2025 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2025

2025 International Conference on Emerging Smart Computing and Informatics (ESCI) | 979-8-3315-1568-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ESCI63694.2025.10987967

Butterfly Optimization and Deep Learning to Classify Heart Sound Signal

Faika Fairuj Preotee, Md Sabbir Hossain, Shuvashis Sarker, Farliha Binte Faisal,
Nawrin Tabassum, Shamim Akhter*
Department of Computer Science and Engineering,
Ahsanullah University of Science and Technology, Dhaka, Bangladesh
{faikafairuj2001, mdsabbir120834, shuvashisofficial,
farliha1722, nawrintabassum14, *shamimakhter}@gmail.com

***Abstract*—Heart sound signals, or PCG recordings, provide a valuable non-invasive alternative to traditional diagnostic methods like CXR and blood tests for detecting cardiovascular disease. In this study, PCG data are collected, and features such as MFCC and WST are extracted to capture both time and frequency information. These features are used to classify the signals into normal and abnormal categories using DL models, including CNN, Bi-LSTM, and Bi-RNN. The study conducts extensive experiments using individual and combined features, showing that combining features enhances classification accuracy. The proposed CNN model, optimized with the Butterfly Optimization Algorithm, achieves an impressive classification accuracy of 99.07%, outperforming other models. This result highlights the effectiveness of using PCG signals and advanced DL techniques for accurate cardiovascular disease detection, offering a non-invasive and efficient alternative for early diagnosis.**

***Index Terms*—Phonocardiogram Record, Wavelet Scattering Transform (WST), Mel-Frequency Cepstral Coefficient (MFCC), Butterfly Algorithm, Deep Learning, Spectrogram.**

I. INTRODUCTION

According to the WHO, in 2016, cardiovascular disease caused nearly seventeen million casualties, accounting for thirty-one percent of all worldwide deaths [1]. A tool called Phonocardiogram(PCG) records these heart sounds visually. It more effectively illustrates changes in the heart's condition than an ECG, which evaluates the electrical activity of the heart. There are several techniques for analyzing heart sounds. Focusing primarily on methods that extract and categorize these sounds. These methods include analyzing the timing of sounds, their wave patterns, and their frequency characteristics [2], as well as using advanced mathematical ML and DL models. These techniques are essential for diagnosing cardiac disorders in a clinical environment by determining the normality or abnormality of heart sounds. This process is essential because it helps doctors to understand the heart's health and guide treatment decisions effectively. The Wavelet Scattering Transform(WST) techniques were used in [3]–[8] for its proficiency in feature extraction from heart sound signals. These methods elucidate the analysis by highlighting and subsequently attenuating the high-frequency components of the signal, ensuring the retention of critical information post filtering [6]. The study [6] employed Scatter Transform methods and TWSVM for cardiac signal categorization, achieving an accuracy of 97%. MFCC and Mel-spectrogram are time-frequency characteristics utilized for cardiovascular sound classification, achieving accuracy rates between 70.57% and 86.02% [9]–[13]. Mostly ML models including SVM, TWSVM, KNN, ANN, XGboost, Decision Tree etc [6], [9], [11], [14] were used as classification models and achieved 90%-95% accuracy. Very few researchers used CNN and LSTM-based deep learning models as heart sound classifiers [2], [3], [10], [12], [15]. The CNN accuracy with MFCC features varied between 70.57%-86.02% [10]. The motivation of our work is to improve the CNN model accuracy with combined features and optimal hyperparameters. Consequently, the ensuing impacts of the present investigation are listed below:

i. Assess performance of Deep learning models with individual and combined features from MFCC, WST, and Mel-Spectrogram.
ii. Implement Butterfly Optimization Algorithm(BOA) to optimize hyperparameter setting for Convolutional Neural Network(CNN) and improve its overall performance.

The remainder of this paper is structured as follows: Section II provides an overview of the *Related Work*. Section III discusses the *Dataset* and the preprocessing techniques used. Section IV outlines the key aspects of the *Proposed Methodology*. Section V presents the *Results Analysis*, evaluating the outcomes. Finally, Section VI summarizes the *Conclusion and Future Work*.

