CHAPTER 7

Synthesis of respiratory signals using conditional generative adversarial networks from scalogram representation

S. Jayalakshmy[a], Lakshmi Priya[b], and Gnanou Florence Sudha[c]
[a]IFET College of Engineering, Villupuram, India
[b]Manakula Vinayaga Institute of Technology, Pondicherry, India
[c]Pondicherry Engineering College, Pondicherry, India

7.1 Introduction

Chronic respiratory disorders are a genre of long-term illness influencing the air passage and the anatomy of the respiratory system. Lung disorders are graded as the biggest donor to the global effect of diseases in the world. Some of the respiratory disorders include chronic obstructive pulmonary diseases (COPD), asthma, occupational lung illness, chronic bronchitis, pneumonia, etc. The diagnostic studies of the World Health Organization (WHO) and Healthy People 2020 revealed that currently over 25 million people in the United States (US) have been diagnosed with asthma and roughly 14.8 million adults have been identified with COPD [1]. In addition, the Forum of International Respiratory Societies (FIRS) has anticipated that more than millions of people live along with increased pressure in the pulmonary system and more than 50 million individuals suffer from occupation-related lung infections, thereby figuring out that more than one billion people are being affected from chronic respiratory conditions [2].

The causes of COPD include deficiency of alpha-1-antitrypsin protein, long-term exposure to air pollution, chemicals, fumes, and dust inhalation in the workplace. This deficiency causes the lungs to deteriorate and also can affect the liver. The common monitoring and diagnostic techniques for pulmonary diseases are spirometry, CT scans, arterial blood gas tests, and auscultation of the lungs. Respiratory auscultation is the most preferred diagnostic tool for the examination of pulmonary disorders and it provides the physiological and pathological information for the medical experts to proceed with the therapeutic procedure [3]. The respiratory sound (RS) is one of the most significant bio-signals used to diagnose certain respiratory abnormalities. RS detected from the chest wall and mouth may be classified as normal (vesicular) and adventitious sounds. Some of the abnormal sounds include wheezing, rhonchus (low-pitched wheezes), stridor, and

Generative Adversarial Networks for Image-to-Image Translation
https://doi.org/10.1016/B978-0-12-823519-5.00024-5

crackles which vary in frequency. The use of a conventional stethoscope to listen lung sound makes lung auscultation technique simple, easy to use, and the most popular non-invasive method for diagnosis.

To assist medical practitioners further in the process of their diagnosis, deep learning-based computer-aided diagnosis (CAD) systems have been extensively used over the past few years. However, for these deep learning algorithms to precisely differentiate even the feeble abnormal breathing patterns, huge volume of training data is required. On the other hand, a very limited number of both normal and abnormal respiratory sound datasets are available in the publicly available datasets. Working with these insufficient resources, deep learning models tend to struggle with few restrictions such as overfitting which performs well only for the trained model but not for other unobserved data. This can be considered as the greatest challenge in conventional deep learning-based CAD systems. Although several patterns of training and architectural designs have been deployed in research findings, training a model with a scarce amount of data is a demanding task. Therefore to acquire more samples, data augmentation is the sensible solution.

For scaling up of datasets, the conventional augmentation approach accomplishes ordinary variations in the given images in order to gain a different facet of original images. Few modifications involve rotational changes, translation, reflection, and diminishing the size of images [4]. Considering from the respiratory signal perspective, transforming the audio signal to the 2D representation using conventional augmentation approaches is not appropriate. In explicit terms, random flipping around the time axis indicates that the signal is reversed in time and random Y flipping will completely change the understanding of frequency. In the same way, random X translation implies that there is only time shift and translation in Y signifies that the frequency spectrum is being modified, which may not really be a true representation of the original signal itself. Scaling randomly in the X direction in order to simulate slower breathing or faster breathing, while not changing the frequency characteristics would be physically meaningful but this in turn adds random noise to the signal representation. To resolve these issues, this study aims at improving the data set artificially with the help of generative adversarial network (GAN). In this study, it is proposed to synthesize a respiratory signal using GAN architecture by the virtue of time-frequency representation which gives a picture of the signal. Scalogram, the visual representation of the energy density of the signal obtained through continuous wavelet transform is found to differentiate well the different classes of respiratory signals [5]. The fact that the continuous wavelet transform is invertible, enables to use of GAN architecture to indirectly synthesize signals.

GAN is a class of deep learning framework which succinctly generates images from a known representation of data called latent space. GAN comprises two deep neural networks called generator and discriminator, competing one against the other in order to learn the probability distribution function of the known training set images and hence the term adversarial. The outcome of this study will undoubtedly address the challenge

faced with restricted training data in deep learning-based respiratory signal classification. Furthermore, the proposed study can be used as a data augmentation technique in the abovementioned signal classification task.

The remaining part of this chapter is structured in this way: related study on the application of GAN in diverse fields and augmentation approaches is presented in Section 7.2; basic GAN, cGAN, and proposed model are explained in Section 7.3, the dataset details, generator, and discriminator network architecture and the classifier results are depicted in Section 7.4; and finally research findings are explained in Section 7.5.

7.2 Related work

Over the past few years, the focus on deep learning models and algorithms has gradually gained significant importance in addressing several issues in the field of medical imaging. Several studies have been reported in the literature using supervised learning techniques, wherein a huge amount of training data is required to prepare a strong model. Owing to the wide range of images in the medical field, the collection of data samples continues to be a great challenge. Introducing minute changes in the original images poses limitations in the classification performance as the augmentation methods induce additional details in the training samples. Furthermore, few proportion of expanded dataset seems to sound in a distinct way compared with the real-world objects resulting in unsuitability to other databases. In order to overcome these issues, Ian Good fellow et al. [6] proposed an alternative approach of data augmentation wherein synthetic images are generated by employing generative adversarial networks (GANs). The implementation of the structured adversarial modeling GAN in turn signified very sharp distributions compared to Markov chain models.

GANs are a kind of unsupervised learning used for mapping the small-scaled hidden vectors to high-dimensional data. In the literature, GANs have been lately put into practice in diverse fields and many initiatives were carried out on medical images which employ image-to-image translation. In the year 2017, Costa et al. [7] explored U-net for generating new fundus images in the retina by vessel segmentation with the help of GANs. The results indicated that both the original and synthetic images were observed differently visibly even though both were a part of the same vessel tree. In addition, the produced synthetic images were of the major part of real image set quality. Lei Bi et al. [8] proposed multichannel generative adversarial networks (M-GAN) for boosting the training data of positron emission tomography (PET) images and provided more realistic images in comparison with conventional GAN. In 2018, Frid-Adar et al. [9] proposed classical data segmentation technique as the first stage to expand the dataset of CT images and synthetic data augmentation as the second stage using GAN for the classification of liver lesions and yielded 78.6% sensitivity and 88.4% specificity for the case of classic data augmentation approach and with the help of synthetic image creation, the classification

accuracy was found to be improved to 85.7% sensitivity and 92.4% specificity. Furthermore, in 2018, Salehinejad et al. [10] have also witnessed the expansion in dataset samples by implementing GAN to produce artificial images for the classification of the lesion in chest X-ray images. The authors utilized a deep convolutional neural network (DCNN) to identify disorders with five different classes of chest X-rays and the performance results are found to be improved.

