International Journal of Research in Engineering, Science and Management
Volume 4, Issue 10, October 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Deep Learning Based Classification of Heart Diseases from Heart Sounds

Alampally Naveen[1*], Parigi Sai Teja Reddy[2], Thenmozhi Thangavel[3]
[1,2]*School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*
[3]*Associate Professor Grade-1, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*

***Abstract*: Heart problems have become one of the prevalent impacts on human health. Heart sound identification techniques play an essential role in predicting heart attacks as a non-invasive assistant diagnostic tool. The phonocardiogram (PCG) signal, a digital recording of heart sounds, was analyzed in this paper and classified into four classes: Normal signal, Murmur signal, Artifact signal, and Extra systole signal. The following are the primary considerations: the relationship between heart sounds and cardiovascular diseases: Denoising, segmentation, feature extraction, and classification are some of the core technologies used in the processing and analysis of heart sound signals, with a focus on the applications of deep learning algorithms in heart sound processing. Mel-frequency Cepstrum (MFCC) was used to extract different features for classifying the dataset. The most conspicuous characteristics are chosen and employed in training the deep learning algorithm for the automatic classification of heart diseases. The deep learning methodologies employed in the study were CNN and LSTM. The findings of the experiment reveal that the efficiency of the procedure appropriate diagnosis of cardiac illnesses using the heart sound dataset is excellent in accuracy, making it appropriate for real-time applications.**

***Keywords*: Artifact signal, Cardiovascular diseases, Denoising, Extra systole signal, Murmur signal, Mel-frequency Cepstrum, Normal signal, Phonocardiogram.**

1. Introduction

Many studies have used audio signal processing to distinguish natural heart sounds from heart automatic sounds with pathological murmurs in recent years. With about 290 million people suffering from cardiovascular diseases in China alone, cardiovascular disease prevention and treatment have become a top priority for health-conscious people. [1] In rural areas, automated and early detection of cardiac problems may be a viable option for reducing mortality. The main reason that the non-invasive heart sound/beats detection methodology has gotten a lot of interest is because of the rise in cardiovascular illness. Cardiac severe pathology may exist without causing symptoms. Clinical physical examination of the human heart using stethoscope auscultations is an easy, accurate, and cost-effective process, but it requires qualified medical experts. [2] The diagnosis of cardiac disorders relies heavily on heart sounds. However, experts find it difficult and time-consuming to distinguish between various heart sounds because of the poor signal-to-noise ratio (SNR).

Cardiovascular infections have gotten perhaps the most predominant danger to human well-being all through the world. As a non-invasive aide demonstrative device, heart sound identification procedures assume a significant part in the expectation of cardiovascular illnesses. The most recent advancement of the PC-supported heart sound discovery strategies throughout the most recent five years has been checked on. Machine learning researchers are particularly interested in the problem because it requires the classification of data, making distinctions between groupings of stimulation is difficult [3]. Data was collected in real-world environments and often includes various types of background noise. Differences in heart sounds that lead to different heart symptoms can be very subtle and difficult to distinguish. Classifying this type of data successfully necessitates the use of highly robust classifiers.

Despite its medical importance, this is a machine learning technology that has yet to be fully explored. There are primarily the accompanying viewpoints: the speculations of heart sounds and the connection between heart sounds and cardiovascular sicknesses [4]; the critical advances utilized in the preparing and investigation of heart sound signs, including de-noising, division, highlight extraction, and order; with accentuation, the utilizations of profound learning calculation in heart sound handling. Eventually, a few territories for future exploration in PC helped heart sound discovery strategies are investigated, wanting to reference the forecast of cardiovascular sicknesses [5].

The main aim of this research is to develop an automatic heart sound signal analysis system that will aid doctors in the early detection of heart murmurs. This study is to uses an analytical viewpoint method to classify heart disease illness based on heart sound. The deep learning algorithm that is being used to process and analyze PCG data is discussed.

