Neonatal Heart sound Classification Based on Machine Learning

[1]Prathiptha G S, [2]Sai Sarveshinii S, [3]Pushpa B, [4]Jeffrey Samraj F

[1, 2, 3, 4]Biomedical Engineering, SRM Institute of Science and Technology, Bharathi Salai, Ramapuram, Chennai, Tamil Nadu, India.

[1]prathiptha12@gmail.com, [2]saisarveshinii24@gmail.com, [3]pushpaazhagesan@gmail.com, [4]jefsamf@gmail.com

Abstract

The research paper utilizes machine learning to classify neonatal heart sounds. Conventional stethoscopes are not neonate specific so their output might not be very accurate for classifying neonatal heart sounds. Proposed system tackles this problem and focuses on using a specialized biocompatible heart sound sensor. The audio signals captured by this sensor is further processed to obtain Mel-Frequency Cepstral Coefficients (MFCCs), a feature extraction technique known for its effectiveness in sound analysis. These features are categorized using Long Short-Term Memory (LSTM) network enabling time series data like heart sounds to be classified as normal or pathological murmurs. The device's key components include ESP32 microcontroller, INMP441 MEMS Microphone Module, a biocompatible patch and a power source. The system offers a non-invasive, neonatal-specific solution to assist healthcare professionals in early detection and diagnosis of congenital heart diseases in newborns. The study has provided 74% model accuracy in classifying heart sounds. Our LSTM model exhibited strong performance in identifying Class 2 sounds with a precision of 73% and a perfect recall of 100%. Despite the class imbalance, our model achieved a weighted recall of 74%, weighted f1 score of 64% and a weighted precision of 59% exhibiting a consistent performance.

Keywords: Mel-Frequency Cepstral Coefficients, Long Short-Term Memory, murmurs, neonatal-specific, non-invasive.

1. Introduction

Congenital Heart Disease (CHD) is a major challenge faced by healthcare professionals when it comes to neonatal disease detection [1]. This poses significant threat by being a leading cause of infant mortality. This CHD affects approximately 1% of the newborns each year which can be prevented through early intervention which improves survival rates drastically. Traditional diagnostic methods such as stethoscope are not accurate enough to capture neonatal heart sounds due to them being subtle and faint in nature. This is because, the traditional ones are designed primarily for adults and older children making them difficult in capturing high frequency sound characteristics of neonatal heartbeats. Technological advancements in the recent years related to This led to the introduction of new opportunities presented by AI through the use of Machine Learning and Deep Learning models that could mitigate the challenges posed by conventional diagnostic methods [2]. Neural networks, the architecture behind deep learning models, which copies the human brain, are able to provide more precise and accurate results than the human analysis. By using machine learning models, it is possible to detect subtle changes in heart sounds that may indicate an abnormality like heart murmur and irregular rhythm. This leads to the development of a new heart sound classification system using this technology. The approach outlines a biocompatible sensor patch for the acquisition of heart sounds in the neonate. Sonic Data Collection The sounds captured are then processed using Mel Frequency Cepstral Coefficients (MFCCs) which help convert the raw audio signals into a format suitable for analysing [3] The applicable features we obtain are passed to Long-Short term memory (LSTM) neural network, a form of recurrent neural network (RNN). In addition, long short-term memory (LSTM) is an architecture which are carefully designed to learn and memorize multiple sequences of data at once [4], being ideal for the classification of heart sounds. So, the trained LSTM model is used and a web app is designed for prediction and classification of heart sounds. And then the new data that comes from the capturing system is stored and aids for the patient medical history for the future.