In 2019, Bhattacharya D et al. [11] suggested deep convolutional generative adversarial network (DCGAN) and experimented on NIH chest X-ray image open database to enhance the efficiency of CNN model using GAN and yielded 65.3% accuracy. With the aid of structure correcting GAN, Dai et al. [12] carried out segmentation between lungs and heart regions in chest X-ray images. In that work, the authors introduced a critic network in order to figure out the higher-level structures to acquire practical segmentation findings to achieve realistic segmentation outcomes. Several comprehensive attempts with this method resulted in real segmentation with high precision. Onishi et al. [13] explored deep CNN (DCNN) and GAN to create the sufficient number of images in order to differentiate malicious and benign lung nodules. In that work, the images were generated using the pixel value distribution present in the mid-portion of the pulmonary nodule. This approach of pretraining and fine-tuning process using DCNN enables to discriminate almost 66.7% of benign nodules and 93.9% of malicious pulmonary nodules and proved with the classification accuracy of 20% more than original images.

Apart from CT and PET images, Chaudhari et al. [14] have trialed the augmentation approach on gene expression dataset using modified generator GAN (MG-GAN) and compared the performance with basic GAN and KNN classifier. The results proved that MG-GAN improved the accuracy by 18.8% and 11.9% and further the loss value of the error function was found to be reduced very drastically from the value 0.6978 to 0.0082 making it suitable for applications with sensitive data. Luo et al. [15] explored the progressive growth of GAN-based augmentation on electroluminescence images for the classification of faulty photovoltaic component cells and improved the performance to the maximum of 14% using an enlarged dataset. Apart from this, Li et al. [16] focused the research on gear safety in the transmission industry for reliability classification using GAN wherein, the authors introduced bounded-GAN to create gear data with different settings and trained the model using ADAM optimizer. The research findings show that the proposed bounded-GAN excels other approaches on operational measures. GAN also finds its application in diverse fields such as magnetic resonance imaging scans, video surveillance, clinical informatics, computational biology, automotive fields, etc.

Furthermore, GAN-based methods strengthen and show the possible gains in the audio synthesis field as well with reference to analysis, processing, and classification of signals. In spite of the latest developments in the field of artificial intelligence and generative models, the compilation of information from inherent sounds through neural nets

remains unresolved. In 2017, Shrivastava et al. [17] substantiated that audio signals can be illustrated at a faster pace with the help of GANs. Here, the authors proposed a combination of simulated and unsupervised learning and attempted numerous changes with the basic GAN such as self-regularization, local loss, and discriminator updation with the past record of perfect images and resulted in good performance improvement. Donahue et al. in 2019 [18] proved that GANs allow the signal generation to take place at any time. The authors introduced WaveGAN and experimented GANs to unsupervised synthesis of raw-waveform audio as an initial attempt and achieved promising results. By the virtue of complexity in neural structure, audio signal generation is a key issue as it depends on several time scales. Therefore, it is advantageous to train a net with a greater degree of illustration rather than utilizing samples in the temporal dimension. Time-frequency analysis are used to distinguish and handle the nonstationary signals in a better way. One such example is Shen et al. [19] demonstrated a novel architecture named Tacotron 2, for combining audio signals through Mel-frequency spectrograms. These Mel spectrograms are fed as input to a network named WaveNet and resulted in a mean opinion score of 4.53.

Even though several achievements have been enabled through TF representations, visual representation of the spectrum of frequencies varying with time (spectrograms) is noninvertible. Marafioti et al. [20] discussed the key points of neural structure explored for producing TF representations especially speech synthesis using STFT. In that, the authors introduced TiFGAN, which unconditionally creates audio. But it poses limitations on producing audio with substantial quality. In addition, the spectrogram type of representation produces only constant resolution. These drawbacks can be resolved by employing the continuous wavelet transform (invertible) in which TF representation named scalograms are produced with variable resolution. The use of an unconditional generative model does not influence the generated modes of data. Several efforts have been made for generating the audio signal in an unconditional manner [21–23]. Despite that, all the techniques utilize the autoregressive method which considers noise samples as input and creates samples of audio signals serially. By exploring the model with some conditioning using extra information, it would be possible to supervise the process of data generation. This conditioning process may rely on class labels or tags on a portion of data or data on a whole.

Mirza et al. [24] introduced the conditional adversarial nets and generated images trained on MNIST class labels. The authors proved the potency of conditional adversarial nets and their useful applications with the tags used individually. Conditional GANs (cGANs) are a kind of GAN, wherein the information concerning the conditions are imposed to basic GAN. The results show superior performance compared to nonconditional GANs.

The majority of the application areas experience the inability to gain access to big data for analysis, in particular, the medical field. Even though several data augmentation

approaches are possible with GAN both for medical images and audio [13–23], some of the technical gaps observed in the existing study are quality of generated data samples, distortion, and its distribution in the data set, which were not good enough resulting in poor classification accuracy as the accuracy is significantly dependent on both the qualitative and quantitative terms. In addition, the speed of the image generation process is very slow specifically in the field of speech synthesis. This poses limitations on high-dimensional data due to the nature of autoregressive modeling. Furthermore, in certain instances, network models trained with artificial generated sample data fail to perform well when fed with real images. All these gaps can be resolved in the proposed study with the help of conditional GAN. Inspired by the performance improvement by conditional GANs in Ref. [24], the proposed study employs the scalogram method of TF representation combined with cGAN for improved targeting and for synthesizing respiratory sounds in order to discriminate the normal and abnormal lung sounds.

7.3 GAN for signal synthesis

In this section, the architectures of simple GAN and conditional GAN and the proposed system model using the conditional GAN to synthesize respiratory sound signals from the wavelet-based time–frequency representation are explained in detail. The conditional GAN is utilized in this proposed study to artificially generate more number of scalogram images. With this proposed model as a data augmentation technique, better prediction accuracy through computer-aided diagnosis is expected.

7.3.1 Simple GAN

A simple GAN comprises two networks named generator and discriminator that are trained concurrently. The generator learns to generate a new image mimicking the data in the latent space by estimating its underlying probability distribution and the discriminator plays the role of binary classifier by mapping the input image either to the real-image dataset or to the generated set of images. The generator model has to be trained such that the generated image very closely resembles the images used for training thereby making the discriminator difficult to distinguish between the original and generated set of images. The discriminator in turn learns to make sure that its performance is better than that of the generator. This adversarial learning behavior of the GAN results in the generation of images which are very close to the real training set images. Fig. 7.1 shows the architecture of simple GAN [25].

