

Article

Heartbeat Sound Signal Classification Using Deep Learning

Ali Raza ¹, Arif Mehmood ¹, Saleem Ullah ^{1,*}, Maqsood Ahmad ¹, Gyu Sang Choi ^{2,*}
and Byung-Won On ³

¹ Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Punjab 64200, Pakistan; meetaliraza@outlook.com (A.R.); arifnhmp@gmail.com (A.M.); maqsood.dba@gmail.com (M.A.)

² Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38542, Korea

³ Department of Software Convergence Engineering, Kunsan National University, Gunsan 54150, Korea; bwon@kunsan.ac.kr

* Correspondence: salimbzu@gmail.com (S.U.); castchoi@ynu.ac.kr (G.S.C.)

Received: 4 October 2019; Accepted: 31 October 2019; Published: 5 November 2019



Abstract: Presently, most deaths are caused by heart disease. To overcome this situation, heartbeat sound analysis is a convenient way to diagnose heart disease. Heartbeat sound classification is still a challenging problem in heart sound segmentation and feature extraction. Dataset-B applied in this study that contains three categories Normal, Murmur and Extra-systole heartbeat sound. In the purposed framework, we remove the noise from the heartbeat sound signal by applying the band filter, After that we fixed the size of the sampling rate of each sound signal. Then we applied down-sampling techniques to get more discriminant features and reduce the dimension of the frame rate. However, it does not affect the results and also decreases the computational power and time. Then we applied a purposed model Recurrent Neural Network (RNN) that is based on Long Short-Term Memory (LSTM), Dropout, Dense and Softmax layer. As a result, the purposed method is more competitive compared to other methods.

Keywords: heart sound; classification; deep learning; RNN

1. Introduction

Before the 19th century, doctors tried to diagnose the disease by listening to the heartbeat sound directly from a patient's chest. This "immediate auscultation" is a very unethical and nonscientific approach for doctors. Laennec invented a medical instrument called the "stethoscope" in 1816 that is used in the medical field to listen to the heartbeat sound. This instrument is widely used in the medical field to diagnose heart disease [1]. Besides, heart failure and coronary heart disease are difficult to diagnose for inexperienced or non-clinical persons by using a stethoscope. Despite accurate auscultating, this requires a long-term practice and several years of experience in a clinic, which is difficult to obtain [2].

The heartbeat sound of a normal heart involves two sound S1 (lub) sound components associated with closing valves systole and S2 (dub) sound component associated with opening valves diastole [3]. Doctors can find heart diseases from listening to the heartbeat by using a stethoscope. A normal heartbeat sound has a clear pattern of "lub dub, dub lub", with the time dub to lub is greater than the time from lub to dub and its rate is 60–100 beats per minute (lub and dub). A murmur heartbeat sound has a noise pattern whooshing, roaring, rumbling or turbulent fluid between lub to dub or dub to lub and symptoms of many heart disease. An extra-systole heartbeat sound has an out of rhythm pattern "lub-lub dub, lub dub-dub" that normally is found in adults and very common in children [4].

Heartbeat sound classification usually comprises three steps. The first step is a pre-processing that cleans the heartbeat signal by passing a band-pass filter to eliminate the noise. The second step is a feature extraction technique that transforms each heartbeat sound signal into a fixed-sized heartbeat sound signal. Then, the Down-sampling [5,6] technique is applied that reduces the heartbeat sound signal frames. However, it helps us to decrease the computation time of our system without affecting the performance of our algorithm performance. The last step is to choose a suitable classifier that extracted features and complete classification tasks. In machine learning, classification algorithms used to classify new data by learns from previous data. Classification algorithms used in this study are Decision Tree (DT), Random Forest (RF) and Linear Support Vector Classifier (LSVC). Deep learning is a subset of Machine Learning. The state-of-the-art methods used in Deep learning algorithms are Multi-Layer Perceptron (MLP) and Recurrent Neural Network (RNN).

The main framework of this study shown in Figure 1 and comprises three steps: pre-processing, feature extraction, and the classification model. This study used Dataset-B from PASCAL Classifying Heart Sounds Challenge [7]. Dataset-B collected from a clinical trial in hospitals using a digital stethoscope (DigiScope). The first step is pre-processing, which cleans the signal by eliminating the undesired frequency. Then, we applied the feature extraction technique, down-sampling [5,6], which decreases the heartbeat sampling rate from 50,000 to 782 frames (divide 8×8 times). Later we applied the classification algorithm, i.e., Decision Tree (DT), Random Forest (RF), Linear Support Vector Classifier (LSVC), Multi-Layer Perceptron (MLP) and Recurrent Neural Network (RNN) in which MLP and RNN are Deep Neural Network. The more efficient classifier is an RNN Deep Neural Network, which gives the highest accuracy as compared to the other classification algorithms DT, RF and LSVC.

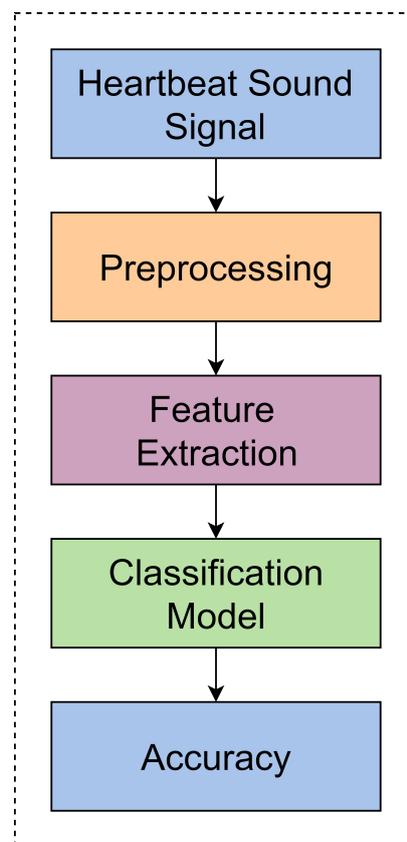


Figure 1. Architecture of this study.

2. Related Work

Many researchers used different techniques, such as segmentation, down-sampling, feature extraction and classification. These techniques are used to predict heart disease. Elsa Ferreria Gomes et al. [7] proposed a system in which they convert the heartbeat signal to segmentation by using Shannon energy also an algorithm used for peak detection. After segmentation, they applied the classification model J48 and MLP algorithm. Unfortunately, they succeed in challenge 1 (segmentation) and does not provide the answer for challenge 2 (classification). J48 and MLP do not provide a good result due to this challenge 2 are not successful.

Elsa Ferreira Gomes et al. [4] describe a methodology for classifying heart sound PASCAL Challenge. They used an algorithm that identified the S1 and S2 heart sound where S1 is lub and S2 is dub. First, they applied the decimate function of MATLAB on the original sound signal and applied a band-pass filter to remove the noise in these signals. After that, they applied the average Shannon energy that is useful to identify the peaks of the heart sound signal easily. They used an algorithm in which they find the maxima and minima points of the sound signal to accomplished the segmentation of heartbeat sound. They used the J48 and MLP algorithm to train the model that predicts the sound signal. These signals are Normal, Murmur, Extra-sound and Artifact.

