EXPERT JURNAL MANAJEMEN SISTEM INFORMASI DAN TEKNOLOGI

ML-based Smart Voice Analysis for Healthy and Pathological Voice Detection

Kamran Saeed [1], M. Fatih Adak [1*], Khalid Javeed [2], Rizwan Saeed [3], Samiah Ijaz [4]
[1] Department of Computer and Information Engineering, Sakarya University, Sakarya, Turkey
[2] College of Computing and Informatics, University of Sharjah, Sharjah, United Arab Emirates
[3] CENTech, Islamabad, Pakistan
[4] Department of Anesthesiology, Shifa International Hospital, Islamabad, Pakistan
[1] kamran.saeed@ogr.sakarya.edu.tr, [1*] fatihadak@sakarya.edu.tr, [2] kjaveed@sharjah.ac.we,
[3] rizwansaeed22@gmail.com, [4] samiahejaz@hotmail.com

ABSTRACT – Voice pathology is increasing dramatically, especially due to unhealthy social habits, being too much talkative, age factor or some kind of pathology in throat. Normally the people who speak a lot, for example, teachers or announcers etc. suffer from voice pathology in their elderly age. A research-oriented simulation project is required, which will help and assist general physicians to identify pathology and to refer the patient to the specialist ENT doctor. Due to the tremendous complexity of speech signal data, a technique for extracting crucial information for pathology classification is required to guarantee both computational effectiveness and model precision. The main objective is to use feature extraction to identify auditorily identifiable elements in the speech stream and produce a series of low-variation feature vectors. For the disease classification model that follows, this compact representation is essential. This research has two phases: firstly, it extracts frequently used features for voice signals and secondly find the best algorithm to classify pathologies. Commonly used features available for feature extraction of voice pathology classification and detection have been investigated and then trained the system using various machine learning techniques to find best pathology detection and classification rate. Developed an accurate Mel frequency Cepstral Coefficient feature extraction technique for classifying and detecting voice pathologies using a German Saarbrucken voice database. The system is trained by using four different kinds of machine learning algorithms which are Naïve Bayes, Random Forest, Support Vector Machine, K nearest neighbor and find out the best method for categorizing speech pathologies is Random Forest Regression. A disease classification system for voice signals that is both accurate and efficient is mostly dependent on feature extraction. The study emphasizes how important Random Forest Regression is for getting better outcomes than other methods. Random Forest Regression's proven effectiveness raises the possibility of its practical implementation, advancing the diagnosis and treatment of voice disease.

Keywords: K-Nearest Neighbors; Mel Frequency Cepstral Coefficient; Naïve Bayes; Random Forest; Saarbrucken Voice Database; Support Vector Machines; Voice Pathology

1. INTRODUCTION

Voice pathology is increasing dramatically, especially due to unhealthy social habits, being too much talkative, age factor or some kind of pathology in throat. Normally the people who speak a lot for example, teachers or announcers etc. suffer from voice pathology in their elderly age. Automatic Voice Pathology Detection and Classification (AVPDC) is a system that can help general physicians for detecting any presence of voice disorders of patients and the kind of other voice disorders from which the individual has to suffer in the initial stages. In effective and normal communication there is a barrier which can be analyzed [1]. Various disorders can be taken as voice pathology detection and for classification four different voice disorders of vocal folds are common in Pakistan i.e. Cyst, Polyp, Paralysis and Laryngitis.

Suffering from several voice disorders like cyst, paralysis, laryngitis and polyp has increased theatrically, with almost 12 million individuals in the US alone tolerating from vocal difficulties [2]. The effect of voice difficulties in teaching professions is larger than any other professions. Different researches in the US have uncovered that the layout of speech pathologies throughout all the lifetime of individuals is 28.8% for non-teachers and 57.7% for teachers [3]. Moreover, for determining the risk and prevalence for voice difficulties in teachers, a survey was done in Gujranwala region Pakistan. This survey is done using 10 different colleges and schools. Results of survey conclude that occurrence of voice disorders was smaller in male teachers as compared to female teachers. The risk factor was the extensive duration of job was more leading [4]. The use of this noninvasive approaches to classify or detect pathological difficulties in voice has amplified over the time, and for the past few years several researches have been made using the automatic classification and detection of vocal cords diseases.

