INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

A Pipeline to Evaluate the Effects of Noise on Machine Learning Detection of Laryngeal Cancer

Mary Paterson[1], *James Moor*[2], *Luisa Cutillo*[1]

[1]University of Leeds
[2]Leeds Teaching Hospitals NHS Trust

scmlp@leeds.ac.uk, j.w.moor@leeds.ac.uk, l.cutillo@leeds.ac.uk

Abstract

Cases of laryngeal cancer are rising, with diagnosis often involving invasive biopsy procedures. An alternate approach is to identify high-risk patients by analysis of voice recordings which can alert clinical teams to those patients that need prioritisation. We propose a pipeline for evaluating speech classifier performance in the presence of noise. We perform experiments using the pipeline with several classifiers and denoising techniques. Random forest classifier performed best with an accuracy of 81.2% on clean data dropping to 63.8% when noise was added to recordings. The accuracy of all classifiers was reduced by added noise, signal denoising improved classifier accuracy but could not fully reverse the effects of noise. The effects of noise on classification is a complex issue which must be resolved for these detection systems to be implemented in clinical practice. We show that the proposed pipeline allows for the evaluation of classifier performance in the presence of noise.

Index Terms: machine learning, laryngeal cancer detection, audio denoising, voice classification

1. Introduction

Laryngeal cancer is the 18th most prevalent cancer world wide with prevalence increasing by 23.8% in the last three decades [1]. While approximately 90% of adults with stage one laryngeal cancer will survive for five or more years after diagnosis, this rate drops significantly to 30% for those with stage four laryngeal cancer [2]. With increasing numbers of cases, it is important that early diagnosis tools are developed to increase oncological cure and reduce disease related mortality.

Current diagnostic techniques include nasendoscopy, laryngoscopy, and biopsy [3]. Nasendoscopy is performed as an outpatient procedure where a small fibre-optic camera is used to view the larynx via the nose. A direct laryngoscopy procedure is performed under general anaesthetic in the operating theatre - both techniques can yield biopsy specimens for histological analysis necessary for formal diagnosis [3].

In order to reduce the need for potentially invasive and uncomfortable medical tests we hope that machine learning (ML) and artificial intelligence (AI) analysis of voice may be used as non-invasive screening tool for the detection of laryngeal cancer. The use of such a tool may also be able to screen patients with concerns regarding their voice quality, prioritise those at highest risk of a cancer diagnosis, expediting their specific care pathway, and increase the accessibility of diagnosis by reducing the need for expensive medical equipment. Indeed several papers have presented ML and AI methods for detecting laryngeal cancer from speech [4, 5, 6, 7, 8]. All of these papers use data from a single dataset where the speech has been recorded in controlled environments. However it is unlikely that the application of these tools in clinical practice would be feasible in an acoustically controlled environment. To ensure these tools are accessible and consistently produce high quality results it is important that they are robust to different recording environments and recording device quality. Therefore, in this work we propose a pipeline which can be used for the evaluation of classifier performance in the presence of noise. We then use this pipeline in experiments evaluating several classifiers and denoising algorithms.

2. Methodology

Figure 1 shows the pipeline proposed in this work. The pipeleine comprises of seven steps - splitting the data, adding noise, denoising, preprocessing, feature extraction, classifier creation, and evaluation as described in the following subsections.

2.1. Test train split

The first step of our pipeline is splitting the data into a test group and a train group. Patients should be randomly allocated to either the train or test group.

2.2. Adding noise

The next step in our pipeline is adding noise to each of the recordings in the test set. Noise should be added to the recordings in the test set to simulate noisy recordings captured in the real world.

2.3. Denoising

The noisy signals created in the previous step are considered as proxies for real noisy data. As such, denoising algorithms are applied to these recordings in an attempt to recover the clean signal.

