Sound & Vibration 2026, 60(4), 1706.
doi: https://doi.org/10.59400/sv1706

Article

Prediction of heart sound using the xLSTM method for hypertrophic cardiomyopathy

Shatiswaran Vigian, R Kanesaraj Ramasamy*, Junaidi Abdullah

Faculty Computing Informatics, Multimedia University, Cyberjaya 63100, Malaysia

*** Corresponding author:** R Kanesaraj Ramasamy, r.kanesaraj@mmu.edu.my

CITATION

Vigian S, Ramasamy RK, Abdullah J. Prediction of heart sound using the xLSTM method for hypertrophic cardiomyopathy. Sound & Vibration. 2026; 60(4): 1706. https://doi.org/10.59400/sv1706

ARTICLE INFO

Received: 9 September 2024
Revised: 28 March 2025
Accepted: 3 April 2025
Available online: 29 June 2026

Abstract: Heart disease has emerged as a major global public health concern, driven by poor dietary habits, unhealthy lifestyle choices, and limited health awareness. Accurate diagnosis of cardiac conditions remains challenging for hospitals and clinical institutions. Hypertrophic cardiomyopathy is an autosomal dominant disorder caused by mutations in sarcomere protein genes that affect the contractile function of cardiac muscle. With the growing adoption of digital health systems, large volumes of patient data are now collected and stored, providing opportunities for computational approaches to support clinical decision making. Machine learning methods have become increasingly important for analysing complex and nonlinear patterns in medical data. This study presents an ensemble-based approach for heart sound classification and introduces an Extended Long Short-Term Memory (xLSTM) model for the detection of cardiac abnormalities. The method was evaluated using acoustic features extracted from phonocardiogram recordings. The proposed model achieved 96.93% accuracy, 93.50% sensitivity, and 99.63% specificity, indicating strong performance in distinguishing normal and abnormal heart sounds. In comparison with previously reported techniques, the ensemble strategy and the xLSTM architecture provided improved accuracy. Model performance was assessed using accuracy, precision, recall, and F1 score, confirming the effectiveness of the proposed approach for automated heart sound analysis.

Keywords: acoustic signal analysis; cardiomyopathy; deep learning; extended long short-term memory; heart sound classification; machine learning; phonocardiogram

1. Introduction

The human heart is a vital organ responsible for circulating blood throughout the body by pumping it through a network of arteries, veins, and capillaries. Despite its essential role, the heart is highly vulnerable to disease and structural damage. As described by King and Lowery [1], heart disease interferes with the pumping mechanism of the heart and contributes to cardiac dysfunction. Cardiovascular disease encompasses a wide range of conditions that impair cardiac performance, and symptoms commonly include fatigue, swollen extremities, muscular weakness, and shortness of breath [2]. Major risk factors include smoking, elevated blood pressure, sedentary lifestyle, and high cholesterol levels [3, 4]. According to the World Health Organization, cardiovascular disease remains the leading cause of mortality worldwide, accounting for approximately 32 percent of all deaths, or nearly 18 million fatalities each year [5]. Coronary artery disease continues to be the most prevalent form of cardiovascular disorder and presents a significant public health concern [6, 7].

Abnormal heart sounds, particularly those associated with hypertrophic cardiomyopathy, arise when turbulent blood flow interacts with cardiac valves or when high-speed flow occurs across the aortic valve during systole [8]. Valve leakage or incomplete closure may also produce distinct acoustic patterns, and additional abnormal sounds can indicate more advanced cardiac conditions due to irregular ventricular flow [9,10]. Other irregularities, including extrasystoles and artifacts, further complicate the acoustic profile of cardiac recordings [11]. Distinguishing normal from abnormal heart sounds is, therefore, a complex task that requires considerable clinical experience and specialized training [9,10].

Advances in computational modelling have encouraged the use of machine learning techniques for automated cardiac sound analysis. Long Short-Term Memory networks have been successfully applied in many sequence-based tasks [12–14], and remained widely used until the emergence of Transformer based models [15]. Their effectiveness has been demonstrated in applications such as text generation [16, 17], handwriting synthesis [16], sequence translation [18], program evaluation [19], image captioning [20,21], rainfall runoff modelling [22,23], and hydrological forecasting [24]. These networks are also used in reinforcement learning systems such as OpenAI Five and AlphaStar [25]. Their strong performance in modelling temporal patterns makes them well suited for analysing the dynamic characteristics of phonocardiogram signals.

Heart sound recordings represent non stationary acoustic signals influenced by physiological, environmental, and instrumental factors. Consequently, computational models must be capable of extracting meaningful information from variable and noisy data. The development of advanced deep learning architectures provides opportunities to enhance automated detection of abnormal heart sounds. This study investigates the use of an Extended Long Short-Term Memory model for heart sound classification and evaluates its performance using multiple acoustic features derived from phonocardiogram recordings.

