Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INNOVATIVE MACHINE LEARNING FRAMEWORK FOR PHONOCARDIOGRAPHY (PCG) USING MFCC AND DEEP EXTREME LEARNING MACHINE (DELM)

ABDULLAH ALTAF[1*], HAIRULNIZAM MAHDIN[2], AWAIS MAHMOOD[3], ABDULREHMAN ALTAF [4]

[1,2,4] Faculty of Computer Science and Information Technology (FSKTM), Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor, Malaysia

[3] Computer Engineering Dept, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

E-mail: [1]hi210007@student.uthm.edu.my, [2]hairuln@uthm.edu.my, [3]mawais@ksu.edu.sa, [4]hi210004@student.uthm.edu.my

ABSTRACT

Cardiovascular Diseases (CVDs) are a significant global cause of mortality, necessitating effective diagnostic techniques. Phonocardiography (PCG) is among the fundamental methods used to analyze heart sounds to detect human heart-related abnormalities. However, in an environment where state-of-the-art PCG equipment is not available, a Machine Learning (ML) based solution can serve as a reliable alternative. However, the main challenges faced by ML-based PCG systems, are the unavailability of balanced and unbiased datasets, the vanishing and exploding gradient a well-known Deep Learning (DL) issue, and inappropriate feature extraction techniques, which often compromise the accuracy and reliability of ML-based PCG systems. This study introduces a novel Deep Extreme Learning Machine (DELM) and Mel-Frequency Cepstral Coefficients (MFCC) based PCG framework for CVD diagnosis. The proposed framework uniquely addresses the above mentioned challenges. The proposed model achieves a remarkable training accuracy of 98.46 % and a test accuracy of 86.80 %, using the Heartbeat Sound dataset with five classes and after class aggregation and dataset normalization the proposed model achieved training accuracy 99.52 % and a test accuracy of 92.30 % demonstrating its potential in PCG diagnostics. This framework represents a significant advancement in ML-based PCG systems for automating heart sound analysis and contributing to improved cardiac healthcare, especially in resource-limited settings.

Keywords: *Cardio Vascular Disease (CVDs), MFCC, Machine Learning, Deep Extreme Learning Machine (DELM), Heart Disease*

1. INTRODUCTION

This CVDs are the leading cause of death globally, taking 17.9 million lives per year, or almost 32 % of all deaths worldwide [1]. A noteworthy 85 % of these deaths are the consequence of heart attacks and strokes. More than 75 % of deaths from CVD occur in low- and middle-income nations. Concerningly, 38 % of the 17 million recorded deaths before the age of 70 occurred only in 2019 as a result of CVDs. It is noteworthy that heart attacks and strokes account for about four out of every five deaths attributable to CVD, and that an alarming one-third of these deaths occur in people under the age of seventy-one. Unfortunately, the lack of cardiovascular disease specialists, the lack of modern diagnostic tools, and the high rate of misdiagnosed cases are a few major causes of these deaths [2].

PCG serves as a fundamental technique utilized to diagnose the state of the human heart and determine whether it is functioning normally or displaying any abnormal patterns. It involves the visual recording of the sounds and murmurs produced by the contracting of the heart, originating from the valves and connected large vessels. In situations where advanced diagnostic equipment is not readily available, general physicians rely solely on the stethoscope to perform PCG [3]. Nonetheless, general physicians could find it difficult to determine whether the heart is beating normally or whether there are any

anomalies in the heart's walls without the assistance of a cardio specialist [4].

There are many strains associated with PCGs as well. For instance, environmental factors like skin friction, electromagnetic interference (EI), and breathing noises can cause considerable disruptions to the PCG diagnosis [3, 5]. Moreover, Heart sounds can differ significantly among humans due to natural factors as well like age, gender, and health condition [6]. The variations in heart sound can also be caused because of anatomical differences, and physiological variations that further challenge PCGs for achieving consistent diagnosis [7]. Furthermore, the lack of standardized protocols and guidelines for recording and analyzing heart sound signals makes PCG diagnosis even more complex and less reliable. Also, inconsistencies in signal acquisition techniques, sensor placement, and signal processing methods make PCG diagnosis deceitful [8]. The practicality and usability of PCG systems can also be hindered by the availability and affordability of suitable hardware devices. Access to high-quality sensors or wearable devices that can reliably capture heart sounds may be limited, particularly with limited resources. Distinguishing between normal and abnormal heart sounds, as well as identifying various pathological conditions, can be intricate and may necessitate the involvement of experienced cardiologists or clinicians. The accurate automation of this process remains a significant challenge [9, 10].

ML-based approaches can provide effective PCG solutions, by overcoming the discussed limitations of conventional PCG techniques [11]. Compared with typical manual PCG diagnosis, ML models can expedite the evaluation of PCG recordings, saving time and lowering the possibility of human error. Furthermore, by training ML models on large datasets of labeled PCG recordings, the accuracy and reliability of PCG diagnosis models can be improved. Also, PCG includes continuously evaluating audio signals and adjusting to new data while ML-based methods are best known for handling vast volumes of data, evolving learning, and eventually improving models. This allows improved performance and the detection of novel circumstances even amongst diverse human populations [8]. Real-time evaluation of PCG recordings by utilizing ML-based PCG models has the potential to offer accurate diagnostic insights. Integrating ML with PCG can produce more precise and unbiased outcomes, ultimately improving the diagnosis and management of cardiac conditions [12]. However, previously proposed ML-based PCG frameworks have indeed faced several challenges that have hindered their effectiveness. Model overfitting, limited interpretability, un-explainability, and the absence of standardized protocols and standards related to the development of ML-based PCG systems are also some other challenges and provocations that make ML-based PCG systems slightly impractical and limit their commercial adoption.

ML-based models need large, balanced, unbiased efficient, and effective datasets to train. However, publicly available datasets are limited in sample size, unbalanced, and biased. The limited size of the datasets restricts the representation of the full variability and complexity of heart sound signals, impacting the generalization and performance of the models. Unbalanced datasets underrepresent or overrepresent certain classes or labels, which can lead to biased predictions and lower accuracy, especially in detecting rare abnormalities. Furthermore, noisy or incomplete labels in the datasets can introduce errors and hinder accurate learning. The lack of standardization in data collection, annotation, and preprocessing has also posed further challenges in comparing results and reproducing experiments.