7.3.2 Conditional generative adversarial networks

In contrast to basic GAN, cGAN utilizes a supervised method where both the generator and discriminator neural networks are adapted to meet the condition during the training

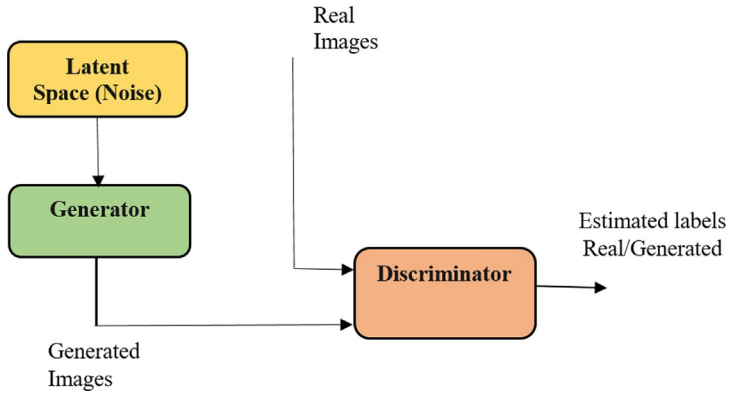

Fig. 7.1 Architecture of simple GAN

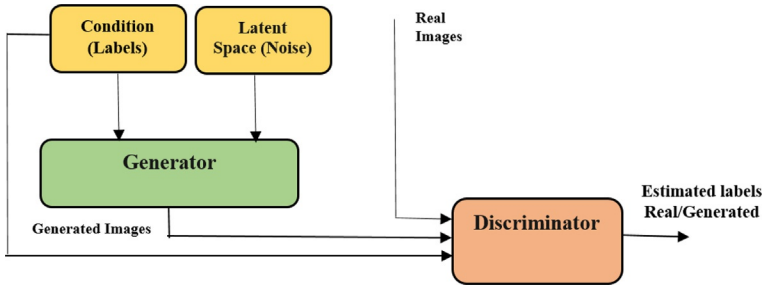

Fig. 7.2 Architecture of conditional GAN.

phase with the help of certain extra information. Fig. 7.2 shows the architecture of conditional GAN [26]. The latent space information in the form of noise and the condition labels are fed as inputs to the generator block. The images generated by the generator along with condition labels and the real images are given as input to the discriminator block. This block detects the similarity between the given labels and images. The data augmentation is achieved by incorporating the conditional variable y into the model.

7.3.3 Conditional GAN for respiratory sound synthesis

The fundamental idea in the proposed study is to train the conditional GAN to generate realistic scalogram images of various respiratory sounds.

7.3.3.1 System model

Fig. 7.3 shows the proposed framework for synthesizing respiratory sounds using cGAN. Different types of lung sound such as vesicular, wheezes, crackles, and low-pitched

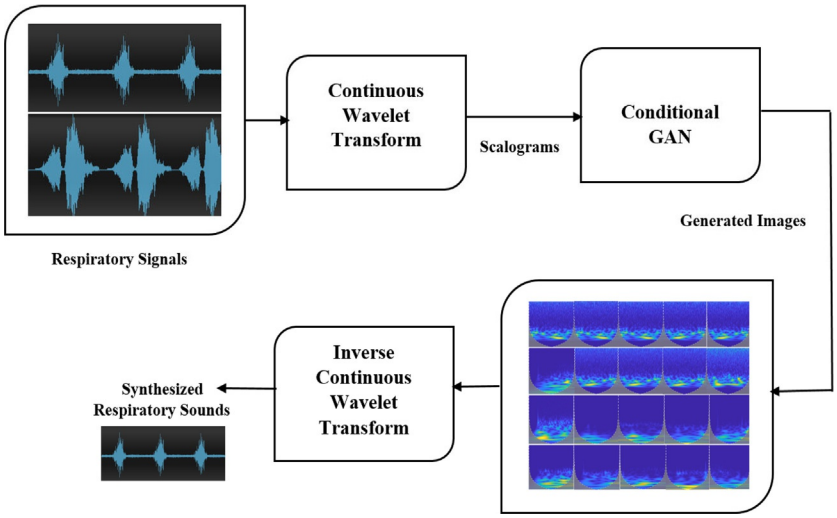

Fig. 7.3 Proposed system model for synthesizing respiratory sounds

wheeze signals are given as input to the time-frequency transform. The continuous wavelet transform transforms the time-domain respiratory signal to time-scale analysis and is represented in the forms of scalogram representation. These real scalograms are fed as input to conditional GAN in order to generate more artificial scalogram images with the help of a generator network in cGAN. Finally, inverse CWT is applied to synthesize the respiratory sound signals in time domain.

7.3.3.2 Time-scale representation using CWT

The continuous wavelet transform (CWT) modifies the temporal length of the basis function for the purpose of achieving a changeable time-frequency localization. To interpret very small changes in frequencies, CWT utilizes longer basis functions at the cost of confined localization in time and uses shorter basis functions ascertaining high localization in time [27]. This time-frequency transform elicits a spectrum with time scale vs amplitude named scalogram. Compared to spectrograms, scalograms are useful for analyzing realistic signals at diverse scales. As the frequencies in CW transform are logarithmic in nature, the obtained scalogram plot also uses a log scale frequency axis.

7.3.3.3 Generator and discriminator network architecture of cGAN

The network architecture for generator and discriminator are shown in Figs. 7.4 and 7.5 and the corresponding analysis result is tabulated in Tables 7.1 and 7.2. In cGAN, the

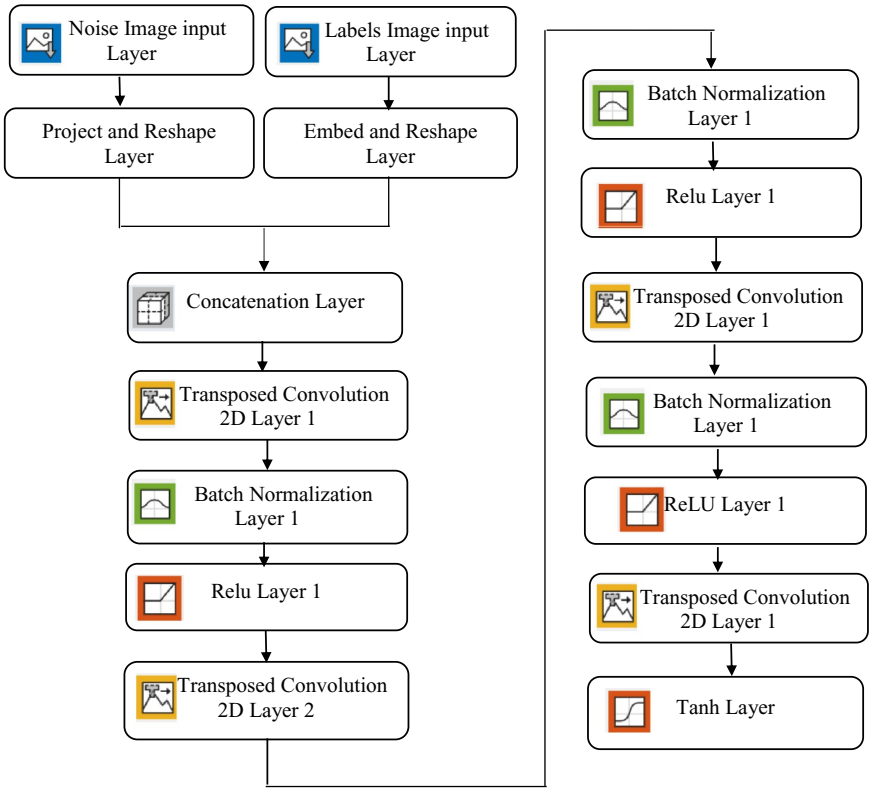

Fig. 7.4 Generator network architecture.

generator network comprises sequential transposed convolutional layers with batch normalization for scaling up the arrays of different dimensions. The small-scale noise vector given to the fully connected section transforms the input to 1024 high-scaled image features which is further reshaped to $4 \times 4 \times 1025$ for feeding to the convolution module. Several successive stages of transposed convolution layers transform the interim features to produce an output image equal to the dimension $(64 \times 64 \times 3)$. This generator network generates synthetic scalograms for all classes of respiratory sounds individually to attain the wider class population.

