Voice Pathologies

The Most Comum Features and Classification Tools

Joana Filipa Teixeira Fernandes
Universidade do Porto
Faculdade de Engenharia
Porto, Portugal
up201911738@fe.up.pt

Diamantino Freitas
Universidade do Porto
Faculdade de Engenharia
Porto, Portugal
dfreitas@fe.up.pt

João Paulo Teixeira
Instituto Politécnico de Bragança
Escola Superior de Tecnologia e Gestão
Bragança, Portugal
joaopt@ipb.pt

Abstract — **Speech pathologies are quite common in society, however the exams that exist are invasive, making them uncomfortable for patients and depending on the experience of the clinician who performs the assessment. Hence the need to develop non-invasive methods, which allow objective and efficient analysis. Taking this need into account in this work, the most promising list of features and classifiers was identified. As features, jitter, shimmer, HNR, LPC, PLP, and MFCC were identified and as classifiers CNN, RNN and LSTM. This study intends to develop a device to support medical decision, however this article already presents the system interface.**

Keywords - Voice Pathologies; Acoustic Parameters; Speech Signal; Neural Networks.

I. Introduction

Voice pathologies are relatively common and can be found in different stages of evolution and severity, affecting approximately 10% of the population [1]. These pathologies directly affect vocal quality, as they alter the phonation process and have increased dramatically in recent times mainly due to harmful habits, such as smoking, excessive consumption of alcoholic beverages, persistent inhalation of dust-contaminated air; and voice abuse [2].

There are a variety of tests that can be performed to detect pathologies associated with the voice, however, they are invasive and become a little uncomfortable for patients, can cause vomiting and are time-intensive. Auditory acoustic analyzes carried out by professionals lack objectivity and depend on the experience of the physician who performs the assessment.

Using an evaluation method that allows easy and quick recognition of voice pathologies, efficient and non-invasive, can be useful in a first evaluation for screening, as a complementary method in the diagnosis of voice pathologies and allows to provide a pre-diagnosis system to regions with limited access to clinical professionals.

Although there are significant studies regarding the classification of voices with pathologies using speech-processing tools, the automatic recognition of these pathologies is still at an early stage. The vast majority of published works refer to binary classifiers, where the test signal is classified as having or not the pathology.

On the other hand, the recognition of voice pathologies consists in the identification of the specific pathology that affects a patient's voice, and this area presents few contributions using very few different pathologies and with results that still do not allow the use of these systems in clinical analysis, also denoting a need for further investigation with varied pathologies.

An example of this is the work of Dibazar, et al., 2002 [3] that feature a robust, fast and accurate system for the automatic detection of normal and pathological speech. They obtained a 99,4% accuracy rate to discriminate between normal speech and pathological speech. In Markaki & Stylianou, 2009 [4] for the detection of vocal pathology they reached an accuracy of 94,1% and a score in the area under the curve (AUC) of 97,8%, already in the classification between four vocal pathologies (vocal cord polyps, spasmodic dysphonia, nodules and vocal cord leukoplakia), obtained an average detection rate and AUC of 88,6% and 94,8% respectively. Carvalho, et al., 2011 [5] used healthy subjects, patients with nodules, Reinke's edema and neurological dysphonia and obtained an average accuracy of 89,14% for all classes, with the best results per group being achieved by the control group (93,33%) , followed by the group with dysphonia (89,4%), Reinke's edema (88,15%) and nodules (86,56%).

As it can be seen, the results for identifying the specific pathology still performed using a reduced number of pathologies.

Therefore, this article intends to review the literature in order to select a set of more prominent parameters and classifiers, and present a graphical interface of the Diagnostic Aid System under development.

Next section presents a state of the art review, followed by an analysis of the most proeminent features and classifiers. The next section presents the devices or systems publicly known to support medical decition related to voice diagnosis. Finaly last section presents the final conclusions.

II. State of the Art Review

Research works that allow the automatic recognition of voice pathologies are still scarce. However, there have been works for several years that have distinguished and classified healthy or pathological voice.

The use of sustained vowels makes it possible to distinguish between subjects with healthy or pathological speech, because, when asked to keep phonation stable for a long period of time, healthy subjects are able to obtain phonation with a more stable frequency and amplitude than individuals with some vocal disturbance [6].

In the recognition of pathological voices, parameters that characterize the movement of the vocal cords are frequently used, such as, for example, jitter and shimmer [7]. However, the results for the identification of the specific pathology do not yet have accuracy rates that allow normal use by clinical staff or patients. So, it is recognized the need to investigate the use of new parameters, more accurate classification models, and use of more complete databases.

