Frontiers in Health Informatics www.ijmi.ir
2024; 13: 205 Open Access

Voice as an indicator for laryngeal disorders using data mining approach

Mohammadjavad Sayadi[1,2], Mostafa Langarizadeh[2]*, Farhad Torabinezhad[3], Gholamreza Bayazian[4]

[1]Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran
[2]Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran
[3]Rehabilitation Research Center, Department of Speech Therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran
[4]ENT and Head &Neck Research Center, The Five Sense Health Institute, Rasoul Akram Medical Complex, Iran University of Medical Sciences, Tehran, Iran

Article Info

Article type:
Research

Article History:
Received: 2024-02-28
Accepted: 2024-04-07
Published: 2024-04-20

**** Corresponding author:***
Mostafa Langarizadeh

Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

Email: langarizadeh.m@iums.ac.ir

Keywords:
Voice
Laryngeal Disorders
Indicator
Data Mining

ABSTRACT

Introduction: Laryngeal disorders are a common health problem that affects people of all ages, genders, and races. One of the main symptoms of laryngeal disorders is changes in the voice, which can be used as an indicator for the presence of such disorders. In this paper, we present a data mining approach for using voice as an indicator for laryngeal disorders.

Material and Methods: We collected a dataset of voice recordings from individuals with and without laryngeal disorders including 434 people from two clinical centers in Tehran. The dataset was created using a powerful signal processing program and then based on the difference between male and female voice, the dataset was separated into two datasets. Finally, a Deep Neural Network was implemented for modelling using Python programming language and F1-score, Accuracy, Sensitivity, Specificity, and AUC as the model's evaluation metrics were reported.

Results: Among all the acoustic features, 23 features were selected for the male dataset and 25 features for the female data set. For the male dataset the final model achieved F1-Score of 0.915 and Accuracy of 0.910. For the female dataset the result was 0.884 of F1-Score and 0.896 of Accuracy.

Conclusion: Our results show that machine learning algorithms can accurately classify voice recordings into two groups: individuals with laryngeal disorders and those without. The high accuracy achieved by the algorithms suggests that voice can be used as an objective and automated diagnostic tool for laryngeal disorders.

Cite this paper as:
Sayadi M, Langarizadeh M, Torabinezhad F, Bayazian G. Voice as an indicator for laryngeal disorders using data mining approach. Front Health Inform. 2024; 13: 205. DOI: 10.30699/fhi.v13i0.605

INTRODUCTION

Laryngeal disorders are a group of health problems that affect the larynx, which is the part of the body responsible for producing sound and protecting the airway during swallowing [1]. These disorders can occur for various reasons, such as infections, trauma, cancer, or neurological conditions. One of the main symptoms of laryngeal disorders is changes in the voice, which can range from hoarseness, breathiness, or weakness to complete loss of voice [2]. Detecting laryngeal disorders early is crucial for effective treatment and prevention of complications.

By providing a platform of targeted clinical knowledge, patient information, and other health information, clinical decision support systems are intended to help diagnose diseases and make medical decisions, thereby improving healthcare delivery. In these systems, the characteristics of a single patient are matched with a computerized clinical knowledge base, and specific evaluations or recommendations regarding the patient are presented to the doctor to make a decision to reach the correct diagnosis. The physician must combine his knowledge with the information or suggestions provided by the clinical decision support system [3]. The use of artificial and data-based intelligence tools and methods in the form of clinical decision support systems in the fields of disease diagnosis and prediction has become very common in recent years. These systems have been very good and accurate in diagnosing certain diseases and enable doctors, final decision makers and health care organizations to make more informed judgments about problems and diseases [4].

In recent years, there has been growing interest in using machine learning techniques to analyze medical data and improve diagnostic accuracy. In this paper, we investigate the use of voice as an indicator for laryngeal disorders using data mining techniques. We aim to develop a model that can accurately classify voice recordings into two groups: individuals with laryngeal disorders and those without.

Several studies have investigated the use of voice analysis for diagnosing laryngeal disorders. These studies have used various techniques, such as acoustic analysis, perceptual evaluation, or a combination of both.

