INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION (IRJMT)

http://www.mapletreejournals.com/index.php/IRJMT

Received 20 September 2019 ISSN 2582-1040
Accepted 23 September 2019 2019; 1(5);1-7
Published online 25 September 2019

# Voice Disorder Recognition and Analysis Using Knowledge Engineering Techniques

**M.Vengateshwaran[1*], N.Gowsalya[2], K.Atchaya[2], R.Nivetha[2]**

[1]*Assistant Professor , Department of CSE, Arasu Engineering College, Kumbakonam,TN,India*
[2,3,4]*UG Student , Department of CSE, Arasu Engineering College, Kumbakonam, TN, India*

**Corresponding author E-Mail ID: mkvengatesh@gmail.com, Mobile: +91 9940857074*

DOI: https://doi.org/10.34256/irjmt1951

## ABSTRACT

Nowadays, the use of mobile application is most important thing in the healthcare sector is increasing rapidly. Mobile technologies not only for communication for multimedia content (e.g. clinical audio-visual notes and medical records) but also promising solutions for people who desire the identification, monitoring, and treatment of their health conditions anywhere and at any time. Mobile E-healthcare systems can contribute to make patient care faster, better, and cheaper. Several pathological conditions can benefit from the use of mobile technologies. In this paper we focus on dysphonia, an alteration of the voice quality that affects about one person in three at least once in his/her lifetime. Voice disorders are rapidly spreading, although they are often underestimated. Mobile health systems can be an easy and fast support to voice pathology detection. The identification of an algorithm that discriminates between pathological and healthy voices with more accuracy is necessary to realize a valid and precise mobile health system. . This technique is evaluated by based on experimental results deep neural networks with machine learning approach to provide an accuracy of 99.89% in detecting voice. In this field to detect any abnormal structure and analysis without human intervention in health care sector to enhance the utility of well beginning system.

**Keywords:** Mobile health systems, machine learning techniques, voice disorders, accuracy

## 1. INTRODUCTION

The introduction of mobile devices for data transmission or disease control and monitoring has been a main attraction of research and business communities. They offer, in fact, numerous opportunities to realise efficient mobile health (mhealth) systems. These solutions can allow patients and doctors to access medical records, clinical audio-visual notes and drug information anywhere and at any time from their mobile devices, such as a tablet or smartphone, to monitor several conditions [1]. M-health solutions can also be used in other important applications such as the detection and prevention of specific diseases, decision making and the management of chronic conditions and emergencies, improving the quality of patient care and reducing the costs of healthcare. Several pathological conditions can be detected and monitored, such as the well known and widespread cardiovascular diseases. In recent years, probably also due to the diffusion of the Internet of Things (IoT) and cloud technologies, there has been a development of monitoring systems in an unobtrusive, portable and easy way using wearable sensors and wireless communications, such as the solutions described in [2]–[7]. These systems are able to achieve health data monitoring and analysis, helpful for patients suffering from cardiovascular diseases or

for their physical therapy. If, on the one hand, health monitoring systems for cardiovascular diseases are so celebrated, on the other hand, there are other little known and often underestimated disorders, such as dysphonia, that could benefit from m-health solutions. Dysphonia is a disorder that occurs when the voice quality, pitch and loudness are altered. About 10% of the population suffer from this disorder [8], caused mainly by unhealthy social habits and voice abuse. Unfortunately, a large number of individuals with voice disorders do not seek treatment. Therefore, m-health systems could be an efficient support for the diagnosis and screening of voice disorders.

Clinical voice pathology detection is performed through the execution of several procedures, such as the acoustic analysis. It consists of an estimation of appropriate parameters extracted from voice signal to evaluate any possible alterations of the vocal tract, according to the guidelines of the SIFEL protocol (Società Italiana di Foniatria e Logopedia), developed by the Italian Society of Logopedics and Phoniatrics, following the instructions of the Committee for Phoniatrics of the European Society of Laryngology. It is a non-invasive examination in clinical practice, complementary to other medical tests, such as the laryngoscopic examination based on the direct observation of the vocal folds. Several acoustic parameters are estimated to evaluate the state of health of the voice. Unfortunately, the accuracy of these parameters in the detection of voice disorders is, often, related to the algorithms used to estimate them. For this reason the main effort of researchers is oriented to the study of acoustic parameters and the application of classification techniques able to obtain a high discrimination accuracy. Recently, speech pathology has focused interest on machine learning techniques. In this work, we want to discuss the application of machine learning algorithms and features selection methods capable of discriminating between pathological and healthy voices with a better accuracy. In detail, we evaluate the pathology recognition using the information data of patients, such as age and gender, and different features extracted from the voice signals. The adopted parameters are those estimated in the clinical acoustic analysis, such as the Fundamental Frequency (F0), jitter, shimmer and Harmonic to Noise Ratio (HNR). In addition, other parameters, the Mel-Frequency Cepstral Coefficients (MFCC), the first and second derivatives, are used due to their wide application both in machine learning techniques and in voice disorders classification as reported in several studies. The performances are evaluated in terms of accuracy, sensitivity, specificity and receiver operating characteristic (ROC) area for each considered machine learning methods.