The analyzes of independent parameters do not reached very conclusive results. However, properly associated with Artificial Intelligence tools are considerably better.

Several authors use acoustic analysis together with artificial intelligence tools in their work.

Markaki & Stylianou, 2009 [4] considered the use of modulation spectra for the detection and classification of four vocal pathologies (vocal cord polyp, nodules, spasmodic dysphonia and vocal cord leukoplakia) for vocal pathology. To reduce the high dimensionality space generated by the modulation spectra, they used Higher Order Singular Value Decomposition (SVD) and proposed a resource selection algorithm based on mutual information between subjective voice quality and resources calculated. They used Support Vector Machine (SVM) with a radial base function (RBF) kernel as a classifier and performed experiments with the MEEI database (Massachusetts Eye and Ear Infirmary).

Wu, et al., 2018 [8] proposed a method with the Convolutional Neuronal Network (CNN), to extract features automatically from the spectrogram of voice recordings for the diagnosis of dysphonia. They used small windows to extract resources from the spectrogram, in order to detect subtle variations between the pathological voice and the normal voice. The speech signal recording from the Saarbrucken Voice Database were used. The classifier they used obtained 88.5%, 66.2% and 77.0% accuracy in the training, validation and test data set, respectively. They used 482 healthy subjects and 482 subjects with dysphonia.

Roy, et al., 2019 [9] tried two different approaches for voice classification and worked with variable size data. They used recurrent neural networks, with the variation Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM), and 2 and 3 layer CNN. In their experimental models, they obtained precision of 73,07% for GRU, 73,39% for LSTM, 74,67% for CNN with 2 layers and 75,64% for CNN with 3 layers. As a database they used the Saarbrucken Voice Database and used the Mel-Frequency Cepstrum Coefficients (MFCCs) as input parameters.

Fang, et al., 2019 [10] in their study, used 60 normal voice signals and 402 pathological voice signals from 8 common clinical voice disorders in a higher education hospital (nodules, polyps, cysts, neoplasia, vocal atrophy, laryngeal dystonia (ie, spasmodic dysphonia and tremor), unilateral vocal cord paralysis and vocal sulcus). They extracted the MFCCs from a sustained vowel and used MFCC, delta MFCC parameters and MFCC (N) with delta parameters as input parameters. The difference between MFCC (N) for MFCC is that the first normalized all MFCC coefficients with zero mean and unit variation. For the classification they used Deep Neural Network (DNN), SVM and the Gaussian mix model. The performance of these three classification models was assessed based on a five-fold cross-validation. The experimental results showed that DNN surpasses the Gaussian mix model and SVM. Its accuracy in the detection of voice pathologies reached 94,26% and 90,52% in male and female individuals, respectively. To validate the results, they used the MEEI database and also the DNN obtained greater precision (99,32%) than the other two classification algorithms.

III. ANALYSIS OF FEATURES AND CLASSIFIERS

Acoustic evaluation or acoustic analysis, when used to assess vocal disorders, allows the non-invasive quantification of human voice characteristics. This technique provides objective measures, such data that are automatically extracted, through an appropriate computational process, such as, a direct estimate of the vibratory patterns of the vocal folds, as well as the shapes of the supraglottic vocal tract and the respective modifications [11]. For this, it is necessary to use different acoustic parameters that make up the signal - periodicity, amplitude, duration and spectral composition - that characterize the physical attributes of the voice in the domains of time, frequency and intensity, in addition to other complex measures that combine the crossing of these domains [12].

The number of publications that can be considered revealing in the recognition of voice pathologies is reduced. Research for the recognition of pathologies is relatively recent, since the first advances occurred in 2000. However, there are many publications that refer to the discrimination of voices that suffer from pathologies without recognizing which pathology.

The 60's and 70's was when the development of digital signal processing started, with the first voice analysis programs, with more precise and clear definitions of the measures to be used.

The signal processing techniques allow to collect and characterize the vibration characteristics of the vocal folds. Digital signal processing allows to analyze, transform or interpret signals using computational algorithms such as FFT (Fast Fourier Transform), LPC (Linear Preditive Coding), or filtering techniques and Cepstrum [13]. Thus, the measurements obtained in the acoustic analysis correspond to defined physical parameters. The glottal signal, source signal, suffers effects along the supraglottic vocal tract until it becomes a speech or voice signal, configuring a filter action. There is a summation of sound waves from the glottal source with others reflected along the vocal tract, the output signal being the signal radiated by the lips [12] [14].