Furthermore, these are required to be examined due to the nonexistence of ordinary approaches and tools for disorder of voices. The most difficult stage is to diagnose the pathalogical voice, classification of disorder and appropriately control voice pathology. The bjective evaluation, which include acoustical examination, is self-determined of human involvement and it can give assistance to doctors in making the right choices. Doctors have the concluding decision concerning about the medical analysis, and an objective evaluation can only be used for the assistance instrument. However, the subjective evaluation of speech feature depends upon the experience of human and can differ it from one human to another. For performing automatic voice disorder classification and detection there are different forms of signal analysis such that short term and long-term signal analysis.

People are increasingly at hazard of voice disorder problems. There are about 25% of the population of world in which individuals professions forces the person to speak extremely louder than the normal level usually suffers from voice disorder diseases [6]. For example, auctioneers. singers, lawyers, actors, aerobics instructors, teachers and industrial administrators are all required to work in occupations that make force to use their voice. As a significance, digital process of working on voice signals has originated to deliver a noninterfering analytical method that assumed to be an effective helping equipment in medical practices when classifying and detecting voice diseases, mainly in their early phases. Voice disorders disturb the vocal cords during the process of phonation. Due to breakdown of different kinds of factors that contribute to voiced vibrations they make the vocal folds to produce irregular vibration. Vocal folds are affected by voice disorder pathologies which results in variation of vibratory series of vocal folds, as their capability to close appropriately is lessened. It also affects the shape or form of the throat vocal tract area and yield asymmetry in spectral properties [7]. Furthermore, voice pathology affects in a different way to vocal fold vibration depending on the location and type of disorders in the vocal folds and it will produce different type of voice tones. The vibration of vocal folds depends on some different factors. For example, the mucus that is existing on the stiffness, muscles in the larynx, tension, vocal folds tissue and the opening & closing of the folds. These types of issues are affected in a different way for several voice disorders. Because of the size and position of voice disorders, vocal folds do not close in a normal way because of vibration. So, these vibrations differ from one to another type of disorder. These vibrations give glottal source frequency excitation and it also affect the below part of vocal tract area which then subsidize the output frequency of speech signal.

With the passage of time voice pathology cases is growing in the people who have to speak a lot in their elderly age. There are fewer common methods and instruments available for the identification and categorization of vocal abnormalities, despite their increasing occurrence. Prior studies have investigated noninvasive techniques with an emphasis on automatic vocal cord illness identification and categorization. But further investigation is necessary due to the incomplete knowledge of the foundation of these methods. By focusing on the in-depth analysis of Automatic Voice Pathology Detection and Classification, this work aims to fill this gap. It expands on earlier research and acknowledges the value of impartial assessments, especially acoustical exams, in supporting doctors in making reasonable decisions. The crucial need for trustworthy diagnostic systems is highlighted by the lack of common methods and instruments for the examination of vocal disorders. Through the exploration of signal analysis, including both short- and long-term signal analysis, this study seeks to advance our knowledge of voice problems and aid in the creation of useful diagnostic instruments. Acknowledging the impact of multiple factors, such as mucus, muscular tension, and tissue state, on vocal fold vibrations, the research aims to distinguish and categorize individual voice illnesses according to their unique vibratory patterns. This study is presented as an important first step in creating a simulation project that will help general practitioners diagnose voice diseases more accurately. This study aims to lay the groundwork for an accurate and effective AVPDC system by carefully examining the current environment and incorporating knowledge from earlier studies.

Detection and Classification Overview

The classification and detection of voice pathology through stroboscopic and Laryngoscopy is a subjective way of visualizing different pathologies of vocal folds which is depend upon the clinical expertise of specialists and there is still a human error. The performance of these equipment's is not very perfect. Voice pathology classification and detection is the job of an ENT specialists. In order to detection numerous equipment is used and classify these pathologies. These kinds of pathologies test are very expensive, time consuming and hard to use.

To overcome this problem, large number of machine learning structures have been designed by the scholars for the classification and detection of voice pathology of vocal fold. These systems are Automatic voice disorder classification and detection systems that are used by general doctors for early analysis. Automatic voice pathology classification and detection structures are noninvasive in real and very easy in use.

