2021 IEEE 2nd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC) | 978-1-6654-4299-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/AESPC52704.2021.9708484

Respiratory signal Classification by cGAN augmented EMD-Scalograms

S.Jayalakshmy
Department of *ECE, Pondicherry Engineering College*
saijai2805@pec.edu

Gnanou Florence Sudha
Professor, ECE,
Pondicherry Engineering College
gfsudha@pec.edu

Abstract— **Respiratory infections are globally accepted as a major threatening factor for death rate in recent years. Even though several technological breakthroughs have progressed, medical experts face the challenges in precise diagnosis of such disorders. Auscultation is one approach used for listening to breathe sounds in order to detect lung disorders. As the publicly available repository for respiratory sounds is limited, the novel method of augmenting the data base is explored using Conditional GAN. In the proposed work, the lung sound signals are decomposed into several IMFs using EMD. It is found that for respiratory signals, IMF 3 is more informative and contributes significantly for classification and therefore only IMF 3 signals are transformed to scalograms using Continuous wavelet transform. The transformed images are augmented using cGAN. Evaluation metrics are used for finding the similarity between original and generated signals. To demonstrate the effect of augmentation in classification, experimentation is carried out with three different classifiers with scalograms as input. The experimental results shows an improvement in accuracy of 98.87% for the augmented data set.**

Keywords— Lung sounds, Empirical Mode Decomposition, Continuous Wavelet transform, Conditional GAN, Pre-trained models

I. Introduction

Deep Structured Learning has excelled significantly in the realm of medical diagnosis compared to conventional machine intelligence [1] owing to the nature of learning features from baseline data. In the field of medicine, extraction of features manually from body organs, cancer lumps and traumas are extremely difficult and labor-intensive using conventional methods of machine learning. Consequently, deep neural nets supersede the manual creation of features by having a full knowledge on the raw input data and thus providing the information into various hidden layers and eventually feeding the output to the last layers in the end-wise procedure. Several research efforts have proven to be effective with this notion of hierarchical learning in examining the bio-signal data. Breathe signals are 1-D sound signals that are heard through the process of auscultation. These signals play a vital role in examining several types of pulmonary disorders in respiratory system including asthma, pneumonia and chronic obstructive pulmonary disease.

Convolutional neural networks (CNN) are best suited for image classification and recognition due to its huge improvement in accuracy. With the help of numerous hidden convolutional layers, CNN extracts the high degree features and builds the network with 2D data. Therefore, to input the bio signal data, into this best performing deep learning model, several efforts have been made to transform 1D signals to 2D images. Several research efforts concentrated on either using short time Fourier transform or wavelet transform to transform the audio signals in time frequency plane. The later method has been proved to perform better in classifying the lung diseases.[2]

A part of the restriction observed in deep learning algorithms is that, in order to provide better accuracy, the network demands massive amount of samples for training. Adversely, the availability of training samples is often scarce in the physical world especially for classifying respiratory signals. The reason for this is that, either the number of subjects for examining the respiratory system is limited in some types of disorders or medical expert knowledge is needed in annotating the labels for each collected sample. For this reason, it is desirable to utilize data augmentation (DA) techniques to increase the number of training samples. Several DA approaches exists in literature for augmenting 2-D images which includes simple transformations such as flipping, rotating, translation and noise addition. The other methods of DA are using generative models. In this modeling, the synthetic creation of samples is extracted from probability distribution which rely on random vectors of noise. In general terms, generative adversarial nets exhibit successful performance only when the constructed model resembles real distribution of data. In this work, it is proposed to generate respiratory images using conditional GAN network using continuous wavelet transform which produces signal representation in the form of scalograms. For experimentation, the signals used in this work are considered from publicly available repository. ICBHI 2017 lung sound dataset [3] is used in which a total of 510 audio files from four categories of lung sounds are considered. The remainder of this article is outlined as follows: Related works is presented in Section II and section III describes the proposed methodology, Section IV demonstrates the result of this work including the comparison with other existing works and concluding remarks are highlighted in section V.

