Journal of the Medical Research Institute

JMRI, 2022; Vol. 43 No. 1: (1-7)

Journal of the Medical Research Institute

Automatic Diagnosis of Heart Sounds Using Bark Spectrogram Cepstral Coefficients

Mohammed Mostafa Azmy[1].

1) Department of Biomedical Engineering, Medical Research Institute, Alexandria University, Egypt.

ABSTRACT:

Auscultation of heart sounds is an essential step for diagnosing heart diseases. Automatic auscultation using computer techniques is used to help physicians in their diagnosing process. Several researches are conducted for analyzing heart sounds using computer techniques. In this paper, new methods of extracting features from heart sounds are presented using Bark spectrogram cepstral coefficients (BSCC) and Mel-spectrogram cepstral coefficients (MSCC). Classification of normal and abnormal heart sounds are based on support vector machine or deep learning neural networks. Database of heart sounds are selected from Physionet challenge database. Signals of heart sounds are detrended and normalized. Then, wavelet transform is applied. After that, energy entropy is calculated. Then BSCC are considered. The classifiers of support vector machine (SVM) and deep learning (DL) with bi-long short term memory (BILSTM) are applied. The maximum obtained accuracy rate of 99.54% is achieved by using BSCC algorithm. The maximum obtained area under curve (AUC) of 0.9846 is achieved when using BSCC algorithm.

Keywords: PCG; MFCC; BFCC; MSCC; BSCC; DL; BILSTM; SVM

1. INTRODUCTION

Heart diseases are the most cause of death around the world. Auscultation of heart sounds is an effective method for diagnosing heart diseases WHO [1]. Automatic auscultation by using computer techniques saves time and efforts of physician. Many papers discussed analysis and recognition of heart sounds using several methods. Shamsuddin N et al. (2005) [2]

Classified heart sounds using mutilayer feed forward neural network. They obtained 100% of correct classification of 11 heart diseases. Garzon JJ and others (2008) (3) used support vector regression to detect murmurs. They obtained 97.85% of accuracy of normal and pathologic phonocardiogram (PCG) signals. Maglogiannisa I and hiscolleagues (2009) (4) used wavelet and SVM to classify heart sounds. Yana Z et. al. (2010) (5) of investigate moment segmentations of heart sounds. Hu X et al. (2011) (6) extracted features from PCG signals using Hilbert transfer. The overall accuracy obtained was 91.95%. Kwak Cand Kwon OW (2012) (7) applied classification of heart sounds using hidden markov models. The maximum accuracy of 85.6% obtained using SVM. Tanga H et al. (2012) (8) used dynamic clustering to make segmentation of heart sounds. Moukadem A and his colleagues (2013) (9) segmented heart sounds using S-transform. Patidar S and Pachori RB (2014) (10) classified heart sounds using wavelet transform. They obtained accuracy of 94.01%. Zheng Y and his colleagues (2015) (11) classified heart sounds into normal and abnormal using SVM. They extracted features from heart sounds by calculating energy fraction and energy entropy of them. They obtained 97.17% of accuracy when comparing between 80 normal heart sounds and 167 systolic heart murmurs. Zahhad MA et al. (2016) (12) extracted features from heart sounds by wavelet packet cepstral coefficients and used it as biometric application. Lubaib P and Muneer KVA (2015) (13) extracted features from heart sounds by calculating mean, energy, variance of MFCC of heart sounds using different classifiers. The maximum obtained accuracy of99.99% was by SVM. Zhang W et al. (2017) (14) obtained features from heart sounds by calculating spectrogram and partial least squares regression and classifying them using SVM. Hamidi M et al. (2019) (15) classified heart sounds using fractal dimension and curve fitting. The maximum obtained accuracy was 98%. Lubis C and Gondawijaya F (2018)(16) obtained features from heart sounds using MFCC, modified MFCC, BFCC and modified BFCC based on back propagation neural network. The obtained accuracy was 95.83% using modified MFCC. Das S et al. (2019) (17) identified fundamental heart sound by using Gamma tone filter bank energy. Chowdhury TH and others (2020) (18) used time frequency analysis to classify, analyze, segment and compress PCG signals. The obtained accuracy was 97.10% using deep learning. Shuvo SB et al. (2021) (19) used novel deep learning algorithm for classification of cardiovascular disease using recording of heart sounds. The obtained accuracy was 99.60%. In this paper, features are extracted from heart sounds using wavelet transform then energy entropy then BSCC. They are classified by deep learning which achieve 99.54% of accuracy. Methods of feature extraction are discussed as wavelet transform and spectrogram. Feature extraction and classification methods used in this work are introduced in section 2. In section 3, results are explained. In section4, discussion is presented. Finally conclusions are presented in section 5