2. Acoustic signal processing background

According to Shinde and Lavanya [26], the Detailed Residual Network, also known as DRN, is trained using the Jellyfish Search Optimizer (JSO) method at the final stage of the heart disease diagnosis process, when the DRN performs the detection of cardiac illness. The proposed DRN JSO approach achieved improved performance measures, including a true positive rate of 82.65%, an F1 score of 83.65%, an accuracy of 84.65%, and a Matthews correlation coefficient of 85.65%.

Ainiwaer et al. [27] reported that the VGG 16 model performed better than ResNet 18 and CNN 7 for heart sound analysis. In the test set, VGG 16 recorded area under the curve values of 0.755 (95% confidence interval 0.614 to 0.819) and 0.652 (95% confidence interval 0.554 to 0.770), while VGG 16 alone achieved an area under the curve of 0.834 (95% confidence interval 0.736 to 0.930). The model demonstrated a sensitivity of 8.04 percent and a specificity of 86.2%. When combined with PTP scores, the diagnostic performance improved, with area under the curve values reaching 0.908 (95% confidence interval 0.845 to 0.971) and 0.915 (95% confidence interval 0.855 to 0.974) for the combined VGG and DF score model. In individuals with three vascular lesions and coronary artery obstruction, VGG 16 demonstrated sensitivity and specificity values greater than 0.85.

According to Ren et al. [28], wavelet transformation presents challenges because selecting an appropriate wavelet function requires extensive empirical analysis. However, recent progress in deep learning makes it possible to extract heart sound features without expert-based selection. Higher level representations of the heart sounds can be obtained automatically from pretrained convolutional neural networks and then provided to a support vector machine classifier. Tolic et al. [29] introduced chrono initialized Long Short Term Memory networks with layer normalization that improved learning stability for sequential acoustic signals. However, Ghosh et al. [30] noted that class imbalance remains a significant challenge in deep learning models for heart sound classification and may require additional optimization strategies

Fernando et al. [31] proposed an attention based deep learning model for heart sound segmentation using a sequence-to-sequence strategy. Their model outperformed several advanced baseline approaches. Amiriparian et al. [32] demonstrated the use of an auto encoder based recurrent neural network for unsupervised representation learning, showing the effectiveness of such methods for heart sound analysis. Similarly, Vinay et al. [33] proposed an RNN Bi LSTM based multi decision generative adversarial approach for cardiovascular disease recognition, emphasizing the importance of feature optimization in improving classification performance. Aji et al. [34] further showed that a combined CNN and Long Short Term Memory architecture can effectively capture both local and temporal patterns in heartbeat sound classification. Rizal et al. [35] demonstrated that Hjorth features combined with an LSTM RNN classifier can accurately differentiate normal and abnormal heart sounds.

Zeinali and Niaki [36] showed that classical signal processing combined with conventional machine learning algorithms remains effective for heart sound classification, particularly in controlled acoustic conditions. According to Latifi et al. [37], the general block diagram of the Multi Branch Deep Convolutional Neural Network model is to enhance the efficiency of convolutional neural networks in audio processing, filters such as 1 by 3 or 1 by 5 are used to capture high frequency patterns and local details. Larger filters such as 1 by 11 or 1 by 9 can capture broader patterns that occur less frequently. Using multiple convolutional layers with different filter sizes and filter counts enables the model to learn a rich multi scale representation of the input audio signal. The optimal filter sizes and configurations are obtained through meta parameter searches and experimentation with various convolutional neural network architectures.

The design of an effective one-dimensional convolutional neural network for audio processing requires careful consideration of the architecture, training strategy, and data preparation procedure. By testing multiple configurations, one dimensional convolutional neural network can accurately represent audio signals and perform reliable classification. The dataset must be divided into training, validation, and test sets. The model is trained using the training set, the validation set is used to tune parameters and select the optimal configuration, and the test set is used to assess final performance.

3. Materials and methods

3.1. Dataset and data preparation

The heartbeat sound dataset used in this research was obtained from the "Kaggle data science community," where the original samples were separated into two datasets (set a and set b) for preprocessing. The complete dataset was then divided into training and test sets, and the unlabeled samples were included in the test set for prediction.

In addition to this dataset, the PhysioNet CinC 2016 dataset [38] was used. This dataset consists of 4,430 recordings from 1,072 subjects, containing a total of 233,512 heart sounds. The recordings originate from six different sources (A, B, C, D, E, and F), with durations ranging from five seconds to more than twenty seconds. These include participants with coronary artery disease, heart valve disease, and healthy individuals. The dataset provides two major classes, normal and abnormal, and was further adapted into four categories for this study, which are artifact (class 0), hypertrophic cardiomyopathy (class 1), normal (class 2), and extrahls (class 3).

All sound recordings were standardized to a duration of five seconds. Any recording shorter than five seconds was removed to ensure uniformity and compatibility with the feature extraction process.