Mostly ML-based PCG prediction model is trained using BP Neural Network (NN) algorithms because of their efficiency, versatility, and adaptability. But using BP has some challenges as well like data sensitivity, time and resource intensive, and prone to model overfitting. Still one of the vital challenges includes vanishing and exploding gradients, which are classical problems in BP algorithm-based trained models. In the BP algorithm, the gradient is a vector that represents the partial derivatives of a function with multiple variables. It measures how the change in weights relates to the change in error. The gradient is used to calculate the rate at which the output of a complex equation changes when the input changes. The gradients calculated during training either become extremely small ("vanishing") or very large ("exploding") as they propagate through the network layers, hindering the learning process and preventing effective weight updates.

Additionally, another significant challenge is a suitable feature extraction technique that extracts features from analog heart sound signals available in different datasets in the form of digital sound files. Selecting appropriate features that capture the distinctive characteristics of different heart conditions is also very critical for ML-based PCG systems.

This study focuses on the three main challenges related to ML-based PCG systems i.e. dataset related issues like data biases, improper class aggregation, etc. BP-based model issued i.e. vanishing and exploding gradients, and use of inappropriate feature extraction techniques. The proposed framework also minimizes the impact of human error, natural factors, environmental issues, and machine-human biases. The proposed study aims to address these challenges in PCG that will assist medical technicians, practitioners, and physicians in providing real-time recommendations to patients with heart abnormalities. By incorporating, a benchmark, high-quality, unbiased, class aggregated and normalized dataset, state-of-the-art Deep Learning (DL) based Feed-Forward Neural Network (FF) learning algorithm i.e. DELM and effective feature extraction techniques i.e. MFCC. The study aims to enhance the accuracy and reliability of heart sound classification by proposing a novel framework.

The research holds significant importance due to its potential to progress the field of PCG and contribute to enhancing the diagnosis and monitoring capabilities of cardiac conditions. The outcomes of this research can greatly impact clinical settings, where timely and accurate identification of cardiac abnormalities is crucial for patient care. Improved classification and detection capabilities can assist healthcare professionals in making more informed decisions, enabling early intervention and leading to better patient outcomes. This research also contributes to the existing knowledge in ML by exploring innovative approaches for feature extraction and learning techniques within the context of PCG. This study contributes to the development of more accurate, reliable, commercially usable, and adoptable ML-based PCG systems. The insights gained from this study can serve as a foundation for future advancements in the field and inspire further research in related areas, ultimately driving innovation and progress in healthcare technology.

The framework's optimal performance assumptions rely on expertise in data preprocessing, noise reduction, and iterative model training. Limitations include reliance on high-quality sensors for data acquisition and challenges in addressing the inherent biases and inconsistencies in publicly available datasets. Furthermore, environmental noise and inter-individual variability in heart sounds represent additional hurdles in achieving a universally robust solution.

The proposed framework aims to address critical gaps in existing PCG systems by integrating MFCC for effective feature extraction and DELM for overcoming gradient issues. Unlike traditional methods, this framework leverages normalized datasets with class aggregation to minimize bias and enhance reliability, ensuring practical applicability in resource-constrained healthcare settings.

2. MATERIAL & METHODS

The proposed PCG framework includes phases like heart sound acquisition, signal denoising, signal segmentation, and feature extraction classification models. Also, the proposed framework is extensively evaluated, the results are rigorously obtained under different experimental settings and the results are comprehensively discussed.

Heart sounds are captured using digital stethoscopes or PCG sensors and converted into digital recordings. Preprocessing is applied on captured heart sounds to remove noise and artifacts, using a signal denoising technique followed by feature extraction using MFCC to capture distinctive features of different heart conditions using MFCC. The DELM algorithm is utilized to train models on labeled datasets. Model performance is evaluated using 5 class, and 3 class aggregation. The proposed model can be deployed in real-world applications for analyzing heart sound recordings and providing predictions or classifications to aid in diagnosing and monitoring cardiac conditions. This study assumes that the optimal performance of the framework relies on expertise in data preprocessing, noise reduction, feature engineering, and iterative model training, as well as the availability of high-quality datasets, standardized protocols, and standards to capture and record accurate heart sound signals. The proposed framework is depicted in the Figure 1.

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

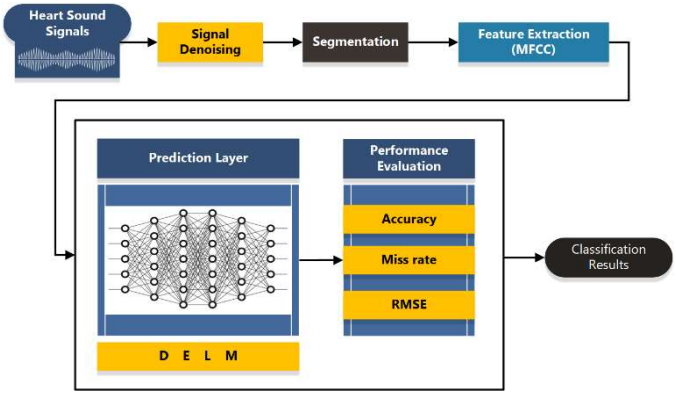

Figure 1: Proposed MFCC and DELM-based PCG Framework

2.1 Heart Sound Signals

Heart sound signals are produced by cardiac events like a valve closing or the chordae tendinous tightening. S1 and S2, the two audible heart sounds, are often present in healthy hearts. S1 is the sound that occurs when the atrioventricular valves close during ventricular contraction. When the semilunar valves close during ventricular diastole, the sound produced is known as S2. Since the two valves close at slightly different times, each sound has two components. There may be other cardiac sounds, such as S3 and S4, which can cause sound wave sensors that translate sound waves into electrical signals, which are subsequently transformed into digital audio data. Heart sound signals are distinct audio vibration bursts that vary in strength, frequency, pitch, length, and quality. Heart sound signals in the form of digital audio files captured by an electronic stethoscope or other audio recording devices act as an input to the proposed MFCC and DELM-based PCG system.

2.2 Signal Denoising

The recording of heart sounds can be readily interfered with by environmental factors including friction between the device and the skin, interference from electromagnetic sources (EI), and background sounds such as breathing, lung or respiratory noises, and surrounding environmental sounds [3]. To filter out-of-band noise or any interference that might be present with the sound signals generated by the human heart must be removed. Noise reduction has a major impact on the subsequent phases like segmentation, feature extraction, and final classification results. Three widely used denoising approaches are wavelet denoising, variational mode deconstruction denoising, and digital filter denoising (DFD) [13]. Proposing a novel signal-denoising technique is not in the scope of this study. However, this framework recommends digital filter denoising techniques for noise reduction. In this study, a wavelet function for human heart impulses is being developed. This wavelet function is built upon prior knowledge of heart sound data [14]. A sound spectrum is an illustration of a sound, typically a brief excerpt of the sound in terms of vibration intensity at each frequency. Typically, it is shown as a power or pressure graph plotted against frequency. Decibels are typically used to measure pressure or power, whereas Hertz, or vibrations per second, is used to measure frequency. After the sound is analyzed, the spectrum is produced, which represents the frequency composition of the sound. A coordinate plane is typically used to describe a sound spectrum. The amplitude A, or intensity, of a harmonic component with a given frequency, is displayed along the axis of ordinates, while the frequency f is plotted along the axis of abscissas.