II. RELATED WORK

Heart sounds can be assessed by several methodologies. In heart sound classification, the integration of wavelet analysis with Deep Neural Networks (DNN) has demonstrated significant enhancements in classification accuracy. Li et al. [6] introduced an innovative approach for classifying heart sound signals through the integration of WST and TWSVM. The proposed approach achieved an assessment performance of 98% on the PhysioNetCinC Challenge 2016 database. In a recent study published in 2021, Mei et al. [7] researched the process of classifying heart sounds based on quality rating and WST. Achieving a measurement of accuracy of 93.64% using SVM, this method didn't require any human experience-related factors for feature selection or signal recognition. Han et al. [5] also published a Multi-path Heart Sound Detection method in 2024, built on WST and an Attention Mechanism that used WST-LSTM, SAN, and TAN and achieved a mean accuracy of 97.12%. In addition, Ghezaiel et al. [8] combined WST with CNN for speaker identification. Their system effectively extracted features from limited short training samples, achieving an 88.04% accuracy on the LibriSpeech dataset. Lee et al. [3] also integrated WST and CWT with one-dimensional & two-dimensional CNN models for cardiovascular sound classification, resulting in enhanced classification performance compared to the most recent research. All the above studies mentioned used WST to extract features. However, all of them proposed a single model with mean accuracy, without any comparative analysis.

Continuing the exploration of advanced methods, Bozkurt et al. [10] examined the use of time-frequency features like MFCC and Mel-Spectrogram for CNN-based automated cardiac sound categorization. Analyzing the performance, the mean accuracy values were between 70.57% to 86.02%. Building on the success of CNNs, in a 2019 study, Abduh et al. [9] employed classical classifiers including SVM, KNN, and ensemble classifiers for heart sound classification in addition to Fractional Fourier transform-based MFCC. With a performance score of 92%, the SVM classifier using a cubic kernel demonstrated the best performance. A recent study, Padhy et al. [12] employed MFCC and LSTM for the classification of cardiac sounds. The MFCC-LSTM model achieved a 98% sensitivity, surpassing the state-of-the-art techniques. Researchers utilized the MFCC approach to improve ML and DL models. But there are still challenges, such as the requirement for large amounts of labeled data, a lot of computing power, and precise recording of intricate cardiac sounds.

To investigate the spectrogram area, Wibawa et al. [15] in their 2018 study on abnormal heart rhythm detection using a CNN applied to spectrogram data and achieved a classification accuracy of 82.75%. The study had limitations despite it's success as it didn't explore alternative CNN configurations or preprocessing techniques for spectrogram creation. Following this advancement, using a radial basis kernel, Tang et al. [16] was able to get sensitivity, specificity, and overall scores of up to 88% when classifying PCG data using multi-domain features and an SVM classifier in 2018. This method achieved a 95% classification accuracy when evaluated on the "PhysioNet/CinC Challenge 2016 dataset". Also in 2020, using the same dataset, Li et al. [2] demonstrated how CNNs can be used as both feature selectors and classifiers in separating typical and atypical cardiac auscultations. The application earned an average score of 86.8%. Lastly, Abbas et al. [11] introduced an AI framework for classifying heart disease from audio signals using Random Forest, MLP, XGBoost, K-Nearest Neighbors, and Decision Tree. The MLP model using MFCC features obtained the best results in tests, with a 95.6% success rate.

The above researches used machine learning models for classification purposes mostly and didn't explore more to gain better results. Also, CNN hyperparameter adjustment was missing. Thus, hyperparameter adjustment may increase CNN accuracy, especially with complex heart sound data. After investigating related research, our proposed approach combines multi-domain feature extraction, including WST and MFCC from PCG data, and classifies them using Deep Learning Models. This hybrid approach aims to enhance the categorization of cardiac signals across diverse & challenging conditions. The importance of hyperparameter optimization cannot be overstated, and the use of an advanced algorithm like the Butterfly Optimization Algorithm (BOA) can significantly enhance the model's performance by automating the tuning to enhance the resilience and precision of our system.