Vocal folds are affected by voice pathology, that produce uneven vibrations because of the malfunctioning of many factors including vocal vibrations. Due to incomplete closure of vocal folds' voice pathology show variations in the vibratory cycles. Voice disorder is also affect the vocal tract shape and there are irregularities in spectral properties. Depending upon the kind of voice disorder and the disorder location in vocal fold, voice pathology damages the vocal fold andto produce different tones on different frequencies. As most of the diseases are caused by the vocal folds vibration so structure of human voice processing system need to be understand. Internal structure of vocal folds is shown in Figure 1.

In this research, four kinds of different diseases which are common in most of the societies. They are named as Cyst, Polyp, Paralysis and Laryngitis. In [5] researchers explained that during speech production airflow vortex is generated. Speech production is also related with nonlinear part and linear plane wave in vortex region. During speech production linear airflow is spread in vocal tract. When the airflow connects with vocal tracts wall, it produces turbulence and vortex. Vocal folds close quickly when the vortex achieved negative pressure close to the vocal tract. The vortex changes the energy range of voice signal. During phonation, the source of excitation is produced by vortices nearby the false vocal folds. NonLinear flow for speech production is illustrated in Figure 2.

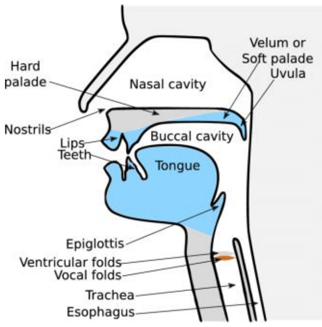

Figure 1. Vocal Fold Internal Structure

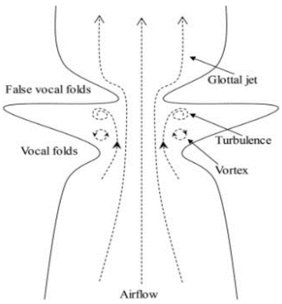

Figure 2. Non-Linear flow for Speech Production

The database of massachusetts eye and ear infirmary (MEEI) has been used by many researchers for classification and detection of voice pathology. For detection it gives good results but it was not clear whether the classifier is classifying the environment or the pathologies in classification machine learning technique. Saarbrucken Voice Dataset (SVD) recorded has been used which is prepared by Institute of Phonetics of Saarland University (IPSU). In this dataset, sustained vowels /a/, /i/ and /u/ pronounced low, normal, high and high-low intonations are available, which is complete set to perform experiments and it is freely downloadable. In this research work, /a/ sound is used, which is simple to pronounce for pathology individual and its peaks are prominent. There are no any past conclusions for voice pathology classification and detection have been available on this database with the given above four diseases and with this work.

The MFCC plays same like the system of human hearing, wherever the internal part of ear of human plays a vital role in differentiating the frequency regions. Survey of literature review shows that different researches uses different diseases and different machine learning algorithm's with SVD dataset and they get dissimilar accuracies.

Background

The recent spread of advanced innovations in digital technologies, with the implementation of some algorithms through telephone systems, detection of pathological voices is possible [8]. With the help of machine learning technologies and wireless communication detection of pathological voice is possible through biomedical instruments [9], [10]. Signals of pathological voice are frequently non-fixed and multi-recurrence and the sounds give significant biomedical understanding. The voice signals can be captured from wearable sensors like microphones, contact sensors. By getting signal information through these sensors it is easy to detect pathological voice through advance signal processing algorithms.

There are four steps for signal processing such as classification, feature extraction and analysis. This [11] research work focused on the extraction of feature. To check the regularity of pathological voice two linear features were proposed such as shimmer and jitter. Shimmer gives the information related to amplitude variation between the two adjacent vibration cycles. Jitter is almost similar to shimmer which gives information related to instability of the vocal folds vibration frequency. Jitter and shimmer totally depended on the fundamental frequency estemation which is very difficult to find out accurate due to irregularity of voices. The distribution technique of cepstral peak prominence smoothed (CPPS) [12] provide better performance in differentiate in between pathological and healthy voice same like spectrogram features [13]. Some other researchers works on the estimation of glottal flow by inverse filtering [14], [15]. In the meantime, some researchers have applied deep learning methods to detect voice disorders, like most famous technique convolutional Neural Network (CNN) [16], [17], [18]. Spectrograms of voices or mel-spectrograms taken as the input to train CNN. Moreover, some other researchers used some other deep learning methods to detect the pathological voice like such as Recurrent Neural Network [19] and Deep Belief Networks [20].

