this domain which is not readily available in remote areas [2]. When sounds and murmurs occur during the cardiac cycle are represented diagrammatically, it is called a phonocardiogram. These vibrations generate the wave, which propagates through the chest wall. A stethoscope, a low-cost handheld digital device, is used to record the information generated through acoustic waves. It gives us an estimation of parameters like heart rate, intensity, tone, quality, frequency, and location of various components of the cardiac sound, which helps in the diagnosis of CVD in a non-invasive manner [3]. Recent advances in computing have enabled researchers to design decision support systems that can be utilized to diagnose CVD at an early stage, even in the absence of an expert. Machine learning and deep learning algorithms have allowed us to create decision

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

support systems that can help doctors and can also be used by laypeople in the absence of doctors [4].

The authors have proposed a hybrid constant-Q transform-based classification model to acquire more detailed information from PCG signals in this work. Acoustic features from the PCG signal are fetched to the ConvNet model for learning. The following are the key contributions of the proposed work:

- Propose hybrid constant-Q transform-based (HCQT) acoustic features for PCG signals.
- Compare the HCQT features to other acoustic features and recommend the best feature set for PCG signal classification.

The following is the paper's structure: Discussion of different models found in the literature for automatic diagnosis of CVD from PCG is given in Section 2. Details of sound features used with the model for classification, classifier, an insight view of the proposed model, and features of the phonocardiogram signal dataset used for the training and testing of the designed model are given in Section 3. Detail of the simulation environment and result generated through the proposed model are given in Section 4. Discussion and analysis of results are presented in Section 5. It is ended with the conclusive remarks given in Section 6.

II. LITERATURE REVIEW

An overview of different types of automatic heart disease diagnostic models from PCG signal along with datasets used and accuracy level achieved by them is given below in Table 1.

Though in the last five years, a lot of research has been carried out in designing of automatic heart disease diagnosis model from PCG signal, yet there are many more areas that are yet to be explored. It has motivated us for the proposed model given in section 3.

III. MATERIAL AND METHODS

This section presents a detailed overview of sound feature extraction methods, classification model, the dataset used, and proposed model utilized in this work.

A. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCCs)

In audio or speech signal processing, The short-term power spectrum of sound is represented by MFC. It is based on a non-linear Mel frequency scale and a linear cosine translation of the logarithmic power spectrum. Collectively MFCCs coefficients make up MFC. The feature extraction process of MFCC is composed of the following steps [23], [24]:

1. Pre-emphasis: It amplifies high frequencies by passing phonocardiogram signals from a high pass filter.
2. Framing: Phonocardiogram signals are separated into overlapping frames. It is implemented to fetch local spectral properties.
3. Windowing: It is implemented on frames for the minimization of discontinuities around edges. An example of a widely used technique is Hamming windowing.
4. Discrete Fourier Transformation: DFT is applied to the sound signal after the third step to obtain the frequency domain signal from the time domain.
5. Mel-Frequency Warping: It's used to calculate the quantity of energy that occurs in various locations of a frequency domain. Mel in this case is a pitch unit. A pitch of 1000 Mels is a pure tone at 1000Hz with a 40 dB strength over the listener's threshold. Mel-scale is used to determine this non-linear frequency result, as presented in (1).

$$M(f) = 1125log\left(1 + \frac{f}{700}\right) \tag{1}$$

Here, the frequency term is denoted by f, while the Mel-scale frequency is denoted by M(f).

6. Discrete Cosine Transform and Log Compression: In this step, the logarithmic function IFFT is applied on filtered bank energies received in step 5. The DCT follows it. Finally, MFCC(n) is computed as shown in (2).

$$MFCC(n) = \frac{1}{T}\sum_{r=1}^{R} log\left[MF(t)\right] cos\left[\frac{2\pi}{T}\left(r + \frac{1}{2}\right)n\right] \tag{2}$$

where MFCC(n) is the nth MFCC coefficient derived from specific audio sections using T triangular filters, and MF(t) is the t-th filter's Mel-spectrum. The heartbeat spectrogram obtained by MFCC is shown in Fig. 1.

B. CONSTANT-Q TRANSFORM (CQT), VARIABLE-Q TRANSFORM, AND HYBRID CONSTANT-Q TRANSFORM (HCQT)

J.C. Brown, in 1988 has introduced CQT. It refers to a technique that transforms a signal from time to frequency domain. However, it is different from Fourier transformation as central frequencies are geometrically spaced, and corresponding Q-factors are equal. CQT is defined as a 1/24 octave filter bank, but it is not restricted to 24 only; it can be varied to 12, 36, or 48 bins per octave also. Unlike DFT, central frequencies of analysis are not uniformly distributed but aligned with equally tempered scale notes; this makes CQT suitable for the processing of sound [25], [26]. Furthermore, the frequency resolution of CQT has a constant Q-factor, which effectively improves resolution accuracy in low-frequency regions. Under the N-th frame of CQT, the frequency component of the K-th semitone can be stated in (3).

$$X_n^{cqt}(k) = \frac{1}{N}\sum_{m=0}^{N_k-1} x(m) w_{N_k}(m) e^{-j2\pi mQ/N_k} \tag{3}$$

where Q is a constant whose value depends on the number of spectral lines of a single octave (β)

$$Q = \frac{1}{2^{1/\beta} - 1}$$

The ability of the constant-Q transform to provide equal frequency support to all semitones and a variable number

TABLE 1. An overview of PCG signal based heart disease diagnosis models.

Important Features	Classifier	Dataset	Accuracy
A. E. F. Malik et al. [5] have designed a Scalogram and CNN-based model to diagnose cardiovascular disease from PCG signals. They have applied the segmentation method [7] to convert each PCG signal from three cardiac cycles to one cardiac cycle. It has increased the length of the dataset. To avoid any loss of frequency and time in the signal, the authors have applied the continuous wavelet transformation in place of Fourier transformation. It has generated scalogram images of size 656 × 875. Through a bicubic interpolation algorithm, the authors have resized the images into 227 × 227 and then feed them to the 2D ConvNet model for final classification. Authors have found that with the increase of the number of the convolutional layer, there is an increase in the accuracy.	2D-CNN	The authors have used the heart sound data set collected by Yaseen et al. [6], and PCG Dataset has 800 abnormal and 200 normal heartbeat sound recordings	99.40% of accuracy in multi-class classification. 99.88% of accuracy in binary class classification The highest accuracy was achieved for five convolution layer ConvNet. Only three classes are considered.
P. Upretee and M. Yuksel [7] have used time-varying spectral features with different classifiers to classify PCG signals in different classes	SVM KNN	The heart sound data set was collected by Yaseen et al. [6].	96.5% of accuracy in multi-class classification with KNN and 99.6% of accuracy in binary class classification with KNN
S. Patidar et al. [8] have designed a model to detect septal defects by analyzing cardiac sound signals. The authors have used the TQWT based advanced signal processing technique to fetch cycles of the heartbeat from cardiac sound signals. Further authors have decomposed the segmented signals through TQWT and used this combination of decomposed signals to extract diagnostic features. It has helped them in the characterization of different types of murmur noise.	LS-SVM	Heart sound signal [9-13]	98.92% in binary classification
A. Gharehbaghi et al. [14] have suggested a classifier to diagnose aortic stenosis (AS), and pulmonary stenosis (PS) among children through their PCG signal.	SVM	Recorded PCG signal of 45 children at the medical center of Tehran university hospital.	93.3% in binary classification
O. Deperlioglue et al. [15] have proposed a decision support system based on IoHT to detect cardiovascular disease through heart sound.	AEN	Pascal [16] PhysioNet [17]	100% accuracy with the Pascal [16] dataset while with the PhysioNet [17] dataset accuracy of 99.8% in binary classification.
M. Banerjee and S. Manjhi [18] have discussed the significance of early detection of heart disease in decreasing the mortality rate. The authors have designed a machine learning-based model to detect heart disease from PCG signals.	2D-CNN	Pascal [16]	83% in multi-class classification.
S. B. Shuvo et al. [19] proposed a hybrid classifier by incorporating the characteristics of representation learning and sequence residual learning to detect CVD from PCG signals. They have used representation learning for the extraction of time-invariant features from PCG signals. For extraction of temporal features, they have used sequential residual learning	CRNN	Heart sound data set collected by Yaseen et al. [6]	99.6% in binary classification.
G. Redlarski et al. [20] have designed a hybrid classifier using SVM and the Cuckoo search algorithm. The proposed model automatically diagnoses heart disease from PCG signals by using the LPC feature extraction approach.	Hybrid of SVM & Cuckoo	A public database of heart sound [21]	93% in multi-class classification
P. Narvaez et al. [22] have used modified EWT for preprocessing of PCG signal. Normalized Shannon average energy was used for segmentation in a single cardiac cycle. They have used six power features for classification	SVM KNN Random forest	Pascal [16]	An accuracy of 99.26% with KNN classifiers in binary classification.