III. Dataset

A. Dataset Overview

The dataset which is utilized in this investigation originates from the "PhysioNet/CinC Challenge 2016" [17], [18] which includes 3,240 phonocardiograms (PCG) records that have been organized into five different databases which are labeled A through E. The dataset includes 665 abnormal recordings and 2,575 normal ones. This compilation of recordings encompasses a diverse array of individuals, including physically active persons and those with cardiac conditions such as "mitral valve prolapse, aortic stenosis, and coronary artery disease." The PCG data are all sampled at a rate of "2,000 Hz" and duration are from "5 seconds to 120 seconds". Each recording, which is saved in WAV format, comes from one of nine(9) possible auscultation sites on the body. These include popular spots like the aortic and pulmonic areas. The dataset has records with different levels of background noise, like people talking, breathing, and the stethoscope moving. This makes it very hard for the algorithms to analyze and classify them. Figures 1 and 2 depict the traditional and atypical cardiac recordings respectively.

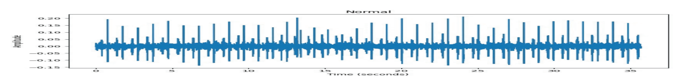


Fig. 1. Standard Cardiac Auscultation Signal

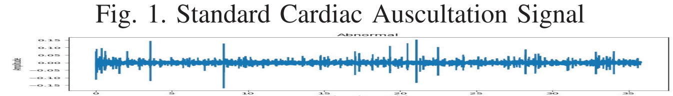


Fig. 2. Irregular Cardiac Auscultation Signal

B. Dataset Preprocessing

This present research employs a three-stage preprocessing technique for the PCG signals. The initial stage entails the resampling of all signals to a consistent rate of $f_s = 2000$. The shorter signals are padded with zeros and the longer signals are trimmed in order to ensure a standardized and consistent duration of $N = 30,000$ samples.

The second stage involves the high-pass filtering of the PCG signals. The low-frequency noise is reduced by applying high-pass filter with a 25 Hz cutoff after the samples have been trimmed to 30,000. It says that the filter is: $H(f) = \frac{f^2}{f^2+f_c^2}$. This improves the clarity of the signal, enabling more precise analysis, by reducing interference from unwanted low-frequency noises.

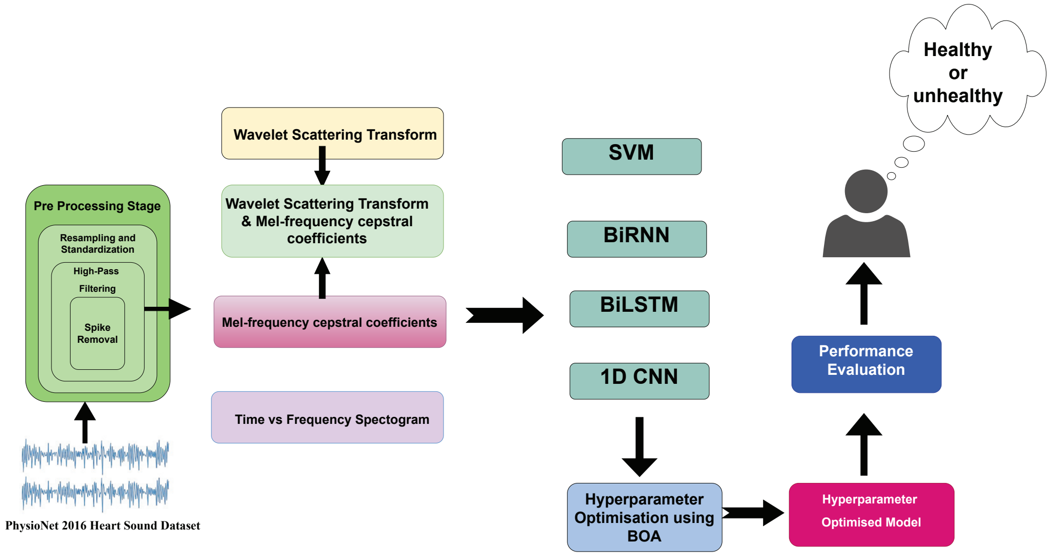


Fig. 3. Model Architecture for our proposed method.