978-1-6654-4299-2/21/$31.00 ©2021 IEEE

II. LITERATURE SURVEY

In several investigational studies, GANs have been newly implemented and trialed in various domains and huge efforts were made on medical images to supersede the performance using deep neural networks. In the year 2020, Pandey et al. proposed 2 levels of approach in order to artificially create and classify the nuclei cell images. Initially the authors used basic GAN in the first stage to generate binary masks and second GAN for conditional mask generation [4]. In 2019, Darius Dirvanauskas et al. [5] proposed a GAN model HEMIGEN to synthetically generate human embryo cells similar to real version. The experimentation resulted with true classification rate of 80 % and false classification rate less than 13%. Frid-Adar et al. [6] introduced two mutations for augmenting the liver lesion images. In this, the initial stage involves traditional method resulting with sensitivity of 78.6% and later stage with 85.7% using GAN.

To produce high quality retinal images, the authors Mahapatra et.al. [7] utilized a novel image super resolution method in accordance with simple GAN and local saliency maps. The outcomes resulted in the accuracy value closer to one when experimenting with original images. Chen et al. in the year 2018, [8] suggested an architecture using dense net based network to synthetically generate high-resolution MRI images. This method involves in generating three dimensional MRI images rather than 2D images with a database comprising of 1,113 patient images. Salehinejad et al., [9] also exhibited the augmentation in chest X-Ray image data set using GAN and classified diseases using deep CNN. In the year 2019, the authors enlarged the NIH chest X-ray image open database using Deep Convolutional Generative Adversarial Network and the experimentation results demonstrated good performance.

Besides the data enhancement with medical images, GAN proved its evidence in signal processing fields. Basically, the acoustic signals are 1D in nature. In order to represent a small frame of audio signal, over thousand floating point numbers are required. For this reason, it is more appropriate to train deep neural nets with 2D representations of audio. Shrivastava et.al.in 2017, [10] validated that GANs can synthetically generate acoustic signal in a rapid rate, wherein the authors developed SimGAN model to experiment a combined version of simulated images and images without labels. Several trials were made with network weights and updations to yield better results. Donahue et.al. [11] made the initial efforts in the synthesis of audio signal using a novel WaveGAN model. Due to the connection complexity in neural network architecture, synthesizing of audio samples is one of the main issues since the signal relies on various scales. As a result, training a model with higher level of representation is beneficial instead of training in time domain.

To meet these needs, Time frequency representations are utilized to deal and to differentiate with non-stationary signals. As an example, Shen et.al. [12] proposed a model termed Tacotron 2, for depicting audio signals in the form of spectrograms represented in Mel Scale. The transformed TFRs in the form of Mel Spectrograms are classified using WaveNet. Most of the recent studies on audio signal classification utilize spectrograms produced through Short time Fourier transform or discrete wavelet transform and simple GAN to augment the data set. Despite the fact that, GANs produce possible gains with TFRs, graphical representation of the frequencies varying with respect to time are non-invertible for spectrogram type of representation. Furthermore, this type of TFR provides only constant resolution. To circumvent this issue and to provide variable resolution, continuous wavelet transform (CWT) is utilized wherein the TFRs named scalograms are produced. Since the basic GANs are unconditional in nature, there is no impact over generated data samples. Therefore Conditional GANs are preferred to have an influence over generated data samples. Numerous research studies have been made on augmenting the data samples in an unconditional manner [13] and have been proved efficient compared to basic GAN. Therefore in the proposed work, cGAN is employed for augmenting scalograms. For this, the lung sound signals are decomposed into several intrinsic mode functions (IMF) using Empirical Mode Decomposition. Based on entropy value, the best signal IMF is transformed to scalograms using Continuous wavelet transform and these scalograms are augmented using cGAN. These are then used for classification using CNN models.