Asadi et al. [5] have investigated the effect of the acoustic characteristics of voice on the identification of people's identity, and the results can be a basis for the diagnosis of some diseases and mental disorders, especially identity disorders, with further development and investigation. In this study, he emphasized the intensity of syllable expression and made it the basis of analysis and diagnosis, and ignored the effect of other characteristics.

Abbaspour et al. [6] tried to diagnose speech disorders by using audio signal processing. In this study, an intelligent tool is presented that uses the acoustic characteristics of the voice to classify people based on the presence or absence of speech disorder.

Arjamandi et al. [7] sought to identify people's voice problems using a different method. In their study, Kivu et al. [8] tried to classify and diagnose speech disorders using deep learning. This study, which was based on Chinese language samples, used the acoustic characteristics of sound. This study, like the previous studies, sought to identify the problem in people's speech.

Other studies have used perceptual evaluation by trained experts to diagnose laryngeal disorders based on the voice quality [9]. While perceptual evaluation is considered the gold standard for diagnosing voice disorders, it is subjective, time-consuming, and requires specialized training. Therefore, there is a need for objective and automated methods for diagnosing laryngeal disorders.

MATERIAL AND METHODS

In this study, the research population included people who referred to the Otorhinolaryngology and Speech Therapy Clinic of Hazrat Rasool Akram Hospital in Tehran and the Speech Therapy Center of the Rehabilitation Faculty of Iran University of Medical Sciences. In total, the data of 434 samples were collected. The collected data of each sample included a personal information questionnaire form, current disease data, the individual's medical history recorded in a paper form, and the patient's larynx voice files, which were recorded by a mobile phone-based application. The records of laryngeal disease were asked from the referring person and the patient's health record was checked to ensure. The current condition of the patient's larynx was examined by an ENT specialist.

The collected data were pre-processed in five steps:

First step: In the first step, the samples that did not meet the minimum quality in their stored data were removed. The meaning of quality was mostly related to the quality of the stored voice of the samples, which was either incomplete or the noise and sounds of the environment dominated the voice. Finally, a small number of samples were removed from the sample collection due to the presence of irrelevant and misleading data in the data collection form. The number of samples that remained at this stage was 289, of which 186 were men and 103 were women. Table 1 shows the distribution of samples in male and female categories.

Table1: Distribution of samples in male and female categories

	Male	Female
Positive	69	38
Negative	117	65

Second step: In this step, the features of the saved voice files of the samples were extracted. A signal processing program was implemented with Python programming language to extract the features of voices. To ensure the correctness of its results, the values obtained by two speech therapists were compared with the output of a specialized voice analysis software. Table 2 shows the list of voice features.

Third step: In this step the dataset file containing the data extracted from voice files and the data collection forms were preprocessed.

Forth Step: In the created data set, there are features that are in the form of an array in which the value of that feature is extracted in the form of a vector for the voice file. To use these features in the modeling process, their statistical description was replaced. Therefore, Maximum, Minimum, Mean and Median values were used for each vector feature.

Fifth step: At this stage, according to the significant difference of the extracted features in male and female samples, the dataset was divided into two sets, one of which contained only male samples and the other contained female samples. After that, for each of these two data sets, firstly, the best subset of features was selected to enter the modeling phase with the Pearson Correlation method, which is suitable for systems with numerical input and qualitative output. In this step, different subsets were evaluated and the features with the highest correlation to the target variable were selected. Finally, by using a deep neural network, the diagnosis model of larynx diseases was created. To validate the models, Holdout and K-Fold methods were examined and due to the better performance of the Holdout validation method, Holdout method was used as validation in these models, where for the models used for the Males, 70% of the data was selected for training and 30% for testing, and for the models that were implemented for females, 80% was selected for model training and 20% for model testing. Finally, according to the results obtained in the model implemented using deep neural network, with the SMOT method, the volume of data increased by about five times. It is important that the model test was performed only with a part of the original data that was selected before increasing the sample size.