The final stage implements an adaptive spike removal method to remove anomalies that are the result of short-term recording issues. Spikes are identified by comparing them to a median-filtered baseline, and any significant deviations, defined by the standard deviation, are eliminated to assure signal stability.

IV. Proposed Methodology

A. Features Extracted in Multi-domain

*1) **WST**:* Wavelet Transform analyzes signal details but may overlook variations over time. The scattering transform enhances this by applying a process that includes wavelet transforms, absolute values, and smoothing [19]. Figure4 shows the wavelet scattering transform process [8].

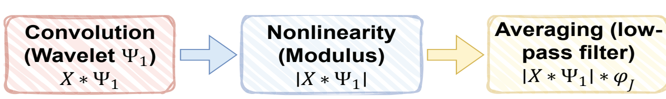


Fig. 4. An average low-pass filter ϕ, a wavelet function ψ, and input data x are used in wavelet scattering to transform processes [8].

Retrieving the processed heart sound signals from the heart sound database, we initialize the WST with a scale parameter $J = 6$, a quality factor $Q = 1$, and adapt it to input signals of length $T = 10,000$ samples. To start a wavelet scattering transform with a maximum scale factor of 2^J, the input signal is mixed with a low-pass filter, which is also called an average function. Each phonocardiogram is normalised, resampled at 2 kHz, and standardized to the same duration. The initial-order scattered indices are computed using the WST, which identifies significant frequency fluctuations in cardiac sounds. These co-efficients, which are vital for the identification of subtle sound pattern variations induced by disease, are then formatted for use in DL and ML models to help in the identification of abnormal heart conditions. Figure5 shows the first-order scattering coefficients transform.

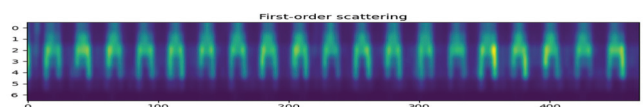


Fig. 5. First-order scattering Transform.

*2) **MFCCs**:* MFCC is a widely employed technique for obtaining features in language and signal treatment.The mathematical modeling of MFCC involves several stages: "signal framing, power spectrum calculation, application of a Mel filter bank, logarithmic translation of the filter banksand execution of the DCT". Figure6 is showing this MFCC framework [20]. Phonocardiogram recordings are processed by

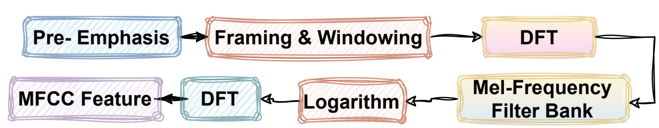


Fig. 6. MFCC Framework [20].

extracting the first seven(7) MFCCs, commonly used in audio signal processing to capture timbral features.The recordings are adjusted to a uniform duration of 30,000 samples (15 seconds) and normalised to a 2 kHz sampling rate. MFCCs are selected for their capacity to show sound power on a mel scale. The features are prepared for DL and ML models to identify heart sounds through post-processing, which includes extracting absolute values and reshaping.

*3) **Spectrogram feature**:* Dynamics signals are like output signals from a communication device that change in intensity over time. There are two main problems with dynamic signals:analyzing them (specially spectral analysis) and handling the signal. Putting the signal's strength on the vertical axis and its time on the horizontal axis can make a waveform [13]. The STFT spectrogram has prominent squared coefficients. The process of normalizing this spectrogram satisfies Parseval's energy-conservation criterion [21]. In this processing stage, audio files are first loaded ensuring that all audios are mono and the same as the original audio. Waveform amplitudes are adjusted to a range of [-1, 1] and then processed using the STFT with an average frame duration of 80 frames and a step size of 40 samples. The spectrogram is produced by the magnitude of the Fourier coefficients and it is later improved for input into convolutional neural networks. This preprocessing guarantees that machine learning tasks are performed with uniform input and effective feature extraction. Figure7 gives us a glimpse of the spectrogram feature.