of bins among them is its main advantage. However, it has drawbacks, one of which being the absence of consistent temporal resolution at lower frequencies. This trade-off can be alleviated by introducing variants of CQT i.e., VQT and HCQT. When compared to the CQT transformation, the VQT transformation provides better temporal resolution at lower frequencies. A new parameter is introduced to allow for an equitable drop of the bins' Q-factors as they approach low frequencies [27], [28].

$$B_k = \alpha f_k + \gamma$$

When $\gamma = 0$, the Q-factor in the constant-Q situation is a constant. The additional parameter γ might be understood as a Hertz offset, and it is normally set to be as low as possible, e.g., around 30 Hz. Instinctively, γ has a stronger relative influence at lower frequencies where the bandwidth

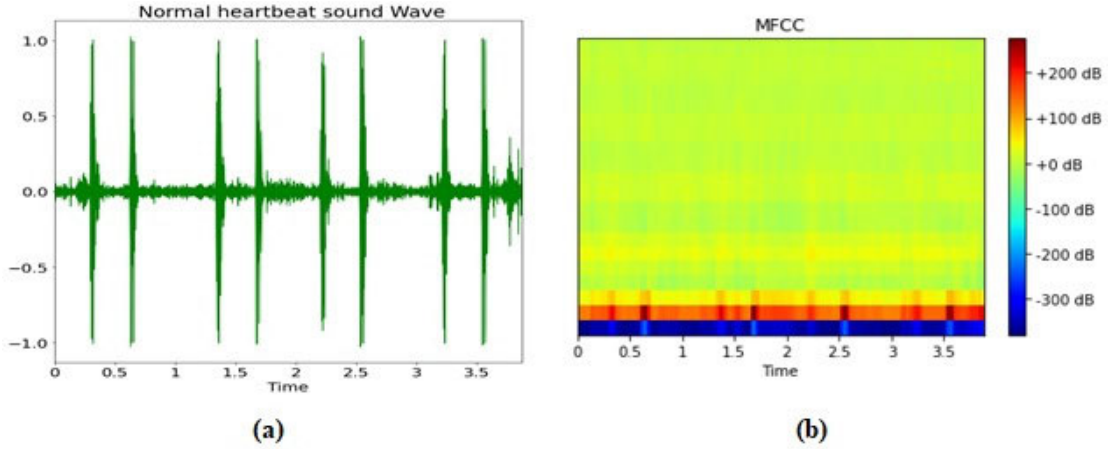

FIGURE 1. (a): A sample waveform for normal phonocardiogram signal, (b): heat map visualization for MFCC of a PCG signal segment. Sliding windows, x, and filter-bank frequencies, y, are represented on the horizontal and vertical axes. MFCC energy information, Ex,y, is represented by pixel color in the heat map. The MFCC is generated with the number of frequency bins = 84 and hop length = 51.

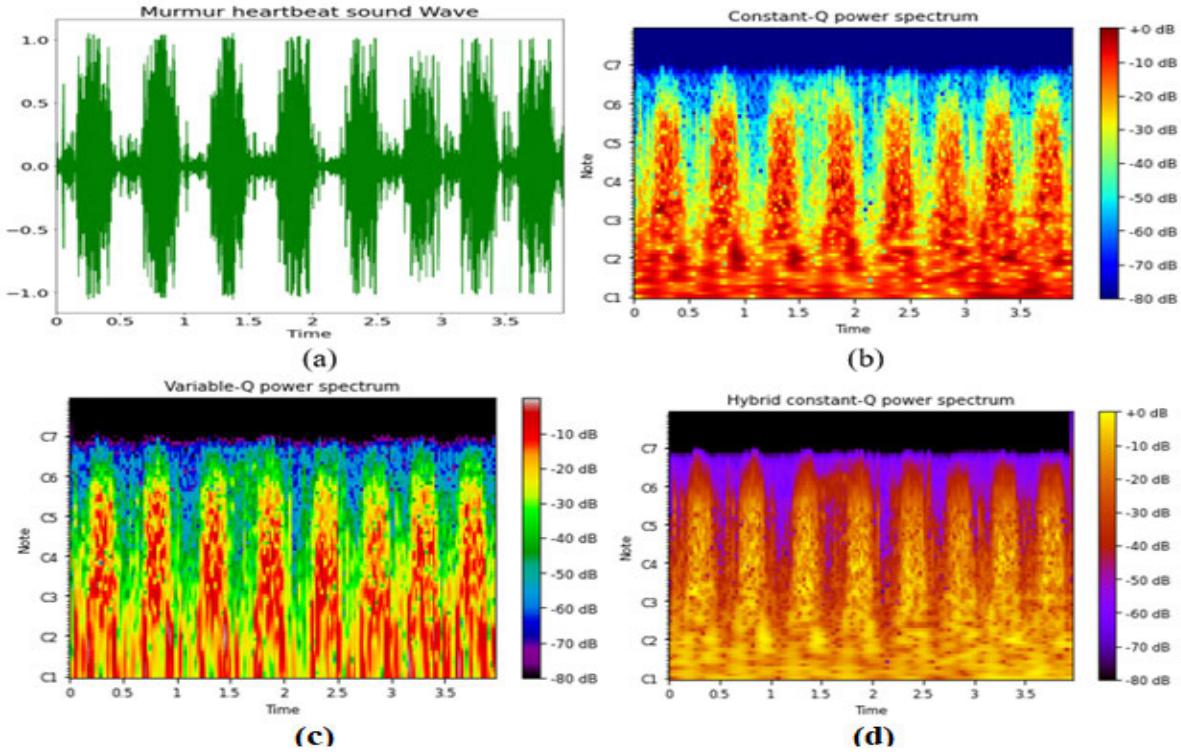

FIGURE 2. (a): A sample waveform fo murmur phonocardiogram signal, (b-d): heat map visualization for CQT, VQT, and HCQT base spectrograms, respectively. Sliding windows, x, and filter-bank frequencies, y, are represented on the horizontal and vertical axes. MFCC energy information, Ex,y is represented by pixel color in the heat map. The MFCC is generated with the number of frequency bins = 84 and hop length = 512.

is insufficient, but fades at higher frequencies. Hybrid CQT, on the other hand, is made up of two CQT varieties. In the temporal domain, the frameshift is thought to include L samples. Then, select the k_c-th filter that fulfills the condition $N[k_c] = 2L$ [29], [30].

High frequencies are those that exceed f_kc, whereas low frequencies are those that are less than f_kc. The high-frequency section of hybrid CQT uses the filter bank of the high-frequency part of CQT to filter the short-term Fourier transform-based spectrogram. The regular CQT is used directly for the low-frequency section of HCQT. In compared to CQT, HCQT is more computationally capable. A visualized comparison of the CQT, VQT, and HCQT is presented in Fig. 2.

