Generative Adversarial Network-based Semi-Supervised Learning for Pathological Speech Classification*

Nam H. Trinh and Darragh O'Brien

ADAPT Centre, School of Computing, Dublin City University
nam.trinh@adaptcentre.ie, darragh.obrien@dcu.ie

Abstract. A challenge in applying machine learning algorithms to pathological speech classification is the labelled data shortage problem. Labelled data acquisition often requires significant human effort and time-consuming experimental design. Further, for medical applications, privacy and ethical issues must be addressed where patient data is collected. While labelled data are expensive and scarce, unlabelled data are typically inexpensive and plentiful. In this paper, we propose a semi-supervised learning approach that employs a generative adversarial network to incorporate both labelled and unlabelled data into training. We observe a promising accuracy gain with this approach compared to a baseline convolutional neural network trained only on labelled pathological speech data.

Keywords: Semi-supervised Learning · Generative Adversarial Networks · Pathological Speech Classification.

1 Introduction

Deep learning for healthcare applications has attracted significant research effort in recent years [14, 24, 26]. One such application is the use of neural networks for pathological speech classification. A challenge in this field is the scarcity of labelled training data [4, 7, 19, 30]. Labelled medical data acquisition often requires significant human expertise and raises privacy and ethical concerns.

While labelled data availability is limited, unlabelled data are typically plentiful. Semi-supervised learning (SSL), incorporating both labelled and unlabelled data [33, 34], presents a potential means of alleviating the labelled data shortage problem and thus improving overall classification performance in pathological speech classification when faced with a limited training dataset. Recently, Generative Adversarial Networks (GANs) (introduced in [9]) have been applied for SSL and have achieved considerable success with benchmark image datasets, e.g. MNIST, CIFAR-10, and SHVN.

In this paper, we explore a GAN-based SSL approach for pathological speech classification that attempts to mitigate the data shortage problem. We evaluate

* Supported by the ADAPT Research Centre and funded by Science Foundation Ireland (SFI) under grant No. 17/RC/PHD/3488.

our proposed approach by comparing its performance with that of a baseline CNN under the same training configuration. Our contributions are:

- An approach to applying GAN-based semi-supervised learning for pathological speech classification,
- An empirical experiment comparing the performance of the proposed approach with that of a baseline CNN using three popular pathological speech datasets.

The paper is organized as follows: we summarize some related work in Section 2. In Section 3, we present the proposed GAN-based SSL approach for pathological speech classification. In Section 4, we describe our experimental design and results. Section 5 concludes the paper.

2 Related Work

In general, pathological speech classification firstly requires salient feature extraction (as illustrated in Figure 1). During feature extraction, raw speech signals are typically converted from the time-domain into frequency-domain features (by means of, for example, the Fourier transform). Frequency-domain features are then fed into a classifier. We summarize in Table 1 related work including relevant details on datasets, features, classifier design and resulting classification accuracy.

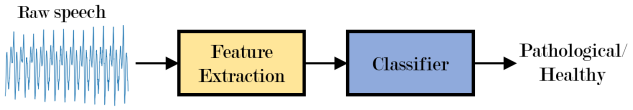


Fig. 1. A general pathological speech classification system

Generative Adversarial Networks (GANs) [9] have been employed in SSL and have been shown capable of contributing considerable improvements in overall classification performance using benchmark image datasets such as MNIST, CIFAR-10 and SHVN. In [25], several new architectural features and training procedures were proposed in order to boost GAN performance in a semi-supervised setting. In [29], SSL incorporating a GAN, specifically a Categorical GAN or CatGAN, was proposed. For the SSL task the GAN's discriminator, a binary classifier, was replaced with a $(K+1)$-class classifier (where K is the number of classes to be classified). This approach demonstrated a significant improvement in accuracy compared to traditional classifiers in image classification tasks. In [20], the proposed GAN-based approach outperformed traditional classifiers at the MNIST classification task. In [13], the proposed GAN-based SSL method with manifold invariance achieved accuracy gains with CIFAR-10 and SHVN datasets. In [6], the proposed GAN method along with a complementary generator improved