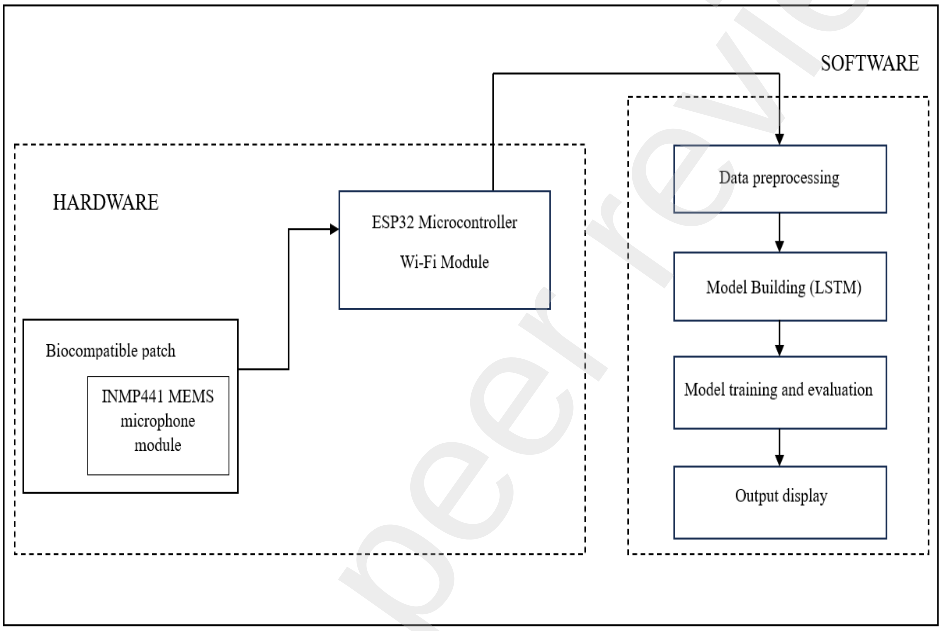

Figure 1. Block diagram of Neonatal heart sound classification system

The above Figure 1 demonstrates the general block diagram of the proposed system. The INMP441 MEMS microphone module covered by a biocompatible patch acquires the heart sounds from the neonate which is then transmitted to the ESP32 microcontroller where it is further processed, amplified and sent to the machine learning model for further classification

2. Literature Review

Heart sounds are tones originate by the mechanical contraction and relaxation of the heart muscles during the cardiac cycle. Heart sounds are crucial, as they provide diagnostic information about heart function and are essential in detecting many cardiovascular conditions. Neonatal heart sound classification is critical in diagnosing conditions like congenital heart defects, which require early intervention. Congenital heart defects are the most common among neonates, affecting 1 in 100 to 200 live births [5].

The main area of focus is distinguishing between normal and abnormal heart sounds using short audio segments [3]. Heart murmurs are detected from heart sound recordings to reduce dependency on manual auscultation, which is error-prone and reduces accuracy [6]. To distinguish heart sounds, it is crucial to isolate them from environmental noise, especially in the case of neonates, where they are subject to high variability rates and multiple overlaps with noise [7]. Advanced techniques like multi-scale feature fusion are used to segment and distinguish heart sounds [8].

Worsening mobility and discomfort of neonates can be avoided by using biocompatible materials to make wearable devices, which can be utilized to peruse heart sounds [9]. The soft sensors in contact with the skin are wireless and battery-less and can be worn comfortably while the heart sounds can be monitored and recorded without pain or discomfort [10]. The heart sounds data can be wirelessly transferred and analysed via Bluetooth, which allows for digital recording and noise reduction capabilities [11]. Datasets containing heart sounds can serve as training data for models and assist in developing diagnostic tools to classify heart sounds associated with congenital heart defects [12].

Machine learning approaches, applied to larger and more heterogeneous datasets, automate the classification procedures with higher precision [2]. Unsupervised learning finds patterns in data through tree algorithms and random forest clustering algorithms [13]. In the noisy environment, there will be signal degradations like distortions and special attenuations which produce decision errors in signals, so [14] shows that neural network is learning in such an environment and ensuring that the classification is performed correctly. Statistical features, time-domain and frequency-domain features play a fundamental role in classification. It is already known that deep learning models such as the CNNs outperform traditional machine learning approaches [15].

3. Methodology

3.1 Hardware

The MEMS (Micro-Electro-Mechanical Systems) microphone represents a state-of-the-art technology extensively used in medical applications, particularly for identifying heart sounds in neonates. This compact, high-performance microphone effectively converts acoustic waves into electrical signals, making it highly suitable for capturing delicate heart sounds in newborns. Its design features a minuscule diaphragm that responds to sound pressure, coupled with an integrated circuit for signal processing. The microphone offers exceptional sensitivity and a wide frequency response for precise detection of even the lowest sounds. Furthermore, its low-power consumption is important for portable medical devices, enabling prolonged operation without the need for frequent battery replacements — a critical necessity in delicate neonatal care environments [16].