C. CONVOLUTIONAL NEURAL NETWORK (ConvNet)

CNN has brought the revolution in the domain of computer vision. It has remarkably achieved better results than the traditional classification algorithms. Deep learning is a sub-class of machine learning which is based on Deep Neural Networks

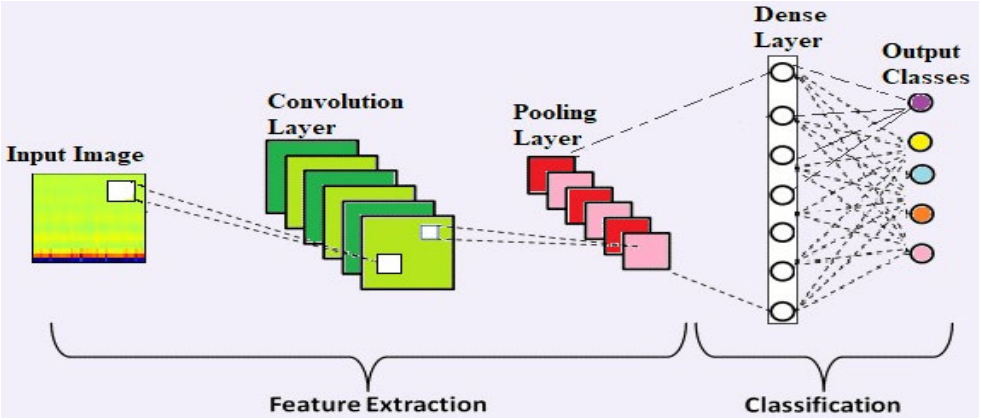

FIGURE 3. The architecture of convolution neural network and spectrogram-based phonocardiogram signal classification model. Inputs are the spectrograms generated through MFCC, CQT, VQT, and HCQT, and output is one of the five classes i.e., artifact, extrahls, extra-systole, murmur, and normal.

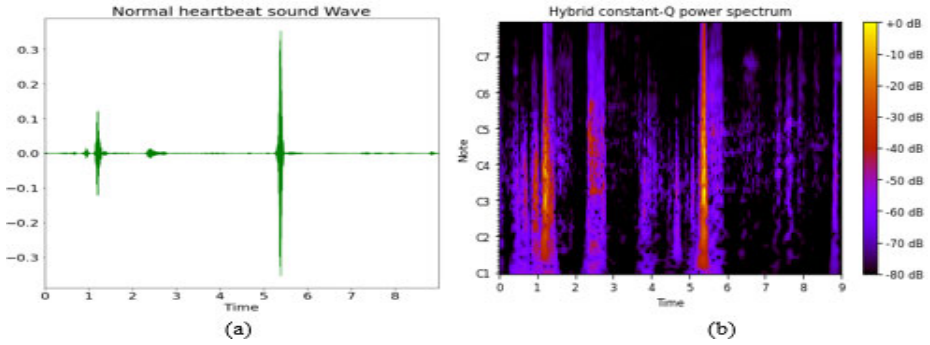

FIGURE 4. (a): A sample waveform for artifact phonocardiogram signal and (b): heat map plot of HCQT power spectrogram for artifact phonocardiogram signal. The spectrogram is generated with the number of frequency bins = 84 and hop length = 512.

(DNNs). Word deep indicates the presence of greater than one hidden later in neural network architecture. CNN is one such type of deep neural network, which is also known as the ConvNet model. It is made up of primarily three layers: a convolution layer, a pooling layer, and a dense layer (fully connected layer) [31], [32]. The first layer i.e., the convolutional layer, is an essential building block of ConvNet. This layer performs the mathematical operation convolution. In a continuous domain, the convolution of two functions f and g is given as in (4):

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau \quad (4)$$

In the discrete case, the same is expressed as in (5):

$$(f * g)(n) = \sum_{m=-\infty}^{\infty} f(m)\, g(n-m) \quad (5)$$

2-D convolution for a digital image can be extended as in (6):

$$(f * g)(x, y) = \sum_{m=-M}^{M} \sum_{n=-N}^{N} f(x-n, y-m)\, g(n, m) \quad (6)$$

The function g represents a filter that is applied to the input image f in this case. 2-D convolution works by applying the convolution filter on the input image. The filter passes over several pixels, which is called a stride. At each spatial location, the convolution between the part of the image and filter is attained. The outcome is a 2-D array which is called a feature map. Softmax, Rectified Linear Unit (ReLU), Randomized Leaky ReLU, and other non-linear activation layers are used to pass this feature map. The pooling layer, also known as the subsampling layer, is another major component of ConvNet. Its purpose is to reduce the spatial size of the activation map to reduce the number of parameters needed for further processing. It applies to all feature maps on its own. Max pooling is the most effective method for the implementation of pooling.

At last, the result of the last pooling layer is received by a fully connected layer and utilized to categorize images into labels. It is the component of ConvNet where discriminative learning is performed. It behaves like a multi-layer perceptron model which can learn weights & identify image classes.

D. PROPOSED PCG SIGNAL CLASSIFICATION MODEL USING ACOUSTIC FEATURES

The offered method for phonocardiogram signal classification using ConvNet is depicted in Fig. 3. The raw data provided is in Waveform Audio File Format (WAV) format, encoding phonocardiogram signals. To pass these sound

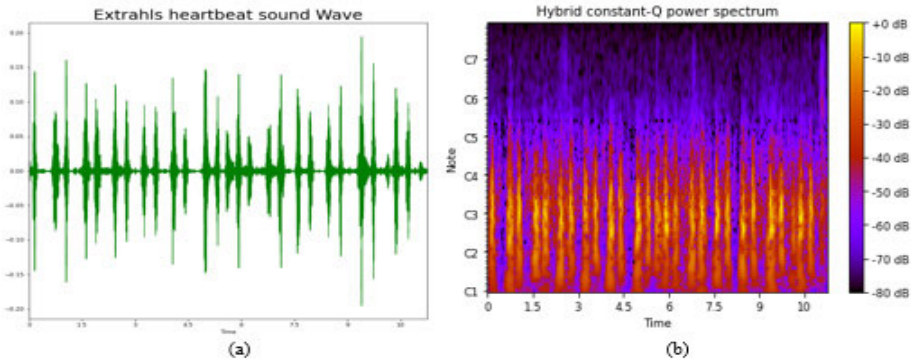

FIGURE 5. (a): A sample waveform for extrahls phonocardiogram signal, (b) heat map plot of HCQT power spectrogram for extrahls phonocardiogram signal. The spectrogram is generated with the number of frequency bins = 84 and hop length = 512.

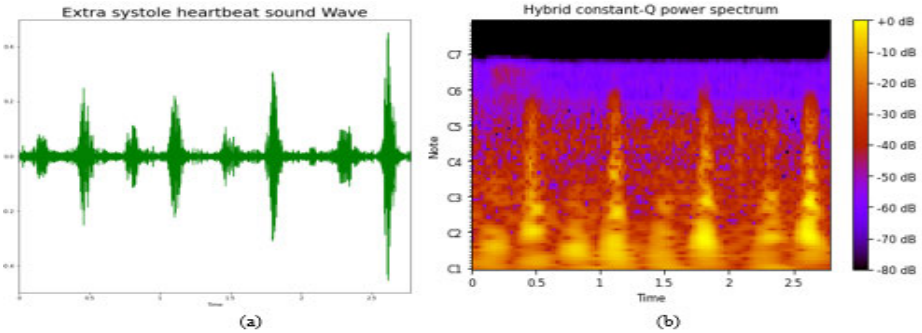

FIGURE 6. (a): A sample waveform for extra systole phonocardiogram signal and (b): heat map plot of HCQT power spectrogram for extra-systole phonocardiogram signal. The spectrogram is generated with the number of frequency bins = 84 and hop length = 512.