3.2. Preprocessing and feature extraction

The preprocessing stage began with down sampling the phonocardiogram signal to 1 kHz using a polyphase antialiasing filter. An example of the preprocessed heart sound waveform from the PhysioNet CinC 2016 dataset is shown in **Figure 1**. This step ensured that the most relevant diagnostic frequency components were retained while reducing computational load.

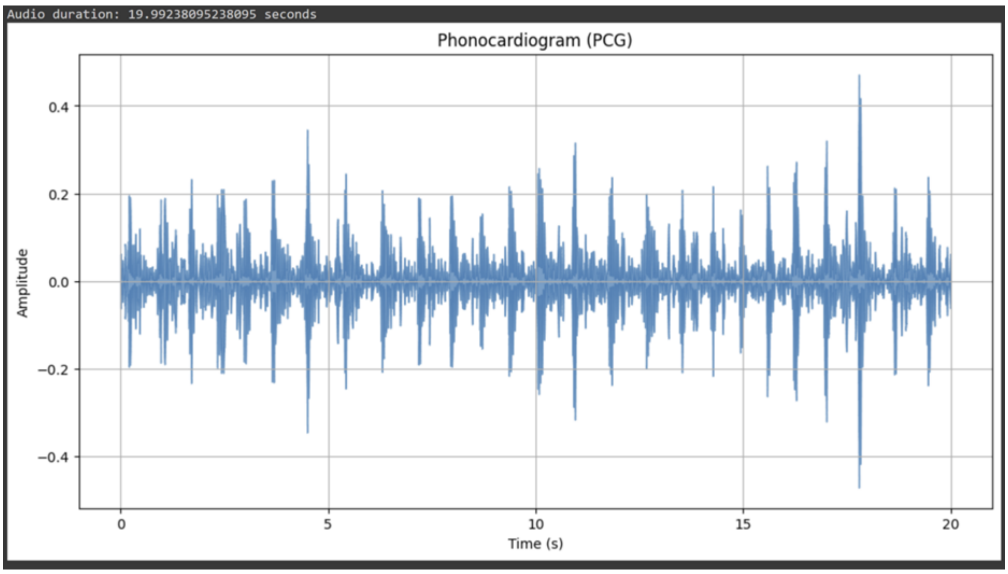

Figure 1. Preprocessed acoustic waveform of a heart sound recording from the PhysioNet CinC 2016 dataset [39].

Mel frequency cepstral coefficients (MFCC) were extracted as the main features. The MFCC represent the cosine transform of the short duration logarithmic power spectrum on a nonlinear Mel frequency scale. As reported by Deng et al. [39] and

Nogueira et al. [40], the MFCC are widely used in audio analysis due to their alignment with human auditory perception and their ability to capture the most informative spectral components. An example MFCC representation of the heart sound is shown in **Figure 2**.

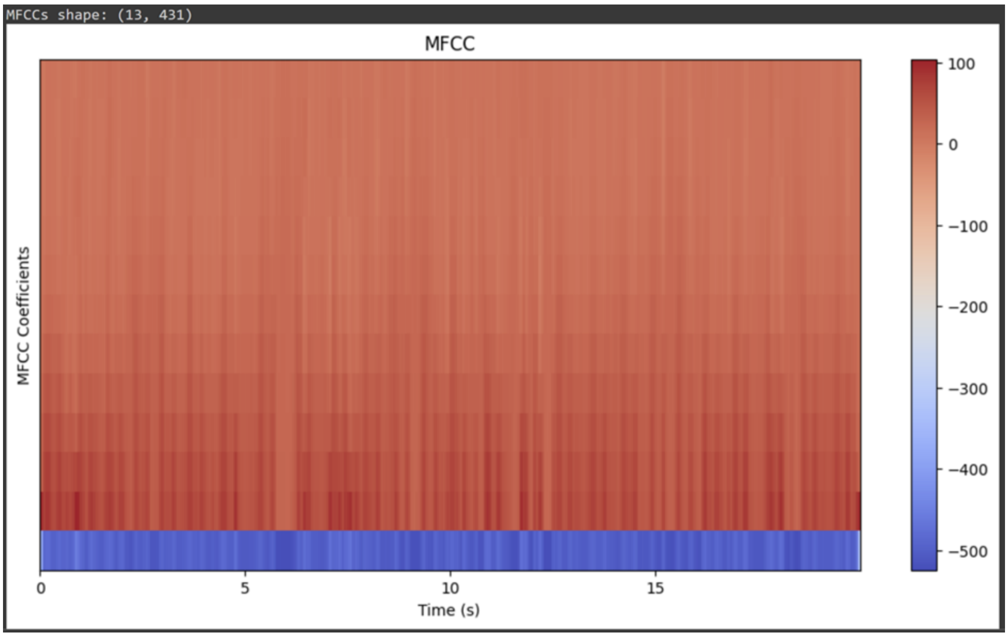

Figure 2. Mel frequency cepstral coefficient features (MFCC) extracted from the heart sound recording.