The acoustic parameters reflect the complex interaction between the glottal source and the vocal tract resonance cavities. Thus, they depend on the biomechanical and aerodynamic forces of the larynx and supraglottic structures, as well as on the complex cortical neuro-motor control. If these components have abnormal anatomical and / or physiological characteristics, then the results obtained are deviated from the expected and, thus, are assumed as indicators of vocal pathology and its respective severity [15] [16].

The identification of pathological speech / speech pathologies is possible, as the pathological voice has characteristics with some differences in healthy speech, such as periodicity, signal / noise ratio, among others. There are several types of parametric representations for the acoustic signals. In recent years, MFCC have become successful parameters for classifying audio segments [17].

Analyzing large data sets with multiple variables can lead to two problems, either that mistakes are made or is sometimes an unaffordable task for humans. Artificial intelligence systems are an asset and can be used in classification systems [18]. Classification problems arise with the need to assign an object to a class or group based on a certain number of parameters related to the object. After training, an artificial intelligence system must have the ability to generalize, that is, before a situation never before seen making a decision based on similarities of things seen previously [19].

In this work we intend to develop a study of acoustic analysis parameters that, later, combined with classification techniques, which are also the object of study, we intend to improve the results in the identification of voice pathologies, in order to support the realization of reliable diagnosis of these pathologies.

Table 1 shows the summary of some research work on the detection of speech pathologies.

Table 1 - Summary of some research work on voice pathology detection detailing the acoustic parameters used, the classification method and the correctness rate.

Author	Features	Classifier	Accuracy (%)
Grzwalski, et al., 2018 [20]	MFCC; Jitter; Turbulent Noise Index; Normalized First; Harmonic Energy	DNN (Deep Neural Network)	77,44
Verde, et al., 2018 [21]	Jitter, Shimmer, HNR; MFCC; first and second derived from MFCC	SVM	85,77%
Gupta, 2018 [22]	MFCC; Spectral Centroid; Chroma; Spectral Contrast	LSTM	97,1
Fang, et al., 2019 [10]	MFCC	DNN	99,32
Alhussein & Muhammad, 2018 [23]	Octave spectrogram and its first- and second-order derivatives	Transfer-Learning from CNN models	97,5
Wester, 1998 [24]	HNR	Hidden Markov models	65
Markaki & Stylianou, 2009 [4]	Modulation Spectra	SVM	94,1
Wu, et al., 2018 [8]	Spectrogram	CNN	88,5
Roy, et al., 2019 [9]	MFCCs	GRU	73,07
		LSTM	73,39
		CNN – 2 layers	74,67
		CNN – 3 layers	75,64
Fang, et al., 2019 [10]	MFCC; ΔMFCC MFCC(N)	DNN	99,32
		SVM	94,26
		GMM	90,52

As it can be seen in Table 1, the results vary a lot depending on the features and classifiers.

Taking these results into account, it is possible to notice that the Deep Neural Networks (DNN), with their ability to assimilate higher levels of abstract concepts using various levels of non-linear nodes, have transformed deep learning algorithms into the best modeling techniques for automatic voice diagnosis.

The fully connected classic feed-forward DNNs are quite efficient, however, more complex methods have led to greater recognition accuracy, since they profit from different connection architectures. Convolutional Neural Networks (CNN), or Recurrent Neural Networks (RNN) using Long Short-Term Memory (LSTM) cells are more complex methods that lead to greater recognition accuracy [25].

Deep-Learning models based on convolutional networks – CNN, are widely used for classification problems using images [26] [27]. It is possible to transform a speech signal over time into a time-frequency representation, as if it were an image, and explore the good performance of these networks for image classification.

Transfer-Learning techniques allow machine learning that, based on networks trained in a given problem, the learning of this one can be reused to carry out the execution of similar problems [28]. This technique allows to take advantage of existing networks for the classification of different types of sounds, and later be adapted for enhanced training by classifying pathological voice signals.

The Recurring Networks (LSTM) configuration is widely used to process sequential and temporal information such as texts, audios and videos [29]. These can be used with continuous speech seeking to identify limitations imposed by some pathologies in the variation of time.

Analyzing the table, it is possible to notice that there are parameters such as MFCC that are repeated in the various studies.

Other authors such as PANEK, et al., 2015 [30] used 28 acoustic parameters. These are the frequency grounds; jitter; shimmer; energy; zero order moments, first, second and third; kurtosis; power factor; amplitude of the first, second and third formants; frequency of the first, second and third formants; maximum and minimum signal values and 10 MFCC coefficients.

