2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) | 978-1-6654-2605-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/IDAACS53288.2021.9660990

The 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
22-25 September, 2021, Cracow, Poland

Data Augmentation using GAN for Sound based COVID 19 Diagnosis

Nishant Yella [1], Bina Rajan [2]
[1] Department of Computer Science and Engineering, Malla Reddy Engineering College, Hyderabad, India
nishanty1219@gmail.com
[2] Department of Electronics and Communication Engineering, Sai Vidya Institute of Technology, Bengaluru, India
binar.17ec@saividya.ac.in

Abstract—**The COVID 19 virus has been mutating at a rapid phase, due to which the golden standard of testing reverse transcription-polymerase chain reaction (RT-PCR) has been producing false negatives at an alarming rate. The inability of the test to detect the mutated strain of the COVID 19 virus using RT-PCR has made it very difficult for diagnosis and hence an alternative solution is needed. Sound-based diagnosis is one effective alternative diagnosis tool. The lack of a large dataset is one challenging aspect for the development of a sound-based diagnosis tool. We look forward to using dataset augmentation as a very effective technique for a selected classification problem: visual perception and also speech recognition tasks. The Generative Adversarial Networks (GANs) have been showing high success for applications in terms of synthesizing realistic images, they're seen rarely in audio generation-based applications Due to the lack of data sets available to develop an accurate model in this paper we showcase an application of WaveGAN, which is a variant of GAN which helps in raw audio synthesis during a supervised setting for the classification task, by developing a method showcasing one of the approaches for augmenting speech datasets by using Generative adversarial networks (GANs). We deploy the WaveGAN on the existing data sets collected from open-source collections to develop synthetic, larger data set to build an accurate sound-based diagnosis tool.**

Keywords—Generative adversarial networks, WaveGAN, Sound-based diagnosis

I. INTRODUCTION

One of the major challenges in the domain of deep learning and its subdomains is the lack of availability of labeled data or datasets. The performance of a multi-layered perceptron (MLPs) and/or Deep Neural Network (DNNs) largely depends upon the quality of training data that is being fed. Data Augmentation is a promising method in terms of generating synthetic data/datasets, these methods and techniques are bound to be very effective especially for generating synthetic images and audio samples (in some cases). The methods of data augmentation introduced in the field of deep learning have given greater scope for generating large amounts of training data for MLPs and DNNs which therefore are intended for solving the issue of lack of availability of data. In this paper/research study we proceed to explore various data-augmentation methods and focus/ narrow down to the applications of generative modeling for augmentation of audio cough samples of covid-19 positive patients with the help of a variant of Generative Adversarial Network (GAN) called waveGAN.

The extended training data from waveGAN will be used in a convolutional neural network (CNN) as a feed-forward neural network for performing a binary classification task to determine a positive/negative covid-19 diagnosis by using cough(speech) as the main input parameter. Traditional machine learning algorithms can be primitive for complex tasks and their method of learning can be restricted to only a certain amount and category of data only, therefore the choice of picking up a DNN would be very appropriate for binary diagnosis classification task with a raw audio sample being the only input parameter. The MLP would help in dividing the raw audio data into multiple segments for each perceptron thus, developing a more refined learning ability for raw audio samples. Backpropagation and dropout would help the neural network develop a better learning rate and prevent overtraining the network respectively. The II section of the paper talks about the methodology used and III describes the modeling of WaveGan followed by III results and IV conclusion.

II. METHODOLOGY

We first briefly summarize how to generate synthetic samples by using Generative adversarial networks (GANs). We Initially generate a set of augmented data from the original dataset (audio data), which is later used in training the classifier. For example, GANs are applied to the original audio set which undergoes basic transformations to generate (augment) a larger dataset.

These audio samples are fed as input net during training time and testing time, the original audio sample is used for validation.

Followed by a method that emphasizes the understanding of data augmentation through a base-level neural network. During training, the neural network inputs two arbitrary audio samples derived from the original audio sample training set, and an output is generated which is modulated into a single “audio” sample, such that the

output audio sample matches in terms of context to a sample from the training data. This output data enters into a second classification feed-forward neural network by the side of the original audio training dataset. The loss generated is back-propagated to train the layers in the convolutional neural network(CNN).

