Comparative Analysis of Feature Selection Based on Metaheuristic Methods for Human Heart Sounds Classification Using PCG Signal

Motaz Faroq A Ben Hamza, Nilam Nur Amir Sjarif
Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia

***Abstract*—Cardiovascular disease is a critical threat to human health, as most death cases are due to heart disease. Although several doctors employ stethoscopes to auscultate heart sounds to detect abnormalities, the accuracy of the approach is considerably dependent upon the experience and skills of the physician. Consequently, optimal methods are required to analyse and classify heart sounds with Phonocardiogram (PCG) signal-based machine learning methods. The current study formulated a binary classification model by subjecting PCG signals to hyper-filtering with low-pass and cosine filters. Subsequently, numerous features are extracted with the Wavelet Scattering Transform (WST) method. During the feature selection stage, several metaheuristic methods, including Harris Hawks Optimisation (HHO), Dragonfly Algorithm (DA), Grey Wolf Optimiser (GWO), Salp Swarm Algorithm (SSA), and Whale Optimisation Algorithm (WOA), are employed to compare the attributes separately and determine the ideal characteristics for improved classification accuracy. Finally, the selected features were applied as input for the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm, simplifying the classification process for distinguishing normal and abnormal heart sounds. The present study assessed three PCG datasets: PhysioNet 2016, Yaseen Khan 2018, and PhysioNet 2022, documenting 94.85%, 100%, and 66.87% accuracy rates with 127-SSA, 168-HHO, and 163-HHO, respectively. Based on the results of the PhysioNet 2016 and 2022 datasets, the proposed method with hyperparameters demonstrated superior performance to those with default parameters in categorising normal and abnormal heart sounds appropriately.**

Keywords*—Cardiovascular Diseases (CVDs); Phonocardiogram (PCG) signal processing; Wavelet Scattering Transform (WST); Metaheuristic Methods; Harris Hawks Optimisation (HHO); Dragonfly Algorithm (DA); Grey Wolf Optimiser (GWO); Salp Swarm Algorithm (SSA); Whale Optimisation Algorithm (WOA); Bidirectional Long Short-Term Memory (Bi-LSTM)

I. Introduction

Diagnosis with heart sounds has drawn much attention in the biomedical research area because of the crucial nature of the heart and the high mortality rates associated with cardiovascular diseases. Heart sounds are the mechanical activities of the heart, which vary according to pathological conditions affecting it [1]. Currently, the two major expensive technologies for detecting heart diseases are echocardiography and cardiac Magnetic Resonance Imaging (MRI) [2]. Auscultation is a basic diagnostic technique commonly used to assess heart function and quality [3]. It involves the listening of heart sounds directly by placing a stethoscope at various points in the chest. Nonetheless, it is subjective and highly dependent on the acuteness of hearing and experience of the physician. Therefore, there is a need to develop a system that can objectively process heart sounds, enabling faster and more accurate diagnoses.

There are two major sounds produced by a normally functioning heart. The first heart sound (S1) and the second heart sound (S2). Closing of the mitral and tricuspid valves causes the first sound, S1, while the second sound, S2, is a result of the aortic and pulmonary valves closing at the end of the systolic phase. Sound identification can be one essential part of the heart disease diagnosis since the nature and other characteristics of the heart murmurs can vary from one to another as per concise disease conditions. In most instances, differences in heart sound patterns between healthy and unhealthy states are distinguished on the basis of changes in intensity, timing, location, etc., among other factors [4].

Several researchers have worked on the study and categorisation of Phonocardiogram (PCG) signals using various machine learning techniques [5]; particularly, two main areas have been focused on: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [6], [7]. However, these studies often require extraction of a wide array of features, which can complicate the training phase when dealing with heart sound signals. Also, these studies involve issues of existing noise in PCG signals and imbalanced datasets, leading to the extraction of unnecessary features [8]. Therefore, the contribution of this study is as follows:

- To introduce low-pass and cosine hyper-filters during preprocessing to reduce background noise from PCG signals.
- To extract various features from the PCG signals with Wavelet Scattering Transform (WST).
- To develop metaheuristic methods with hyperparameters to select optimal features for refined features from an initial set of features, therefore reducing computational complexity and improving classification performance by diminishing the search space.
- To execute high-performance heart sound classification with the Bidirectional Long Short-Term Memory (Bi-LSTM) classifier.

Section II of this study discusses related articles on heart sound classification. Details on the WST characteristics, the proposed methodology with several metaheuristic methods, and the Bi-LSTM algorithm are included in Section III. In Section IV, the results, including the dataset employed for comparison with existing literature, are discussed. The conclusions of this study are mentioned in Section V. Recommendations for future research are also included.

II. Relevant Work

In the biomedical field, several techniques have been proposed to classify normal and abnormal heart sounds, emphasising Phonocardiogram (PCG) signal processing. Generally, the signals are preprocessed to filter out extraneous high-frequency noises. Subsequently, characteristic features are extracted and utilised as input data to construct or develop mathematical models intended for diagnosing heart conditions [8], [9].

Preprocessing is fundamental in PCG categorisation. The step involves three primary tasks: baseline wander removal, background noise reduction, and normalisation. Baseline wander is a common issue in low-frequency PCG recordings. This issue shifts the baseline reference level, which hinders accurate signal property extraction [10], [11]. Background noise is prevalent in high-frequency PCG recordings [12] and can degrade signal quality. Normalisation scales sample values in the dataset to a standard range (typically 0-1 or -1 to 1), ensuring that amplitude discrepancies do not distort feature extraction and classification processes [13]. This standardisation prevents bias in the model due to variations in recording amplitudes [14].

Potes et al. [15] classified heart sound recordings from the PhysioNet Challenge 2016 into normal and abnormal classes with an ensemble classifier. The report applied the Butterworth bandpass filter and extracted 124 features from PCG signals with Mel Frequency Cepstral Coefficients (MFCCs). The approach achieved an 86.02% accuracy rate without employing any feature reduction technique. The study also achieved first place in the PhysioNet Challenge 2016.

In 2017, Kay and Agarwal [16] applied a hidden semi-Markov model for segmentation before extracting temporal and spectral features utilising continuous WST and MFCCs from the PhysioNet Challenge 2016 dataset. Subsequently, Principal Component Analysis (PCA) was employed to reduce data dimensionality. Finally, the information was fed into an Artificial Neural Network (ANN), yielding an 85.2% accuracy. In another study, Bao et al. revealed that the Bi-LSTM classifier performed better than CNN on an identical dataset. Accuracy, sensitivity, and specificity rates of 92.64%, 84.77%, and 95.14%, respectively, were noted [17].

Li et al. [18] utilised the Twin SVM (TWSVM) classifier to compare the performance of multi-dimensional scaling and PCA for feature selection on the PhysioNet Challenge 2016 dataset. The Multi-Dimensional Scaling (MDS) outperformed PCA with 98.58%, 98.58%, 98.57%, and 99% in respective accuracy, sensitivity, specificity, and F1 score. Meanwhile, Alshamma et al. (2019) applied a high-pass filter on each PCG signal before employing a normalisation method based on zero mean and standard deviation. Subsequently, different K-Nearest Neighbours (KNN) and Support Vector Machines (SVM) classifiers were compared with the PhysioNet Challenge 2016 dataset. The fine-KNN classifier documented interesting results with a 93.5% accuracy [11].

In 2020, Singh and Majumder obtained short and non-segmented signals from the PhysioNet Challenge 2016 dataset. High noise frequencies were filtered with the Butterworth low-pass filter before filtering 27 features with MFCCs. Post-training with an ensemble classifier, the features recorded 92.47% accuracy, 94.08% sensitivity, and 91.95% specificity [19].

An energy envelogram was employed to extract PCG signals from the PhysioNet Challenge 2016 dataset (Potdar, 2021) [20]. The traits were procured with Discrete Wavelet Transform (DWT), selected utilising PCA, and fed into a fine-tuned Bayesian-optimised SVM algorithm. On the other hand, Milani et al. (2021) applied Springer's segmentation algorithm to preprocess the PhysioNet Challenge 2016 dataset, extracting the attributes with MFCCs. Subsequently, Linear Discriminant Analysis (LDA) was applied to further diminish dimensionality. The ANN-based classification recorded a 93.33% accuracy [12].

In 2022, Zhang et al. extracted numerous features in two-dimensional convolution from the PhysioNet Challenge 2016 dataset. Subsequently, the Particle Swarm Optimisation (PSO) was subjected to feature selection before being trained with the convolutional Bi-LSTM network. The report documented accuracy, sensitivity, and specificity of 91.93%, 91.58%, and 92.27%, respectively [14].

A recent study [21] employed a VGG16 model during the spectrogram trait extraction of 1,330 PCG signals from the PhysioNet Challenge 2016 dataset, achieving an 88.84% accuracy. Nevertheless, the authors concluded that the accuracy level was insufficient for reliable patient diagnosis, requiring a more substantial and balanced dataset for performance enhancement. Although the original dataset was extensive, its imbalance might result in biased results. Consequently, implementing more sophisticated balancing techniques rather than simple down-sampling could improve the outcomes. Moreover, the report did not remove noise, a critical aspect of real-world PCG analysis.

Son and Kwon (2018) [22] constructed the Yaseen Khan 2018 dataset, consisting of 1,000 PCG signals. The report classified heart sounds via three algorithms: SVM, KNN, and Deep Neural Network (DNN). The MFCCs and DWT were also applied during feature extraction. The SVM exhibited the best performance, recording accuracy, sensitivity, and specificity of 97.9%, 98.2%, and 99.4%, respectively.

Flores-Alonso et al. applied a smoothing filter, normalised, and segmented the noisy Yaseen Khan 2018 dataset. The report integrated features for classification utilising CNN and Multi-Layer Perceptron (MLP), including MFCCs, DWT, and Continuous Wavelet Transform (CWT). The report recorded an accuracy of 99.8% [23]. In another study, [24] utilised 957 PCG signals from an identical dataset. Nonetheless, the study

excluded samples under 2 s. Multiple features were extracted with MFCCs and DWT. The data obtained was then employed to train classification models with five machine learning algorithms: Random Forests (RF), KNN, SVM, Naive Bayes (NB), and MLP. A remarkably high accuracy of 99.89% in diagnosing normal and abnormal heart sounds with the RF algorithm was procured.

McDonald et al. [25] focused on segmentation techniques in PCG signals, separating data into S1, S2, systole, and diastole. The study also applied hidden semi-Markov models to the PhysioNet Challenge 2022 dataset. Furthermore, a classification model was developed utilising a Bi-GRU algorithm. The 60.2% accuracy documented led the study to win the competition. Meanwhile, Singh et al. [26] applied Mel-spectrograms to extract features from the same dataset. The extracted features were classified with the U-Net method. The report achieved 56.8%, 54.77%, 59.29%, and 58.33% accuracy, sensitivity, specificity, and F1 score, respectively.