We implement three denoising algorithms: bandpass filter, Wiener filter, and wavelet filter. A bandpass filter is a passive filter which removes frequencies outside of a given range [9]. In this application the $Butterworth$ filter is used from the scipy signal library [10]. The Wiener filter assumes that the input signals can be modelled as stationary stochastic processes with known power spectral densities [11]. The scipy function $Wiener$ was used to implement the filter in this work [10]. Discrete wavelet transforms allow time series to be analyzed into wavelet coefficients. These coefficients can then be used to denoise signals [12]. In this work the package $PyWavelets$ was used for denoising [13]. The Wiener and Butterworth filters were chosen due to their use in similar works [14, 15]. While wavelet denoising has not been applied in this field it is widely used in audio denoising [16, 17, 18].

10.21437/Interspeech.2023-1574

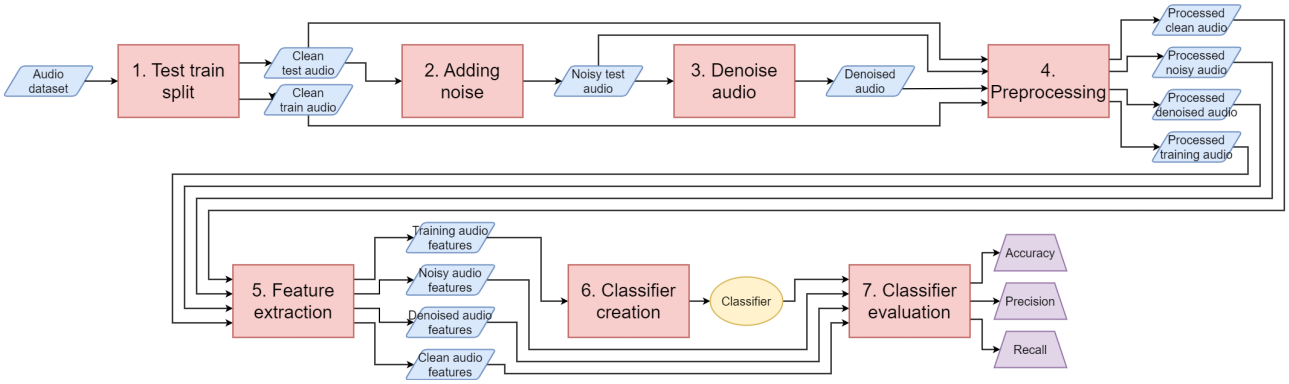


Figure 1: *The methodology pipeline used in this work.*

2.4. Preprocessing

Each of the recordings is preprocessed before features are extracted. Common preprocessing techniques include endpoint detection and normalization. Endpoint detection is commonly used to remove the leading and trailing silences in the recordings [19, 20, 21]. Normalizing signals such that that their amplitude is between -1 and 1, reduces effects caused by the distance between the microphone and the patient [22, 23].

2.5. Feature Extraction

Before classification, features are extracted from each signal. We extracted 22 acoustic features using *Parselmouth* [24]. These features include 1 pitch feature, 4 pulse features, 3 voicing features, 5 jitter features, 6 shimmer features, and 3 harmonicity features. For some patients the hoarseness of their voice means that no features can be extracted. For these patients the prediction is set to 'cancer'.

2.6. Classifier Creation

The classifier is built using the training set. We implement four classifiers: support vector machine (SVM), decision tree (DT), random forest (RF), and logistic regression (LR). These methods were chosen as they are simple, work well with small datasets, and have been used in similar applications [4, 25, 26].

2.7. Classifier Evaluation

To evaluate classifier performance three metrics are calculated: accuracy, precision, and recall. Each classifier is evaluated using these metrics on the clean, noisy, and denoised audio test sets. By calculating precision and recall as well as accuracy the specific effects of noise on false positive and false negative rates can be investigated.