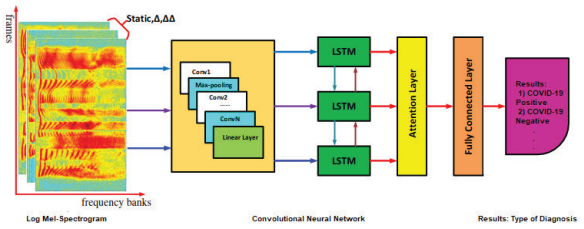

Figure 1. Experiment Diagram

The generated output, which is an augmented audio sample generated by the neural network, proceeds further into the next classification network with the original training audio dataset.

During the time of testing, audio samples from the validation set are run through a classification network.

A. *Generative Adversarial Networks(GANs):*

For every audio sample which is used as an input, we select synthesized audio from a subset. A synthesized transformation of the original audio data sample is developed. The original and synthesized audio sets are fed in training the neural network. There are two differentiable functions in the framework of GAN. The first function is known to be the generator. The generator develops samples that are essentially using the training set. The second function is called the discriminator, the discriminator's task is to examine samples to detect if they are fake or real. The discriminator understands utilizing supervised learning methods and splitting inputs into classes named real and fake. [6]

B. *Convolutional Neural Network(CNN):*

This task involves building a Convolutional Neural Network (CNN) for performing a binary classification task for determining the positive or negative diagnosis of covid-19 just from an audio sample of a cough of an individual. A convolutional neural network is the extension of a multi-layered perceptron. A generic CNNs architecture primarily consists of the following segments which are

1) Input Layer 2) Pooling Layer 3) Hidden Layer/Layers (completely connected) 4) Output Layer.

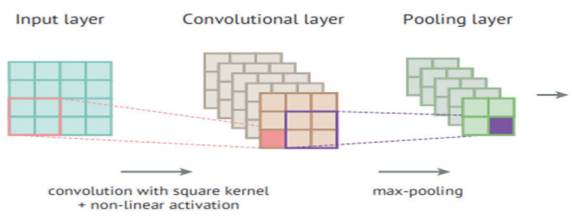

Figure 2. Representation of CNN

The CNN network is trained on the original audio dataset followed by augmented data and will be tested against the validation set of the original audio dataset.

One very effective approach/method of tackling this type of task is by introducing the neural network for dropout training. The dropout is a technique that is used for preventing the hidden layers of the neural network from the problem of overfitting. During the training, if the data is tending towards overfitting before the specified epochs value, the dropout method prevents the network from further training. Therefore the dropout technique acts as a regularization method for a neural network.

An architecture that is involved with dropout attempts to make sure every single hidden unit understands feature representation, which in turn is favorable for the correct output of a classification problem.

III. MODELLING

A. *Wave GAN Architecture:*

WaveGAN is a variant of GAN which is originally derived from DCGAN. Unlike DCGAN WaveGAN has been specifically built to run in a single dimension. Here we extend the overall single dimension filter of the waveGAN instead of using multi-dimensions. Having a variation in the discriminator and the filters, the waveGAN is capable of producing an extended audio sample of the covid-19 positive recording from the original audio data, which is longer than the original sample. Hence waveGAN would help generate extended training data for more precise audio classification. [6]

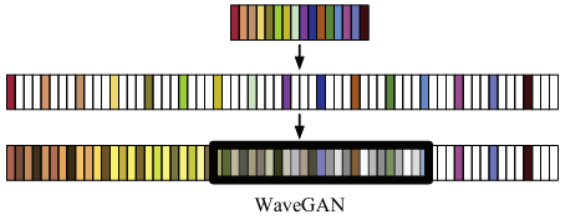

Figure 3. WaveGAN

B. *Experiment setup(CNN):*

In the point of view of evaluating the performance of augmented audio data samples along with the training set. This neural network will be using 50 epochs and would have a learning rate of 0.001 along with the adam optimizer. We have set up a convolutional neural network referred from Jason Wang et al.[7]

The augmented audio-set is then fed into the neural network which performs the binary classification task.

C. *Datasets & Features:*

Since there is a shortage of availability of labeled data. We have collected the cough audio sample of a person who has been tested positive for COVID-19. The length of this sample is 12 seconds long which is considered the primary training dataset. An unconventional method i.e data augmentation has been put to generate an extended training

sample from the original data. The objective of this method is for expanding the original labeled training data as a deep convolutional neural network performs more efficiently in learning when provided with more labeled datasets, especially in the category of supervised learning for a binary classification task. Hence, the research statement has to be expanded in this context where convolutional neural networks(CNN) are used effectively for external environmental sound classification tasks with significantly lower amounts of training data.