Yineng Zheng et al. [8] purposed a novel feature for detecting abnormality of heart sound and the features include first and second heart sound (S1-S2EF), energy fraction of heart murmur (HMEF), max energy fraction of heart sound frequency sub-band (HSEFmax), the entropy of S1-S2EF (S1-S2sampen). After that, they have normalized the sound signal then decomposed by wavelet packet. They applied a support vector machine (SVM) for a small dataset that consists of 80 normal heart sound and 167 murmur heart sound. The performance of the purposed method is good because of a small dataset that consists of 2 categories.

Shi-Wen Deng et al. [9] present a framework for heart sound classification based on an auto-correlation feature without using segmentation. The auto-correlation feature extracted from sub-band coefficients of heart signal with Discrete Wavelet decomposition (DWT) after that auto-correlation are fused to get the final features than they applied Support Vector Machine (SVM) on these final features and in Dataset-B [7] they get the precision of Normal (77), Murmur (76) and extra-systole (50) which is not too good.

Wenjie Zhang et al. [10] purpose a method to extract the discriminative feature of heartbeat sound classification by using scaled spectrogram and tensor decomposition. The spectrograms detected the heart cycles of fixed size to extract the discriminative feature. Tensor decomposition is used to reduce the dimension to get a more discriminative feature. After that, they applied the Support Vector Machine (SVM) to these features. The SVM model that gets the normalized precision is 0.74.

Yaseen et al. [11] used multiple feature extraction techniques for heartbeat sound classification. They used an electronics stethoscopes device to get the digital recording of the heart sound called a phonocardiogram (PCG). Mel Frequency Cepstral Coefficient (MFCCs) and Discrete Wavelets Transform (DWT) are the features extraction techniques used in this study. Furthermore, they combine the MFCCs and DWT feature extraction to enhance the results. They used various classification algorithms in this study, such as Support Vector Machine (SVM), Deep Neural Network and K-Nearest Neighbor (KNN).

Many researchers usually involve three steps of heartbeat sound classification. The first step is heart sound segmentation that detects by amplitude threshold-based method [12–14] and probabilistic-based method [15,16]. The second step is feature extraction based on time [17], frequency [18] and time-frequency [10,17,19] and the last step is the classification model. The commonly classification model used is SVM [8–10] and MLP [7] as shown in Table 1.

Deep Neural Network is the most emerging field in data mining. Deep learning is the subset of machine learning that used different types of layers to perform different tasks on data. These tasks are object detection and voice recognition. Many researchers have worked on deep neural networks to perform classification on the audio dataset.

Table 1. Summary of Previous Work on Dataset A & Dataset B.

Authors	Dataset	Model	Overall Results
[7]	Dataset A	MLP	2.12
[7]	Dataset B	MLP	1.67
[9]	Dataset A	SVM	0.74
[9]	Dataset B	SVM	2.03
[10]	Dataset A	SVM	0.76
[10]	Dataset B	SVM	0.74
[10]	Dataset A	2D-PCA	0.75
[10]	Dataset B	2D-PCA	0.68
[10]	Dataset A	MLP	0.59
[10]	Dataset B	MLP	0.62

Shawn et al. [20] have worked on a large-scale audio dataset and proved that its results are good by using Convolutional Neural Network. Yong Xu et al. [21] presented a gated Convolutional Neural Network that won the 1st place in the large-scale weakly supervised sound event detection and Classification of Acoustic Scenes and Events (DCASE) 2017 challenge. Yan Xiong Li et al. [22] used the BLSTM Network on Acoustic Scenes to get a better result. Kele Xu et al. [23] purpose a novel ensemble-learning system consists of CNN that gets a superior classification performance.

Most of the author [20,21,23] used CNN that performs well on image classification also RNN [24] performs well on a series of data. Image data gives the data as a 2D array that consists of the pixel value. Audio data give the data as a 1D array that consists of frequency value. In this study, Dataset-B consists of heartbeat sound samples that converted into 1D arrays. So, that is why the RNN model is used instead of CNN.

3. Methodology

The main framework of this study is shown in Figure 2 and comprises three steps: pre-processing, feature extraction and the classification model. First of all, split the Dataset-B [7] into two sets with a ratio of 70 and 30 percent training and test data. Then we applied the pre-processing techniques that eliminate the noise by applying a band-pass filter. After that, we fixed the size of the heartbeat sound signal by converting sampling frames into a 50,000 frame for each heartbeat sound file. Then we applied Down-sampling techniques that reduce the sampling frame rate 50,000 to 782 frames of the heartbeat sound signal by using decimate techniques. Now the classification model is ready to be applied and we predicted the results of our model.

The main framework of heartbeat sound classification has divided into two sections. Sections (a) and (b) are the training and test phases shown in Figure 2. In the training phase, the data insert as a signal and label then applied the pre-processing technique on these training data and applied feature extraction technique based on data framing and down-sampling. In the last, we have trained the purposed model.

In the test phase, the data inserted as a signal in which applied different techniques. These techniques consist of pre-processing, data framing and down-sampling. In RNN classification, get the trained model from the training phase and predict the heartbeat sound signal. After that, evaluate the purposed model by using the evaluation parameter i.e.accuracy.

