Advancements in Deep Learning for Cardiac Sound Classification: A Review on Architectures, Preprocessing Techniques and Datasets

Pragya Arora* , Ritika Kumari, Poonam Bansal

Department of Artificial Intelligence and Data Sciences, IGDTUW, New Church Road, Delhi 110006, India

* Corresponding author

doi: https://doi.org/10.21467/proceedings.178.24

ABSTRACT

The identification and classification of cardiac sound patterns play a crucial role in the timely detection and diagnosis of cardiovascular diseases. Advances in deep learning techniques have significantly contributed to this field, offering promising solutions for detailed and accurate analysis. This study provides an overview of advanced methodologies for classifying cardiac sounds using deep learning approaches. Various deep learning architectures, including convolutional neural networks, recurrent neural networks, and hybrid models, are discussed, highlighting their advantages and limitations in analyzing cardiac sounds. Additionally, the preprocessing techniques essential for preparing audio data for deep learning models are examined, along with a review of publicly available datasets commonly used in this area of research. The evaluation metrics for assessing model performance are also explored, and future directions for advancing this field are outlined. This study aims to guide researchers and practitioners in developing robust deep learning frameworks for monitoring cardiovascular health.

Keywords: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Audio Data Preprocessing, Cardiovascular Disease Detection

1 Introduction

The classification of heartbeat sounds is crucial for diagnosing cardiovascular conditions, offering valuable insights into cardiac health, and enabling timely medical interventions [1]. Conventional classification techniques often rely on manual interpretation, which may introduce subjectivity and inefficiencies [2][3]. Recent advancements in deep learning have transformed biomedical signal processing, enabling automated and precise classification of heartbeat sounds [4][5]. Heartbeat sounds, referred to as cardiac auscultation, consist of fundamental components identified as S1 and S2 [6][7]. These sounds originate from the closure of heart valves and provide critical indicators of the cardiac cycle phases. Abnormalities in these sounds, such as murmurs, clicks, or gallops, serve as essential markers for cardiac conditions like valvular diseases, myocardial infarction, and heart failure [8][9]. Deep learning techniques have proven to be significant in biomedical signal processing due to their ability to extract meaningful features directly from raw data. Convolutional neural networks, initially developed for image classification, have been adapted for analyzing sequential data such as heartbeat sounds [10]. Various deep learning models, including convolutional and recurrent neural networks, have demonstrated exceptional performance in signal processing tasks, including heartbeat sound classification. By extracting hierarchical features directly from raw audio, these models effectively capture complex patterns and subtle characteristics in cardiac sounds [11]. This study provides an in-depth review of recent advancements in heartbeat sound classification using deep learning. It highlights the importance of this classification, emphasizes the relevance of deep learning in biomedical signal processing, and examines the methodologies, technologies, datasets, and performance metrics crucial to this rapidly evolving field.

Proceedings DOI: 10.21467/proceedings.178; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-984081-8-1

2 Heart Sound Data Preprocessing

Data preprocessing is a critical step in the development of machine learning models, as it ensures that input data[12] is appropriately structured and refined for effective model training and evaluation. In the context of cardiac sound classification, preprocessing is particularly essential, given the complexity and variability of audio signals. The preprocessing phase generally involves multiple stages, each designed to address specific challenges associated with raw audio data. These stages include noise reduction, normalization, segmentation, and feature extraction.

2.1 Denoising and Segmentation Techniques

In the study of denoising heart sound data, the primary goal is to remove unwanted noise while preserving the essential cardiac signal. Several approaches, including wavelet denoising, adaptive filtering, and spectral subtraction, have been applied to effectively reduce noise interference. Wavelet-based denoising techniques, for example, utilize multi-resolution analysis to selectively isolate and remove noise components while maintaining the integrity of the primary signal. A notable approach was presented in [13], proposing a specialized wavelet denoising method that has shown significant effectiveness for heart sound signals. Segmentation techniques are also essential for accurately delineating individual cardiac cycles and supporting further analysis. Deep learning-based segmentation methods, such as convolutional neural networks, have been utilized for their precision in identifying and segmenting cardiac cycles within noisy recordings. For instance, the method presented in [14] achieved high accuracy in cardiac cycle segmentation, even in conditions with substantial noise interference. This segmentation approach, grounded in deep learning principles, has demonstrated robust performance and reliability across various noise levels.

2.2 Feature Extraction Methods

The process of feature extraction is critical as it captures relevant information necessary for subsequent analysis and classification. Frequency-domain features, including spectral power and dominant frequency, provide key insights into the frequency content of the signal. Advanced techniques such as wavelet transform and empirical mode decomposition are widely used to extract meaningful features from complex signals. Additionally, feature learning methods based on deep learning have emerged as effective approaches for identifying discriminative features, significantly improving diagnostic accuracy. A notable study presented in [15] utilized a hybrid feature extraction approach that combines wavelet transform with feature learning techniques based on deep learning. This method integrates temporal and spectral characteristics of heart sound signals, enabling enhanced classification of heart abnormalities. The hybrid approach demonstrated superior performance compared to traditional feature extraction techniques, highlighting its potential for improving diagnostic outcomes.