Dave, 2013 [31] in his work he uses LPC, Perceptual Linear Prediction (PLP), MFCC as acoustic parameters. In this work it is possible to notice that the PLP and MFCC parameters obtain better response compared to the LPC parameters, since they are similar to the concept of the human auditory system.

Ai, et al., 2012 and Wong & Sridharan, 2001 [32] [33] in their work used LPCC and MFCCs. Although several authors refer to MFCCs as one of the best acoustic parameters, these two studies show that LPCCs slightly outperform MFCCs.

Through the reading of several works it was possible to notice that the parameters, jitter, shimmer, HNR, LPC, PLP and MFCC are the most used parameters in the detection of speech pathologies. As a rule, the LPC, PLP and deltas MFCC end up getting better results. Taking into account the works described above and works described in the state of the art, it is possible to notice that these parameters are widely used in the detection of speech pathologies.

IV. COMMERCIAL SYSTEMS

Some authors have combined the techniques mentioned above, and developed applications that allow classifying the voice to assist in pre-diagnosis, accompanying pharmacological and even post-surgical treatments.

Dibazar, et al., 2002 [3] created a robust, fast and accurate system for the automatic detection of normal and pathological speech. The MFCCs and the measures of pitch dynamics were modeled by Gaussian mixtures in a Markov hidden model classifier (HMM). For the evaluation of this method, the vowel / a / of more than 700 normal individuals and of different pathologies from the MEEI database was used. They obtained a 99,44% correct rate of correct classification rates to discriminate normal speech from pathological speech.

Alhussein & Muhammad, 2018 [23] developed a voice pathology detection system using Deep Learning in a mobile healthcare structure. In this system the voices are captured using smart mobile devices. Voice signals are processed before being sent to a CNN. They used Transfer Learning to use existing robust CNN models. They used the Saarbrucken Voice Database. The experimental results showed that the accuracy of the detection of vocal pathology reaches 97,5% using Transfer Learning from CNN models.

Muhammad, et al., 2017 [34] integrated IoT and cloud computing for a voice pathology detection framework. In a binary detection, they obtained 98.1% accuracy.

Although there are several speech pathology detection systems, these systems have not been converted into assistive devices for clinical personnel. This led to the development of a device to support medical decision making.

A system to aid in the diagnosis of speech pathologies was started and then the interfaces under development are presented.

In Figure 1 it is possible to observe the front-end of the system been developed.

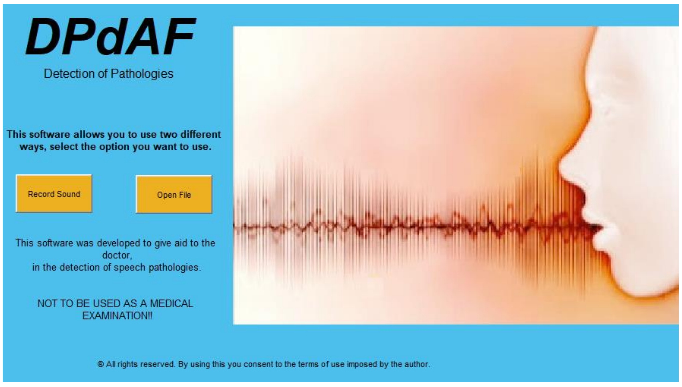

Figure 1- Medical decision support device

This software has two options, we can record the speech directly at the moment or we can already have the sounds and just load them. If we select “Record Sound” we will have a window similar to Figure 2. In this window we can hear an example of the sounds to be recorded, when recording the sounds the signal is being represented at the same time. Finally, we can hear the signals, it is possible to record again, see the signal and obtain acoustic parameters for each signal. Finally, we can record these sounds and make a pre-diagnosis. This is obtained through the acoustic parameters obtained and performed using DNN [35, 36].

Figure 2 - Record Sound Window

Chosing "Open File", Figure 3, it is possible to open previously recorded sound files. This window allows the same features as the "Record Sound" window when loading sound files.