Dynamic and static Welch methods were applied to enhance power spectrum estimation. These feature maps were then used as input representations for the deep learning model.

3.3. Model architecture

Sequential data recorded across multiple time intervals were processed using a stacked Long Short-Term Memory (LSTM) model. The architecture of the proposed stacked LSTM model is shown in **Figure 3**, where multiple concealed LSTM layers are arranged to capture temporal patterns. Each layer contains multiple memory cells that allow the model to generate a single representative output for each processed three-dimensional input sequence.

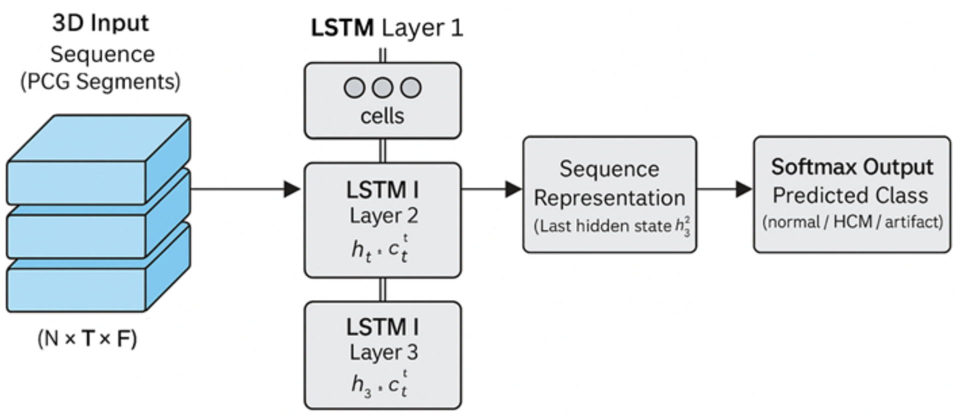

Figure 3. Flow of the proposed stacked LSTM model for heart sound classification.

To address the limitations of the standard Long Short-Term Memory, the Extended Long Short-Term Memory (xLSTM) architecture was introduced. This architecture enhances the standard LSTM formulation (Equation (1)) by including two new variants:

- sLSTM, which uses scalar memory with scalar updates and memory mixing;
- mLSTM, which uses matrix memory with a parallelizable covariance update method.

$$\begin{aligned} i_t &= \sigma\left(W_i x_t + U_i h_{\{t-1\}} + b_i\right) \\ f_t &= \sigma\left(W_f x_t + U_f h_{\{t-1\}} + b_f\right) \\ o_t &= \sigma\left(W_o x_t + U_o h_{\{t-1\}} + b_o\right) \\ \hat{c}_t &= \tanh\left(W_c x_t + U_c h_{\{t-1\}} + b_c\right) \\ c_t &= f_t \odot c_{\{t-1\}} + i_t \odot \hat{c}_t \\ h_t &= o_t \odot \tanh\left(c_t\right). \end{aligned} \tag{1}$$

Both sLSTM and mLSTM apply exponential gating to improve memory regulation. The mLSTM removes hidden to hidden memory mixing to enable parallel execution, while the sLSTM retains memory mixing across scalar memory cells. These components, together with the residual blocks and multi head structures, form the Extended Long Short-Term Memory family, illustrated in **Figure 4**.

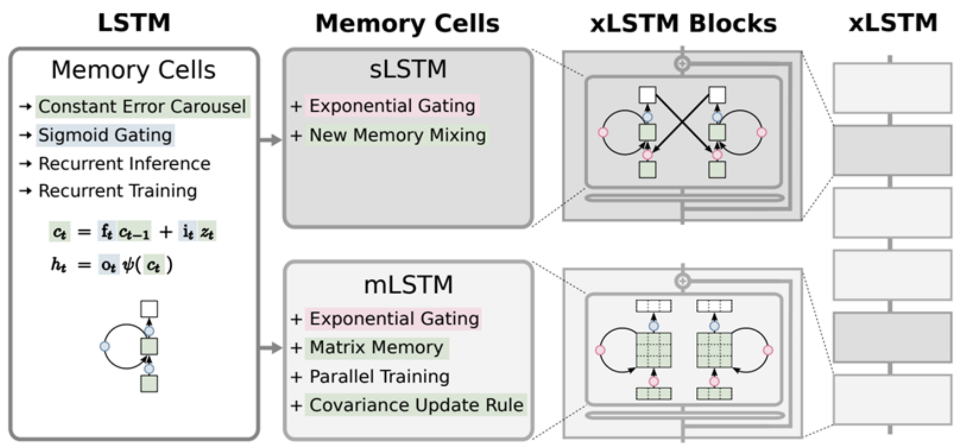

Figure 4. Overview of the extended long short term memory (xLSTM) architecture and its variants.

3.4. Training and evaluation procedure

The dataset was divided into training, validation, and test sets before model training. The stacked Long Short-Term Memory (LSTM) and Extended Long Short-Term Memory (xLSTM) models were trained in a supervised manner. The models were applied to both labeled and unlabeled samples, where a random unlabeled sample was selected to estimate its class label.