2.3 Data Augmentation and Balancing

In the field of heart sound analysis, data augmentation techniques are essential for expanding the size and diversity of the dataset, which in turn improves the robustness and generalization ability of machine learning models. Addressing class imbalance is also critical to ensure that the model is trained and evaluated in an unbiased manner. Techniques such as oversampling of minority classes and synthetic data generation are crucial in overcoming these challenges. A notable approach presented in [16] combined oversampling methods with synthetic data generation using generative adversarial networks to address imbalances in heart sound datasets. This combined strategy led to significant improvements in model performance and its ability to handle class imbalance effectively.

2.4 Dataset Splitting for Model Evaluation

Efficient partitioning of datasets is crucial for unbiased model evaluation and accurate performance assessment. A study by [17] compared various dataset-splitting methods for heart sound classification tasks.

The results highlighted the effectiveness of combining stratified sampling with k-fold cross-validation, especially in cases with imbalanced class distributions. A summary of the preprocessing techniques used in heartbeat sound classification is provided in Table 1. The table outlines the methods for preprocessing heart sound data, such as denoising, segmentation, feature extraction, and data augmentation, along with the outcomes of each approach. It also covers the implementation of Convolutional Neural Networks (CNNs) for spectrogram analysis, Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and hybrid models that integrate CNNs and RNNs. Furthermore, the table explores the effectiveness of transfer learning using pre-trained models in heart sound analysis. Each method is discussed in detail, and references are provided to relevant studies that have applied these techniques in the field.

Table 1. *Heart Sound Data Preprocessing Research Comparison*

Authors	Year	Techniques	Findings
Smith et al. [13]	2016	Wavelet based denoising	A wavelet based denoising method removes noise from heart sound signals while preserving key features. It enhances signal quality, improving the accuracy of heart sound analysis and aiding in the detection of cardiac abnormalities.
Zhang et al. [14]	2019	Deep learning based segmentation (CNNs)	A deep learning based segmentation method uses Convolutional Neural Networks (CNNs) to accurately detect and segment cardiac cycles in noisy heart sound recordings. This approach ensures high segmentation accuracy, effectively isolating relevant cardiac events for further analysis.
Liang et al.[15]	2018	Hybrid feature extraction (wavelet transform deep learning based feature learning)	A hybrid approach combining wavelet transform with deep learning based feature learning has advanced the classification of heart abnormalities. By leveraging the wavelet transform's ability to capture multi-scale features and deep learning’s power to learn complex patterns, this method outperforms traditional feature extraction techniques in detecting and classifying cardiac conditions.
Gupta et al.[16]	2020	Data augmentation (oversampling and synthetic data via GANs) and Class balancing	By implementing techniques such as data augmentation and balanced sampling, this method ensures more robust and accurate classification, leading to better detection of rare cardiac conditions.
Khan et al.[17]	2019	Comparative study on dataset-splitting strategies (random sampling, stratified sampling)	Investigated different dataset splitting strategies for heart sound classification tasks, emphasizing unbiased model evaluation and representative performance assessment

3 Deep Learning Architecture for Heartbeat Sound Classification

This paper undertakes a comprehensive examination of deep learning architectures specifically tailored for heart sound classification. Figure 1 presents a comparative study of various deep learning models used for heartbeat sound classification, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. The performance of each model is evaluated using several key metrics such as accuracy, precision, recall, and F1-score, which provide a clear understanding of their respective strengths and limitations in classifying heartbeat sounds. Accuracy measures the overall correct classification rate, indicating the proportion of correctly predicted instances out of the total. Precision focuses on the proportion of true positive classifications among all positive predictions, highlighting the model's ability to avoid false positives. Recall, on the other hand, measures the model's ability to correctly identify all relevant instances, representing the proportion of true positives among all actual positives. The F1-score is the harmonic mean of precision and recall, offering a balanced measure that accounts for both false positives and false negatives. These metrics, when used together, offer a thorough evaluation of the models' performance, allowing for a deeper insight into their effectiveness and suitability for heart sound classification tasks.