On the other hand, the discriminator network is modeled using multiple convolutional layers with leaky ReLu to produce prediction values. From the given input image to the discriminator block, which has a dimension of $(64 \times 64 \times 3)$, the high-scaled features with the dimension $(4 \times 4 \times 512)$ are extracted by the series of convolution layers. Further, these features are evened and are given as input to the fully connected network. This network gradually assigns the features to a low scale for classification.

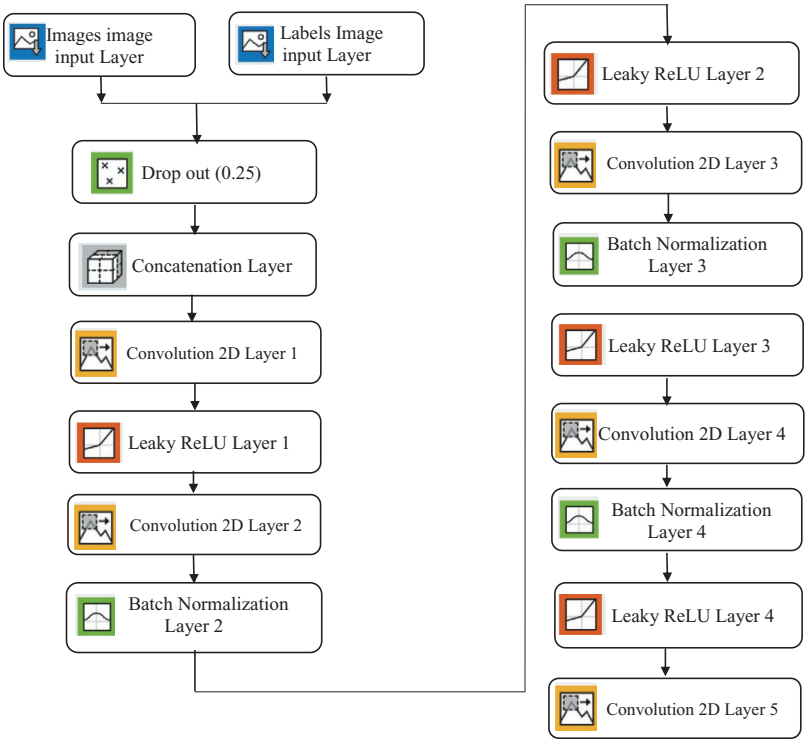

Fig. 7.5 Discriminator network architecture.

7.3.3.4 Algorithm

Algorithm 7.1. Generation of Scalograms using cGAN for the training process with Epochs = 500, 1000, 1500, and 2000, learn rate = 0.0002 and no. of classes = 4

Input: Real scalograms $S_0,S_1,S_2…S_n$ and noise input vector: Y, Conditional metrics: $C_0,C_1,C_2…C_n$

Formulate bounds for G and D of cGAN with input dimension [64 64 3]

for num of training phases **do**

- Update the discriminator using $S_0,S_1,S_2…S_n$ with $C_0,C_1,C_2…C_n$
- Generate scalograms $Z_0,Z_1,Z_2…Z_n$ using noise input vector with $C_0,C_1,C_2…C_n$
- Revise the D network using $Z_0,Z_1,Z_2…Z_n$ with $C_0,C_1,C_2…C_n$
- Revise the G network using $Z_0,Z_1,Z_2…Z_n$ with $C_0,C_1,C_2…C_n$

end for

Output: 200 no. of observations for each class.

Table 7.1 Analysis result of generator network.

Sl. no.	Name	Type	Activations	Learnables
1	Noise $1 \times 1 \times 100$ images	Image input	$1 \times 1 \times 100$	–
2	Proj Project and Reshape Layer with output size $4 \times 4 \times 1024$	Project and reshape	$4 \times 4 \times 1024$	Weights $16{,}384 \times 100$ Bias $16{,}384 \times 1$
3	Labels $1 \times 1 \times 1$ images	Image input	$1 \times 1 \times 1$	–
4	emb Reshape Layer with output size 4×4	Embed and reshape layer	$4 \times 4 \times 1$	Embedding weights 50×4 Fully connecting weights 16×50 Fully connecting Bias 16×1
5	Cat Concatenation of 2 inputs along dimension 3	Concatenation	$4 \times 4 \times 1025$	–
6	tconv1 256 $5 \times 5 \times 1025$ transposed convolutions with stride[1 1] and cropping [0 0 0 0]	Transposed convolution layer	$8 \times 8 \times 256$	Weights $5 \times 5 \times 256 \times 1025$ Bias $1 \times 1 \times 256$
7	bn1 Batch normalization with 256 channels	Batch normalization	$8 \times 8 \times 256$	Offset $1 \times 1 \times 256$ Scale $1 \times 1 \times 256$
8	Relu1 Relu	Relu	$8 \times 8 \times 256$	–
9	tconv2 128 $5 \times 5 \times 256$ transposed convolutions with stride [2 2] and cropping same	Transposed convolution layer	$16 \times 16 \times 128$	Weights $5 \times 5 \times 128 \times 256$ Bias $1 \times 1 \times 128$
10	bn2 Batch normalization with 128 channels	Batch normalization	$16 \times 16 \times 128$	Offset $1 \times 1 \times 128$ Scale $1 \times 1 \times 128$
11	Relu2 Relu	Relu	$16 \times 16 \times 128$	–
12	tconv3 64 $5 \times 5 \times 128$ transposed convolutions with stride [2 2] and cropping same	Transposed convolution layer	$32 \times 32 \times 64$	Weights $5 \times 5 \times 64 \times 128$ Bias $1 \times 1 \times 64$
13	bn3 Batch normalization with 64 channels	Batch normalization	$32 \times 32 \times 64$	Offset $1 \times 1 \times 64$ Scale $1 \times 1 \times 64$

Continued

Table 7.1 Analysis result of generator network—cont'd

Sl. no.	Name	Type	Activations	Learnables
14	Relu3 Relu	Relu	$32 \times 32 \times 64$	–
15	tconv4 3 $5 \times 5 \times 64$ transposed convolutions with stride [2 2] and cropping same	Transposed convolution layer	$64 \times 64 \times 3$	Weights $5 \times 5 \times 3 \times 64$ Bias $1 \times 1 \times 3$
16	tanh Hyperbolic tangent	Tanh	$64 \times 64 \times 3$	–

Table 7.2 Analysis result of discriminator network.

