How to Cite:
Thendral, P., Karkuzhali, S., Siyon, V. A., & Sweeton, C. J. K. (2022). Prediction of abnormalities in heart beat sounds using convolutional neural networks. *International Journal of Health Sciences*, *6*(S4), 9844–9855.
https://doi.org/10.53730/ijhs.v6nS4.11312

Prediction of abnormalities in heart beat sounds using convolutional neural networks

Dr P Thendral
Assistant Professor Senior, Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, Sivakasi, India
Corresponding author email: thendralp@mepcoeng.ac.in

Dr. S. Karkuzhali
Assistant Professor, Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, India
Email: karkuzhali@mepcoeng.ac.in vijikarkuzhali@gmail.com

V. A. Siyon
Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, Sivakasi, India
Email: eugeneronald_ai@mepcoeng.ac.in

C. Jeyanth Kallis Sweeton
Department of Artificial Intelligence and Data Science, Mepco Schlenk Engineering College, Sivakasi, India
Email: sweeton2001_ai@mepcoeng.ac.in

***Abstract*---**Worldwide physicians prefer to use physical stethoscope and they listen to the heart beat sound and its rhythm to diagnose various heart conditions. In this work various abnormalities that happened in heart will be reflected in the sound of the heartbeat. In this work we have created a classification system which is based on Convolutional Neural Network (CNN) to analyze the heart beat sound to predict the abnormalities. Heart beat sound is converted to spectrogram images and then CNN is trained with those images. In order to reduce the computational time, pooling is done, so that it reduces the parameters by taking particular pixel from particular part of pixels. The parameters in CNN are varied in the convolution and pooling layers to enhance the accuracy of classification of heart beat sound. Experiment is carried out by varying number of convolutional layers and by changing the pooling methods for various combinations of CNN models and the results are analyzed to find the optimal combination which can suit for sound analysis.

Keywords---convolutional neural network, deep learning, medical acoustics analysis, spectrogram, artificial intelligence healthcare.

Introduction

The electrical impulse generated from atria and ventricles is called as heartbeat. In right atrium, an electrical impulse starts with bundles of specialized cells called SA node. It is also known as natural pacemaker. The heartbeat is generated by this electrical impulse. The SA node is responsible for setting the rate and rhythm of the heartbeat. There are many types of heartbeat sounds. Sounds are created or made when an object vibrates. For heartbeat sound, atria and ventricles work together by contracting and relaxing. Different series of sound are produced by heart based on their characteristics, which can be normal or abnormal. Normal sinus rhythm is characterized as normal heart rhythm because the sinus node fires regularly. There are different types of heart beat sounds that varies on the basis of electrical impulse generated by contraction and relaxation of atria and ventricles. This electrical impulse generally produces two kinds of sounds that is lub and dub. The lub is technically called S1 and the dub is called S2. The closing of valves in our heart produces these lub and dub sounds. The abnormal sounds are produced due to some problem in heart. The symptoms for abnormal sounds in heart include ventricular gallop, early diastolic gallop, ventricular filling sound, protodiastolic sound. The normal heart sound is S1 and S2 and the cardiac cycle timing for this is start of systole and end of systole. The abnormal heart sound is S3 and S4 and the cardiac cycle timing for this is early diastole and late diastole. The types of heart beat sounds are categorized into murmur, normal, extrasystole, artifacts.

Speaking of the different types of the heart beat sounds in Fig.1, normal heart beat sounds are the sounds generated by the health heart which is more commonly seen in the body of a healthy adult. A normal heart beat generates two sounds, a lub (also called as s1) and a dub (also called as s2). The problems in the heart may generate different heart sounds other than the normal heart beat. The heart also generates murmur sounds Fig.6 like whooshing and swishing, which is caused due to the rapid flow of blood to the heart. This is also caused due to the noisy blood flows to the heart. The next type of heart sound is the extrasystole, this is produced due to the extra heart beat in each heart cycle, unlike the normal heart beat cycle. The occurrence of this type of sound in certain ages of people can be an indication of the disease tachycardia. Tachycardia is an irregular generation of electric signal formed in the upper or lower chambers of heart which makes the heart to beat faster. The heart rate of a tachycardia patient may be over 100 beats per minute. The next type is artifact, they are the electrocardiographic alterations which are formed inside a heart but are not a type of cardiac electrical activity. Due to the artifact heart beat the parameters of ECG (Electrocardiogram) like the waves and the baseline are both disorted. The shaking of the rhythmic movement causes the motional artifacts.