During the past five years, most research employed the Butterworth filter for background noise reduction [19], [27] and segmentation techniques [12], [20] for preprocessing. Nonetheless, segmenting non-linear PCG datasets remains challenging as it might lead to valuable information loss through abstract feature extractions [19], [28]. Moreover, a Wavelet Transform (WT) method has been broadly applied to extract PCG signal features. Nevertheless, the technique often extracts redundant traits, further degrading model performance [29], resulting in numerous research studies focusing on traditional feature extraction selection methods, such as LDA and PCA [12], [16]. The recent review by study [8] suggested using the WT method with metaheuristic methods for feature analysis from PCG signals, which allows the LSTM algorithm to generate the optimal model that can achieve a high classification accuracy.

Based on the literature reviewed in this section, heart sound classification was performed through various approaches. An increasing interest in feature selection and optimisation techniques to improve classification accuracy has also been observed. Although substantial strides have been achieved, handling and minimising non-linear PCG dataset feature redundancies are still challenging.

Consequently, advanced preprocessing and feature selection methodologies necessitate further exploration to improve the accuracy and reliability of heart sound classification systems. Preprocessing heart sounds to eliminate noise is a promising avenue for future studies. Furthermore, combining multiple extracted features into a single representation and training the model on the enriched data might yield superior results.

III. Proposed Methodology

The current study reduced PCG signal noise in the preprocessing with Butterworth and cosine filters. Subsequently, the signal attributes were automatically extracted utilising WST. The characteristics were optimally selected through several metaheuristic methods, including Harris Hawks Optimisation (HHO), Dragonfly Algorithm (DA), Grey Wolf Optimiser (GWO), Salp Swarm Algorithm (SSA), and Whale Optimisation Algorithm (WOA), contributing to feature selection.

The results revealed that the proposed method could achieve excellent classification accuracy with a few feature sets. The suggested model also involved three processes: preprocessing, feature extraction and selection, and classification, as shown in Fig. 1. Each procedure is detailed in subsections A to D.

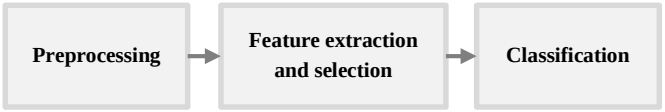

Fig. 1. The general process diagram of the current study.

A. Preprocessing of the PCG Signal

The preprocessing phase is vital in improving Phonocardiogram (PCG) signal quality before extracting and classifying any features. The following lists the steps necessary during the procedure.

1) Baseline wander removal: Baseline wander refers to low-frequency noise typically arising from respiratory movements or patient motion. The noises might distort PCG signals. Consequently, a high-pass Butterworth filter with a 0.5 Hz cut-off frequency was employed to overcome the matter. The filter effectively attenuated frequencies associated with baseline drift. Moreover, heart sound signals predominantly detected between 20 and 150 Hz were preserved. The findings indicated that removing low-frequency noise is critical to avoid interference during the feature extraction phase.

2) Background noise removal: The PCG signals are susceptible to various noise sources, including ambient, lung, and muscle contraction sounds. Accordingly, this study applied a two-stage filtering process to enhance the Signal-to-Noise Ratio (SNR). Firstly, a Butterworth low-pass filter at a 150 Hz cut-off frequency was employed to eliminate high-frequency noises. Subsequently, the Adaptive Noise Cancellation (ANC) utilised allowed noise component estimations within the PCG signal by referencing a separate noise signal and subtracting them from the PCG data. The two-phase approach effectively isolated relevant heart sound components from the background noise, facilitating accurate feature extraction and classification.

3) Data normalisation: The PCG signal amplitudes vary significantly depending on the recording environment, device, and patient physiology. Consequently, this study employed data normalisation to address the issue. The procedure scaled signal amplitudes to a common range, typically between 0 and 1 or −1 and 1. The min-max normalisation in the present study was calculated according to following the equation.

$$x' = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

Where x' refers to normalised PCG signals, x denotes the original PCG signal, $max(x)$ represents the maximum PCG value, and $min(x)$ is the minimum PCG value.

B. Feature Extraction

Feature extraction transforms preprocessed PCG signals into classification-appropriate representations. Various methods, including MFCCs, DWT, and WST, have been employed in previous studies.

1) Mel-frequanct Cepstral Coefficient (MFCC): The MFCCs capture spectral PCG signal properties by emphasising perceptually relevant frequency bands. The process involves the following steps [19]:

a) Framing - Dividing the filtered PCG signal into overlapping 5-s frames.

b) Spectrum estimation - Calculating the amplitude spectrum of each frame.

c) Windowing - Multiplying each frame by a Hamming window to reduce spectral leakage through signal end section attenuation to zero.

d) Fourier transform - Applying the Fast Fourier Transform (FFT) to convert the time-domain signal into the frequency domain.

e) Mel Filterbank - Passing the spectrum through a series of triangular bandpass filters spaced according to the Mel scale.

f) Logarithm – Computing the Mel Filterbank output algorithm to mimic the human auditory system perception.

g) Discrete Cosine Transform (DCT) – Applying the DCT to the log-filterbank outputs to decorrelate the coefficients. Resultantly, 27 MFCCs in the time-frequency domain for each PCG signal were obtained. Fig. 2 demonstrates the MFCCs extraction phases, which were determined based on following the equation.

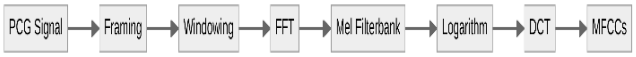

Fig. 2. The MFCC method for feature extraction.

$$MFCC_n = \sum_{k=1}^{K} log(S_k) \cdot cos\left(\frac{\pi n(k-0.5)}{K}\right) \quad (2)$$

Where S_k is the log power at each Mel frequency k, K refers to the total number of Mel frequency bands, and N denotes the number of MFCCs to retain.

2) Discrete Wavelet Transform (DWT): The current study utilised DWT to analyse the non-stationary nature of the PCG signals by iteratively decomposing them into time-frequency domains through wavelet filters [30]. The technique fuses low- and high-pass filters, providing approximation and detail coefficients from the PCG signal [31].

The DWT enables high-frequency energy in systolic and diastolic heart sound alteration representations [32]. The approach captures high-frequency components (details) and low-frequency components (approximations), crucial for detecting anomalies in heart sounds. Meanwhile, following the equation determines wavelet coefficients by projecting x(t) signals onto scaled and shifted wavelet ψ(t) function versions.

$$DWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^*\left(\frac{t-b}{a}\right) dt \quad (3)$$

Where $DWT(a,b)$ represents the wavelet coefficient at scale a and position b, x(t) is the input signal, ψ^* denotes the complex conjugate of the wavelet function ψ, a represents the scaling parameter, and b is the translation parameter.

In this study, the approximation coefficient (c, *A-n.*) and the detail coefficient (*c*, *D-n.*) of each level, n, was established with following the equation. The DWT applied in the present study had 15 decomposition levels for each cycle. The signal energy, entropy, and standard deviation were procured as 16-time features, while the waveform length and wavelet variance estimates were obtained as 16-frequency features.

$$A = cA_n \sum_{k=0}^{n} cD_n \quad (4)$$

Where *A* represents the wavelet coefficient values, while *n* indicates the approximate level number.

Utilising the DWT method, 80 features were extracted from the 15 approximation coefficients for each cycle per the guidelines outlined by [31]. The coefficients were calculated based on the Shannon energy. Fig. 3 represents a generalised flow of the wavelet decomposition from the PCG signal.

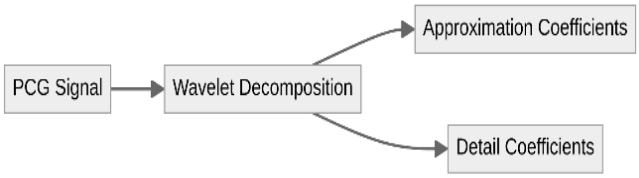

Fig. 3. The wavelet decomposition from the PCG signal.

3) Wavelet Scattering Transform (WST): The WST is an advanced technique that captures stable, hierarchical time-frequency features from PCG signals. The method involves the following steps [18][33]:

- Convolution of PCG signals with a wavelet family of varying scales.
- The application of non-linear modulus operators to the convolved signals to obtain scattering coefficients.
- The coefficients are averaged over time to produce the final scattering form, retaining translation-invariant and stable attributes.

In WST, the x(t) signal is convolved with a wavelet function, ψ1, before being further convolved with additional wavelet functions, ψ2, …, ψJ. Eq. (5) is the mathematical representation of WST. The phase enables multi-scale and multi-resolution feature extractions as illustrated in Fig. 4.

$$S_J\big(x(t)\big) = |x * \psi_1| * \psi_2 * \ldots * \psi_J \quad (5)$$

Where $S_J\big(x(t)\big)$ denotes the wavelet scattering coefficients at level J, $x(t)$ represents the input signal, and ψ_1, ψ_2, …, ψ_J are the wavelet functions at different scales.

Fig. 4. The wavelet scattering transforms on PCG signal.

C. Feature Selection

Feature selection aims to eliminate duplicate or non-informative features, which reduces the excess amount of input features, leading to achieving an optimal model. Hence, feature selection is critical in ensuring that only the most relevant information is passed to the classifier [8]. In the current study, five metaheuristic algorithms for feature selection were employed: HHO, DA, GWO, SSA, and WOA.

1) Harris Hawks Optimisation (HHO): The HHO is a swarm intelligence algorithm published by [34] in 2019. The model was inspired by nature, mimicking the cooperative hunting strategy of Harris hawks. Balancing exploration and exploitation, the algorithm simulates different prey-hunting phases, such as surprise pounce and soft besiege.

The current study employed HHO to determine the most suitable optimal features. The approach is characterised by rapid large solution space explorations with substantial accuracy and rapid convergence, maintaining exploration and exploitation equilibrium. Post-feature extraction, the initial parameters were applied to establish a random position for HHO within the search bounds.

The HHO algorithm determines the best position for prey. The exploration phase is denoted by prey energy values over 1 (E > 1), where HHO is still in the searching mode. Meanwhile, a slight energy drop below 1 (E < 1) denotes the model entering the exploitation stage. During the soft siege (E ≥ 0.5), HHO continually updates the position of the prey demonstrating energy to escape. Nevertheless, HHO dives around the target in a hard siege once the energy falls below 0.5 with a significant escape probability.