Several machine learning algorithms were used as comparison methods, including Naive Bayes, Convolutional Neural Network, Support Vector Machine, Random Forest, Bagging, and Boosting. An artificial neural network was also used to extract MFCC features for classifying normal and abnormal heart sounds.

To improve generalization, dropout layers, sigmoid gating with L2 regularization, and batch normalization were incorporated. Greater model depth allowed extraction of both low frequency and high frequency features, resulting in improved classification accuracy.

3.5. Baseline models for comparison

Multiple baseline methods from recent literature were used to benchmark performance. These include one dimensional Convolutional Neural Network, multilayer perceptron, and multi branch Convolutional Neural Network architectures, along with classical machine learning classifiers. Studies by Shinde and Lavanya [26], Ainiwaer et al. [27], and Ren et al. [28], demonstrate the effectiveness of these models in heart sound analysis tasks. These baselines provided a framework for evaluating the robustness of the proposed Extended Long Short-Term Memory model.

3.6. Performance metrics

The performance was evaluated using accuracy (ACC), sensitivity (SEN), specificity (SPC), and precision (PPV), which were computed using Equations (2)–(5). The F measure (F1 score), derived from sensitivity and precision, was computed using Equation (6). The kappa coefficient (Equation (7)) was also used to measure classifier agreement, where a value of 1 represents perfect agreement. Before presenting the equations, the following terms are used throughout the metric definitions:

- TP (True Positive): Cases where abnormal heart sounds were correctly classified as abnormal.
- TN (True Negative): Cases where normal heart sounds were correctly classified as normal.
- FP (False Positive): Cases where normal heart sounds were incorrectly classified as abnormal.
- FN (False Negative): Cases where abnormal heart sounds were incorrectly classified as normal.

The indications listed above may be expressed as follows:

$$\mathrm{ACC}\ (\%) = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \times 100, \tag{2}$$

$$\mathrm{SEN}\ (\%) = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \times 100, \tag{3}$$

$$\mathrm{SPC}\ (\%) = \frac{\mathrm{TN}}{(\mathrm{TN} + \mathrm{FP})} \times 100, \tag{4}$$

$$\mathrm{PPV}\ (\%) = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})} \times 100, \tag{5}$$

where the numbers for TP, TN, FP, and FN, respectively, are represented by the variables, F-measure, also known as the F1-score, is another well-known metric that was influenced by SEN and PPV metrics. It is computed as follows:

$$\text{F1-score}(\%) = 2 \times \left(\frac{\mathrm{PPV} \times \mathrm{SEN}}{\mathrm{PPV} + \mathrm{SEN}}\right) \times 100. \tag{6}$$

And the kappa coefficient, defined as follows, is also used:

$$\mathrm{Pe} = \Sigma \left[\frac{(\mathrm{row_i} \times \mathrm{column_i})}{\mathrm{N}^2} \right]. \tag{7}$$

4. Experimental results

Information and preparation of the heart sound recordings were carried out using data collected from a wide range of participants worldwide in both clinical and non-clinical environments. These participants included individuals with coronary artery disease and heart valve disease as well as healthy subjects. The PhysioNet CinC 2016 dataset [39], which contains 4,430 recordings from 1,072 subjects and a total of 233,512 heart sounds, was used in this research. The dataset consists of two classes, normal and abnormal, and includes recordings from six different sources (A, B, C, D, E, and F) with lengths ranging from five seconds to more than twenty seconds.

The inclusion of both normal and abnormal heart sound recordings allowed assessment of the model's capability to detect and categorize anomalies, which is essential for improving the precision of any machine learning model. All signals were standardized to a duration of five seconds, and recordings shorter than five seconds were excluded before the experimental analysis.

Evaluation results

The proposed Extended Long Short-Term Memory (xLSTM) model achieved strong performance across several evaluation metrics. Using the accuracy, sensitivity, specificity, precision, F measure, and kappa coefficient, the model attained 96.93% accuracy, 93.50% sensitivity, and 99.63% specificity on the test dataset. These results demonstrate that the model was able to correctly classify a high proportion of both normal and abnormal heart sound samples.

The confusion matrix for the preprocessed xLSTM model is presented in **Figure 5**. This matrix illustrates the distribution of true and predicted classes and highlights how effectively the model distinguishes between the four categories used in this study. The high number of true positive and true negative predictions supports the reliability of the model in identifying heart sound patterns.

The model demonstrated superior performance in feature extraction, network design, and classification compared to several alternative methods. It effectively addressed challenges such as low signal to noise ratio, variability in heart sound recordings, and limited labeled data. The deeper architecture of the model contributed to improved performance, as larger filters extracted low frequency features while smaller filters extracted high frequency features. The multi branch structure further enhanced classification accuracy, and the use of dynamic and static Welch methods provided improved power spectrum estimates that served as valuable input features for the deep learning approach.