Proposed works within the field of voice disorder classification and detection started early in 1980s. At that time the pattern recognition and machine learning techniques weren't introduced. After 1995, a correct research work started on automatic voice pathology classification system. Many researchers proposed their system using SVD, MEEI and native dataset. In [21] researcher used database of King Abdul Aziz University Hospital in Riyadh. He used extraction technique which is Mel Frequency Cepstral Coefficient (MFCC) features. He obtained differing types of voice disorders and acquire an accuracy of 91.66%. [22] used Vocal Tract Area feature extraction technique using the identical SVD database obtained an accuracy of 94.7%. He uses Support Vector Machine Learning Algorithms. [23] implemented Naive Bayesian Network machine learning algorithm using Paralysis voice disorder type and acquire a result of 90% with SVD German dataset. This survey shows us that it should require to see more into recommendations for testing the feature vector that perhaps perform well in AVPDS. [24] used three kinds of datasets with interlaced feature extraction technique and acquired a result of 88.5% for detection and 90.3% for classification using SVM classifier. [25] take out MFCC and its delta features. He has achieved results with the accuracy of 86%. [26] from King Saud University used MEEI dataset with four ML algorithms using different voice disorders and acquire a result of 72%, 99.91%, 99.78% and 93.72%. [27] used SVD data set of 685 pathological voices and 685 healthy voices and obtained an accuracy of 86%. [28] used SVD and MEEI dataset with SVM classifier and obtained accuracy of 92.79% and 99.69% for detection. Different experimenters have been using different features and compared their outcomes with other proposed work. Not any one of the experimenters achieved the exact similar accuracy for VPDS (Voice pathology detection system). Instead, their outcomes with the feature MFCC differ. MFCC features extraction technique has produced different detection and classification results, which demonstrate that MFCC are very fine enough; however, excellent results can be produced using acceptable machine learning algorithms. After the survey of these literature reviews, it is notifying that experimenters whom used features as MFCC and nearly every one of them testified different voice disorder classification and detection accuracies. This conclude that the features are the similar but modelling methods produce different results. The overall research flowchart is given in Figure 3.

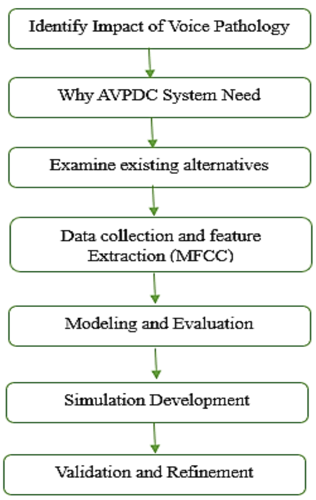

Figure 3. Research Flowchart

2. RESEARCH METHODOLOGY

Extracting of features is the method of maintaining useful data of the voice signal while neglecting unwanted and redundant information. It is the parameterization of the voice signal. This is projected to make a perceptually significant representation of the voice signal. It involves the process of altering the voice signal into a digital form by determining some of the important variables/characters of the voice signal for example frequency response or energy. It also includes altering the voice signal into such kind of form that is appropriate for our model used for classification of voice pathologies. In this research the given input information is altered into a set of features that provides the related information for executing a certain task without the want of the full-size dataset but using the compact dataset. The main goal of extraction of features is to untangle the voice signal into dissimilar acoustically recognizable components & for obtaining that set of features that has low rate of variation in order to keep the calculation feasible. The outcome of the feature extraction technique process is the arrangement of feature vectors. In pathological voice classification system, the feature extraction method is the main aim to calculate a sequence of feature vectors that provide a compacted signal representstion which is given. The research has two parts. First part is Feature Extraction, and commonly used feature that are used for voice is extracted. Second part is modeling, which is the part of machine learning, one or two different modeling techniques just to make a model for pathology classification for four diseases are used. In the research, only one model for voice pathology classification is proposed. Overall proposed method is show in Figure 4.