## 2. RELATED WORK

Speech or, in general, the voice signal is used in several kinds of application ranging from emotion recognition to patient healthcare state recognition. Several m-health solutions, such as, adopt these signals to estimate the state of voice health, as well as systems that use voice signals to evaluate emotional condition. Voice disorder detection has, often, been achieved through expert systems techniques, and over recent years, several approaches have been developed to improve the performance in terms of accuracy in the discrimination between healthy and pathological voices. These studies are focused on the identification of parameters to measure the voice condition and new techniques able to detect voice disorders. Among several expert systems techniques existing in literature, Support Vector Machine (SVM) has been widely used in voice signal processing. Godino-Llorente et al, for example, focused on the classification of pathological and healthy voices based on MFCC to train and test an SVM classifier. These have obtained a good accuracy (95%). However, the poor numerosity of the used dataset composed of only 173 pathological and 53 healthy voices selected by the Massachusetts Eye and Ear Infirmary voice and speech lab (MEEI) database should be obsverved. Additionally, important information, as for example the pathologies of the selected voices, is not available in this work. The SVM technique was also used in to estimate the presence of dysphonia, investigating four types of pathology: chronic laryngitis, cysts, Reinke's edema and spasmodic dysphonia. The authors

proposed an algorithm based on the use of MFCC and Linear Discriminant Analysis (LDA) as a dimensionality reduction method. This algorithm identifies the presence of a pathology with a discrete accuracy (86%). However, it was tested on a very limited dataset. In fact, only 70 pathological and 40 healthy voices were selected by the Saarbruechen Voice Database (SVD) .


***Fig.1 System Architecture***

***Table.1 Voice Disorder detection***

| Database used | #of voices (pathological + healthy) | #of voices used for training (pathological + healthy) | #of voices used for testing (pathological + healthy) | Features used | Classifier used | Accuracy(%) over the testing set | Accuracy(%) over the training set |
|---|---|---|---|---|---|---|---|
| MEEI | 226 (173+53) | 158 (n.a.) | 68 (n.a.) | MFCC, Noise Features and Temporal Derivatives | SVM | 95 | n.a. |
| SVD | 110 (40+70) | 77 (n.a.) | 33 (n.a.) | MFCC and Temporal Derivatives | SVM | 86 | n.a. |
| MEEI | 154 (118+36) | 4-fold | classification | MFCC | GMM | 77.90 | n.a. |
| SVD | 100 (40+60) | 75 (35+40) | 25 (10+15) | MFCC | SVM, GMM | 96.5 - 95.5 | n.a. |
| SVD | 101 (38+63) | 81 (29+52) | 20 (9+11) | MFCC, jitter and shimmer | GMM | 82.37 | n.a. |
| Private | 60 (30+30) | 48 (n.a) | 12 (n.a.) | Wavelet Transform | LS-SVM | 90 | n.a. |
| Private | 152 (111+41) | 101 (n.a.) | 51 (n.a.) | jitter, shimmer, NHR, SPI, APQ and RAP | GMM | 95.2 | 98.4 |
| Private | 77 (n.a.) | n.a. | n.a. | Spectral Features | GMM | 94 | n.a. |
| Private | 400 (300+100) | 200 (150+50) | 200 (150+50) | Perturbation, noise and energy parameters | K-NN, LDA | 93.5 | n.a. |

## 3. SYSTEM METHODS

In this study we analysed the accuracy in the discrimination of pathological from healthy voices of the main machine learning techniques to identify the most reliable one. The idea has been to integrate the best one in a valid m-health system, where the voice signal can be acquired by a mobile device, such as a smartphone or table, processed in realtime to extract the voice features, and analyzed by using the machine learning classifier to detect the presence or not of a voice disorders, as shown in Figure 1. In detail, we have evaluated the performance of SVM, the principal adopted technique in literature in relation to the Kernel function, and of some other machine learning algorithms used to identify the presence of voice disorders. The analysis has been performed using the WEKA tool, one of the most commonly used tools for data mining tasks, selected for the data analysis due to its efficiency, versatility and affordability. In the following subsections we introduce the dataset used in this study, the features extracted from the voice signal and used for the classification, and the machine learning techniques compared.