Although the sensitivity and specificity values remain below the thresholds required for clinical diagnostic use, Short Time Fourier Transform based features clearly improved the differentiation of normal and pathological heart sound signals.

The comparison between the Multi Branch Deep Convolutional Network and the xLSTM model confirmed the strong performance of the proposed design and demonstrated its potential usefulness in automated cardiac sound analysis.

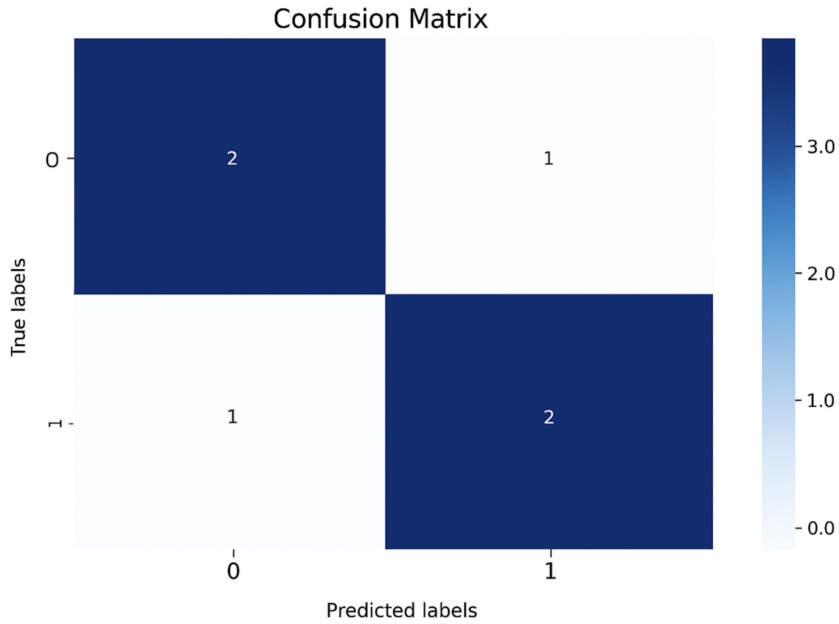

Figure 5. Confusion matrix showing the classification performance of the preprocessed extended long short term memory (xLSTM) model.

5. Discussion

The results show that the proposed Extended Long Short-Term Memory (xLSTM) model achieved strong performance in the classification of heart sounds, with an accuracy of 96.93%, a sensitivity of 93.50%, and a specificity of 99.6%. These values indicate that the model can reliably distinguish between normal and abnormal cardiac sounds across multiple recording conditions [41]. The deeper architecture, which incorporates both small and large filters, helped extract low-frequency and high-frequency acoustic patterns, while the dynamic and static features from the Welch method improved the power spectrum representation of the phonocardiogram signal. These elements contributed to the overall improvement in classification performance [42].

When comparing the proposed method to previous techniques, the findings align with earlier work demonstrating the effectiveness of deep neural networks for cardiac sound analysis. Several prior studies, including the models proposed by Shinde and Lavanya [26], Ainiwaer et al. [27], and Ren et al. [28], reported that convolutional and recurrent neural network architectures outperform traditional feature-based approaches. Consistent with these earlier findings, the xLSTM model used in this study demonstrated the benefit of improved memory representation and temporal modeling in acoustic classification tasks. As noted in the comparative approaches evaluated in this research, these techniques mimic the processing of complex sound information in the human auditory system and can be used to identify abnormal sounds associated with cardiovascular disease.

Although the results are promising, the sensitivity and specificity values have not yet reached a level suitable for clinical diagnosis. Variability in recording conditions, differences in patient physiology, and the presence of noise within the heart sound recordings may affect performance. Future work may focus on incorporating additional acoustic features, expanding the dataset, and refining the model to improve its generalization in real world environments.

6. Conclusion

This study presented an Extended Long Short-Term Memory (xLSTM) model for the classification of heart sounds using acoustic features derived from phonocardiogram recordings. The model achieved 96.93% accuracy, 93.50% sensitivity, and 99.63% specificity, demonstrating strong capability in identifying normal and abnormal cardiac sounds across multiple recording conditions. The use of deeper memory-based architecture, combined with Mel frequency cepstral coefficients and enhanced power spectrum estimates, contributed to improved feature representation and more reliable classification.

The findings confirm the value of advanced deep learning approaches in processing non stationary acoustic signals such as heart sounds. Compared with previous methods reported in the literature, the proposed model shows competitive and, in several cases, superior performance, highlighting its potential use as a supportive tool in automated cardiovascular screening. Although the current outcomes do not meet clinical diagnostic requirements, the approach offers a solid foundation for further development.

Future work may include expanding the dataset, incorporating additional acoustic features, exploring attention-based mechanisms, and evaluating performance in real world environments. These improvements may enhance the generalization of the system and support the design of more robust and clinically relevant heart sound analysis tools.