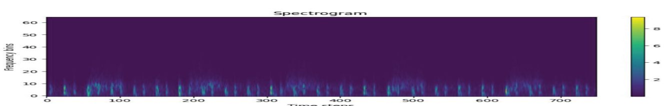


Fig. 7. Spectrogram Feature.

*4) **Combined feature of MFCC and WST**:* MFCCs and WST features are combined into one set of features. This should be improved how well our models could classify things. By putting these features together, we have used both spectral analysis and invariant feature representation, which are different but work well together. We have organized these combined features into a three-dimensional shape to match our neural networks which is needed for training and testing. This has made it easier and faster to train and test our models.

Following feature extraction, cardiac resonances are categorized into two distinct sets: a learning set and a monitoring set. Seventy percent of cardiac impulses are allotted for learning, while the rest thirty percent are designated for monitoring.

B. Model Application

Figure 3 shows a summary of the suggested model in this study. We have pre-processed the signals first and then pulled out the features in a different domain. After that, for a machine learning model, using a SVM on 3D data brings both challenges and opportunities compared to regular 2D data. We have used a linear kernel and tested the performance of our model using 42 different random states. We have used a type of classifier called SVC, which tries to find the best possible way to separate different classes in the data. For DL models, we implement CNN, Bi-LSTM and Bi-RNN for performance analysis. A CNN is specifically designed to process two-dimensional data. The network architecture commences with a convolutional layer including 16 filters utilizing 2x2 kernels and ReLU activation, succeeded by an additional convolutional layer with 32 filters to extract more intricate features. Subsequent to convergence, the data is streamlined and transmitted to a compact layer comprising 128 neurons with activation of ReLU, ending in a sigmoid activated layer of output for binary sorting. The model is compiled using the Adam optimizer with a learning rate of 1×10^{-4} and employs binary cross-entropy for loss computation. Likewise, Bi-LSTM and Bi-RNN are enhanced with the application of the Adam optimizer, utilizing a learning rate of 0.0001. The models are trained 100 times with a batch size of 16, utilizing the binary cross-entropy loss function. The Bi-LSTM and Bi-RNN models each have a solitary dense layer comprising 64 neurons, triggered by the ReLU function.

C. Hyperparameter Optimization(BOA)

The hyperparameter optimization of a convolutional neural network is carried out in this study using the Butterfly Optimization Algorithm (BOA) to classify phonocardiogram signals. Butterflies mostly share personal information through their senses. They use their antennas to pick up fragrances in the environment which helps them find food and mates. The BOA works similarly. It uses an objective function to guide the search for the best solutions and rate the quality of those solutions.

Fragrance Calculation:

$$f = cI^a \tag{1}$$

where c is the sensory modality, I is the stimulus intensity, a is the power exponent dependent on modality [22]

Movement Calculation: In the global search phase, the butterfly updates its position according to:

$$x_i^{(t+1)} = x_i^{(t)} + r^2 \times (g^* - x_i^{(t)}) \times f_i \tag{2}$$

where $x_i^{(t)}$ is the i-th butterfly's position at iteration t, g^* is the best solution, f_i is the fragrance, and r is a random number in [0, 1] [22].

BOA, which is derived from swarm intelligence, enhances the efficacy of models by improving learning rates, kernel sizes, and layer numbers. The classification accuracy was considerably enhanced by the tuned hyperparameters, which effectively distinguished distinction between physiological and pathological cardiac sounds in the sample. With such optimization, the optimal model configuration includes 32 and 16 filters in the convolutional layers, a learning rate of 0.0001, strides of [2,2], 100 epochs of training, and a batch size of 16 and 128 units in the dense layer.

