Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-6 Year 2023** Page **355:358***

A Review: Voice Pathology Classification Using Machine Learning

Zubin Nizam Khan[1*], Nisha Balani[2], Uma Patel Thakur[3]

[1]M. Tech Scholar, Jhulelal Institute Of Technology, Nagpur, Maharashtra, India
[2,3]Professor, Jhulelal Institute of Technology Nagpur, Maharashtra, India
-7571-8400[4]

**Corresponding author's: Zubin Nizam Khan*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 25 Nov 2023

Abstract

Voice pathology detection requires the presence of a specialist doctor and time to treat each patient, but it is not always possible to have a doctor who can treat all patients at once and at one precise time. For residents of remote areas, it is all expensive equipment that must be provided. Or even for people who may not be aware of having any voice pathology. Our goal is to design a diagnostic aid system to detect whether the voice is pathological or healthy, so that the patient can be referred to a doctor or not without being moved from the start. Our system is based on the classification, by SVM "Support Vector Machine", using MFCCs "Mel Frequency Cepstral Coefficients" extracted from the patient's voice. The learning and testing of our system are done using the SVD database "Saarbruecken Voice Database"

Keywords: *Voice Disorder; Machine Learning*

1. Introduction

Sound is the most important communication tool among human beings. Vocal communication is a basic skill that people need it to express their feelings for daily social interactions and earn it during their life span. Professionals such as singers, actors, auctioneers, lawyers, and teachers who use their voices at a higher level than normal levels are at risk of pathological voice problems. Also, people who have voice disturbances due to the voice misuse, neurological disorders, drug use and unhealthy social habits may encounter many problems when communicating with people and lead to many social and personal complications. Vocal cord vibration is affected in different ways depending on the type and location of the disease in the vocal fold. Voice pathologies resulting from changes in sound quality, pitch and volume can be clinically determined by performing several procedures, such as acoustic analysis which consists of estimation of appropriate parameters extracted from an audio signal to assess possible changes in the audio path in accordance with the rules of the SIFEL protocol

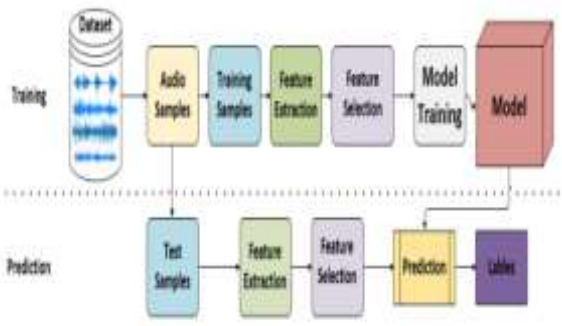

Figure.1. Illustration of voice pathology detection system

Acoustic parameters are then used to evaluate sound health status. The accuracy of these parameters in detecting voice disturbances plays an important role in the detection of voice disturbances. The accuracy of the acoustic parameters is related to the algorithms used to estimate them. There has been an interest recently in machine learning techniques in the field of sound pathology in order to increase the accuracy of these parameters. In this study, application of feature selection techniques and machine learning

algorithms with the ability to distinguish the patient and healthy sounds with the best accuracy is investigated. In detail, we evaluate the pathology using different features obtained from voice signals in addition to patients' age and gender information The performance of the used machine learning methods is evaluated in terms of accuracy, sensitivity and specificity

The detection of the voice pathology is a very important task, which must not be neglected before its progression. There are many precancerous pathologies that present a great danger, and other pathologies that are benign but have a negative effect on the voice. The voice becomes incomprehensible, which makes the communication with people very difficult. In order to avoid the progression of voice pathologies, it is necessary to follow a treatment in the first steps of the pathology. The problem that arises here is the early detection of these vocal pathologies, we can distinguish two types of methods of detection of pathologies; subjective and objective.

The subjective methods are those based on specialized medical techniques, but doctors are not always able to detect the pathology before reaching an advanced stage of evolution. Objective methods are the methods based on voice signal processing. In this work, we used a subjective method for the early detection of two pathologies 'dysphonia' and 'laryngitis', using machine learning techniques.