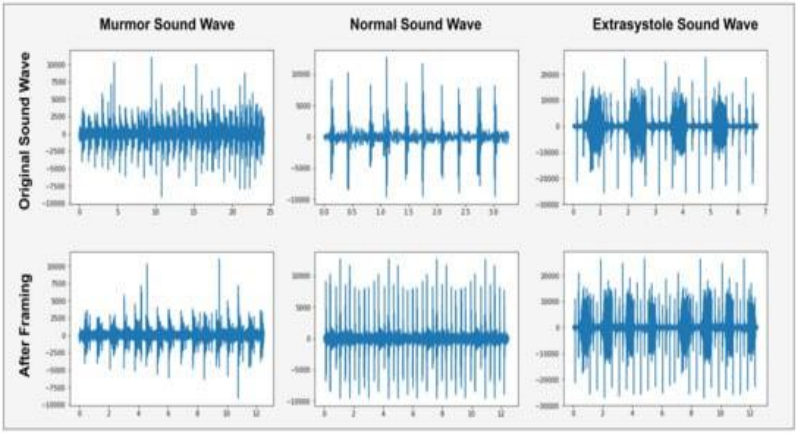

Fig.1. Types Of Sound Waves Of Heart Sounds

These different types of heart beat sounds are in the form of audio file. These audio files are in the duration of eight seconds. The audio is perceived either by sensor or the microphone is attached to the stethoscope, when microphone is vibrated, it generates an electrical signal and that is converted into voltage. The voltage is sent to computer and thereby it recognizes these audios. The intensity of the sound is called amplitude. In one second, the audio signal produce number of waves is termed as frequency. The audio signal can be digitalized by measuring amplitude at fixed interval of time. The audio files are in the form of .wav, .mp3,.wma and so on. For audio processing, python has library called librosa. These audio files are converted to spectrogram. Deep Learning model mostly have the ability to recognize the audio file as spectrogram. Spectrogram is simply an image, which is used to represent audio signal to image and the Fourier transform is used to generate the spectrogram. Spectrum of Frequency over time graph is called spectrogram and in x axis, it plots time and in y axis, it plots frequency.

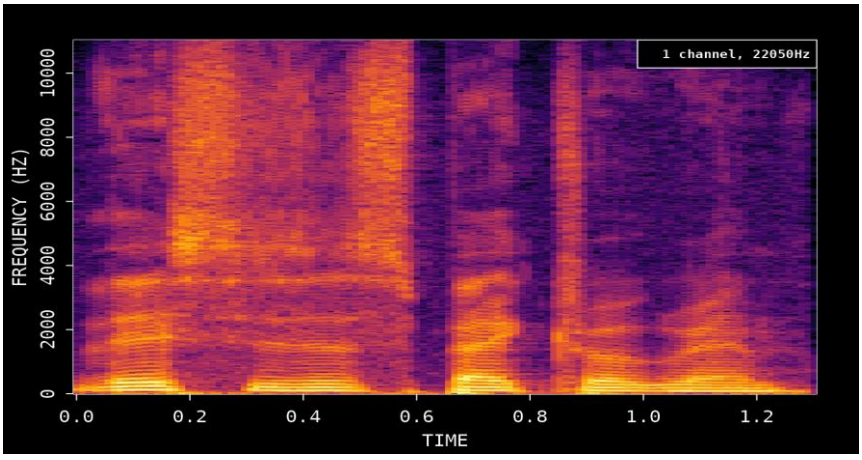

Fig.2. Spectrogram Image

The brighter color in spectrogram Fig.2 indicates that the sound is listened concentratedly and the darker color in spectrogram indicates dead sound. For

creating Spectrogram, we need to chunk the audio file and the compressed audio signal is transformed into spectrogram by Fourier Transform. Human percept frequency in log scale, most of them it's preferred to use log scale because it is audible for human. Normal Spectrogram captured by observing frequency in linear scale, so it's converted to Mel Spectrogram. Mel Spectrogram is obtained by converting using Mel Scale. Mel Scale measures human perception of sound. In Mel spectrogram, the plot in y axis is Mel scale and the plot in x axis is Decibel scale. For human perception, Mel spectrogram is used instead of normal spectrogram because it uses Mel Scale, capture frequency not in linear manner.