2.3 Segmentation

The diastole, second heart sound S2, and first heart sound S1 are separated into four segments as part of the segmentation process. Important information that helps differentiate between various heart sounds can be found in each area. Nevertheless, erroneous PCG signal segmentation can result from individual differences in the length of the pulse cycle, the quantity of heart sounds, and the kinds of heart murmurs. As a result,a crucial stage in the automated processing of PCG signals is segmenting the Fatal Heart Sound (FHS). Envelope-based approaches have emerged as some of the most popular strategies for heart sound segmentation in recent years [15]. Several significant segmentation techniques include the electrocardiogram (ECG) [16], time-frequency analysis approaches [18], feature-based methods [17], and probabilistic model methods [19, 20, 21, 22]. This diastolic interval is longer than the systolic period, which is the fundamental assumption of the proposed framework. It is

crucial to remember that this presumption is not always accurate for an aberrant heart sound in infants and cardiac patients. [23, 29]. Due to the similarities between ECG and heart signals, research has shown that combining ECG data with the cardiac cycle enhances the accuracy of segmentation algorithms. However, they do require more sophisticated technology in terms of hardware and software.

In sound signals pitch and frequency are related i.e. higher pitch means high frequency and vice versa. To produce a representation that is essentially the same as the human brain processable, the representation gains the frequency dimension from the spectrogram. A spectrogram is a visual representation of the frequency spectrum of a signal that varies with time. When applied to an audio source, spectrophotograms are sometimes called sonography, voiceprints, or voice grams. Spectrograms are widely utilized in many disciplines, including seismology; voice processing, sonar, radar, music, and linguistics. Audio spectrograms can be used to study animal cries and phonetically identify spoken phrases. It can be produced using a wavelet transform, a bank of band-pass filters, an optical spectrometer, or a Fourier transform.

2.4 Feature Extraction

Davis and Murmelstein first proposed the MFCC method in 1980 [24]. In this study, the MFCC technique is proposed for feature extraction. MFCC is suitable to capture a compact representation of the spectral envelope of an audio signal, which can then be utilized further. The signal's power spectrum is simply smoothed to create the spectral envelope, which is a representation of the signal's energy distribution over several frequency bands [25] seen in Figure 2.

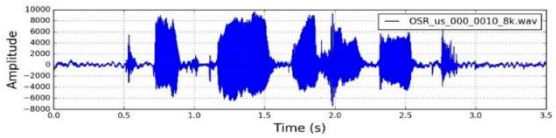

Figure 2: Signal in the Time Domain

The signal is first divided into brief, overlapping frames by the MFCC algorithm; these frames typically last 20 to 40 milliseconds. Each frame is then transformed to the frequency domain using a Fourier transform. The resulting frequency spectrum is then filtered using a sequence of similarly spaced Mel-scale filter banks.

Compared to the linear frequency scale, this perceptual scale is more closely associated with how humans perceive pitch [26]. To obtain the MFCCs, the data are changed using the Discrete Cosine Transform (DCT) after each filter bank's output has been logarithmically scaled. The DCT de-correlated the filter bank outputs to help provide a set of coefficients more appropriate for analysis [27].

The quantity of extracted MFCCs depends on the specific use case, but for speech recognition tasks, 12–13 Usually, coefficients are used. When used as features, the resultant MFCCs can assist machine learning techniques such as Hidden Markov Models (HMMs), which are frequently used in voice recognition.

According to [3, 24], MFCC entails the following crucial actions depicted in Figure 2.1.

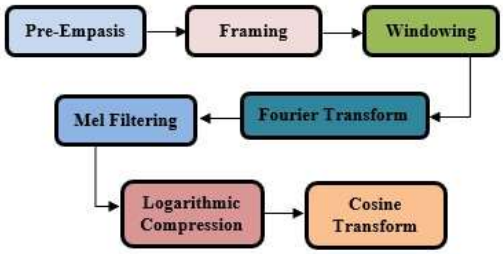

Figure 2.1: MFCC-based Feature Extraction

2.4.1 Pre-emphasis

Pre-emphasis, the initial stage of MFCC processing, is filtering the audio signal to increase the frequency of higher frequencies and decrease the frequency of lower frequencies. As a result, the MFCCs produced may have a stronger signal-to-noise ratio. Pre-emphasis filters have several significant uses. It can also improve the Signal-to-Noise Ratio (SNR).

1. High frequencies typically have smaller magnitudes than lower frequencies to balance the frequency spectrum.
2. Steer clear of numerical issues when doing the Fourier transformations.
3. The SNR might also be improved.

A first-order pre-emphasis filter is applied to the signal x, as represented by the equation:

$$y(t)=x(t)-\alpha x(t-1) \quad (1)$$

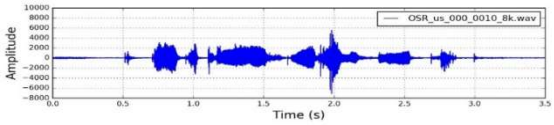

Figure 3: Signal representation in the time domain following pre-emphasis.

2.4.2 Framing

Signal into short time intervals. This is carried out due to a signal's fluctuating frequency throughout time. Therefore, applying the Fourier transform to the full signal is impractical since doing so would eventually result in the loss of frequency information. We can fairly assume that a signal's frequencies stay consistent for a relatively brief length of time to prevent this loss. By joining adjacent frames, we can achieve an accurate estimate of the signal's frequency information by applying a Fourier transform over this brief time interval. Speech processing typically uses frame widths between 20 and 40 milliseconds, with a 50% overlap between successive frames. For the frame size, 25 milliseconds is the recommended value.