2. Literature Review

Hussein M. A. Mohammed four different machine learning algorithms are used to detect pathology in voice signals. For the classification problem, all the models are trained, validated and tested using continuous voice /a/ sound produced with normal pitch voice obtained from Saarbruecken Voice Data set containing 71 types of pathology. All tests are performed for the entire SVD dataset as well as 1370 equally distributed samples between healthy and pathological classes. These samples are selected making sure that all age groups and all 71 different disease types are included in the balanced (equally distributed) dataset. For balanced data set and whole data set, different sound pathology identification performances of various techniques such as Support Vector Machine, Decision Tree, Bayes Classification, Logistic Model Tree and KNN are compared

João Paulo Teixeira presents the ANN response to parameters extracted by an algorithm developed by15, and an analysis of the best predictors for dysphonia identification. The parameters used as inputs were 4 parameters of jitter, 4 parameters of shimmer and HNR. All 9 parameters for 3 vowels at 3 tones formed an input vector of length 81. First de accuracy for all 81 input variables was analyzed and then the response to models obtained by dimension reduction and variable selection techniques. This allow us to reduce the processing time and improve the accuracy by eliminating redundant information and leave on the best predictors. The results shows a significant improvement by applying the second method (multilinear regression analysis). For the female case an accuracy of 100%, was achieved, and for male case 90%. The first method (hierarchical clustering) and PCA method were capable of getting the same accuracy of an input vector with all parameters. For the male case the PCA get the same accuracy of second method. Analyzing only

S. L. Salzberg It would have been especially interesting to read what Quinlan thinks of the numerous experiments comparing C4.5 and its predecessors to the other classification algorithms that are commonly used in the research community. Even at its current length, though, will be very useful to researchers and practitioners of machine learning. Anyone intending to use C4.5 in their experiments or on a practical problem will find this book an invaluable resource the best results models (second method) there is a preference for Shimmer, vowel /a/ and N tones predictors, for Female group. For the male case the Jitter parameters and vowel /a/ seems to get best response

Sumner, Marc proposed two modifications to the SimpleLogistic algorithm employed by LMT that are designed to improve training time. The use of AIC instead of cross-validation to determine an appropriate number of LogitBoost iterations resulted in a dramatic speedup. It resulted in a small but significant decrease in accuracy in only two cases when performing stand-alone logistic regression. The simple heuristic of weight trimming consistently improved the training time while not affecting accuracy at all. The use of AIC and weight trimming in LMT have resulted in training times up to 55 times faster than the original LMT algorithm while, in most cases, not significantly affecting classification accuracy. These results were measured on datasets of relatively low size and dimensionality. We would expect the speedup to be even greater on larger and high-dimensional datasets

Xu, Zhixiang, Gao Huang s paper introduces GBFS, a novel algorithm for nonlinear feature selection. The algorithm quickly train very accurate classifiers while selecting high quality features. In contrast to most prior work, GBFS is based on a variation of gradient boosting of limited depth trees This

approach has several advantages. It scales naturally to large data sets and it combines learning a powerful classifier and performing feature selection into a single step. It can easily incorporate known feature dependencies, a common setting in biomedical applications medical imaging and computer vision This has the potential to unlock interesting new discoveries in a variety of application domains. From a practitioner's perspective, it is now worth investigating if a data set has inter-feature dependencies that could be provided as additional side-information to the algorithm

Harar, Pavol, et al was to carry out a preliminary study which would clarify whether the use of Deep Neural Network model, especially combination of convolutional and LSTM layers would prove itself worthy of further exploration in case of Voice Pathology Detection problem using only sustained vowel. Using just recordings of vowel /a/ produced at normal pitch, the examined method achieved 71.36 % accuracy on validation data and 68.08 % accuracy on testing data. Since this result is comparable to that published in [6] we assume that further investigation is in place and could lead to much better results.

Verde, Laura, proposed approach detects carefully voice disorders. It identifies with a better accuracy (about 82%) the presence of a voice disorder than the approaches required by the standard medical protocols. Additionally, a comparison with other regression algorithms has been performed, which has confirmed the greater accuracy of our approach in detecting a voice disorder than that achieved by using the other algorithms

Wu, Kebin, et al propose a regression-based model JOLL4R to fuse features from two different audios for voice-based disease detection. Four key factors 5 contribute to the success of JOLL4R. First of all, the fusing model couples the regression losses from two views together to adjust the weights adaptively so that the more discriminative view can be emphasized in the final classification. Secondly, the -dragging technique facilitates to construct flexible regression targets and to enlarge the margins between different classes, by which more 10 discrimination projections can be learned

Wu, Huiyi, et al novel algorithm for pathological voice detection is introduced in this work. Convolutional neural network is shown to effectively extract features from spectrograms of voice recordings and diagnose voice disorders. Convolutional deep belief network helps initialize the weights and make the system more robust. However, a tradeoff must be struck between robustness and accuracy.

Musaed Alhussein and Ghulam Muhammad proposed voice pathology detection system was embedded to the framework to constantly assess the voice condition of a patient. A deep learning in the form of the CNN models was used in the system. Several popular CNN models were investigated. In the experiments on the SVD database, the system achieved 98.77% accuracy with the CaffeNet CNN model followed by the SVM classifier. The result is promising because it outperforms some of the previous results reported in literature.