Table 1. Related work in pathological speech classification: features, classifiers and reported accuracy

Reference	Dataset	Features	Classifier	Accuracy
Poorjam et al. (2018) [22]	Data collected by the authors in collaboration with Sage Bionetworks	MFCCs	SVM	88.0%
Moon et al. (2018) [17]	SVD [3]	Jitter, shimmer and MFCCs	MLP	87.4%
Smitha et al. (2018) [28]	Supplied by the Nitte Institute of Speech and Hearing Mangaluru	MFCCs	MLP	95.0%
Shia et al. (2017) [27]	SVD	Wavelet Sub-band Energy Coefficients	MLP	93.3%
Alhussein et al. (2018) [1]	SVD	Spectrogram (after framing and applying STFT)	CNN	97.5%
Trinh et al. (2019) [31]	SVD SPDD [18]	Spectrogram	CNN	99.0% 96.7%

the overall performance in image classification tasks. Recently, MarginGAN [8] (based on margin theory) achieved high accuracy compared to other SSL methods. Besides GANs, variational inference generative methods such as Variational Autoencoders (VAE) have also been tested in an SSL context [12]. In [2], the proposed approach using sequence to sequence autoencoders for representation learning achieved a promising accuracy gain with the acoustic scene classification task.

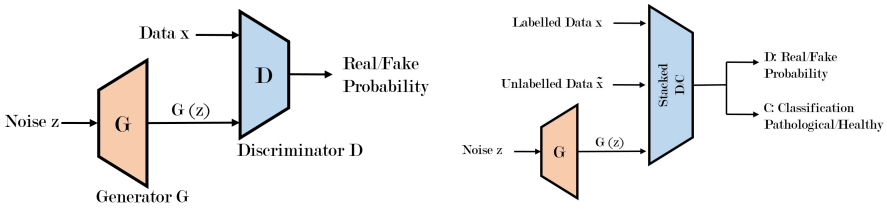
Semi-supervised approaches have been applied in medical imaging. In [4], the authors report a significant improvement in medical imaging segmentation thanks to SSL. In [30], a graph-based SSL approach incorporating a CNN was proposed for breast cancer diagnosis. In [19], an attention-based SSL approach achieves state-of-the-art results on real clinical segmentation datasets. Work to-date in pathological speech classification has typically assumed an adequate corpus of pathological speech data. In our previous work [32], we presented preliminary results where a semi-supervised method was applied to mitigate the data shortage problem. In this paper, we further explore and extend the GAN-based SSL approach by testing against three popular pathological speech datasets.

3 Methodology

In this section, we describe our method for modifying the traditional GAN architecture to fit the task of semi-supervised pathological speech classification.

3.1 Architecture Overview

The original GAN [9] architecture is illustrated in Figure 2a. A GAN is a generative model taking random noise as input and seeking to generate a real data distribution. A vanilla GAN consists of a discriminator and a generator. The generator takes random noise as input and generates new data samples. The discriminator's objective is to discriminate between real and generated samples (provided by the generator), classifying them as real or fake, respectively. The two networks compete with each other until an equilibrium is reached where the discriminator cannot reliably discriminate between real and fake data.



(a) The original GAN (b) The proposed Semi-Supervised GAN

Fig. 2. Architecture Overview

Let D be the discriminator and G be the generator. The minimax game between D and G is modelled mathematically as follows:

$$\min_G \max_D V(G,D) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[log(1-D(G(z)))] \quad (1)$$

where $E_{x\sim p_{data}(x)}$ is the expected value over all real data samples with a data distribution $p_{data}(x)$, $D(x)$ is the probability that a real data sample is classified as real, $E_{z\sim p_z(z)}$ is the expected value over all noise samples with a prior noise distribution $p_z(z)$, $G(z)$ is the generated output from the generator from input noise z. The objective of the training process is to train D to maximize the probability of classifying generated samples $G(z)$ as fake and data samples x as real and to train G to convince D that generated samples, $G(z)$, are real. In other words, D is trained to maximize the loss function (1) while G is trained to minimize (1).