In our experimental tests, to perform the experiments on a well-balanced database containing both pathological and healthy voices, we have selected a total of 1370 /a/ vocalizations. In detail, we have chosen: • 685 pathological voices (257 male and 428 female); and • 685 healthy voices (257 male and 428 female).

***Table.2 voice signal***

| Category | Gender | Age Group | No. | % |
|---|---|---|---|---|
| Healthy | Female | 17-29 | 359 | 26.20 % |
| | | 30-39 | 27 | 1.98 % |
| | | 40-49 | 15 | 1.09 % |
| | | 50-59 | 13 | 0.95 % |
| | | 60+ | 14 | 1.02 % |
| Healthy | Male | 17-29 | 138 | 10.07 % |
| | | 30-39 | 62 | 4.53 % |
| | | 40-49 | 37 | 2.70 % |
| | | 50-59 | 9 | 0.66 % |
| | | 60+ | 11 | 0.80 % |
| Pathological | Female | 17-29 | 58 | 4.24 % |
| | | 30-39 | 94 | 6.86 % |
| | | 40-49 | 85 | 6.20 % |
| | | 50-59 | 87 | 6.35 % |
| | | 60+ | 104 | 7.59 % |
| Pathological | Male | 17-29 | 23 | 1.68 % |
| | | 30-39 | 24 | 1.75 % |
| | | 40-49 | 43 | 3.14 % |
| | | 50-59 | 74 | 5.40 % |
| | | 60+ | 93 | 6.79 % |
| ***Total*** | *Female* | *17-60+* | *856* | *62.48%* |
| | *Male* | *17-60+* | *514* | *37.52%* |

## 3.1 FEATURES EXTRACTION

Feature extraction is an important task that allows an improvement of the analysis and classification. The choice of which features of the speech signal to use in our study was made by taking into account two considerations. On the one hand, we have used the main parameters adopted by the specialist during the clinical evaluation; on the other, we have chosen the main features used in several correlated studies existing in literature concerning the use of machine learning techniques for the voice classification.

In detail, the parameters used in clinical practice are:

• Fundamental Frequency (F0): this represents the rate of vibration of the vocal folds constituting an important index of laryngeal function. It is at the basis of the other parameters calculated in the acoustic analysis and most noise estimation methods.

• Jitter: this describes the instabilities of the oscillating pattern of the vocal folds, quantifying the cycle-to-cycle changes in fundamental frequency.

$$Jitter(\%) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}(|T_i - T_{i+1}|)}{\frac{1}{N}\sum_{i=1}^{N}(T_i)} \qquad (1)$$

• Shimmer: this indicates the instabilities of the oscillating pattern of the vocal folds, quantifying the cycle-to cycle changes in amplitude.

$$Shimmer(dB) = \frac{1}{N-1}\sum_{i=1}^{N-1}|20log\frac{A_{i+1}}{A_i}| \qquad (2)$$

• Harmonic to Noise Ratio (HNR): this quantifies the ratio of signal information over noise due to turbulent airflow, resulting from an incomplete vocal fold closure in speech pathologies.

The parameters used in other correlated studies are:

• Mel-Frequency Cepstral Coefficients (MFCC): these coefficients try to analyse the vocal tract independently of the vocal folds that can be damaged due to voice pathologies. In this work, the experiments were conducted using 13 MFCC coefficients. • First and second derivatives of cepstral coefficient: these are useful to investigate the properties of the dynamic behaviour of the speech signal.

### 3.2 CLASSIFICATION

Classification includes a board range of decision theoretic approaches to the identification of voice datasets. All classification algorithms are based on the assumption that the voices in question depicts one or more features and that each of these features belongs to one of several distinct and exclusive classes. These classes may be specified a priori by an analysis or automatically clustered is the another form of component labelling.

Classification algorithms typically employ two phases of processing:

1] Training phase.
2] Testing phase.