2. Materials &Methods:

2.1 Database

Signals of heart sounds HSs are selected from the PhysioNet 2016 Challenge (20). These signals are sampled at 2000 Hz. The collected signals are from training type e database. Heart sounds signals are divided into 2 classes: normal heart sounds (N) and abnormal heart sounds (A). The number of normal heart sounds is 2423 patients. The number of abnormal heart sounds is 469 patients.

2.2 Wavelet Transform

Wavelet transform is suitable for heart sounds because they are non-stationary signals. Wavelet transform presented in time-frequency analysis method so it is preferred over Fourier transform which is a frequency analysis method only. Wavelet transforms look like collection of band-pass filters. Here, signals are decomposed into different bands. Lower and higher frequency components of the decomposed signal are computed by low and high band pass filers. The detailed coefficients are calculated from the series of high pass filters. The approximation coefficients are computed from the series of low pass filters. The selected mother wavelet is chosen to be almost like signals of heart sounds, Goa RX and Yan RQ (2011) (21) Theodoridis S and his colleage (2003) (22). In this work, the used mother wavelets are Daubechies (db7). The number of selected levels is four levels. Equation (1) shown below describes wavelet transform for signal x(t).

$$WT_x^{\varphi}(\tau,s)=\varphi_x^{\varphi}(\tau,s)=\frac{1}{\sqrt{|s|}}\int x(t)\cdot\varphi_x^{*}\left(\frac{t-\tau}{s}\right)dt \quad (1)$$

Where φ is the mother wavelet, x(t) is the function of which wavelet transform is obtained, s is scale s>0, τ is translation value.

2.3 Spectrogram

Spectrogram is the Short-Time Fourier Transform (STFT) of a signal. Fourier transform describes signal in the frequency domain only. STFT overperform Fourier transform in describing the signal in the time domain and frequency domain by using a small section of Fourier transform of the signal at a time which is called windowing the signal, Vapnik VN (1998) (23). The used window here is Hamming window. The number of frequency points used to calculate the discrete Fourier transforms is equal to the maximum of 256 points. Equation 2 defines STFT of a signal

$$STFT_x^{(\omega)}(t',f)=\int[x(t)\cdot\omega^{*}(t-t')]\cdot e^{-i2\pi ft}dt \quad (2)$$

Where ω (t) is the window function, x(t) is the function of which STFT is calculated, t' is the time index.

2.4 Feature Extraction Techniques

2.4.1 MFCC

Mel-frequency Cepstral Coefficient (MFCC) is widely used for obtaining features of human speech as shown in Fig.1. MFCC can distinguish between redundant noise and energy of audio signal components and their frequencies. The scale of Mel is logarithmic scale which is suitable for human perception of frequency. MFCC is suitable for extracting features of heart sounds which are heard by humans also. Computation of MFCCs includes pre-emphasizing, framing and windowing of the input signal. Then, Fast Fourier Transform is calculated as shown in equation 3. Then, these coefficients are converted to Mel scale as shown in equation 4-6. After that, these vectors are logarithmized. Then, DCT is computed to remove redundant information as shown in equation 7. Finally, cepstral coefficients are selected, Lubaib P and Muneer KVA (2015) (13)