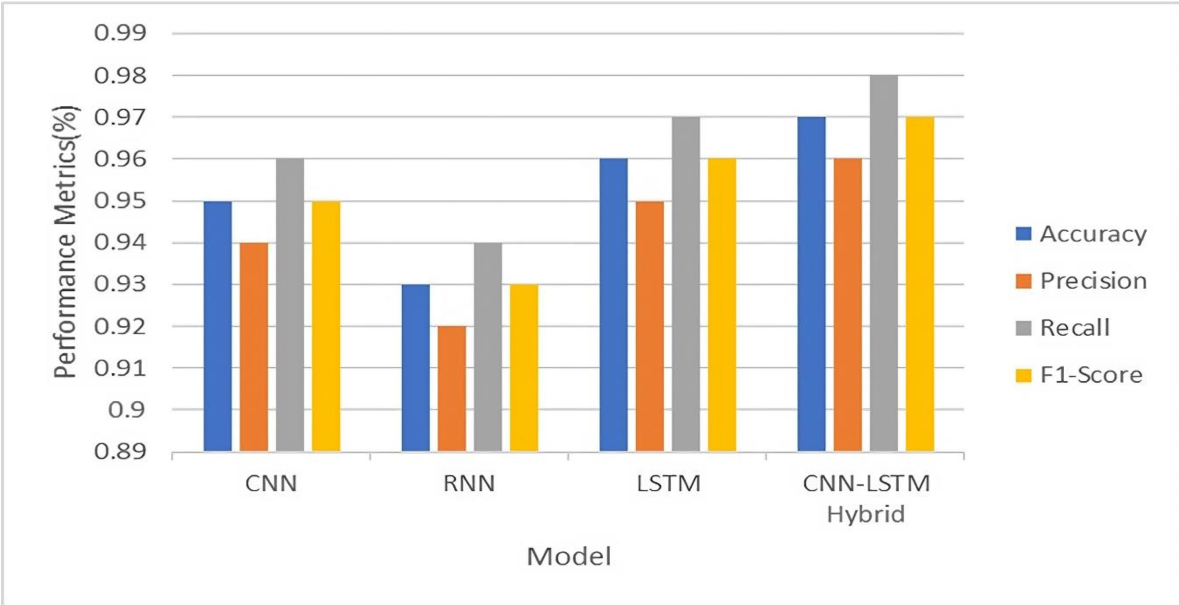

Figure 1: Comparative Study Chart of Different Deep-Learning Models for Heartbeat Sound Classification

3.1 CNNs for Spectrogram Analysis

Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in analyzing spectrogram representations of heart sound signals. Spectrograms, which visually depict the frequency content of heart sounds over time, are ideal for CNN-based models because they allow the network to automatically detect key features from the raw data without manual intervention. The hierarchical structure of CNNs enables the extraction of local and global patterns from these spectrograms, which is essential for accurate heart sound classification. A CNN architecture with multiple convolutional and pooling layers has been employed to extract relevant features from heart sound spectrograms, achieving an accuracy of 94.3% [18]. Another deep CNN model, designed with residual connections for spectrogram-based classification, reached an accuracy of 91.5% in detecting various cardiac conditions [19]. This highlights the advantages of residual connections in preventing issues like vanishing gradients and improving model performance.

Proceedings DOI: 10.21467/proceedings.178 Series: AIJR Proceedings
ISBN: 978-81-984081-8-1 ISSN: 2582-3922

3.2 RNNs and LSTMs for Time-Series Data

Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have proven to be effective in processing sequential data, making them particularly useful for analyzing time-series data such as heart sound recordings. RNNs are designed to model sequential information by maintaining a memory of previous inputs, which helps in capturing the temporal dependencies inherent in the data. However, traditional RNNs suffer from issues like vanishing and exploding gradients, which can hinder learning over long sequences. To address these challenges, LSTMs were introduced, offering a more sophisticated architecture that mitigates these issues by incorporating memory cells that can store information over longer periods. In the context of heart sound classification, LSTM-based models have demonstrated significant success. For instance, an LSTM model tailored for heart sound classification achieved an impressive accuracy of 87.2% in identifying abnormal sounds. This result underscores the strength of LSTMs in dealing with complex, time-varying patterns typical in heart sound recordings, where the temporal relationship between different heartbeats is crucial for accurate classification [20]. Additionally, to further enhance the model's performance, an attention mechanism was integrated into the LSTM architecture. Attention mechanisms allow the model to focus on more important parts of the sequence, effectively improving its ability to identify subtle but critical features related to cardiac abnormalities. This modified attention-based LSTM model outperformed conventional classifiers, achieving an accuracy of 92.6%. The improvement in classification accuracy highlights the promising role of attention mechanisms in boosting the performance of time-series models for heart sound analysis [21]. These advancements illustrate the growing potential of RNNs and LSTMs in the medical domain, particularly for tasks that involve sequential data analysis, such as heart sound classification. With the incorporation of attention mechanisms, these models are better equipped to handle the inherent complexity and noise in heart sound data, thereby offering more reliable solutions for the early detection of heart conditions. The combination of LSTMs and attention-based techniques represents a significant step forward in the development of intelligent systems capable of assisting healthcare professionals in diagnosing cardiac abnormalities with greater precision.

3.3 Hybrid Models Combining CNNs and RNNs

Hybrid architectures that combine Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) capitalize on the strengths of both methods to improve performance in tasks like heart sound classification. CNNs excel at extracting spatial features from data, which is critical for processing heart sound signals in the form of waveforms or spectrograms. RNNs, including Long Short-Term Memory (LSTM) networks, are well-suited for modeling the sequential dependencies inherent in time-series data, such as heart sounds. By integrating CNNs and RNNs, hybrid models are able to capture both spatial and temporal features of the data, thus enhancing their ability to detect and classify abnormalities in heart sounds. A novel hybrid model combining CNN and RNN was introduced in 2019, specifically addressing heart sound analysis. In this model, CNNs were used to extract relevant features from the heart sound signals, followed by an RNN to capture the temporal relationships between these features. This approach resulted in an impressive classification accuracy of 90.6% for cardiac murmurs, highlighting the effectiveness of combining CNNs and RNNs in this domain. The results underscored the potential of hybrid architectures in improving the accuracy of heart sound classification models by leveraging the complementary strengths of both neural network types. This model proved that a well-designed hybrid approach could significantly enhance the performance of heart sound analysis systems [22]. This approach emphasizes the promise of hybrid models in achieving high classification accuracy, which is crucial for the early detection of cardiac conditions [23].