Table2: List of extracted Voice features

No	Feature's Name	Feature's Description
1	Chroma Cense	The index of displaying melodic sound information
2	Jitter	The cycle-to-cycle variability of the period duration of the acoustic signal coming from voice production
3	Mel-Frequency Cepstrum Coefficients (MFCC)	The representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency
4	Pitch	The feature of determining the amount of change in the frequency of the sound signal
5	Shimmer	The feature of determining the amount of change in the audio signal peak
6	Cepstral Peak Prominence (CPP)	Criterion for determining sound quality
7	Formant	Frequency peaks in the sound spectrum that have a high degree of energy
8	Vocal Tract Estimate	Is a cylindrical tube extending from the larynx to the lips
9	Intensity	The power carried by sound waves per unit area in a direction perpendicular to that area
10	Signal to Noise Ratio	A standard measure of the amount of background noise present in a speech (or other) signal
11	Spectral Shape	An object that calculates several audio descriptors and bundles them together

Python programming language was used for feature selection and modeling. At this phase, the F1-Score was used for analysis, and the main measures of Accuracy, Sensitivity, Specificity, and the AUC were calculated and reported.

RESULTS

The selected features for male and female samples using Pearson Correlation method are shown in Table 3. This method selected 25 features for female dataset and 23 features for male dataset.

Table3: Selected feature for male and female

Female	Male
Mfcc1_delta_min	Spectral_flux_median
Mfcc3_max	Mfcc2_delta_min
Mfcc8_accelerate_mean	Spectral_flux_mean
Mfcc0_delta_median	Mfcc18_accelerate_median
Chroma4_median	Spectral_bandwidth_3_median
Mfcc11_max	Mfcc1_delta_min
Chroma5_mean	Chroma8_mean
Mfcc3_mean	Mfcc3_median
Mfcc11_mean	Chroma7_median
Mfcc17_min	spectral_bandwidth_3_mean
Mfcc10_delta_mean	Mfcc8_mean
Chroma1_median	Mfcc14_delta_max
Mfcc4_mean	Spectral_bandwidth_2_mean
Mfcc11_median	Mfcc3_mean
Chroma11_mean	Mfcc13_max
Mfcc6_max	Mfcc5_accelerate_max
Perceptual_shock_wave_median	Spectral_centroids_mean
Mfcc2_accelerate_max	Mfcc7_max
Spectral_flux_median	Mfcc14_delta_min
Mfcc0_min	Mfcc16_max
Spectral_rolloff_median	Mfcc4_max
Mfcc6_median	Chroma11_mean
Mfcc6_mean	Mfcc13_delta_median
Mfcc3_accelerate_min	
Mfcc13_accelerate_min	

The implemented Deep Neural Network model for male dataset had eight layers that received 23 features (Table 3) as input and was trained in 39 rounds. The structure of the deep neural network of this stage is shown in Fig 1.

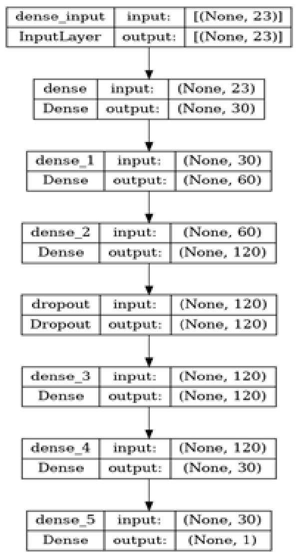

Fig 1: The structure of implemented deep neural network for male dataset

The implemented model, out of 56 samples that were used for the model test, in 45 cases, the model was correctly diagnosed and 11 cases were wrongly diagnosed. Table 4 shows the confusion matrix of this model.

Table 4: Confusion matrix for male dataset

	Diagnosis of laryngeal disease	Diagnosing the absence of laryngeal disease
Presence of laryngeal disease	23	6
Absence of laryngeal insufficiency	5	22

According to the model evaluation criteria shown in Table 5, the model implemented at this stage has detected larynx insufficiency with 80% accuracy. The F1 score is a relatively acceptable value of about 0.81, the area under the ROC diagram (AUC) shown in Fig 2 was 0.93.