Normally training phase, enhance the signal features are isolated in each classification categories, and then formulated the training class. In the testing phase, feature space partitions are used to classify voice signal features.

Minimum (mean) distance classifier:

$$m_{j=\frac{1}{Nj}} \sum_{x \in wj} x \; for \; j = 1,2, \ldots\ldots, M$$

Deep neural networks is a subset of machine learning algorithm that is very good at recognizing patterns but typically requires a large number of data. To recognize each object in an voice and then, implemented using Deep neural networks (DNN). Each layer are extracting further feature of the voice.

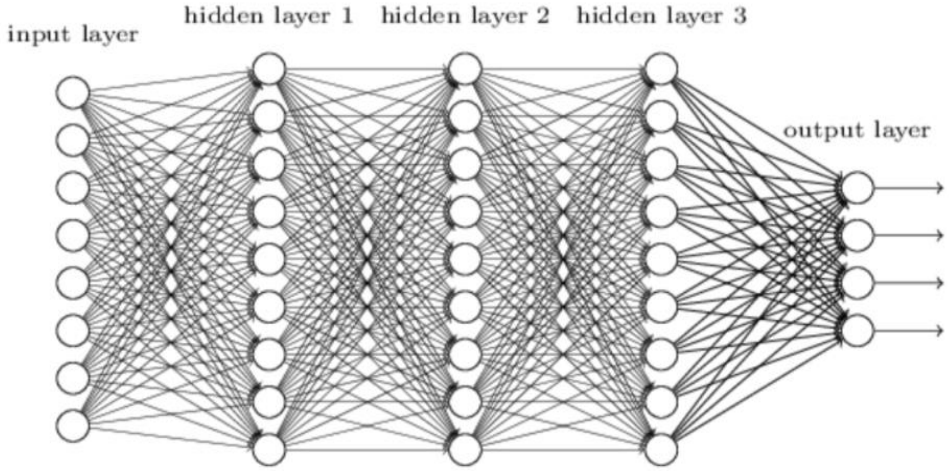


***Fig.2 Classification using DNN***

### 4. RESULTS AND DISCUSSION

The performance of the selected machine learning classification techniques was evaluated in terms of accuracy, sensibility, specificity and ROC area by using the following measurements:

- True Positive (TP): the voice sample is pathological and the algorithm recognizes this;
- True Negative (TN): the voice sample is healthy and the algorithm recognizes this;
- False Positive (FP): the voice sample is healthy but the algorithm recognizes it as pathological;
- False Negative (FN): the voice sample is pathological but the algorithm recognizes it as healthy

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (4)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (5)$$

***Table.3 Classification result of different algorithm***

| *Considering all parameters* | | | | | | |
|---|---|---|---|---|---|---|
| | ***SMO*** | ***DT*** | ***BC*** | ***LMT*** | ***Ibk*** | ***Kstar*** |
| Accuracy (%) | **85.77** | 82.04 | 78.61 | 82.48 | 76.28 | 49.71 |
| Sensitivity (%) | **87.59** | 81.02 | 79.85 | 86.28 | 77.37 | 64.09 |
| Specificity (%) | **83.94** | 83.07 | 77.37 | 78.69 | 75.18 | 35.33 |
| ROC Area | 0.858 | 0.832 | **0.863** | 0.885 | 0.763 | 0.524 |
| *Considering only the parameters selected with the InfoGainAttribute Eval features selection* | | | | | | |
| | ***SMO*** | ***DT*** | ***BC*** | ***LMT*** | ***Ibk*** | ***Kstar*** |
| Accuracy (%) | **84.16** | 82.48 | 78.61 | 82.12 | 76.20 | 50.22 |
| Sensitivity (%) | 85.26 | 80.58 | 79.85 | **86.13** | 77.23 | 64.82 |
| Specificity (%) | 83.07 | **84.38** | 77.37 | 78.10 | 75.18 | 35.62 |
| ROC Area | **0.842** | 0.830 | 0.863 | 0.883 | 0.762 | 0.535 |
| *Considering only the parameters selected with the Correlation method features selection* | | | | | | |
| | ***SMO*** | ***DT*** | ***BC*** | ***LMT*** | ***Ibk*** | ***Kstar*** |
| Accuracy (%) | 82.99 | **83.58** | 80.58 | 82.26 | 77.45 | 81.82 |
| Sensitivity (%) | 83.07 | 83.80 | 81.61 | **86.72** | 78.39 | 84.23 |
| Specificity (%) | 82.92 | **83.36** | 79.56 | 77.81 | 76.50 | 79.42 |
| ROC Area | 0.83 | 0.857 | 0.873 | **0.884** | 0.774 | 0.873 |
| *Considering only the parameters selected with the PCA features selection* | | | | | | |
| | ***SMO*** | ***DT*** | ***BC*** | ***LMT*** | ***Ibk*** | ***Kstar*** |
| Accuracy (%) | **71.75** | 69.20 | 69.12 | 69.12 | 64.53 | 67.66 |
| Sensitivity (%) | **82.34** | 62.92 | 65.99 | 74.45 | 64.67 | 72.12 |
| Specificity (%) | 61.17 | **75.47** | 72.26 | 63.80 | 64.38 | 63.21 |
| ROC Area | 0.718 | 0.737 | 0.745 | **0.753** | 0.645 | 0.737 |