Features within the optimal range and low error rate are selected at each stage. During the final cycle, the selected PCG attributes are optimised and employed as input for the classification model. The HHO process [35] is illustrated in Fig. 5.

The technique is particularly effective for selecting features from the WST as it efficiently handles the hierarchical and multi-scale data. Feature selection was calculated according to following the equation.

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1.|X_{rand}(t) - 2r_2.X(t)| & E \geq 0.5 \\ X_{prey}(t) - X(t) - r_3.\left|X_{prey}(t) - X(t)\right| & E < 0.5 \end{cases} \quad (6)$$

Where $X(t+1)$ refers to the next position of the hawk, X_{rand} and X_{prey} represent random and prey positions, while r1, r2, and r3 are random numbers.

2) Dragonfly Algorithm (DA): Introduced in 2016 by [36], DA is a metaheuristic optimisation algorithm based on the static and dynamic swarming behaviours of dragonflies. The model utilises five key factors: separation, alignment, cohesion, attraction to food, and distraction from enemies.

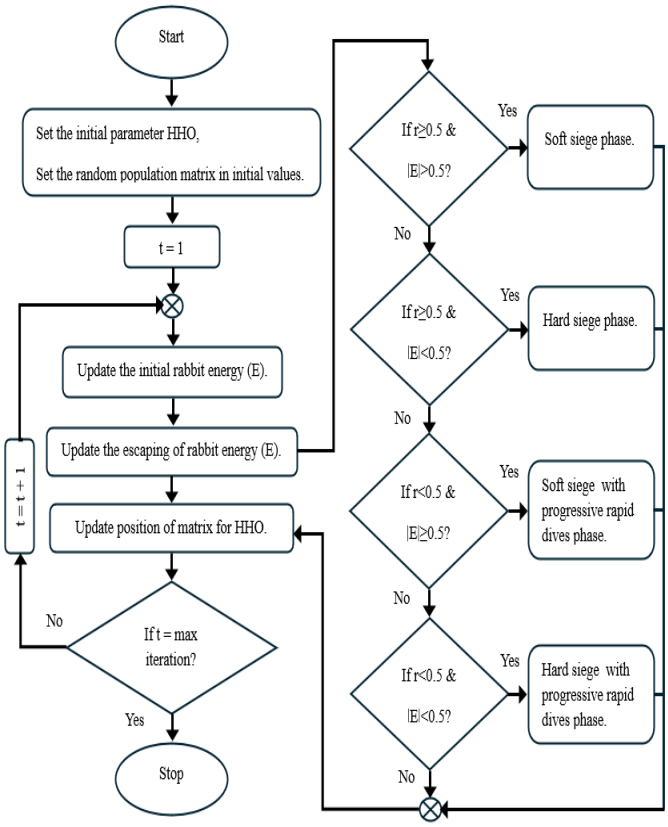

Fig. 5. The HHO working principle.

The DA defines a neighbourhood of radius r according to how it works. Neighbourhood sizes are scaled up based on a linear relation with the iteration counter during exploration to exploitation transitions. Consequently, static swarms are transformed into dynamic swarming. In the final optimisation stage, each dragonfly solution is joined to form an active unified swarm, which converges towards the best global solution at the end of the convergence [37], [38], [39]. The feature selection for the technique is represented by following the equation. The DA was also applied in this study during the feature selection stage (see Fig. 6).

$$\Delta X = \sum_{i=1}^{N}(S_i + A_i + C_i + F_i + E_i) \quad (7)$$

Where S is separation, A denotes alignment, C represents cohesion, F is food attraction, and E denotes enemy distraction. The working principle of DA.

3) Gray Wolf Optimiser (GWO): Mirjalili et al. [40] proposed GWO, a swarm intelligence-based metaheuristic algorithm, in 2014. The algorithm was based on the grey wolf leadership hierarchy and hunting strategy. Based on the grey wolves hierarchical system [35], the alpha (α), or the most dominant wolf in the pack, leads the other wolves during food hunting and finding. During the absence of the alpha wolf, the beta (β) becomes the pack leader. In the hierarchy, the power levels of the delta (δ) and omega (ω) groups are significantly less apparent than their nearest rival (see Fig. 7).

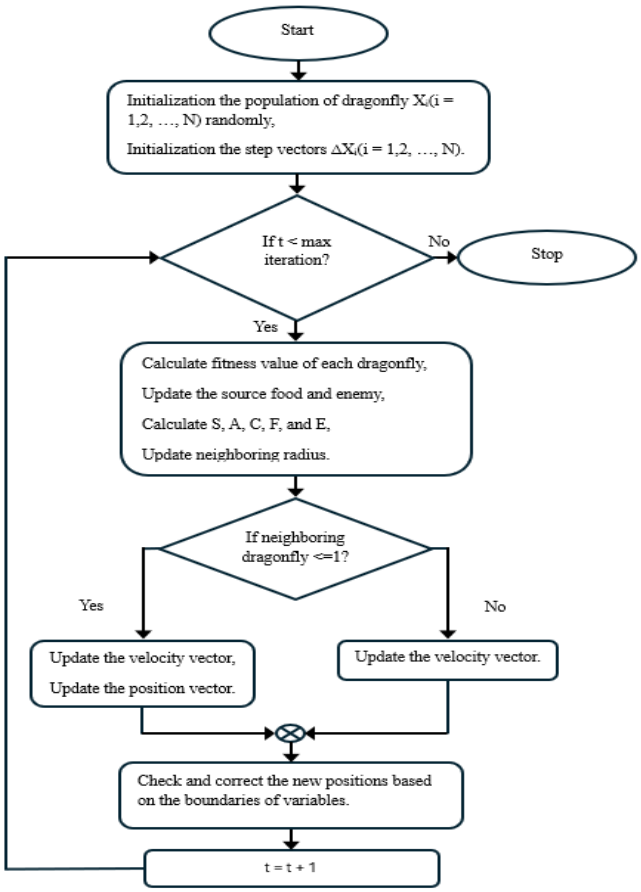

Fig. 6. The working principle of DA.

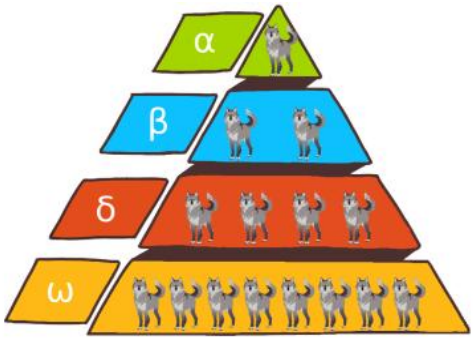

Fig. 7. The GWO hierarchical levels.

The mechanism of the GWO algorithm in the sociological agenda is complex social intelligence. Grey wolves exhibit remarkable hunting strategies by chasing, encircling, and attacking their prey [41], [42], [43]. Successful pursuits lead them to the optimal solution through successive phases with distinctive efficiencies, encouraging others to adopt similar cooperative actions [35]. Fig. 8 outlines the GWO feature selection process, while the following equation was employed to determine feature selection.

$$\vec{X}(t+1) = \frac{\vec{X}_\alpha + \vec{X}_\beta + \vec{X}_\delta}{3} \quad (8)$$

Where $X(t+1)$ refers to the next position of the wolf, and $\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta$ are the positions of the top three wolves.

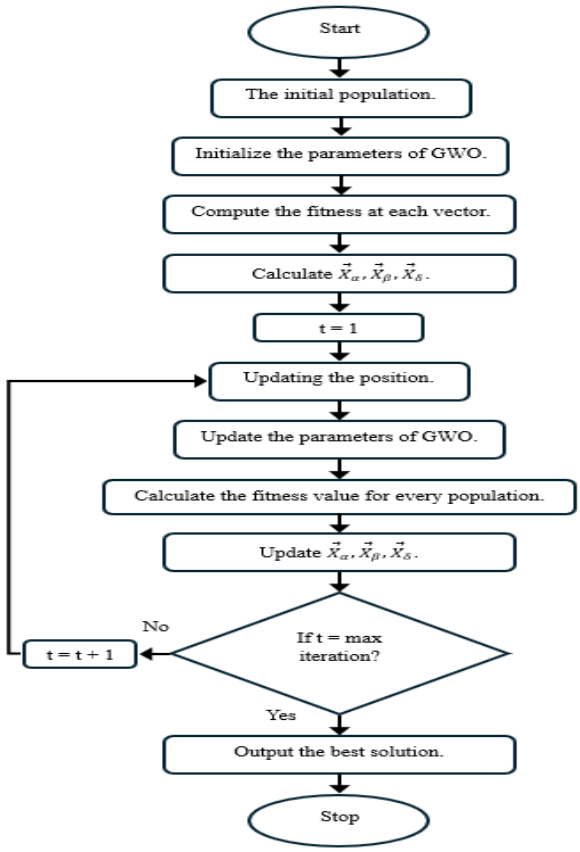

Fig. 8. The GWO working principle.

4) Salp Swarm Algorithm (SSA): In 2017, SSA, a novel nature-inspired optimiser, was proposed by Mirjalili et al. [44]. The model mimics the collective behaviour of salps, which are marine wildlife. The SSA is versatile, efficient, straightforward, and applicable to parallel and serial modes. Furthermore, the algorithm has a single adaptively decreasing parameter, contributing to optimal diversification and intensification tendencies balancing.

Salps move dynamically and update their positions through mutual interactions to avoid being trapped in local optima. The SSA behaviour is recognised by the salp chain algorithm, searching and selecting optimal food sources effectively. The swarm aims to identify and locate a specific food source within the search space.

The salps in the SSA approach are categorised as either "leaders" or "followers" based on their position in the chain. Follower salps rely on the actions of their leader for guidance. The flowchart of the SSA processes is demonstrated in Fig. 9 [45]. In this study, feature selection utilising SSA was established according to following the equation.

$$X_{i,j}(t+1) = \begin{cases} X_j(t) + c_1.\left(U_j - L_j\right) + L_j \, i = 1 \\ \frac{X_{i,j}(t) + X_{i-1,j}(t)}{2} \; i > 1 \end{cases} \quad (9)$$

Where X_j is the food source in the j dimension, U_j and L_j refer to the upper and lower bounds of the j dimension, respectively, and c_1 denotes a random number.