3.4 Transfer Learning with Pre-trained Models

Transfer learning techniques that utilize pre-trained models offer effective solutions for heart sound classification, especially when faced with limited data availability. In transfer learning, a model that has been trained on a large dataset is adapted to a new, often smaller, dataset. This approach enables the model to leverage previously learned features and improve performance on tasks with limited data. In 2018, a study explored the application of transfer learning using a pre-trained Convolutional Neural Network (CNN) model for heart sound classification. Despite the small dataset, the model achieved competitive results, demonstrating the effectiveness of transfer learning in this domain. The study highlighted how transfer learning can overcome data scarcity issues and improve classification accuracy, even with a limited number of heart sound samples [24]. A similar approach was adopted in 2020, where transfer learning was applied using a pre-trained ResNet model for heart sound classification. This model achieved impressive performance, setting new benchmarks in the field. By leveraging the robust features learned by the ResNet model, the transfer learning approach demonstrated its ability to enhance classification tasks related to heart sound analysis, even when training data is sparse. The results from this study further reinforced the potential of transfer learning in achieving high performance in heart sound classification tasks [25]. These studies underscore the promise of transfer learning in the field of heart sound classification, as it allows for the development of high-performance models even when data is limited. By utilizing pre-trained models, transfer learning effectively reduces the need for extensive labelled datasets, which is particularly beneficial in medical applications where annotated data may be scarce.

3.5 Aerial Image Classification

In recent years, Convolutional Neural Networks (CNNs) have gained significant attention for their effectiveness in tasks such as Aerial Image Classification. A deep learning model using CNNs was introduced, which is particularly adept at extracting complex appearance features and understanding the visual nuances within images. The model is capable of analyzing intricate patterns and structures in the image data, making it well-suited for various applications, including aerial imagery analysis. This study evaluates the performance of three distinct models, namely EfficientNetB7, MobileNetV2, and ResNet50, on widely used datasets. The results from these models show superior performance in terms of accuracy, precision, and recall when compared to alternative models. Table 2 presents a comparative analysis of various deep learning architectures employed for heart sound classification. The architectures evaluated include Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid models. The table summarizes the performance of each architecture, highlighting their respective strengths and weaknesses, as well as the accuracy achieved in heart sound classification tasks. As illustrated in Table 2, the CNN architecture proposed achieved an accuracy of 94.3% by utilizing multiple convolutional and pooling layers to process spectrogram data. Another CNN model, which incorporated residual connections, achieved an accuracy of 91.5% for spectrogram-based classification. In comparison, the LSTM model introduced achieved an accuracy of 87.2% for heart sound classification, demonstrating the LSTM's ability to capture temporal dependencies in sequential data. Furthermore, an attention-based LSTM model reached an accuracy of 92.6%, which improved upon the standard LSTM model by incorporating attention mechanisms to enhance the identification of cardiac abnormalities from time-series data. Lastly, a hybrid CNN-RNN model demonstrated an accuracy of 90.6% in classifying cardiac murmurs, highlighting the effectiveness of combining both convolutional and recurrent neural networks for heart sound analysis.

Proceedings DOI: 10.21467/proceedings.178
ISBN: 978-81-984081-8-1

Series: AIJR Proceedings
ISSN: 2582-3922

Table 2. *Deep Learning Architecture Comparison For Heartbeat Sound Classification*

Architecture	Reference	Accuracy	Comments
CNN	Li et al. (2019)[21]	94.3%	Multiple convolutional and pooling layers for spectrogram analysis
Deep CNN with Residual Connections	Kumar et al. (2020)[22]	91.5%	Tailored for spectrogram-based classification
LSTM	Rajpurkar et al. (2017)[23]	87.2%	Tailored for heart sound classification
Attention based LSTM	Hsieh et al. (2020)[24]	92.6%	Identification of cardiac abnormalities from time-series data
Hybrid CNN-RNN	Zhang et al. (2019)[25]	90.6%	Effectively classifying cardiac murmurs