## 5. CONCLUSION

In recent years, the use of mobile multimedia services and applications in healthcare sector has been increasing significantly. Mobile health applications allow people to access medical information and data of interest at any time and anywhere, useful for the monitoring and detection of specific diseases, such as dysphonia, a voice disorder often underestimated that affects a great percentage of people. Research on mobile automatic systems to estimate voice disorders has received considerable attention in the last few years due to its objectivity and non-invasive nature. Machine learning techniques can be a valid support to investigate new approaches to signal processing in an easy and fast way that can be implemented in an m-health solution. This study compares the performance of different voice pathology identification methods, taking into account the main machine learning techniques. Several techniques are applied such as the Support Vector Machine, Decision Tree, Bayesian Classification, Logistic Model Tree and Instance-based Learning algorithms. Moreover, in this work we focus on identifying appropriate voice signal features by using the comparative study of different classifiers. All analyses are performed on a wide dataset of 1370 voices selected from the Saarbruecken Voice Database.

## REFERENCES

[1]. B. M. C. Silva, J. J. C. Rodrigues, I. de la Torre Díez, M. López-Coronado, and K. Saleem, ‘‘Mobile-health: A review of current state in 2015,’’ J. Biomed. Inform., vol. 56, pp. 265–272, Aug. 2015.

[2]. S. Naddeo, L. Verde, M. Forastiere, G. De Pietro, and G. Sannino, ‘‘A real-time m-health monitoring system: An integrated solution combining the use of several wearable sensors and mobile devices,’’ in Proc. HEALTHINF, 2017, pp. 545–552.

[3]. M. S. Hossain and G. Muhammad, ‘‘Cloud-assisted Industrial Internet of Things (IIoT)—Enabled framework for health monitoring,’’ Comput. Netw., vol. 101, pp. 192–202, Jun. 2016.

[4]. G. Sannino, I. De Falco, and G. De Pietro, ‘‘A supervised approach to automatically extract a set of rules to support fall detection in an mHealth system,’’ Appl. Soft Comput., vol. 34, pp. 205–216, Sep. 2015.

[5]. J. Mohammed, C.-H. Lung, A. Ocneanu, A. Thakral, C. Jones, and A. Adler, ‘‘Internet of Things: Remote patient monitoring using Web services and cloud computing,’’ in Proc. IEEE Int. Conf. Green Comput. Commun. (GreenCom), IEEE Cyber, Phys. Social Comput. (CPSCom), Internet Things (iThings), Sep. 2014, pp. 256–263.

[6]. R. Gravina and G. Fortino, ‘‘Automatic methods for the detection of accelerative cardiac defense response,’’ IEEE Trans. Affect. Comput., vol. 7, no. 3, pp. 286–298, Jul./Sep. 2016.

[7]. S. Iyengar, F. T. Bonda, R. Gravina, A. Guerrieri, G. Fortino, and A. Sangiovanni-Vincentelli, ‘‘A framework for creating healthcare monitoring applications using wireless body sensor networks,’’ in Proc. ICST 3rd Int. Conf. Body Area Network. (ICST), 2008, p. 8.

[8]. R. H. G. Martins, H. A. do Amaral, E. L. M. Tavares, M. G. Martins, T. M. Conclaves, and N. H. Dias, ‘‘Voice disorders: Etiology and diagnosis,’’ J. Voice, vol. 30, no. 6, pp. 761.e1–761.e9, Nov. 2016.

### Conflict of Interest

None of the authors have any conflicts of interest to declare.