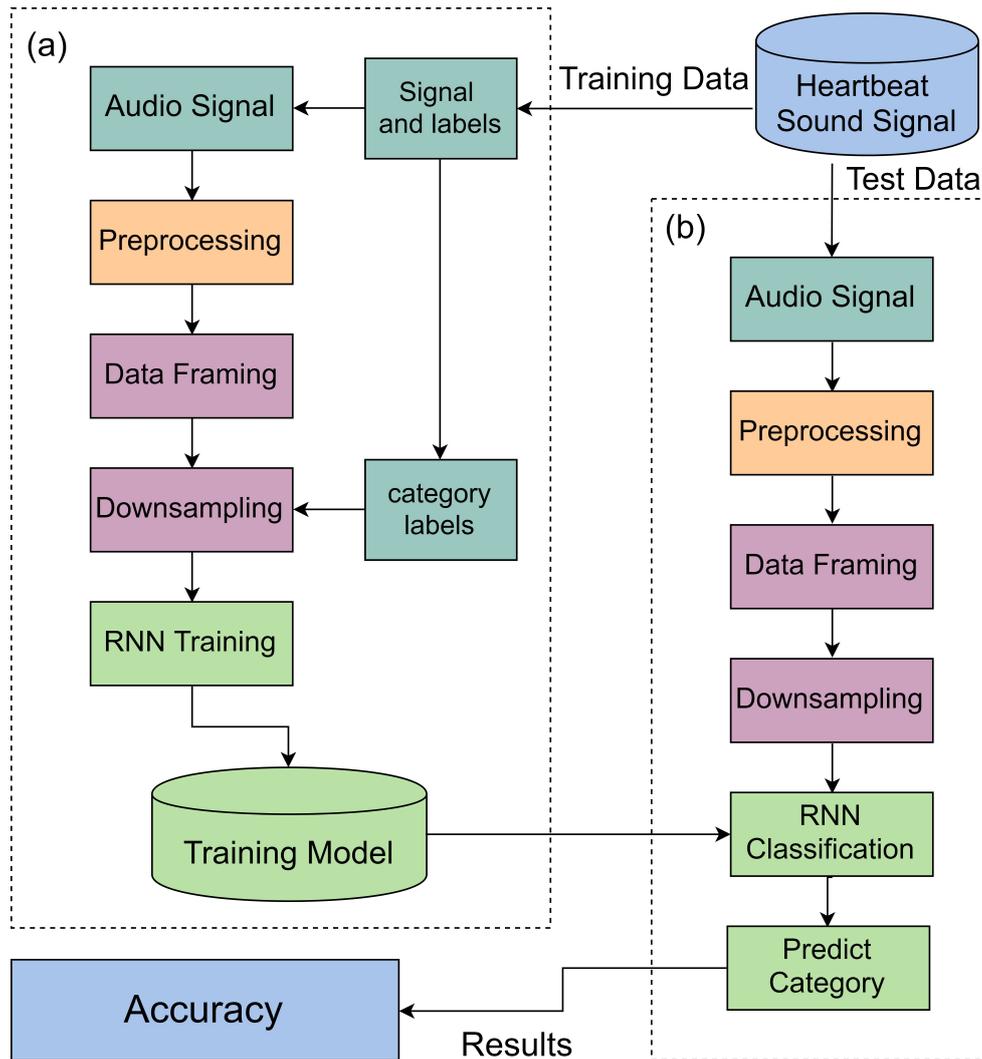


Figure 2. The main Framework of Heartbeat sound Classification, (a) is the training phase and (b) is the test phase.

3.1. Data Collection

Dataset-B [7] used in this study for classifying heart sounds PASCAL challenge competition. This Dataset-B contains 461 samples subdivided into three groups i.e., Normal, Murmur and Extra-systole. Each sample of Dataset-B carries 4000 sampling frequency rates i.e., file 1 contains total sample 26,551 (6.6 s), file 2 contains total samples 38,532 (9.6 s). The Dataset-B is displayed in Table 2.

Table 2. Dataset-B Detail.

Category	No of Samples
Normal	320
Murmur	95
Extra-systole	46

A normal heartbeat sounds like e.g., lub dub or dub lub, and a murmur heartbeat sound has a noise between dub to lub or lub to dub and extra-systole heart sound is out of beats e.g., lub-lub dub or lub dub-dub [4]. These heartbeats sound waves are shown in Figure 3.

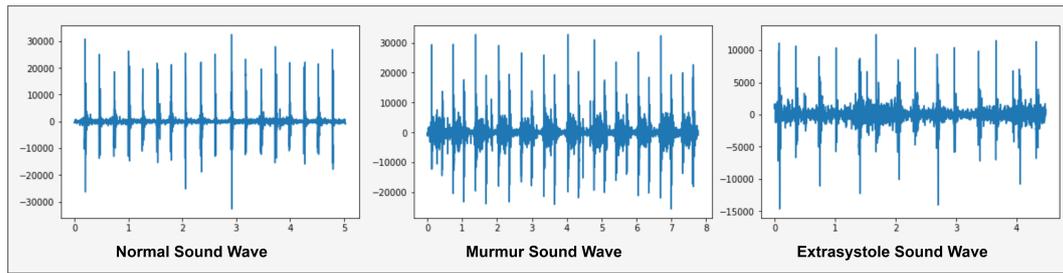


Figure 3. Graphical representation of Heartbeat sound signal.

Figure 3 shows the normal sound wave with a precise shape of lub dub that has no noise. In murmur sound wave, showed a noise between lub to dub or dub to lub and extra-systole has a different shape like a lub-lub dub or lub dub-dub as appearing in Extra-systole sound wave. Spectrogram of these shapes, as presented in Figure 4.

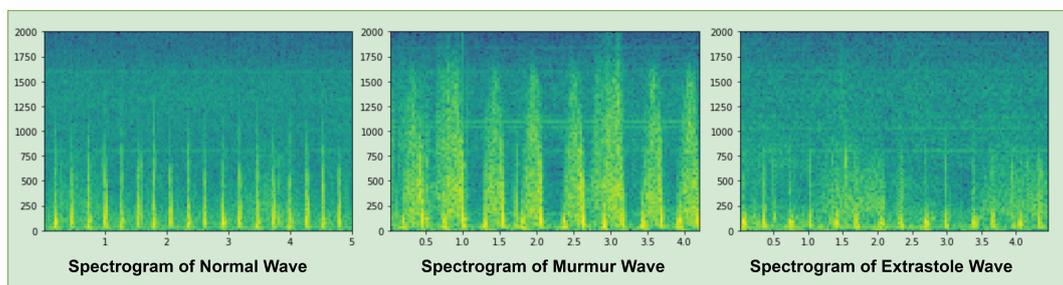


Figure 4. Spectrogram of Heartbeat sound signal.

A spectrogram is a visual representation of the sound signal waves, as shown in Figure 4, that presents three types of spectrograms i.e., Normal, Murmur and Extra-systole waves. Green shades described the amplitude of a sound wave in a spectrogram. The spectrogram of a normal wave is a clear sequence of amplitude i.e., lub dub. In the murmur wave, it shows a noise sequence of amplitude that is greater than normal and extra-systole sound waves. In extra-systole wave, the amplitude of a sound wave is higher than the normal sound wave but smaller than the murmur sound wave.

3.2. Pre-processing

The heartbeat sound signal $x(i)$ takes a 4 kHz sampling frequency rate and then cleaned heartbeat sound signal $x(i)$ with a band-pass, zero-phase filter to get rid of the noise. However, filter the 50–800 Hz frequency to get rid of the noise in the heartbeat sound signal.

$$x'[i] = f(x[i]). \quad (1)$$

where $x'[i]$ is the filtered heartbeat sound wave signal.

3.3. Data Framing

The Dataset-B [7] consists of 461 sound files in which each file is a specific second period with a 4000 sampling (Frame) rate. This sampling rate is the frame values of a sound file within 1-s and gets the overall frames by multiply the sampling rate to the time of a sound file.

$$total\ frame = samplingrate \times time \quad (2)$$

For example, if a sound file f1 has a 5-s time, then the total frame rate of this file can be calculated by this formula.

$$f1 = 4000 \times 5.5 = 22000 \quad (3)$$

Due to the different sound file times, the frame rate of each file is not equal and do not apply any classification algorithm because of different features length. So, data framing is used to fix the sampling rate of each record file. In data framing, converts the sampling frequency of each record file to a 50,000 (12.5 s) frame rate as shown in Figure 5.