Sl. no.	Name	Type	Activations	Learnables
1	Images $64 \times 64 \times 3$ images	Image input	$64 \times 64 \times 3$	–
2	Dropped 25% dropout	Drop out	$64 \times 64 \times 3$	–
3	Labels $1 \times 1 \times 1$ images	Image input	$1 \times 1 \times 1$	–
4	emb Reshape Layer with output size $64 \times 64 \times 3$	Embed and reshape layer	$64 \times 64 \times 1$	Embedding weights 50×4 Fully connecting weights 4096×50 Fully connecting Bias 4096×1
5	Cat Concatenation of 2 inputs along dimension 3	Concatenation	$64 \times 64 \times 1$	–
6	conv1 64 $5 \times 5 \times 4$ convolutions with stride [2 2] and padding same	Convolution	$32 \times 32 \times 64$	Weights $5 \times 5 \times 4 \times 64$ Bias $1 \times 1 \times 64$
7	lrelu1 Leaky relu with scale 0.2	Leaky Relu	$32 \times 32 \times 64$	–
8	Conv2 128 $5 \times 5 \times 64$ convolutions with stride [2 2] and padding same	Convolution	$16 \times 16 \times 128$	Weights $5 \times 5 \times 64 \times 128$ Bias $1 \times 1 \times 128$
9	bn2 Batch normalization with 128 channels	Batch normalization	$16 \times 16 \times 128$	Weights $1 \times 1 \times 128$ Bias $1 \times 1 \times 128$

Table 7.2 Analysis result of discriminator network—cont'd

Sl. no.	Name	Type	Activations	Learnables
10	lRelu2 Leaky ReLU with scale 0.2	Leaky Relu	$16 \times 16 \times 128$	–
11	Conv3 256 $5 \times 5 \times 128$ convolutions with stride [2 2] and padding same	Convolution	$8 \times 8 \times 256$	Weights $5 \times 5 \times 128 \times 256$ Bias $1 \times 1 \times 256$
12	bn3 Batch normalization with 256 channels	Batch normalization	$8 \times 8 \times 256$	Weights $1 \times 1 \times 256$ Bias $1 \times 1 \times 256$
13	lRelu3 Leaky ReLU with scale 0.2	Leaky Relu	$8 \times 8 \times 256$	–
14	Conv4 512 $5 \times 5 \times 256$ convolutions with stride [2 2] and padding same	Convolution	$4 \times 4 \times 512$	Weights $5 \times 5 \times 256 \times 512$ Bias $1 \times 1 \times 512$
15	bn4 Batch normalization with 512 channels	Batch normalization	$4 \times 4 \times 512$	Weights $1 \times 1 \times 512$ Bias $1 \times 1 \times 512$
16	lRelu4 Leaky ReLU with scale 0.2	Leaky Relu	$4 \times 4 \times 512$	–
17	Conv5 1 $4 \times 4 \times 512$ convolutions with stride [1 1] and padding [0 0 0 0]	Convolution	$1 \times 1 \times 1$	Weights $4 \times 4 \times 512$ Bias 1×1

7.3.3.5 Steps

The stages for the generation of original, realistic scalograms, and synthesis of respiratory sounds are outlined as follows:

***Step 1*:** Examine and analyze the given respiratory signals having integral number of time-varying frequencies using continuous wavelet transform.

Step 2: Produce Morse scalogram representations for various lung sounds with the aid of MATLAB wavelet tool box.

Step 3: Input the obtained scalograms of the original lung sound signals to Conditional GAN.

Step 4: Generate realistic scalogram images with the help of modeled generator network.

Step 5: Synthesize the original respiratory signal by giving the generated scalograms to inverse CWT.

Step 6: Provide the original scalogram images extracted through continuous wavelet transform and generated scalogram images through cGAN to the pretrained models Alexnet CNN [5], GoogLenet [28], and ResNet 50 to quantify the performance.

7.4 Results and discussion

To demonstrate the performance of the proposed data augmentation technique using conditional GAN, the original scalogram images extracted through continuous wavelet transform and generated scalogram images through cGAN are input to different pretrained models. The classification performance is compared for all the classes of respiratory sounds with and without augmentation.

7.4.1 Dataset

The dataset used in this proposed study was acquired from various sources namely RALE (Respiration acoustics Laboratory Environment) repository [29], Think labs Lung sound library [30], and ICBHI [31] benchmark publicly available databank. These archives comprise gender-based normal and abnormal lung sounds of several kinds. For the training and testing phase, the entire lung sound database has been arbitrarily split into 70% and 30%. By and large, the database has 73 normal files, 281 crackle sound files, 33 rhonchi files, and 122 wheeze files.

7.4.2 Data augmentation using conditional GAN

The training phase of the data augmentation process using cGAN is explained in this section. For the process of experimentation,

(i) The number of latent inputs for the generator network are considered to be 100. Typically the generator produces RGB noise as scalograms at random.

(ii) From the modeled convolutional filters, the discriminator network attempts to figure out the difference between random noise scalograms and real-scalogram images of respiratory sounds.

(iii) To mystify the discriminator network, the generator learns from the transposed convolution filters.

(iv) The process is continued endlessly, as long as the discriminator is confused to the greatest extent.

In the proposed approach, the network is trained for four different epochs namely 500, 1000, 1500, and 2000 and the cost function is observed for all cases. The visual representation of the progress of training with scores of both the networks are shown in Fig. 7.6.

To check for the convergence of the network during the process of training, the scores are plotted on a scale from 0 to 1. The score of the generator network is defined