III. METHODOLOGY

Fig. 1 shows the proposed framework for classifying the respiratory sounds using synthetic creation of scalogram images. Generally the biomedical signals are nonlinear and non-stationary in nature. Traditionally, wavelet transforms are utilized for accomplishing time-frequency analysis. Since, the wavelet transform approaches are not suitable for calculating the momentary transitions in the signal, to estimate all the minute variations in the signal, Empirical Mode Decomposition (EMD) method is used in the proposed work. EMD is adaptive in nature and the different oscillatory modes and instantaneous frequency in the signal can be captured with this transform. The dataset totally has four categories of audio files viz. 73 normal signals, 281 crackle, 33 low pitched signals (rhonchi) and 122 wheeze signals.

In the initial stage, all the audio signals from the respiratory dataset under consideration are pre-processed to eliminate the trend.

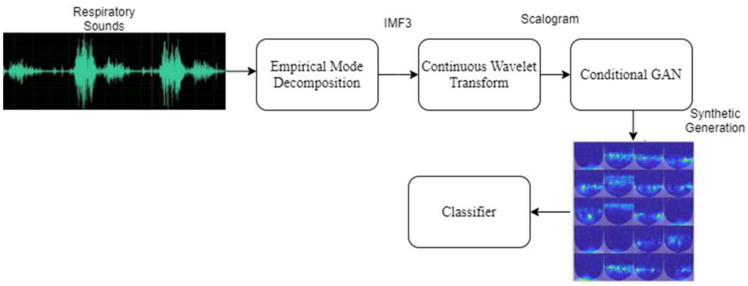

Fig. 1. Block Diagram of the Proposed Methodology

A. *Empirical Mode Decomposition*

Subsequent to the pre-processing, both the normal and abnormal signals are decomposed into several intrinsic mode functions (IMFs) using EMD. The first IMF represents the high frequency component, second represents the next high frequency component and so on. Totally seven IMFs are produced for all the time domain signals. Out of 7 IMFs, first three IMFs are considered as the entropy value is larger compared to other lower oscillatory modes from IMF 4 to IMF 7. The selected IMFs (1-3) are transformed into time frequency plane using Continuous Wavelet transform. Fig. 2 and Fig.3 shows the plot of original signal and its corresponding scalograms for a sample of both normal and abnormal signals.

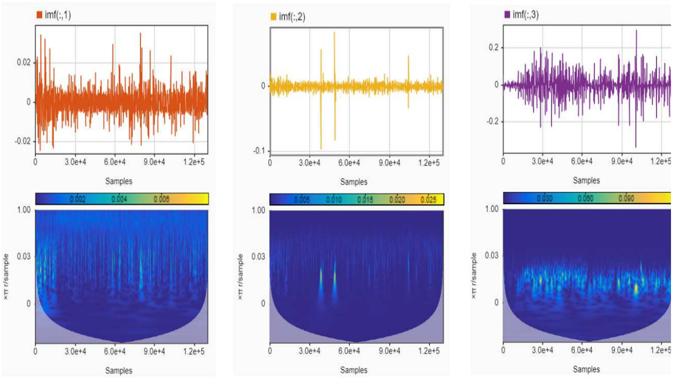

Fig.2 Sample Scalogram plot of IMF (1-3) for normal signal

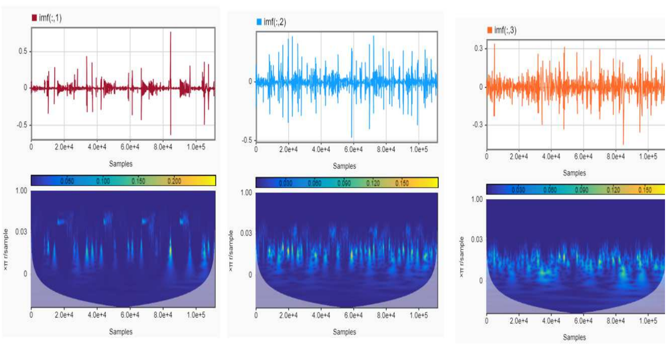

Fig.3 Sample Scalogram plot of IMF (1-3) for abnormal signal

B. *Conditional Generative Adversarial Networks*

Mirza et.al [14] developed an improved version of basic GAN named CGAN . CGAN method of augmentation acts dissimilar to basic GAN and utilizes supervised approach to have the controllability over created data samples. cGAN considers the random vectors of noise y and the class labels C as inputs to the generator network. Using these two inputs, the generator network generates the fake samples. Subsequently, the fake samples from generator and real samples along with class labels are fed as input to second network discriminator.