$$X(k)=\sum_{n=0}^{N-1}x(n)e^{\frac{-j2\pi nk}{N}} \qquad 0\le k\le N-1 \quad (3)$$

$$Mel(f)=2595.\log(1+f/700) \quad (4)$$

$$s(m)=\sum_{k=0}^{N-1}[|X(k)|^2H_m(k)]; \qquad 0\le m\le M-1 \quad (5)$$

Where M is total number of triangular mel weighting filters, H(k) is the weight giving to the kth energy spectrum bin contributing to the mth output and can be calculated as:

$$H_m(k)=\begin{cases}0 & k<f(m-1)\\ \frac{2(k-f(m-1))}{f(m)-f(m-1)} & f(m-1)\le k\le f(m+1)\\ \frac{2(f(m+1)-k)}{f(m+1)-f(m)} & f(m)<k\le f(m+1)\\ 0 & k>f(m+1)\end{cases} \quad (6)$$

M can be changed from 0 to M-1

$$c(n)=\sum_{m=0}^{M-1}\log(s(m))\cos\left(\frac{\pi n(m-0.5)}{M}\right);$$

$$n=0,1,2,\dots\dots,C-1 \quad (7)$$

Where c(n) are the cepstral coefficients, C is the number of MFCCs.

2.4.2 MSCC

Mel-Spectrogram Cepstral Coefficient (MSCC) is a new method of feature extraction as shown in Fig.2. It is derived from MFCC. MSCC is computed by pre-emphasizing, framing and windowing of the input signal. Squared spectrogram of the windowed framed of the input energy signals is calculated and then converted to the Mel scale. Then, these vectors are logarithmized. Finally, DCT is applied. Then the 1st 13cepstral coefficients are selected, Lubaib P and Muneer KVA (2015) (13).

2.4.3 BFCC

Bark Frequency Cepstral Coefficient (BFCC) as shown in Fig.3is implemented by pre-emphasizing and framing of the input signals. Fast Fourier transform of the windowed framed of the input signals is calculated and then converted to the bark scale as shown in equation 8. Then, these vectors are logarithmized. Finally, DCT is applied. Then the 1st 13 cepstral coefficients are chosen, Lubis C and his colleage (2018) (16). All equations of calculating BFCCs are the same equations of MFCC from equation (3) to equation (7) except equation (4) is replaced by equation (8).

$$Bark(f)=\{13.atan(0.76f/100)+3.5\} \tag{8}$$

2.4.4 BSCC

Bark Spectrogram Cepstral Coefficient (BSCC) is a new method of feature extraction as shown in Fig.4. BSCCs are implemented by pre-emphasizing and framing of the input signal. Squared spectrogram of the windowed framed is calculated and then converted to the bark scale. Then, these vectors are logarithmized. Finally, DCT is applied. Then the 1st 13 cepstral coefficients of DCT are selected, Lubis C and his colleage (2018) (16).

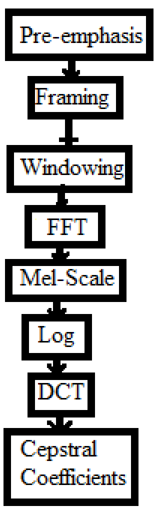

Fig1 MFCC

Fig.2 MSCC

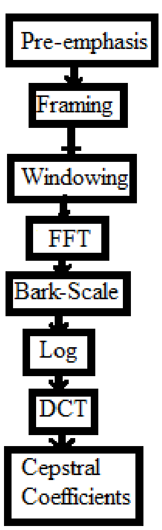

Fig.3 BFCC

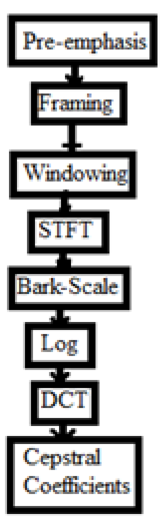

Fig.4 BSCC

2.5 Classification Methods

2.5.1 Support Vector Machine

Support vector machine (SVM) is a classifier that was created by Vapnik VN (1998) (23). SVM is designed for classifying two classes' problems. Now, SVM can be used for multiclass problems after some modifications. It can be also used for recognizing nonlinearity separable classes. If the system is classified into two classes which have group of points, a linear SVM will look for the hyperplane putting the biggest possible fraction of points of the identical class on the same side, but the distance of either class will be maximized from the hyperplane. Kernel function of SVM can be changed to be 1st, 2nd or 3rd order to achieve maximum recognition percent. In this work, the used kernel function is a linear function.