3. Experiments

In order to test this pipeline we conducted several experiments. Experiments were conducted using python 3.9.16, all library versions are stated in the github repository: github.com/mary-paterson/Interspeech2023-EvaluationPipeline

3.1. Dataset

The experiments carried out in this work use the *Saarbruecken Voice Database* (SVD). This is a German dataset containing recordings of more than 2000 people with over 60 different pathologies [27]. Each patient is recorded producing three vowel sounds /i/, /a/, and /u/. From this dataset six malignant and two pre-malignant conditions were identified by an experienced clinician. The eight pathologies and the number of male and female patients with each of the pathologies can be seen in Table 1.

Table 1: *The number of patients in the dataset per condition. This is also split into the number of male and female patients per condition.*

Condition	Condition Type	Male	Female	**Total**
Vocal cord cancer	Malignant	21	1	**22**
Hypopharyngeal tumor	Malignant	6	0	**6**
Larynx tumor	Malignant	4	1	**5**
Epiglottic cancer	Malignant	0	1	**1**
Mesopharyngeal tumor	Malignant	1	0	**1**
Carcinoma in situ	Malignant	1	0	**1**
Dysplastic dysphonia	Pre-malignant	1	0	**1**
Dysplastic larynx	Pre-malignant	1	0	**1**

The cancer patients range in age from 38 to 75 years with a mean of 59, while healthy patients range in age from 9 to 84 years with a mean age of 28. A t-test shows that there is a significant difference between the ages of the cancer and healthy patients (p=1.758e-51).

The number of male and female subjects in the cancer and healthy groups can be seen in Table 2. A Barnard exact test shows an association between gender and cancer (p=0.002). Cancer statistics from Germany in the year the majority of the recordings were taken show a similar gender and age distribution to that found in the SVD. The majority of people diagnosed with throat cancer were between the ages of 60 and 64 with only 13% of cancer diagnoses being in women [28].

Table 2: *The gender balance between the healthy and cancer groups before and after matching takes place.*

	Unmatched		Matched	
	Male	Female	Male	Female
Healthy	437 (49.9%)	438 (50.1%)	26 (68.4%)	12 (31.6%)
Cancer	35 (92.1%)	3(7.9%)	35 (92.1%)	3 (7.9%)

3.1.1. Matching

To mitigate the effects of the gender and age imbalance between the cancer and healthy group, propensity score matching was used to select patients for the healthy group. After matching the age range of the healthy patients was 38-84 with a mean age of 60. A t-test shows that there is no significant difference of age between the two groups after matching (p=0.859).

Table 2 shows the distribution of male and female patients in the healthy and cancer groups after matching. A Barnard exact test shows that there is still an association between gender and cancer after matching (p=0.010), however, the difference in proportion between the unmatched and matched groups is almost halved (0.422 to 0.237). This suggests a weaker association after matching.

By using matching to select the healthy patients we have created a set with the same number of cancer patients and healthy patients (38 in each group). In clinical settings it is likely that there would be many more non-cancer patients compared to cancer patients, however, due to the limited data available we feel that it is most important that the two groups have a similar demographic distribution.

3.1.2. Test Train Split

Patients were randomly allocated to either the test or train set with 30% of the patients being used for testing and the remaining 70% being used for training. This meant that 53 patients were used for training the classifiers and the remaining 23 patients were used for testing. Since each patient was recorded making three different vowel sounds this meant that 69 recordings were available for testing and 159 recordings were used for training.

3.2. Classifiers

For these experiments four classifiers are implemented: DT, RF, SVM, and LR. The DT classifier used the entropy function and had a maximum depth of 7 this classifier trained in 0.004s. The RF consisted of 45 trees and trained in 0.077s. The SVM used a linear kernel with a C value of 100 and gamma of 1 and trained in 8.193s. The LR used an L2 penalty and C value of 1 and trained in 0.015s.

4. Results and Discussion

In this section we discuss the results of the experiments. First we discuss the results of the classification on clean data and discuss any effects that patient demographic may have on misclassification. Secondly we discuss the classification of noisy data. Thirdly the effects of denoising algorithms on signal quality are discussed. Finally we discuss the effects of denoising algorithms on the classification of noisy signals.