Semi-supervised GAN To mitigate the problem of a shortage of training data, unlabelled and labelled data are incorporated into the training process in order to enhance the classification decision boundary (depicted in Figure 3). By incorporating unlabelled data, the semi-supervised model can shift the decision boundary to better cluster the data distribution [34]. This can be viewed as the model attempting to first cluster the data before subsequently identifying the decision boundary by assuming that unlabelled data points carry the same label as the labelled data region they reside closest to.

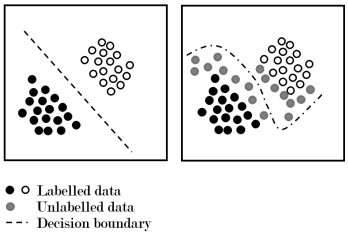


Fig. 3. Data points in supervised learning with a limited amount of labelled data (left) and in semi-supervised learning with labelled data and unlabelled data (right) [34]

A GAN-based approach for semi-supervised learning (as illustrated in Figure 2b) incorporates data supplied from the GAN's generator and feeds the latter, along with labelled and unlabelled data, into the discriminator. In this work, we modify the discriminator to not only classify a data sample as real or fake (as in the original GAN formulation) but to also classify that sample as healthy or pathological. Using the same method outlined in [25], we modify the discriminator's architecture by adding an output layer in parallel with the output layer responsible for real/fake classification in order to classify speech data as pathological or healthy. This can be considered as a stacking of a discriminator D (for real/fake discrimination) and a classifier C (for healthy/pathological classification).

As shown in Figure 5, the weights of the two networks (D and C) are shared across the input layer to the last hidden layer. Following the hidden layer, the output layers of D and C are separated. A detailed description of this implementation is presented in Section 4. The shared weight structure ensures that as D learns a feature representation from the unlabelled data, D shares that representation with C and helps C improve its feature learning compared with C being trained on only limited labelled data.

3.2 Loss Functions

We train D to maximize the probability that D classifies both labelled data x and unlabelled data $\tilde{x}$ as real but generated data $G(z)$ as fake. We train C to classify the labelled data as healthy or pathological. We train G to maximize the probability that D will classify generated samples $G(z)$ as real. We derive the loss functions for D, C and G as follows:

$$\begin{aligned} Loss(D) = -(&\mathbb{E}_{x\sim p_l(x)}[logD(x)] + \mathbb{E}_{\tilde{x}\sim p_u(\tilde{x})}[logD(\tilde{x})] \\ &+\mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))]) \end{aligned} \tag{2}$$

$$Loss(C) = -\mathbb{E}_{(x,y)\sim p_l(x,y)}[ylogC(x)] \tag{3}$$

$$Loss(G) = \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))] \tag{4}$$

where $p_u(\tilde{x})$ and $p_l(x)$ are unlabelled data and labelled distributions, $p_z(z)$ is the prior Gaussian noise distribution, $\mathbb{E}_{x\sim p_l(x)}$ is the expected value over all labelled data, $\mathbb{E}_{\tilde{x}\sim p_u(\tilde{x})}$ is the expected value over all unlabelled data, $\mathbb{E}_{z\sim p_z(z)}$ is the expected value over all noise samples, $\mathbb{E}_{(x,y)\sim p_l(x,y)}$ is the expected value over all labelled data points (x, y), $G(z)$ is the generated sample from the generator G, D is the probability that the discriminator classifies a data sample as real and $C(x)$ is the pathological/healthy classification result. The minimax game equation for the proposed semi-supervised GAN model is as follows:

$$\begin{aligned}\min_G \max_{D,C} J(G, D, C) = \mathbb{E}_{x\sim p_l(x)}[logD(x)] + \mathbb{E}_{\tilde{x}\sim p_u(\tilde{x})}[logD(\tilde{x})] \\ + \mathbb{E}_{(x,y)\sim p_l(x,y)}[ylogC(x)] \\ + \mathbb{E}_{z\sim p_z(z)}[log(1 - D(G(z)))]\end{aligned} \tag{5}$$

4 Experiments and Results

In this section, we describe our experiments applying the approach above to three popular pathological speech datasets. We compare the performance of the GAN-based SSL approach with that of a baseline CNN (that shares the same architecture as the GAN's discriminator in order to ensure results produced by the two approaches are comparable).