2.5.2 Deep Learning

One of the earliest methods of machine learning technique is deep learning. The method of learning is achieved by permitting the network to learn and select features from each hidden layer of neurons. From word "deep", it is concluded that the artificial neural network (ANN) has large number of hidden neurons. The architecture of conventional neural network (CNN)has ensuring translation and shift invariance which are not found in ANN. The Recurrent Neural Network (RNN) is construction of deep learning. RNN is recurrent network that makes a routine task with the output which depends on the previous computations. The unit which performs this task is called memory. The types of RNN are Long Short-Term Memory (LSTM) and the bidirectional LSTM (BILSTM). The learning process of LSTM could be long-term dependencies. The number of gates of LSTM from memory block is three gates. The names of these gates are input, output, and forget gates. The operation of these gates is to control forgetting or storing information from the network which is repeated for every input. The input gate chooses to store the new information and update it in the cell state. Output gate chooses what information is used according to the cell state. The redundant information is removed from the cell state by the forget gate. In LSTM, the sequence of time is considered only in the forward direction where as it is considered in both forward and backward directions in BILSTM, Faust O et al.(2018) (24).

2.6 The Proposed Algorithm

The steps of the proposed method are as follows: first, detrended normalized heart sounds DNHSs are obtained as shown in equations 9-11. Then one dimension wavelet transform is applied with number of levels 4 with mother wavelet Daubechy"db7". 4th approximation which is obtained from wavelet transform is squared and energy entropy is computed. Extracted features from the obtained energy entropy are compared using four algorithms MFCC, MSCC, BFCC and BSCC. Two classifiers are compared SVM deep learning (BILSTM) as shown in Fig.5. The database of HSs signals are distributed between 70% for training data and 30% for testing data.

$$x_{normalized} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{9}$$

$$x_{mean} = \frac{\sum x}{length(x)} \tag{10}$$

$$x_{\det rended} = x_{normalized} - x_{mean} \tag{11}$$

The database of HSs is divided into small batches during learning process. The size of input signals of BILSTM is 13 features. Numbers of neurons of BILSTM is 100 neurons. The final output is prepared to be two which is specified by fully connected layer. Then next layers are followed softmax layer and classification layer. Then, the options of the classifier are determined. The number of maximum epochs is chosen to be 100 to permit the network to compose 100 passes during the training process. Minimum batch size is selected to be 150 to allow the network to deal with 150 training signals at a time. The initial learning rate is 0.01 to accelerate the process of training. The gradient threshold is set to be 1 to make the training process stable by stopping gradients from increasing too high. Adaptive moment estimation solver is applied which improve the performance of RNN.

MATLAB 2019 is used as a platform for data analysis using operating system windows 7