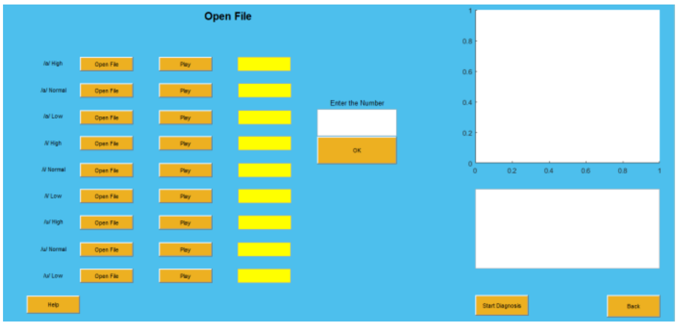

Figure 3 - Open File Window

V. Conclusions

Speech pathologies are quite common and the tests that allow them to be detected are invasive and uncomfortable for patients. Therefore, a non-invasive exam can make a very useful contribution to the correct and early diagnosis of these pathologies.

This work outlined and explored some parameters, tools and techniques that make it possible to diagnose vocal pathologies.

In this work, acoustic parameters were studied that allow the characterization of the voice signal to be later combined with Artificial Intelligence and thus obtain a reliable diagnosis of voice pathologies. The analyzed parameters are the parameters that have the highest incidence in works performed.

The MFCC parameters are the parameters most used in speech pathology detection work, hence they are also considered here. However, once these parameters are extracted, it is possible to obtain their deltas, which are also very promising in the detection of speech pathologies. In addition to the MFCC, the parameters jitter, shimmer, HNR, LPC and PLP are parameters that revealed high hit rates when used, especially if the LPC and PLP deltas are used.

Regarding artificial intelligence tools, those that have obtained the best results were analyzed and, lately, they have been more used in the detection of speech pathologies. Taking into account the works described, the tools that have obtained better accuracy are LSTM, CNN and Transfer-Learning.

Research on the commercialization of vocal pathology detection systems has been investigated, however, despite the fact that there are enough systems, none are yet commercialized.

Acknowledgment

This work has been supported by Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

References

[1] R. Martins, M. Santana e E. Tavares, "Vocal Cysts: Clinical, Endoscopic, and Surgical Aspects," *Journal of Voice*, vol. 25, nº 1, pp. 107-110, 2011.

[2] J. Godino-Llorente, P. Gomez-Vilda e M. Blanco-Velasco, "Dimensionality Reduction of a Pathological Voice Quality Assessment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters," *IEEE Transactions on Biomedical Engineering*, vol. 53, nº 10, pp. 1943-1953, 2006.

[3] A. A. Dibazar, S. Narayanan e T. W. Berger, "Feature analysis for automatic detection of pathological speech," *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society [Engineering in Medicine and Biology]*, vol. 1, pp. 182-183, 2002.

[4] M. Markaki e Y. Stylianou, "Using modulation spectra for voice pathology detection and classification,," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2514-2517, 2009.

[5] R. T. S. Carvalho, C. C. Cavalcante e P. C. Cortez, "Wavelet transform and artificial neural networks applied to voice disorders identification," *2011 Third World Congress on Nature and Biologically Inspired Computing, Salamanca*, pp. 371-376, 2011.

[6] A. Tsnas, M. Little, P. McSharry e L. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," *Journal of the Royal Society Interface*, pp. 842-855, 17 Novembro 2010.

[7] F. Teixeira, J. Fernandes, V. Guedes, A. Junior e J. P. Teixeira, "Classification of Control/Pathologic Subjects with Support Vector Machines," *Procedia Computer Science*, vol. 138, pp. 272-279, 2018.

[8] H. Wu, J. Soraghan, A. Lowit e G. D. Caterina, "Convolutional Neural Networks for Pathological Voice Detection,," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1-4, 2018.

[9] S. Roy, M. I. Sayim e M. A. H. Akhand, "Pathological Voice Classification Using Deep Learning," *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1-6, 2019.

[10] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin e C.-T. Wang, “Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach,” *Journal of Voice,* vol. 33, nº 5, pp. 634-641, 2019.

[11] S. N. Awan e N. Roy, “Outcomes Measurement in Voice Disorders: Application of an Acoustic Index of Dysphonia Severity,” *Journal of Speech, Language, and Hearing Research,* vol. 52, pp. 482-499, 2009.

[12] E. Buder, “Acoustic Analysis of Voice Quality: A Tabulation of Algorithms 1902-1990,” *Voice Quality Measurement,* pp. 119-244, 2000.

[13] R. Baken e R. Orlikoff, Clinical Measurement of Speech and Voice, 2º ed., San Diego:: Singular, 2000.

[14] R. Kent e C. Read, The Acoustics Analysis of Speech, San Diego: Singular , 1992.