2. Related Work

The heart sound signal can be acquired using an electronic stethoscope; in each complete cycle of heart sound signal, we have S1 to S4 intervals; S3 and S4 are rare heart sounds and are not ordinarily audible but are often shown on the graphic recording in the form of phonocardiogram waves. [6] The S3

*Corresponding author: alampallynaveen123@gmail.com

and S4 intervals are known as murmur sounds, and the heart sound signal that contains murmurs is known as strange heart sounds, and it can be identified based on the position of the murmur within the signal. Deep learning has shown good superiority within the computer-aided classification of heart sound signals. [7]

The physiological data of heart sounds are contained in varied domains like time domain, frequency domain, time-frequency domain, making the feature extraction method harder. In addition, the extracted options area unit time-varying in most cases. [8] Consequently, there is an expansion of feature extraction strategies exploiting completely different aspects of the heart sound signal. Among these, strategies based chiefly on applied mathematics approaches, strategies using wave transforms, and strategies exploiting spectral properties represent the main approaches to the matter.

The GAP layer carries out average pooling on the feature map; therefore, it can learn information concerning every feature map. Every feature map generates a feature node, and every one of those feature maps or mapped into an eigenvector. The GAP layer plays the role of a bridge between the Convolutional block and a classification task while not requiring additional parameters to be trained. [9] Besides time-frequency features, MFCCs also are widely used options for phonocardiogram analysis with the mix of convolutional neural networks and different machine learning classification algorithms. The frequency options area unit captured exploitation quick Fourier transform (FFT), and a convolutional primarily based model is employed to develop a cardiovascular diseases (CVD) detection system. The experiments area unit was performed with the pascal dataset. [10] Similarly, a typical feature set for the centre sound classification was produced for phonocardiograms and employed in the Spectrograms area unit. Using the Physio Net dataset, the authors developed a reliable algorithm for diagnosing cardiac disorders.

The Convolutional neural networks (CNN) hold accuracy and loss over train and looking at knowledge. It does not generalize from train to unseen; take a look at data and memorize the main points that do not affect on the performance. It is going to indicate that the model is redundant. Because of its repeated structure, the model is additional computationally intensive and vulnerable to overfitting, though it has less trainable parameters than CNN. As a result, the model is exceptionally likely to be lowered without incurring significant performance reduction; other regularization approaches may be beneficial. [11]

While recording, several noises other than the artifacts area unit were also recorded. So, denoising is that the initial and crucial step for proper analysis of the heart sound. For this purpose, a bandpass filter is employed. [12] The three primary building phases of the proposed heart valve disease detection technique are as follows: subsequent three elementary building phases: Preprocessing as well as a discretization of the attributes, Generate the reduces with less number of attributes in conjunction with the importance of the attributes, Rule generation for the classification and generate a listing of rules, compute the general accuracy of the generated rules; this phase utilizes the principles generated from the previous section to predict the classification accuracy. [13]

3. Methodology

The heart sound signal carries valuable information about the heart's functioning and health status. It has been analyzed using a signal processing methodology to diagnose various cardiac disorders before they worsen. [14] The primary phase is to gather the publicly accessible datasets and organize them in a very compatible type for more processes. The consecutive phase is to perform necessary preprocessing on the datasets. Signals are chunked into fixed window sizes during preprocessing.

A. Data Collection

Data utilized in this paper is an assortment of heart sound. The commotions made by the pumping heart and the blood flow through it are known as heart noises. Every heartbeat in solid grown-ups is accompanied by two common heart noises, which are commonly portrayed as a lub and a name (or dub). The atrioventricular and semilunar valves produce the primary and secondary heart sounds, respectively. File augmentation .wav, the sound records are of changing lengths, between 1 second and 30 seconds (some have been cut to decrease exorbitant clamour and give the particular section of the sound).

Natural, Artifact, Murmur, and Extrasystole are four categories of heart disorders data in phonocardiogram waves that are utilized in the development of a deep learning model. Based on heart sound, the model has been trained to classify heart illness. Each pulse creates two natural heart sounds in fit individuals, which are generally referred to as a lub and a dub.

Most data in heart sounds are contained in the low recurrence segments, with commotion in the higher frequencies. It isn't unexpected to apply a low-pass channel at 195 Hz. Quick Fourier changes are likewise liable to give valuable data about volume and recurrence over the long haul. More space explicit information about the contrast between the classes of sounds is provided underneath. The types of cardiovascular diseases which are used are given below:

1) Normal

There are commonly safe heart sounds in the Normal category. After completion of recording, when the device is out of the body, there may be a disturbance. Random noises can also be heard, like inhalation, recorder brushing during the process. The period between "lub" and "dub" is less than the period between "dub" and the succeeding "lub" in a regular heart sound.