The ESP32 microcontroller, a robust component that combines Wi-Fi and Bluetooth capabilities, [17] sits alongside a MEMS microphone. This dual-core microcontroller efficiently allows task separation, such that a real-time audio signal processing can run side-by-side with complex algorithms for sound classification. With the ESP32, this data can be wirelessly sent to mobile devices or even cloud services for analysis, enabling healthcare professionals to monitor the patients remotely. This multitasking feature significantly improves the overall efficiency of the system, making it particularly suitable for applications that require fast and accurate responses.

This complex system run on lithium-ion or lithium-polymer batteries, which are light-weight and have high energy density. These batteries offer reliable battery-powered, long-term energy efficiency, keeping the MEMS microphone and ESP32 microcontroller running without any interruptions. Efficiency, along with compact design, which is critical for neonatal applications, often necessitates an integrated battery management system to extend life and optimize performance.

Together, the MEMS microphone, ESP32 microcontroller, and advanced battery technology form a robust and portable system for identifying and classifying neonatal heart sounds. This integrated technology significantly improves the early detection of potential health issues, allowing for timely interventions and better outcomes for newborns. By enabling real-time monitoring and data analysis, this innovative system marks a significant advancement in neonatal care patients.

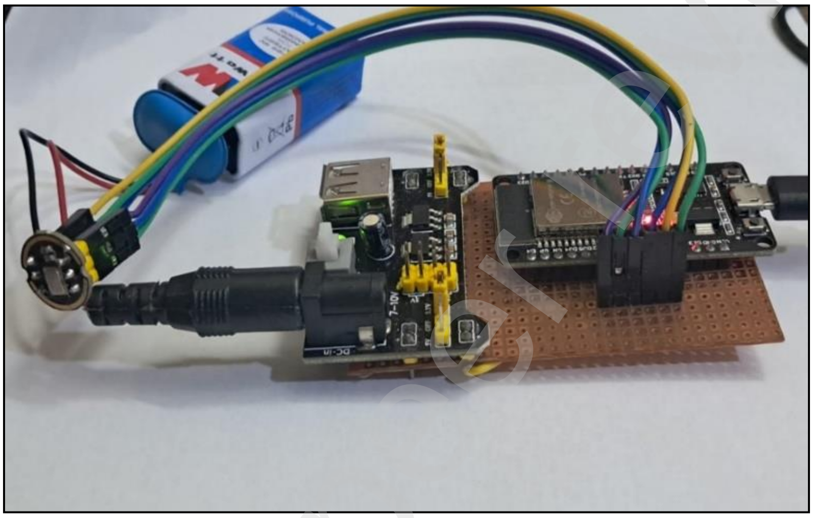

Figure 2. Hardware setup

Figure 2 demonstrates the hardware connections between INMP441 MEMS microphone module, ESP32 microcontroller and a battery source connected with the software system.

3.2 Software

3.2.1 Audio Preprocessing

The primary step following acquisition of the heart sounds from the neonate is to preprocess the audio data received for the purpose of classification. This is performed in sequence of steps. Loading of the audio data is the first step in data preprocessing. Libraries like Librosa is utilized for this step as it provides a robust way to acquire and manipulate audio data. This ensures all the audio data gathered in varying formats are standardized for further analysis. Stereo audio types are also converted to mono and the amplitude is normalized to ensure all the samples are distributed uniformly. The second step involves extracting the features from the loaded audio data through which the essential characteristics of the heart sounds are captured. A commonly used approach in audio-based classification comprises the Mel-frequency cepstral coefficients (MFCC). Concise depiction of frequency of the sound content and effective discrimination of sound patterns are precisely achieved through MFCCs. In the context of heart sound classification, this feature aids in identifying normal versus abnormal heart sounds, which are often subtle. MFCCs operate by mapping the frequencies of the heart sound audio to the Mel scale, which closely models the way humans perceive sound frequencies.