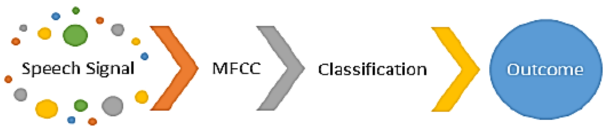

Figure 4. Proposed Method

AVPC Systems

With the help of machine learning algorithms automatic voice pathology classification system of voice pathologies can be developed and designed. Before applying machine learning techniques preprocessing and transformation on features are applied. Following are point that are required to perform. (1) To label the four diseases data by the researchers manually, MP3 format is required for the available speeches; (2) Convert the audio information further subdivided into short frames, featured are extracted from each frame. (3) Features that are extracted by previous step to be considered as input that are further analyzed by the machine learning approaches.

Split the data into two sections: training and testing for detection of pathological voices. Training samples are used to make model much able to identify the unknown samples and testing data is used to calculate the accuracy of the trained model. In whole process accuracy of the model is calculated and for evaluating the performance of automatic voice pathology classification system calculated accuracy is taken as metric. Formula 1 equations is used to check the accuracy analysis.

$$Accuracy = {TP + TN}/{TP + TN + FP + FN}$$

(Formula 1)

The SVD Voice Disorder Dataset

In the research work, a dataset of voice samples from the Saarbrucken Voice Database (SVD) is taken. SVD dataset contains of 2302 voice recordings which includes healthy and pathological individuals voices.It has been published by Institute of Phonetics of the University of Saarland. Soundtracks contains resolution of 16-bit and sampled at 50 kHz. This study consists of soundtracks of vowels /a/. In total, 408 pathological voice samples of four different kinds of diseases of random ages female and male are carried out. Table 1 shows additional details of the selected soundtracks, in which number of considered voices for each one of disease is shown.

Table 1. Dataset

No	Voice Disorder	No. of Voice Sample
1	Cyst	8
2	Paralysis	215
3	Polyp	45
4	laryngitis	140
	Total	408

Carefully dividing the dataset is necessary to make the creation and assessment of machine learning models easier. A split ratio of 70% for training, 15% for validation, and 15% for testing is a standard procedure in this area. There would be 1611 training samples, 345 validation samples, and 346 testing samples after this split. This helps preserve the integrity of the dataset throughout model building and evaluation by ensuring a representative distribution of the 2302 voice recordings among the training, validation, and testing sets.

Features Used for Classification: Mel Frequency Cepstral Coefficient (MFCC)

Detection of noise, speaker identification, detection, and classification of speeches MFCC is mainly used to get desire features. Coefficients that are generated by MFCC may assist for vocal tract analysis irrespective of the vocal fold that could be suffered by pathological voice. In this research analysis is computed over 13 coefficients. By calculating log compression of the /a/ vowels and Discrete Cosine Transform (DCT) over the cepstral of the speech in frequency domain. Figure 5 shows Block diagram of implemented stages.

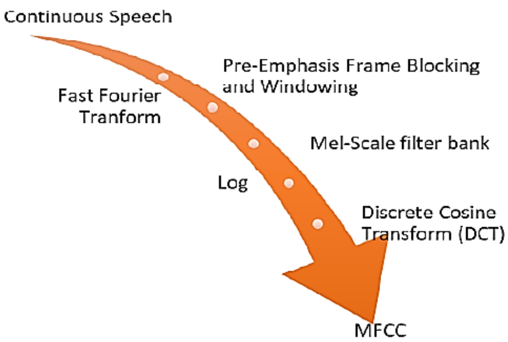

Figure 5. Block Diagram

There are various steps involved in the MFCC computation for speech/speaker recognition: (1) Pre-emphasis: To highlight high-frequency components in the speech stream, a high-pass filter is employed. The original and after pre-emphasis wave plotting is given in Figure 6.

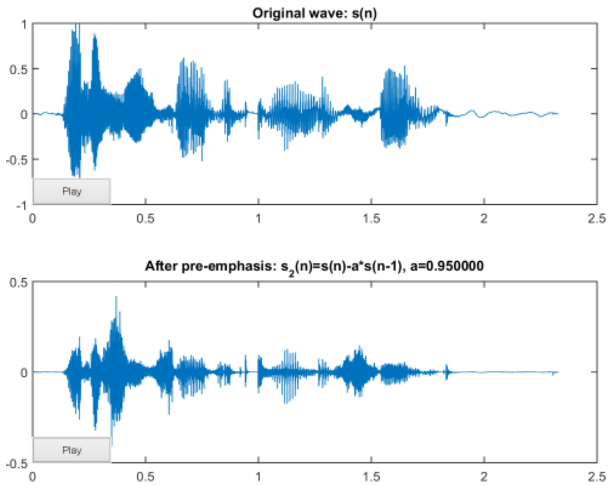

Figure 6. Original and After Pre-Emphasis Signal Plotting

(2) Frame blocking: The signal is divided into frames, which might overlap and are usually 20–30 ms in length. For FFT, frames are frequently power-of-two lengths. (3) Hamming windowing: To preserve continuity at frame boundaries, a Hamming window is multiplied by each frame. Genrelized hamming window is given in Figure 7.