Ex:

…lub……….dub……………..lub……….dub……………..lub……….dub…
………….lub……….dub…

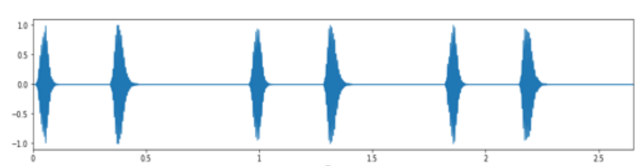

Fig. 1. Normal heart sound wave plot

2) *Murmur*

In one of two temporal places, a "whooshing, booming, rumbling, or turbulent fluid" sounds can be heard: (1) between inhalation and exhalation or (2) between exhalation and inhalation That may indicate range for cardiac issues, out of which they may be saviour, but a lub and a dub will persist. Murmurs can take place between inhalation and exhalation only. One of the characteristics that non-medically trained people find perplexing is this.

Ex:

...lub..****...dub................lub..****..dub................lub..****..dub......

..........lub..****..dub...

(or)

...lub..........dub...******....lub..........dub...******....lub..........dub...*

*****....lub..........dub...

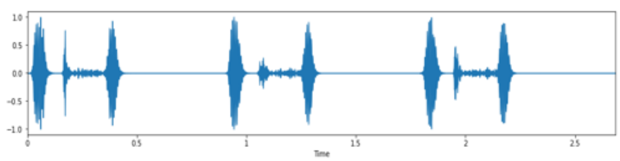

Fig. 2. Murmur heart sound wave plot

3) *Artifact*

The Artifact group contains a variety of audios. These frequently may not be detectable cardiac audio at frequencies below 195 Hz, and as a result, there was next to no sequential repetition. This is perhaps the special type among several.

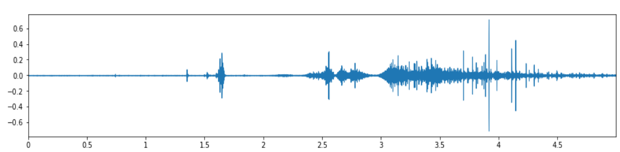

Fig. 3. Artifact heart sound wave plot

4) *Extrasystole*

Extrasystole sounds, sometimes known as "lub-lub dub" or "lub dub-dub," are characterized by an out-of-sync heart tone due to more or fewer heartbeats. Extrasystole isn't usually a sign of something wrong. It can affect anyone, even adults, and is more prevalent in youngsters. Extrasystoles can, however, be affected by heart disease in some cases. Treatment is more likely to be successful if these diseases are diagnosed sooner.

Ex:

lub..........dub............lub...........dub.........................lub.....................dub

.................lub.lub.................dub.......

(or)

lub.........dub............lub...........dub..........................lub.................

......dub.dub.........................lub...................dub........

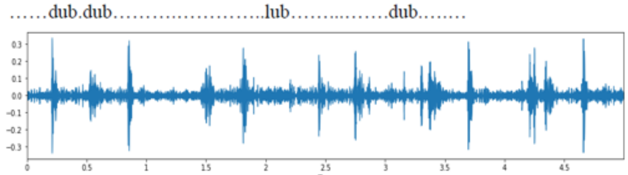

Fig. 4. Extrasystole heart sound wave plot

B. *Preprocessing*

Automatic classification of heart sounds in the ordinary and abnormal categories may be a difficult task due to the insignificant variance in heart sounds that are coming from normal or abnormal subjects. Because of this, the development of a strategic methodology that might properly extract out the outstanding and discriminatory features from the centre signals and which can any assist the machine learning-based classifiers in higher classification of the centre sound in terms of traditional and abnormal heart sound signals are very much needed. [17] Within the projected work, audio samples of heartbeat are recorded with the assistance of PCG. While recording, several noises and different artifacts area unit also recorded. Denoising is the first and most crucial step in adequately analyzing the heartbeat. A bandpass filter is used for this purpose. The vary of the used bandpass filter is taken from 20 to 500 Hz. Lower cut-off frequency, i.e., 20 Hz, is prescribed within the user info. The filtered signal preserved all the knowledge values, such as the above-named frequency ranges.