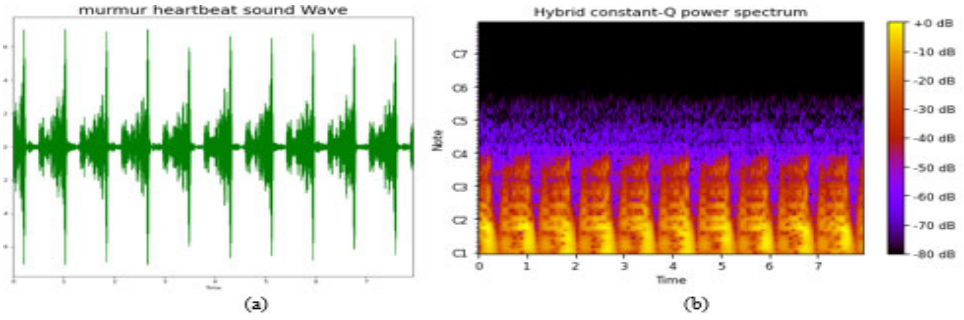

FIGURE 7. (a) A sample waveform for murmur phonocardiogram signal and (b): heat map plot of HCQT power spectrogram for murmur phonocardiogram signal. The spectrogram is generated with the number of frequency bins = 84 and hop length = 512.

waves to ConvNet model, these phonocardiogram signals are converted into an image, i.e. 2-D spectrogram. Spectrograms are convenient for representing these heartbeat recordings because they capture the intensity of the frequencies throughout a given sound. Thus, these spectrograms are effective representations of an audio recording. In this work, the authors have proposed the use MFCC, CQT, VQT, and HCQT based spectrograms for phonocardiogram signal classification.

E. PHONOCARDIOGRAM SIGNAL DATABASE

The authors have used the freely available open access dataset on Kaggle [33], originating through the PASCAL heart sounds classification challenge. Two datasets named A & B were generated through the PASCAL heart sound classification challenge [16]. Dataset A contains the variable-length (varying from 1 to 30 seconds) sounds recorded through a digital stethoscope in a real-time situation having background noise. Dataset A was partitioned into four classes named normal, extra heart sound, murmur, and artifact, while dataset B was partitioned into three classes: normal, extra-systole, and murmur. The authors have merged both datasets into a single dataset consisting of all five classes in this work.

The number of phonocardiogram signals in normal, murmur, artifact, extra-systole, and extrahls classes are 255, 114, 40, 37, and 16. Since the number of heartbeat signals in each class is very low, audio augmentation is performed over raw audio signals. We have applied noise injection, shifting time, varying pitch, and speed to generate augmented data for phonocardiogram signals. After audio augmentation, the number of phonocardiogram signals in normal, murmur,

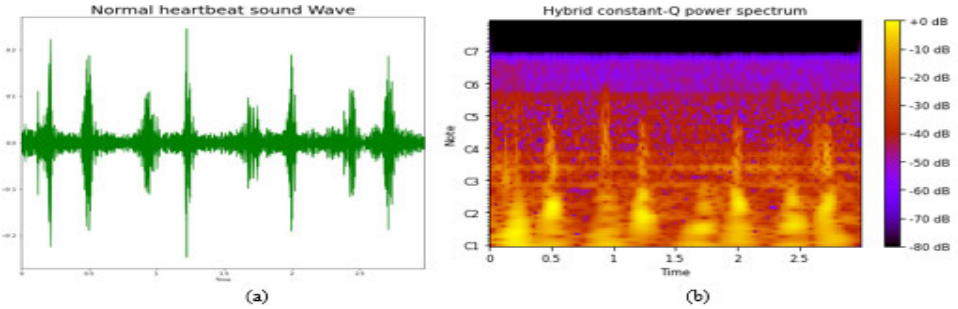

FIGURE 8. (a) A sample waveform for normal phonocardiogram signal and (b): heat map plot of HCQT power spectrogram for normal phonocardiogram signal. The spectrogram is generated with the number of frequency bins = 84 and hop length = 512.

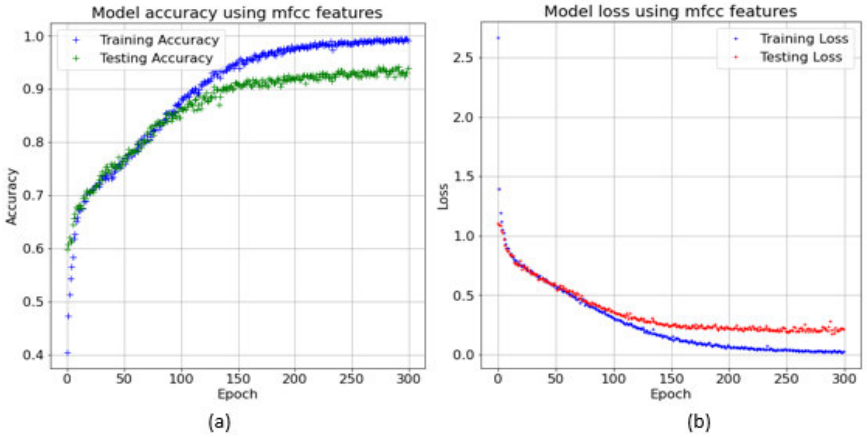

FIGURE 9. (a): Evolution of classification accuracy with the training & validation image datasets throughout the training of ConvNet-MFCC model. Accuracy increases abruptly for the first 200 repetitions & becomes stable after 250 repetitions. (b): Evolution of classification loss with the training & validation image datasets throughout the training of ConvNet-MFCC model. Loss decreases abruptly for the first 200 repetitions and becomes stable after 250 repetitions.

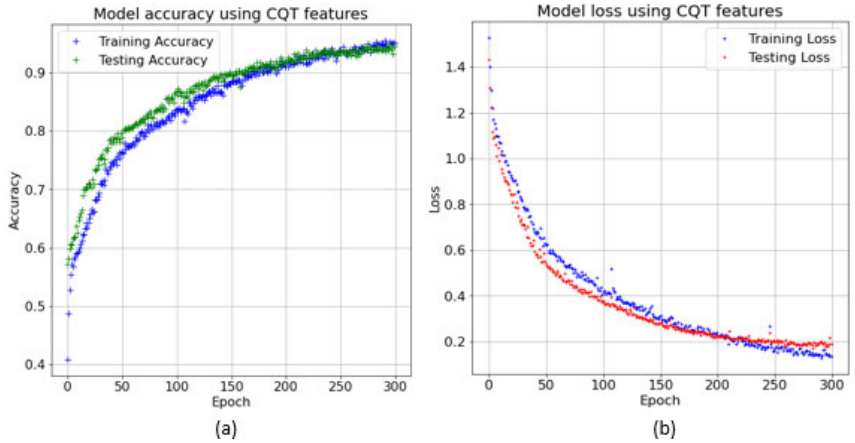

FIGURE 10. (a): Evolution of classification accuracy with training and validation image datasets throughout the training of ConvNet-CQT model. Accuracy increases abruptly for the first 200 repetitions and becomes stable after 250 repetitions. (b): Evolution of classification loss with training and validation image datasets throughout the training of ConvNet-CQT model. Loss decreases abruptly for the first 200 repetitions and becomes stable after 250 repetitions.

artifact, extra-systole, and extrahls classes are 2555, 1146, 400, 378, and 158, respectively. The augmented dataset is partitioned into training and testing datasets with an 80:20 ratio. A spectrogram represents the PCG signal waves, as shown in Fig. (4-8), that presents five types of HCQT spectrograms for the artifact, extrahls, extra-systole, murmur, and normal in that order. Red shades described the amplitude of a PCG signal in a spectrogram. The spectrogram of a normal PCG signal is a strong sequence of amplitude, i.e., lub dub. It displays a noise sequence of amplitude in the murmur PCG signal greater than normal and extra-systole PCG signals. The amplitude of a PCG signal is greater than the normal PCG

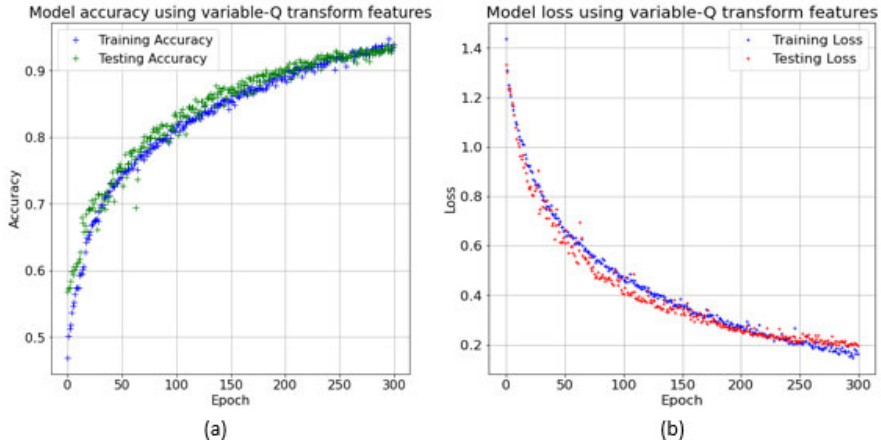

FIGURE 11. (a): Evolution of classification accuracy with training and validation image datasets throughout the training of ConvNet-VQT model. Accuracy increases abruptly for the first 200 repetitions and becomes stable after 250 repetitions. (b): Evolution of classification loss with training and validation image datasets throughout the training of ConvNet-VQT model. Loss decreases abruptly for the first 200 repetitions and becomes stable after 250 repetitions.