This network learns the similarity between the class labels and images. The process of data generation can be controlled by providing a conditional variable z. To standardize the inputs fed to the different layers, the generator uses a series of transposed convolutions and batch normalizations. Similarly the discriminator is designed with diverse convolutional layers along with leaky ReLu activations to produce future values. To generate the scalogram images for higher IMFs namely IMF 1, IMF2 and IMF 3 the cGAN is trained using custom training with MATLAB 2020.

C. *Classification*

The performance of DA using cGAN can be accessed using CNN architectures. The original and generated scalogram images are experimented with three classifiers namely AlexNet, GoogLeNet and Resnet 50 for classifying the lung sounds. The dimension of the images generated using conditional GAN are resized to 227x227x3 and 224x224x3 for classifying using the three classifiers.

IV. RESULTS AND DISCUSSION

The outcomes of the proposed classification model is outlined in this section. To generate the scalogram images using conditional GAN, the network is trained with different epochs. A sample of generated images and score plot for IMF 1, IMF2 and IMF 3 is depicted in Fig.4 (a-c) for 150 Epochs.

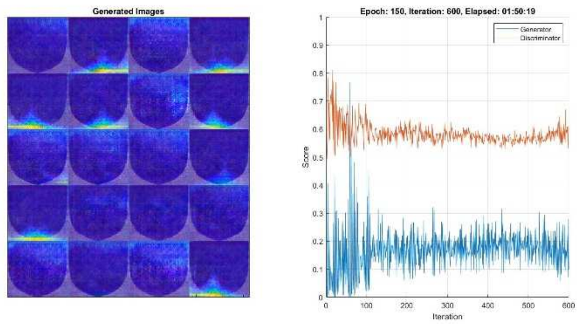

Fig.4a Generated Images and Score plot of IMF 1

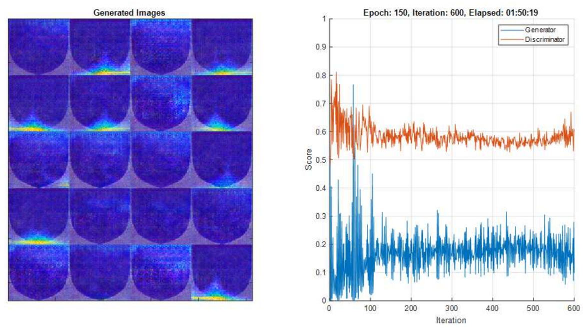

Fig.4b Generated Images and Score plot of IMF 2

978-1-6654-4299-2/21/$31.00 ©2021 IEEE

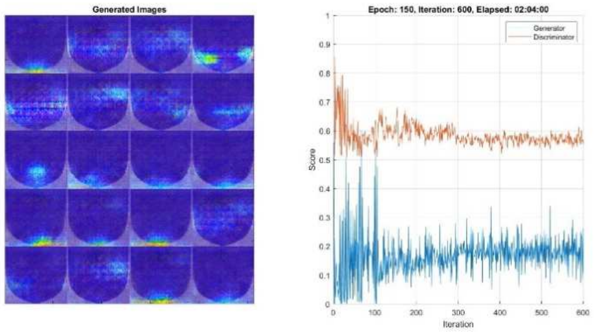

Fig.4c Generated Images and Score plot of IMF 3

During the training phase, the network convergence can be monitored visually with the help of score plot ranging from 0 to 1. From the results based on entropy value [15], since high information content is found in IMF 3, only IMF 3 augmented scalogram images are considered for classification. The above score plot indicates that generator network generates lesser number of variations in the images which leads to mode collapse. This in turn can be changed by enhancing the capability of generator network either through more number of dimensions input samples or the filter size variations. Further to assess the similarity between original time frequency scalogram images and generated scalogram images, the evaluation metrics such as structural similarity index (SSIM) and normalized mutual information(NMI) are computed. SSIM is computed by multiplying the terms viz. contrast, luminance and structure in an image and NMI is used for evaluation the information present in an image. Table 1 shows the metrics evaluated for first three IMFs of both normal and abnormal class of respiratory sounds. The findings from Table 1 indicates that the similarity metrics is found to be high for IMF 3 in both cases of normal and abnormal signals.

TABLE I. EVALUATION OF SIMILARITY BETWEEN ORIGINAL AND GENERATED SCALOGRAM IMAGES

Metrics	Normal			Abnormal		
	IMF1	IMF2	IMF 3	IMF1	IMF2	IMF 3