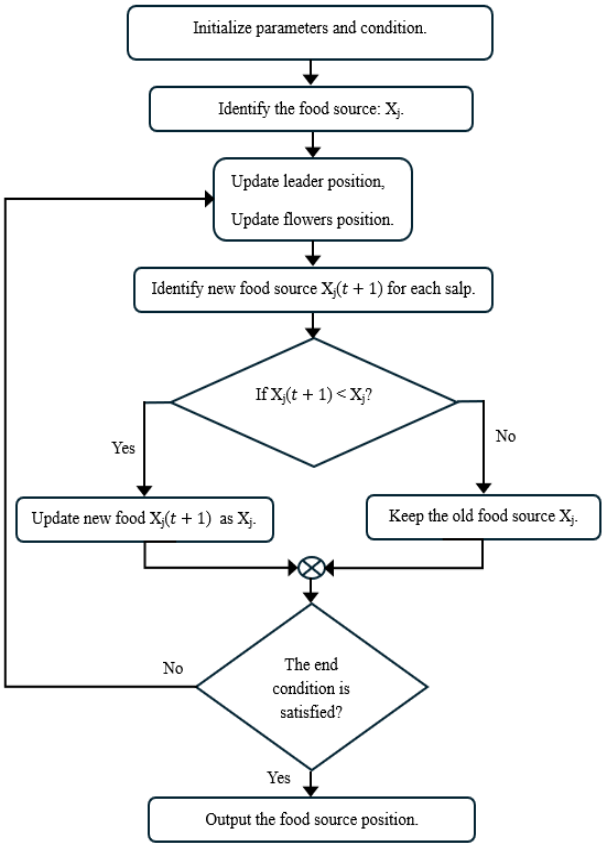

Fig. 9. The working principle of SSA.

5) Whale Optimisation Algorithm (WOA): The WOA algorithm was developed by Mirjalili and Lewis [46] in 2016. The model was the first metaheuristic approach applicable to comprehensive optimisations. The WOA algorithm involves two primary phases based on the bubble-net hunting strategy of humpback whales.

Firstly, humpback whales hunt and encircle their prey. Similarly, the WOA algorithm assumes that the best solution is unknown, thus identifying the optimal candidate solution as the target prey or close to it. Subsequently, additional search agents adjust their positions to match the best search agent, shrinking the search space effectively.

Fig. 10 depicts the second phase of the nut strategy, the bubble-net attacking stage. During the step, the whales reduce the predator chain around the prey and move in a spiral attacking pattern. Meanwhile, random humpback whales calculate new positions instead of relying solely on the globally best-known position in the update phase [47], [48]. Eq. (10) was employed to determine the feature selection in this study, and the selection process is illustrated in Fig. 11.

$$\vec{X}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - A.D & p < 0.5 \\ D'.e^{bl}.cos(2\pi l) + \overrightarrow{X^*}(t) & p \geq 0.5 \end{cases} \quad (10)$$

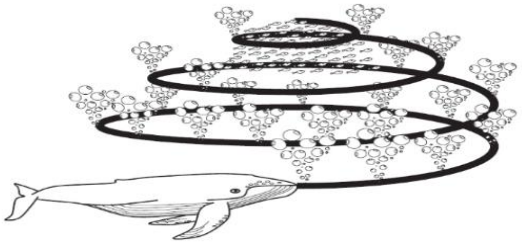

Fig. 10. The WOA behaviour.

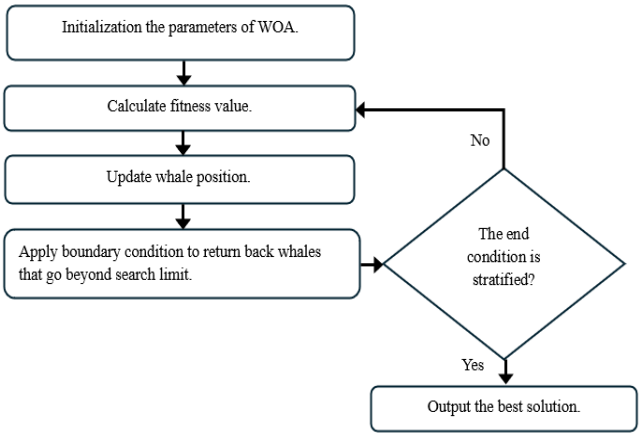

Fig. 11. The working principle of WOA.

Where X(t + 1) is the updated position of the whale, $X^*(t)$ represents the position of the best solution (prey), A and D are coefficient vectors calculated as $A = 2a \cdot r - a$ and $D = |C \cdot X^*(t) - X(t)|$, $D' = |X^*(t) - X(t)|$ denotes the distance to the prey, p is a random number between 0 and 1 that determines the exploitation or exploration phase, b is attributed to a constant defining the shape of the logarithmic spiral, and l represents a random number between −1 and 1.

D. Classification

During the classification model design, the selected features were classified with a 17-layered Recurrent Neural Network (RNN), which included Bi-LSTM. The architecture was designed to capture dependencies in sequential heart sound data (past and future), enabling the model to differentiate normal and abnormal sounds with significant accuracy.

Fig. 12. The Bi-LSTM block digram.

The model design is illustrated in Fig. 12. The description of each layer in the proposed model is presented as follows:

- The input layer receives selected features.
- The input features were converted into dense vectors in the embedding layer.
- Bi-LSTM Layer: Processes input in the forward and backward directions to capture temporal dependencies.

- The dropout layers were applied following the Bi-LSTM and dense layers to prevent overfitting.
- A total of 13 dense layers followed were included to transform the features.
- A final dense layer, the output layer, contains softmax activation for normal or abnormal heart sound categorisation.
- The forward and backward Bi-LSTM passes are presented by equations (11) and (12), respectively. This study utilised the Bi-LSTM layer model due to its bidirectional dependencies in PCG signals, essential for accurately categorising heart sounds. Its 17-layer structure also allows the model to learn complex patterns, improving classification performance. The proposed model architecture is illustrated in Fig. 13.

$$\vec{h}_t = \sigma\left(W_h.\left[\vec{h}_{t-1}, x_t\right] + b_h\right) \quad (11)$$

$$\acute{h}_t = \sigma\left(W_h.\left[\acute{h}_{t+1}, x_t\right] + b_h\right) \quad (12)$$

Where $\vec{h}_t$ and $\acute{h}_t$ represent hidden states in the forward and backward directions, respectively.

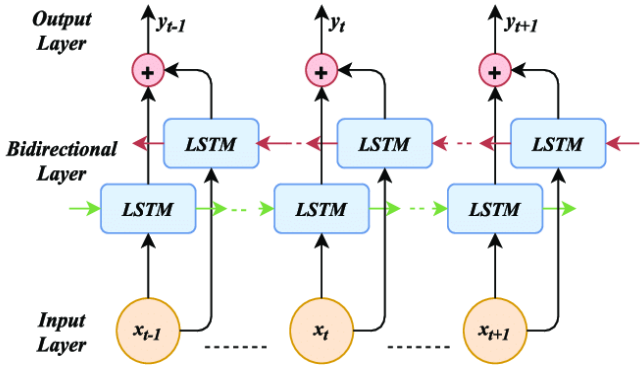

Fig. 13. The Bi-LSTM-incorporated RNN architecture.

IV. Experiments and Results

The dataset and evaluation criteria employed in this study are discussed in this section. The experiments were conducted according to the specified criteria, and the results obtained are also included.

A. Dataset

The current study explored three datasets commonly applied in related research: PhysioNet 2016 [49] and 2022 [50] and Yaseen Khan 2018 [22]. The PhysioNet 2016 dataset consists of 3,240 PCG signals from global volunteers. The duration of the heart sounds in the dataset ranged between 5 and 120 sec (s) recorded at 2,000 Hz. Furthermore, the dataset includes normal and abnormal heart sounds.

The Yaseen Khan 2018 dataset comprises 1,000 2-s PCG signals. The sounds were sampled at 8,000 Hz with binary and multi-class labels. The third dataset, PhysioNet 2022, contains 3,163 PCG signals of 5 to 65 s recorded at 4,000 Hz. In this study, each dataset was split with the k-fold and stratified k-fold (K = 5 and 10) Cross-Validation (CV) methods.

B. Performance Evaluation Metrics

The accuracy, sensitivity, specificity, and F1 score were the evaluation criteria of the proposed model. Each criterion was determined according to Eq. (13)–(16). Meanwhile, Fig. 14 demonstrates the confusion matrix-based evaluation procedure applied in the present study [50].

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

$$Sensitivity = \frac{TP}{TP+FN} \quad (14)$$

$$Specificity = \frac{TN}{TN+FP} \quad (15)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (16)$$

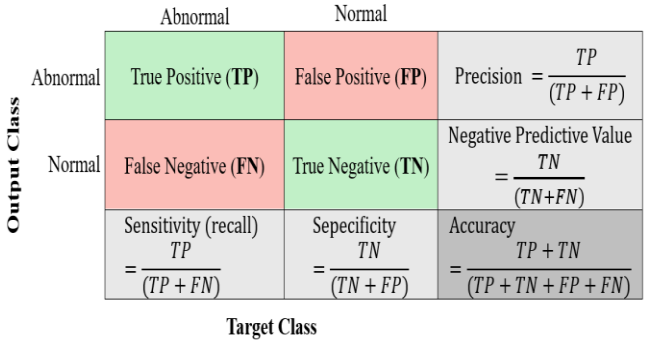

Fig. 14. Evaluation measures using confusion matrix.

C. Experimental Results

A comprehensive explanation of the results is included in this section, documenting the performance of the proposed method. The suggested model was implemented with MATLAB 2023, while training and testing were conducted on a system equipped with an Intel Core i7 processor (2.6 GHz), 32 GB RAM, and NVIDIA RTX 960M GPU (4 GB).

During the initial experiments (baseline), 3,240, 1,000, and 3,163 PCG signals from PhysioNet 2016, Yaseen Khan 2018, and PhysioNet Challenge 2022, respectively, were assessed. The datasets were split with hold-out CV, with 85% utilised as training, while 15% were employed during the testing stage.

A 5-s time window applied on the PhysioNet 2016 dataset produced 10,000 samples at 2,000 Hz frequency, while the PhysioNet Challenge 2022 generated 20,000 records at 4 kHz. Meanwhile, 8,000 samples recorded at 8 kHz were procured from the Yaseen Khan 2018 with a 1-s interval.

Several methods, including MFCC, DWT, and WST, were employed to extract raw features from each dataset at varying window sizes (5, 10, and 8 WS). The present study also utilised the Bi-LSTM algorithm with 17 layers for training, applying K-fold and stratified (S) K-fold (K = 5 and 10) CV approaches to mitigate data imbalance. Classification performance was then evaluated based on accuracy, sensitivity (Sen%), specificity (Spec%), and F1 score (F1%). Whereas, WL stands for window length, WS stands for window size, and L stands for the number of decomposition levels of DWT.