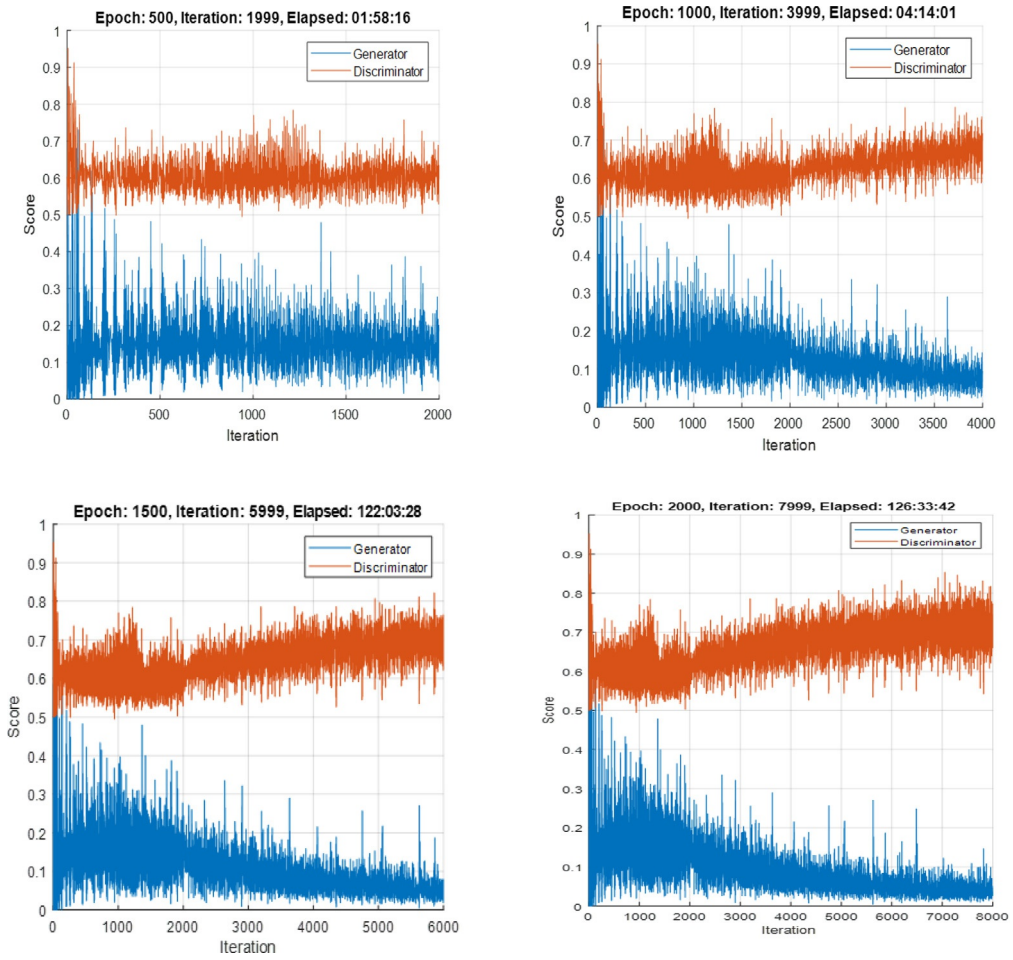

Fig. 7.6 Training plots of generator and discriminator for different epochs.

as the average of the likelihood images analogous to the discriminator output for the generated images. In case 1 of Fig. 7.6, i.e., 500 epochs, the concept of mode collapse happens which indicates that the generator is incapable of learning the scalogram representation corresponding to diverse inputs. Therefore, in order to increase the ability of the generator to produce more outputs, the number of epochs is increased. The training plots of cases (2, 3, 4) indicate that the generator score reaches the value 0 and the score of discriminator extends to almost one which signifies that the discriminator network is dominating the generator network and therefore classifies most of the images correctly. Since the plots are almost stable in cases 3 and 4, the training phase is stopped with 1500 epochs. Further increasing the number of iterations, increases the computational time of the network.

7.4.3 Samples of generated scalogram images for different classes

Fig. 7.7 shows the samples of scalogram images generated by the generator network for 1500 epochs.

7.4.4 Synthesis of respiratory sounds using inverse CWT

The inverse CWT is applied to the generated scalograms and the acquired respiratory sound signal for the case of normal and abnormal lung sounds are plotted using the signal analyzer app and shown in Fig. 7.8.

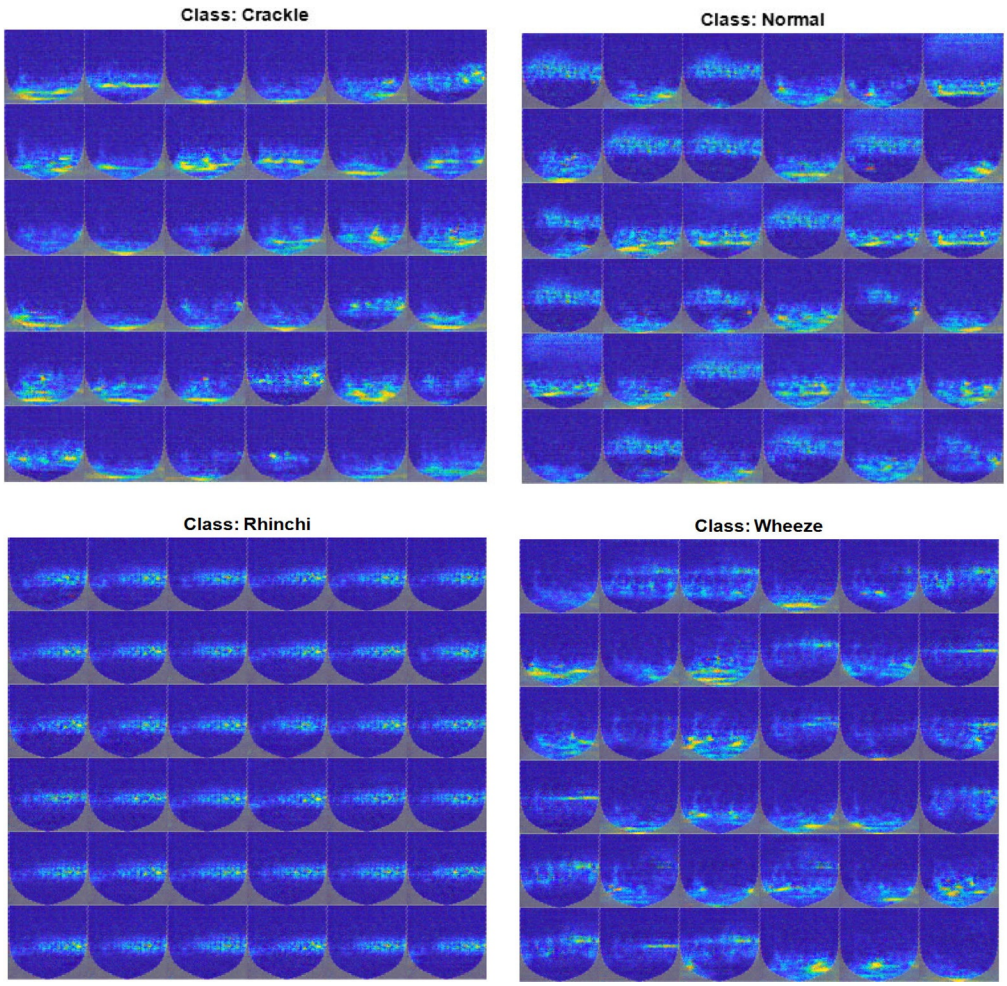

Fig. 7.7 Samples of generated scalogram images for different classes using 1500 epochs.

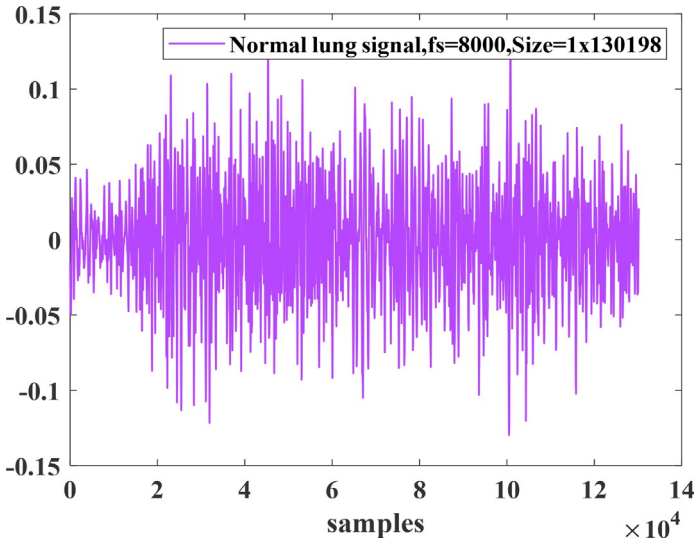

Fig. 7.8 Sample of normal synthesis from scalogram using ICWT.

7.4.5 Performance results

To evaluate the performance of data augmentation using conditional GAN, pretrained deep learning models such as AlexNet, GoogLeNet, and Resnet 50 are used for classification. The classification is performed for all the classes of respiratory sounds without augmentation and with augmentation and the results are compared for all the deep learning models. For classifying the data set without augmentation, the number of images considered for training are 357 and 153 images are used for testing. Similarly, for classification with augmentation, 200 images are generated for each class of respiratory sounds using cGAN. Out of which, 500 images are considered for training and 300 for testing. The experimental settings for modeling the network are listed in Table 7.3.