Structural Similarity Index	0.763	0.783	0.853	0.759	0.76	0.79
Normalized Mutual Information	0.69	0.71	0.73	0.663	0.768	0.779

Classification is carried out for cases of with and without augmentation for all classes of lung sounds. The original signal classification considers 357 audio files for training (70% of original) and 153 files (30% of original) for testing. In the same way, for the case of categorization using augmentation, totally 200 scalogram images are generated for IMF 3 signal in each class using cGAN irrespective of number of signals in each class. From the total 800 images generated, training phase considers 500 scalogram images and testing phase considers 300 images. The hyper parameter settings used for modeling the network is tabulated below in Table II.

TABLE II. HYPER PARAMETER SETTINGS FOR CLASSIFICATION

Sl.No	Hyperparameters	Values
1	Momentum	0.9
2	Learning Rate	0.0001
3	Learning Rate Drop Factor	0.2
4	Number of Epochs	20
5	Batch Size	27
6	Optimizer	Sgdm

The performance of the proposed system model is evaluated with three trained deep neural nets and the classification accuracy is tabulated in Table III for without and with augmentation. The non-linearity function used in the training model is ReLu which will output the input directly if it is negative. The results show that network with 50 layers i.e. ResNet 50 provides better accuracy compared to other two networks Alexnet and GoogLeNet. This in turn indicates that deeper the layers, the model learns more number of features thereby yielding better classification. In addition, compared to original images, augmentation has higher impact in classification. The classification accuracy is found to be 84.88 % for the case of original images and 98.14% for the case of with augmentation for ResNet 50 which is high compared to other two networks. Furthermore, the proposed model experimented with EMD based scalogram generation has comparatively higher accuracy with the existing method of scalogram generation without EMD [16].

TABLE III CLASSIFICATION ACCURACY FOR THE VARIOUS PRE-TRAINED MODELS WITH AND WITHOUT AUGMENTATION

Classifier	Original Scalogram Images	Augmented IMF 3 Image using cGAN		
		10 Epochs	20 Epochs	30 Epochs
AlexNet	79.14%	91.13%	93.73%	94.22%
GoogLeNet	82.7%	93.32%	94.08%	95.43%
Resnet 50 (Proposed)	**84.88%**	**96.17%**	**97.9%**	**98.87%**
Resnet 50 (Without EMD) [16]	81.37%	95.23%	97.82%	98.75%

Table IV shows the confusion matrix for the ResNet 50 classifier. From the confusion matrix, the class wise metrics

978-1-6654-4299-2/21/$31.00 ©2021 IEEE

namely. Precision, Recall, and F1 score are listed for all classes of respiratory sound signals in Table V. The results from table V shows that the accuracy of all classes are exceeding 98%. Furthermore, the true positive rate, positive predictive value and F1 score is found to be high in all classes which shows that the testing accuracy is superior for augmented dataset.