4.1. Classification

Table 3 shows the accuracy, precision, and recall of all four classification algorithms on clean, noisy, and denoised data. The best performing classifier on the clean data is the RF with an accuracy of 81.2%. This is followed by the DT classifier with an accuracy of 75.4%.

For each classifier we investigated whether the gender of the patients had an impact on the misclassification rate via a Barnard exact test. Similarly we checked if the age of the patients was significantly different in the correctly and incorrectly classified patients via a t-test. The results of these test are shown in Table 4. The p-values calculated indicate that there is not a significant association between the gender of the patient and the rate of misclassification. Moreover, there is not a significant difference between the ages of the correctly and incorrectly classified patients.

Table 3: *The results of classification on clean, noisy, and denoised recordings. The bold figures show the highest accuracy found after denoising. RF - random forest, DT - decision tree, SVM - support vector machine, LR - logistic regression.*

Model	Metric	Clean	Noisy	Bandpass	Wiener	Wavelet
DT	Accuracy	75.4%	63.8%	66.7%	**68.1%**	63.8%
	Precision	77.5%	63.0%	65.4%	67.3%	62.5%
	Recall	79.5%	87.2%	87.2%	84.6%	89.7%
RF	Accuracy	81.2%	63.8%	71.0%	69.6%	**73.9%**
	Precision	81.0%	62.1%	67.3%	66.1%	69.1%
	Recall	87.2%	92.3%	94.9%	94.9%	97.4%
SVM	Accuracy	60.9%	53.6%	55.1%	**58.0%**	50.7%
	Precision	71.4%	56.6%	57.4%	59.6%	54.5%
	Recall	51.3%	76.9%	79.5%	79.5%	76.9%
LR	Accuracy	63.8%	47.8%	52.2%	53.6%	**59.4%**
	Precision	75.0%	52.6%	55.4%	55.9%	59.0%
	Recall	53.8%	76.9%	79.5%	84.6%	92.3%

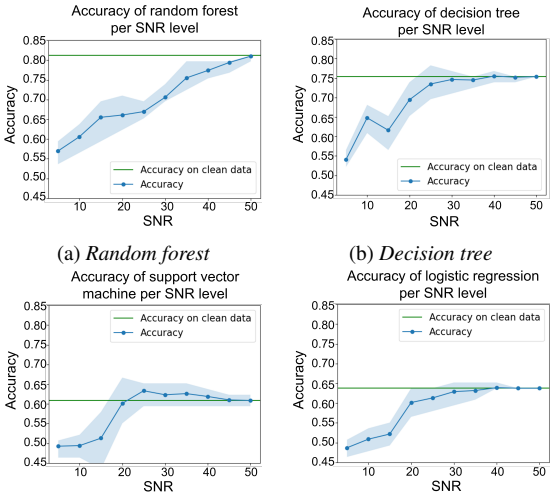
Table 4: *The p-values calculated in the Barnard exact test and the t-test when investigating the misclassification of patients based on age and gender.*

Model	Age p-value	Gender p-value
Decision tree	0.286	0.859
Random forest	0.946	0.468
Support vector machine	0.581	0.735
Logistic regression	0.809	0.836

4.2. Effect of noise on classification

To investigate the effects of noisy signals on classification we added Gaussian noise to each recording in the test set at 10 different signal to noise ratio (SNR) levels between 5 and 50. We then evaluated each of the classifiers on these noisy signals. Since there is an element of randomness when generating Gaussian noise we repeated this 10 times and recorded the accuracy for each attempt. Figure 2 shows how the accuracy varied for each model with different levels of noise. The shaded area shows the minimum and maximum accuracy found across the 10 runs. As is expected the general trend for each model is that the lower the SNR (more noisy signal) the lower the accuracy. It can be seen however, that in some instances the presence of noise increases the accuracy of the classifier most notably in the SVM. This may be because the addition of noise enhances some frequencies in the signal which improves the classification performance of some algorithms. It may also be caused by overfitting in the models. It can be seen that even when the mean accuracy is above the clean accuracy the minimum accuracy is below meaning that the noise does not consistently improve accuracy.