Figure 4 Represents the Covid positive sound sample as a Log-Mel Spectrograph, which is plotted across frequency and time domain.

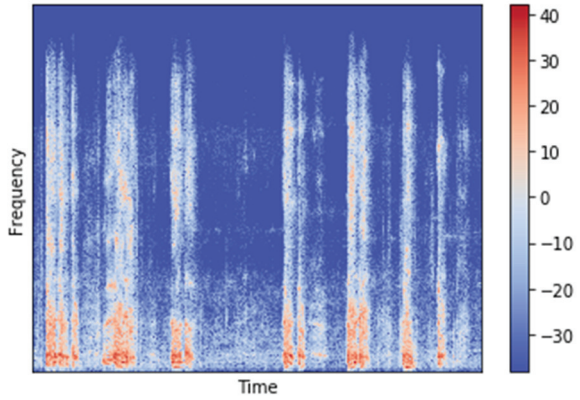

Figure 4. Time-Frequency Graph of the Audio sample

Figure 5 is a graphical representation of the Covid positive sound sample that was collected.

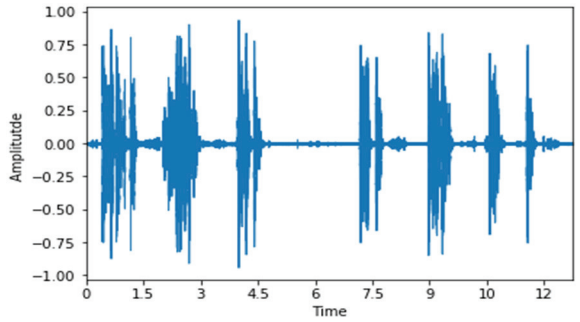

Figure 5. Time-Amplitude Graph of the Audio sample

Figure 6 is the graphical plot of the Frequency vs Magnitude.

When compared to the study carried out by B.Lange et al, where the use of Support Machine Vectors was used to detect COVID-19 using audio cough samples as the parameter[6]. The model showcased an accuracy of 0.59 or 59%. Another group study carried out in project N.Sharma et al lacked COVID-19 positive audio cough samples, overall the Coswara project was not able to detect COVID-19 mainly due to lack of training data[7]. In comparison, our neural network has shown tremendous improvement in terms of accuracy and precision during the phase of evaluation displayed is about 0.79 (or 79%) and precision to be of 0.81 (or 81%).

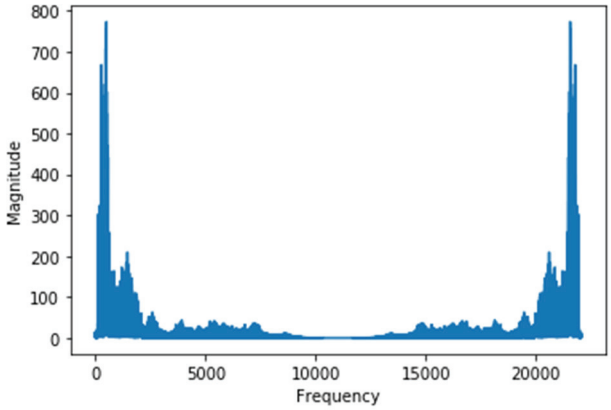

Figure 6. Frequency-Magnitude Graph of the Audio sample

IV. Results and Discussions

The final output, which has been trained on the augmented audio set, has proven to increase the overall accuracy and precision of the convolutional neural network (CNN). The accuracy displayed is about 0.79 (or 79%) and precision to be of 0.81 (or 81%). We can infer from this result that audio samples of a covid-19 positive cough can be an effective feature for this binary classification diagnosis task. Here the convolutional neural network has shown to have a better performance over any traditional supervised machine learning algorithms. The network's performance can degrade due to the lack of availability of training data and containing minimum features. However, we can be mindful about our feature selections and also include shallow coughs, breath audio samples, and even include the breathing patterns of individuals. Introducing multiple features can certainly make the diagnosis task more efficient, but to do so would require collecting more samples and recordings taken in where the background noise is minimum.

V. Conclusion

In this paper, we have shown that generating augmented audio data of cough of covid-19, using a feed-forward GAN has significantly improved the overall accuracy and precision of an audio classifying neural network which has been created to diagnose covid-19 symptoms using audio/speech as the input parameter. The application of data augmentation techniques has been showcased as a method to eliminate the shortage of available labeled data, especially when dealing with classification tasks of speech and image recognition. However, this is just the tip of the iceberg for the method and variant that has been put to practice in this paper.