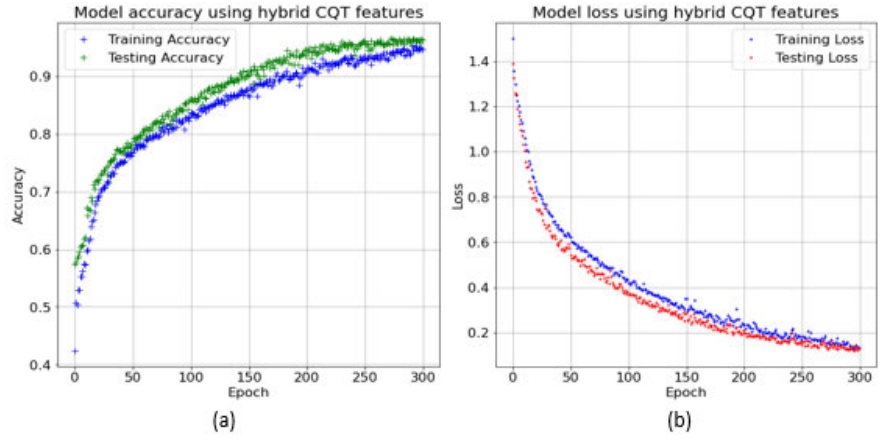

FIGURE 12. (a): Evolution of classification accuracy with training and validation image datasets throughout the training of ConvNet-HCQT model. Accuracy increases abruptly for the first 200 repetitions and becomes stable after 250 repetitions. (b): Evolution of classification loss with training and validation image datasets throughout the training of ConvNet-HCQT model. Loss decreases abruptly for the first 200 repetitions and becomes stable after 250 repetitions.

signal but lesser than the murmur PCG signal in the extra-systole PCG signal.

IV. EXPERIMENT & RESULTS

Four separate ConvNet models termed ConvNet-MFCC, ConvNet-CQT, ConvNet-VQT, and ConvNet-HCQT are designed with MFCC, CQT, VQT, and HCQT spectrograms, respectively. To build the proposed ConvNet models, Keras, an open-source Python library, has been used that can run on top of different machine learning libraries like TensorFlow. In addition, the Librosa library in Python is used for generating MFCC, CQT, VQT, and HCQT spectrograms.

ConvNet models used in this phonocardiogram signal classification model using these spectrograms have four convolutional layers. The first convolution layer has a size of 32-5 $\times$ 5, the second convolution layer has a size of 64-5 $\times$ 5, the third convolution layer has a size of 64-5 $\times$ 5, and the last layer has a size of 32-5 $\times$ 5. A subsampling layer using max-pooling follows the first two convolution layers. The size of these max-pooling layers is 2 $\times$ 2 with a stride of size 2 $\times$ 2. The final layer of the ConvNet model is a fully connected layer with a softmax non-linear activation function with five units. These five units in the last layer are essential for this five-class phonocardiogram signal classification problem.

Additionally, two dropout layers are also used to avoid overfitting with a 0.4 drop rate. The size of the MFCC spectrogram images is 128 $\times$ 130. The model is compiled after design. The optimizer is the gradient descent algorithm based on 'Adam' optimizer and cross-entropy loss to calculate the prediction error rate. The values 0.0001 are used as the learning rate. This optimizer uses backpropagation to update the weights of the neurons. It computes the derivative of the loss function regarding each weight and deducts it from the weight. A categorical cross-entropy loss function is utilized due to the multi-class nature of the problem, which has the

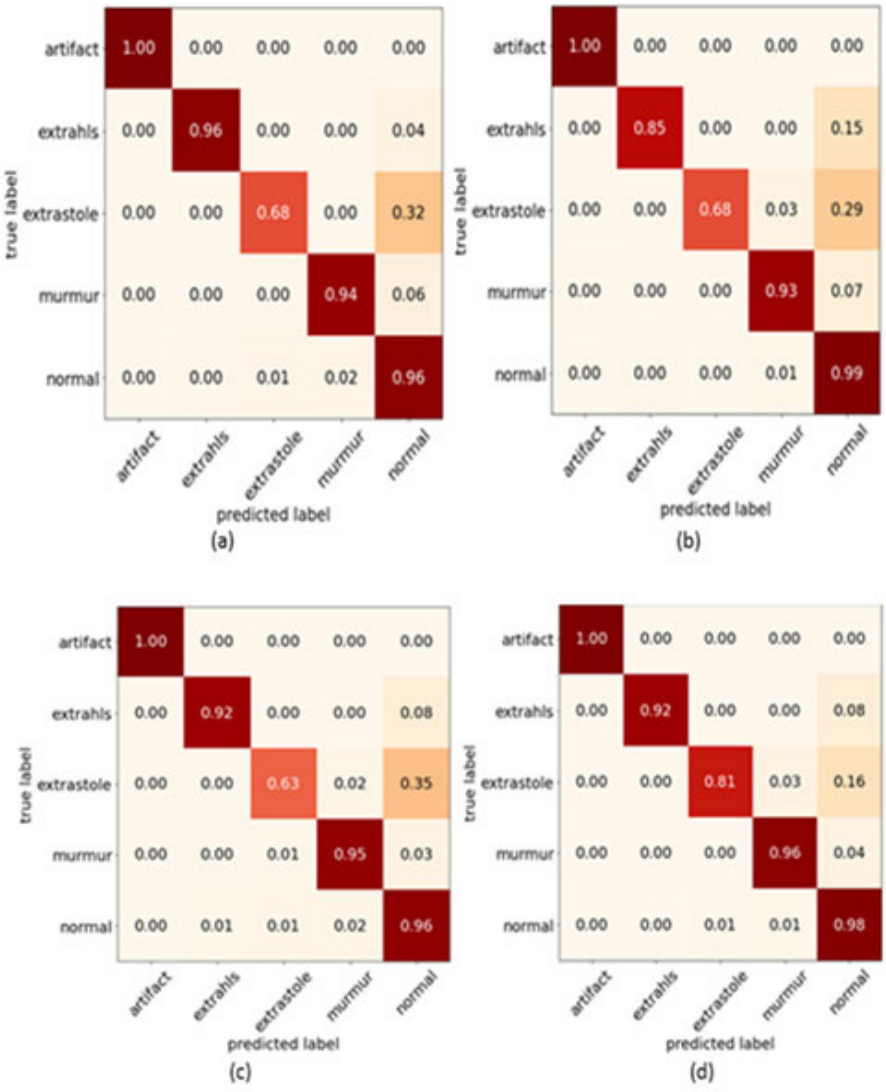

FIGURE 13. Confusion matrix of ConvNet models on the test subset (a) ConvNet-MFCC model, (b) ConvNet-CQT model, (c) ConvNet-VQT model and (d) ConvNet-HCQT model.

form given by (7):

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N} log\frac{e^{W_{y_i}^T x_{i+} b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}} \tag{7}$$

W = weight matrix, x_i = i^{th} training sample, y_i = class label for the i^{th} training sample, b = bias term, N = sample count, $W_{j,}$ and W_{yi} are the j^{th} and y_i^{th} column of W. 300 epochs with batch size 128 are used for training. Fig. (9-12) shows the accuracy and loss curves for the train and test set during the training of ConvNet models. The shape and dynamics of these learning curves are studied to diagnose the behavior of a ConvNet model. Three common dynamics observed in these learning curves are under-fitting, overfitting, and optimal fitting. From these plots, it can be verified that the ConvNet-HCQT model has offered optimal fit in comparison to other models.