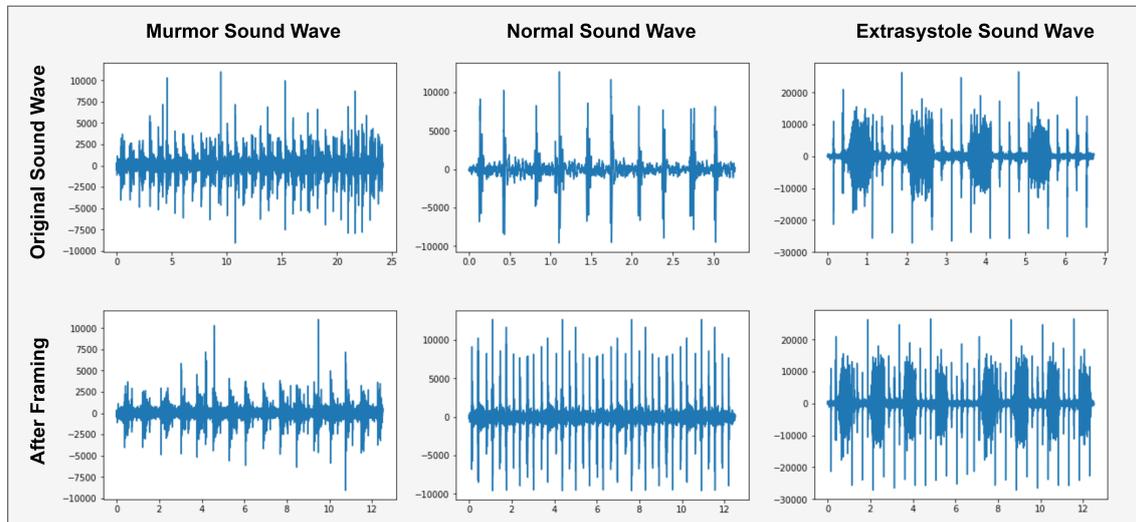


Figure 5. Graphical representation of Data Framing Heartbeat sound signal.

Figure 5 shows the sound samples with a precise time 3-s, 7-s and 25-s. Then we convert each sound sample to 12.5-s (50,000 frames) samples. The sound file of 3-s repeats itself until it reaches 12.5-s. The sound file of 25-s can store only 12.5-s file parts.

Then, we transform each sample of a sound file into a 50,000 (12.5-s) frame rate that converts each sound file into a fixed size. Dataset-B gets the same feature of each sound file after that applied machine learning classification algorithms.

3.4. Down-Sampling

Each sound sample of Dataset-B contains 50,000 frame rates. It is a highly challenging task, dealing with this vast amount of features on our GPU. Now, it takes very high power and time to complete this task. To overcome this situation, reduce the features of a sound file to perform tasks more quickly, in less time.

For this purpose, applied down-sampling techniques [5,6] helped to reduce the sound signal frequency with a minor effect on its performance. Down-sampling (decimate) is a signal reducing technique applied to reduce the features of sound files.

Furthermore, the decimate technique applied to our sound files and its sampling frame rate 50,000 converts into a 782 frame rate by applying a low-pass filter (divide 8×8).

$$x'[i] = \frac{x[i]}{\max(|x[i]|)} \quad (4)$$

where $x'[i]$ is the normalized signal that converted 50,000 frame rate into 782 frame rate of heartbeat sound wave as shown in Figure 6.

We reduce the frame rate of heartbeat sound signal waves using down-sampling to extract the more discriminative features and further use these discriminative features in the classification model. In the x-axis, the features of the sound signal wave convert 50,000 to 800 frames, as displayed in Figure 6. Now we have to apply a classification algorithm in which it takes less computational power and less time to complete the task.

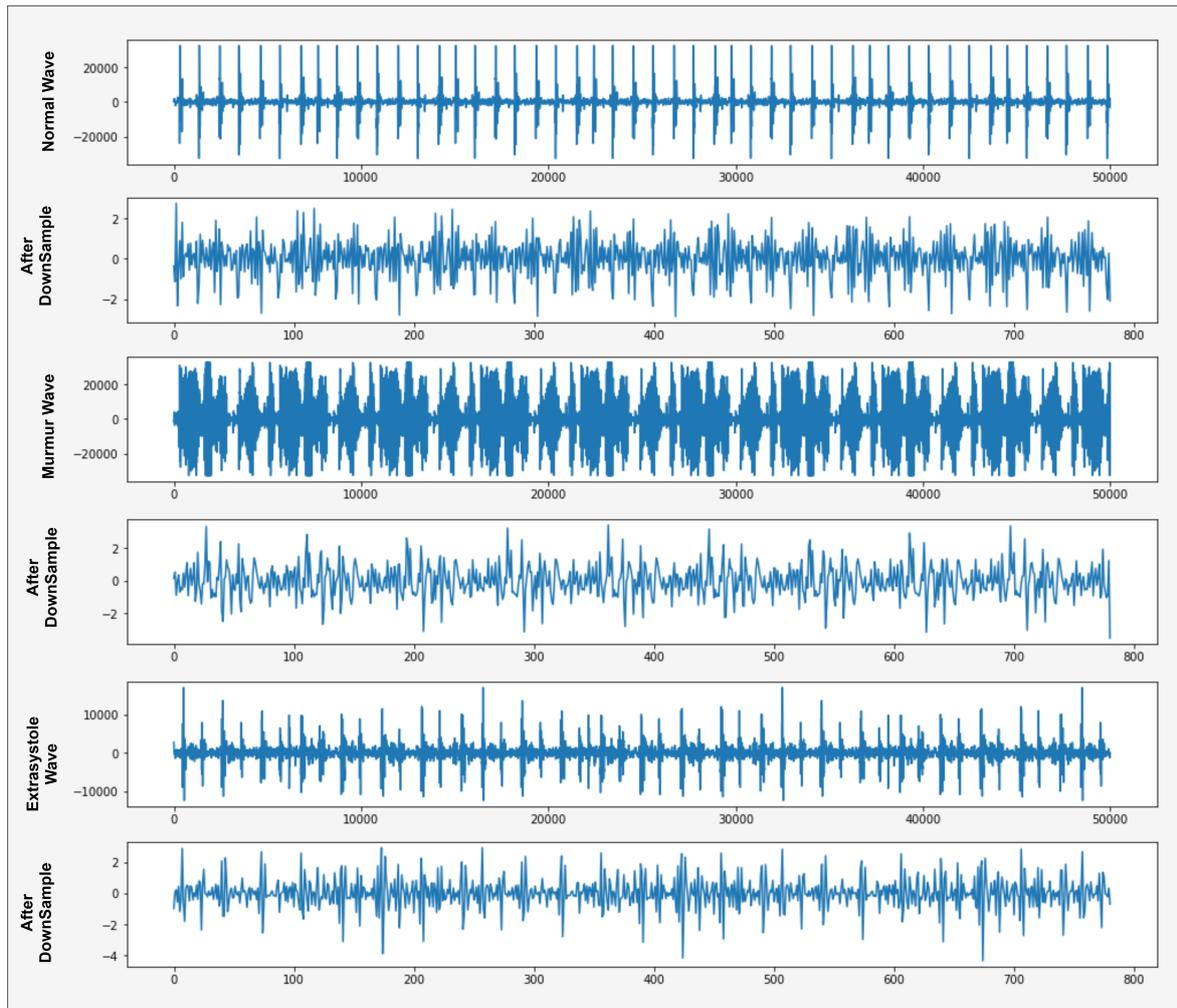


Figure 6. Graphical representation of Down-sampling Heartbeat sound signal.