Table 7.3 Parameters settings for trained network model.

Sl. No	Hyperparameters	Values
1	Momentum	0.9
2	Initial learning ate	0.0001
3	Learning rate drop factor	0.2
4	Learning rate drop period	5
5	Number of epochs	20
7	Batch size	10
8	Optimizer	Sgdm

Table 7.4 Classification accuracy for the various pre-trained models with and without cGAN.

Classifier	Without augmentation (%)	With augmentation using cGAN		
		500 epochs (%)	1000 epochs (%)	1500 epochs (%)
AlexNet	68.63	93.13	95.13	96.38
GoogLeNet	73.86	93.45	96.88	96.88
Resnet 50	**81.37**	95.23	97.82	**98.75**

Resnet 50 provides high accuracy (indicated in *bold*) compared to other two methods

The performance metrics, accuracy of the pretrained network is usually estimated by calculating the testing accuracy with the help of a confusion matrix. Accuracy measures the number of correctly classified normal and abnormal sound files corresponding to the total number of test samples. Table 7.4 shows the classification accuracy obtained for various deep network models without and with augmentation for different epochs. The results in Table 7.4 indicate that the pretrained CNN model Resnet50 performs well for both the cases with and without augmentation. For the case of classification with real images, i.e., without augmentation, the Resnet 50 model produces an accuracy of 81.37% which is high compared to AlexNet and GoogLeNet classifiers. Furthermore, generating new images with cGAN and training the deep network model with ResNet 50 produces the highest classification accuracy of 98.67% compared with other deep network models at 1500 epochs.

The training progress and confusion matrices of ResNet 50 network model with real images and generated images are shown in Figs. 7.9 and 7.10 and Tables 7.5 and 7.6.

The columns of the confusion matrix plotted in Tables 7.5 and 7.6, indicate the true cases for the classes and the rows indicate the cases that are belonging to the class. To be more specific, in Table 7.5, the number of actual cases in the class crackle is 77, 31 in normal, 16 in rhonchi, and 29 in wheeze. The number of cases correctly classified as belonging to the particular class are 71 for crackle, 18 for normal, 10 for rhonchi, and 25 for wheeze. With data augmentation, in Table 7.6, the number of actual cases in the class crackle is 72, 79 in normal, 75 in rhonchi, and 74 in wheeze, and the number of cases correctly classified as belonging to the particular class are 72 for crackle, 75 for normal, 75 for rhonchi, and 74 for wheeze.

From the confusion matrices tables, other metrics, such as precision, recall, and F1 score are also calculated for all types of lung sounds and are tabulated in Tables 7.7 and 7.8. Precision calculates the total number of positive class forecasts which is actually positive. The parameter recall computes the same number of positive class forecasts with all positive samples in the dataset while the F1 score is the weighted average of precision and recall.

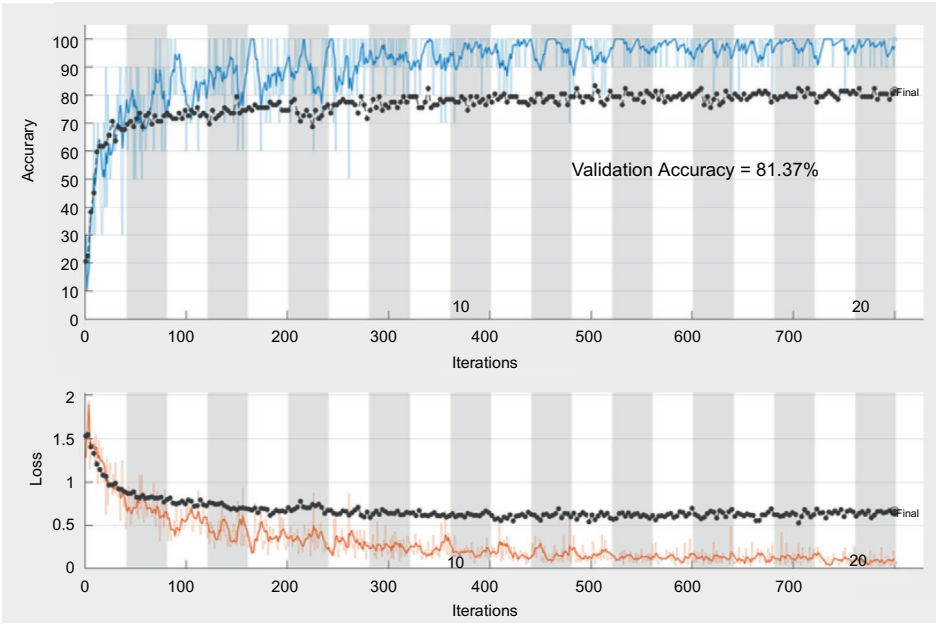

Fig. 7.9 Training progress of ResNet 50 network for the case of real images (without augmentation).

From the tables, it is observed that the class-wise accuracy is found to be high for all classes almost more than 98% for the case of classification with augmentation, whereas it is comparatively low for the case of without augmentation. In addition, the F1 score is high for all classes in Table 7.8 which reveals that the validation accuracy is better for augmented data in all cases.

7.4.6 Analysis

The proposed method in this chapter demonstrates the training of CNNs with an alternative method for data augmentation by way of generating synthetic scalogram images by using conditional generative adversarial networks. The proposed method is experimented with three pretrained models namely Alexnet, GoogLeNet, and ResNet50 classifiers. The pretrained Alexnet model with five convolutional and three fully connected layers produces an accuracy of 68.63% for the case of original scalogram images whereas the network trained with 1500 epochs using cGAN data augmentation approach yields an accuracy of 96.38%. The same number of images when trained with GoogLeNet classifier with 22 layers deep yields an accuracy of 73.86% without augmentation and 96.88% with data augmentation. The third model ResNet 50 comprises 49 convolutional layers and a fully connected layer. This network when trained, yields an accuracy of 81.37%