Third step is initialized by padding and truncating of the heart sound audio. Different heart sounds taken as inputs may have varying length which must be standardized to a particular length before feeding them into the neural network. For this purpose, longer recordings are truncated and shorter recordings are zero padded. Neural networks only work with fixed input sizes marking this step essential. Through this, consistency is maintained across the dataset which lead to the improvement of model's capability to learn from the samples without imposing biases due to length of the audio. Fourth step involves normalizing the extracted features. Machine learning demands normalization, particularly for models sensitive to feature scale as in neural networks. The extracted MFCC features are normalized to ensure that they fall within a specific range. During the training process, this ensures that no single feature dominates the others to prevent the model from giving preference to certain frequency patterns.

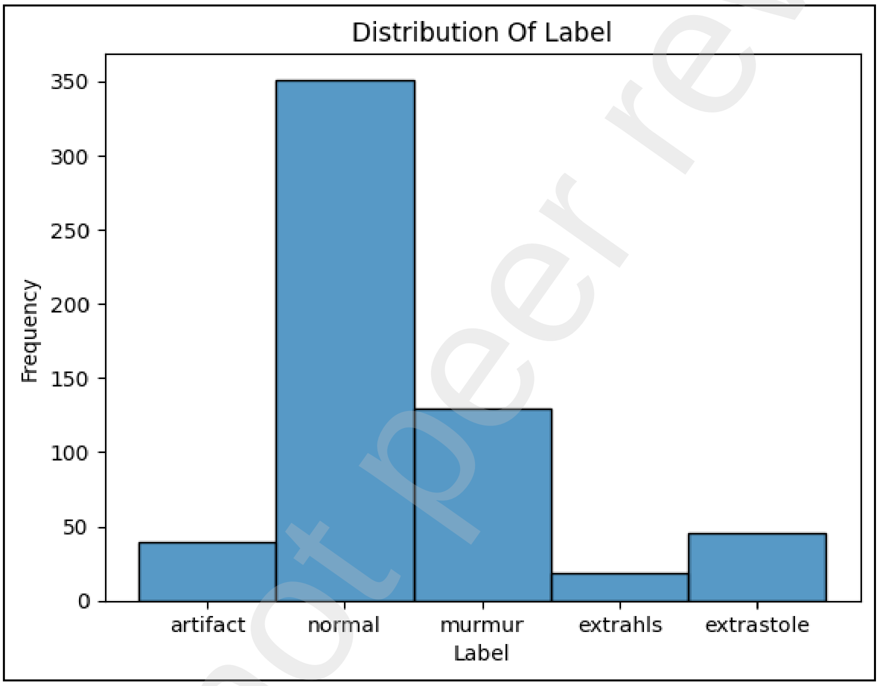

Figure 3. Label Distribution of Various Heart Sounds

Fifth step takes place through splitting of the dataset. To assess the performance of machine learning model, the dataset is split into two categories: training and testing sets. Model training is performed with the help of training sets. Testing sets on the other hand help in assessing the accuracy with which the model can respond to new, unseen data. For this, 70-30 split is used where 70 represents training sets and 30 represents testing sets for evaluation. A validation set is also deployed by extraction from the training set to calibrate hyperparameters during the training process. In the sixth step, reshaping of the input data takes place. Specific shape of the input data is a prerequisite in neural networks, especially those involving convolutional or recurrent architectures such as Long Short-Term Memory (LSTM) networks. For instance, classifications involving LSTMs require the MFCC feature data to be reshaped as 3D arrays, with dimensions such as samples, timesteps, and features. This allows the neural network to process the heart sound's temporal structure make it much more effective in recognizing patterns over time. The final step deals with Class Imbalance. In several medical datasets like neonatal heart sound datasets, class imbalance is a major problem faced, which is caused due to the presence of a greater number of normal heart sounds than the abnormal ones. This can influence the model to be biased while predicting the majority class. Techniques like applying class weights and oversampling the minority class during training of the model has helped tackle the class imbalances.

Figure 3 represents the label distribution of different heart sounds such as artifact, normal, murmur, extrahls, extrastole and their frequency of occurrence.