4.1 Datasets

The Spanish Parkinson's Disease Dataset (SPDD) [21] SPDD consists of speech samples from 50 Parkinson's disease patients and 50 healthy controls, 25 men and 25 women per group. All subjects are Colombian native Spanish speakers. Several types of speech recordings are included in the dataset:

- sustained vowels including /a/, /u/, /i/, /e/ and /o/,
- some specific words and phonemes,
- conversational speech.

We use speech data extracted from sustained vowel /a/ recordings at 44100 Hz as labelled data and from other sustained vowels /u/, /i/, /e/ and /o/ as unlabelled data in the experiments described below.

The Saarbrucken Voice Database (SVD) [3] SVD is a collection of speech samples from more than 2000 people including healthy and pathological speech samples (with 71 different voice pathologies). There are three types of recordings in the dataset:

- sustained vowel sounds (/a/, /u/ and /i/) at normal, high and low pitch,
- sustained vowel sounds (/a/, /u/ and /i/) at rising-falling pitch,
- a conversational sentence in German.

In our work, we make use of a subset of SVD data comprising of 50 pathological speech samples and 53 healthy speech samples from the sustained vowel /a/ as labelled data and sustained vowels /u/ and /i/ at different pitches as unlabelled data.

The Arabic Voice Disorder Dataset (AVPD) [16,18] AVPD is a collection of 353 normal and disordered speech samples. Types of voice disorders in this dataset are cysts, nodules, paralysis, polyps and sulcus. Three types of speech recordings are included:

- sustained vowel sounds (/a/, /u/ and /i/),
- isolated words including Arabic digits and common words,
- continuous speech.

Similar to SVD and SPDD, we also use sustained vowel /a/ samples as labelled data and sustained vowel /u/ and /i/ samples as unlabelled data.

For all three datasets, we include both healthy and pathological samples in both labelled and unlabelled sets.

4.2 Experimental Design

Speech Spectrogram Extraction Our chosen feature representation is the spectrogram. To extract spectrograms from raw speech, we use the librosa [15] speech processing framework. The Short-time Fourier Transform is calculated with 128 frequency components. The extracted feature matrices (with a shape $(128, 96)$) are then zero-padded to obtain $(128, 128)$ square matrices.

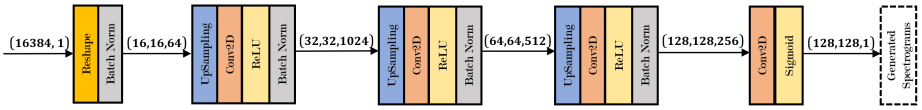


Fig. 4. The generator

Semi-Supervised GAN The proposed semi-supervised GAN includes a stacked discriminator/classifier and a generator as shown in Figure 2b. Our GAN's architecture is inspired by that of the DCGAN [23]. The architectures of the generator and the discriminator are shown in Figures 4 and 5.

The architecture of the generator is depicted in Figure 4. The generator's input is a Gaussian noise vector of shape $(16384, 1)$. The latter is reshaped to a square tensor of shape $(16, 16, 64)$. Next, three stages of upsampling are applied to increase the data dimension from $(16, 16, 64)$ to $(128, 128, 256)$. Each stage includes an UpSampling layer followed by a convolutional layer with ReLU activation and a batch normalization layer [10]. We finally apply a convolutional layer with a sigmoid activation function. The output of the generator is a tensor of shape $(128, 128, 1)$.