Organize the data that have been chosen by encoding, washing, and sampling it. Three common data preprocessing steps are:

- Formatting: The information that has been chosen may not be organized in a way that we can work with. The data may be in a social knowledge set, which would want it to be in a level document, or it might be in a restrictive record arrangement, which would want it to be in a social data set or a content record.
- Cleaning: Data cleaning is the process of removing or replacing missing data. There may be instances where information is lacking and does not transmit the information that has been trusted to resolve the problem. These examples could be removed. Furthermore, sensitive data may exist in a fraction of the ascribes, and these credits may be anonymized or removed from the data.
- Sampling: There may be significantly more relevant information available than is required for work. More data will encourage more calculations to be done, as well as more significant process and memory conditions. We can perform a smaller delegate test on the selected data, which will be faster for studying and prototyping arrangements before examining the complete dataset.

C. *Feature Extraction*

The subsequent stage is to do Feature extraction is a method of lowering quality. Unlike feature determination, which places current ascribes in order of predictive value, highlight extraction alters the credits. The modified characteristics, also known as features, are simple blends of the original qualities. [18]

1) *Mel-Frequency cepstral coefficients*

From the heart signal that we used in our investigation, we extracted features as a set of measured, non-redundant, and derived values. The first step in any automated audio recognition system is to extract features or classify the audio signal components that are useful for identifying linguistic content while discarding all else, such as background noise, emotion, etc. An audio signal is moulded to smaller windows using the Hamming window in the MFCC method, resulting in signal splitting into frames. [19]

The MFCC may be calculated using the following formula:

- Take (a windowed snippet of) a signal and perform the Fourier transform on it.
- Using overlapping triangular windows, map the powers of the spectrum acquired above onto the Mel scale.
- Take a look at the power logs for each Mel frequency.
- As if it were a signal, compute the discrete cosine transform of the list of Mel log powers.
- The resulting spectrum from the amplitudes is the MFCCs.

MFCC can be calculated in Python using the liborsa module, which returns two attributes: data and sample rate. Features and labels will be extracted from Filename and kept in x_data.

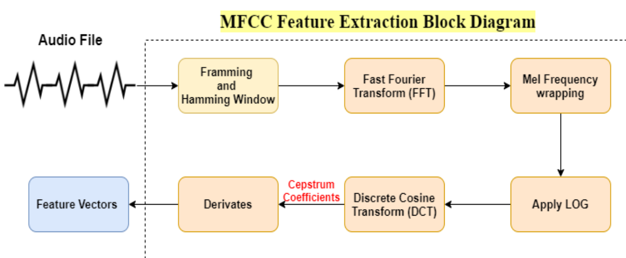

Fig. 5. MFCC block diagram

Model evaluation aids in the selection of the best model to reflect our data and the prediction of how well the chosen model will do in the future. In data science, evaluating model output with the data used for training is not a good idea because it can lead to over-optimistic and overfitted models. [20] The averaged performance of each classification model is used to estimate its performance. The result will be in the shape that has been visualized. Graphs are used to represent data that has been categorized.

This huge dataset is divided into three parts in different proportions to train, test and validate the model. They are

- Training Dataset
- Testing Dataset

Training Dataset is partitioned in such a way that it consists of the major part of the overall dataset. This dataset comprises around 80% of the data i.e., around 404 phonocardiogram waves had used only for the training process. The reason for the maximum fraction for this dataset is as this training process will be the heart of the process and also this is the major part that is used for detecting the images. So, the maximum we take the training dataset, the maximum will be the effectiveness and accuracy.

The testing dataset uses the 101 heart sound of the original dataset and the main theme of this testing dataset is to measure how good the training process was done; it will measure the accuracy rates accordingly. If the testing dataset is more random then obviously the effectiveness of the model will be judged accordingly.

D. Algorithms

1) Long Short-term Memory Algorithm

The LSTM neural network is a type of RNN that can learn long-term dependencies and avoids the problem of vanishing gradients. Memory chunks, also known as subnets, make up the LSTM. Input, output, and forget gates, as well as a memory cell, are all included in each block. According to the equations, the LSTM layer transfers the input sequence X to the output sequence Y.