This set of hyperparameters has been associated with appreciable improvement in the model's accuracy and further confirmed the potential of nature-inspired algorithms in boosting machine learning models for clinical diagnostic applications.

V. Result Analysis

The present investigation evaluates various DL models employing various feature harvesting techniques for cardiovascular sound detection, comparing their performance across several metrics.

The Table I summarizes the performance of the models using WST, MFCC and Spectrogram features individually. For

TABLE I: Performance evaluation of different feature types.

Feature Type	Model	Accuracy	Precision	Recall	F-1 Score
Wavelet Scattering Coefficient	**CNN**	**0.95**	**0.94**	**0.97**	**0.95**
	BiLSTM	0.93	0.93	0.89	0.91
	BiRNN	0.94	0.92	0.98	0.95
	SVM	0.85	0.87	0.96	0.91
MFCC Spectrum	**CNN**	**0.96**	**1.00**	**0.92**	**0.96**
	BiLSTM	0.95	0.95	0.94	0.94
	BiRNN	0.95	0.99	0.91	0.95
	SVM	0.89	0.90	0.96	0.93
Spectrogram	**CNN**	**0.98**	**0.97**	**1.00**	**0.98**
	BiLSTM	0.94	0.95	0.92	0.94
	BiRNN	0.94	0.92	0.95	0.93
	SVM	0.82	0.88	0.88	0.88

TABLE II: Performance evaluation of the proposed method.

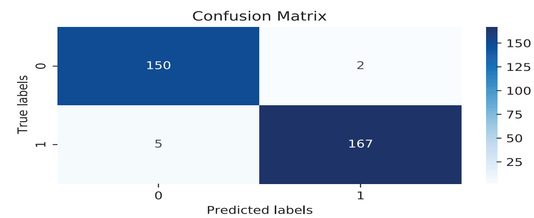
Feature Type	Model	Accuracy	Precision	Recall	F-1 Score
Combined Feature (WST and MFCC)	**CNN (With BOA)**	**0.99**	**0.99**	**0.99**	**0.99**
	CNN	0.98	0.97	0.98	0.98
	BiLSTM	0.98	0.97	0.99	0.98
	BiRNN	0.96	0.96	0.95	0.95
	SVM	0.91	0.92	0.97	0.95

WST features, the highest accuracy is achieved by the CNN which is 94.75%,While the performance of the Bi-LSTM and Bi-RNN models is comparable, that of the SVM model is weaker. When using MFCC Spectrum features, the CNN again reaches the highest accuracy of 95.68%, with Bi-LSTM and Bi-RNN models also performing well while the SVM shows lower performance. For Spectrogram features, the highest accuracy is 98.14% which is achieved by the CNN model, with Bi-LSTM and Bi-RNN models performing comparably well and the SVM showing the weakest performance. Thus in overall cnn is the best and svm is the worst performer.

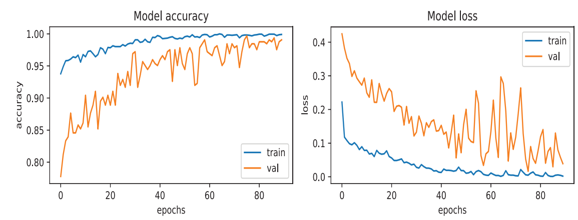
The results in Table II highlight the significant benefits of integrating MFCC and WST characteristics for heart sound classification across various models. The CNN model, optimized with BOA, achieved an accuracy of 99%, up from 98% without BOA, demonstrating the effectiveness of the optimized hyperparameters.