2.4.3 Windowing

The next step is to multiply each frame by a window function, usually a Hamming window, to lower spectral leakage and boost FFT accuracy. The following is the Hamming window.

$$w[n]=0.54-0.46\cos(2\pi n/N-1) \quad (2)$$

Where $0 \leq n \leq N-1$, with N representing the window length. When plotted, the equation generates the following graph:

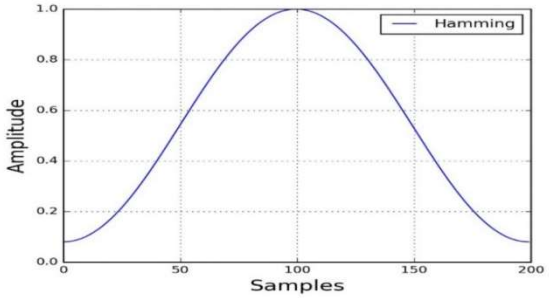

Figure 4: Hamming Window

2.4.4 Fourier-Transform

The FFT is then used to convert the windowed frames from the time domain to the frequency domain. As a result, each frame's audio signal is represented spectrally. The frequency spectrum, also known as the Short-Time Fourier-Transform (STFT), is computed on each frame using an N-point FFT, Typically, N is set to 256 or 512, with NFFT = 512. The power spectrum (periodogram) is then calculated using the equation below.

$$P=(|FFT(xi)|)/ \quad (3)$$

2.4.5 Mel Filtering

A series of triangular Mel filters are then used to convert the spectral representation into the Mel-frequency domain, which is an approximation of the non-linear frequency response of the human auditory system. The next step involves analyzing the power spectrum and extracting frequency bands using a Mel scale and triangular filters to construct the filter banks, typically about 40 filters (n-filt = 40). The Mel-scale is more accurate at lower frequencies and less accurate at higher frequencies to mimic the non-linear manner in which the human ear perceives sound. Additionally, it makes conversion between Mel (m) and Hertz (f) easier with the following formulas:

$$m=2595\log10(1+f700) \quad (4)$$

Triangular filters with a response of 1 at the central frequency make up each filter bank. When the response hits the middle frequencies of the two adjacent filters, it declines linearly toward zero and turns into zero. This is seen in the following Figure 5.

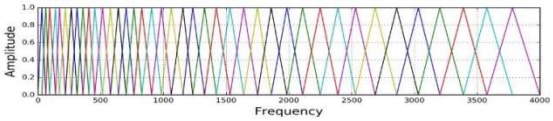

Figure 5: Filter on Mel Scale

2.4.6 Logarithmic Compression

The non-linear behavior of the human auditory system is then further approximated by compressing the Mel-frequency coefficients logarithmically.

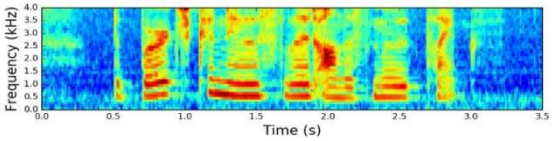

Figure 6: Spectrogram of the Signal after Compression

2.4.7 Discrete Cosine Transform

Finally, the logarithmically compressed coefficients undergo a discrete cosine transform (DCT) to generate the MFCCs. The generated MFCCs are utilized as features for speech an d audio processing tasks and describe the spectral envelope of the audio stream.

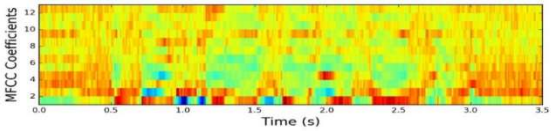

Figure 7: Discrete Cosine Function using MFCC

3. PREDICTION LAYER: DELM-BASED PROPOSED MODEL

This guide Using DELMs for classification is a prominent feed-forward (FF) approach especially for high-dimensional data like audio [30]. Traditional BP algorithms have slow learning rates and require more samples, they may overfit the model. DELM is commonly applied to classification problems across various domains due to its fast learning rate and computational efficiency. The DELM model consists of three main layers: an input layer, several hidden layers, and an output layer. The structure of the DELM model is shown in Figure 8.

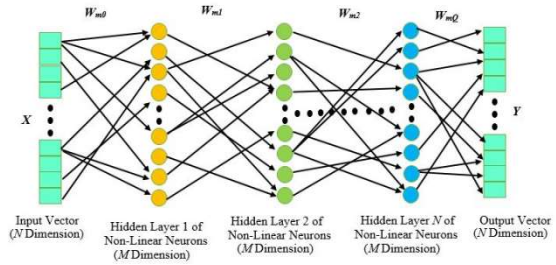

Figure 8: Architecture of Deep Extreme Learning Machines DELM [28]

$$XVT = Wm0\ (VVT + R\ J) \quad (5)$$

Here R is regularization for parameter optimization by another parameter. First consider {Y, Z} = {Xi, Zi} where j ranges from 1 to N, and the input feature Y = [yj1, yj2, yj3… yin] and desired matrix Z = [zj1, zj2, zj3, …..zjn] consists of training examples, the matrix

Y and Z can be denoted as:

$$Y = \begin{bmatrix} y_{11} & y_{12} \dots \dots \dots & y_{1N} \\ y_{21} & y_{22} \dots \dots \dots & y_{2N} \\ y_{n1} & y_{n2} \dots \dots \dots & a_{nN} \end{bmatrix}$$

$$Z = \begin{bmatrix} z_{11} & z_{12} \dots \dots \dots & z_{1N} \\ z_{21} & z_{22} \dots \dots \dots & z_{2N} \\ z_{n1} & z_{n2} \dots \dots \dots & z_{nN} \end{bmatrix} \quad (6)$$

N denotes the dimensions of the input and output matrices, respectively. In ELMs, weights are randomly assigned between the input and hidden layers.

$$W = \begin{bmatrix} w_{11} & w_{12} \dots \dots \dots & w_{1N} \\ w_{21} & w_{22} \dots \dots \dots & w_{2N} \\ w_{n1} & w_{n2} \dots \dots \dots & w_{nN} \end{bmatrix} \quad (7)$$

The weights *W* are shared by the *jth* input and the output are hidden layer neurons. ELM takes into account the following Connections between the hidden layer and the output layer of neurons are represented by weights:

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} \dots \dots \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} \dots \dots \dots & \alpha_{2N} \\ \alpha_{lz} & \alpha_{lz} \dots \dots \dots & \alpha_{nN} \end{bmatrix} \quad (8)$$

This represents the weights between the neurons in the j^{th} hidden layer and those in the j^{th} output layer. The bias for the hidden layer is then randomly assigned by the ELM as follows:

$$\beta = [\beta_1, \beta_2, \beta_3 \dots \dots \dots \beta_n]^T \quad (9)$$

Next, the ELM chooses an activation function (x). The output matrix O can be represented as follows:

$$T = [T_1, T_1, T_3, \dots \dots, T_n]_{N*k} \quad (10)$$

The vector for each column of matrix O is expressed as follows:

$$Oj = \begin{bmatrix} O_{1K} \\ O_{2K} \\ \vdots \\ O_{Nk} \end{bmatrix} = \begin{bmatrix} \sum_1^l \alpha_{i1} g\,(w_1 x_1 + \beta_1) \\ \sum_2^l \alpha_{i2} g\,(w_2 x_2 + \beta_2) \\ \dots \\ \dots \\ \sum_N^l \alpha_{im} g\,(w_N x_k + \beta_N) \end{bmatrix} \quad (11)$$