Table I summarises the time and frequency domains of the PCG signals acquired from the three datasets evaluated. The findings indicated that each dataset required separate training. The WST technique extracted numerous features from the PCG signals, achieving the highest baseline accuracy. Nonetheless, the accuracy of the model required improvement due to background noise in the PCG signals and irrelevant extracted features, which affect classification precision. Moreover, the stratified 10-fold CV recorded a notable accuracy rate when the datasets were individually assessed. Consequently, the approach was utilised in the proposed experiment, applying different window sizes for each dataset (see Table II).

TABLE I. THE BASELINE EXPERIMENTAL RESULTS FOR HEART SOUND ANALYSIS

Dataset	WL-feature extraction method		CV	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	5s-WL, 27-MFCC		5-folds	84.74	60.90	92.53	66.29
			10-folds	88.29	67.90	94.95	74.07
			S-5-folds	89	76.33	93.14	77.37
			S-10-folds	87.32	68.72	93.40	72.76
	5s-10WS-DWT	265 (L52)	5-folds	91.09	72.18	95.74	76.19
		115 (L22)	10-folds	91.41	70.62	96.53	76.48
		75 (L14)	S-5-folds	91.54	72.18	96.30	77.12
		80 (L15)	S-10-folds	91.25	76.04	95.00	77.45
	5s-10WS	**263-WST**	5-folds	92.48	75.93	96.56	79.97
			10-folds	91.76	71.14	96.84	77.34
			S-5-folds	92.53	75.10	96.82	79.88
			S-10-folds	**92.69**	74.79	97.10	80.17
Yaseen Khan 2018, 1,000 PCG signals	1s-WL, 27-MFCC		5-folds	99.66	100	97.61	99.80
			10-folds	99.66	99.61	100	99.80
			S-5-folds	99.66	99.61	100	99.80
			S-10-folds	99.66	99.61	100	99.80
	1s-8WS-DWT	30 (L5)	5-folds	100	100	100	100
		40 (L7)	10-folds	100	100	100	100
		45 (L8)	S-5-folds	100	100	100	100
		30 (L5)	S-10-folds	100	100	100	100
	1s-8WS	**261-WST**	5-folds	99.91	99.90	100	99.95
			10-folds	99.91	99.90	100	99.95
			S-5-folds	100	100	100	100
			S-10-folds	**100**	100	100	100
PhysioNet 2022, 3,163 PCG signals	5s-WL, 27-MFCC		5-folds	50.61	30.36	69.58	37.28
			10-folds	50.56	49.57	51.49	49.23
			S-5-folds	53.79	59.97	48.01	55.66
			S-10-folds	51.64	36.51	65.80	42.20
	5s-5WS-DWT	20 (L3)	5-folds	57.17	47.43	66.04	51.36
		30 (L5)	10-folds	59.49	59.11	59.83	58.18
		200 (L39)	S-5-folds	60.59	78.23	44.51	65.43
		65 (L12)	S-10-folds	61.56	65.30	58.14	61.83
	5s-5WS	**348-WST**	5-folds	57.84	58.23	57.50	56.84
			10-folds	61.43	60.44	62.33	59.91
			S-5-folds	59.87	61.15	58.70	59.23
			S-10-folds	**62.82**	65.13	60.72	62.55

TABLE II. THE WINDOW SIZE RECOMMENDATION FOR THE DATASETS

Dataset	Time and frequency domains	Feature extraction method	Window size	Classifier	CV
PhysioNet 2016	5 sec, 2,000 Hz	WST	10	Bi-LSTM algorithm (17 layers)	Stratified 10-fold
Yaseen Khan 2018	2 sec, 8,000 Hz		8		
PhysioNet 2022	5 sec, 4,000 Hz		5		

TABLE III. THE METAHEURISTIC RESULTS UNDER DEFAULT PARAMETERS

Dataset	Preprocessing	WL-feature extraction and selection method		Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	Butterworth filter (low-pass), cosine filter (low-pass), normalisation (-1, 1)	**5s-10WS-263WST**	132-HHO	92.44	77.39	96.15	80.19
			141-DA	92.55	75.41	96.76	80.00
			122-GWO	92.32	78.02	95.84	80.06
			126-SSA	92.24	75.72	96.30	79.41
			132-WOA	**92.59**	78.95	95.94	80.81
Yaseen Khan 2018, 1,000 PCG signals	Butterworth filter (low-pass), cosine filter (low-pass), normalisation (-1, 1), and zero padding	**2s-8WS-330WST**	**168-HHO**	**100**	100	100	100
			175-DA	**100**	100	100	100
			155-GWO	**100**	100	100	100
			151-SSA	**100**	100	100	100
			59-WOA	**100**	100	100	100
PhysioNet 2022, 3,163 PCG signals	Butterworth filter (low-pass), cosine filter (high-pass), normalisation (-1, 1)	**5s-5WS-348WST**	195-HHO	60.67	69.20	52.90	62.66
			193-DA	59.62	60.61	58.70	58.87
			185-GWO	57.17	56.54	57.74	55.73
			175-SSA	**61.68**	55.39	67.41	57.96
			181-WOA	60.59	57.87	63.06	58.34

During the evaluations, each dataset was assessed independently and split with holdout CV (see Table III). Low-pass Butterworth and cosine filters were applied during preprocessing to reduce high-frequency intensities from the PhysioNet 2016 and Yaseen Khan 2018 datasets at 200 Hz and 800 Hz, respectively, cut-off frequencies. Meanwhile, the PhysioNet 2022 dataset was subjected to low-pass Butterworth and high-pass cosine filters within the 15–400 Hz cut-off frequency.

After preprocessing, each dataset was normalised to a range of −1 to 1. A 5-s duration was applied to the PhysioNet 2016 and 2022 datasets, while Yaseen Khan 2018 was subjected to 2-s intervals with zero padding. Subsequently, the features were extracted with the WST. Resultantly, 263 WST features with 10-WS were procured from the PhysioNet 2016 dataset, 330 WST features with 8-WS from Yaseen Khan 2018, and 348 WST features with 5-WS from PhysioNet 2022. Metaheuristic techniques were employed during feature selection, including HHO, DA, GWO, SSA, and WOA. Default parameters and hyperparameters were applied to enhance classification accuracy.

Initially, all metaheuristic approaches were employed under default parameters of 10 maximum iterations, a 10-population size, and a 5 K-value for the KNN classifier. Holdout CV of 80% training and 20% testing, 0.99 alpha (α), and 0.01 beta (β) were also applied per the information reported in previous studies.

The 17-layer Bi-LSTM architecture developed in this study was utilised to train the classification model. A stratified 10-fold CV was also applied to overcome the class imbalance. The metaheuristic methods were then evaluated with test data with classification accuracy as the key performance metric. The results are listed in Table III.

Under default parameters, the proposed model documented significant results with 100% accuracy when applied to the Yaseen Khan 2018 dataset utilising all metaheuristic techniques. Conversely, the accuracy of the PhysioNet 2016 and 2022 datasets was not considerably higher than the baseline. Consequently, the current study employed hyperparameters for all metaheuristic methods in the second task of the proposed model as indicated in Table IV.

Preprocessing methods under hyperparameter settings and training step with the Bi-LSTM classifier (17 layers) were applied to the PhysioNet 2016 and 2022 datasets to create the classification model in this study. A stratified 10-fold CV was employed during training. Fig. 15 and Fig. 16 demonstrate the findings from the PhysioNet 2016 and 2022 datasets, respectively.

TABLE IV. THE HYPERPARAMETERS-USED FOR THE PHYSIONET 2016 AND 2022 DATASETS

Parameters	PhysioNet 2016	PhysioNet 2022
Cross-Validation	Holdout (85% training, 15% testing)	Holdout (80% training, 20% testing)
Fitness Value	β = 1e-10 - 1e-1	β = 0.01 - 0.1
Population Size	10	
Classifier	KNN	
k-value	11	5
Max iterations (M-iters)	10	

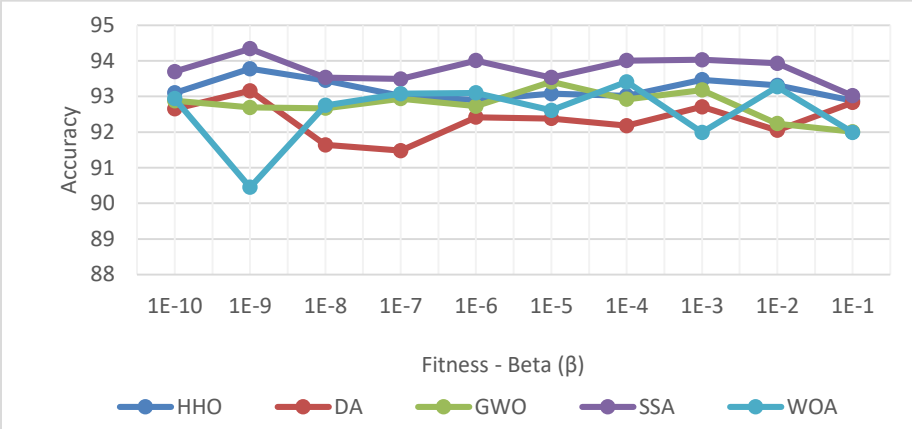

Fig. 15. The best fitness value and accuracy from the PhysioNet 2016 dataset when hyperparameter settings were applied for HHO, DA, GWO, SSA, and WOA.

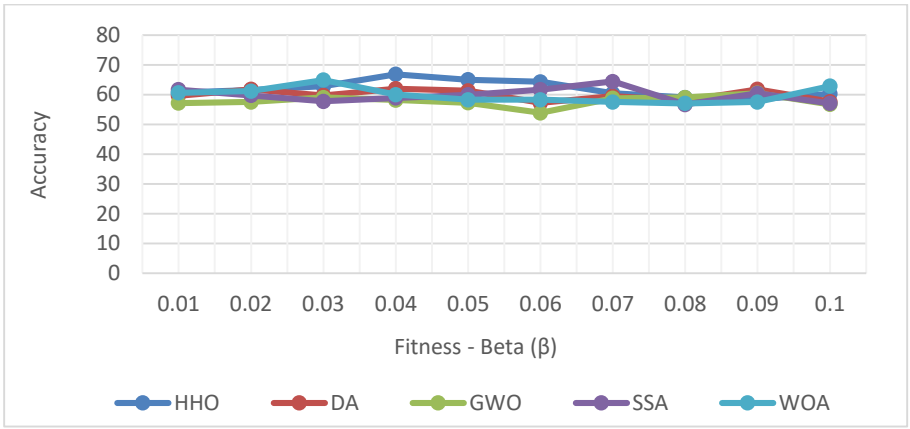

Fig. 16. The optimal fitness value and accuracy of the PhysioNet 2022 dataset were evaluated with HHO, DA, GWO, SSA, and WOA under hyperparameter settings.