3.5. Proposed Method

RNN (Recurrent Neural Network) architecture in this research that consists of various layers such that Input, LSTM, Dropout, Dense and Softmax layer. Figure 7 displays the comprehensive architecture of the RNN model.

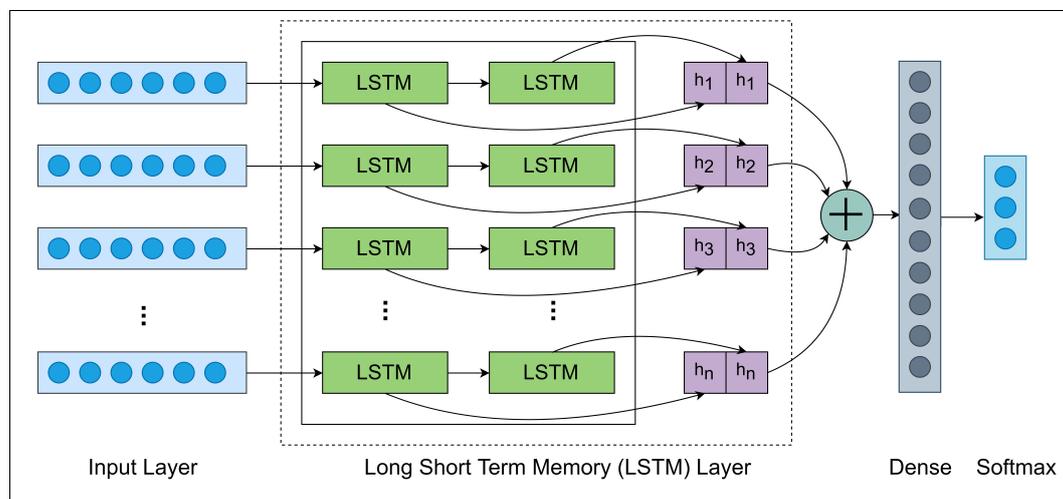


Figure 7. Overall Architecture of Recurrent Neural Network (RNN).

Table 3 displays the dimension and operations of each layer using by the RNN model. Involving the feature extraction techniques, thus it took 782 features of each sound file that passed into an input layer with shape $1 \times 782 \times 1$. In the LSTM layer, the $1 \times 782 \times 1$ shape changes into this $1 \times 782 \times 64$ shape and applied the dropout layer. In the second LSTM layer, the shape has changed into $1 \times 1 \times 32$ then applied dropout layer in which shape has not changed.

Table 3. Dimension and Operations of Layers.

Layer	Operator	Output Height	Output Width	Output Depth
Input	-	1	782	1
LSTM	-	1	782	64
Dropout	Rate = 0.35	1	782	64
LSTM	-	1	1	32
Dropout	Rate = 0.35	1	1	32
Dense	-	1	1	3
Softmax	-	1	1	3

In the dense layer, the shape has converted into $1 \times 1 \times 3$ which helps to classify the dataset because the dataset consists of three category classes. In the second to last layer, we only need three neurons equals to categorize the class length. Then we applied a Softmax layer that is used in the last layer to get the probabilities of each sample.

3.5.1. LSTM Layer

The Long Short-Term Memory [25–27] consists of a connected block known as memory block are a part of the Recurrent Neural Network (RNN) and used in the artificial neural network. Each memory block in LSTM contains three gates, input, output and forget that perform different operations read, write and reset function. More precisely, the input of the cell multiplied by the input gate. The output of the cell multiplied by the output gate and the previous cell value multiplied by the forget gate, as shown in the following equations.

$$f_t = \sigma_s(w_f x_t + U_f h_{(t-1)} + b_f) \quad (5)$$

$$i_t = \sigma_s(w_i x_t + U_i h_{(t-1)} + b_i) \quad (6)$$

$$o_t = \sigma_s(w_o x_t + U_o h_{(t-1)} + b_o) \quad (7)$$

$$c_t = f_t \times c_{(t-1)} + i_t \times \sigma_h(W_c x_t + U_c h_{(t-1)} + b_c) \quad (8)$$

$$h_t = o_t \times \sigma_h(c_t) \quad (9)$$

where f_t is the forget gate activation vector, i_t is the input gate activation vector, o_t is the output gate activation vector, h_t is the hidden state vector, c_t is the cell state vector, σ_s is the sigmoid function, σ_h is the hyperbolic function, w is the weight, b is the bias vector and the initial values of $c_0 = 0$ and $h_0 = 0$.

3.5.2. Dropout Layer

Over-fitting is a big issue in the field of machine learning. It occurs when a classification algorithm trains data and gives satisfying results. Then applied the classification algorithm on test data that gives the unsatisfied result. This happens when two or more neurons detect the same results repeatedly [28] and needs to drop the neurons that affect the results.

$$z_i^L = w_i^L y^L + b_i^L \quad (10)$$

where L is a hidden layer in which $L \in l^1, l^2, l^3, \dots, l^n$, z^L is the input layers, y^L is the vector output, w^L is the weights and b^L is the biases.

3.5.3. Dense Layer

The layer is a set of neurons that get input as a weight, perform some linear function then passes the output to its next layer. The dense layer is the most basic neural network in deep learning that is used to change the dimensions of a layer. In the dense layer, all the neurons connected to the input and output layer.

$$X = f(Y \times w + b) \quad (11)$$

where X is the output layer, Y is the input layer, w is the weight, f is the activation function and b is the bias vector [17].

3.5.4. Softmax Layer

Softmax is an activation function that is extremely important in an artificial neural network and decides whether the neurons are active or not. Softmax is an effective way that handles multi-class classification problems in which output represents in categorical ways [29]. The main goal of Softmax is to highlight the maximum value in neurons. Assigned the maximum neuron weight is one and assigned other neurons weight is zero. The activation function of Softmax defined as

$$S(y_i) = \frac{e^{y_i}}{\sum_k e^{y_k}}, k = 1, 2, \dots, k \quad (12)$$

where y and S is the input and output. The Softmax function is used in the last layer of the neural network to obtain the probabilities of the category class of each input.

3.6. Evaluation Criteria

There is a different type of several measures to evaluate the performance of classifier i.e., Accuracy Precision, Recall, specificity, sensitivity and F-Measure [30]. The evolution criteria used in Dataset-B is accuracy. To measure the accuracy, we need the number of true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) values defined as

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

4. Results and Discussion

The objective of this study is to classify different heartbeat signals automatically. These results give the preliminary diagnoses that help to determine further processes in medicine. The dataset applied in this study is Dataset-B [7] that comprises three categories i.e., Normal, Murmur and Extra-systole. First, converted our Dataset-B into two datasets that consist of two different times i.e., 27.8-s and 12.5-s. Each sample of 27.8-s datasets contains 111,468 frames and 12.5-s dataset contains 50,000 frames.