Author contributions: Conceptualization, SV and RKR; methodology, SV and RKR; validation, SV, RKR and JA; formal analysis, SV; writing—original draft preparation, SV; writing—review and editing, RKR and JA; supervision, RKR and JA; project administration, RKR; funding acquisition, RKR. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS), Project ID: FRGS/1/2022/ICT08/MMU/03/1.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

AI use statement: During the preparation of this manuscript, the authors used

AI-assisted tools to support language refinement, gram-mar checking, clarity improvement, and formatting of selected sections. he authors subsequently reviewed and edited the content as necessary and take full responsibility for the final content of the published article.

References

1. King J, Lowery DR. Physiology, Cardiac Output. StatPearls Publishing; 2023. Available online: https://pubmed.ncbi.nlm.nih.gov/29262215/
2. Powell-Wiley TM, Poirier P, Burke LE, et al. Obesity and Cardiovascular Disease: A Scientific Statement From the American Heart Association. Circulation. 2021; 143(21). doi: 10.1161/CIR.0000000000000973
3. Reddy RK, Charles WN, Sklavounos A, et al. The effect of smoking on COVID-19 severity: A systematic review and meta-analysis. Journal of Medical Virology. 2021; 93(2): 1045–1056. doi: 10.1002/jmv.26389
4. Iglesias MJ, Kruse LD, Sanchez-Rivera L, et al. Identification of Endothelial Proteins in Plasma Associated With Cardiovascular Risk Factors. Arteriosclerosis, Thrombosis, and Vascular Biology. 2021; 41(12): 2990–3004. doi: 10.1161/ATVBAHA.121.316779
5. Nishiga M, Wang DW, Han Y, et al. COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. Nature Reviews Cardiology. 2020; 17(9): 543–558. doi: 10.1038/s41569-020-0413-9
6. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. Circulation Research. 2016; 118(4): 564–578. doi: 10.1161/CIRCRESAHA.115.306566
7. Nakao T, Natarajan P. Clonal hematopoiesis, multi-omics and coronary artery disease. Nature Cardiovascular Research. 2022; 1(11): 965–967. doi: 10.1038/s44161-022-00154-1
8. Bouril A, Aleinikava D, Guillem MS, et al. Automated Classification of Normal and Abnormal Heart Sounds using Support Vector Machines. In: Proceedings of the 2016 Computing in Cardiology Conference; 11–14 September 2016; Vancouver, BC, Canada. doi: 10.22489/CinC.2016.158-329
9. Dao AT. Wireless laptop-based phonocardiograph and diagnosis. PeerJ. 2015; 3: e1178. doi: 10.7717/peerj.1178
10. Cester L, Starshynov I, Jones Y, et al. Remote laser-speckle sensing of heart sounds for health assessment and biometric identification. Biomedical Optics Express. 2022; 13(7): 3743. doi: 10.1364/BOE.451416
11. Rahmawati D. Accuracy of the NADAS Criteria in Helping Establish the Diagnosis in Children With Suspected to Have Congenital Heart Disease. Paediatrica Indonesiana. 2023; 63(4): 267–73. doi: 10.14238/pi63.4.2023.267-73
12. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997; 9(8): 1735–1780. doi: 10.1162/neco.1997.9.8.1735
13. Hochreiter S, Heusel M, Obermayer K. Fast model-based protein homology detection without alignment. Bioinformatics. 2007; 23(14): 1728–1736. doi: 10.1093/bioinformatics/btm247
14. Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks. 2015; 61: 85–117. doi: 10.1016/j.neunet.2014.09.003
15. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. arXiv preprint. 2023. doi: 10.48550/arXiv.1706.03762
16. Graves A, Jaitly N, Mohamed A. Hybrid speech recognition with Deep Bidirectional LSTM. In: Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding; 8–12 December 2013; Olomouc, Czech Republic. pp. 273–278. doi: 10.1109/ASRU.2013.6707742
17. Karpathy A, Johnson J, Fei-Fei L. Visualizing and Understanding Recurrent Networks. arXiv preprint. 2015. doi: 10.48550/arXiv.1506.02078
18. Sutskever I, Vinyals O, Le QV. Sequence to Sequence Learning with Neural Networks. arXiv preprint. 2014. doi: 10.48550/arXiv.1409.3215
19. Zaremba W, Sutskever I. Learning to Execute. arXiv preprint. 2015. doi: 10.48550/arXiv.1410.4615
20. Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 7–12 June 2015; Boston, MA, USA. pp. 3128–3137. doi: 10.1109/CVPR.2015.7298932
21. Hossain MZ, Sohel F, Shiratuddin MF, et al. A Comprehensive Survey of Deep Learning for Image Captioning. ACM Computing Surveys. 2019; 51(6): 1–36. doi: 10.1145/3295748