$x = (x_1, x_2, \dots x_T)$

$y = (y_1, y_2, \dots y_T)$

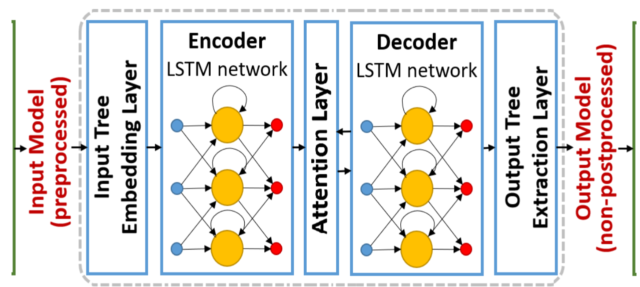

Fig. 6. Layered architecture of LSTM

In a multi-layer structure, the previous layer's output is the next layer's input. A dense level will accompany two LSTM layers with SoftMax activation feature in our heart sound classification model. LSTM generates a series, however only the last value will be sent to the output layer. The first two levels contain 64 and 32 units respectively, whereas the last layer is dense, as is appropriate for a good class. To reduce overfitting, dropout with a rate of 0.25 is applied to the output of the LSTM layers. B using Adam optimizer, we minimized the training categorical cross-entropy loss function. Because of the long training time, a full search of hyper parameters is infeasible. Thus, the most promising combination was found using single-fold evaluation by using one-hot encoding. The input of our model is magnitude Mel-spectrogram with 128 bands that cover a frequency range from 0 Hz to 22050 Hz, which has .wav extension. Audio files are evaluated at a sample rate of 44100 Hz using a 1024 sample window and a hop size of the same width. The length of the input sequence is variable and depends upon audio clip duration. Later on, it is padded to the same time interval for all the audio files in the dataset.

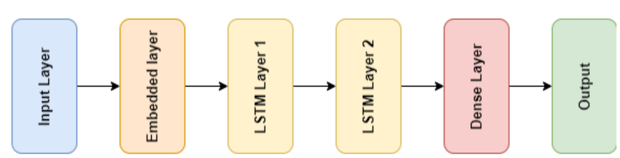

Fig. 7. Architecture of LSTM algorithm

Embedding layer: in simple terms, it creates word vectors of each word in the word index and groups words that are related or have similar meanings by analyzing other words around them.

LSTM Layer: This layer will decide whether to preserve or discard data based on the present input, prior output, and memory.

Dense layer: Using an activation function and the weight matrix and bias (optional), compute the input. The optimizer is the Adam optimizer.

For training, it is simple. We only need to fit our x train (input) and y train (output/label) data. I use a mini-batch learning method with a batch size of 32 and 100 epochs for this training. To assess the model, we must use our x test data to forecast sentiment and compare the predictions to the y test anticipated output data. Then, we calculate the model's accuracy by dividing the numbers of correct predictions by the total data.

2) Convolutional Neural Networks

Deep learning has had considerable success in neural network fields, image identification, speech recognition, and medical image and signals analysis, particularly for convolutional neural networks (CNN). Several researchers have recently employed deep learning methods to categorize cardiac sounds mechanically. Raw heart sound signal time-frequency options were sent into the model to identify healthy heart sound and pathological cardiac sounds as Murmur, Artifact, and Extrasystole. Our convolutional neural network model uses Mel-frequency spectral coefficients (MFFCs) as inputs.

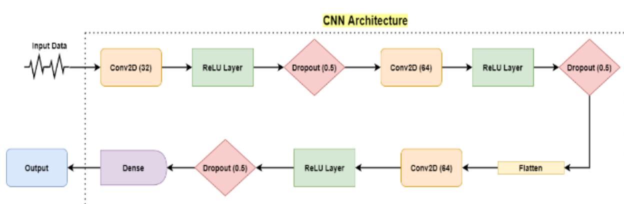

Fig. 7. CNN architecture algorithm

The Dropout Layer prevents overfitting by setting the weights of a section of the data to zero at random, and the Dense units have hidden layers connected to the degrees of freedom the model has while fitting the data. Overfitting can occur in the training dataset when all features are connected to the convolutional layer. Overfitting happens when a model performs poorly on new data but performs well on training data. A dropout layer is used to solve this problem, in which a few neurons are removed from the neural network during training, resulting in a model. After passing a dropout of 0.5, 50% of the nodes in the neural network are dropped out at random. The Flatten Layer condenses every feature map data into a single column, which is then sent into the Dense layer, which produces the species that the model is expected to classify the audio recordings into the model. Finally, the activation function is one of the most crucial elements in the CNN model. [24] The activation functions provide the model with the ability to be nonlinear. The Rectified Linear Unit (ReLU) function is used to zero out negative weights in this case. Other activation functions can be found here, but this is an excellent place to start. SoftMax is the activation function type for the last Dense layer, and it outputs a probability for each class. The loss functions are used to evaluate how different the expected and actual data are and penalize the model for poor predictions.