Based on Fig. 15 and Fig. 16, improved classification accuracy was observed when the hyperparameters for the Physionet 2016 and 2022 datasets were adjusted. The datasets recorded excellent accuracy rates at SSA and HHO tuned to 94.34% and 66.83%, respectively. The data were considered more accurate than the baseline and default parameters. The findings are summarised in Table V. Fig. 17 to Fig. 19 and Table VI illustrate the results of real-world experiments through majority voting of the datasets employed.

According to Table VI, a 94.85% final accuracy rate was recorded by the PhysioNet 2016 dataset in real-world evaluations without voting ties (NoUniqueMode = 0). During the assessment, WST (263 extracted features, 5 sec) and SSA (127 selected features) with the Bi-LSTM (17 layers) algorithm were applied (see Fig. 17).

TABLE V. SUMMARY OF THE METAHEURISTIC RESULTS UNDER HYPERPARAMETERS

Dataset	WL-feature extraction and selection method		Optimal-fitness value (β)	k/M-iters	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	**5s-10WS-263WST**	142-HHO	1E-9	11/10	93.78	81.87	96.71	83.88
		137-DA	1E-9		93.16	81.77	95.97	82.54
		128-GWO	1E-5		93.41	81.45	96.35	83.01
		127-SSA	**1E-9**		**94.34**	82.29	97.30	85.17
		147-WOA	1E-4		93.41	79.47	96.84	82.66
PhysioNet 2022, 3,163 PCG signals	**5s-5WS-348WST**	**163-HHO**	**0.04**	5/10	**66.83**	60.70	72.41	63.57
		178-DA	0.04		62.02	57.61	66.04	59.12
		183-GWO	0.09		60.50	55.92	64.67	57.45
		173-SSA	0.07		64.43	62.74	65.96	62.71
		181-WOA	0.03		64.89	61.59	67.90	62.58

TABLE VI. SUMMARY OF THE FINAL EXPERIMENTAL RESULTS OF THE PROPOSED METHOD

Dataset	Feature extraction and selection method	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	263-WST, 127-SSA	**94.85**	83.33	97.69	86.48
Yaseen Khan 2018, 1,000 PCG signals	330-WST, All metaheuristic methods	**100**	100	100	100
PhysioNet 2022, 3,163 PCG signals	348-WST, 163-HHO	**66.87**	60.61	72.58	63.57

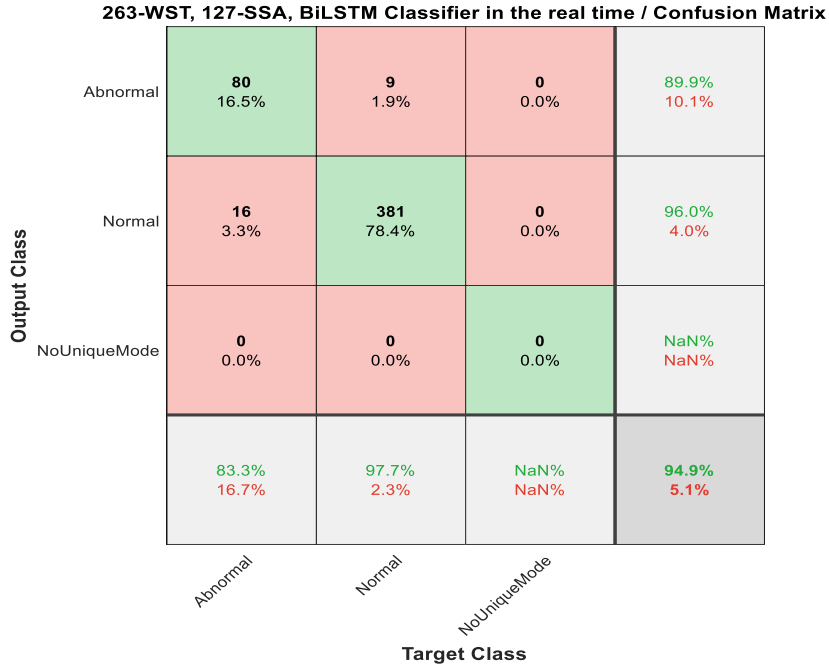

Fig. 17. The final experimental results for the PhysioNet 2016 dataset evaluated under hyperparameters by 127-SSA.

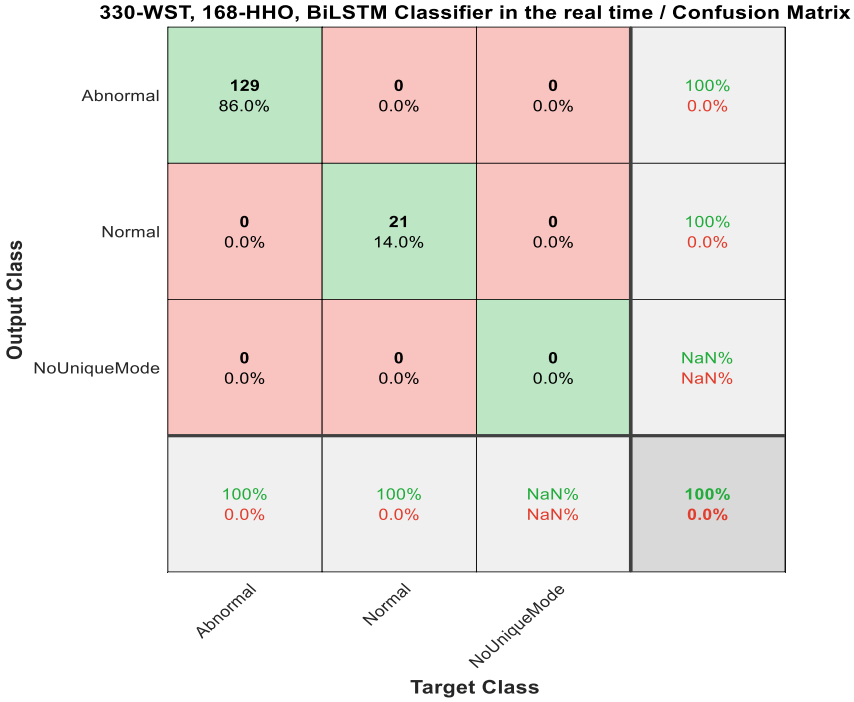

Fig. 18. The final experimental findings for the Yaseen Khan 2018 dataset with default parameters in all metaheuristic methods.

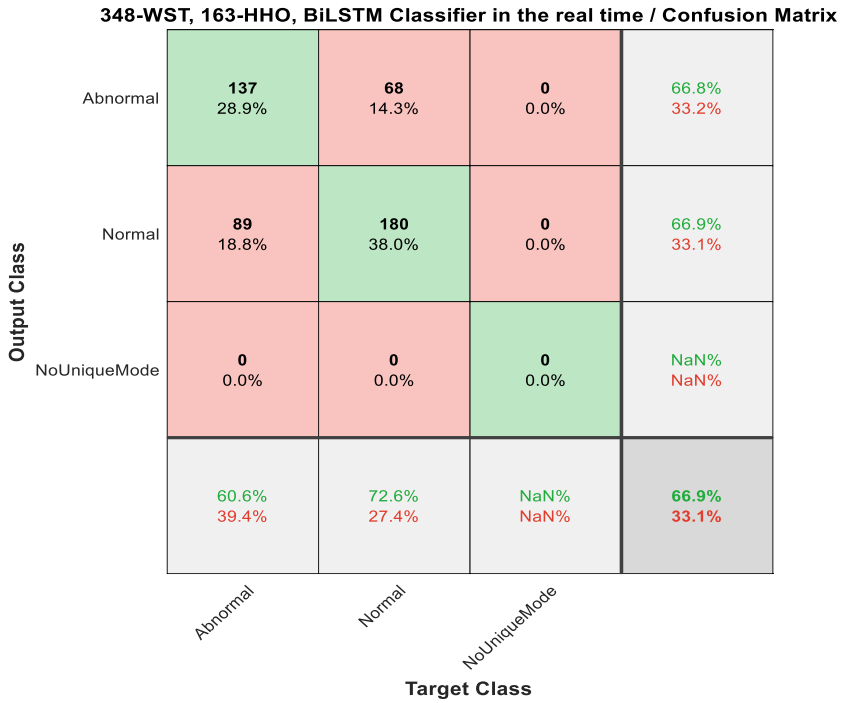

Fig. 19. The final experimental results for the PhysioNet 2022 dataset assessed with hyperparameter settings by 163-HHO.

In the Yaseen Khan 2018 real-world evaluation, the 2-s proposed model with WST and all metaheuristic methods achieved substantial results with 100% accuracy without any prediction ties. The results are summarised in Fig. 18. Meanwhile, the PhysioNet 2022 dataset achieved a final accuracy of 66.87% in the real-world assessment with the WST (348 extracted features, 5 sec) and HHO (163 selected features) methods. The data was considered more accurate than the baseline and default parameter values. The superiority was due to reduced redundant features without any ties in the voting classes, as illustrated in Fig. 19.

D. Discussion

Overall, HHO, DA, GWO, SSA, and WOA algorithms were more sensitive and negatively affected in imbalanced datasets. The metaheuristic techniques revealed similar efficiency and convergence speed limitations and global solution procurement issues. Furthermore, randomisation is crucial during the exploration and exploitation phases. Accordingly, increased randomisation would lead to classification accuracy from elevated computational time.

Based on the results, the SSA approach improved the accuracy and performance of the classification model in the PhysioNet 2016 dataset. The method exhibited superior exploration abilities for features at low frequencies than the other metaheuristic methods employed in this study.

All metaheuristic techniques utilised in the current study achieved a considerable accuracy rate when applied to the Yaseen Khan 2018 dataset. Meanwhile, the HHO approach performed better than GWO, SSA, and WOA on the PhysioNet 2022 dataset. The findings might be due to the significant exploration capabilities for features at high frequencies of HHO.

In this study, the performance of the proposed method utilising the metaheuristic techniques was compared to recently published reports that aimed to diagnose heart sounds utilising PCG signals (see Table VII). The highest accuracy rates achieved with the PhysioNet 2016 dataset were reported by [11] and [12] at 93.5% and 93.33%, respectively. Nonetheless, [11] primarily focused on extracting 527 features without applying any technique to eliminate irrelevant features. Whereas [12] procured 130 features with MFCCs and the traditional LDA technique to select optimal attributes.

For the 1,000 PCG Yaseen Khan 2018 dataset, [18] and [47] documented the highest scores, 99.80% and 99.90%, respectively. Nevertheless, [23] only employed MFCC, DWT, and CWT to extract several features. Conversely, the proposed method was based on the WST. All metaheuristic approaches also recorded a significant accuracy rate of 100% when applied to the same dataset.