Since real world noisy recordings are not affected by a constant noise level, for the rest of our experiments noise was added to each recording at an SNR of either 5, 10, or 20. This means that the noisy test set contains a mix of different noise levels. Table 3 shows that the precision of each of the classifiers is reduced when noise is added meaning that it over classifies patients as having cancer. The SVM classifier was least affected



(a) *Random forest* (b) *Decision tree*

(c) *Support vector machine* (d) *Logistic regression*

Figure 2: *The classification accuracy of all four classifiers on speech with different levels of Gaussian noise added.*

by the added noise with a 11.9% percentage reduction in accuracy when noise is added. The LR classifier was the most affected by noise with a 25.0% reduction in accuracy when noise was added.

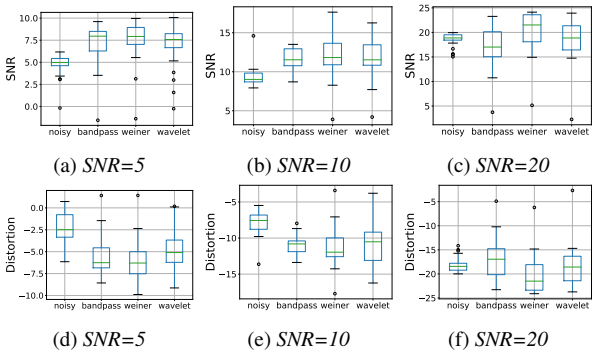
4.3. Denoising

We implement three denoising algorithms: bandpass filter, Wiener filter, and wavelet filter. In order to find the best parameters for the filters a grid search was used. The SNR and distortion of the signals calculated for each denoised signal when compared to the clean signal [29]. SNR is the measure of the strength of a signal relative to background noise [30]. Distortion is a measure of how much the signal is changed from the original waveform [29]. The SNR calculated should be maximised while the distortion should be minimized. The bandpass filter was found to be optimal with a lower cutoff of 10Hz, upper cutoff of 3500Hz, and an order of 1. The Wiener filter was optimized with a Wiener filter window size of 3 and noise-power of 0.1. The wavelet filter was best with a biorthogonal 3.9 wavelet, level of 1, and hard threshold function.

Figure 3 shows the SNR and distortion calculated for the best performing filters as well as the noisy signals before denoising. Since the noise was added to the test set at three different SNR levels (5, 10, and 20) the results of the denoising can be seen for each level. Figures 3a 3b 3d, and 3e show that all three filtering methods improve the recording quality when the SNR of the noisy signal is 5 or 10. The Wiener filter has the best performance for both measures at both levels. However, Figures 3c and 3f show that the bandpass filter reduces the quality of the recordings with an SNR of 20, while the Wiener and wavelet filters improve the quality of some of the recordings.

4.4. Effect of Denoising on Classification

The Wiener filter best improved the performance of the DT and SVM classifiers whereas, the wavelet filter best improved the performance of the RF and LR classifiers. This is in contrast to the results of the denoising where the Wiener filter best increased the SNR and reduced the distortion. The wavelet filter can be seen not to improve the accuracy of the DT and to reduce



(a) *SNR=5* (b) *SNR=10* (c) *SNR=20*

(d) *SNR=5* (e) *SNR=10* (f) *SNR=20*

Figure 3: *The performance of each of the three denoising algorithms at each of the SNR levels. (a-c) - SNR measured for each recording, (d-f) - distortion measured for each recording.*

the accuracy of the SVM.

These results show that a single type of denoising algorithm cannot be used across all classifier types. It also shows that the use of signal quality metrics cannot be assumed to translate to classifier performance.