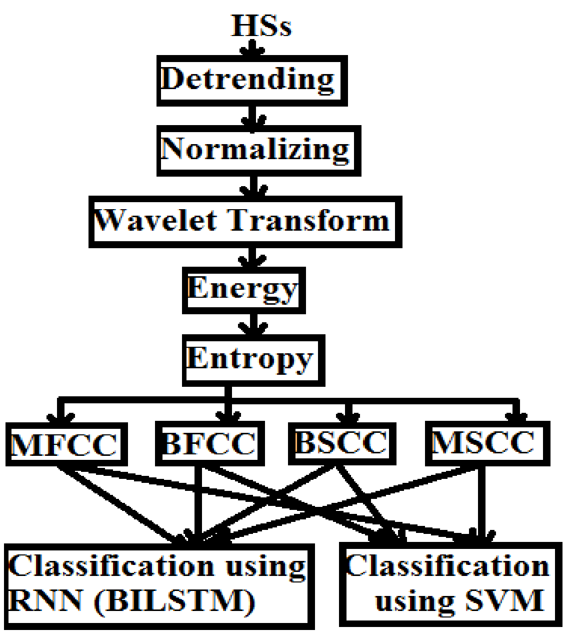

Fig.5 The proposed Method

2.7 Evaluation Metrics

For evaluating performance of the proposed method, several metrics are calculated. All performance measures used are calculated from parameters obtained from confusion matrix shown in table I.

Table I Parameters of Confusion Matrix

		Actual Values	
		1	0
The predicted Values	1	TP	FP
	0	FN	TN

Where TP is true positive, TN is true negative, FP is false positive, FN is false negative. As shown in equation (12) TP rate is obtained. As shown in equation (13) TN rate is obtained. Receiver operating characteristic ROC curve is the relation between TP rate and TN rate. AUC is obtained from the area under curve of ROC curve. Accuracy is calculated as shown in equation (14). Sensitivity is obtained as shown in equation (15). Specificity is calculated as shown in equation (16). Precision is obtained as shown in equation (17). Recall is obtained as shown in equation (18). F_measure is obtained as shown in equation (19). Gmean is calculated as shown in equation (20). F1_score is calculated as shown in equation (21)

$$TP\ rate = \frac{TP}{FN+TP} \tag{12}$$

$$TN\ rate = \frac{TN}{Fp+TN} \tag{13}$$

$$Accuracy = 100 \cdot \frac{TP+TN}{FP+FN+TP+TN} \tag{14}$$

$$Sensitivity = 100 \cdot \frac{TP}{FN+TP} \tag{15}$$

$$Specificity = 100 \cdot \frac{TN}{FP+TN} \tag{16}$$

$$Precision = 100 \cdot \frac{TP}{FP+TP} \tag{17}$$

$$Recall = sensitivity \tag{18}$$

$$F\_measure = 2 \cdot \frac{precision \cdot recall}{precision+recall} \tag{19}$$

$$Gmean = 100.\sqrt{TP\ rate.TN\ rate} \tag{20}$$

$$F1\_score = \frac{2.TP}{2.TP+FP+FN} \tag{21}$$

3. Results

As shown in Table. II, the maximum obtained accuracy of BSCC is 99.54% using BILSTM. The accuracy of BSCC is 85.48% using SVM. These values of accuracy are the maximum obtained values. As shown in Table III, metrics of classification using BILSTM are illustrated. AUC of BSCC is 0.9846. The percent of F-measure of BSCC is 98.43%. Sensitivity and recall are 100%. The specificity is 99.46%. The obtained G-mean is 99.73%. The calculated F1 score of BSCC is 0.9844. The AUC of using BSCC is 0.9846. As shown in Fig.6, ROC of using MFCC is illustrated. As shown in Fig.7 ROC of using BFCC is presented. As shown in Fig.8 ROC of using MSCC is obtained. As shown in Fig.9, ROC of using BSCC is presented.

Table II Accuracy of using different feature extraction methods using SVM and RNN (BILSTM)

Method of feature extraction	SVM	RNN (BILSTM)
MFCC	83.41%	91.36%
BFCC	82.72%	96.08%
MSCC	85.14%	89.98%
BSCC	**85.48%**	**99.54%**

Table III Metrics of using different feature extraction methods with BILSTM

Performance measurements				
	MFCC	BFCC	MSCC	BSCC
Accuracy	91.36%	96.08%	89.97%	**99.54%**
Sensitivity	47.92%	100%	42.64%	**100%**
Specificity	100%	95.26%	98.24%	99.46%
Precision	100%	81.52%	80.88%	96.92%
Recall	47.92%	100%	42.64%	**100%**
F_measure	**64.78%**	89.82%	55.84%	**98.43%**
G-mean	69.22%	97.6%	64.72%	**99.73%**
F1_score	0.65	0.898	0.56	**0.9844**
AUC	0.9531	0.9076	0.8582	**0.9846**