4. Implementation

Transfer Learning modules from deep learning were with the support of python programming, Transfer Learning modules from deep learning were implemented. The deep learning library of TensorFlow 2.3.2 is used to incorporate all Transfer Learning, and the training and testing procedures are carried out on the Jupyter Notebook platform. All the experiments presented in this paper have been performed on Jupyter with the Windows operating system where we use the CPU, Tesla K80 Graphical Processing unit, or Tensor Processing Unit hardware from the online cloud service that is available for free. The Convolutional neural networks (CNN) and long short-term memory (LSTM) models pre-trained on the Heart sound dataset are used with random initialization weights and adaptive moment estimation (ADAM) optimization of the cross-entropy function has been performed. For all experiments, a batch size of 32, a learning rate of 0.00001, and several epochs of 100 were used, respectively. There has been a random division of data into two separate datasets (train & test) of 80 % and 20% for training and testing, respectively.

5. Experimental Results

The proposed architecture is to demonstrate that the pipeline model that we assembled would optimize the outcome and evaluate all model results, then compare model accuracy and training time. The classification model was analyzed in terms of precision, recall, and F1-score and is applied to the dataset after data set collection and using different data preprocessing techniques. These are some of the results after applying all the algorithms. We obtained the following accuracies after successfully training the model using each of the strategies. The computed accuracies obtained after training the majority of the images in the dataset, as well as the validation dataset, are utilized to verify the results [25]. Because both methods use the same dataset, they can be compared to see which one performed better and which one is more effective. It is clear that both techniques require a GPU for training, therefore the computational cost will be higher, possibly even higher in LSTM due to the higher number of filters and layers in this approach when compared to CNN.

A. LSTM

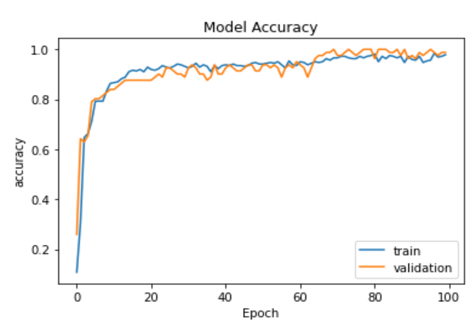

Fig .8. Epoch vs. Accuracy

In Fig. 8, In this Model Accuracy, the x-axis indicates the Epoch, and the y-axis indicates the Accuracy of training and testing of data, which displays the graph of our dataset.

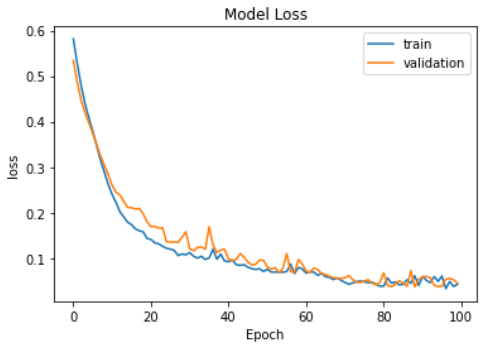

Fig. 9. Epoch vs. Loss

In Fig. 9, In this Model Loss, the x-axis indicates the Epoch, and the y-axis indicates the Loss of training and testing of data, which displays the graph of our dataset.

B. CNN

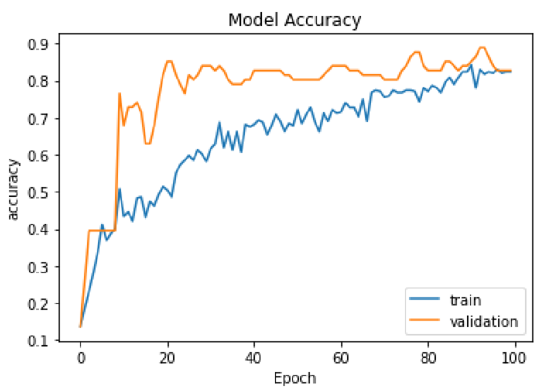

Fig. 10. Epoch vs. Accuracy

In Fig. 10, In this Model Accuracy, the x-axis indicates the Epoch, and the y-axis indicates the Accuracy of training and testing of data, which displays the graph of our dataset.