The architecture of the discriminator is shown in Figure 5. The input to the discriminator has shape $(128, 128, 1)$. We employ successive 2D convolutional layers with filter numbers of $32, 64, 128, 256$ and 512 respectively. To the output of each convolutional layer, we apply LeakyReLU with an alpha of 0.2, a drop-out

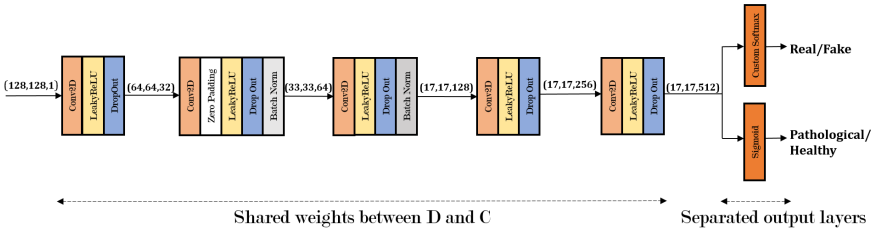


Fig. 5. The stacked discriminator D and classifier C

layer with a rate of 0.25 and a batch normalization layer with a momentum of 0.8. The final output is then flattened and a copy sent in two directions: to a discriminator for classification as fake or real and to a second classifier for pathological/healthy classification. For pathological speech classification, the final output layer is a single neuron with a sigmoid activation function for binary classification. For real/fake discrimination, we create a custom softmax layer to output the probability of data being real.

Baseline CNN To implement the baseline CNN, we reuse the GAN's discriminator architecture. This ensures results produced by the SSL and baseline approaches are comparable. The baseline is trained only on labelled data.

Training Configuration For each dataset, we train our models in 100 epochs, with a batch size of 32, with the Adam optimizer [11] and with a learning rate of 0.00002. Across experiments, we reduce the number of labelled spectrogram samples for training from 1000 through 800, 600, 400 and 200 and test on 800 spectrogram samples. We use 20000 unlabelled spectrograms (without healthy/pathological labels) as unlabelled data to train the proposed SSL approach.

4.3 Results

Generative Results We present, for visual inspection, in Figure 6 sample spectrograms produced by the generator trained on the SPDD alongside original spectrograms extracted from the dataset. Similar frequency content is observed.

Classification Accuracy Accuracies obtained for the three datasets SPDD, SVD and AVPD are presented in Tables 2, 3 and 4 respectively. We compare the classification accuracy of the proposed semi-supervised GAN approach with that of the baseline CNN. We also compare the accuracy of the proposed approach against two additional classifiers previously proposed in the literature [1, 31]. We observe an accuracy gain across all three datasets. The accuracy gains with only 400 and 200 labelled data samples are significant across all three datasets.

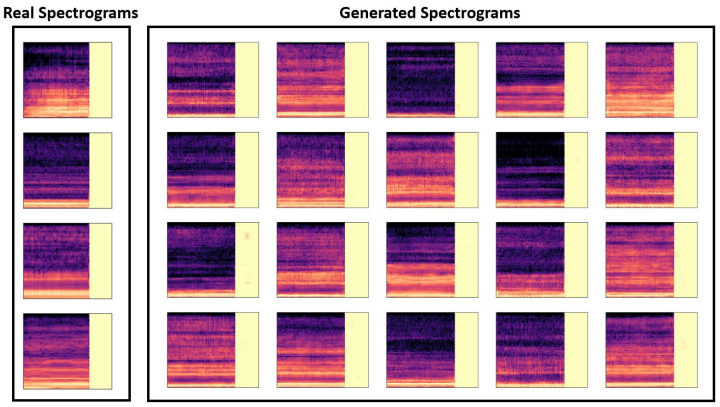


Fig. 6. Original spectrograms (left) and generated spectrograms (right)

Table 2. SPDD classification accuracy

Approach	Number of labelled data samples				
	1000	800	600	400	200
CNN [31]	0.896	0.835	0.851	0.798	0.705
VGG16-based CNN [1]	0.925	0.923	**0.929**	0.873	0.769
Baseline CNN	0.914	0.874	0.855	0.788	0.746
Proposed GAN-based SSL	**0.951**	**0.942**	0.919	**0.890**	**0.833**