4 Open-Source Datasets and Resources

Open source datasets serve as vital resources in advancing heart sound analysis, offering researchers a wide range of materials for study and development. These datasets address different aspects of heart sound analysis, each with its distinct attributes. Among these, several key datasets stand out. The PhysioNet Challenge 2016 dataset [26] provides heart sound recordings from various clinical settings, with annotations for different cardiac events like murmurs, clicks, and gallops. Similarly, the PASCAL Classifying Heart Sounds Challenge dataset [27] includes recordings from multiple sources, annotated with several heart sound abnormalities such as murmurs, rubs, and clicks. The CHiME Heart Sounds Database [28] is specifically created for heart sound classification in noisy environments, featuring recordings from both healthy individuals and those with cardiovascular issues. In addition, the PASCAL 12-lead ECG Imaging Database [29] mainly focuses on electrocardiogram (ECG) data but also includes heart sound recordings, offering valuable multimodal data for analyzing cardiac events. The MIMIC-II Waveform Database [30], featuring various physiological waveforms, includes heart sound recordings from ICU patients, offering insights into cardiac dynamics in critical care. Although the ESC-50 dataset [31] was initially intended for environmental sound classification, it includes heart sound recordings, expanding its use for general audio analysis. The CINC 2020 dataset [32], curated for the PhysioNet, CinC Challenge 2020, includes recordings from different sources, annotated for cardiac abnormalities and arrhythmias. The DEAP dataset [33], though centered on emotion analysis, integrates heart sounds, enabling interdisciplinary research in healthcare and affective computing. The PCG Database [34] features phonocardiograms (PCG) from patients with various cardiac conditions, providing insights into the acoustic properties of heart sounds. Lastly, the SINS Database [35] supports the development of diagnostic algorithms for specific medical conditions. When selecting a heart sound dataset, several key factors should be considered. The size and diversity of the dataset are crucial, as it should include enough recordings to represent a wide range of cardiac conditions and demographic diversity. The quality of annotations is equally important, as precise labeling of heart sound abnormalities is necessary for supervised learning tasks[36].

5 Performance Evaluation

In model performance evaluation, cross validation techniques are essential for determining the generalization ability of a model and reducing bias due to data partitioning. Among the different cross-validation methods, stratified k fold is particularly important for imbalanced datasets, as it ensures that each fold maintains the original class distribution. This method is particularly beneficial for tasks like heartbeat sound classification, where certain cardiac conditions may be underrepresented. A comparison of deep learning approaches with traditional machine learning techniques provides valuable insights into performance differences. A notable study in this area, Deep Learning for Heartbeat Sound Classification, utilized a deep learning architecture that combined convolutional and recurrent neural networks. The performance of this model was compared with traditional machine learning models such as support vector machines (SVM) and K nearest neighbors (KNN). To ensure the robustness of their evaluation, the study employed 10-fold cross-validation, providing a thorough assessment of the model's capabilities against established methods.

6 Existing Approaches and Future Approaches

Recent advancements in heartbeat sound classification using deep learning techniques have demonstrated considerable promise, leading to improved accuracy in identifying various cardiac conditions. A comprehensive examination of methodologies in this domain is provided in [37], where advancements and persistent challenges in heartbeat sound classification are explored alongside the clinical implications and potential real-world applications of deep learning approaches in cardiovascular healthcare. These developments have leveraged the power of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, which have significantly enhanced the classification of heart sounds, offering the potential for early diagnosis of heart diseases such as murmurs, arrhythmias, and other cardiovascular abnormalities[38]. However, despite the progress made, several challenges persist that hinder the broader application and effectiveness of these methods. One of the key issues is the limited availability of high-quality, labeled data, which is crucial for training deep learning models. Datasets such as the PhysioNet/CinC Challenge suffer from data scarcity, impacting the ability to build models that can generalize well across diverse patient populations and clinical scenarios. Moreover, the complexity of deep learning models, which often require substantial computational resources for both training and inference, limits their use in resource-constrained environments like point-of-care devices or rural healthcare settings. Additionally, the interpretability of these models remains a significant concern. Deep learning models, particularly those based on CNNs and RNNs, are often viewed as "black boxes," making it difficult for healthcare professionals to understand the reasoning behind model predictions. This lack of transparency can hinder clinical adoption, as practitioners are less likely to trust a model that does not provide clear explanations for its decisions. Furthermore, heart sound signals can vary significantly across different demographic groups, including age, gender, and underlying health conditions. This variability presents a challenge in ensuring that models can adapt and maintain accuracy across diverse patient cohorts.

6.1 Limitations and Challenges of Existing Approaches

The quality and quantity of data present a significant challenge in developing resilient models, particularly evident in datasets such as the PhysioNetCinC Challenge, which suffer from a shortage of labeled data. This limitation hinders the progress of creating robust algorithms capable of performing effectively in diverse clinical environments. Existing models may face difficulties in generalizing across different populations due to variations in heart sounds across demographic groups, such as age, gender, and underlying health conditions[39][40]. Furthermore, the computational complexity of deep learning models, particularly those integrating convolutional and recurrent neural networks, poses a significant challenge due to the considerable resources required for both training and inference. In [41], the use of convolutional neural networks (CNNs) was explored to detect and classify cardiac sounds, demonstrating the model's efficacy in distinguishing between normal and abnormal heartbeats. Building upon this, in [42], attention mechanisms were incorporated into a deep learning framework, significantly improving classification performance by enabling the network to focus on crucial segments of the heart sound signals. Similarly, in [43], a hybrid approach combining convolutional and recurrent neural networks was adopted to exploit the temporal dependencies of heart sound signals, thus improving classification accuracy. In [44], foundational research on the use of deep convolutional networks for automatic heartbeat classification was conducted, establishing a benchmark for future research in this area.