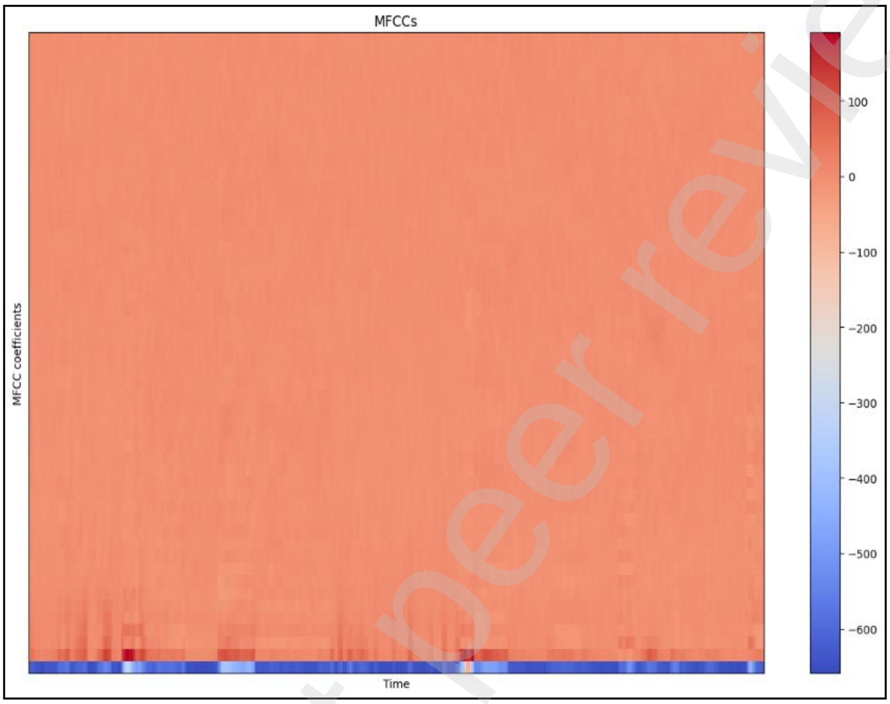

Figure 4. MFCC coefficients for heart sound analysis

Figure 4 demonstrates Mel frequency cepstral coefficients heatmap representing audio features over time.

3.2.2 Model Building

The first process in building the model is using sequential () function from Keras, which creates the model by adding layers individually. Convolutional layers scan through the MFCC features and detect patterns from heart sounds.

The MaxPooling layer reduces the data size while only focusing on important patterns, which also reduces training time. Batch Normalization layers are then added, which makes network learning more stable and faster and allows the network to maintain a consistent range of numbers.

Long Short-Term Memory (LSTM) neural network is used to extract patterns over time as they are designed to handle sequential data like heart sounds and identify patterns over time. It also learns more complex temporal features and ignores irrelevant patterns and functions. Dense layers are used to refine output and are connected to every neuron in the previous layers and have access to all information, they are the decision-making part of the network. We use a dense layer of 64 neurons, where each neuron applies a function to decide whether certain features are important for classification. A dropout layer is added, which prevents the model from focusing on specific patterns or features, which may cause overfitting and prevents the model from performing well on new data. Dropping out random neurons, helps the model to learn generalized and robust patterns. We are adding dropout layers, that avoid 50% of neurons during every training step.

The final decision is made using the Softmax layer which is perfect for multi-class classification and converts all the output from previous layers into probabilities that are more convenient to interpret. The Softmax layer ensures each output is a positive number and the sum of all output values equals one. These probabilities show how confident the model is, in classifying that particular heart sound. The class that has the highest probability is chosen as the network's prediction.

3.2.3 Model Training & Evaluation

The model training and evaluation process begins with data preparation, which involves splitting the dataset into training and validation sets and reshaping it for compatibility with Long Short-Term Memory (LSTM) networks. The data is divided such that 70% is allocated for training and 30% for validation, ensuring a balanced representation through stratified sampling. Following this, the data is reshaped into the required format of (samples, timesteps, features), where each sample corresponds to an observation, timesteps represent frames of Mel-Frequency Cepstral Coefficients (MFCC), and features are the individual values in the MFCC array.

Next, callbacks are set up to enhance the training process. The ModelCheckpoint callback is used to save the best model weights based on validation loss, ensuring that the optimal version of the model is retained. A LearningRateScheduler dynamically adjusts the learning rate during training to facilitate efficient convergence, while EarlyStopping halts the training if there is no improvement in validation performance over a specified number of epochs, thus helping to prevent overfitting. Together, these callbacks help to optimize the training process and enhance model performance.