5. Conclusions and Further Work

In this work we have developed a pipeline for the evaluation of laryngeal cancer classifiers in the presence of noise, using machine learning methods. We show how to use this pipeline to investigate the effects of Gaussian noise on laryngeal cancer detection. We also investigate the effectiveness of simple denoising solutions in restoring classifier performances.

Our experiments showed that Gaussian noise can render classifiers unusable. Added noise caused a decrease in precision of the tested classifiers, with an increase in healthy patients being classified as having cancer. We show that using simple denoising algorithms reduces misclassification rates, however the performance of the classification models can not be fully restored using these denoising algorithms. We found that different types of classifiers are best paired with different denoising algorithms, meaning that there is no one-size fits all solution.

We show that our proposed pipeline allows for the evaluation of speech classifier performance in the presence of noise. Since this is early work, we have only investigated the effects of Gaussian noise. The effects of other more complex types of noise, such as reverberation or other real world noise, can be easily investigated using our pipeline in future work. In future work we would also like also to classify patients with benign voice pathologies rather than healthy controls in order to better mirror real world clinical settings.

Our results indicate that it is essential to consider the impact of noisy recordings when implementing ML classifiers for detecting laryngeal cancer from speech. Our experiments suggest that simple denoising methods alone can mitigate but cannot fully reverse the effects of noisy recording environments on classification. In order for laryngeal cancer classifiers to be implemented in clinical practice these effects must be considered, and mitigating steps must be implemented.

6. Acknowledgements

This research was funded in part by the UKRI Engineering and Physical Sciences Research Council (EPSRC) [EP/S024336/1].

7. References

[1] R. Nocini, G. Molteni, C. Mattiuzzi, and G. Lippi, "Updates on larynx cancer epidemiology," *Chinese Journal of Cancer Research*, vol. 32, no. 1, pp. 18–25, Feb. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7072014/

[2] Cancer Research UK, "Survival | Laryngeal Cancer | Cancer Research UK," 2019. [Online]. Available: https://www.cancerresearchuk.org/about-cancer/laryngeal-cancer/survival

[3] NHS, "Laryngeal (larynx) cancer - Diagnosis," Oct. 2017, section: conditions. [Online]. Available: https://www.nhs.uk/conditions/laryngeal-cancer/diagnosis/

[4] Hyun Bum Kim, H.-B. Kim, J. Jeon, Y. J. Han, Young-Hoon Joo, Y. H. Joo, J. Lee, S. Lee, and S. Im, "Convolutional Neural Network Classifies Pathological Voice Change in Laryngeal Cancer with High Accuracy," *Journal of Clinical Medicine*, vol. 9, no. 11, p. 3415, Oct. 2020, mAG ID: 3094604100.

[5] C.-T. Wang, Z.-Y. Chuang, C.-H. Hung, Y. Tsao, and S.-H. Fang, "Detection of Glottic Neoplasm Based on Voice Signals Using Deep Neural Networks," *IEEE Sensors Letters*, vol. 6, no. 3, pp. 1–4, Mar. 2022, conference Name: IEEE Sensors Letters.

[6] K. Ezzine, A. Ben Hamida, Z. Ben Messaoud, and M. Frikha, "Towards a computer tool for automatic detection of laryngeal cancer," in *2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, Mar. 2016, pp. 387–392.

[7] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach," *Journal of Voice*, vol. 33, no. 5, pp. 634–641, Sep. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089219971730509X

[8] A. Ben Aicha and K. Ezzine, "Cancer larynx detection using glottal flow parameters and statistical tools," in *2016 International Symposium on Signal, Image, Video and Communications (ISIVC)*, Nov. 2016, pp. 65–70.

[9] J. C. Das, "Passive Filters," in *Power System Harmonics and Passive Filter Designs*, 2015, pp. 685–743.

[10] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[11] E. Hänsler and G. Schmidt, "Wiener Filter," in *Acoustic Echo and Noise Control: A Practical Approach*, 2004, pp. 53–60.