Reports on applying feature selection methods to PCG signals, particularly metaheuristic approaches, are limited. For instance, [14] extracted deep features and employed PSO for improvement before utilising the Bi-LSTM algorithm, recording a 91.93% accuracy rate. Consequently, the present study aimed to close the knowledge gap, focusing on selecting optimal features based on metaheuristic approaches to improve the performance of the model.

This study achieved the best performance with 94.85%, 83.33%, 97.69%, and 86.48% accuracy, sensitivity, and specificity rates and F1 score. The WST based on the SSA method was employed. The data supported the superiority of the suggested model.

TABLE VII. RESULTS COMPARISON BETWEEN THE PROPOSED METHOD AND PREVIOUS STUDIES

Author	Dataset used	Feature extraction method	Feature selection method	Classifier	CV method	Acc (%)	Sen (%)	Spe (%)	F1 (%)
[11]	PhysioNet 2016 dataset (3,240 PCG signals)	Multi-domain features (527 features)	Not used	Fine KNN	Hold-out	93.5	-	-	-
[12]	PhysioNet 2016 dataset (3,126 PCG signals)	MFCCs (130 features)	LDA	ANN	Hold-out	93.33	-	-	-
[14]	PhysioNet 2016 dataset (3,240 PCG signals)	Deep features	PSO	Bi-LSTM	5-fold	91.93	91.58	92.27	-
[16]	PhysioNet 2016 dataset (3,153 PCG signals)	CWT and MFCC (675 features)	PCA	ANN	10-fold	85.2	-	-	-
[19]	PhysioNet 2016 dataset (3,240 PCG signals)	MFCCs (27 features)	Not used	Ensemble classifier	5-fold	92.47	94.08	91.95	-
[21]	PhysioNet 2016 (1,330 PCG signals)	Spectrogram feature	Not used	VGG16	-	88.84	87.23	86.55	87.87
[22]	Own dataset	MFCCs + DWT (43 features)	Not used	SVM	5-fold	97.9	98.2	99.4	99.7
[23]	Yaseen Khan 2018 dataset (1,000 PCG signals)	MFCC, DWT, and CWT	Not used	CNN and MLP	Hold-out	99.8	99.8	99.8	-
[24]	Yaseen Khan 2018 dataset (957 PCG signals)	Multiple features using MFCC and DWT methods	Not used	RF	-	99.89	99.90	99.60	99.90
[25]	PhysioNet 2022 dataset (3,163 PCG signals)	Spectrogram feature	Not used	Bi-GRU	5-fold	60.2	-	-	54.9
[26]	PhysioNet 2022 dataset (3,163 PCG signals)	Mel-spectrograms	Not used	U-Net	Hold-out	56.80	54.77	59.29	58.33
[28]	PhysioNet 2016 (2,435 PCG signals)	Merage short- and long-term features (33 features)	Not used	Subspace K-Nearest Neighbor	Hold-out	92.7	96	82	-
[51]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Deep features	Not used	Vision Transformer (ViT)	10-fold	99.90	99.90	99.90	99.90
[52]	PhysioNet 2016 dataset (3,153 PCG signals)	Multiple features (515 features)	Selected features randomly (400 features)	SVM	10-fold	88	88	87	-
[53]	PhysioNet 2016 dataset (3,126 PCG signals)	Zero crossing rate (ZCR), discrete fourier transform (DFT), and MFCCs (315 features)	Genetic Algorithm (GA) (15 features)	LightGBM (Light Gradient Boosting)	10-fold	92.3	91.06	93.54	-
[54]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Time-varying spectral feature (35 features)	Not used	KNN	5-fold	99.6	99.79	98.83	99.75
[55]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Chirplet Transform (CT) (300 features)	Not used	Composite Classifier	Hold-out	98.33	-	-	-
[56]	PhysioNet 2022 dataset (3,163 PCG signals)	MFCCs (25 features)	Not used	Ensemble classifier	10-fold	56.8	-	-	52.8
Propose Methods	**PhysioNet 2016 dataset (3,240 PCG signals)**	**WST (263 features)**	**SSA (127 features)**	**Bi-LSTM (17 layers)**	**Stratified 10-fold**	**94.85**	**83.33**	**97.69**	**86.48**
	Yaseen Khan 2018 (1,000 PCG signals)	**WST (330 features)**	**HHO, DA, GWO, SSA, and WOA**			**100**	**100**	**100**	**100**
	PhysioNet 2022 (3,163 PCG signals)	**WST (348 features)**	**HHO (163 features)**			**66.87**	**60.61**	**72.58**	**63.57**

The study in [25] documented an accuracy of 60.2% for the PhysioNet Challenge 2022 dataset with spectrogram features. The less accurate results were from the substantially noisy PCG signals, rendering the extraction of vital features and improvement of the classification model challenging. Nonetheless, the proposed design achieved an optimal solution by obtaining an accuracy rate of 66.87% when applied to the same dataset. Furthermore, the current study effectively determined the sensitivity, specificity, and F1 scores, yielding 60.61%, 72.58%, and 63.57%, respectively. The suggested method diminished the number of extracted features, positively affecting convergence speed during the training process and achieving excellent performance.

V. Conclusion and Future Directions

Medically, diagnosing heart sounds faces significant obstacles, particularly PCG signal processing, due to noise, imbalanced datasets, and the considerable extracted feature search space. Numerous researchers have addressed the issues and developed various solutions. Nonetheless, related reports on applying metaheuristic techniques to PCG signals are scarce. The non-linear nature of the data and its complicated relationships have contributed to the knowledge gap.

The current study proposed a model employing WST to address the challenges posed by noise signals and irrelevant extracted features. Hyper-filters were applied to mitigate the impact of noise, while metaheuristic optimisation techniques, HHO, DA, GWO, SSA, and WOA, were utilised to select the most informative WST features. The selected features then served as input to a Bi-LSTM algorithm to produce the classification model. Moreover, a stratified 10-fold cross-validation was implemented to mitigate the effects of imbalanced datasets and overfitting. Resultantly, the metaheuristic methods documented potential, exhibiting 100% accuracy with the Yaseen Khan 2018 dataset with default parameters. Meanwhile, the classification accuracy of the suggested model on the PhysioNet 2016 and 2022 datasets under hyperparameter settings was 94.85% and 66.87% with SSA and HHO, respectively.

Overall, the proposed method documented superior results to previous research. The suggested model might also improve clinical finding reliability. However, the limitations are still having in this study, such as noise in PCG signals, imbalanced datasets, and unnecessary features. Consequently, future studies should enhance the attribute by focusing on several key areas. For instance, utilising a more significant PCG signal dataset and refining the preprocessing techniques by applying deep filters to reduce noise. The imbalanced classes issue could also be mitigated by enhancing the stratified K-fold CV process. Furthermore, extracting multiple features with CWT, DWT, and WST techniques could be considered. Finally, an improved model performance might be achieved by applying hybrid metaheuristic optimisation methods, such as HHO-SSA, to select optimal features efficiently.

References

[1] J. Chen, Z. Guo, X. Xu, G. Jeon, and D. Camacho, “Artificial intelligence for heart sound classification: A review,” Expert Syst., vol. 41, no. 4, pp. 1–20, Apr. 2024, doi: 10.1111/exsy.13535.

[2] A. Javeed, S. S. Rizvi, S. Zhou, R. Riaz, S. U. Khan, and S. J. Kwon, “Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification,” Mob. Inf. Syst., vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/8843115.

[3] H. Yadav et al., “CNN and Bidirectional GRU-based heartbeat sound classification architecture for elderly people,” Mathematics, vol. 11, no. 6, pp. 1–25, Mar. 2023, doi: 10.3390/math11061365.

[4] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, “Deep learning methods for heart sounds classification: A systematic review,” Entropy, vol. 23, no. 6, pp. 1–18, May 2021, doi: 10.3390/e23060667.

[5] Z. Ren, Y. Chang, T. T. Nguyen, Y. Tan, K. Qian, and B. W. Schuller, “A comprehensive survey on heart sound analysis in the deep learning era,” arXiv Prepr. arXiv2301.09362, pp. 1–16, May 2024, [Online]. Available: https://arxiv.org/abs/2301.09362

[6] F. Li, Z. Zhang, L. Wang, and W. Liu, “Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning,” Front. Physiol., vol. 13, pp. 1–16, Dec. 2022, doi: 10.3389/fphys.2022.1084420.

[7] N. B. Aji, K. Kurnianingsih, N. Masuyama, and Y. Nojima, “CNN-LSTM for heartbeat sound classification,” JOIV Int. J. Informatics Vis., vol. 8, no. 2, pp. 735–741, May 2024, doi: 10.62527/joiv.8.2.2115.

[8] M. F. A. Ben Hamza and N. N. A. Sjarif, “A comprehensive overview of heart sound analysis using machine learning methods,” IEEE Access, vol. 12, pp. 117203–117217, Jul. 2024, doi: 10.1109/ACCESS.2024.3432309.

[9] M. U. Khan, S. Zuriat-e-Zehra Ali, A. Ishtiaq, K. Habib, T. Gul, and A. Samer, “Classification of multi-class cardiovascular disorders using ensemble classifier and impulsive domain analysis,” in 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), IEEE, Jul. 2021, pp. 1–8. doi: 10.1109/MAJICC53071.2021.9526250.

[10] R. M. Potdar, M. R. Meshram, and R. Kumar, “Optimal Parameter Selection for DWT based PCG Denoising,” Turkish J. Comput. Math. Educ., vol. 12, no. 10, pp. 7521–7532, Apr. 2021, doi: 10.17762/turcomat.v12i10.5658.

[11] O. Alshamma, F. H. Awad, L. Alzubaidi, M. A. Fadhel, Z. M. Arkah, and L. Farhan, “Employment of multi-classifier and multi-domain features for PCG recognition,” in 2019 12th International Conference on Developments in eSystems Engineering (DeSE), IEEE, Oct. 2019, pp. 321–325. doi: 10.1109/DeSE.2019.00066.

[12] M. G. M. Milani, P. E. Abas, L. C. De Silva, and N. D. Nanayakkara, “Abnormal heart sound classification using phonocardiography signals,” Smart Heal., vol. 21, pp. 1–18, Jul. 2021, doi: 10.1016/j.smhl.2021.100194.

[13] Y. He, W. Li, W. Zhang, S. Zhang, X. Pi, and H. Liu, “Research on segmentation and classification of heart sound signals based on deep learning,” Appl. Sci., vol. 11, no. 2, pp. 1–15, Jan. 2021, doi: 10.3390/app11020651.