Fig. 13 offers the results for these experiments in terms of the confusion matrix. Confusion Matrix is a $N \times N$ matrix, in which rows represent the true categories and the columns represent the classified category by the model. The number $n_{i,j}$ at the intersection of i-th row and j-th column is identical to the number of cases from the i-th phonocardiogram signal class which have been categorized as belonging to the j-th phonocardiogram signal class. It is extremely useful for measuring precision, recall, F-score, accuracy, and most importantly AUC-ROC curve. All these performance metrics are computed and presented in the next section to compare all four models.

V. RESULT ANALYSIS AND DISCUSSION

Statistical performance measures, namely precision, F-score, sensitivity, and accuracy, are computed from the confusion matrix as given in Section 4 to evaluate the performance of all four models, i.e., ConvNet-MFCC, ConvNet-CQT,

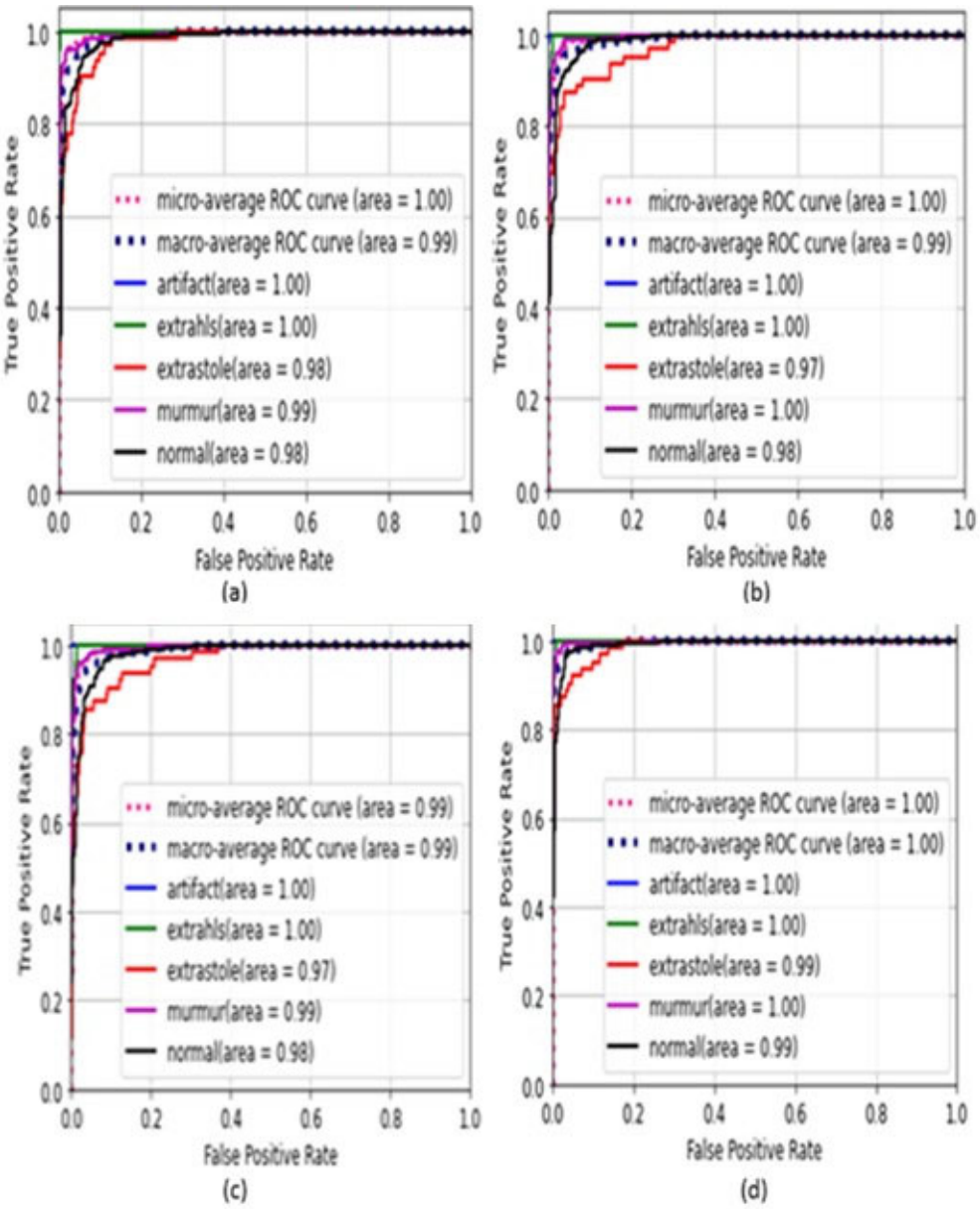

FIGURE 14. ROC curve obtained using acoustic features demonstrating AUC for the artifact, extrahls, extra-systole, murmur, and normal classes separately, micro average and macro average performance measures (a) ConvNet-MFCC model, (b) ConvNet-CQT model, (c) ConvNet-VQT model, and (d) ConvNet-HCQT model.

ConvNet-VQT, and ConvNet-HCQT. These measures are defined in (8-11) [34].

$$\text{Precision(P)} = \frac{T^{+}}{(T^{+} + F^{+})} \tag{8}$$

$$\text{Sensitivity(S)} = \frac{T^{+}}{(T^{+} + F^{-})} \tag{9}$$

$$F - \text{Score} = \frac{(2 * P * S)}{(P + S)} \tag{10}$$

$$\text{Accuracy} = \frac{(T^{+} + T^{-})}{(T^{+} + T^{-} + F^{-} + F^{+})} \tag{11}$$

where T^{+}, T^{-}, $F^{+,}$ and F^{-} are the truly projected positive, truly negative cases, false-positive cases, and false-negative cases, respectively. The results in terms of the above performance measures are offered in Table 2. The results clearly show that ConvNet-HCQT beats other models. The average accuracies achieved using HCQT is 96%, whereas it is 93%, 94%, and 94%, respectively, for ConvNet-MFCC, ConvNet-CQT, and CovNet-VQT models. The performance of ConvNet-CQT and CovNet-VQT models is the same but superior to ConvNet-MFCC. MFCC features are widely used features in the past for heartbeat sound classification. In comparison to earlier work, the experimental results show that the proposed strategy achieves good outcomes. The proposed method outperforms the PhysioNet/Computing in Cardiology Challenge2016's stated best accuracy of 0.86 for normal/abnormal binary classification. Table 3 provides

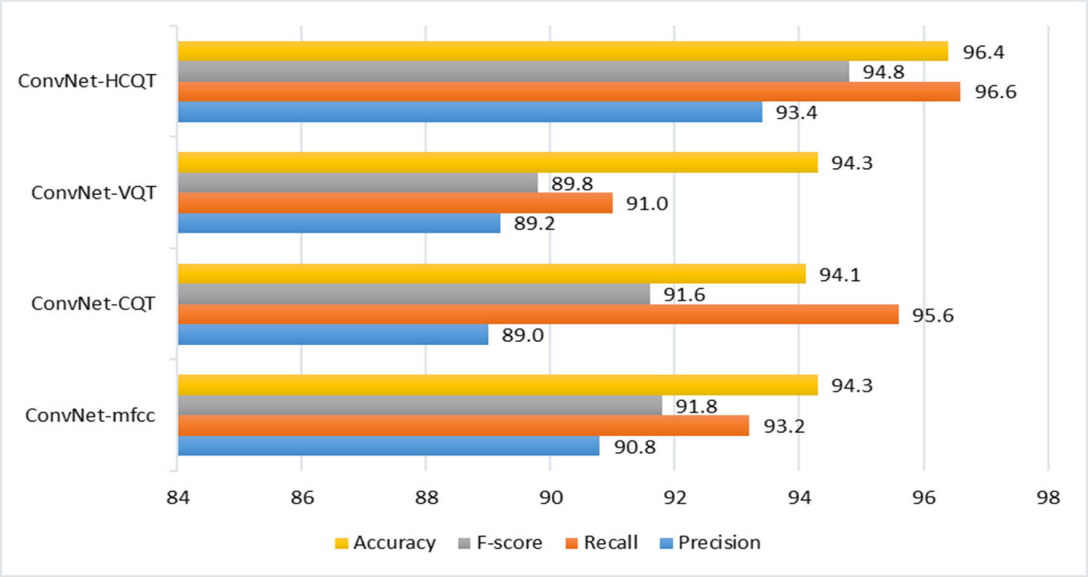

FIGURE 15. Performance comparison in terms of precision, recall, F-score, and accuracy of the HCQT-based ConvNet model with others.