Laura Verde, Giuseppe De Pietro 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,

Ghulam Muhammad, Mohammed F. Alhamid, Mansoud e proposed system was successfully deployed in the framework. Experimental results show that the proposed system could achieve high detection and classification accuracy. Based on the success of the proposed system

Victor H. Téllez Arrieta, presents how to detect determinism in a time series by a statistical direction measurement of neighboring trajectories. Determinism is detected by identifying whether the TDM value of the original time series is significantly different with that of its surrogates. It should be mentioned that evolvement analogue of neighboring trajectories in a deterministic time series actually.

4. Conclusion

In this paper we propose a new multi-parametric acoustic marker able to evaluate globally the voice quality and detect possible disorders, that can be embedded in a mobile health solution, to monitor and support the correct diagnosis of voice disorders. This marker combines, in fact, data provided by each acoustic parameter considered, such as information about the laryngeal function provided by Fundamental Frequency, cycle-to-cycle instabilities in frequency and amplitude by jitter and shimmer, and the presence of noise due to a voice disorder by HNR, to detect any alterations of voice quality due to a possible disorder of the pneumo phonoarticulatory apparatus. The proposed approach detects carefully voice disorders. It identifies with a better accuracy (about 82%) the presence of a voice disorder than the approaches required by the standard medical protocols. Additionally, a comparison

with other regression algorithms has been performed, which has confirmed the greater accuracy of our approach in detecting a voice disorder than that achieved by using the other algorithms.

References:

1. Hussein M. A. Mohammed," Voice Pathology Classification Using Machine Learning",ISASE2021
2. Smith, J. O. and Abel, J. S., ``Bark and ERB Bilinear Transforms", *IEEE Trans. Speech and Audio Proc*., 7(6):697-708, 1999
3. Lee, K.-F., *Automatic Speech Recognition: The Development of the SPHINX SYSTEM,* Kluwer Academic Publishers, Boston, 1989.
4. Rudnicky, A. I., Polifroni, Thayer, E H., and Brennan, R. A. "Interactive problem solving with speech", *J. Acoust. Soc. Amer*., *Vol. 84, 1988, p S213(A)*
5. Muhammad, Ghulam, et al. "Edge computing with cloud for voice disorder assessment and treatment." IEEE Communications Magazine 56.4 (2018): 60-65
6. Hemmerling, Daria. "Voice pathology distinction using autoassociative neural networks." 2017 25th European signal processing conference (EUSIPCO). IEEE, 2017.
7. Verde, Laura, Giuseppe De Pietro, and Giovanna Sannino. "Voice disorder identification by using machine learning techniques." IEEE Access 6 (2018): 16246-16255
8. Alhussein, Musaed, and Ghulam Muhammad. "Voice pathology detection using deep learning on mobile healthcare framework." IEEE Access 6 (2018): 41034-41041.
9. Wu, Huiyi, et al. "A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks." Interspeech. Vol.
10. Wu, Kebin, et al. "Joint learning for voice based disease detection." Pattern Recognition 87 (2019): 130-139
11. Verde, Laura, et al. "Dysphonia Detection Index (DDI): A New Multi-Parametric Marker to Evaluate Voice Quality." IEEE Access 7 (2019): 55689-55697.
12. Ezzine, Kadria, and Mondher Frikha. "Investigation of glottal flow parameters for voice pathology detection on SVD and MEEI databases." 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE, 2018.
13. Harar, Pavol, et al. "Voice pathology detection using deep learning: a preliminary study." 2017 international conference and workshop on bioinspired intelligence (IWOBI). IEEE, 2017.
14. Xu, Zhixiang, Gao Huang, Kilian Q. Weinberger, and Alice X. Zheng. "Gradient boosted feature selection." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 522-531. ACM, 2014.
15. S B. Schölkopf, C. J. Burges, and A. J. Smola,Advances in kernelmethods: support vector learning.MIT press, 1999.
16. Sumner, Marc, Eibe Frank, and Mark Hall. "Speeding up logistic model tree induction." European conference on principles of data mining and knowledge discovery. Springer, Berlin, Heidelberg, 2005.
17. S. L. Salzberg, "C4. 5: Programs for machine learning by j. ross quinlan.morgan kaufmann publishers, inc., 1993,"Machine Learning, vol. 16,no. 3, pp. 235–240, 1994.
18. Teixeira, João Paulo, Paula Odete Fernandes, and Nuno Alves. "Vocal Acoustic Analysis–Classification of Dysphonic Voices with Artificial Neural Networks." Procedia Computer Science 121 (2017): 19-26