Considering equation (10) and equation (11):

$$T' = \alpha H \quad (12)$$

T represents the transposed matrix *T*, and *H* refers to the result produced by the hidden layer. The Least Squares Method (LSM) is applied to calculate the values, which correspond to the weighted sum.

$$\alpha = H + T \quad (13)$$

The regularization term is introduced to increase generalization capacity and result in stability.

$$H_{L1} = T\ \alpha^{-1} \quad (14)$$

The inverse of matrix β is represented as β^{-1}. Therefore, Layer 2 can be determined using equation (14).

$$H_{L1} = g(WH_1 + \beta_1) \quad (15)$$

The values W_1, H_{L1}, and H_1 in equation 14 indicate the initial two hidden layers weight matrices, the

first neuron's hidden level preference, the output from the first hidden layer, and the anticipated input for the second hidden layer, respectively.

$$W_{HLE} = g^{-1}(H_1)\, H_E^{+} \quad (16)$$

In solving Equation 8, (g(x)) represents the activation function, and HLE+ is the inverse of HE. Thus, by selecting the appropriate g(x) activation function, the output of the second hidden layer is modified as follows:

$$H_{L2} = g(W_{HE}\, H_E) \qquad Where,\ W_{HE}H_E = N_E th_2$$

$$H_{L2} = g(N_E\, th_2) \quad (17)$$

$H_{L2}+$ is equivalent to H_{L2} after updating the weight matrix between the second and third layers using Equation 17. The predicted outcomes for Layer 3 are shown in Equation 18.

$$\beta_{new} = H_{L2}\, N^{+} \quad (18)$$

$$H_{L3} = \beta^{+}_{new} \quad (19)$$

The inverse of matrix β_{new} is represented as β^{+}_{new}. The DELM assigns the matrix $W_{HLE1} = [\beta_2, W_2]$, and Equations 11 and 12 allow the output of the third layer to be determined as shown in Equation 21.

$$\beta_{new} = H_{L2}^{+}\, N \quad (20)$$

$$H_{L3} = \beta^{+}_{new} \quad (21)$$

DELM then sets the matrix $W_{HE2} = [\beta_3, W_3]$. The fourth layer output is made possible by Equations 11 and 12.

$$H_2 = g^{-1}(H_2\ W_2) = g(N_E\ th_2) \quad (22)$$

$$W_{HE1} = \beta^{-1}(H_{L2})\, H_{LEI}^{+} \quad (23)$$

In equation (19), the hidden layer H_{L2} displays the intended result, and the hidden layer W_2 represents the weight distribution between the second and third hidden layers. The logistic Sigmoid Function (LSF) was applied in Equation 24. The output of the third hidden layer is computed in Equation 25 as follows:

$$g(x) = \frac{1}{1+e^{-x}}$$

$$H_{L2} = g(W_{HE1}\, H_{LE1}) \quad Where\ W_{HE1}H_E = N_E\, th \quad (24)$$

$$H_{L2} = g(N_E\ th_2) \quad (25)$$

Equation 26 calculates the weighted matrix for both the third hidden layer and the final layer.

$$\beta_{new} = H_{L4}^{\ t}\left(\frac{1}{\lambda} + H_{L4}^{\ t} H_{L4}\right)^{-1} N \quad (26)$$

$$H_{L4} = N\beta^{+}_{new} \quad (27)$$

The transposed matrix β_{new} is represented as $N\beta^{+}_{new}$. Following this, the DELM constructs the matrix. $W_{HE2} = [\beta, W_3]$. Equations (14) and (15) are then applied to compute the output for the fourth hidden layer.

$$H_{L4} = g^{-1}(H_{L3} W_3 + \beta_3) = g(N_E\, th_{4,1}) \quad (28)$$

$$W_{ME2} = \beta^{-1}\left((H_{L4})\, M^{+}_{N2}\right) \quad (29)$$

The sigmoidal logistic function is applied in equation 29, and the subsequent measurement of values for the 3rd and 4th hidden layers is as follows:

$$H_4 = g(Neth_{4,2}) \quad (30)$$

Equation (28) calculates the output matrix connecting the *nth* hidden layer to the layers that produce the final output and displays the nth layer's evaluated result. The needed output of the DELM framework is shown in equation (31).

$$\beta_{new} = H_{Lnt}^{\ t}\left(\frac{1}{\lambda} + H_{Lnth}^{\ t} H_{Lnth4}\right)^{-1} N \quad (31)$$

$$M_{nth} = N\,\beta^{+}_{new} \quad (32)$$

$$f(x) = m_{nth}\,\beta_{new} \quad (33)$$

Following the DELM framework nth hidden layer computation process. You may recompute equations (32) and (33) to obtain the hidden-layer parameters. The outcome of the DELM-based network is then computed. When hidden layers are multiplied, the output of additional hidden layers is computed using Equation 34.

$$op = \frac{1}{1+e^{-Nethj}} \quad \text{where } j = 1,2,3, \ldots\ldots, r \quad (34)$$

4. OBTAINED RESULTS

The "Heartbeat Sound" benchmark dataset available at Kaggle is used to train and assess the model. The instance was collected from the general public via the stethoscope (iPhone Pro application), having 176 instances, and from a

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

clinical trial in hospitals using the digital stethoscope having audio 656 files. In total, there are 832 instances in the dataset. The instances in the dataset are audio files which are 2 to 30 seconds of real human heartbeat sound audio files in .wav format.

The original dataset with 832 instances has six classes or labels i.e., 'Normal', 'Artifact', 'Extrastole', 'Extrahls', 'Murmur', and 'unlabelled' are real human heartbeat sound files. Only 575 labeled instances are considered for the trainingand testing of the model. Figure 9 represents the labeled instances w.r.t class distribution in the dataset in the form of a pie chart.