[12] D. B. Percival and A. T. Walden, "Wavelet-Based Signal Estimation," in *Wavelet Methods for Time Series Analysis*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000, pp. 393–456.

[13] G. R. Lee, R. Gommers, F. Waselewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *Journal of Open Source Software*, vol. 4, no. 36, p. 1237, Apr. 2019. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.01237

[14] L. Gavidia-Ceballos and J. H. Hansen, "Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection," *IEEE transactions on bio-medical engineering*, vol. 43, no. 4, pp. 373–383, Apr. 1996.

[15] T. Grzywalski, A. Maciaszek, A. Biniakowski, J. Orwat, S. Drgas, M. Piecuch, R. Belluzzo, K. Joachimiak, D. Niemiec, J. Ptaszynski, and K. Szarzynski, "Parameterization of Sequence of MFCCs for DNN-based voice disorder detection," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 5247–5251.

[16] M. A. Ali and P. M. Shemi, "An improved method of audio denoising based on wavelet transform," in *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Dec. 2015, pp. 1–6.

[17] H. Hammam, A. A. Elazm, M. E. Elhalawany, and F. E. Abd El-Samie, "Blind separation of audio signals using trigonometric transforms and wavelet denoising," *International Journal of Speech Technology*, vol. 13, no. 1, pp. 1–12, Mar. 2010. [Online]. Available: https://doi.org/10.1007/s10772-010-9066-0

[18] M. Saric, L. Bilicic, and H. Dujmic, "White Noise Reduction of Audio Signal using Wavelets Transform with Modified Universal Threshold," Feb. 2005.

[19] M. Ju, Z. Jiang, Y. Chen, and S. Ray, "A Multi-Representation Ensemble Approach to Classifying Vocal Diseases," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 5258–5262.

[20] M. Pham, J. Lin, and Y. Zhang, "Diagnosing Voice Disorder with Machine Learning," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 5263–5266.

[21] J. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 380–384, Feb. 2004, conference Name: IEEE Transactions on Biomedical Engineering.

[22] I. Miliaresi, K. Poutos, and A. Pikrakis, "Combining acoustic features and medical data in deep learning networks for voice pathology classification," in *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, pp. 1190–1194, iSSN: 2076-1465.

[23] J. D. Arias-Londoño, J. Andrés Gómez-García, L. Moro-Velázquez, and J. I. Godino-Llorente, "ByoVoz Automatic Voice Condition Analysis System for the 2018 FEMH Challenge," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 5228–5232.

[24] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, Nov. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095447017301389

[25] K. Degila, R. Errattahi, and A. E. Hannani, "The UCD System for the 2018 FEMH Voice Data Challenge," in *2018 IEEE International Conference on Big Data (Big Data)*, Dec. 2018, pp. 5242–5246.

[26] M. Pishgar, F. Karim, S. Majumdar, and H. Darabi, "Pathological Voice Classification Using Mel-Cepstrum Vectors and Support Vector Machine," Dec. 2018, arXiv:1812.07729 [cs, eess, stat]. [Online]. Available: http://arxiv.org/abs/1812.07729

[27] Manfred Pützer and William, J Barry, "Saarbruecken Voice Database," May 2007, publisher: Institut für Phonetik, Universität des Saarlandes. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4

[28] German Centre for Cancer Registry Data and Robert Koch Institute, "Database Query with estimates for cancer incidence, prevalence and survival in Germany, based on data of the population based cancer registries. Mortality data provided by the Federal Statistical Office." Sep. 2022. [Online]. Available: www.krebsdaten.de/database

[29] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of Blind Signal Separation Methods," Mar. 2000, pp. 261–266.

[30] A. Dave, "Signal-to-Noise Ratio," in *Practical Communication Theory*, 2nd ed. Institution of Engineering and Technology, 2014, pp. 1–1.