Deep Learning model uses CNN because CNN known for its image classification by extracting hidden features from images. Spectrogram is given as input to CNN model, either in grayscale image or red blue green image. Feature is extracted from images by passing through different convolutional layers and pooling methods like max pooling, min pooling and so on. By changing different types of convolutional layers and pooling methods, we can improve extraction of features extraction. Once features are extracted from the images, we use classifier to classify it based on trained model label.

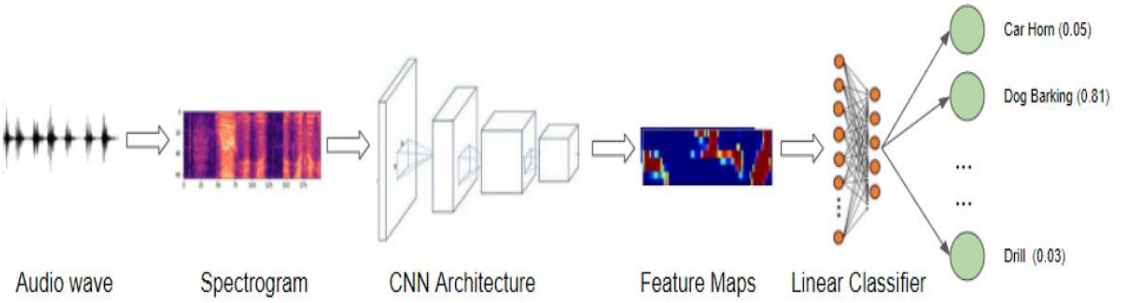

Fig.3.Formulating the CNN Model

After passing through convolutional layers, it's then connected to fully connected layers and by using classifier, we can classify it to specific labels by trained model. Before sending the audio signal to CNN model, we have to pre-process the audio files, i.e. by removing background noise and compressing audio files and so on to improve exact features extraction from obtained spectrogram using CNN which is shown in Fig.3.

Literature review

In order to analyze respiratory sound, the CNN-RNN hybrid model is used to track the particular respiratory and pulmonary sounds, but the drawback is its accuracy is comparatively low which is 66.31% because of downsampling the dataset to 4Hz and they didn't do different pooling methods to improve accuracy (Acharya, J., & Basu, A. (2020). The concept is to analyze the environmental and the urban sounds, they used CNN to analyze the sound. In order to analyze the urban sound and environmental sound, we need to do some preprocessing of audio signals that is, by removing background noise and due to less datasets, they used data augmentation to increase the datasets, but it is not efficient way

to get an improved accuracy (Salamon, J., & Bello, J. P. (2017). The CNN model is used to recognize the sound event and they used gammatonegram as training dataset. The drawback of gammatonegram is using linear filters but human perceive sound not in linear, for human perception we mostly use Mel spectrogram as Mel filters convert the sound to Mel spectrogram according to human perception (Greco et al., (2020). The CNN model is used to audio recapture detection, since the audio clips of 2 second duration, it's not enough to establish an improved accuracy and so they used ENF and its harmonic. But the drawback is excessive computational time compared to normal CNN procedure (Lin, X et al., (2016).

The audio signal classification is done using CNN which is deep learning techniques and explained about various layers of CNN and different pooling methods (Peeters, G., & Richard, G. (2021). Audio depression recognition is done by CNN and GAN (Generative Antagonism Network), they preprocess the audio dataset such as removing background noise and so on, and then splice the audio file into new audio file. But they didn't talk about different pooling methods are not applied to improve the accuracy (Wang, Z. (2020). The proposed method of regularizing the RF (Receptive Field) of CNN, but in emotion and theme detection in music require timestamp-based analysis model such as RNN and LSTM plays a role but they didn't mention about those networks and the performance measure used is Precision-Recall under Curve (Koutini, K. et al., (2021). The audio processing is done by Machine Learning instead of Deep Learning and Clustering techniques. The polyphonic sound is detected by Machine Learning. But the drawback is using machine learning instead of using Deep Learning Techniques like CNN, RNN and so on. Deep Learning is more efficient way for handling audio classification than machine learning because in ML we use features to classify the data but in multimedia data it is not efficient way to specify a particular feature to classify the audio efficiently (Rong, F. (2016).