Table 5: Evaluation metrics of implemented model for male dataset

F1-Score	Accuracy	Sensitivity	Specificity	AUC
0.807	0.808	0.821	0.793	0.931

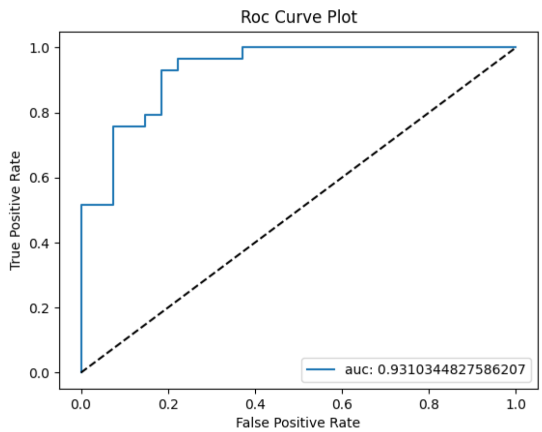

Fig 2: ROC curve diagram for male dataset

To improve the implemented model, the number of samples was increased with the SMOTE method and the Deep Neural Network model was trained again. The evaluation criteria of the new implemented model are shown in Table 6, where the accuracy rate of the model has increased to 91% and the area under the ROC chart (AUC) shown in Fig 3 has increased to 0.96.

Table 6: Evaluation metrics of improved model for male dataset

F1-Score	Accuracy	Sensitivity	Specificity	AUC
0.915	0.910	0.9	0.923	0.965

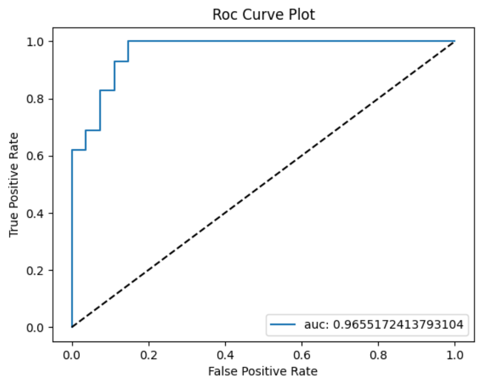

Fig 3: ROC curve diagram for the improved model for male dataset

The implemented Deep Neural Network model for female dataset had eight layers that received 25 features (Table 3) as input and was trained in 39 rounds. The structure of the deep neural network of this stage is shown in Fig 4.

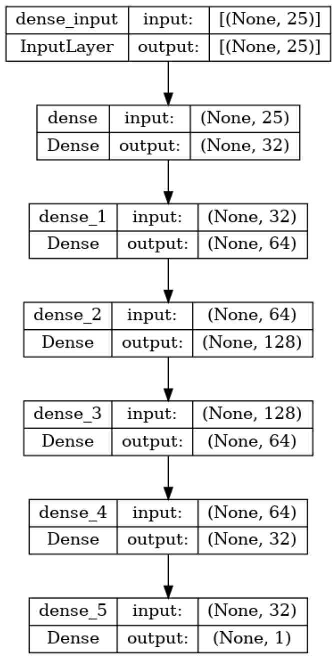

Fig 4: The structure of implemented deep neural network for female dataset

The implemented model, out of 26 samples that were used for the model test, in 20 cases, the model was correctly diagnosed and 6 cases were wrongly diagnosed. Table 7 shows the confusion matrix of this model.

According to the model evaluation criteria shown in Table 8, the model implemented at this stage has detected larynx insufficiency with 80% accuracy. The F1 score is a relatively acceptable value of about 0.81, the area under the ROC diagram (AUC) shown in Fig 5 was 0.93.