In this study classification algorithm applies to the Dataset-B. Decision Tree (DT) and Random Forest (RF) is the Machine learning algorithm applied to heartbeat sound dataset to classify the heartbeat diseases. Decision Tree is the earlier classification algorithm used in a diverse area of classification. Random forest is an ensemble learning method used for text classification and generating random decision tree [31]. These algorithms perform well on categorical and text datasets. Their performance on voice and image dataset is unsatisfied. As a result, DT and RF have gained less accuracy displayed in Table 4 as compare to the neural network.

Table 4. Overall Accuracy of 12.5-s and 27.8-s samples.

Classifiers	27.8-s	12.5-s
Decision Tree	57.5	48.9
Random Forest	68.3	71.2
Multi Layer Perceptron (MLP 6 Layer)	66.1	67.6
Multi Layer Perceptron (MLP 16 Layer)	67	69
Recurrent Neural Network (RNN)	77.2	80.8

Deep learning is an artificial neural network that acts like a human brain to learn and make a decision on its own. For classifying audio and image datasets, authors [20,21,32,33] prefers the deep learning algorithm and their performance very well. Deep learning is the subset of machine learning, as well as machine learning, which is a subset of artificial intelligence. In deep learning, data passes through each layer. That is the output of the previous layer and input to its next layer. The first and last layer is called the input and output layer. Another layer between the input and output layer is considered to be a hidden layer in which each layer performs an activation function. The deep learning algorithm extracts the discriminative feature itself but extracts the feature separately in the machine learning algorithm. The deep learning algorithm extracts the feature itself. However, machine learning needs to extract the feature separately.

Multi-Layer Perception is a feed-forward network consisting of input, output and hidden layers. It trained by back-propagation where the weight connected with layer and mostly used for regression and classification problems [34]. It is the simple neural network that uses Dense, Dropout and Softmax layer. In this study, Multi-Layer Perceptron gained 67% and 69% accuracy by using 6 and 16 layers. Increasing the hidden layers in Multi-Layer Perceptron also increases the accuracy. But at some level, increasing the layers does not affect the accuracy anymore.

Recurrent Neural Network (RNN) gains the highest accuracy in our study based on Long Short Term Memory (LSTM), Dropout, Dense and Softmax layers. LSTM layer is used for time-series analysis and remember the past data in a memory cell. Each time step, a few gates are used to deal with the passing of information along the sequence that can capture long-range information more precisely [35,36]. This is the reason that our purposed model has gained the highest accuracy because it remembers the previous information for a long period that connects to the present task. Adam optimizer used in our RNN model that gives the best result compared to other optimizers Adagrad and SGD. The dropout layer is an effective way to avoid over-fitting because it randomly drops some of the connections between layers that help to prevent over-fitting [37]. Furthermore, our hyper-parameter model tuned here by applying the different dropout rates {0.05, 0.20, 0.35}. Each dropout rates applied to our RNN model that gives the different results shown in Table 5. Hence, it shows that increasing the dropout rate increases the accuracy and decreases the loss. 0.35 is the ideal dropout rate that gives the best results in our model.

Table 5. Hyper-parameter tuning of 12.5-s samples by Applying different Dropout rates.

Dropout Rates	Accuracy	Loss
0.05	76.10	52.30
0.20	77.93	49.59
0.35	80.81	47.05

During the training of the 12.5-s dataset of our purposed model, RNN used 322 samples in which 289 samples used for training and 33 samples used for test data.

The validation accuracy and loss in training shown in Figure 8 help to find out the model working is fine or not. If validation loss increases and validation accuracy decreases, that means our model is not in the learning state. If validation loss and accuracy are increases, that means our model is over-fitting. If validation loss decreases and validation accuracy increases, then it means model learning and working is fine. Figure 8 shows increases validation accuracy and decreases validation

loss. It means our model learning is fine and gained the highest accuracy as compared to other classification algorithms.

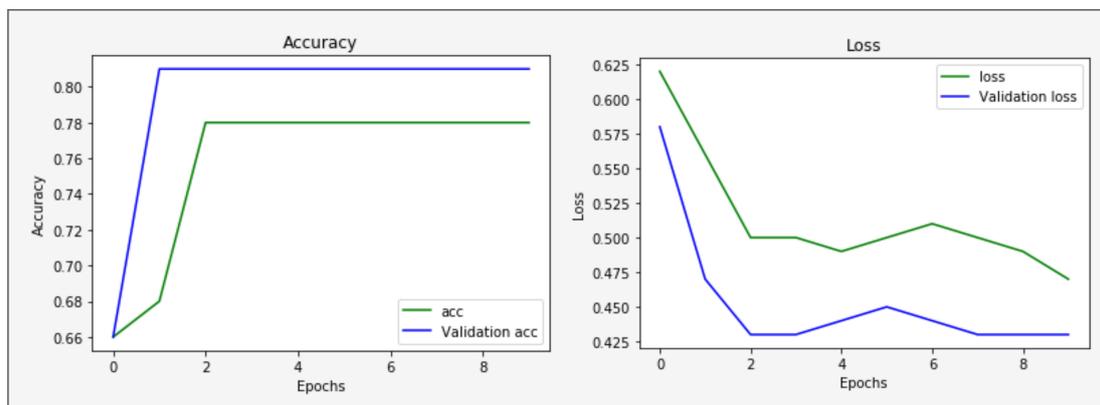


Figure 8. Validation accuracy and loss during the training of the RNN model.

Figure 9 shows the results in which the Recurrent Neural Network (RNN) gained the highest accuracy on both 12.5-s and 27.8-s sample files in Dataset-B. RNN gained 80.8 Accuracy with 12.5-s sample files and 77.2 Accuracy with 27.8-s sample files in Dataset-B. Decision Tree gained the lowest accuracy in both 12.5-s and 27.8-s sample files in Dataset-B. It also shows that the overall results of 12.5-s sample files are higher than the results of 27.8-s sample files. Recurrent Neural Network (RNN) is the overall best performing classification algorithm in our model.