The GMM (Gaussian Mixture Model) is used to classify the respiratory sounds into normal and wheeze classes. For feature extraction, the Fourier Transform, linear predictive coding, wavelet transform and Mel-frequency cepstral coefficient. But the drawbacks are using machine learning techniques such as GMM for audio classification, but we used CNN. As CNN is good for audio classification and image classification, it extracts hidden features from images by using many different convolutional filters working and it also reduces dimensionality of images and so it reduces the computational time and its main purpose is to process the pixel data and also increase the accuracy (Bahoura, M. (2009). To classify the lung sound, they used Noise masking Recurrent Neural Network (NMRNN). For audio classification, at particular clip of sound we can understand the abnormalities of lung sound and the lung sound is categorized into four, that are normal, containing wheezes, crackles and both wheezes and crackles, RNN is preferred only when we need time-stamp to be kept on record and CNN is used to classify audio and image efficiently at particular clips (Kochetov, et al., (2018, October).

The Hidden Markov model is used to classify the respiratory sound into normal, wheeze and crackle classes.Spectral subtraction is used to remove noise from audio files. But the accuracy of classifying the respiratory sound is only 39.56 ,

its low , since Hidden Markov model is a technique of machine learning , it is not much efficient in classifying audio signal unless all the features are given exactly but its not feasible in real time, in order to overcome we used CNN for audio classification , since CNN is good for extraction pixel from images , so by converting audio into spectrogram, its efficient way to classify audio and it produces high accuracy(Jakovljević, N., & Lončar-Turukalo, T. (2017, November). CNN is used in biomedical field for image classification and genomic sequencing. Here respiratory sound is being classified using CNN. We classified the heartbeat sounds by using CNN model, we are giving input as Spectrogram, which is obtained from audio file. We used Mel-Spectrogram in order to capture audio in non-linear manner which is used for percept by human. Human generally percept sound in non-linear manner (Perna, D. (2018, December).

A linear classifier is used to classify the different respiratory sound. Linear Predictive Coding is used as feature vector. But the drawback of this classifier is human don't perceive sound or audio not in linear manner. So, we prefer CNN, for CNN model we give Mel Spectrogram as input, as Spectrogram converted into Mel Spectrogram using Mel filters, by implementing this we can increase our accuracy than linear classifier (Pramono, R. X. A., et al., (2019). CNN and RNN plays a major role in sound events, CNN is used to classify the particular sound from extracting features from spectrogram and RNN (Recurrent Neural Network) learns context in audio signals by keeping track of previously seen images of spectrogram. But in addition to that, we can implement the model by different pooling methods to improve the accuracy or performance of model (Cakır, et al., (2017). CNN model is used to classify heart sound recordings and it aims to predict whether the heart beat sound is normal or abnormal. Here the heart beat sound is preprocessed by removing noise by Windowed-Sinc Hamming filter algorithm and then it is segmented. But its accuracy score is nearly around 70 to 85. We used Mel-Spectrogram which percepts the heartbeat sound in non-linear manner, that is human perception and using different types of pooling methods also increases accuracy in classification (Ryu, et al., (2016, September).

Data Set

Come to think of it, the human heart sounds vary from person to person, ages to ages and also regions to regions. Speaking of regions, warm blooded animals such as dogs have higher heart beats than the cold blooded animals such as humans. On considering them in account, the heart sounds are divided into several types such as murmur, artifacts, extrasystole and normal. There were only few datasets available on the internet platform. Strictly speaking about the dataset, it consists of nearly 456 rows (Normal – 231, Murmur – 157, Extrasystole – 46, Artifacts – 19) which is clearly shown in the Fig.4. These datasets were provided from the website PETERJBENTY. The input data is in the form of mp3 sound file which are with duration of 5-10 seconds. Some sounds files are quite large, so they were chopped to the current need. The dataset consists of only two columns such as the Name of the Audio File and the Label of that audio files are taken into account as the data. As the population of the human race is quite a large, so if we can take different sounds of the heart beats of different people, we can increase the accuracy rate and the system will be more and more sustainable to the unknown data's classification.