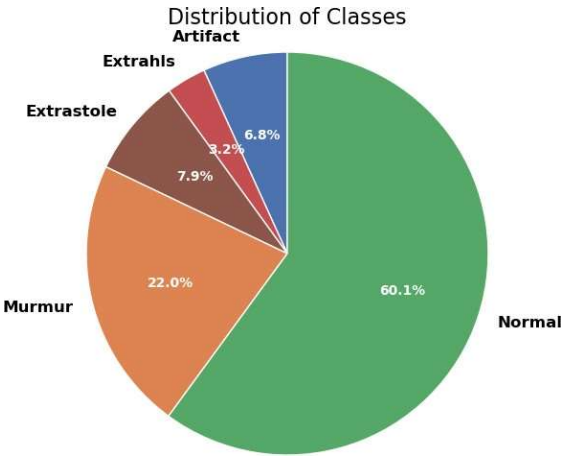

Figure 9: Distribution of Classes in the Dataset with Five Classes MFCC

The dataset is split into an 80 % training set with 468 instances and a 20 % test set with 107 instances using a random split. The training results are presented in Table I and Table II.

Table I: Confusion Matrix Obtained From Training Results With Five Classes.

Confusion Matrix						
Training		Predicated values				
		Arti-fact	Mur-mur	Extra-hls	Extras-tole	Normal
Actual Values	Artifact	30	1	0	0	1
	Extrahls	0	13	0	2	0
	Extrastole	0	0	37	0	0
	Murmur	0	2	0	99	2
	Normal	0	7	0	3	271

Table Ii: Evaluation Matrix Of Training Results With FiveClasses

Class	Accuracy	Precision	Recall	F1-Score
Artifact	99.57 %	0.94	1.0	0.97
Extrahls	97.44 %	0.87	0.57	0.68
Extrastole	100 %	1.0	1.0	1.0
Murmur	98.08 %	0.96	0.95	0.96
Normal	97.22 %	0.96	0.99	0.98

The test results are presented in Table III and Table IV.

Table III: CONFUSION MATRIX BTAINED FROM TESTING RESULTS WITH FIVE CLASSES.

Confusion Matrix						
Testing		Predicted Values				
		Arti-fact	Mur-mur	Extr-ahls	Extra-stole	Normal
Actual Values	Artifact	4	1	0	2	1
	Extrahls	0	4	0	0	0
	Extrastole	0	0	2	0	7
	Murmur	0	1	1	9	15
	Normal	0	2	2	6	60

Table IV: EVALUATION MATRIX OBTAINED OF TEST RESULTS WITH FIVE CLASSES

Class	Accuracy	Precision	Recall	F1-Score
Artifact	96.58%	0.5	1	0.67
Extrahls	96.58%	1	0.5	0.67
Extrastole	91.45%	0.22	0.4	0.29
Murmur	78.63%	0.35	0.53	0.42
Normal	71.79%	0.86	0.72	0.78

Figure 10: Comparison between Average Training and Testing Results with Five Classes

The main reason to get high training results and low testing results is model overfitting. The original dataset is highly biased and imbalanced because it contains 60 % of instances belonging to the 'Normal' class only. To obtain maximum accuracy the dataset class aggregation is applied. Now the dataset contains three classes depicted in Figure 12. The trainingresult with 3 labled classes are depicted in Table V and Table VI. The test results with 3 labled classes are illustrated in Table VII and Table VIII.

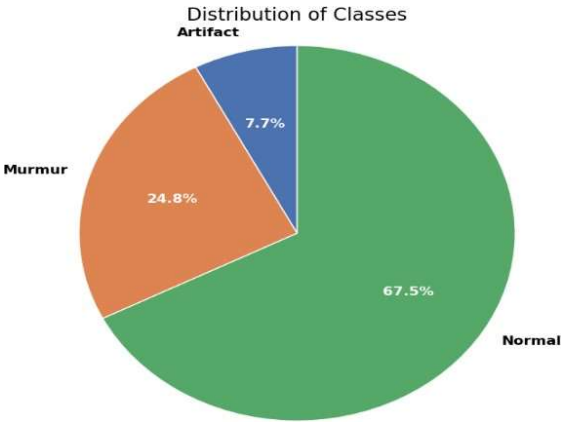

Figure 11. Distribution of Classes in the Dataset with Three

Table V. Confusion Matrix Obtained From Training Results With Three Classes

Confusion Matrix				
Training		Predicted Values		
		Artifact	Murmur	Normal
Actual Values	Artifact	31	1	0
	Murmur	0	103	0
	Normal	1	1	280

Table Vi. Evaluation Matrix Obtained From Training Results With Three Classes

Class	Accuracy	Precision	Recall	F1-Score
Artifact	99.52 %	0.97	0.97	0.97
Murmur	99.52 %	1.0	0.98	0.99
Normal	99.52 %	0.99	1.0	1.0

Table Vii. Confusion Matrix Obtained From Test Results With Three Classes

Confusion Matrix				
Training		Predicted Values		
		Artifact	Murmur	Normal
Actual Values	Artifact	7	1	0
	Murmur	0	22	4
	Normal	0	7	63

Table Viii. Evaluation Matrix Obtained From Test Results With Three Classes

Class	Accuracy	Precision	Recall	F1-Score
Artifact	99.04 %	0.88	1.0	0.93
Murmur	86.54 %	0.77	0.71	0.64
Normal	87.50 %	0.90	0.86	0.88

The training and test results are compared in Figure 12.

Figure 12. Comparison between Average Training and Testing Results with

5. DISCUSSIONS

Metrics such as accuracy, precision, recall, and F1- score play crucial roles in providing a comprehensive understanding of model performance. Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total instances. Precision assesses the proportion of true positive predictions among all positive predictions, indicating how well the model avoids false positives. Conversely, recall evaluates the model's ability to identify all relevant instances, focusing on the true positives among actual positives and highlighting the importance of minimizing false negatives. The F1-score serves as a harmonic mean of precision and recall, offering a single metric that balances both concerns, particularly useful in scenarios with imbalanced datasets. Together, these metrics provide a nuanced perspective on a model's strengths and weaknesses, guiding improvements and ensuring effective deployment in real-world applications.

The 'Heart-Sound' dataset with 575 labeled instanceswith 5 classes was considered for the training and testing ofthe model. The average training accuracy reached up to 98.46 % with precision 94.6 %, recall 90.2 %, and F1 Score 91.8 % as illustrated in Table 1 and Table 2. The test resultsachieved by the proposed model trained with 5 label classes were recorded as average test accuracy obtained up to 88.80% with precision of 29.33 %, Recall 60.31 %, and F1 Score 56.66 %

shown in Table 3 and Table 4. The obtained results highlighted two keen observations. First, the training resultis very high and the testing result i.e. 88.8 % is slightly lowas compared with training results i.e. 98.46 %. It was a clear indication of model overfitting. Secondly, the test results with 5 classes also indicated that high accuracy with low precision, recall, and F1 score is a clear indication demonstrated by Figure 10 that the dataset is highly imbalanced as can be observed in Figure 9. This is a very common issue in all PCG-related datasets because some ofthe heart conditions are very rare and difficult to record because of very low frequency and pitch.