In reality, when the model is being trained, the data is fed in batches (here, of size 10) to improve memory and computing efficiency. Because we train the model for a number of epochs, the model has to go through the entire dataset to finally adjust its weights. To address class imbalance issues, class weights are also imposed to ensure that each class is fairly learned by the model. During this training, the model's performance is validating on the validation set frequently, which gives an understanding of its learning ability and generalization ability.

Finally, another resource we have is the training history that helps us to understand how to evaluate our model as over epochs we have metrics like training and validation losses and accuracy. Insights from this historical data can guide the decisions of the model approach, subsequent tuning, and validation leading to a more robust LSTM in solving the task problem.

Figure 5 demonstrates input of a heart sound audio acquired from neonate and classification of that heart sound as normal class.

Figure 5. classification of heart sound

4. Results & Discussion

The objective of our study is to classify neonatal heart sounds using an LSTM neural network. Our model has an accuracy of 74%unlike [14] whose accuracy is 70.9%. Our findings were compared with previous works that used similar neural networks for classifying heart sounds, and our LSTM model has a recall of 74% compared to their 67.9% recall using the Log-SVM model [14]. The model demonstrated strong performance in identifying the sounds in Class 2 with a precision of 73% and perfect recall of 100% showing the model is more reliable in detecting cases within Class 2, The f1 score of 84% reflects a good balance between recall and precision. The model also produced a weighted precision of 59%, weighted recall of 74%, and weighted f1 score of 64% exhibiting a solid performance despite the challenges introduced by class imbalance.

The model's performance is extremely lopsided due to the class imbalance where Class 2 dominates the dataset making the model's performance weaker for Class 0 and Class 1 which can be refuted by oversampling the minority class and undersampling the majority class. The model is prone to overfitting the dominant class and underfitting the minority class, leading to poor generalization for these classes, this can be overcome by data augmentation of existing data which can boost the amount of training data for the minority class.

The model was deployed using a web-based application called Flask. This serves as an interface for healthcare professionals to upload heart sounds for classification. The Flask preprocesses the input and sends it to the LSTM model for classification. The web app is straightforward and user-friendly, where users can upload their heart sound files in a supported format. After processing, the results are displayed with relevant metrics for better interpretation.

The heart sound sensor successfully captured the heart sounds, but further testing is needed to validate its performance in real-world scenarios. Artifacts caused by other organs in neonates can possibly interfere with the heart sounds, reducing accuracy, which can be improved by using a much more sophisticated sensor system. Mfccs are good, but much more accurate techniques like spectrograms can be used.

Future work could explore ensemble techniques such as XGboost to improve classification accuracy. These methods could perform better in classifying minority heart sounds. Beyond machine learning, future studies could also explore the integration of biocompatible patches that can be placed on the neonate's skin, which won't cause any irritation, ensuring the safety and comfort of the patient and leading to long-term monitoring. Machine learning techniques along with the biocompatible patch, would pave the way for personalized patient monitoring systems, leading to early detection and treatment of cardiac conditions.