DATA SET	
Types Of Heart Beats	**Number Of Audio Files**
Normal	251
Murmur	137
Artifact	46
Extrasystole	19
Total	453

Fig.4. Rows of the Data Set

Proposed Method

Sounds are generated or created when an object vibrates in due course of time. Similarly, different series of sounds which are made by the heart are based on their characteristics. Deep learning algorithm of CNN is applied for the training the model respectively. The process of this approach is deeply shown the Fig.5 consists of three major steps. First is the extraction of suitable inputs of various heart sounds and chopping them accordingly in the range of 5 to 10 seconds respectively. Secondly the audio files of the heart are converted to spectrogram images by using minimum pooling and maximum pooling, different combinations and the different convolutions. The last and the final step is these spectrogram images are fed to the CNN model to classify the newly approaching sounds. The pooling factors, different combinations and more over convolutions are varied for every trail to find the respective accuracy at each process. All the suitable combinations are taken into the account for the validation of the given CNN model in classifying the heat beat sound of the human beings. Increasing the dataset can increase the given accuracy drastically and also can be improved further.

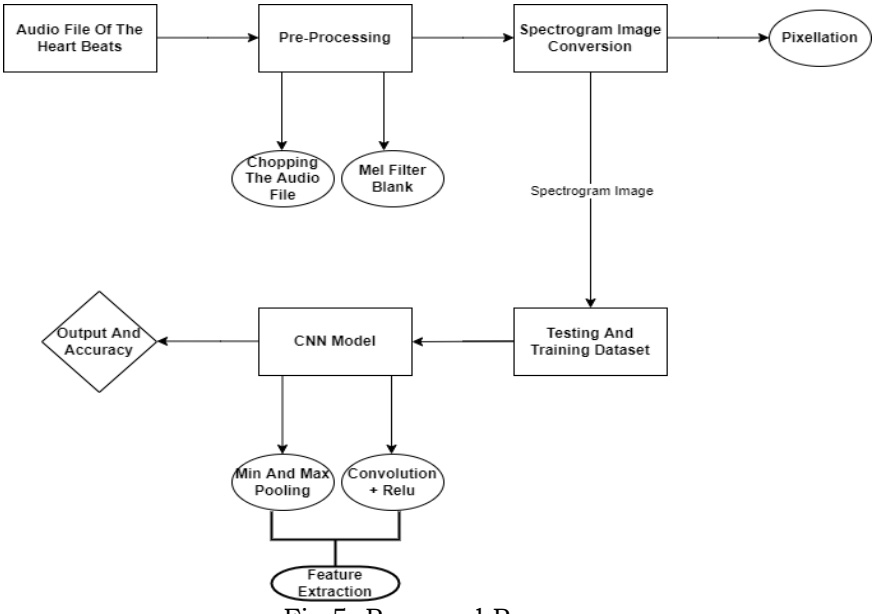

Fig.5. Proposed Process

Input

Speaking early of this, the input data for this approach is the audio files of the human beings at various beats which are classified as normal, murmur, artifacts and extrasystole. The given inputs were chopped according to the need for the process say 5 to 10 seconds. These inputs were clear without any noise in the background as they are already pre-processed for the convenient input given by the user to the model. These audio files were named by the series of the numbers and the respective labels were also given to it. Some audio files had noise and were not clear a lot. This can cause a slight change in the pattern of the spectrogram image. A slight change cannot affect a lot in the output accuracy of the model but speaking of the dataset is huge a lot, they are trained with several numbers of audio files. Each audio file with this noise can affect the output accuracy of the CNN model significantly. So, in order to remove the noise of the audio files, all the audio files were pre-processed once for their sustainability in the input of the model. Even during the chopping of the audio files, the data were restored so that significant beat sounds of the hearts were not lost.