Table 3. SVD classification accuracy

Approach	Number of labelled data samples				
	1000	800	600	400	200
CNN [31]	0.976	0.967	0.974	0.942	0.862
VGG16-based CNN [1]	**1.00**	**1.00**	0.993	0.984	0.946
Baseline CNN	1.00	0.998	0.985	0.973	0.939
Proposed GAN-based SSL	**1.00**	**1.00**	**0.999**	**0.998**	**0.960**

Table 4. AVPD classification accuracy

Approach	Number of labelled data samples				
	1000	800	600	400	200
CNN [31]	0.984	0.939	0.939	0.920	0.870
VGG16-based CNN [1]	**0.991**	0.991	0.978	0.963	0.860
Baseline CNN	0.990	0.986	0.966	0.944	0.818
Proposed GAN-based SSL	**0.991**	**0.998**	**0.993**	**0.971**	**0.889**

Ablation Study by removal of unlabelled data To study the effect of unlabelled data on the training, we remove the unlabelled data in an experiment using the SPDD data to observe any drop in the classification performance. The result of the ablation study is presented in Table 5. We observe a significant drop

in the accuracy obtained, especially when training on only 400 and 200 labelled samples and without unlabelled data. This result further validates the effect of unlabelled data on improving the classification performance.

Table 5. SPDD Ablation Study

Proposed GAN-based SSL	Number of labelled data samples				
	1000	800	600	400	200
w/ unlabelled	**0.951**	**0.942**	**0.919**	**0.890**	**0.833**
w/o unlabelled	0.934	0.940	0.899	0.866	0.734

5 Conclusion and Future Work

This paper describes a proposed GAN-based semi-supervised approach for pathological speech classification tasks. Results are presented that indicate the approach has the potential to mitigate the labelled data shortage problem faced by certain medical applications of deep learning. A GAN is incorporated into SSL by replacing the former's traditional binary discriminator with a multi-class discriminator that not only classifies a sample as real or fake but also categorizes that sample as healthy or pathological. We test the approach against three commonly used pathological speech datasets: SPDD, SVD and AVPD. Comparing the performance of our GAN-based approach with a baseline CNN and two additional classifiers previously proposed in the literature [1, 31], we observe a promising improvement in accuracy when we decrease the number of labelled training samples from 1000 through 800, 600, 400 and 200.

Future work will evaluate the performance of alternative GAN architectures (e.g. infoGAN [5] and marginGAN [8]) in semi-supervised pathological speech classification. Feature matching [25] will be explored as a means to improve discriminator performance. The proposed approach has potential applications not only in pathological speech classification but also across other audio classification tasks.

Acknowledgement

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme(Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. Nam H. Trinh is funded by Science Foundation Ireland under grant No. 17/RC/PHD/3488.

References

1. Alhussein, M., Muhammad, G.: Voice pathology detection using deep learning on mobile healthcare framework. IEEE Access **6**, 41034–41041 (2018)