To obtain a highly accurate and reliable PCG model the dataset is aggregated from 5 labeled classes to 3 labeledclasses. The dataset instance distribution w.r.t label can be observed using Figure 11. The proposed model obtained 99.52 % accuracy with 98.66 % precision, 98.33 % recall, and 98.66 % F1-score illustrated in Table 5 and Table 6. The test results of the model reached up to 92.30 % accuracy with 87.66 % precision, 89.01 % recall, and 88.0 % F1- score presented in Table 6 and Table 7. The comparison between training and testing with average accuracy, precision, recall, and F1 score is demonstrated using Figure 12. The obtained training accuracy i.e. 99.52 % and test accuracy of 92.30 % depicts that now the proposed model for DELM-based PCG framework is not overfitted. Also, the other evaluation parameters like precision, recall, and F1 have achieved high results indicating that the model is not biased. The experimental results advocate the efficiencyof the trained model with high accuracy, least overfitting, and high unbiased. The comparison between these improved results can be visualized using Figure 13.

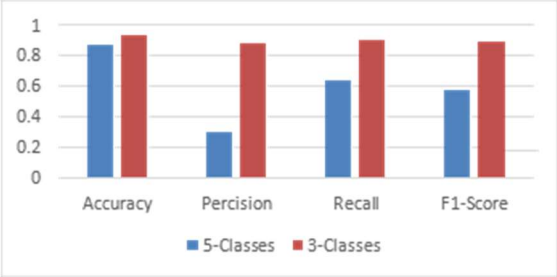

Figure 13: Average Test Results Comparison Obtained From Test Results With Five Classes And Three Classes

The proposed model for DELM based PCG system is highly accurate and reliable as shown by the experimental outcomes. DELM is based on Extreme Learning Machines (ELM) and has many hidden layers, random weights, and biases. The efficiency of the proposed system is also very important to expect which hugely impacts its feasibility, applicability, and usability. The DELM-based model was trained using differenthidden layers. With six hidden layers, the proposed DELM- based PCG framework's model produced an optimal training accuracy of 98.6 % and a Test accuracy of 92.3 % with$1.31*10^{-3}$ and $1.28*10^{-3}$ Root Mean Square Error (RMSE) respectively. With the introduction of a seventh and eighth hidden layer, only training results are improving. The accuracyof the model is primarily evaluated based on the test results, which do not improve with additional layers. Instead, introducing more hidden significantly increases the computational complexity of the proposed model. Theperformance of the proposed model based on hidden layers isillustrated in Table IX. Where HL stands for a number of hiddenlayers. 'Acc' represents accuracy, MR stands for miss-rate, and Time (training and testing) is mentioned in milli seconds.

Table Ix. Performance Of Proposed Delm-Based Pcg System

	Training Results				Testing Results			
HL	**Acc %**	**MR %**	**Time**	**RMSE**	**Acc %**	**MR %**	**Time**	**RMSE**
4	97.7	0.9	7.6	$1.36*10^{-3}$	91.6	8.4	5.5	$1.39*10^{-3}$
5	98.2	0.8	8.9	$1.33*10^{-3}$	92.1	7.9	5.9	$1.36*10^{-3}$
6	**98.6**	**0.5**	**9.2**	**$1.31*10^{-3}$**	**92.3**	**7.7**	**6.4**	**$1.28*10^{-3}$**
7	98.7	0.3	11.6	$1.29*10^{-3}$	92.3	7.6	8.1	$1.27*10^{-3}$
8	98.8	0.3	13.7	$1.28*10^{-3}$	92.3	7.5	8.7	$1.26*10^{-3}$

The DELMs-based suggested framework achieved remarkable training results and testing results. The results of the effectiveness of the proposed MFCC & DELM-based PCG framework. The proposed framework is a reliable contribution to the field of ML-based PCG systems.

6. FUTURE WORK

In the future, the following improvements can be made to make the framework more effective and accurate. The integration of more precise audio sound recording, noise reduction, and

segmentation, techniques, etc. Multiple benchmark datasets can also be involved to train and evaluatethe model performance. With more hardware resources the model could be trained on large datasets and different ML- based classifiers and learning architectures, such as convolutional neural networks, etc., can be used to build the model for more accuracy and reliability. Future studies could also address other important issues like environmental and natural factors.

Compared to existing PCG frameworks, the proposed model exhibits superior performance in addressing dataset imbalance and mitigating model overfitting. Unlike BP-based neural networks, DELM significantly reduces training complexity while achieving higher test accuracies. A Plus-Minus-Interesting (PMI) as Table X analysis revealed that while DELM enhances generalization, its reliance on high-quality datasets poses limitations, which can be mitigated through further integration of advanced noise reduction techniques.

Table X. Table X: PMI Analysis of the Novel MFCC-DELM Framework

Aspect	Strengths (Plus)	Weaknesses (Minus)	Interesting Observations
Dataset Handling	Mitigates imbalance through class aggregation	Limited generalizability to unseen conditions	Aggregation significantly improves recall
Feature Extraction	MFCC captures nuanced frequency features effectively	Computationally expensive for larger datasets	Enhances detection accuracy for anomalies
Learning Algorithm	DELM minimizes gradient issues and overfitting	Requires precise tuning of parameters	Achieves high accuracy in resource-limited settings

7. CONCLUSION

The findings underscore the novelty of combining MFCC and DELM in addressing critical challenges such as dataset bias and model overfitting. This research provides a foundational framework for developing scalable and reliable ML-based PCG systems tailored for low-resource healthcare environments. In resource-intensive environments where a lack of cardio specialists, and state-of-the-art equipment is not accessible PCG is the only technique used to predict cardiovascular- related medical emergencies. Besides its usefulness, PCG faces some challenges as well. ML-based PCG systems couldbe a better solution for tackling these challenges but ML-based PCG systems have their limitations and strains as well. This study focuses on three main issues like unavailability of appropriate datasets, the vanishing and exploding of gradients, and inappropriate feature extraction techniques. This study presented a novel MFCC & DELM-based framework for PCG systems. This study argues MFCC as a suitable feature extraction technique, DELM as an FF learning algorithm, anddataset aggregation. The proposed framework remarkably achieved a training accuracy of 98.6 % and a Test accuracy of 92.3 % with $1.31*10^{-3}$ and $1.28*10^{-3}$ RMSE respectively. This study contributes to the development of more accurate, reliable, and usable PCG systems for cardiovascular disease and emergency diagnosis.