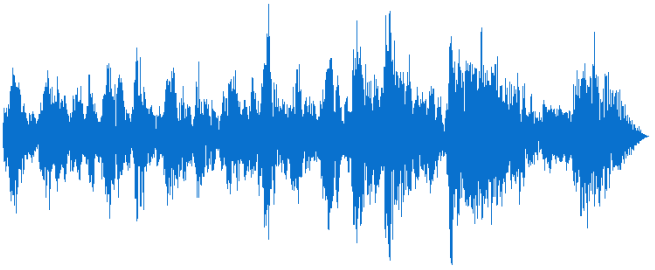

Fig.6.Audio File for Murmur Sound

Converting To Spectrogarm Images

The given mp3 files of the human heart beats are converted to spectrogram images. Going on through the process, the default value of the n_fft is set to 2048 samples as they are really good for the audio files. They correspond to a physical duration of 93 milliseconds in time and also at a sample rate of 22050 Hz. In this approach the win_length is set equal to the n_fft. As speaking of the hop_lenght, they were set to the value of 512 samples respectively. The SR (Sampling Rate) is set to the value of 22050 respectively. The SR (Sampling Rate) was changed with several combinations in due course of time and checked for the accuracy. Here the Mel Filter Blanks are used as the sounds of certain heart beans are rally lower in frequency and sound and some of them are really loud and higher in frequency, to maintain a constant perception of sound ie providing a better resolution at low frequencies and less resolution at higher frequencies of sound. The value of n_mels is set to 10. The Fig.7 depicts the mel filter blank of the murmur sound of the heart beat. The extraction of features is also done by a specific approach.

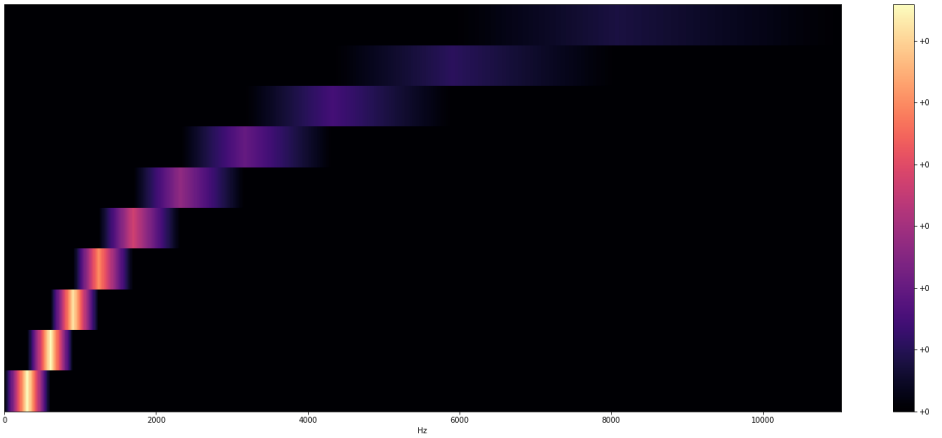

Fig.7.Mel Filter Blank of Murmur Sound

After dealing with this, the spectrogram images of the various audio files of the heart beat sounds are validated and are stored as images. In Fig.8 the spectrogram image of the murmur audio sound of the heart beat is generated by the approach. These images will be fed as the input to the CNN model. The conversion of the audio file to a Spectrogram images nearly takes 0.02 seconds respectively. This computation time changes from every audio file as they are chopped at different courses of time based on the user need. The process of generating the mel spectrogram image is show in Fig.7.