[15] A. Mendes, H. Rothman, C. Sapienza e W. B. Jr., “Effects of vocal training on the acoustic parameters of the singing voice,” *Journal of Voice,* vol. 17, nº 4, pp. 529-543, 2003.

[16] Y. Zhang e J. J. Jiang, “Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies,” *Journal of Voice,* vol. 22, nº 1, pp. 1-9, 2008.

[17] J. S. Vuppalapati, S. Kedari, S. Kedari, A. Ilapakurti e C. Vuppalapati, “Artificial Intelligent (AI) Clinical Edge for Voice disorder Detection,” *IEEE Fifth International Conference on Big Data Computing Service and Applications,* pp. 340-345, 2019.

[18] S. B. Kotsiantis, “Supervised machine learning: a review of classification techniques,” *Emerging Artificial Intelligence Application in Computer Engineering,* vol. 31, pp. 249-268, 2007.

[19] T. L. Lanc, “The Importance of Input Variables to a Neural Network Fault-diagnostic System for Nuclear Power Plants. Retrospective Theses and Dissertations,” 1992.

[20] T. Grzwalski, A. Maciaszek, A. Biniakowski, J. Orwat, S. Drgas, M. Piecuch, R. Belluzzo, K. Joachimiak, D. Niemiec, J. Ptaszynski e K. Szarzynski, “Parameterization of Sequence of MFCCs for DNN-based voice disorder detection,” *IEEE,* 2018.

[21] L. Verde, G. D. Pietro e G. Sannino, “Voice Disorder Identification by using Machine Learning Techniques,” *IEEE Acess,* vol. PP, pp. 1-1, 2018.

[22] V. Gupta, “Voice Disorder Detection Using Long Short Term Memory (LSTM) Model,” *Cornell University,* 2018.

[23] M. Alhussein e G. Muhammad, “Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework,” *IEEE Access,* vol. 6, pp. 41034-41041, 2018.

[24] M. Wester, “Automatic classification of voice quality: comparing regression models and hidden Markov models,” *Proceedings of Voicedata,* pp. 92-97, 1998.

[25] G. Pironkov, S. Dupont e T. Dutoit, “Speaker-aware long short-term memory multi-task learning for speech recognition,” *2016 24th European Signal Processing Conference (EUSIPCO),* pp. 1911-1915, 2016.

[26] Y. Lecun, L. Bottou, Y. Bengio e P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE,* vol. 86, nº 11, pp. 2278-2324, 1998.

[27] Y. Bengio, “Learning Deep Architectures for AI,” *Foundations,* vol. 2, pp. 1-55, 2009.

[28] L. Torrey e J. Shavlik, “Transfer learning,” *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques,* pp. 242-264, 2010.

[29] E. Bezerra, “Introdução à aprendizagem profunda,” Rio de Janeiro, 2016.

[30] D. PANEK, A. SKALSKI, J. GAJDA e R. TADEUSIEWICZ, “ACOUSTIC ANALYSIS ASSESSMENT IN SPEECH PATHOLOGYDETECTION,” *Int. J. Appl. Math. Comput. Sci.,* vol. 25, nº 3, pp. 631-643, 2015.

[31] N. Dave, “Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition,” *INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY,* vol. 1, nº VI, pp. 1-5, 2013.

[32] O. C. Ai, M. Hariharan, S. Yaacob e L. S. Chee, “Classification of speech dysfluencies with MFCC and LPCC features,” *Expert Systems with Applications,* vol. 39, nº 2, pp. 2157-2165, 2012.

[33] E. Wong e S. Sridharan, “Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification,” *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing ISIMP 2001 (IEEE Cat. No.01EX489),* pp. 95-98, 2001.

[34] G. Muhammad, S. M. M. Rahman, A. Alelaiwi e A. Alamri, “Smart Health Solution Integrating IoT and Cloud: A Case Study of Voice Pathology Monitoring,” *IEEE Communications Magazine,* vol. 55, nº 1, pp. 69-73, 2017.

[35] V. Guedes, F. Teixeira, A. Oliveira, J. Fernandes, L. Silva, A. Junior e J. P. Teixeira, “Transfer Learning with AudioSet to Voice Pathologies Identification in Continuous Speech,” *Procedia Computer Science,* vol. 164, pp. 662-669, 2019.

[36] V. Guedes, A. Junior, J. Fernandes, F. Teixeira e J. P. Teixeira, “Long short term memory on chronic laryngitis classification,” *Procedia Computer Science,* vol. 138, pp. 250-257, 2018.