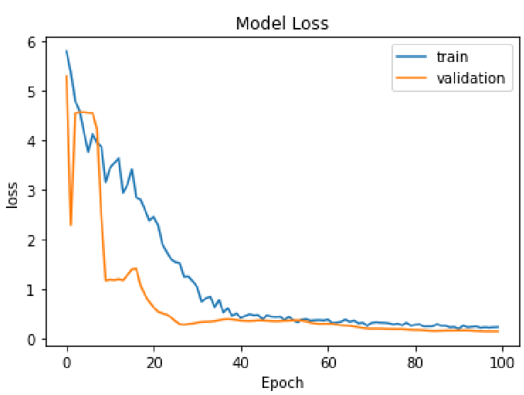

Fig. 11. Epoch vs. Loss

In Fig. 11, In this Model Loss, the x-axis indicates the Epoch, and the y-axis indicates the Loss of training and testing of data, which displays the graph of our dataset.

In the convergence graph field, it is a standard measure to check how quickly the model can classify the data. Precision, recall, F1-score, and Accuracy is defined, and the following equations will describe them.

Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

F1-score = 2*(precision*recall)/ (precision + recall)

Accuracy: It is defined as the number of true positives and true negatives by the number of true positives, true negatives, false positives, and false negatives.

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Table 1
Classification report of all algorithms

Algorithms	Precision	Recall	F1-Score	Accuracy
Long short-term memory (LSTM)	0.94	0.93	0.94	0.94
Convolutional neural networks (CNN)	0.90	0.89	0.88	0.89

In the above model, the Precision, recall, and F1-score, and Accuracy performances are listed after processing via various algorithms in Table 1. Long short-term memory (LSTM) and Convolutional neural networks (CNN) algorithms are trained and probed on the Heart sound phonocardiogram dataset of the pre-trained models. In the above figures, the model accuracy and model loss of the trained models are given. [26] It can be said that for the initiation, the highest model training accuracy is obtained for Long short-term memory, and the training is carried out for 100 epochs. We can analyze from the model loss figures that the loss values are reduced in all pre-trained models, and the comparison of all algorithms is listed based on performance from the classification table. Comparing the Precision, recall, f1-score from the above table, the Long short-term memory gives the best results in both the training and testing phase and shows better output than the other model.

6. Conclusion

In this developed algorithm, a deep learning-based classification technique is applied to classify the PCG recorded heart sounds as normal/abnormal also in abnormal subtypes as a murmur, artifact, and extrasystole. MFCC is used as feature extraction, and various statistical features are extracted from filtered heart sound signals. In the proposed work, the features were strategically considered which are likely to be affected during an abnormal heart functioning. Those features improved the performance of deep learning algorithms in terms of time complexity and accuracy. We experimented with different convolutional neural networks and long short-term memory (LSTM). By fine-tuning two pre-trained models on our training set, we developed a deep learning system for heart disease detection from a heart sound dataset. Comparing all the algorithms, we conclude that the Long short-term memory gives the most accurate results for model accuracy and gives less model loss. Based on the confusion matrix, long short-term memory (LSTM) obtained the highest accuracy and f1 score with 94% among all the algorithms. In the Future, to validate our proposed model, more datasets will be collected and added to the existing dataset of heart sound can be considered, and we would like to research and develop a more effective and precise image classification model.

7. Future Work

Some of the improvements further we can do is improving the quality of the generated audios. And some major drawbacks are it consumes too much time to train the model even with GPU. We can further optimize the model to decrease the training time.

Acknowledgment

This Project provides us with an opportunity to have a greater understanding of the subject and explore beyond. This work was supported in part by the Vellore Institute of Technology.

References

[1] Suboh, Mohd Zubir & Mashor, Mohd & Saad, A & Mohamed, Mohd Sapawi. (2008). Heart valve disease classification based on heart sound.

[2] Redlarski, G., Gradolewski, D., & Palkowski, A. (2014). A system for heart sounds classification. *PloS one*, *9*(11), e112673.

[3] Randhawa, S. K., & Singh, M. (2015). Classification of heart sound signals using multi-modal features. *Procedia Computer Science*, *58*, 165-171.

[4] Amiri, A. M., & Armano, G. (2013). Heart sound analysis for diagnosis of heart diseases in newborns. APCBEE procedia, 7, 109-116.

