State-of-the-art review on automatic techniques employing speech to discover disorder/ illness

S.B. Kindler von Knobloch Luengo[1]

[1] Biomedical Engineering Degree at University CEU San Pablo, Madrid, Spain, sb.kindler@usp.ceu.es

Abstract

In today's ever-developing technological world, it has become more and more evident how the use of machine learning and artificial intelligence can be employed as means to improve the overall health of our society. Lately, it has become apparent in the realm of clinical diagnosis of a wide range of diseases. This review serves as an example of the latter, as it will showcase how from a speech signal, analysed through machine learning tools, the detection of Parkinson's Disease (PD) , Coronavirus Disease (COVID-19) and Laryngeal pathologies (LP) is effectively achieved. The present study will therefore explore the steps that are taken from the signal acquisition, through its ***pre-processing*** *stage (to facilitate feature extraction), the following* ***feature extraction*** *(where the features that will be analysed are collected) and finalising with automatic* ***classification****. To achieve this task, a wide range of scientific articles have been studied. Although the three pathologies exposed above come from different biological standpoints, one being neurological, the second respiratory and the third pertaining specifically to the vocal system, after extensive research it has become clear that the pre-processing, feature extraction and classification of the three diseases is extraordinarily similar. Within the articles studied, the features most commonly extracted for analysis are the Mel Frequency Cepstral Coefficients (MFCC) and the machine learning algorithm of choice is the Support Vector Machine (SVM) achieving over 90% accuracy rate of detection of the three diseases.*

1. Introduction

Parkinson's Disease (PD) is the second most common neurodegenerative disorder and is characterised by the loss of dopaminergic neurons in the midbrain, which alters the nigral complex to different extents. It's clinical detection is achieved through the unified Parkinson's disease rating scale (UPDRS). Symptoms depend on the degree of development of the disease, but in general terms include tremor, stiffness, postural instability and slowness.

Although it is a neurological disorder, it affects different speech production dimensions such as breathing, phonation, articulation and prosody which are represented through reduced vocal tract volume, tongue flexibility, significantly narrower pitch range, voice intensity level and articulation rate. These symptoms have an effect on the resulting speech signal, manifesting themselves in the form of features used for automatic analysis.

In terms of the Coronavirus disease (COVID-19) , which is characterised by dysfunctions in respiratory physiology including the diaphragm and other parts of the lower respiratory tract, there is also an effect on speech production, which can be analysed to distinguish between a healthy subject and an ill patient. The impairment of the respiratory apparatus leads to significant effects on voice production, whereby vocal fold oscillations, which refer to the vibration of the vocal folds as air passes through, become more asynchronous, asymmetrical and restricted.

Since its outbreak, COVID-19 has been declared as a global pandemic by the World Health Organisation (WHO) and has rapidly spread over the globe, accounting for more than 6 million deaths since the beginning of 2019. For this reason, it is of great interest to achieve its automatic detection through machine learning algorithms.

In the case of laryngeal disorders (LP), the implementation of machine learning algorithms for automatic detection has been in question for a longer period of time, due to the direct relationship between symptoms and its representation in speech signals. The symptoms in the voice of patients with laryngeal pathologies (LP) mainly relate to abnormal vibration of vocal folds, hoarseness and breathy voice. The early detection of laryngeal pathologies significantly increases the effectiveness of treatment. The use of machine learning for this process can have a positive impact creating a rapid detection process.

2. Methodology

The article selection process has been achieved through two search engines of academic content: Google Scholar and Science Direct. In order to maintain the specificity of the content without missing the originality and variety in scientific perspective, the search process is filtered into two steps. The first is done directly in the search engines themselves, and consists of reducing the scope of the search using:

1. Keywords in the search bar: Using the words and phrases "speech signal", "Parkinson's", "COVID-19", "Laryngeal pathologies", combined with "machine learning" in a concise manner.
2. Restricting the article date: Overall searching for articles written after 2010, mainly choosing those that are more recent. In the case of the articles reviewing automatic detection of COVID-19, there is not a great variety in terms of date due to how recent it has been.

Once the articles have been collected, the number of citations in the article and the origin of the database used is studied. Those articles with the least number of sources are eliminated, and the ones with repeated databases are less likely to be chosen for reviewing as the greater the variety in data used, the more a reliable and widespread the classification can be.

3. Speech-based detection of Parkinson's disease, COVID-19 and laryngeal disorders

There are a wide range of approaches to automatically detect Parkinson's disease, COVID-19 and laryngeal disorders using speech signals. Nevertheless, they all follow the same general pattern mentioned above and exposed in figure 1.

Figure 1: Block diagram of general methodology used

Each of these steps is explained in detail hereafter.