This paper will provide scope for further research and studies in the field of covid-19 and pandemic for evaluating options on non-contact audio/speech-based diagnosis. The audio set of a cough can be one of the significant input

parameters to remotely diagnose the presence of a virus in the human body. We believe that the method of diagnosis has the omnipotent potential of not just limiting itself for covid-19 but also other viruses that tend to infect the human body in the future.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Pavan Kumar Jakkepalli (NIT Warangal) for guiding us constantly and Dr. Chaya BM (Sai Vidya Institute of Technology) for domain-specific knowledge and relentless support.

REFERENCES

[1] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, "WaveNet: A Generative Model for Raw Audio," 2016; URL:http://arxiv.org/abs/1609.03499

[2] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, Y. Bengio, "SampleRNN: An unconditional end-to-end neural audio generation model," 2016; URL: http://arxiv.org/abs/1612.07837

[3] N. Jaitly, & E. Hinton, Vocal Tract Length Perturbation (VTLP) improves speech recognition, 2013.

[4] O. Abdel-Hamid, A. Mohamed, H. Jiang and G. Penn, "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," *Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 4277-4280, doi: 10.1109/ICASSP.2012.6288864.

[5] K. J. Piczak, "Environmental sound classification with convolutional neural networks," *Proceedings of the 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2015, pp. 1-6, doi: 10.1109/MLSP.2015.7324337.

[6] B. Lange, D. Li, E. Nehoran, E. Tuzhilina, & M. Lu, "Early detection of COVID-19 from cough sounds, symptoms, and context," CS 472: Data science and AI for COVID-19.

[7] N. Sharma, P. Krishnan, R.hit Kumar, S. Ramoji, S. R. Chetupalli, R. Nirmala, P. K. Ghosh, S. Ganapathy, "Coswara – A database of breathing, cough, and voice sounds for COVID-19 diagnosis," 2020. https://dx.doi.org/10.21437/Interspeech.2020-2768 10.21437/Interspeech.2020-2768

[8] I. Goodfellow, "NIPS 2016 Tutorial: Generative Adversarial Networks," 2016; URL: http://arxiv.org/abs/1701.00160

[9] L. Perez, J. Wang, "The effectiveness of data augmentation in image classification using deep learning," Proceedings, 2017. URL:http://arxiv.org/abs/1712.04621

[10] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time-series," in the Handbook of Brain Theory And Neural Networks, A. Arbib, Ed. 1995, MIT Press.

[11] V. Turchenko, E. Chalmers, & A. Luczak, "A deep convolutional auto-encoder with pooling – unpooling layers in Caffe," *International Journal of Computing*, vol. 18, issue 1, pp. 8-31, 2019. https://doi.org/10.47839/ijc.18.1.1270

[12] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng and F. -Y. Wang, "Generative adversarial networks: introduction and outlook," IEEE/CAA Journal of Automatica Sinica, vol. 4, no. 4, pp. 588-598, 2017, doi: 10.1109/JAS.2017.7510583.

[13] V. Golovko, Y. Savitsky, T. Laopoulos, A. Sachenko, L. Grandinetti, "Technique of learning rate estimation for efficient training of MLP," *Proceedings of the International Joint Conference on Neural Networks*, 2000, 1, pp. 323-328.

[14] B. M. Pavlyshenko, "Sales time series analytics using deep Q-learning," *International Journal of Computing*, vol. 19, issue 3, pp. 434-441, 2020. https://doi.org/10.47839/ijc.19.3.1892

[15] M. Komar, P. Yakobchuk, V. Golovko, V. Dorosh, A. Sachenko, "Deep neural network for image recognition based on the Caffe framework," *Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP'2018*, 2018, pp. 102-106.

[16] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322-36333, 2019, doi: 10.1109/ACCESS.2019.2905015.

[17] S. Anfilets, S. Bezobrazov, V. Golovko, A. Sachenko, M. Komar, R. Dolny, V. Kasyanik, P. Bykovyy, E. Mikhno, & O. Osolinskyi, "Deep multilayer neural network for predicting the winner of football matches," *International Journal of Computing*, vol. 19, issue 1, pp. 70-77. https://doi.org/10.47839/ijc.19.1.1695

[18] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi and N. Harada, "Unsupervised detection of anomalous sound based on deep learning and the Neyman–Pearson lemma," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212-224, 2019, doi: 10.1109/TASLP.2018.2877258.