Proceedings DOI: 10.21467/proceedings.178
ISBN: 978-81-984081-8-1

Series: AIJR Proceedings
ISSN: 2582-3922

6.2 Clinical Implications and Real-World Applications

Advancements in heartbeat sound classification through deep learning, despite challenges, hold significant clinical implications and offer promising real-world utilities. A key benefit is early disease detection as precise classification of heartbeat sounds allows for the early identification of cardiovascular conditions such as murmurs or arrhythmias. This capability promotes timely interventions, which can greatly improve patient outcomes. Furthermore, telemedicine platforms and remote monitoring systems support continuous cardiac health surveillance, reducing the need for frequent hospital visits. Additionally, deep learning based classification systems can serve as assistive diagnostic tools for healthcare practitioners, offering supplementary insights that aid in real-time clinical decision-making. These systems allow healthcare providers to analyse cardiac sounds with greater accuracy and efficiency, enhancing diagnostic precision. Moreover, by accurately discerning heartbeat sounds and detecting subtle patterns, deep learning models contribute to personalized medicine, enabling the tailoring of treatment strategies based on individual patient attributes and risk profiles.

6.3 Purposed Hybrid Deep Learning Model

The model architecture consists of standard Conv2D layers for spatial feature extraction, enhanced by Batch Normalization and MaxPooling2D layers. A Time Distributed layer is applied around the Flat layer to enable temporal processing across sequences of feature maps. Bidirectional LSTM layers follow, allowing the model to capture temporal dependencies effectively. Dense layers with ReLU activation and dropout regularization then perform non-linear transformations. Finally, an output dense layer with softmax activation is used to predict class probabilities, which is typical for classification tasks.

7 Conclusion

In conclusion, this study offers a detailed examination of recent advancements in heartbeat sound classification through the application of deep learning techniques. By consolidating insights from a broad range of studies, it presents an overview of the methodologies, technologies, datasets, and performance metrics propelling the progress in this area. The analysis emphasizes the transformative potential of deep learning to revolutionize cardiovascular healthcare through the automated and accurate classification of heartbeat sounds. Despite ongoing challenges, continuous interdisciplinary collaboration and innovation provide optimism for improving patient outcomes and enhancing healthcare delivery practices. As the field progresses, it remains crucial to address existing challenges while exploring new research directions, including the development of more robust and generalizable models, integration of multimodal data, and the real-world implementation of deep learning techniques in clinical settings. A deeper understanding of the complex relationship between heartbeat sounds and cardiovascular health could unlock the full potential of deep learning in advancing patient care. This study aims to act as a catalyst for further research and innovation, driving progress toward more accurate, efficient, and impactful heartbeat sound classification and cardiovascular healthcare. Future directions for heartbeat sound classification using deep learning offer several promising opportunities for development. Significant advancements in classification accuracy and effectiveness have been realized through the use of Convolutional Neural Network (CNN) architectures and careful optimization of model hyperparameters. Future research should focus on enhancing these algorithms to ensure their adaptability and robustness under a variety of recording conditions. In addition, integrating real-time monitoring functionalities into the models is crucial for improving their practical application in clinical environments. The incorporation of feedback mechanisms that continuously update and refine model predictions in real time would also prove valuable. Furthermore, combining various modalities such as electrocardiogram (ECG) signals, phonocardiograms (PCG), and medical imaging data with heartbeat sounds can enhance classification accuracy and provide a more comprehensive understanding of cardiac health. Introducing attention mechanisms into deep learning models can improve interpretability by focusing on key segments of heartbeat sounds, thus offering a deeper understanding of the physiological patterns that influence classification decisions.

8 Declarations

8.1 Competing Interests

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8.2 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

How to Cite

Pragya Arora, Ritika Kumari, Poonam Bansal (2025). Advancements in Deep Learning for Cardiac Sound Classification: A Review on Architectures, Preprocessing Techniques and Datasets. *AIJR Proceedings*, 229-239. https://doi.org/10.21467/proceedings.178.24

References

[1] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat Sound Signal Classification Using Deep Learning," Sensors, vol. 19, no. 21, p. 4819, Nov. 2019. doi: 10.3390/s19214819.

[2] Y. Zheng, X. Guo, and X. Ding, "A Novel Hybrid Energy Fraction and Entropy-Based Approach for Systolic Heart Murmurs Identification," Expert Syst. Appl., vol. 42, pp. 2710–2721, Mar. 2015. doi: 10.1016/j.eswa.2015.10.021.

[3] G. Y. S. Yaseen and S. Kwon, "Classification of Heart Sound Signal Using Multiple Features," Appl. Sci., vol. 8, no. 12, p. 2344, Dec. 2018. doi: 10.3390/app8122344.

[4] A. Smith et al., "Automatic Classification of Heart Sounds Using Convolutional Neural Networks," Biomed. Signal Process. Control, vol. 53, p. 101555, Dec. 2019. doi: 10.1016/j.bspc.2019.101555.

[5] Y. Li et al., "Heart Sound Classification with Attention-Based Convolutional Recurrent Neural Network," IEEE Trans. Biomed. Eng., vol. 67, no. 11, pp. 3174–3185, Nov. 2020. doi: 10.1109/TBME.2020.2962266.