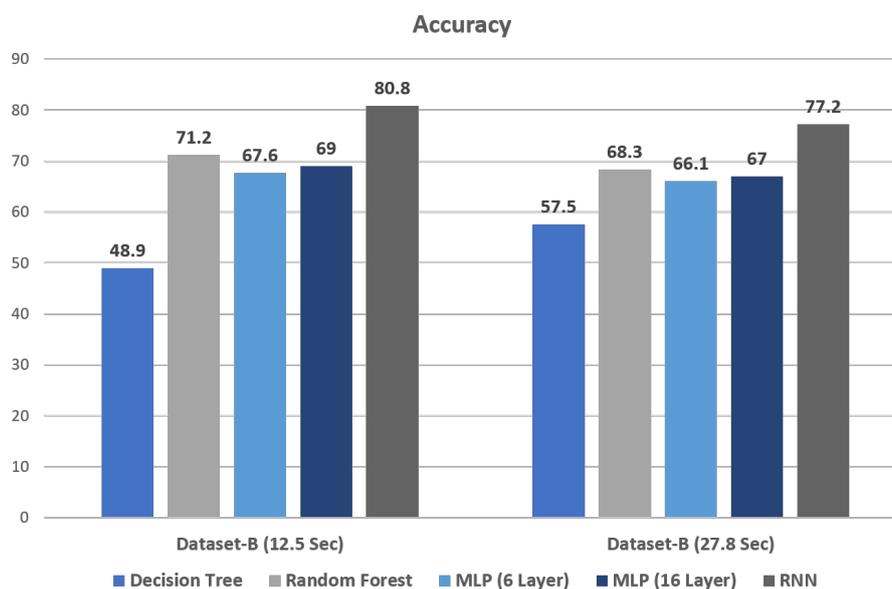


Figure 9. Accuracy Score comparison among 12.5-s and 27.8-s samples.

K-Fold Cross-Validation is a re-sampling technique used to evaluate the classification algorithm on limited data samples. The result of K-Fold Cross-Validation is less biased or less optimistic compared to other methods, such as train/test split. K represents the number of folds is randomly partition of k equal-sized sub-samples. The 5-fold cross-validation is used in our classification algorithm to estimate the performance of the RNN model. Figure 10 shows the result of 5-fold cross-validation in which the RNN model gained the weighted accuracy and loss is 80.45 and 47.37.

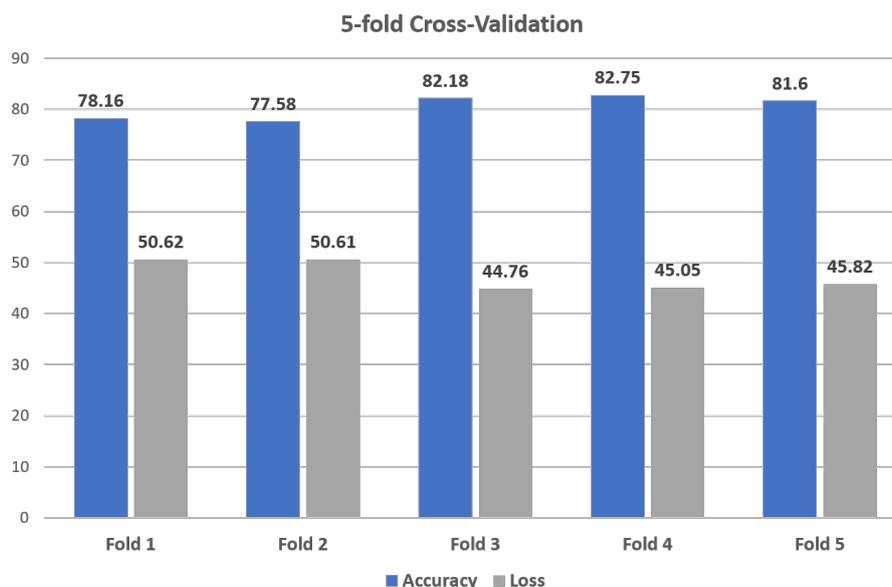


Figure 10. K-Fold Cross-Validation of 12.5-s samples.

The experimental results indicate that our proposed method has gained good results as compared to the previous work. Gomes et al. [7], Shi-Wen Deng et al. [9] and Wenjie Zhang et al. [10] gained the overall result on Dataset-B is 0.70 (MLP), 0.74 (SVM) and 0.76 (SVM). Our method has a good ability to extract the discriminative feature for heartbeat sound classification that gained the highest result is 80.8.

5. Conclusions

This research proposed a deep learning model for heartbeat sound classification based on data framing, Down-sampling, and RNN for Dataset-B. This proposed method can efficiently detect the heartbeat signal and gives information for deciding whether further treatment is necessary or not. Dataset-B divided into three categories Normal, Murmur and Extra-Systole. The heartbeat sound signal is cleaned by filtering the noise. The data framing converts the sampling frame rate of each audio file into a fixed-size frame rate than the down-sampling decrease the dimension of the heartbeat sound signal wave to extract the more discriminative features. The RNN proposed model was applied to the Dataset-B in this study that gained the highest accuracy 80.80. The experiment shows that our method is more competitive and efficient.

In future work, we will apply spectrogram techniques to get more discriminative features and applied other deep learning techniques like Convolutional Neural Network (CNN).

Author Contributions: All of the authors contributed in literature search, figures, study design, data collection, data analysis, data interpretation, and writing etc. Such as, A.R. and A.M. designed and carried out the experimental work. A.R., A.M., S.U., M.A., B.-W.O. and G.S.C. analyzed the data and interpreted the results. A.R. and A.M. drafted the manuscript with the input from the others. G.S.C. and A.M. supervised the project.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1A2C1006159), and MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2019-2016-0-00313) supervised by the IITP(Institute for Information & communications Technology Promotion), and the 2019 Yeungnam University Research Grant.