[14] L. Zhang, C. P. Lim, Y. Yu, and M. Jiang, “Sound classification using evolving ensemble models and particle swarm optimization,” Appl. Soft Comput., vol. 116, pp. 1–28, Feb. 2022, doi: 10.1016/j.asoc.2021.108322.

[15] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, “Ensemble of feature based and deep learning based classifiers for detection of abnormal heart sounds,” in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada: IEEE, Sep. 2016, pp. 621–624. doi: 10.22489/CinC.2016.182-399.

[16] E. Kay and A. Agarwal, “Drop connected neural networks trained on time-frequency and inter-beat features for classifying heart sounds,” Physiol. Meas., vol. 38, no. 8, pp. 1–15, Mar. 2017, doi: 10.17863/CAM.12452.

[17] X. Bao, Y. Xu, and E. N. Kamavuako, “The effect of signal duration on the classification of heart sounds: A deep learning approach,” Sensors, vol. 22, no. 6, pp. 1–14, Mar. 2022, doi: 10.3390/s22062261.

[18] J. Li, L. Ke, Q. Du, X. Ding, X. Chen, and D. Wang, “Heart sound signal classification algorithm: A combination of wavelet scattering transform and twin support vector machine,” IEEE Access, vol. 7, pp. 179339–179348, Dec. 2019, doi: 10.1109/ACCESS.2019.2959081.

[19] S. A. Singh and S. Majumder, “Short unsegmented PCG classification based on ensemble classifier,” Turkish J. Electr. Eng. Comput. Sci., vol. 28, no. 2, pp. 875–889, Mar. 2020, doi: 10.3906/elk-1905-165.

[20] R. M. Potdar, M. R. Meshram, and R. Kumar, “Optimization of automatic PCG analysis and CVD diagnostic system,” Turkish J. Comput. Math. Educ., vol. 12, no. 11, pp. 3738–3751, May 2021, [Online]. Available: https://turcomat.org/index.php/turkbilmat/article/view/6456

[21] K. E. K. Blitti, F. G. Tola, P. Wangdi, D. Kumar, and A. Diwan, “Heart sounds classification using frequency features with deep learning approaches,” in 2024 IEEE Applied Sensing Conference (APSCON), IEEE, Jan. 2024, pp. 1–4. doi: 10.1109/APSCON60364.2024.10465862.

[22] Yaseen, G.-Y. Son, and S. Kwon, “Classification of heart sound signal using multiple features,” Appl. Sci., vol. 8, no. 12, pp. 1–14, Nov. 2018, doi: 10.3390/app8122344.

[23] S. I. Flores-Alonso, B. Tovar-Corona, and R. Luna-García, "Deep learning algorithm for heart valve diseases assisted diagnosis," Appl. Sci., vol. 12, no. 8, pp. 1–18, Apr. 2022, doi: 10.3390/app12083780.

[24] S. Swaminathan, S. M. Krishnamurthy, C. Gudada, S. K. Mallappa, and N. Ail, "Heart sound analysis with machine learning using audio features for detecting heart diseases," Int. J. Comput. Inf. Syst. Ind. Manag. Appl., vol. 16, no. 2, pp. 131–147, May 2024.

[25] A. McDonald, M. JF Gales, and A. Agarwal, "Detection of heart murmurs in phonocardiograms with parallel hidden semi-markov models," in Conference: 2022 Computing in Cardiology (CinC), Tampere, Finland, Sep. 2022, pp. 1–4. doi: 10.22489/CinC.2022.020.

[26] G. Singh, A. Verma, L. Gupta, A. Mehta, and V. Arora, "An automated diagnosis model for classifying cardiac abnormality utilizing deep neural networks," Multimed. Tools Appl., vol. 83, no. 13, pp. 39563–39599, Apr. 2024, doi: 10.1007/s11042-023-16930-5.

[27] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, "Classification of heart sounds using convolutional neural network," Appl. Sci., vol. 10, no. 11, pp. 1–17, Jun. 2020, doi: 10.3390/app10113956.

[28] M. Guven and F. Uysal, "A new method for heart disease detection: long short-term feature extraction from heart sound data," Sensors, vol. 23, no. 13, pp. 1–15, Jun. 2023, doi: 10.3390/s23135835.

[29] M. Wang, B. Guo, Y. Hu, Z. Zhao, C. Liu, and H. Tang, "Transfer learning models for detecting six categories of Phonocardiogram recordings," J. Cardiovasc. Dev. Dis., vol. 9, no. 3, pp. 1–17, Mar. 2022, doi: 10.3390/jcdd9030086.

[30] S. Ismail, I. Siddiqi, and U. Akram, "Localization and classification of heart beats in phonocardiography signals —a comprehensive review," EURASIP J. Adv. Signal Process., vol. 2018, no. 1, pp. 1–27, Dec. 2018, doi: 10.1186/s13634-018-0545-9.

[31] R. Khushaba, "Feature extraction using multisignal wavelet transform decom." Github, Aug. 2020. [Online]. Available: https://github.com/RamiKhushaba/getmswtfeat

[32] N. Dia, J. Fontecave-Jallon, P.-Y. Gumery, and B. Rivet, "Heart rate estimation from phonocardiogram signals using non-negative matrix factorization," in ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, May 2019, pp. 1293–1297. doi: 10.1109/ICASSP.2019.8682343.

[33] B. Soro and C. Lee, "A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization," Sensors, vol. 19, no. 8, pp. 1–12, Apr. 2019, doi: 10.3390/s19081790.

[34] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," Futur. Gener. Comput. Syst., vol. 97, pp. 1–36, Aug. 2019, doi: 10.1016/j.future.2019.02.028.

[35] M. Al-Kaabi, V. Dumbrava, and M. Eremia, "Multi criteria frameworks using new meta-heuristic optimization techniques for solving multi-objective optimal power flow problems," Energies, vol. 17, no. 9, pp. 1–37, May 2024, doi: 10.3390/en17092209.

[36] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," Neural Comput. Appl., vol. 27, no. 4, pp. 1053–1073, May 2016, doi: 10.1007/s00521-015-1920-1.

[37] M. Alshinwan et al., "Dragonfly algorithm: a comprehensive survey of its results, variants, and applications," Multimed. Tools Appl., vol. 80, no. 10, pp. 14979–15016, Apr. 2021, doi: 10.1007/s11042-020-10255-3.

[38] P. T. Prinson and A. Geetha, "Dragonfly algorithm and variants for feature selection: A review," in 2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHESS), IEEE, Sep. 2023, pp. 1–5. doi: 10.1109/iQ-CCHESS56596.2023.10391693.

[39] Y. Meraihi, A. Ramdane-Cherif, D. Acheli, and M. Mahseur, "Dragonfly algorithm: A comprehensive review and applications," Neural Comput. Appl., vol. 32, no. 21, pp. 16625–16646, Nov. 2020, doi: 10.1007/s00521-020-04866-y.

[40] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," Adv. Eng. Softw., vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[41] S. N. Makhadmeh et al., "Recent advances in grey wolf optimizer, its versions and applications: Review," IEEE Access, vol. 12, pp. 22991–23028, Feb. 2024, doi: 10.1109/ACCESS.2023.3304889.

[42] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, "Grey wolf optimizer: a review of recent variants and applications," Neural Comput. Appl., vol. 30, no. 2, pp. 413–435, Jul. 2018, doi: 10.1007/s00521-017-3272-5.

[43] Y. Liu, A. As'arry, M. K. Hassan, A. A. Hairuddin, and H. Mohamad, "Review of the grey wolf optimization algorithm: variants and applications," Neural Comput. Appl., vol. 36, no. 6, pp. 2713–2735, Feb. 2024, doi: 10.1007/s00521-023-09202-8.

[44] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems," Adv. Eng. Softw., vol. 114, pp. 1–29, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.

[45] L. Abualigah, M. Shehab, M. Alshinwan, and H. Alabool, "Salp swarm algorithm: a comprehensive survey," Neural Comput. Appl., vol. 32, no. 15, pp. 11195–11215, Aug. 2020, doi: 10.1007/s00521-019-04629-4.

[46] S. Mirjalili and A. Lewis, "The whale optimization algorithm," Adv. Eng. Softw., vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.

[47] Ş. Ay, E. Ekinci, and Z. Garip, "A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases," J. Supercomput., vol. 79, no. 11, pp. 11797–11826, Jul. 2023, doi: 10.1007/s11227-023-05132-3.

[48] D. Guha, P. K. Roy, and S. Banerjee, "Whale optimization algorithm applied to load frequency control of a mixed power system considering nonlinearities and PLL dynamics," Energy Syst., vol. 11, no. 3, pp. 699–728, Aug. 2020, doi: 10.1007/s12667-019-00326-2.

[49] C. Liu et al., "An open access database for the evaluation of heart sound algorithms," Physiol. Meas., vol. 37, no. 12, pp. 2181–2213, Nov. 2016, doi: 10.1088/0967-3334/37/12/2181.

[50] M. A. Reyna et al., "Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022," PLOS Digit. Heal., vol. 2, no. 9, pp. 1–22, Sep. 2023, doi: 10.1371/journal.pdig.0000324.

[51] S. Jamil and A. M. Roy, "An efficient and robust phonocardiography (PCG)-based valvular heart diseases (VHD) detection framework using vision transformer (ViT)," Comput. Biol. Med., vol. 158, pp. 1–15, May 2023, doi: 10.1016/j.compbiomed.2023.106734.

[52] H. Tang, Z. Dai, Y. Jiang, T. Li, and C. Liu, "PCG classification using multidomain features and SVM classifier," Biomed Res. Int., vol. 2018, no. 2, pp. 1–14, Jul. 2018, doi: 10.1155/2018/4205027.

[53] N. K. Sawant, S. Patidar, N. Nesaragi, and U. R. Acharya, "Automated detection of abnormal heart sound signals using Fano-factor constrained tunable quality wavelet transform," Biocybern. Biomed. Eng., vol. 41, no. 1, pp. 1–44, Jan. 2021, doi: 10.1016/j.bbe.2020.12.007.

[54] P. Upretee and M. E. Yuksel, "Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results," in Conference: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), IEEE, Apr. 2019, pp. 1–4. doi: 10.1109/EBBT.2019.8741730.

[55] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, "Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals," Comput. Biol. Med., vol. 118, pp. 1–17, Mar. 2020, doi: 10.1016/j.compbiomed.2020.103632.

[56] Z. Imran, E. Grooby, C. Sitaula, V. Malgi, S. Aryal, and F. Marzbanrad, "A fusion of handcrafted feature-based and deep learning classifiers for heart murmur detection," in Conference: 2022 Computing in Cardiology (CinC), Tampere, Finland: IEEE, Dec. 2022, pp. 1–4. doi: 10.22489/CinC.2022.310.