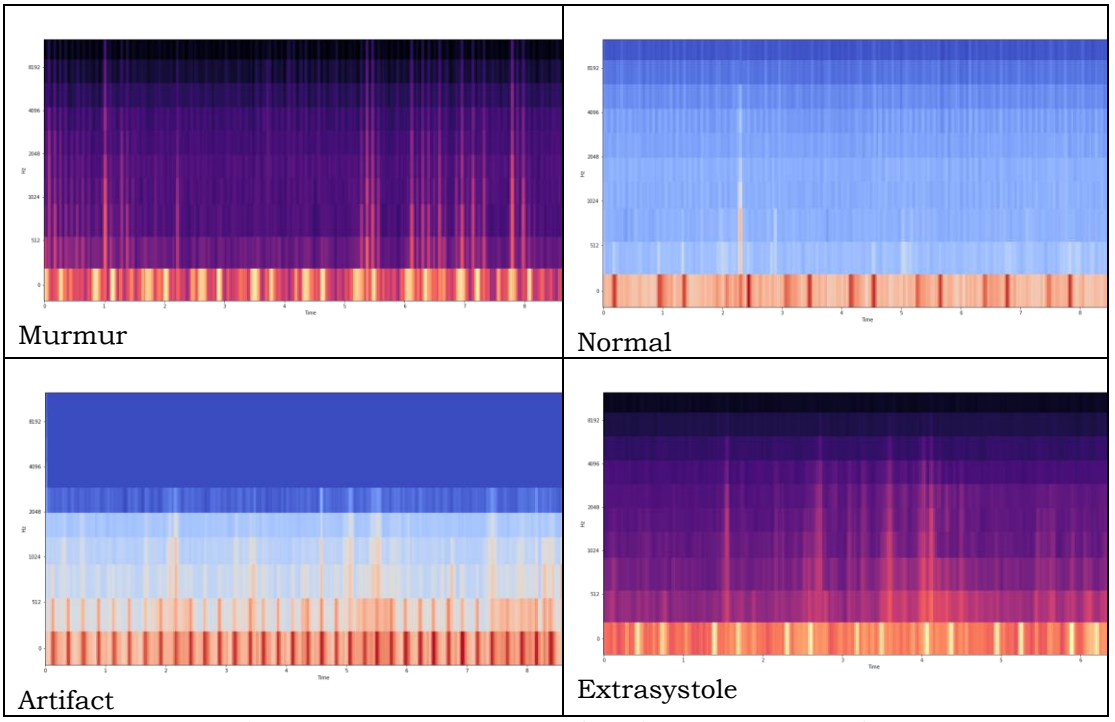

Fig.8.Spectrogram Image of Various Heart Sounds

Speaking of the spectrogram images of the audio files that we generated are both Time and Frequency variant like model. Different ranges of spectrogram images are formed for different sources of heart beat sounds and are sent for the training of the CNN model. The greater the audio files given as input to the CNN model, greater will be the accuracy. In some of the spectrogram images, there was some of pixellation decrement. In order to improve the quality of the images, the generated images from the output of the spectrogram generation were sent to image pixellation improvement. This was done by the online software called Gi. These improved images can effectively increase the quality of the CNN model on its accuracy a lot.

Process In CNN Model

The spectrogram images were sent as coloured, there was no change of their colour gradient of the images generated by them. The converted audio files i.e. the spectrogram images of the respective heart beats are classified as both training and testing data. The number of spectrogram images used for the training dataset was 400 images respectively. The remaining 56 images were used for the testing data. The CNN model is fed up with the training dataset for training the model effectively. In order to make all the spectrogram images more evenly to the total loss, rescale is used and is varied to 1/255 which converts the pixel range from [0, 255] to [0, 1]. The rescale value of both the training and the validating datasets are set to the same value as 1/255. Speaking of the batch size here, its value is set to 32 to 64. The accuracy varied visually a lot while changing the batch size. The higher the value of the batch size made the noise in the gradients to be less or minimum and also made so better in the estimation of the gradient. After the training is complete, the training accuracy and the training losses were generated by the model.

The spectrogram images are really 1D binary label so the value of the class_mode is set to binary. The target size is maintained to be at 200px x 200px. The values of target_size, class_mode and the batch_size for the training and the testing datasets were set the same respectively. The activation is set to relu which will output the input data directly if the value is positive or the output will be zero. As the datasets that are fed to CNN as inputs are 400, the epochs were initially set to 10 and are increased gradually for each and every iterations. Moreover the increase in the epochs doesn't affect the accuracy rate of the output drastically. In this approach of CNN, both the maximum pooling and also the minimum pooling is done and checked with the accuracy of the output of this given model.

Result

After fetching out the outputs of several iterations of both min and max pooling, varying the steps per epochs, using different convolution techniques and combinations, we came to know that on an average the accuracy of the model ranges between 85.3 to 95.2. When we increase the number of epochs and use max pooling for our approach, the accuracy of the CNN model improved a lot and increase to 93.7. On the other hand, when we approach the model with minimum pooling and also with decreased value of the number of epochs, the accuracy decreased with the value of average by 81.6. When a new test data approaches

the model, the test data i.e. the audio file of the heart beat is first converted to spectrogram image. Then the spectrogram image of that audio file is fed to the CNN as the input. The CNN model tests the data with several aspects of combinations, pooling and different combinations are generated. Then the test data is classified as any of the four classifications of heart beat sounds respectively.