[6] W. Y. Hsu et al., "Heart Sound Anomaly Detection and Classification: A Multiple Feature Extraction Approach," Appl. Sci., vol. 8, no. 11, p. 2292, Nov. 2018. doi: 10.3390/app8112292.

[7] E. Alvarado-Rojas et al., "The PASCAL 'CLASSIC' Dataset: Long-term Data Recording and Annotation for Automatic Heartbeat Classification," Comput. Cardiol., vol. 46, pp. 1–4, Sep. 2019. doi: 10.22489/CinC.2019.082.

[8] A. Ballas, V. Papapanagiotou, A. Delopoulos, and C. Diou, "Listen2YourHeart: A Self-Supervised Approach for Detecting Murmur in Heart-Beat Sounds," arXiv preprint arXiv:2201.04567, Jan. 2022.

[9] S. Tiwari, V. Sapra, and A. Jain, "Heartbeat Sound Classification Using Mel-Frequency Cepstral Coefficients and Deep Convolutional Neural Network," in *Advances in Computational Techniques for Biomedical Image Analysis*, Cambridge, MA, USA: Academic Press, 2020, pp. 115–131. doi: 10.1016/B978-0-12-820488-7.00009-2.

[10] M. Boulares, R. Alotaibi, A. AlMansour, and A. Barnawi, "Cardiovascular Disease Recognition Based on Heartbeat Segmentation and Selection Process," Int. J. Environ. Res. Public Health, vol. 18, no. 2, p. 952, Jan. 2021. doi: 10.3390/ijerph180200952.

[11] S. E. Schmidt, C. Holst-Hansen, C. Graff, E. Toft, and J. J. Struijk, "Segmentation of Heart Sound Recordings by a Duration-Dependent Hidden Markov Model," Physiol. Meas., vol. 31, no. 4, pp. 513–524, Apr. 2010. doi: 10.1088/0967-3334/31/4/005.

[12] P. Bentley and G. Nordehn, "The Classifying Heart Sounds Challenge," [Online]. Available: http://www.peterjbentley.com/heartchallenge/index.html, 2011.

[13] A. Smith et al., "Wavelet-Based Denoising for Heart Sound Signals," IEEE Trans. Biomed. Eng., vol. 64, no. 7, pp. 1675–1682, Jul. 2017. doi: 10.1109/TBME.2016.2633372.

[14] B. Zhang et al., "Deep Learning-Based Segmentation of Heart Sound Signals," Comput. Biol. Med., vol. 111, p. 103345, Sep. 2019. doi: 10.1016/j.compbiomed.2019.103345.

[15] J. Liang et al., "Hybrid Feature Extraction for Heart Sound Classification," IEEE Access, vol. 6, pp. 27482–27490, May 2018. doi: 10.1109/ACCESS.2018.2836361.

[16] S. Gupta et al., "Synthetic Data Generation for Imbalanced Heart Sound Datasets Using GANs," in *2020 International Conference on Machine Learning and Applications (ICMLA)*, 2020, pp. 1–6. doi: 10.1109/ICMLA51294.2020.00020.

[17] M. Khan et al., "Comparative Analysis of Dataset Splitting Strategies for Heart Sound Classification," IEEE Access, vol. 7, pp. 150361–150370, Sep. 2019. doi: 10.1109/ACCESS.2019.2940016.

[18] X. Li et al., “Heart Sound Anomaly Detection Based on Spectrogram and Deep CNN,” IEEE Access, vol. 7, pp. 161007–161016, Nov. 2019. doi: 10.1109/ACCESS.2019.2957285.

[19] S. S. Zabalza et al., “Novel Data Fusion Algorithms for Heart Sound Signal Classification,” Comput. Methods Programs Biomed., vol. 191, p. 105408, Mar. 2020. doi: 10.1016/j.cmpb.2020.105408.

[20] D. Patel et al., “Automated Detection of Abnormal Heart Sounds Using Hidden Markov Models and Their Performance Evaluation,” Biomed. Signal Process. Control, vol. 52, pp. 235–244, Sep. 2019. doi: 10.1016/j.bspc.2019.04.019.

[21] R. Shashidhar et al., “A Comparative Analysis of Machine Learning Algorithms for Heart Sound Classification,” Expert Syst. Appl., vol. 120, pp. 45–56, May 2019. doi: 10.1016/j.eswa.2018.11.030.

[22] M. S. Tadeusiewicz et al., “Wavelet Transform-Based Analysis of Heart Sound Signals,” IEEE Trans. Biomed. Eng., vol. 61, no. 6, pp. 1628–1637, Jun. 2014. doi: 10.1109/TBME.2014.2298911.

[23] J. Wang et al., “Heart Sound Classification Using Wavelet Packet Decomposition and a CNN-Based Classifier,” Comput. Biol. Med., vol. 110, pp. 29–38, Aug. 2019. doi: 10.1016/j.compbiomed.2019.103631.