ACKNOWLEDGMENT

I want to extend my deepest gratitude to my supervisor, Prof. Dr. Hairulnizam Mahdin, for his invaluable guidance and support, the Computer Science Department (FSKTM) at UTHM for their resources and encouragement, my colleagues for their insightful discussions, the participants for their cooperation and unwavering support, all of which made this research possible.

REFERENCES:

[1] Asmare, M. H., Filtjens, B., Woldehanna, F., Janssens, L., & Vanrumste, B. (2021). Rheumatic heart disease screening based on phonocardiogram. Sensors, 21, 1-17.

[2] World Health Organization. (2023, January 8). Cardiovascular diseases (CVDs). Retrieved from https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[3] Li, S., Li, F., Tang, S., & Xiong, W. (2020). A review of computer-aided heart sound detection techniques. BioMed Res., 1-10.

[4] Morris, K. (2021). Design of an Esophageal Deflection and Thermal Monitoring Device for Use During Cardiac Ablation Procedures. University of California, San Diego.

[5] Chakir, F., Jilbab, A., Nacir, C., & Hammouch, A. (2018). Phonocardiogram signals processing approach for PASCAL Classifying Heart Sounds Challenge. Signal Image Video Process, 12, 1149-1155.

[6] Ijaz, A., Nabeel, M., Masood, U., Mahmood, T., Hashmi, M. S., Posokhova, I., ... & Imran, A. (2022). Towards using cough for respiratory

disease diagnosis by leveraging Artificial Intelligence: A survey. Informatics in Medicine Unlocked, 100832.

[7] Xiao, B., Xu, Y., Bi, X., Zhang, J., & Ma, X. (2020). Heart sound classification using a novel 1-D convolutional neural network with extremely low parameter consumption. Neurocomputing, 392, 153-159.

[8] Chen, J., Sun, S., Zhang, L. B., Yang, B., & Wang, W. (2021). Compressed sensing framework for heart sound acquisition in Internet of medical things. IEEE Transactions on Industrial Informatics, 18(3), 2000-2009.

[9] Chowdhury, M. E., Khandakar, A., Alzoubi, K., Mansoor, S., Tahir, A., Reaz, M. B. I., & Al-Emadi, N. (2019). Real-time smart-digital stethoscope system for heart disease monitoring. Sensors, 19(12), 2781.

[10] Omarov, B., Saparkhojayev, N., Shekerbekova, S., Akhmetova, O., Sakypbekova, M., Kamalova, G., ... & Akanova, Z. (2022). Artificial intelligence in medicine: Real-time electronic stethoscope for heart diseases detection. Computers, Materials & Continua, 70(2).

[11] Jamil, S., & Roy, A. M. (2023). An efficient and robust phonocardiography (PCG)-based valvular heart diseases (VHD) detection framework using Vision Transformer (ViT). Computers in Biology and Medicine, 106734.

[12] Athreya, A. M., Paramesha, K., Avani, H. S., & Madhu, S. (2022). Neural networks for detecting cardiac arrhythmia from PCG signals. In Intelligent Vision in Healthcare (pp. 103-115). Springer, Singapore.

[13] Thalmayer, A., Zeising, S., Fischer, G., & Kirchner, J. A. (2020). Robust and real-time capable envelope-based algorithm for heart sound classification: Validation under different physiological conditions. Sensors.

[14] Kapen, P. T., Youssoufa, M., Kouam, S. U. K., Foutse, M., Tchamda, A. R., & Tchuen, G. (2020). Phonocardiogram: A robust algorithm for generating synthetic signals and comparison with real-life ones. Biomed. Signal Process. Control, 60, 101983.

[15] Wei, W., Zhan, G., Wang, X., Zhang, P., & Yan, Y. (2019). A novel method for automatic heart murmur diagnosis using phonocardiogram. In Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM) (pp. 1-6).

[16] Malarvili, M., Kamarulafizam, I., Hussain, S., & Helmi, D. (2003). Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram. In Comput. Cardiol. (pp. 327-330).

[17] Liu, Q., Wu, X., & Ma, X. (2018). An automatic segmentation method for heart sounds. Biomed. Eng. Online, 17, 22-29.

[18] Chen, T. E., Yang, S. I., Ho, L. T., Tsai, K. H., Chen, Y. H., Chang, Y. F., & Wu, C. C. (2017). S1 and S2 heart sound recognition using deep neural networks. IEEE Trans. Biomed. Eng., 64, 372-380.

[19] Oliveira, J. H., Renna, F., & Mantadelis, T. (2018). Adaptive sojourn time HSMM for heart sound segmentation. IEEE J. Biomed. Health Inform., 23, 642-649.

[20] Kamson, A. P., Sharma, L., & Dandapat, S. (2019). Multi-centroid diastolic duration distribution based HSMM for heart sound segmentation. Biomed. Signal Process. Control., 48, 265-272.

[21] Renna, F., Oliveira, J. H., & Coimbra, M. T. (2019). Deep convolutional neural networks for heart sound segmentation. IEEE J. Biomed. Health Inform., 23, 2435-2445.

[22] Liu, C., Springer, D., & Clifford, G. D. (2017). Performance of an open-source heart sound segmentation algorithm on eight independent databases. Physiol. Meas., 38, 1730-1745.

[23] Ding, X., Zhang, Y., & Tsang, H. K. (2016). Impact of heart disease and calibration interval on accuracy of pulse transit time–based blood pressure estimation. Physiological measurement, 37(2), 227.

[24] Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process., 28(4), 357-366.

[25] Ghosh, S., Tripathy, S., Saha, G., & Paul, S. (2021). An improved feature extraction method for speech recognition using deep learning. In 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 728-733).

[26] Wang, M., Guo, B., Hu, Y., Zhao, Z., Liu, C., & Tang, H. (2022). Transfer learning models for detecting six categories of phonocardiogram recordings. Journal of Cardiovascular Development and Disease, 9(3), 86.

[27] Nguyen, T. T., Wang, R., & Nguyen, T. T. (2020). A comparative study of feature extraction techniques for Vietnamese speech recognition. In 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS) (pp. 188-193).

[28] Rizvi, S. S. R., Khan, M. A., Abbas, S., Asadullah, M., Anwer, N., & Fatima, A. (2022). Deep extreme learning machine-based optical character recognition system for Nastalique Urdu-like script languages. The Computer Journal, 65(2), 331-344.

[29] Deng, S. W., & Han, J. Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. Future Gener. Comput. Syst., 60, 13-21.

[30] Rizvi, S. S. R., Khan, M. A., Abbas, S., Asadullah, M., Anwer, N., & Fatima, A. (2022). Deep extreme learning machine-based optical character recognition system for nastalique urdu-like script languages. The Computer Journal, 65(2), 331-344.