Conclusion

On behalf of the model generation, the major approach of this model is to maintain the accuracy in the classification of the different heart sounds given to the CNN model as the input. They should be time efficient and the accuracy should also be maintained at the same time. They decrease the error rate as there is no human interventions. The information are retrived by the model from the heart in fast manner and are classified relatively by the model efficiently. This approach can also be used in the classification of different sounds of stomach, respiration, pancreatic sounds, etc. This work also examined the core relationship between the classification of the CNN model and the effect of min and max pooling, also with several convolution techniques using the Deep Learning algorithm.

References

1. Acharya, J., & Basu, A. (2020). Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning. *IEEE transactions on biomedical circuits and systems*, *14*(3), 535-544.
2. Bahoura, M. (2009). Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes. *Computers in biology and medicine*, *39*(9), 824-843.
3. Cakır, E., Parascandolo, G., Heittola, T., Huttunen, H., & Virtanen, T. (2017). Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(6), 1291-1303.
4. Fitria, F., Ahmad, M., Hatijar, H., Argaheni, N. B., & Susanti, N. Y. (2022). Monitoring combination of intermittent auscultation and palpation of contractions on oxygen saturation of newborns. *International Journal of Health & Medical Sciences*, *5*(3), 221-227. https://doi.org/10.21744/ijhms.v5n3.1930

5. Greco, A., Petkov, N., Saggese, A., & Vento, M. (2020). Aren: A deep learning approach for sound event recognition using a brain inspired representation. *IEEE transactions on information forensics and security*, *15*, 3610-3624.
6. Jakovljević, N., & Lončar-Turukalo, T. (2017, November). Hidden markov model based respiratory sound classification. In *International Conference on Biomedical and Health Informatics* (pp. 39-43). Springer, Singapore.
7. Kochetov, K., Putin, E., Balashov, M., Filchenkov, A., & Shalyto, A. (2018, October). Noise masking recurrent neural network for respiratory sound classification. In *International Conference on Artificial Neural Networks* (pp. 208-217). Springer, Cham.
8. Koutini, K., Eghbal-zadeh, H., & Widmer, G. (2021). Receptive field regularization techniques for audio classification and tagging with deep convolutional neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1987-2000.
9. Lin, X., Liu, J., & Kang, X. (2016). Audio recapture detection with convolutional neural networks. *IEEE Transactions on Multimedia*, *18*(8), 1480-1487.
10. Peeters, G., & Richard, G. (2021). Deep Learning for Audio and Music. In *Multi-Faceted Deep Learning* (pp. 231-266). Springer, Cham.
11. Perna, D. (2018, December). Convolutional neural networks learning from respiratory data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 2109-2113). IEEE.
12. Pramono, R. X. A., Imtiaz, S. A., & Rodriguez-Villegas, E. (2019). Evaluation of features for classification of wheezes and normal respiratory sounds. *PloS one*, *14*(3), e0213659.
13. Rong, F. (2016, December). Audio classification method based on machine learning. In *2016 International conference on intelligent transportation, big data & smart city (ICITBS)* (pp. 81-84). IEEE.
14. Ryu, H., Park, J., & Shin, H. (2016, September). Classification of heart sound recordings using convolution neural network. In *2016 Computing in Cardiology Conference (CinC)* (pp. 1153-1156). IEEE.
15. Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal processing letters*, *24*(3), 279-283.
16. Suryasa, I. W., Rodríguez-Gámez, M., & Koldoris, T. (2021). Health and treatment of diabetes mellitus. *International Journal of Health Sciences*, *5*(1), i-v. https://doi.org/10.53730/ijhs.v5n1.2864
17. Wang, Z., Chen, L., Wang, L., & Diao, G. (2020). Recognition of audio depression based on convolutional neural network and generative antagonism network model. *IEEE Access*, *8*, 101181-101191.