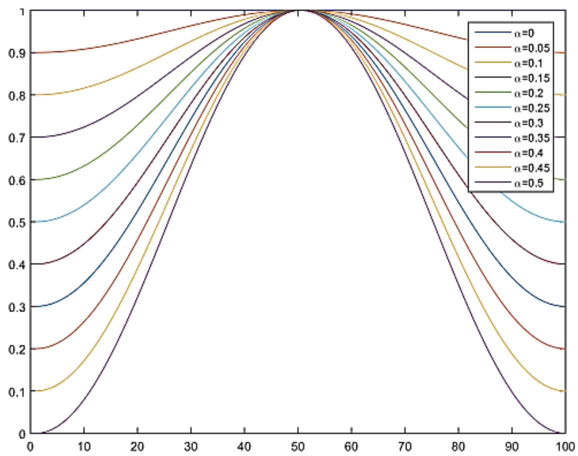

Figure 7. Generalized Hamming Window Plotting

(4) Fast Fourier Transform (FFT): The magnitude frequency response of each frame is obtained by the application of FFT in spectral analysis. (5) Triangular Bandpass Filters: To obtain the log energy of various frequency bands, the magnitude spectrum is multiplied by a collection of triangular bandpass filters. (6) Discrete Cosine Transform (DCT): Mel-scale cepstral coefficients (MFCCs) are generated by applying DCT to the log energy values. Usually, twelve coefficients are employed. (7) Log Energy: A frame's energy is calculated and frequently added as a new feature. Figure 8.

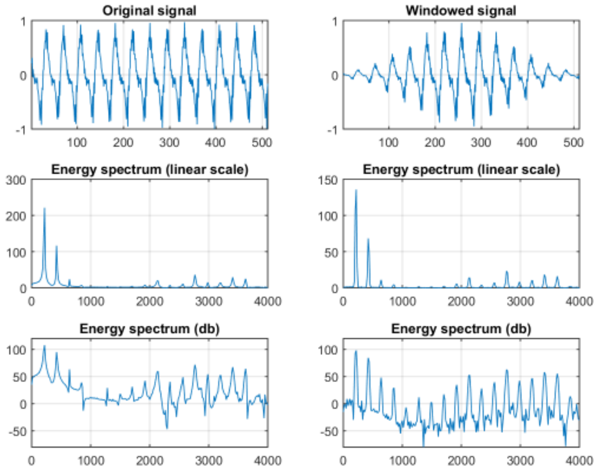

Figure 8. Original and windowed signal plotting

(8) Delta Cepstrum: Features of acceleration and velocity can be represented by computing the time derivatives of the energy and MFCCs. Figure 9.

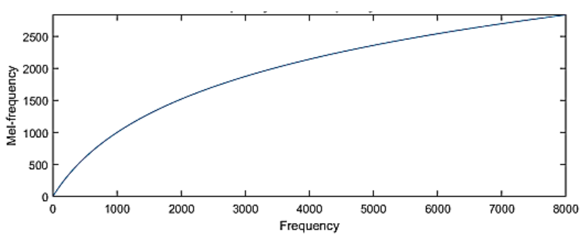

Figure 9. Frequency to Mel-frequency curve

Together, these phases make up the MFCC feature extraction procedure, which is popularly used in speech and speaker identification systems and is adapted to human auditory perception.

Machine Learning Classifier:
Naive Bayes classifier

Machine learning most rich algorithm Naïve Bayes is based on conditional probability. Object that is under consideration is allocated a class from predetermined sets on the different value of descriptive attributes. After determining the conditional probabilities that have high probability it calculates the most frequent. Bayesian belief probability is applied in NBC for detection and classification. Conditional probability of each attribute is calculated, given that the labeled class from the data. Byes rule applied over the probability that attribute with all features which are predicted by the highest rise posterior probability. It is assumed in Naïve Bayes value of predictor over labeled class is independent of other predictors and this is known as class conditional independence. Formula 2 is given below which is used in this algorithm.

$$P(c|x) = {}^{p(x|c)p(c)}\!/_{p(x)}$$

(Formula 2)

Random Forest Classifier

Random forest requires limited parameters and it is known for its processing speed that why it has elevated in interest from couples of years. It looks like collection of trees that train model and get accuracies by the voting process.