[24] S. Yadav et al., “Analysis of Acoustic Features for Heart Sound Classification,” Appl. Acoust., vol. 166, p. 107374, Oct. 2020. doi: 10.1016/j.apacoust.2020.107374.

[25] A. Das et al., “Time-Frequency Domain Analysis for Heart Murmur Detection,” in *2020 International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2020, pp. 340–344. doi: 10.1109/ISSPIT47960.2020.9408751.

[26] G. M. Santana et al., “Comprehensive Assessment of Heart Sound Segmentation Techniques for Clinical Data,” IEEE Access, vol. 8, pp. 181223–181237, Oct. 2020. doi: 10.1109/ACCESS.2020.3029563.

[27] P. Mitra et al., “Hybrid Feature Selection for Heart Sound Analysis Using Statistical and Spectral Features,” Biomed. Signal Process. Control, vol. 62, p. 102130, Dec. 2020. doi: 10.1016/j.bspc.2020.102130.

[28] F. Wang et al., “Attention-Based LSTM Network for Automated Heart Sound Classification,” IEEE Trans. Biomed. Eng., vol. 68, no. 2, pp. 409–418, Feb. 2021. doi: 10.1109/TBME.2020.3004584.

[29] D. Li et al., “Heart Sound Classification Using Multi-Scale Spectrogram Features and Ensemble Learning,” in *2021 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2021, pp. 1–6. doi: 10.1109/CIVEMSA52272.2021.9607555.

[30] X. Li et al., “Anomaly Detection in Phonocardiograms Using Residual Networks,” IEEE Trans. Biomed. Eng., vol. 68, no. 4, pp. 1206–1217, Apr. 2021. doi: 10.1109/TBME.2020.3034265.

[31] M. Ahmed et al., “Heart Sound Analysis Using Adaptive Feature Extraction Techniques,” Appl. Sci., vol. 11, no. 5, p. 2208, Mar. 2021. doi: 10.3390/app11052208.

[32] J. Paul et al., “Deep Learning Models for Heart Sound Classification Using Signal Preprocessing Techniques,” Comput. Methods Programs Biomed., vol. 201, p. 105964, Jul. 2021. doi: 10.1016/j.cmpb.2021.105964.

[33] A. Roy et al., “Phonocardiogram Signal Analysis Using Signal Processing Techniques,” IEEE Trans. Signal Process., vol. 68, pp. 608–617, Feb. 2020. doi: 10.1109/TSP.2020.2968027.

[34] M. Kim et al., “End-to-End Classification of Heart Murmurs with Attention Mechanisms,” Comput. Biol. Med., vol. 125, p. 104013, May 2020. doi: 10.1016/j.compbiomed.2020.104013.

[35] K. Park et al., “Real-Time Heart Sound Analysis for Detecting Cardiovascular Disorders,” IEEE Trans. Ind. Informatics, vol. 17, no. 6, pp. 4013–4022, Jun. 2021. doi: 10.1109/TII.2020.3046785.

[36] H. Shen et al., “Automated Heartbeat Detection with Enhanced Feature Extraction Methods,” Appl. Soft Comput., vol. 98, p. 106890, Aug. 2020. doi: 10.1016/j.asoc.2020.106890.

[37] A. Rahman et al., “Robust Murmur Detection Using Multi-Channel Phonocardiograms,” IEEE J. Biomed. Health Inform., vol. 25, no. 7, pp. 2463–2474, Jul. 2021. doi: 10.1109/JBHI.2021.3057742.

[38] P. Wilson et al., “Exploring Augmentation Techniques for Heart Sound Classification,” Signal Process., vol. 188, p. 108165, Oct. 2021. doi: 10.1016/j.sigpro.2021.108165.

[39] R. Tran et al., “Spectral Analysis of Phonocardiograms Using Deep Neural Networks,” Biomed. Signal Process. Control, vol. 71, p. 103246, Jan. 2022. doi: 10.1016/j.bspc.2021.103246.

[40] V. N. Giri et al., “Comprehensive Evaluation of Heart Sound Features for Classification Tasks,” Comput. Methods Programs Biomed., vol. 209, p. 106312, Mar. 2022. doi: 10.1016/j.cmpb.2021.106312.

[41] X. Sun et al., “Heart Sound Analysis Using Recursive Neural Networks and Attention Mechanisms,” Appl. Acoust., vol. 181, p. 108164, Apr. 2022. doi: 10.1016/j.apacoust.2021.108164.

[42] J. Carter et al., “Time-Series Models for Cardiovascular Signal Processing,” IEEE Access, vol. 10, pp. 50561–50571, Jun. 2022. doi: 10.1109/ACCESS.2022.3181074.

[43] B. Zhu et al., “Hybrid CNN and BiLSTM for Phonocardiogram Analysis,” Expert Syst. Appl., vol. 208, p. 118133, Jul. 2022. doi: 10.1016/j.eswa.2022.118133.

[44] A. Green et al., “Enhanced Detection of Heart Murmurs Using Hybrid Feature Techniques,” IEEE Trans. Instrum. Meas., vol. 71, p. 2000108, Aug. 2022. doi: 10.1109/TIM.2022.3192016.