[5] Jiang, Z., & Choi, S. (2006). A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. *Expert Systems with Applications*, *31*(2), 286-298.

[6] Rubin, J., Abreu, R., Ganguli, A., Nelaturi, S., Matei, I., & Sricharan, K. (2017). Recognizing abnormal heart sounds using deep learning. arXiv preprint arXiv:1707.04642.

[7] Li, S., Li, F., Tang, S., & Xiong, W. (2020). A review of computer-aided heart sound detection techniques. *BioMed research international*, *2020*.

[8] Upretee, P., & Yüksel, M. E. (2019, April). Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1-4). IEEE.

[9] Lee, C.C., Mower, E., Busso, C., Lee, S., Narayanan, S. (2011). Emotion recognition using a hierarchical binary decision tree approach. Speech Communication, 53(9-10):1162-1171

[10] Khan, K. N., Khan, F. A., Abid, A., Olmez, T., Dokur, Z., Khandakar, A., ... & Khan, M. S. (2020). Deep Learning Based Classification of Unsegmented Phonocardiogram Spectrograms Leveraging Transfer Learning.

[11] Palaz, D., Doss, M.M., Collobert, R. (2015). Raw speech signal-based continuous speech recognition using convolutional neural networks. 2015 IEEE International 122 Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, pp. 4295-4299.

[12] Yadav, A., Singh, A., Dutta, M. K., & Travieso, C. M. (2019). Machine learning-based classification of cardiac diseases from PCG recorded heart sounds. *Neural Computing and Applications*, 1-14.

[13] Hamdy, A., El-Bendary, N., Khodeir, A., Fouad, M. M. M., Hassanien, A. E., & Hefny, H. (2013, September). Cardiac disorders detection approach based on local transfer function classifier. In *2013 Federated Conference on Computer Science and Information Systems* (pp. 55-61). IEEE.

[14] Brunese, L., Martinelli, F., Mercaldo, F., & Santone, A. (2020). Deep learning for heart disease detection through cardiac sounds. Procedia Computer Science, 176, 2202-2211.

[15] Son, G. Y., & Kwon, S. (2018). Classification of heart sound signal using multiple features. *Applied Sciences*, *8*(12), 2344.

[16] Kotb, M. A., Nabih, H., El Zahraa, F., El Falaki, M., Shaker, C. W., Refaey, M. A., & Rjoob, K. W. (2016). Improving the recognition of heart murmur. Int. J. Adv. Comput. Sci. Appl, 7(7), 283-287.

[17] Chen, T., Xiang, L., & Zhang, M. (2015). Recognition of heart sound based on distribution of Choi-Williams. Research on Biomedical Engineering, (AHEAD).

[18] Banerjee, P., & Mondal, A. (2015). An irregularity measurement based cardiac status recognition using support vector machine. Journal of medical engineering, 2015.

[19] Alhlffee, M. (2020). MFCC-based feature extraction model for long time period emotion speech using CNN. Revue d'Intelligence Artificielle, Vol. 34, No. 2, pp. 117-123.

[20] Cabral, F. S., Fukai, H., & Tamura, S. (2019). Feature Extraction Methods Proposed for Speech Recognition Are Effective on Road Condition Monitoring Using Smartphone Inertial Sensors. *Sensors*, *19*(16), 3481.

[21] Lezhenin, I., Bogach, N., & Pyshkin, E. (2019, September). Urban sound classification using long short-term memory neural network. In *2019 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 57-60). IEEE.

[22] Redlarski, G., Gradolewski, D., & Palkowski, A. (2014). A system for heart sounds classification. *PloS one*, *9*(11), e112673.

[23] Li, F., Tang, H., Shang, S., Mathiak, K., & Cong, F. (2020). Classification of heart sounds using convolutional neural network. *Applied Sciences*, *10*(11), 3956.

[24] Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., & Wilson, K. (2017, March). CNN architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 131-135). IEEE.

[25] Jiang, Z., & Choi, S. (2006). A cardiac sound characteristic waveform method for in-home heart disorder monitoring with electric stethoscope. *Expert Systems with Applications*, *31*(2), 286-298.

[26] Babaei, S., & Geranmayeh, A. (2009). Heart sound reproduction based on neural network classification of cardiac valve disorders using wavelet transforms of PCG signals. *Computers in biology and medicine*, *39*(1), 8-15.