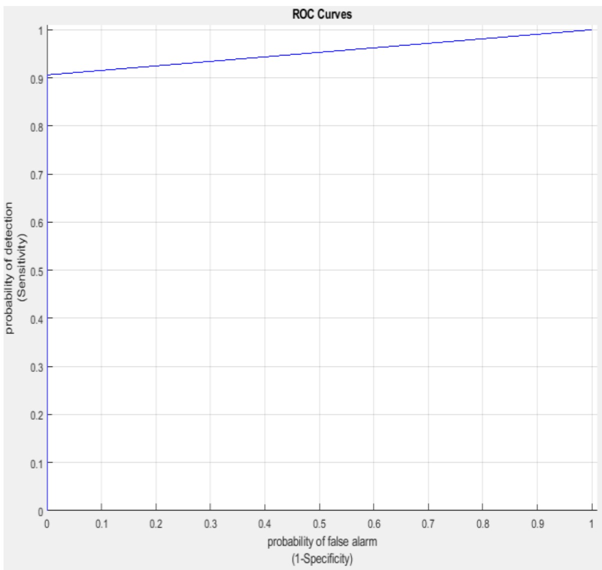

Fig.6 ROC of using MFCC feature extraction technique with RNN (BILSTM) in Diagnosing Heart Sounds

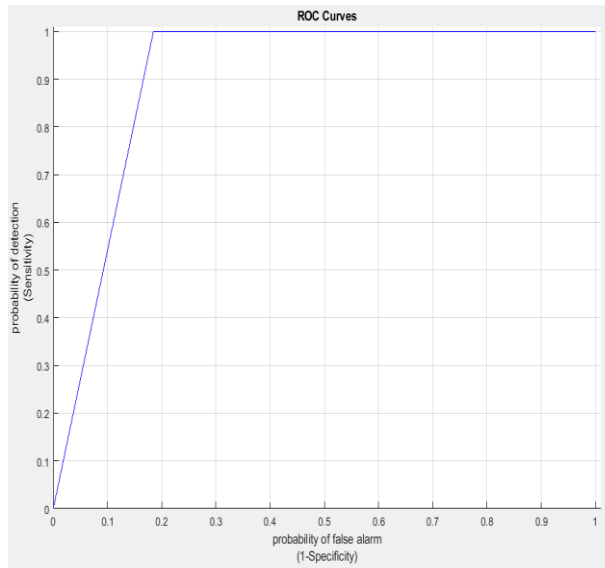

Fig.7 ROC of using BFCC feature extraction technique with RNN (BILSTM) in Diagnosing Heart Sounds

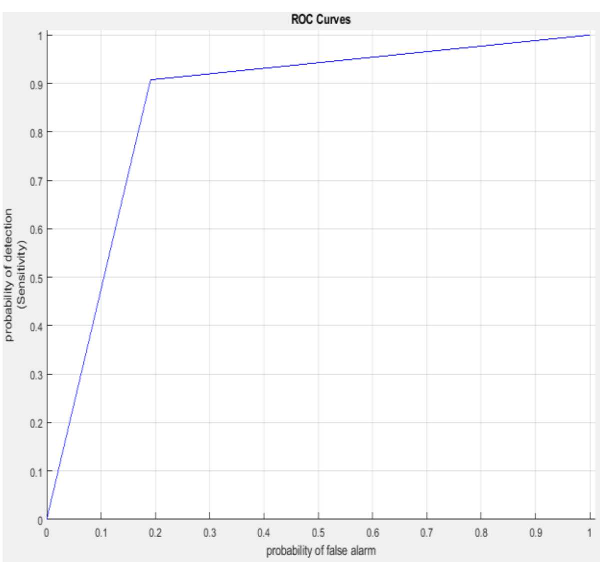

Fig.8 ROC of using MSCC feature extraction technique with RNN (BILSTM) in Diagnosing Heart Sounds