3.1. Database

What is displayed in the block diagram as "signal acquisition" is the process of obtaining a speech signal, from which speech features will be extracted to provide a diagnosis. The set of signals from all the subjects becomes the database of the study. The greater the variance in terms of sex, health state (control or ill subject), and origin of the patients, the greater the accuracy and similarity to the actual epidemiological condition of the diseases. Likewise, bigger databases will provide more precise classifications as there is more training data for the machine learning algorithms. All of these factors were studied in the three diseases investigated.

- Parkinson's Disease

The reviewed papers on Parkinson's automatic detection showed a great variety of the database in terms of patient origin, ranging from Spanish-speaking subjects [1] [2] to India [3], or even multiple datasets (of Spanish, German and Czech origin) [4].

- Coronavirus Disease

Similarly, in terms of COVID-19 articles, there is multiple datasets per paper [21] and there is also sex-distinction between positive and control COVID-19 patient-databases [5].

- Laryngeal Disorders

Most of the papers on automatic detection of laryngeal disorders however, have a common database that is used for different feature extraction and classification methodologies, pertaining to the Massachussets Eye Infirmary (MEEI) Voice and Speech Lab [6] [7] . The commonality between databases comes as a disadvantage, as it reduces the scope of the investigation.

3.2. Preprocessing

Once the speech signal has been recorded, the features that will later be automatically analysed through machine learning tools must be chosen. Nevertheless, the "quality" or viability of feature extraction is not always optimal due to the presence of artifacts that contaminate the signal. These artifacts are sometimes referred to as noise, that is eliminated to improve the quality of feature extraction through multiple manners, two of them are exposed below:

I. Filtering of the speech signal: Signal denoising is achieved through Least Mean Square (LMS) filtering whereby the filter coefficients to produce the least mean square error of the signal are found. Moreover, the technique of spectral subtraction. Here, the noise spectrum is estimated during speech pauses and subtracted from the noise speech spectrum estimating the clean speech.

II. Decomposition of the speech signal: The Variational Mode Decomposition (VMD) tool decomposes the signal into different modes, the noise in the signal will correspond to a high frequency mode which can be removed, decreasing the contamination of the signal.

Furthermore, feature extraction can be simplified if the speech signal is properly segmented, and downsampled to a specific frequency. In the case of Mel-frequency cepstral coefficients (MFCC) feature extraction, pre-emphasis is done boosting the energy in higher frequencies through a first order high pass filter and hamming windows are used to section the signals downsampled to 8kHz or 16kHz.

- Parkinson's Disease

Downsampling [8] and the hamming window function are most commonly applied [3] [1]. VMD is also employed [9], as well as the overlap of frames [10].

- Coronavirus Disease

Pre-processing for COVID-19 detection also applies downsampling [11] [12] [13], as well as Least Mean Square Filtering (LMS) [14] .

- Laryngeal Disorders

Papers that cover pre-processing of speech signals for automatic laryngeal disorder diagnosis apply pre-emphasis and Hamming window [15] .

3.3 Feature Extraction

Feature extraction is a key step in the process of automatic diagnosis, as the correct classification of the signal depends entirely on it. For this reason, it is fundamental to understand how speech is produced given that the features extracted are best classified when they mimic actual biological parameters and functionalities.

Speech production is achieved when the vocal folds vibrate with the air coming from the lungs creating sound, also known as phonation. After passing the larynx and pharynx, the air reaches the nasal or the oral cavities, the latter being where, through the help of articulators (lips and teeth) speech sounds are distinguished from one

another (articulation). The variation in the pitch, loudness and time to articulate is known as prosody. The biological parameters of phonation, articulation and prosody are reflected in the speech signal and can be extracted as features. Phonation factors are expressed through the first and second derivatives of jitter, pitch, fundamental frequency, the latter also representing prosody, which is also parametrised through energy and duration measurements of the signal. Articulation features are extracted in the form of Bark band energies (BBE) and Mel-frequency cepstral coefficients (MFCCs).

Mel-frequency cepstral coefficients (MFCCs) is the most common method of feature extraction within the pathology detection patterns of the three disorders. It is based on the idea of mimicking the human auditory system to assess pathological speech. To obtain this, the speech signal must be windowed into overlapping frames, its discrete Fourier Transform calculated producing the mel-scale filtering which is then taken the logarithm of, finally its discrete Cosine Transform is calculated producing the mel-frequency cepstral coefficients.

- Parkinson's Disease

MFCC's feature extraction is most common [3] [1] [2] [8] [20] although Linear prediction coefficients (LPC) is also present [3].