TABLE III: Comparison of Our Proposed Approach with Existing Methods

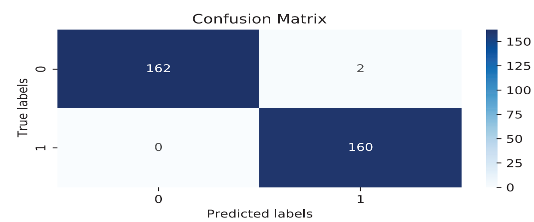
Study	Method	Features	Accuracy
Hong Tang [16]	SVM classifier with RBF kernel	Multidomain	88%
Jonathan Rubin [23]	CNN + MFCC heat maps	MFCC	83.99%
Cristhian Potes [24]	An ensemble model with the AdaBoost-abstain classifier and CNN	Time-Domain, Frequency-Domain	86.02%
Siddique Latif [25]	Bi-LSTM	MFCC	97.63%
Our Proposed Method	**CNN(Using BOA)**	**Combination of WST+ MFCC**	**99.07%**



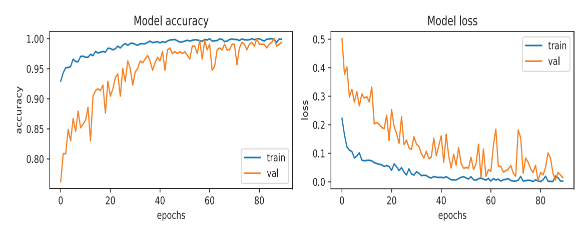
(a) Confusion Matrix of CNN without BOA.



(b) Accuracy vs loss curve of CNN without BOA.



(c) Confusion Matrix of CNN with BOA.



(d) Accuracy vs loss curve of CNN with BOA.

Fig. 8. Comparison of CNN performance with and without BOA

After a closer look, the Bi-LSTM model shows good performance in correctly identifying true positives, remaining robust even with slightly different measurements, and boasting excellent recall. However, the performance of the Bi-RNN decreases significantly at a level of 96%, suggesting challenges in managing temporal complexities. Despite its lower effectiveness compared to other models, the SVM nevertheless produces commendable outcomes, highlighting the potential for improvement through combining features. All of these results confirm that combining different feature types increases the accuracy and consistency of models and increases their usefulness in clinical settings. Therefore, BOA is utilized exclusively for combined features rather than for individual ones. Table III shows that our proposed model, which obtains competitive accuracy in comparison to existing methods, optimizes parameters using the Butterfly Optimization Algorithm (BOA) within a CNN architecture and combines WST and MFCC features. Many of these methods do not incorporate this combination of features or utilize BOA. Figure8 display the confusion matrices and accuracy vs. loss curves for a CNN model before and after implementing hyperparameter tuning using the Butterfly Optimization Algorithm (BOA).

Since the BOA tuning, there have been fewer false negatives in the confusion matrices, and the general accuracy has gone up. The accuracy and loss curves are more stable now, with smoother training and validation trends. This means that the model can generalize better and is less likely to overfit.

VI. Conclusion & Future Work

Our study has utilized the extensive PhysioNet database and has shown that deep learning techniques with the combined features including WST and MFCC significantly improves the accuracy of heart sound classification. This combination helps the creation of non-invasive diagnostic tools, which are very important for finding heart conditions quickly and treating them effectively, thereby enhancing the overall quality of diagnostic solutions. In future research, improvements such as incorporating transfer learning and transformer models can enhance heart sound analysis, making it more applicable across diverse datasets and real-world scenarios. Employing Explainable AI methods would also increase transparency, revealing key diagnostic features and building trust among medical professionals. Additionally, simplifying the user interface could further facilitate adoption in clinical settings.

References

[1] W. H. Organization, "Cardiovascular diseases (cvds)." https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds), 2024. Accessed: 2024-05-14.

[2] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, "Classification of heart sounds using convolutional neural network," *Applied Sciences*, vol. 10, no. 11, p. 3956, 2020.

[3] J.-A. Lee and K.-C. Kwak, "Heart sound classification using wavelet analysis approaches and ensemble of deep learning models," *Applied Sciences*, vol. 13, no. 21, p. 11942, 2023.

[4] Y. Chen, S. Wei, and Y. Zhang, "Classification of heart sounds based on the combination of the modified frequency wavelet transform and convolutional neural network," *Medical & Biological Engineering & Computing*, vol. 58, pp. 2039–2047, 2020.