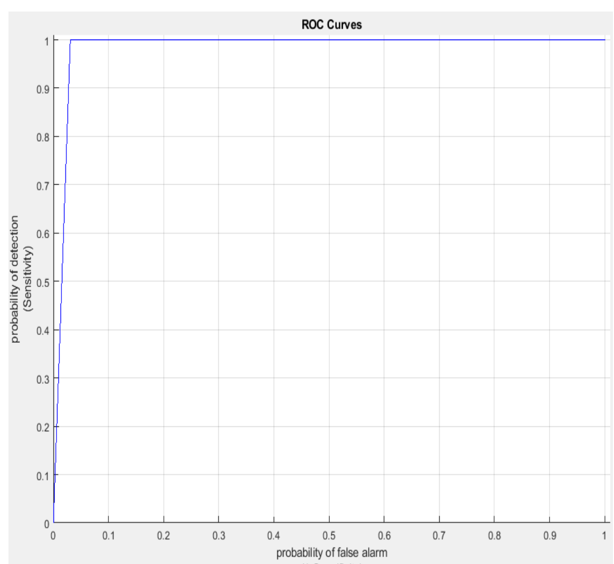

Fig.9ROC of using BSCC feature extraction technique with RNN (BILSTM) in Diagnosing Heart Sounds.

4. DISCUSSION

Several algorithms are presented such as MFCC, BFCC, MSCC and BSCC. The accuracy of BSCC is the most accuracy rate when comparing it to other algorithms (MSCC,BFCC ,MFCC). The AUC of BSCC when compared to AUC of other algorithms is the best As shown in Fig.9, AUC of BSCC is near to 1. STFT is a step of BSCC. STFT is more suitable for extracting features of heart sounds than FFT because heart sounds is non-stationary signals. STFT is used with BSCC. From all these reasons, BSCC is the best algorithm for extracting signals of heart sounds.

5. CONCLUSIONS

In this paper, new algorithms are presentedfor analyzingHSs signals. HSs signals are collected from Physionet challenge 2016 database. First, signals are detrended and normalized. Second, HSs signals are decomposed using wavelet transform. Entropy of energy is obtained. MFCC, MSCC, BFCC and BSCC are calculated to extract features.The classification phase is conducted for all 4 algorithms by SVM and Deep Learning with BILSTM. The compared simulation results demonstrated that classification using BILSTM has an accuracy percent of 99.54% using BSCC with highest value of AUC of 0.9846 . In the future, it is suggested to find a new algorithm achieving higher accuracy rate than this algorithm.

6. List of Abbreviations

BSCC barkspectrogram cepstral coefficients.
MSCC mel spectrogram cepstral coefficients.
BFCC bark frequency cepstral coefficients.
MFCCmel frequency cepstral coefficients.
Daubechies db.
STFT short-time Fourier transform
SVM support vector machine.
DL deep learning.
CNN conventional neural network.
BILSTM bi-long short term memory.
LSTMlong short-term memory.
RNN recurrent neural network.
HSs heart sounds.
DNHSs detrendednormalized heart sounds.
N normal heart sounds.
A abnormal heart sounds.
ROC receiver operating characteristics.
AUC Area under curve.

7. Declarations

Availability of data and material: available (20).
Competing interests:Not applicable.
Funding:Not applicable.
Authors' contributions:Not applicable.
Acknowledgements: Not applicable.