Correlation may influence the results of the final map, but accuracy and processing speed are the two main asserts of this algorithm. N-tree bootstraps samples from the given data are withdrew by the random forest and m-tree of the predictors are randomly selected by the index at each of the nodes, such as information gain and GINI. One of the best splits is selected, amongst all the variables. At last majority of the votes from all of trees produced the detection and classification rate. For conducting the experiments 10 folds were used. In the given equation of random forest, N are the data points, value that is returned by the model is noted as Fi and Yi are the actual value for the given data. Formula 3 is given below which is used in this algorithm.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2$$

(Formula 3)

Support Vector Machine (SVM) Classifier

Hyper plane or linear classifier is used in Support vector machine to separate the given data and known as linear binary classifier. Though, Maximus separation is achieved by one of them. In SVM, for maximizing the margin between two of each classes a separating hyperplane is selected. Orientation of hyperplane which is split up over the smallest line that is separating the convex hulls of the training data must be perpendicular for each class.

K-Nearest Neighbor (KNN)

KNN focused on feature similarity which make is lazy and it assumes that similar data are exiting in neighbors or you can say thing that are closer to each other are present in its surroundings. Classification is carried out by voting system and that object assigned the label that wins the majority vote for its neighborhood by the factor K. Formula 4 is given below which is used in this algorithm.

$$\sqrt{\sum_{i=1}^{K}(xi - yi)^2}$$

(Formula 4)

3. RESULTS AND DISCUSSION

For evaluating the automatic voice pathology classification, Saarbrucken Voice Database (SVD) is used. In order to make all samples sampling frequency same, its recordings are sampled down before extracting the features from 50 kHz to 25 kHz. The duration of voice pathologies is one second because of pathologies. All the samples of voice pathologies cyst, Paralysis, Polyp and laryngitis. In the proposed research four different types of machine learning algorithms have been used and two different kinds of experimentations have been done. As defined before, only MFCC features are extracted. First 13 MFCC features for conducting our experiments carried out. In first type, comparison of four machine learning algorithms are conducted. The second type of experiment is also conducted using first 13 MFCC features but there is comparison with previous research work. Naïve Bayes, Support Vector Machine, Random Forest and KNN machine learning algorithms are used in this experiment. Both types of experiment are conducted using K-fold cross validation. 10-fold cross validation approach is used for these four-disease classifications. In the k-fold, data is randomly divided into 10 groups. In each of its iteration, as a training nine groups is used while for testing remaining groups were used. All the groups are tested after the 10 iterations.

The first type of experiment is conducted using 10 k-fold cross validation with four machine learning algorithms having first 13 MFCC feature. The results are shown below in Table 2.

Table 2. 13 MFCC Accuracy

No	ML Techniques	Accuracy (%)	ROC Area
1	Naïve Bayes	66.419%	0.834
2	SVM	65.8364 %	0.774
3	KNN	99.8968 %	1.00
4	Random Forest	100%	1.00

The confusion matrix is shown below in Table 3 so that the effectiveness of the classification algorithms may be evaluated. A thorough analysis of each algorithm's prediction accuracy, including true positive (TP), false positive (FP), true negative (TN), and false negative (FN) numbers, can be found in the confusion matrix. The outcomes of the Naïve Bayes, SVM, KNN, and Random Forest algorithms that were used to classify the dataset form the basis of the findings. The percentages indicated by the values in the table are in relation to the total number of positive and negative cases in the dataset. It should be noted that the placeholders in the computations should be replaced with the real total positive (Total Pos) and total negative (Total Neg) instances in the dataset.