Table 7: Confusion matrix for female dataset

	Diagnosis of laryngeal disease	Diagnosing the absence of laryngeal disease
Presence of laryngeal disease	13	1
Absence of laryngeal insufficiency	5	7

Table 8: Evaluation metrics of improved model for female dataset

F1-Score	Accuracy	Sensitivity	Specificity	AUC
0.812	0.769	0.722	0.875	0.851

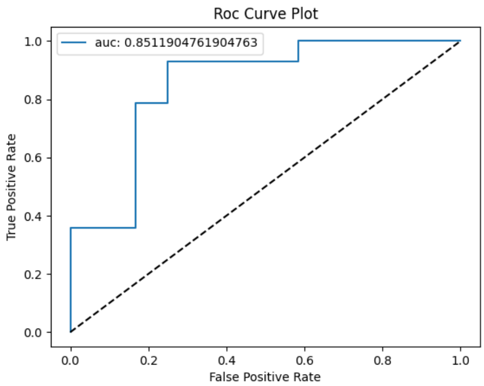

Fig 5: ROC curve diagram for male dataset

To improve the implemented model, the number of samples was increased with the SMOTE method and the Deep Neural Network model was trained again. The evaluation criteria of the new implemented model for female dataset are shown in Table 9, where the accuracy rate of the model has increased to 88% and the area under the ROC chart (AUC) shown in Fig 6 has increased to 0.9.

Table 9: Evaluation metrics of improved model for female dataset

F1-Score	Accuracy	Sensitivity	Specificity	AUC
0.896	0.884	0.866	0.909	0.904

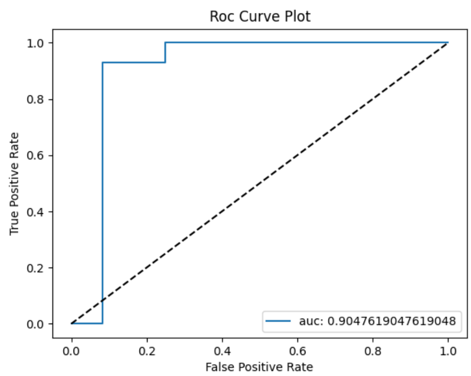

Fig 6: ROC curve diagram for the improved model for female dataset

DISCUSSION

In this article, models were presented on the data set containing characteristics related to the voice of the larynx, which are related to the diagnosis of the presence of diseases of the larynx, and according to the results obtained, only the acoustic features were effective in this model. At this stage, male and female samples were separated from each other and modeled separately.

In order to improve the efficiency of the implemented model, the data was developed using the Smot method and the deep learning model was re-implemented. Hold Out validation method was used in all stages. In previous similar studies, models have been presented to diagnose laryngeal diseases using the patient's voice, which are described below.

In 2015, Qasimzadeh et al. conducted a study to diagnose the problems of patients' vocal cords. In this study, the acoustic features of the patient's voice were used to detect the presence of a problem using the SVM algorithm [10]. In this study, an acoustic feature was used. has been used and used the sound of the larynx to train the system. In the present study, it has been tried to include all the acoustic features that are correlated with the existence of laryngeal disease in the modeling. Therefore, a range of features has been used instead of a single feature. Another noteworthy point is that in this study, the validation results of a model were reported and the sample was not separated based on gender.

Mahmoud et al. [11], Fonseca et al. [12], Varikas et al. [13] and Jarolin et al. Its accuracy is reported to be 99.3%. These studies have also used the acoustic features of sound for modeling, and deep learning, support vector machine, support vector machine, and K-nearest neighbor have been used, respectively, and have generally reported acceptable accuracy. But the current research has tried by using a range of features. In all of these studies, one feature was used for diagnosis, while in this study, more than 20 acoustic features were used simultaneously for modeling, and the effect of laryngeal diseases on the changes of those features was investigated.

Our results demonstrate that voice can be used as an effective indicator for the presence of laryngeal disorders. The high accuracy achieved suggests that the acoustic features extracted from the voice recordings are informative and can differentiate between healthy and disordered voices. The results also suggest that a data mining approach can be used to develop an objective and automated diagnostic tool for laryngeal disorders.

One limitation of our study is the small size of the dataset, which may affect the generalizability of the results. Therefore, further studies with larger and more diverse datasets are needed to validate our findings. Additionally, future research can explore the use of other acoustic features or combine them with other clinical or imaging data to improve the diagnostic accuracy.