REFERENCES

1. https://www.who.int/
2. Shamsuddin N, Mustafa MN, Husin S and Taib MN. Classification of heart sounds using a multilayer feed-forward neural network: 2005 Asian Conference on Sensors and the International Conference on New Techniques in Pharmaceutical and Biomedical Research. 2005; 87-90, doi: 10.1109/ASENSE.2005.1564512.
3. Garzon JJ, Manrique AQ, Godino-LlorenteI and Castellanos-Dominguez CG. Feature extraction for murmur detection based on support vector regression of time-frequency representations: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.2008; 1623-26, doi: 10.1109/IEMBS.2008.4649484.
4. Maglogiannisa I, Loukisb E, Zafiropoulosb E, Stasisb A. Support Vectors Machine-based identification of heart valve diseases using heart sounds: computer methods and programs in biomedicine. 2009;95, 47–61.
5. Yana Z, Jiangb Z,MiyamotobA ,Wei Y. The moment segmentation analysis of heart sound pattern: computer methods and programs in biomedicine. 2010; 98, 140–50.
6. Hu X, Zhang J, Cao G, Zhu H and Li H. Feature extraction and choice in PCG based on Hilbert Transfer: 4th International Congress on Image and Signal Processing. 2011; 2159-63, doi: 10.1109/CISP.2011.6100614.
7. Kwak C, Kwon OW. Cardiac disorder classification by heart sound signals using murmur likelihood and hidden Markov model state likelihood: IET Signal Process. 2012; 6, 326–34.
8. Tanga H, Li T , Qiua T, Parkc Y. Segmentation of heart sounds based on dynamic clustering: Biomedical Signal Processing and Control 2012; 7, 509–16.
9. Moukadema A, Dieterlena A, Hueberb N, Brandtc C. A robust heart sounds segmentation module based on S-transform: Biomedical Signal Processing and Control. 2013; 8,273–81.
10. Patidar S, Pachori RB. Classification of cardiac sound signals using constrained tunable-Q wavelet transform: Expert Systems with Applications. 2014; 41, 7161–70.
11. Zheng Y, Guo X, Ding X. A novel hybrid energy fraction and entropy-based approach for systolic heart murmurs identification: Expert Systems with Applications. 2015; 42, 2710–21.
12. Zahhad MA, Ahmed SM, AbbasSN. Biometrics from heart sounds: Evaluation of a new approach based on wavelet packet cepstral features using HSCT-11 database: Computers and Electrical Engineering. 2016; 53, 346–58.
13. LubaibP ,Muneer KVA. The Heart Defect Analysis Based on PCG Signals Using Pattern Recognition Techniques: International Conference on Emerging Trends in Engineering, Science and Technology. 2015; ICETEST, 1024–31.
14. Zhanga W, Hana J, Dengb S. Heart sound classification based on scaled spectrogram and partial least squares regression: Biomedical Signal Processing and Control. 2017; 32,20–8.

15. Hamidi M, Ghassemian H, Imani M. Classification of heart sound signal using curve fitting and fractal dimension: Biomedical Signal Processing and Control. 2019; 39,351–59.
16. Lubis C and Gondawijaya F. Heart Sound Diagnose System with BFCC, MFCC, and Backpropagation Neural Network: IOP Conference Series: Materials Science and Engineering,Indonesia,TICATE. 2018;1-6.
17. Das S, Pal S, Mitra M. Supervised model for Cochleagram feature based fundamental heart sound identification: Biomedical Signal Processing and Control. 2019; 52, 32–40.
18. Chowdhury TH,Poudel KN and HuY. Time-Frequency Analysis, Denoising, Compression, Segmentation, and Classification of PCG Signals: IEEE Access. 2020; 8, 160882-90, doi: 10.1109/ACCESS.2020.3020806.
19. Shuvo SB, Ali SN, Swapnil SI, Al-RakhamiMS and Gumaei A. CardioXNet: A Novel Lightweight Deep Learning Framework for Cardiovascular Disease Classification Using Heart Sound Recordings: IEEE Access. 2021; 9, 36955-67, doi: 10.1109/ACCESS.2021.3063129.
20. www.physionet.org/challenge/2016.
21. Gao RX, Yan RQ. Wavelets: Theory and Applications for Manufacturing, first edition, Springer, New York.2011.
22. Theodoridis S, Koutroumbas K. Pattern Recognition: Academic Press, 2nd edition.2003.
23. Vapnik VN. Statistical Learning Theory, New York: Wiley.1998.
24. Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review: Computer Methods and Programs in Biomedicine. 2018; 161, 1-13, ISSN 0169-2607,