TABLE 2. Performance measures for phonocardiogram signal classification using MFCC, CQT, VQT, and HCQT features in terms of precision, sensitivity, F-score, macro, weighted and average accuracy.

	ConvNet-MFCC			ConvNet-CQT			ConvNet-VQT			ConvNet-HCQT		
Class/metric	*Precision*	*sensitivity*	$F-score$	*Precision*	*sensitivity*	$F-score$	*Precision*	*sensitivity*	$F-score$	*Precision*	*Sensitivity*	$F-score$
Artifact	1	1	1	1	0.98	0.99	1	1	1	1	1	1
Extrahls	0.96	0.93	0.94	0.85	0.92	0.88	0.92	0.83	0.87	0.92	0.96	0.94
Extrasystole	0.68	0.84	0.75	0.68	0.98	0.8	0.63	0.83	0.72	0.81	0.94	0.87
Murmur	0.94	0.96	0.95	0.93	0.97	0.95	0.95	0.95	0.95	0.96	0.97	0.96
Mormal	0.96	0.93	0.95	0.99	0.93	0.96	0.96	0.94	0.95	0.98	0.96	0.97
Macro Average	0.91	0.93	0.92	0.89	0.96	0.89	0.91	0.89	0.91	0.93	0.97	0.95
Weighted Average	0.93	0.94	0.94	0.95	0.94	0.95	0.94	0.94	0.94	0.96	0.96	0.96
Average Accuracy		0.93			0.94			0.94			0.96	

the overall accuracies for the best models presented in PhysioNet Computing Cardiology challenge [35]. Accuracies provided by these models are very much inferior to the proposed multi-class classification model using HCQT.

To further confirm the robustness of these phonocardiogram signal classification models, ROC curves are also plotted in Fig. 14. The false-positive rate on the x-axis and the true positive rate on the y-axis is plotted on ROC curves. This implies that the top left corner of the plot is the ''perfect'' point where a true positive rate of one and a false positive rate of zero. It means that a larger AUC is generally superior [36]. It is evident from the ROC curves that the ConvNet-HCQT model performs better than other models, which the AUC of these ROC plots confirms. For the MFCC-based ConvNet model, the micro-average area and macro-average area are 1.00 and 0.99, respectively. With the HCQT-based ConvNet model, which is 1.00, the metric macro-average area is slightly improved. The area under the curve for the artifact, extrahls, extra-systole, murmur, and normal classes are 1.00, 1.00, 0.99, 1.00, and 0.99. It can be noticed that AUC is slightly improved for these classes with HCQT based features in comparison to other features.

Commonly used time-frequency transformations and features such as DFT, DWT, and MFCC have extensively supported various acoustic recognition systems. Though they are appreciated for most acoustic analyses, they are still not customized to any particular problem. So, it may be valuable to investigate features from other time-frequency transformations such as CQT, VQT, and HCQT. CQT is a dominant feature in acoustic signal processing analysis. CQT transforms a series of time-domain signals to the frequency domain signal. It is similar to the Short Term Fourier Transform (STFT) and almost identical to the complex Morlet wavelet transform. Hybrid CQT is a more computationally efficient version of CQT. It utilizes the pseudo-CQT for higher-order frequencies where the hop length is larger than half the filter size and full CQT for the lower frequencies. The findings

TABLE 3. Selected results from the 2016 PhysioNet computing in cardiology challenge [35].

Rank	Overall Accuracy	Description
1	0.8602	AdaBoost & CNN
2	0.859	Ensemble of SVMs
3	0.852	Regularized Neural Networks
4	0.8454	MFCCs, Wavelets, Tensors
5	0.8448	KNN Random Forest + LogitBoost
6	0.8415	Unofficial entry
7	0.8411	Probability-distribution based
8	0.8399	Heatmaps+CNN
9	0.8282	Approach Unknown
10	0.8263	Approach Unknown

of the experiments show that HCQT is more effective than traditional CQT and variable CQT.

In this study, an effort is made to suggest the best acoustic features for phonocardiogram signal classification. Fig. 15 presents the comparison of the HCQT-based ConvNet model with others. Results have proved that HCQT outperforms other acoustic features of the time-frequency domain. It would be fascinating to investigate a larger number of architectural configurations and filter banks, as well as hyperparameter sets, in the future.

VI. CONCLUSION

Diagnose at an early stage is the only way to decrease the mortality rate occurring due to CVD. However, due to a lack of awareness for routine health checkups and unavailability of all resources at low cost, there are major hurdles in the early diagnosis of CVD. The situation worsens in developing countries where population density is high, and a doctor is not available in remote locations. To target these issues, the authors have offered a design of a decision support system that utilizes the PCG signals for the early diagnosis of CVD. PCG signals can be captured by a small, low-cost handheld device called a stethoscope. In this work, a multi-class phonocardiogram signal database with five classes, namely, extra heart sound, artifact, extra-systole, normal, and murmur heartbeat, are used to design the phonocardiogram signal, classification model. The authors have designed a PCG signal classification model with a new acoustic feature HCQT. HCQT has been formed by combining two CQTs consisting of dissimilar resolutions for treating the high-frequency bins of the conventional CQT. Analysis of results has proved that HCQT is a superior feature that generally applies acoustic features like MFCC, CQT, and VQT. Through the proposed work, the authors have achieved an accuracy of 96% in the multi-class classification of PCG signals.

In future work, authors have planned to ensemble multiple spectrograms to get more discriminative stacked features. Also, classification accuracy may further be improved by using other deep learning architecture like Recurrent Neural Network (RNN). Moreover, the authors have also planned to use an ECG signal with the PCG signal to design the multimodality model using these acoustic features.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Prince Sultan University, Riyadh, Saudi Arabia for partially supporting this project and for paying the Article Processing Charges (APC) of this publication.