CONCLUSION

In this paper, we present a data mining approach for using voice as an indicator for laryngeal disorders. Our results show that machine learning algorithms can accurately classify voice recordings into two groups: individuals with laryngeal disorders and those without. The high accuracy achieved by the algorithms suggests that voice can be used as an objective and automated diagnostic tool for laryngeal disorders. Further research is needed to validate the findings and explore other potential features or data sources for improving the diagnostic accuracy.

ACKNOWLEDGMENT

This study was part of a PhD thesis supported by Iran University of Medical Sciences (grant No: 1400-1-37-20909).

AUTHOR'S CONTRIBUTION

All authors contributed to the literature review, design, data collection and analysis, drafting the manuscript, read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

ETHICS APPROVAL

This study was approved by the researcher's institute review board at Iran University of Medical Sciences. The approval code number was IR.IUMS.REC.1400.327.

REFERENCES

1. Yavari A, Fatehi F, Dalvand H, Valizadeh A, Moradzadeh R, Mirhoseini FS. Prevalence of speech disorders in Arak primary school students, 2014-2015. Journal of Arak University of Medical Sciences. 2016; 19(111): 87-94.
2. Rufo MJ, Martin J, Perez CJ, Paniagua S. A bayesian decision analysis approach to assess voice disorder risks by using acoustic features. Biom J. 2019; 61(3): 503-13. PMID: 30408226 DOI: 10.1002/bimj.201700233 [PubMed]
3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: Benefits, risks, and strategies for success. NPJ Digit Med. 2020; 3: 17. PMID: 32047862 DOI: 10.1038/s41746-020-0221-y [PubMed]
4. Musen MA, Middleton B, Greenes RA. Clinical decision-support systems. In: Shortliffe EH, Cimino JJ (Eds). Biomedical informatics: Computer applications in health care and biomedicine. 5th ed. Springer; 2021.
5. Asadi H, AliNejad B. Speech rhythm measures: Acoustic cues for speaker identification. Journal of Language Research. 2022; 12(2): 29-49.
6. Abaspur Kazerouni I, Moslemi H, Hourali F. The intelligent diagnosis of speech disorder: Using of audio signal processing system. Medical Journal of Tabriz University of Medical Sciences. 2018; 40(5): 55-62.
7. Arjmandi MK, Pooyan M, Mikaili M, Vali M, Moqarehzadeh A. Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. J Voice. 2011; 25(6): e275-89. PMID: 21186096 DOI: 10.1016/j.jvoice.2010.08.003 [PubMed]
8. Kuo Y-M, Ruan S-J, Chen Y-C, Tu Y-W. Deep-learning-based automated classification of chinese speech sound disorders. Children (Basel). 2022; 9(7): 996. PMID: 35883979 DOI: 10.3390/children9070996 [PubMed]
9. Sayadi M, Varadarajan V, Langarizadeh M, Bayazian G, Torabinezhad F. A systematic review on machine learning techniques for early detection of mental, neurological and laryngeal disorders using patient's speech. Electronics. 2022; 11: 4235.
10. Ghasemzadeh H, Khass MT, Arjmandi MK, Pooyan M. Detection of vocal disorders based on phase space parameters and lyapunov spectrum. Biomedical Signal Processing and Control. 2015; 22(1): 135-45.
11. Mahmoud SS, Kumar A, Li Y, Tang Y, Fang Q. Performance evaluation of machine learning frameworks for aphasia assessment. Sensors (Basel). 2021; 21(8): 2582. PMID: 33916993 DOI: 10.3390/s21082582 [PubMed]
12. Fonseca ES, Guido RC, Scalassara PR, Maciel CD, Pereira JC. Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. Comput Biol Med. 2007; 37(4): 571-8. PMID: 17078942 DOI: 10.1016/j.compbiomed.2006.08.008 [PubMed]
13. Verikas A, Gelzinis A, Bacauskiene M, Hållander M, Uloza V, Kaseta M. Combining image, voice, and the patient's questionnaire data to categorize laryngeal disorders. Artif Intell Med. 2010; 49(1): 43-50. PMID: 20338736 DOI: 10.1016/j.artmed.2010.02.002 [PubMed]