TABLE IV CONFUSION MATRIX OF RESNET 50 CLASSIFIER WITH IMF 3 IMAGE AUGMENTATION

Classes	Crackle	Normal	Rhonchi	Wheeze
Crackle	73	1	0	1
Normal	0	75	0	0
Rhonchi	0	1	74	0
Wheeze	0	1	0	74

TABLE V CLASS WISE METRICS FOR RESNET 50 CLASSIFIER WITH IMF 3 IMAGE AUGMENTATION

Class	Accuracy	Precision	Recall	F1 Score
Crackle	99.33%	0.97	1	0.98
Normal	100%	1	0.97	0.99
Rhonchi	98.87%	0.99	0.99	0.99
Wheeze	99.33%	0.99	0.99	0.99

V. CONCLUSION

In this study, an EMD- Scalogram based respiratory signal augmentation is experimented using conditional GAN. Due to the limitations in the traditional augmentation approaches, supervised learning structure is utilized for augmenting the data set. As the respiratory signals are non-linear and non-stationary in nature, to detect the minute variations in the signals, EMD is preferred in the proposed model. The decomposed IMFs using EMD specifically IMF 1-3 are transformed into 2D scalograms with the help of CWT. The transformed scalograms are augmented using cGAN and classified using three trained deep learning models namely Alexnet, GoogleNet and ResNet 50. To assess the similarity between the original and generated scalograms, structural similarity index and normalized mutual information is calculated. The research findings prove that the overall accuracy, class wise accuracy and other class wise metrics are found to be superior for ResNet 50 classifier with augmentation. The results are also compared with original image classification and existing method of scalogram generation without EMD.

REFERENCES

[1] Mu, R., & Zeng, X. (2019). A review of deep learning research. KSII Transactions on Internet and Information Systems (TIIS), 13(4), 1738-1764.

[2] Jayalakshmy, S., & Sudha, G. F. (2020). Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks. Artificial intelligence in medicine, 103, 101809.

[3] https://bhichallenge.med.auth.gr/

[4] Pandey, S., Singh, P. R., & Tian, J. (2020). An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation. Biomedical Signal Processing and Control, 57, 101782.

[5] Dirvanauskas, D., Maskeliūnas, R., Raudonis, V., Damaševičius, R., & Scherer, R. (2019). Hemigen: human embryo image generator based on generative adversarial networks. Sensors, 19(16), 3578.

[6] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018, April). Synthetic data augmentation using GAN for improved liver lesion classification. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018) (pp. 289-293). IEEE.

[7] Mahapatra, D., & Bozorgtabar, B. (2017). Retinal vasculature segmentation using local saliency maps and generative adversarial networks for image super resolution. arXiv preprint arXiv:1710.04783.

[8] Chen, Y., Shi, F., Christodoulou, A. G., Xie, Y., Zhou, Z., & Li, D. (2018, September). Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 91-99). Springer, Cham.

[9] Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., & Barfett, J. (2018, April). Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 990-994). IEEE.

[10] Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., & Webb, R. (2017). Learning from simulated and unsupervised images through adversarial training. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2107-2116).

[11] Donahue, C., McAuley, J., & Puckette, M. (2018). Adversarial audio synthesis. arXiv preprint arXiv:1802.04208.

[12] Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., & Saurous, R. A. (2018, April). Natural TTS synthesis by conditioning wavenet on Mel spectrogram predictions. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4779-4783). IEEE.

[13] Dieleman, S., van den Oord, A., & Simonyan, K. (2018). The challenge of realistic music generation: modelling raw audio at scale. In Advances in Neural Information Processing Systems (pp. 7989-7999).

[14] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[15] Jayalakshmy, S., & Sudha, G. F. (2021). GTCC-based BiLSTM deep-learning framework for respiratory sound classification using empirical mode decomposition. Neural Computing and Applications, 1-12.

[16] Jayalakshmy, S., Priya, L., & Sudha, G. F. (2021). Synthesis of respiratory signals using conditional generative adversarial networks from scalogram representation. In Generative Adversarial Networks for Image-to-Image Translation (pp. 161-183). Academic Press.

978-1-6654-4299-2/21/$31.00 ©2021 IEEE