- Coronavirus Disease

Through MFCC feature extraction, the features analysed most for COVID-19's automatic detection include Harmonics to Noise Ratio (HNR) [5] [4] and the derivatives of the mel-frequency cepstral coefficients from which further statistical features are extracted [18].

- Laryngeal Disorders

Laryngeal disorder classification also makes use of complexity measures such as correlation dimension (CD) and fractal dimension (FD) [16] as well as spectral and temporal characteristics including Mean Square Residue (MSR) or Excess Coefficient [19].

3.3. Classification and results

Overall in the set of reviewed papers, more than a tenfold of machine learning algorithms have been utilised. In many cases, more than one is implemented to compare their effectiveness in pathology detection. Within those that achieve the highest classification rate are the Support Vector Machine (SVM), the k-Nearest Neighbour (kNN) , the Linear Discriminant Analysis (LDA) and the Gaussian Mixture Model (GMM).

On one hand, The Support Vector Machine (SVM) algorithm generates the automatic classification result solving an optimisation problem with the goal of maximising the margin between two or more data groups. This is achieved from a supervised perspective, meaning it requires a training set where the data is already classified into healthy and ill patients. On the other hand, the k-Nearest Neighbour classifier is a semi-supervised algorithm based on calculating the distances between neighbouring data values, and proceed to their classification based on their proximity. In a similar manner, GMM groups data points that belong to a single Gaussian Distribution together. LDA works in a similar manner to the GMM except it is recognised as a supervised algorithm.

- Parkinson's Disease

Classifiers for Parkinson's detection range from SVM [8] [4] [20] to kNN [10], sometimes both appear in the same paper for comparison purposes [8], although they do not obtain the maximum detection rate (82.89% for kNN and 82.23% for SVM) [20].GMM classification is also prominent, achieving the highest classification accuracy of 91% in Parkinson's detection through automatic speech signal classification.

- Coronavirus Disease

Coronavirus classification rates reach maximum values through SVM classification ranging from 85% to 99% depending on the language and speech task [4]. Furthermore, there is a range of tree-machine learning algorithms: Random Forest [13] [21] and Boosting Trees [18] with accuracy values of 88.5%.

- Laryngeal Disorders

Speech signals for laryngeal pathology diagnosis are mostly classified with kNN [22] [23] and SVM [6] [16] [17] with accuracy above 90%.

4. Conclusions

In conclusion, voice in the form of a signal has proved to be a potent digital biomarker for early detection and monitoring of various diseases.

Although there is no evidence that these automatic techniques are currently being applied for clinical detection purposes, this review serves to showcase the broad range of possibilities this technique can offer. Not only in terms of economic resources, but also in terms of the overall health of our society, specially with COVID-19. In a similar manner, the early detection of laryngeal disorders and Parkinson's Disease increases effectiveness of treatment.

Interestingly enough, given the spectrum of pre-processing, feature extraction and classification techniques, many not mentioned in this article itself, most scientific papers for the detection of the three pathologies use the same tools. Overall, the Mel Cepstral Frequency coefficients are the features extracted the most, and the SVM algorithm the machine learning tool used most often for classification, being the most accurate for both COVID-19 and laryngeal pathology classification.

Lastly, it seems significant to mention that the automatic classification through speech signal has only been reviewed for three diseases in this report. Nevertheless, there are multiple other disorders such as Alzheimer's Disease (AD), Heart Failure (HF), Sleep Apnea (OSA) that have been studied with the same purpose. This exemplifies the great potential that speech signal has in pathology detection.

Acknowledgments

I would like to thank my tutor Javier Tejedor Noguerales for his support and counselling creating the review.

References

[1] Vaiciukynas, Evaldas, et al. “Detecting Parkinson’s Disease from Sustained Phonation and Speech Signals.” PLOS ONE, vol. 12, no. 10, 2017, https://doi.org/10.1371/journal.pone.0185613.

[2] ER, Mehmet Bilal, et al. “Parkinson's Detection Based on Combined CNN and LSTM Using Enhanced Speech Signals with Variational Mode Decomposition.” 2021, https://doi.org/10.21203/rs.3.rs-305818/v1.

[3]Vikas, and R. K. Sharma. “Early Detection of Parkinson's Disease through Voice.” 2014 International Conference on Advances in Engineering and Technology (ICAET), 2014, https://doi.org/10.1109/icaet.2014.7105237.

[4] Orozco-Arroyave, J. R., et al. “Automatic Detection of Parkinson's Disease in Running Speech Spoken in Three Different Languages.” The Journal of the Acoustical Society of America, vol. 139, no. 1, 2016, pp. 481–500., https://doi.org/10.1121/1.4939739.