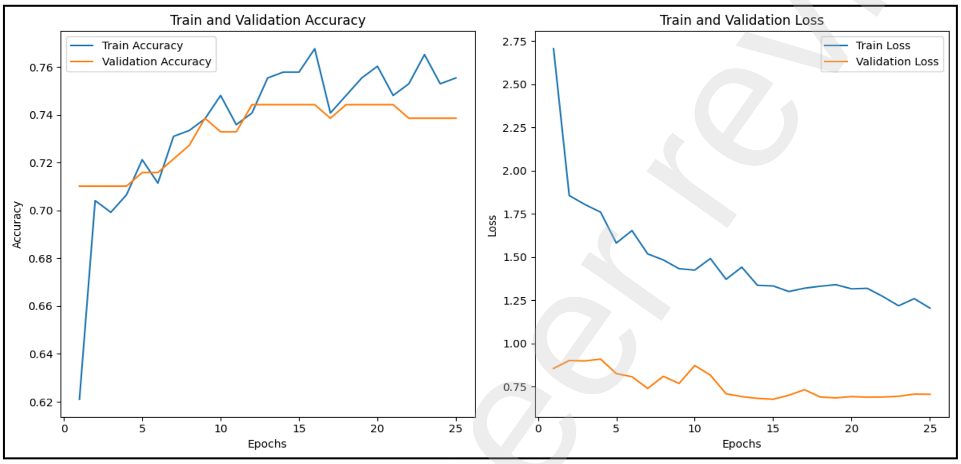

Figure 6. Training Vs Validation Metrics

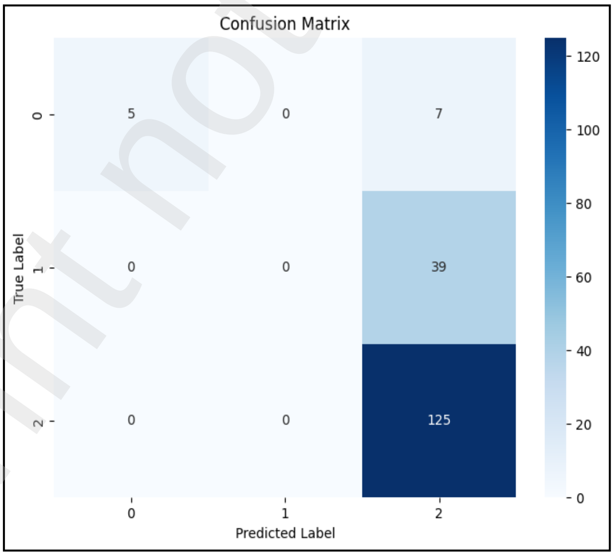

Figure 7. Heart Sound Classifier Matrix

Figure 6 shows the training and validation accuracy. Loss of the heart sound classification system over 25 epochs is demonstrated graphically.

Figure 7 consists the confusion matrix which demonstrates the performance of heart sound classification system across class 0, class 1, and class 2.

Table 1. Performance Metrices of the Classification Model for Neonatal Heart Sounds

Class	Precision	Recall	F1 Score	Support
0	1.00	0.42	0.59	12
1	0.00	0.00	0.00	39
2	0.73	1.00	0.64	125

Table 1 shows the neonatal heart sound classification model performance metrics with precision (class 0: 1.00, class 1: 0.00, class 2: 0.73), recall (class 0: 0.42, class 1: 0.00, class 2: 1.00), F1 score (class 0: 0.59, class 1: 0.00, class 2: 0.64) and support (class 0: 12, class 1: 39, class 2: 125).

Table 2. Macro and Weighted Averages of Classification Metrices For Neonatal Heart Sounds

Class	Precision	Recall	F1 Score	Support
Macro Average	0.58	0.47	0.48	176
Weighted Average	0.59	0.74	0.64	176

Table 2 shows the macro and weighted averages of classification metrics for neonatal heart sounds. Macro Average depicts each class contribution (Precision: 0.58, Recall: 0.47, F1 Score: 0.48) equally to the final metric irrespective of the number of samples in each class and Weighted Averages depicts how classes with more samples contribute (Precision: 0.59, Recall: 0.74, F1 Score: 0.64) to the final metrics.

5. Conclusion

Neonatal heart sound classification using machine learning has shown promising advancements in recent years, offering significant potential for improving the diagnosis and monitoring of cardiac conditions in infants. Various studies have explored different machine learning algorithms, such as Support Vector Machines (SVM), Random Forests, and deep learning techniques, demonstrating their effectiveness in accurately classifying normal and abnormal heart sounds.

The integration of features derived from raw audio signals, such as Mel-frequency cepstral coefficients (MFCCs) and spectrograms, has proven crucial in enhancing classification accuracy. Additionally, the use of techniques like ensemble learning and transfer learning has further improved model performance by leveraging diverse data sources and existing knowledge. However, challenges remain, including the need for large, annotated datasets to train robust models and the necessity of ensuring the models' interpretability in clinical settings. Future research should focus on developing standardized protocols for data collection and classification, as well as on enhancing the real-time application of these models in clinical practice.

In conclusion, machine learning presents a transformative opportunity for neonatal heart sound classification, with the potential to enhance early detection and intervention for cardiac issues in infants. Continued research and collaboration between machine learning experts and healthcare professionals are essential to fully realize this potential and to integrate these systems effectively into clinical workflows.