Acknowledgments: The authors would like to thank the Editor and the anonymous reviewers for their insightful comments and constructive suggestions that certainly improved the quality of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hanna, I.R.; Silverman, M.E. A History of Cardiac Auscultation and Some of Its Contributors. *Am. J. Cardiol.* **2002**, *90*, 259–267. [[CrossRef](#)]
2. Jiang, Z.; Choi, S. A Cardiac Sound Characteristic Waveform Method for In-Home Heart Disorder Monitoring with Electric Stethoscope. *Expert Syst. Appl.* **2006**, *31*, 286–298. [[CrossRef](#)]
3. Kumar, D.; Carvalho, P.; Antunes, M.; Gil, P.; Henriques, J.; Eugenio, L. A New Algorithm for Detection of S1 and S2 Heart Sounds. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings, Toulouse, France, 14–19 May 2006; Volume 2, pp. 1180–1183.
4. Gomes, E.F.; Pereira, E. Classifying Heart Sounds Using Peak Location for Segmentation and Feature Construction. *Aistats* **2012**, 1–5.
5. Díaz-García, J.; Brunet, P.; Navazo, I.; Vázquez, P.P. Downsampling Methods for Medical Datasets. In Proceedings of the International Conferences on Computer Graphics, Visualization, Computer Vision and Image Processing 2017 and Big Data Analytics, Data Mining and Computational Intelligence 2017 - Part of the Multi Conference on Computer Science and Info, Lisbon, Portugal, 23 July 2017; pp. 12–20.
6. Genussov, M.; Cohen, I. Musical Genre Classification of Audio Signals Using Geometric Methods. *Eur. Signal Process. Conf.* **2010**, *10*, 497–501.
7. Gomes, E.F.; Bentley, P.J.; Coimbra, M.; Pereira, E.; Deng, Y. Classifying Heart Sounds: Approaches to the PASCAL Challenge. In Proceedings of the HEALTHINF 2013- Proceedings of the International Conference on Health Informatics, Barcelona, Spain, 11–14 February 2013; pp. 337–340.
8. Zheng, Y.; Guo, X.; Ding, X. A Novel Hybrid Energy Fraction and Entropy-Based Approach for Systolic Heart Murmurs Identification. *Expert Syst. Appl.* **2015**, *42*, 2710–2721. [[CrossRef](#)]
9. Deng, S.W.; Han, J.Q. Towards Heart Sound Classification without Segmentation via Autocorrelation Feature and Diffusion Maps. *Future Gener. Comput. Syst.* **2016**, *60*, 13–21. [[CrossRef](#)]
10. Zhang, W.; Han, J.; Deng, S. Heart Sound Classification Based on Scaled Spectrogram and Tensor Decomposition. *Expert Syst. Appl.* **2017**, *84*, 220–231. [[CrossRef](#)]
11. Yaseen, G.Y.S.; Kwon, S. Classification of Heart Sound Signal Using Multiple Features. *Appl. Sci.* **2018**, *8*, 2344. [[CrossRef](#)]
12. Chen, T.; Kuan, K.; Celi, L.; Clifford, G.D. Intelligent HeartSound Diagnostics on a Cellphone Using a Hands-Free Kit. *AAAI Spring Symp. Ser.* **2010**, *2010*, 26–31.
13. Liu, Y.; Poon, C.C.Y.; Zhang, Y.T. A Hydrostatic Calibration Method for the Design of Wearable PAT-Based Blood Pressure Monitoring Devices. In Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'08 - "Personalized Healthcare through Technology, Vancouver, BC, Canada, 22–25 August 2008; Volume 24, pp. 1308–1310.
14. Moukadem, A.; Dieterlen, A.; Hueber, N.; Brandt, C. A Robust Heart Sounds Segmentation Module Based on S-Transform. *Biomed. Signal Process. Control* **2013**, *8*, 273–281. [[CrossRef](#)]
15. Schmidt, S.E.; Holst-Hansen, C.; Graff, C.; Toft, E.; Struijk, J.J. Segmentation of Heart Sound Recordings by a Duration-Dependent Hidden Markov. *Model. Physiol. Meas.* **2010**, *31*, 513–529. [[CrossRef](#)] [[PubMed](#)]
16. Springer, D.B.; Tarassenko, L.; Clifford, G.D. Logistic Regression-HSMM-Based Heart Sound Segmentation. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 822–832. [[CrossRef](#)] [[PubMed](#)]
17. Ari, S.; Hembram, K.; Saha, G. Detection of Cardiac Abnormality from PCG Signal Using LMS Based Least Square SVM Classifier. *Expert Syst. Appl.* **2010**, *37*, 8019–8026. [[CrossRef](#)]
18. Safara, F.; Doraisamy, S.; Azman, A.; Jantan, A.; Abdullah Ramaiah, A.R. Multi-Level Basis Selection of Wavelet Packet Decomposition Tree for Heart Sound Classification. *Comput. Biol. Med.* **2013**, *43*, 1407–1414. [[CrossRef](#)]
19. Thiyagaraja, S.R.; Dantu, R.; Shrestha, P.L.; Chitnis, A.; Thompson, M.A.; Anumandla, P.T.; Sarma, T.; Dantu, S. A Novel Heart-Mobile Interface for Detection and Classification of Heart Sounds. *Biomed. Signal Process. Control* **2018**, *45*, 313–324. [[CrossRef](#)]
20. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN Architectures for Large-Scale Audio Classification. In Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 5–9 March 2017; pp. 131–135.

21. Xu, Y.; Kong, Q.; Wang, W.; Plumbley, M.D. Large-Scale Weakly Supervised Audio Classification Using Gated Convolutional Neural Network. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 121–125.
22. Li, Y.; Li, X.; Zhang, Y.; Wang, W.; Liu, M.; Feng, X. Acoustic Scene Classification Using Deep Audio Feature and BLSTM Network. In Proceedings of the ICALIP 2018—6th International Conference on Audio, Language and Image Processing, Shanghai, China, 16–17 July 2018; pp. 371–74.
23. Xu, K.; Zhu, B.; Kong, Q.; Mi, H.; Ding, B.; Wang, D.; Wang, H. General Audio Tagging with Ensembling Convolutional Neural Networks and Statistical Features. *J. Acoust. Soc. Am.* **2019**, *145*, EL521–EL527. [[CrossRef](#)]
24. Keren, G.; Schuller, B. Convolutional RNN: An Enhanced Model for Extracting Features from Sequential Data. In Proceedings of the International Joint Conference on Neural Networks, Vancouver, BC, Canada, 24–29 July 2016; pp. 3412–3419.
25. Gastaldo, P.; Redi, J. Machine Learning Solutions for Objective Visual Quality Assessment. In Proceedings of the 6th International Workshop on Video Processing and Quality Metrics for Consumer Electronics. VPQM, Scottsdale, AZ, USA, 19–20 January 2012; Volume 12, pp. 2451–2471.
26. Graves, A.; Schmidhuber, J. Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)]
27. Hochreiter, S.; Schmidhuber, J.U. Lstm. *Neural Comput.* **1997**, *9*, 1735–1780. Available online: <http://www7.informatik.tu-muenchen.de/~hochreit%0Ahttp://www.idsia.ch/~juergen> (accessed on 15 November 1997). [[CrossRef](#)]
28. Nitish, S.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
29. Yu, M.; Huang, Q.; Qin, H.; Scheele, C.; Yang, C. Deep Learning for Real-Time Social Media Text Classification for Situation Awareness—Using Hurricanes Sandy, Harvey, and Irma as Case Studies. *Int. J. Dig. Earth* **2019**, 1–18. [[CrossRef](#)]
30. Sokolova, M.; Japkowicz, N.; Szpakowicz, S. AI 2006: Advances in Artificial Intelligence. In Proceedings of the 19th Australian Joint Conference on Artificial Intelligence, Hobart, Australia, 4–8 December 2006.
31. Kowsari, K.; Meimandi, K.J.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text Classification Algorithms: A Survey. *Information* **2019**, *10*, 1–68. [[CrossRef](#)]
32. Minaee, S.; Abdolrashidi, A. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. *arXiv* **2019**, arXiv:1902.01019.
33. Minaee, S.; Bouazizi, I.; Kolan, P.; Najafzadeh, H. Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection In Videos. *arXiv* **2018**, arXiv:1806.08612.
34. Ravi, V.; Pradeepkumar, D.; Deb, K. Financial Time Series Prediction Using Hybrids of Chaos Theory, Multi-Layer Perceptron and Multi-Objective Evolutionary Algorithms. *Swarm Evol. Comput.* **2017**, *36*, 136–149. [[CrossRef](#)]
35. Zhao, Z.; Chen, W.; Wu, X.; Chen, P.C.Y.; Liu, J. LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast. *IET Intell. Transp. Syst.* **2017**, *11*, 68–75. [[CrossRef](#)]
36. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors* **2017**, *17*, 273. [[CrossRef](#)] [[PubMed](#)]
37. Reimers, N.; Gurevych, I. Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks. *arXiv* **2017**, arXiv:1707.06799.