22. Kratzert F, Klotz D, Brenner C, et al. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences. 2018; 22(11): 6005–6022. doi: 10.5194/hess-22-6005-2018
23. Kratzert F, Klotz D, Herrnegger M, et al. Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. Water Resources Research. 2019; 55(12): 11344–11354. doi: 10.1029/2019WR026065
24. Nearing G, Cohen D, Dube V, et al. Global prediction of extreme floods in ungauged watersheds. Nature. 2024; 627(8004): 559–563. doi: 10.1038/s41586-024-07145-1
25. Lokman A, Ramasamy RK. Smart Toilet: Threats and Challenges Identifying Human Presence Using IoT Sensors. In: Proceedings of the 3rd International Conference on Big Data and Internet of Things; 22–24 August 2019; Melbourne VIC Australia. pp. 56–60. doi: 10.1145/3361758.3361765
26. Shinde R, Lavanya V. Heart Disease Prediction using Jellyfish Search Optimization and Deep Residual Networks. International Journal of Electrical and Electronics Engineering. 2023; 10(5): 165–173. doi: 10.14445/23488379/IJEEE-V10I5P115
27. Ainiwaer A, Hou WQ, Qi Q, et al. Deep learning of heart-sound signals for efficient prediction of obstructive coronary artery disease. Heliyon. 2024; 10(1): e23354. doi: 10.1016/j.heliyon.2023.e23354
28. Ren J, Mao X, Fu S. Image-based flow decomposition using empirical wavelet transform. Journal of Fluid Mechanics. 2021; 906: A22. doi: 10.1017/jfm.2020.817
29. Tolic A, Boshkoska BM, Skansi S. Chrono Initialized LSTM Networks With Layer Normalization. IEEE Access. 2024; 12: 115219–115236. doi: 10.1109/ACCESS.2024.3445329
30. Ghosh K, Bellinger C, Corizzo R, et al. The class imbalance problem in deep learning. Machine Learning. 2024; 113(7): 4845–4901. doi: 10.1007/s10994-022-06268-8
31. Fernando T, Gammulle H, Denman S, et al. Deep Learning for Medical Anomaly Detection—A Survey. ACM Computing Surveys. 2022; 54(7): 1–37. doi: 10.1145/3464423
32. Amiriparian S, Schmitt M, Ottl S, et al. Deep Unsupervised Representation Learning for Audio-Based Medical Applications. In: Deep Learners and Deep Learner Descriptors for Medical Applications, Intelligent Systems Reference Library. Springer International Publishing; 2020. pp. 137–164. doi: 10.1007/978-3-030-42750-4_5
33. Vinay NA, Vidyasagar KN, Rohith S, et al. An RNN-Bi LSTM Based Multi Decision GAN Approach for the Recognition of Cardiovascular Disease (CVD) From Heart Beat Sound: A Feature Optimization Process. IEEE Access. 2024; 12: 65482–65502.doi: 10.1109/ACCESS.2024.3397574
34. Aji NB, Kurnianingsih K, Masuyama N, et al. CNN-LSTM for Heartbeat Sound Classification. JOIV : International Journal on Informatics Visualization. 2024; 8(2): 735. doi: 10.62527/joiv.8.2.2115
35. Rizal A, Fauzi H, Wulandari DR, et al. Classification of Normal and Abnormal Heart Sounds Using Hjorth Features and LSTM – RNN Algorithm. Journal of Physics: Conference Series. 2023; 2673(1): 012003. doi: 10.1088/1742-6596/2673/1/012003
36. Zeinali Y, Niaki STA. Heart sound classification using signal processing and machine learning algorithms. Machine Learning with Applications. 2022; 7: 100206. doi: 10.1016/j.mlwa.2021.100206
37. Latifi SA, Ghassemian H, Imani M. Multi-branch convolutional network and LSTM-CNN for heart sound classification. Physical and Engineering Sciences in Medicine. 2025; 49(1): 263–287. doi: 10.1007/s13246-025-01664-5
38. Jing H, Jie R, Siqi L, et al. A cardiac audio classification method based on image expression of multidimensional features. Scientific Reports. 2024; 14(1): 23196. doi: 10.1038/s41598-024-73237-7
39. Deng M, Meng T, Cao J, et al. Heart sound classification based on improved MFCC features and convolutional recurrent neural networks. Neural Networks. 2020; 130: 22–32. doi: 10.1016/j.neunet.2020.06.015
40. Nogueira DM, Ferreira CA, Gomes EF, et al. Classifying Heart Sounds Using Images of Motifs, MFCC and Temporal Features. Journal of Medical Systems. 2019; 43(6): 168. doi: 10.1007/s10916-019-1286-5
41. Khan FA, Abid A, Khan MS. Automatic heart sound classification from segmented/unsegmented phonocardiogram signals using time and frequency features. Physiological Measurement. 2020; 41(5): 055006. doi: 10.1088/1361-6579/ab8770
42. Chen W, Sun Q, Chen X, et al. Deep Learning Methods for Heart Sounds Classification: A Systematic Review. Entropy. 2021; 23(6): 667. doi: 10.3390/e23060667