2. Amiriparian, S., Freitag, M., Cummins, N., Schuller, B.: Sequence to sequence autoencoders for unsupervised representation learning from audio. In: Proc. of the DCASE 2017 Workshop (2017)
3. Barry, W., Pützer, M.: Saarbrucken voice database. Institute of Phonetics, Universität des Saarlandes, http://www. stimmdatenbank. coli. uni-saarland. de (2007)
4. Baur, C., Albarqouni, S., Navab, N.: Semi-supervised deep learning for fully convolutional networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 311–319. Springer (2017)
5. Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., Abbeel, P.: Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In: Advances in neural information processing systems. pp. 2172–2180 (2016)
6. Dai, Z., Yang, Z., Yang, F., Cohen, W.W., Salakhutdinov, R.R.: Good semi-supervised learning that requires a bad gan. In: Advances in neural information processing systems. pp. 6510–6520 (2017)
7. Deo, R.C.: Machine learning in medicine. Circulation **132**(20), 1920–1930 (2015)
8. Dong, J., Lin, T.: Margingan: Adversarial training in semi-supervised learning. In: Advances in Neural Information Processing Systems. pp. 10440–10449 (2019)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
10. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Kingma, D.P., Mohamed, S., Rezende, D.J., Welling, M.: Semi-supervised learning with deep generative models. In: Advances in neural information processing systems. pp. 3581–3589 (2014)
13. Kumar, A., Sattigeri, P., Fletcher, T.: Semi-supervised learning with gans: Manifold invariance with improved inference. In: Advances in Neural Information Processing Systems. pp. 5534–5544 (2017)
14. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
15. McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O.: librosa: Audio and music signal analysis in python. In: Proceedings of the 14th python in science conference. vol. 8 (2015)
16. Mesallam, T.A., Farahat, M., Malki, K.H., Alsulaiman, M., Ali, Z., Al-nasheri, A., Muhammad, G.: Development of the arabic voice pathology database and its evaluation by using speech features and machine learning algorithms. Journal of healthcare engineering **2017** (2017)
17. Moon, J., Kim, S.: An approach on a combination of higher-order statistics and higher-order differential energy operator for detecting pathological voice with machine learning. In: 2018 International Conference on Information and Communication Technology Convergence (ICTC). pp. 46–51 (Oct 2018). https://doi.org/10.1109/ICTC.2018.8539495
18. Muhammad, G., Alsulaiman, M., Ali, Z., Mesallam, T.A., Farahat, M., Malki, K.H., Al-nasheri, A., Bencherif, M.A.: Voice pathology detection using interlaced derivative pattern on glottal source excitation. Biomedical signal processing and control **31**, 156–164 (2017)

19. Nie, D., Gao, Y., Wang, L., Shen, D.: Asdnet: Attention based semi-supervised deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 370–378. Springer (2018)
20. Odena, A.: Semi-supervised learning with generative adversarial networks. arXiv preprint arXiv:1606.01583 (2016)
21. Orozco-Arroyave, J.R., Arias-Londoño, J.D., Bonilla, J.F.V., Gonzalez-Rátiva, M.C., Nöth, E.: New spanish speech corpus database for the analysis of people suffering from parkinson's disease. In: In Proc. Of the International Confer- ence on Language Resources and Evaluation (lrec). pp. 342–347. Reykjavik, Iceland (2014)
22. Poorjam, A.H., Little, M.A., Jensen, J.R., Christensen, M.G.: A parametric approach for classification of distortions in pathological voices. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 286–290. IEEE (2018)
23. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
24. Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., Yang, G.Z.: Deep learning for health informatics. IEEE journal of biomedical and health informatics **21**(1), 4–21 (2016)
25. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in neural information processing systems. pp. 2234–2242 (2016)
26. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering **19**, 221–248 (2017)
27. Shia, S.E., Jayasree, T.: Detection of pathological voices using discrete wavelet transform and artificial neural networks. In: 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). pp. 1–6 (March 2017). https://doi.org/10.1109/ITCOSP.2017.8303086
28. Smitha, Shetty, S., Hegde, S., Dodderi, T.: Classification of healthy and pathological voices using mfcc and ann. In: 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC). pp. 1–5 (Feb 2018). https://doi.org/10.1109/ICAECC.2018.8479441
29. Springenberg, J.T.: Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv preprint arXiv:1511.06390 (2015)
30. Sun, W., Tseng, T.L.B., Zhang, J., Qian, W.: Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Medical Imaging and Graphics **57**, 4–9 (2017)
31. Trinh, N., O'Brien, D.: Pathological speech classification using a convolutional neural network. In: IMVIP 2019: Irish Machine Vision & Image Processing. Technological University Dublin, Dublin, Ireland (August 2019). https://doi.org/10.21427/9dnc-n002
32. Trinh, N., O'Brien, D.: Semi-supervised learning with generative adversarial networks for pathological speech classification. In: 31st Irish Signals and Systems Conference (ISSC2020). Letterkenny, Ireland (Virtual) (June 2020)
33. Zhu, X., Goldberg, A.B.: Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning **3**(1), 1–130 (2009)
34. Zhu, X.J.: Semi-supervised learning literature survey. Tech. rep., University of Wisconsin-Madison Department of Computer Sciences (2005)