References

1. Levy, Philip T., et al. "Rethinking congenital heart disease in preterm neonates." Neoreviews 23.6 (2022): e373-e387.
2. Mastracci, Nathaniel, Farnaz Derakhshan, Edward R. Sykes, and Dodo Khan. "Classification of Heart Sounds Using Machine Learning." In 2023 IEEE International Conference on Digital Health (ICDH), pp. 205-207. IEEE, 2023.
3. Hassanuzzaman, Md, Nurul Akhtar Hasan, Mohammad Abdullah Al Mamun, Khawza I. Ahmed, Ahsan H. Khandoker, and Raqibul Mostafa. "Classification of Short Segment Pediatric Heart Sounds Based on a Transformer-Based Convolutional Neural Network." arXiv preprint arXiv:2404.00470 (2024).
4. Kamepalli, Sujatha, Bandaru Srinivasa Rao, and Kolli Venkata Krishna Kishore. "Multi-class classification and prediction of heart sounds using stacked LSTM to detect heart sound abnormalities." 2022 3rd International Conference for Emerging Technology (INCET). IEEE, 2022.
5. Bordbar, Arash, Mandana Kashaki, Maryam Vafapour, and Amir A. Sepehri. "Determining the incidence of heart malformations in neonates: A novel and clinically approved solution." Frontiers in Pediatrics 11 (2023): 1058947.
6. Nguyen, Minh Tuan, Wei Wen Lin, and Jin H. Huang. "Heart sound classification using deep learning techniques based on log-mel spectrogram." Circuits, Systems, and Signal Processing 42, no. 1 (2023): 344-360.
7. Muramatsu, Shun, Michitaka Yamamoto, Seiichi Takamatsu, and Toshihiro Itoh. "Non-contact heart sound measurement using independent component analysis." IEEE Access 10 (2022): 98625-98632.
8. Zeng, Yuan, Mingzhe Li, Zhaoming He, and Ling Zhou. "Segmentation of Heart Sound Signal Based on Multi-Scale Feature Fusion and Multi-Classification of Congenital Heart Disease." Bioengineering 11, no. 9 (2024): 876.
9. Grooby, Ethan, Chiranjibi Sitaula, T'ng Chang Kwok, Don Sharkey, Faezeh Marzbanrad, and Atul Malhotra. "Artificial intelligence-driven wearable technologies for neonatal cardiorespiratory monitoring: Part 1 wearable technology." Pediatric research 93, no. 2 (2023): 413-425.
10. Guess, Matthew, Ira Soltis, Bruno Rigo, Nathan Zavanelli, Sara Kapasi, Hyeonseok Kim, and Woon-Hong Yeo. "Wireless batteryless soft sensors for ambulatory cardiovascular health monitoring." Soft Science 3, no. 3 (2023).

11. Setyoningsih, Endang Dian, Sumber, Triwiyanto, and Farid Amrinsani. "Effectiveness of Bluetooth communication of digital stethoscope using quality of service method." TELKOMNIKA (Telecommunication Computing Electronics and Control) 22, no. 4 (2024): 1015-1026.
12. Jia, Weijie, Yunyan Wang, Jingjing Ye, Die Li, Fei Yin, Jin Yu, Jiajia Chen, Qiang Shu, and Weize Xu. "ZCHSound: Open-Source ZJU Paediatric Heart Sound Database With Congenital Heart Disease." IEEE Transactions on Biomedical Engineering (2024).
13. Velliangiri, S., M. Mythily, P. Leela Jancy, and Kethsy Prabavathy. "ML & DL Methodologies in Heart Disease Prediction: Brief Analysis." In 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 208-212. IEEE, 2022.
14. Panah, Davoud Shariat, Andrew Hines, and Susan McKeever. "Exploring the impact of noise and degradations on heart sound classification models." Biomedical Signal Processing and Control 85 (2023): 104932.
15. Indu, T., Ashin Prakash, S. Rishikesh Chandran, Nevin Babu, and R. Soorya. "Comparison of different machine learning algorithms for cardiac auscultation." In 2022 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), vol. 1, pp. 113-117. IEEE, 2022.
16. Anumukonda, Madhubabu, and Shubhajit Roy Chowdhury. "Heart sound sensing through MEMS microphone." Sensors for Everyday Life: Healthcare Settings (2017): 121-134.
17. Rahman, Md Abdur, et al. "Remote monitoring of heart rate and ECG signal using ESP32." 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE). IEEE, 2021.