[5] B. Han, M. Chen, Y. Li, L. Liu, S. Wei, and X. Jiang, "Multi-path heart sound detection based on wavelet scattering and attention mechanism," in *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–7, IEEE, 2023.

[6] 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, 2019.

[7] N. Mei, H. Wang, Y. Zhang, F. Liu, X. Jiang, and S. Wei, "Classification of heart sounds based on quality assessment and wavelet scattering transform," *Computers in biology and medicine*, vol. 137, p. 104814, 2021.

[8] W. Ghezaiel, B. Luc, and O. Lézoray, "Wavelet scattering transform and cnn for closed set speaker identification," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6, IEEE, 2020.

[9] Z. Abduh, E. A. Nehary, M. A. Wahed, and Y. M. Kadah, "Classification of heart sounds using fractional fourier transform based mel-frequency spectral coefficients and traditional classifiers," *Biomedical Signal Processing and Control*, vol. 57, p. 101788, 2020.

[10] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection," *Computers in biology and medicine*, vol. 100, pp. 132–143, 2018.

[11] S. Abbas, S. Ojo, A. Al Hejaili, G. A. Sampedro, A. Almadhor, M. M. Zaidi, and N. Kryvinska, "Artificial intelligence framework for heart disease classification from audio signals," *Scientific Reports*, vol. 14, no. 1, p. 3123, 2024.

[12] S. K. Padhy, A. Mohapatra, and S. Patra, "Heart sound classification based on mfcc feature extraction and long-short term neural networks," in *2023 2nd International Conference on Ambient Intelligence in Health Care (ICAIHC)*, pp. 1–6, IEEE, 2023.

[13] Z. Khodzhaev, "A practical guide to spectrogram analysis for audio signal processing," *arXiv preprint arXiv:2403.09321*, 2024.

[14] M. Rahmandani, H. A. Nugroho, and N. A. Setiawan, "Cardiac sound classification using mel-frequency cepstral coefficients (mfcc) and artificial neural network (ann)," in *2018 3rd International Conference on Information Technology, Information System and Electrical Engineering (ICITISEE)*, pp. 22–26, IEEE, 2018.

[15] M. S. Wibawa, I. M. D. Maysanjaya, N. K. D. P. Novianti, and P. N. Crisnapati, "Abnormal heart rhythm detection based on spectrogram of heart sound using convolutional neural network," in *2018 6th international conference on cyber and IT service management (CITSM)*, pp. 1–4, IEEE, 2018.

[16] H. Tang, Z. Dai, Y. Jiang, T. Li, C. Liu, *et al.*, "Pcg classification using multidomain features and svm classifier," *BioMed research international*, vol. 2018, 2018.

[17] C. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. Johnson, *et al.*, "An open access database for the evaluation of heart sound algorithms," *Physiological measurement*, vol. 37, no. 12, p. 2181, 2016.

[18] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[19] E. Oyallon, S. Mallat, and L. Sifre, "Generic deep networks with wavelet scattering," *arXiv preprint arXiv:1312.5940*, 2013.

[20] Z. K. Abdul and A. K. Al-Talabani, "Mel frequency cepstral coefficient and its applications: A review," *IEEE Access*, vol. 10, pp. 122136–122158, 2022.

[21] N. Instruments, "Stft spectrogram." https://www.ni.com/docs/en-US/bundle/labview-advanced-signal-processing-toolkit-api-ref/page/lvasptconcepts/aspt_stft_spectrogram.html#:~:text=The%20STFT%20spectrogram%2C%20a%20Cohen's,frequencies%20in%20the%20STFT%20spectrogram., 2024. Accessed: 2024-05-07.

[22] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," *Soft computing*, vol. 23, pp. 715–734, 2019.

[23] J. Rubin, R. Abreu, A. Ganguli, S. Nelaturi, I. Matei, and K. Sricharan, "Recognizing abnormal heart sounds using deep learning," *arXiv preprint arXiv:1707.04642*, 2017.

[24] 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)*, pp. 621–624, IEEE, 2016.

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