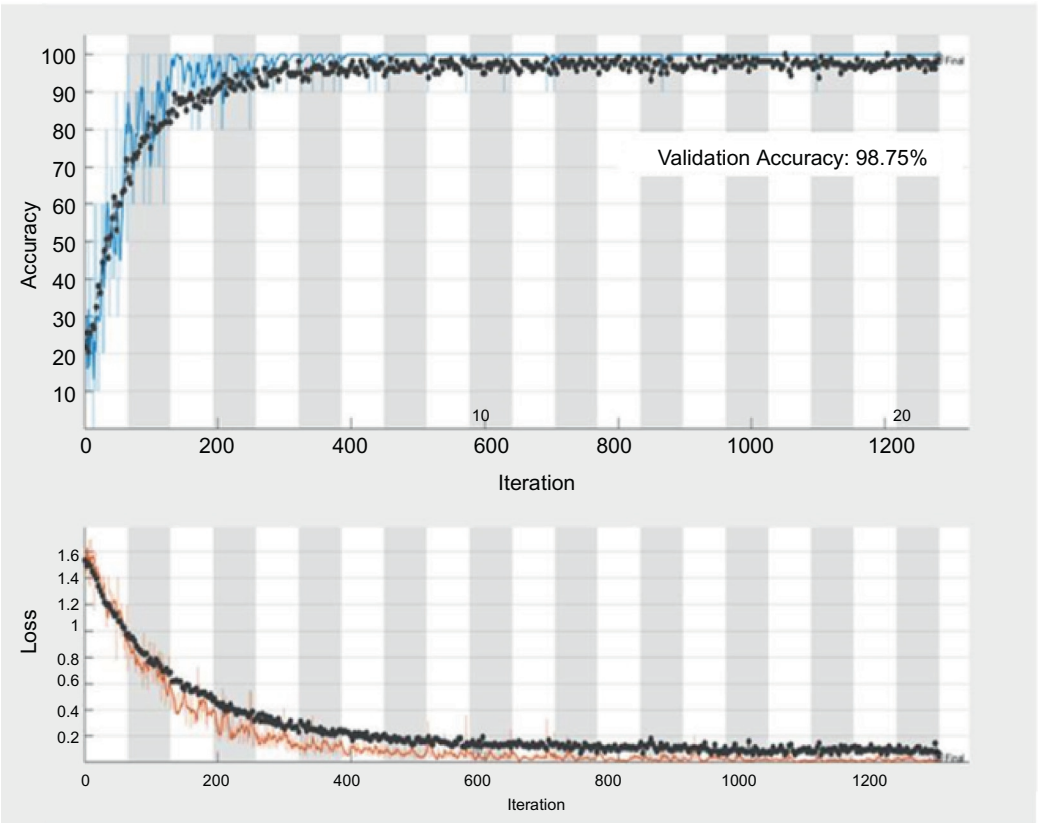

Fig. 7.10 Training progress of ResNet 50 network with augmentation.

Table 7.5 Confusion matrix of ResNet 50 network without augmentation.

	Crackle	Normal	Rhonchi	Wheeze
Crackle	71	9	1	3
Normal	3	18	0	1
Rhonchi	0	0	10	0
Wheeze	3	4	5	25

Table 7.6 Confusion matrix of ResNet 50 network with augmentation.

	Crackle	Normal	Rhonchi	Wheeze
Crackle	72	3	0	0
Normal	0	75	0	0
Rhonchi	0	0	75	0
Wheeze	0	1	0	74

Table 7.7 Precision and recall for ResNet 50 model without augmentation.

Class	Accuracy (%)	Precision	Recall	F1 score
Crackle	87.58	0.85	0.92	0.88
Normal	88.89	0.82	0.58	0.68
Rhonchi	96.08	1	0.63	0.77
Wheeze	89.54	0.68	0.86	0.76

Table 7.8 Precision and recall for ResNet 50 model with augmentation.

Class	Accuracy (%)	Precision	Recall	F1 score
Crackle	99	0.96	1	0.98
Normal	98.67	1	0.95	0.97
Rhonchi	100	1	1	1
Wheeze	99.67	0.99	1	0.99

without augmentation and 98.75% for the case of with augmentation. This indicates that deeper networks prove efficient both in terms of computation and the number of parameters. In addition, the model with good accuracy both in case of with and without augmentation, i.e., ResNet 50 is assessed with various metrics namely accuracy, precision, recall, and F1 score. Based on results from Table 7.8, the high values of precision, recall, and F1 score shows that the validation accuracy is better for augmented data in all classes of respiratory sounds.

7.5 Conclusion and future scope

Owing to the challenges in incorporating the conventional data augmentation techniques for time-frequency representation of the signal, a novel data augmentation approach has experimented for the signal under study. In this chapter, GAN an unsupervised learning structure is utilized to generate the synthetic images for the different classes of respiratory sounds. For improved targeting on the image generation, the conditional information is imposed to basic GAN. The contradictory learning behavior of conditional GAN gives rise to the generation of scalogram images really close to original scalogram images of respiratory sounds. It is also found that the performance of modeled discriminator network predominates the generator network and therefore categorizes the majority of the images accurately. In addition, the performance of the data augmentation approach is evaluated with different pretrained deep learning classifiers and compared with original images without augmentation. The results show that there is a significant improvement in the classification accuracy of all models in the data augmentation approach in comparison with without cGAN. Furthermore, the testing accuracy of ResNet 50 model produces an increased accuracy of 98.75% with high values of precision, recall, and F1 score for all

classes of respiratory sounds resulting in better prediction. This study can be further extended with other types of GAN such as cycle GANs and Wasserstein GANs for the synthetic generation of images. The same setup can be compared with the generation of images using variational convolutional autoencoder to produce a better prediction model.

References

[1] https://www.healthypeople.gov/2020/topics-objectives/topic/respiratory-diseases (Respiratory Diseases—Accessed 10 May 2020).
[2] https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory_Disease.pdf (Global impact of Respiratory diseases—Accessed 05 May 2020).
[3] M. Sarkar, I. Madabhavi, N. Niranjan, M. Dogra, Auscultation of the respiratory system, Ann. Thoracic Med. 10 (3) (2015) 158.
[4] A. Sharif Razavian, H. Azizpour, J. Sullivan, S. Carlsson, CNN features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014, pp. 806–813.
[5] S. Jayalakshmy, G.F. Sudha, Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks, Artif. Intell. Med. 103 (2020) 101809.
[6] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial networks, 2014. arXiv preprint arXiv:1406.2661.
[7] P. Costa, A. Galdran, M.I. Meyer, M.D. Abràmoff, M. Niemeijer, A.M. Mendonça, A. Campilho, Towards Adversarial Retinal Image Synthesis, 2017. arXiv preprint arXiv: 1701.08974.
[8] L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs), in: Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment, Springer, Cham, 2017, pp. 43–51.
[9] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using GAN for improved liver lesion classification, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, April, pp. 289–293.
[10] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, J. Barfett, 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), IEEE, 2018, April, pp. 990–994.
[11] D. Bhattacharya, S. Banerjee, S. Bhattacharya, B.U. Shankar, S. Mitra, GAN-based novel approach for data augmentation with improved disease classification, in: Advancement of Machine Intelligence in Interactive Medical Image Analysis, Springer, Singapore, 2020, pp. 229–239.
[12] W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, E.P. Xing, Scan: structure correcting adversarial network for chest x-rays organ segmentation, arXiv (2017). arXiv preprint arXiv: 1703.08770.
[13] Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, H. Fujita, Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks, BioMed Res. Int. 2019 (2019) 6051939, https://doi.org/10.1155/2019/6051939.
[14] P. Chaudhari, H. Agrawal, K. Kotecha, Data augmentation using MG-GAN for improved cancer classification on gene expression data, Soft Comput. 24 (2019) 11381–11391.
[15] Z. Luo, S.Y. Cheng, Q.Y. Zheng, GAN-based augmentation for improving CNN performance of classification of defective photovoltaic module cells in electroluminescence images, in: IOP Conference Series: Earth and Environmental Science, vol. 354 (1), IOP Publishing, 2019, October, p. 012106.
[16] J. Li, H. He, L. Li, G. Chen, A novel generative model with bounded-GAN for reliability classification of gear safety, IEEE Trans. Ind. Electr. 66 (11) (2019) 8772–8781.