[5]Verde, Laura, et al. “Exploring the Use of Artificial Intelligence Techniques to Detect the Presence of Coronavirus Covid-19 through Speech and Voice Analysis.” IEEE Access, vol. 9, 2021, pp. 65750–65757., https://doi.org/10.1109/access.2021.3075571.

[6]Orozco-Arroyave, Juan Rafael, et al. “Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases.” IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 6, 2015, pp. 1820–1828., https://doi.org/10.1109/jbhi.2015.2467375.

[7]Akbari, Ali, and Meisam Khalil Arjmandi. “An Efficient Voice Pathology Classification Scheme Based on Applying Multi-Layer Linear Discriminant Analysis to Wavelet Packet-Based Features.” Biomedical Signal Processing and Control, vol. 10, 2014, pp. 209–223., https://doi.org/10.1016/j.bspc.2013.11.002.

[8]Sarria Paja, Milton Orlando. “Automatic Detection of Parkinson's Disease from Components of Modulators in Speech Signals.” Computer and Electronic Sciences: Theory and Applications, vol. 1, no. 1, 2020, pp. 71–82., https://doi.org/10.17981/cesta.01.01.2020.05.

[9]ER, Mehmet Bilal, et al. “Parkinson's Detection Based on Combined CNN and LSTM Using Enhanced Speech Signals with Variational Mode Decomposition.” 2021, https://doi.org/10.21203/rs.3.rs-305818/v1.

[10]Belalcazar-Bolanos, E. A., et al. “Automatic Detection of Parkinson's Disease Using Noise Measures of Speech.” Symposium of Signals, Images and Artificial Vision - 2013: STSIVA - 2013, 2013, https://doi.org/10.1109/stsiva.2013.6644928.

[11]Xia, Tong, et al. “Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data.” Interspeech, 2021, https://doi.org/10.21437/interspeech.2021-1320.

[12]Dash, Tusar Kanti, et al. “Detection of Covid-19 from Speech Signal Using Bio-Inspired Based Cepstral Features.” Pattern Recognition, vol. 117, 2021, p. 107999., https://doi.org/10.1016/j.patcog.2021.107999.

[13]Al Ismail, Mahmoud, et al. “Detection of Covid-19 through the Analysis of Vocal Fold Oscillations.” ICASSP - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, https://doi.org/10.1109/icassp39728.2021.9414201.

[14]Al-Dhlan, Kawther A. “An Adaptive Speech Signal Processing for COVID-19 Detection Using Deep Learning Approach.” International Journal of Speech Technology, 2021, https://doi.org/10.1007/s10772-021-09878-0.

[15]Optimizing Laryngeal Pathology Detection by Using Combined Cepstral ... https://www.researchgate.net/publication/234059582_Optimizing_laryngeal_pathology_detection_by_using_combined_cepstral_features.

[16] Vaziri, Ghazaleh, et al. “Pathological Assessment of Patients’ Speech Signals Using Nonlinear Dynamical Analysis.” Computers in Biology and Medicine, vol. 40, no. 1, 2010, pp. 54–63., https://doi.org/10.1016/j.compbiomed.2009.10.011.

[17]Fang, Shih-Hau, et al. “Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach.” Journal of Voice, vol. 33, no. 5, 2019, pp. 634–641., https://doi.org/10.1016/j.jvoice.2018.02.003.

[18]Brown, Chloë, et al. “Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data.” Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, https://doi.org/10.1145/3394486.3412865.

[19]De Oliveira Rosa, M., et al. “Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis.” IEEE Transactions on Biomedical Engineering, vol. 47, no. 1, 2000, pp. 96–104., https://doi.org/10.1109/10.817624.

[20]Nissar, Iqra, et al. “Voice-Based Detection of Parkinson’s Disease through Ensemble Machine Learning Approach: A Performance Study.” EAI Endorsed Transactions on Pervasive Health and Technology, vol. 5, no. 19, 2019, p. 162806., https://doi.org/10.4108/eai.13-7-2018.162806.

[21]Despotovic, Vladimir, et al. “Detection of Covid-19 from Voice, Cough and Breathing Patterns: Dataset and Preliminary Results.” Computers in Biology and Medicine, vol. 138, 2021, p. 104944., https://doi.org/10.1016/j.compbiomed.2021.104944.

[22]Shama, Kumara, et al. “Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology.” EURASIP Journal on Advances in Signal Processing, vol. 2007, no. 1, 2006, https://doi.org/10.1155/2007/85286.

[23]Hadjitodorov, S., et al. “Laryngeal Pathology Detection by Means of Class-Specific Neural Maps.” IEEE Transactions on Information Technology in Biomedicine, vol. 4, no. 1, 2000, pp. 68–73., https://doi.org/10.1109/4233.826861.