REFERENCES

[1] (Jun. 2021). *CVD Data as Cited on 27th*. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
[2] A. Jain, S. Tiwari, and V. Sapra, "Two-phase heart disease diagnosis system using deep learning," *Int. J. Control Autom.*, vol. 12, no. 5, pp. 558–573, 2019.
[3] 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.
[4] P. Rani, R. Kumar, N. M. S. Ahmed, and A. Jain, "A decision support system for heart disease prediction based upon machine learning," *J. Reliable Intell. Environ.*, pp. 1–13, Jan. 2021, doi: 10.1007/s40860-021-00133-6.
[5] A. E. F. Malik, S. Barin, and M. E. Yuksel, "Accurate classification of heart sound signals for cardiovascular disease diagnosis by wavelet analysis and convolutional neural network: Preliminary results," in *Proc. 28th Signal Process. Commun. Appl. Conf. (SIU)*, Oct. 2020, pp. 1–4.
[6] G. Y. Son and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, pp. 2344–2358, 2018.
[7] P. Upretee and M. E. Yuksel, "Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results," in *Proc. Sci. Meeting Elect.-Electron. Biomed. Eng. Comput. Sci. (EBBT)*, Apr. 2019, pp. 1–4.
[8] 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, 2015.
[9] C. Cable. (1997). *The Aucultation Assistant*. Accessed: Apr. 3, 2014. [Online]. Available: www.med.ucla.edu/wilkes/intro.html
[10] Stillman. (2007). *ALDMD Clinical Cardiology Tools from Hennepin County Medical Center*. Accessed: Apr. 8, 2014. [Online]. Available: http://www.aldmd.com
[11] R. Lichtenberg. (May 1, 2014). *Heart Sounds*. [Online]. Available: http://www.loyolauniversity.adam.com
[12] (Mar. 16, 2014). *University of Michigan Heart Sound and Murmur Library*. [Online]. Available: http://www.med.umich.edu
[13] J. M. Wilson. (2009). *Heart Sound Podcast Series*. Accessed: Apr. 6, 2014. [Online]. Available: www.texasheartinstitute.org
[14] A. Gharehbaghi, A. A. Sepehri, A. Kocharian, and M. Linden, "An intelligent method for discrimination between aortic and pulmonary stenosis using phonocardiogram," in *Proc. World Congr. Med. Phys. Biomed. Eng.*, Toronto, ON, Canada, 2015, pp. 1010–1013.
[15] 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.
[16] P. Bentley, G. Nordehn, M. Coimbra, S. Mannor, and R. Getz. *Classifying Heart Sounds Challenge*. Accessed: May 3, 2018. [Online]. Available: http://www.peterjbentley.com/heartchallenge/#downloads
[17] *PhysioNet/Computing in Cardiology Challenge. Classification of Normal/Abnormal Heart Sound Recordings*. Accessed: May 31, 2018[. Online]. Available: https://www.physionet.org/challenge/2016/
[18] M. Banerjee and S. Majhi, "Multi-class heart sounds classification using 2D-convolutional neural network," in *Proc. 5th Int. Conf. Comput., Commun. Secur. (ICCCS)*, Oct. 2020, pp. 1–6.
[19] S. B. Shuvo, S. N. Ali, S. I. Swapnil, M. S. Al-Rakhami, and A. Gumaei, "CardioXNet: A novel lightweight deep learning framework for cardiovascular disease classification using heart sound recordings," 2020, *arXiv:2010.01392*. [Online]. Available: http://arxiv.org/abs/2010.01392
[20] G. Redlarski, D. Gradolewski, and A. Palkowski, "A system for heart sounds classification," *PLoS ONE*, vol. 9, no. 11, pp. 1–12, 2014.

[21] *3M Poland Microphone Samples*. Accessed: Jun. 2012. [Online]. Available: http://www.littmann.in/wps/portal/3M/en_IN/Littmann/stethoscope/education/heart-lungsounds/
[22] P. Narváez, S. Gutierrez, and W. S. Percybrooks, "Automatic segmentation and classification of heart sounds using modified empirical wavelet transform and power features," *Appl. Sci.*, vol. 10, no. 14, pp. 4791–4812, 2020.
[23] F. Zheng, G. Zhang, and Z. Song, "Comparison of different implementations of MFCC," *J. Comput. Sci. Technol.*, vol. 16, no. 6, pp. 582–589, Nov. 2001.
[24] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan, "Heart sound classification based on improved MFCC features and convolutional recurrent neural networks," *Neural Netw.*, vol. 130, pp. 22–32, 2020.
[25] P. B. Bachhav, M. Todisco, M. Mossi, C. Beaugeant, and N. Evans, "Artificial bandwidth extension using the constant q transform," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 5550–5554.
[26] M. Todisco, H. Delgado, and N. W. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," *Odyssey*, vol. 2016, pp. 283–290, Jun. 2016.
[27] E. Benetos and T. Weyde, "An efficient temporally-constrained probabilistic model for multiple-instrument music transcription," presented in 16th Int. Soc. Music Inf. Retr. Conf., 2015.
[28] S. Abidin, R. Togneri, and F. Sohel, "Spectrotemporal analysis using local binary pattern variants for acoustic scene classification," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 11, pp. 2112–2121, Nov. 2018.
[29] M. Wang, R. Wang, X.-L. Zhang, and S. Rahardja, "Hybrid constant-Q transform based CNN ensemble for acoustic scene classification," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2019, pp. 1511–1516.
[30] M. Wan, R. Wang, B. Wang, J. Bai, C. Chen, Z. Fu, and S. Rahardja, "Ciaic-ASC system for DCASE 2019 challenge task1," DCASE2019 Challenge, New York, NY, USA, Tech. Rep., 2019.
[31] S. Tiwari and A. Jain, "Convolutional capsule network for COVID-19 detection using radiography images," *Int. J. Imag. Syst. Technol.*, vol. 31, no. 2, pp. 525–539, Jun. 2021.
[32] I. Ahmad and K. Pothuganti, "Analysis of different convolution neural network models to diagnose Alzheimer's disease," *Mater. Today, Proc.*, pp. 1–5, Oct. 2020, doi: 10.1016/j.matpr.2020.09.625.
[33] *Heartbeat Sounds Classifying Heartbeat Anomalies From Stethoscope Audio*. Accessed: Jun. 27, 2021. [Online]. Available: https://www.kaggle.com/kinguistics/heartbeat-sounds
[34] S. Tiwari, "A blur classification approach using deep convolution neural network," *Int. J. Inf. Syst. Model. Design*, vol. 11, no. 1, pp. 93–111, Jan. 2020.
[35] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing abnormal heart sounds using deep learning," 2017, *arXiv:1707.04642*. [Online]. Available: http://arxiv.org/abs/1707.04642
[36] A. C. J. W. Janssens, "ROC curves for clinical prediction models—Part 2. The ROC plot: The picture that could be worth a 1000 words," *J. Clin. Epidemiology*, vol. 126, pp. 217–219, Oct. 2020.

SHAMIK TIWARI is currently working as a Senior Associate Professor with the Department of Virtualization at SoCS, University of Petroleum and Energy Studies, Dehradun. He has rich experience of around 18 years as an academician. His research interests include digital image processing, computer vision, biometrics, machine learning especially deep learning, and health informatics. He has written many national and international publications, including books in these fields.

ANURAG JAIN is currently working as an Associate Professor with the Department of Virtualization at SoCS, University of Petroleum and Energy Studies, Dehradun. He is in the field of academia for the last 18 years. Under his guidance, 14 students have successfully defended their M. Tech. thesis. He has published around 44 research articles in renowned journals and conferences. He is also guiding five Ph.D. students in their research work. His research interests include scheduling and load balancing in cloud computing, healthcare, machine learning, and data science.

AKHILESH KUMAR SHARMA (Member, IEEE) received the B.E., M.E., and Ph.D. degrees. He is currently working as an Associate Professor with Manipal University Jaipur, India. He is with CIDCR Laboratory, India. He chaired sessions and acted as an Expert for keynotes in IITs, NITs, Vietnam, Thailand, Malaysia, Australia, and China. He is affiliated with IEEE, ACM, CSI, (IUCEE), and MIR Laboratory, USA. He has organized various FDP's, events and conferences, and workshops, with GOI funding and established. Due to his meritorious contributions in the field of education, he has been awarded many awards. He has written three books and two are in progress. He is guiding six Ph.D. candidates as five guides and one co-guide. He is also a Joint Secretary with ACM Professional Chapter Jaipur and ACM Student Chapter Faculty Coordinator. He established many active MOU's with Warsaw University (Poland), KMUTNB (Thailand), and Prince Sultan University (Saudi Arabia). He holds three patents and two copyrights to his credit and has set up a Cognitive Intelligence Research Laboratory, Jaipur, Rajasthan, India.

KHALED MOHAMAD ALMUSTAFA received the B.E.Sc. degree in electrical engineering, and the M.E.Sc. and Ph.D. degrees in wireless communication from the University of Western Ontario, London, ON, Canada, in 2003, 2004, and 2007, respectively. He is currently working as an Associate Professor with the Department of Information Systems (IS), College of Computer Science and Information Systems (CCIS), Prince Sultan University (PSU), Riyadh, Saudi Arabia. He served as a General Supervisor for the Information Technology and Computer Services Center (ITCS), PSU, the Chairman of the Department of Communication and Networks Engineering (CME), and the Vice Dean for the College of Engineering at PSU, the Director for the Research and Initiatives Center, PSU, and he is also serving as the CITO with PSU. His research interests include error performance evaluation of MIMO communication systems in partially known channels, adaptive modulation, and channel security, text recognition models, and control systems.