According to the available studies most of the researchers had not discuss about time and speed i.e., how much an algorithm is taking time for achieving the results. If someone had to implement it on real world application this will be an important point. Accuracy of Naïve Bayes is low for feasible value of the attributes, but it is efficient in time complexity and take less time to get desire results. Uncertainties in numeric values and worst results could be occurred as its outcome in probability fashion ranging from zero to one. As SVM classifier is also have low accuracy classification rate. From previous work SVM only performs well when model contains a two or three class problem. It does not perform well if it is more than three class problem. Our work contains model of four class problem that is why it is not performing well. Moreover, SVM classifier has not much feature space for this dataset. It gives good performance and results if data or features are transfer to another space. Figure 10 shows th graphical representation of comparison of different algorithms.

Table 3. Confusion Matrix

No	ML Techniques	TP	FN	FP	TN
1	Naïve Bayes	189	42	59	56
2	SVM	156	72	92	26
3	KNN	248	97	0	1
4	Random Forest	248	98	0	0

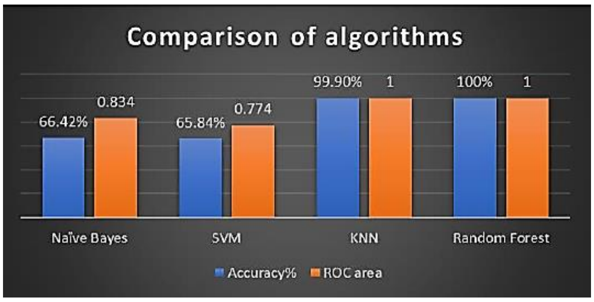

Figure 10. Classification accuracy of algorithms

For KNN classifier, k=5 value is used. As it is slow and lazy learner, so it takes much more time than any of the above algorithms. Furthermore, as in the table from KNN algorithms, accuracy of 99.89% is achieved and its ROC is also one. It is clearly observed that automatic voice pathology classification rate is obtained a very good accuracy for Random forest algorithm. Random forest performed better as compared to the other ML algorithms and attained 100% accuracy result and AUC is also one. These results demonstrate that 100% voice pathologies were classified. By achieving this came to know that if right features with appropriate ML techniques are used, outstanding results could be achieved.

Table 4. Accuracy of 13 MFCC

No	Approaches	Accuracy%
1	Muhammad et al [10]	90.3%
2	Muhammad et al [11]	94.70%
3	Ahmed Al-Nasheri et al [7]	99.53%
4	Godino Llorente et al. [5]	86%
5	Proposed model	100%

Table 2 and Table 4 shows the differentiation of proposed system with previously available results. Classification of pathological voice could be seen how the system performs best in their results. Figure 11 shows that how our results improved with RF algorithm of our proposed model.

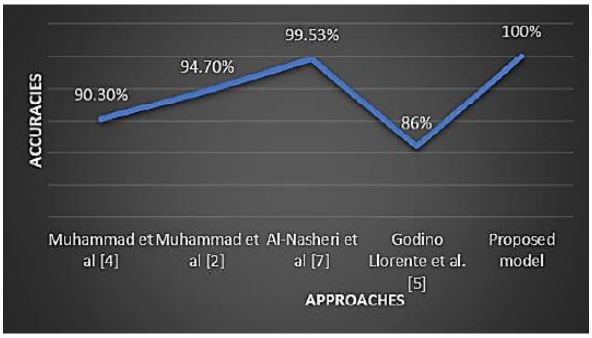

Figure 11. Comparison Graph of Previous Research Work

In Figure 12 ROC curve is shown, while comparing four of our machine learning algorithms it is clearly seen that highest detection rate is observed using MFCC features with random forest algorithm. X-axis indicates True positive and y-axis is false negatives.

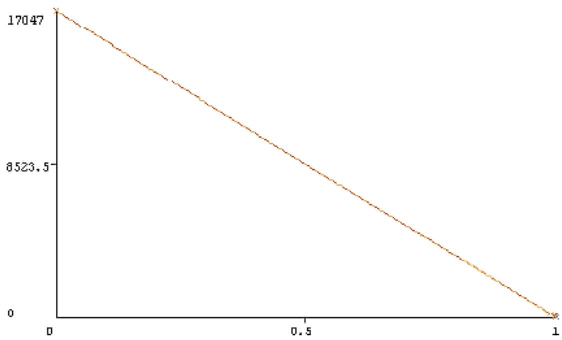

Figure 12. ROC Curve of Random Forest

Using the Saarbrucken Voice Database (SVD) and the Naïve Bayes, SVM, KNN, and Random Forest algorithms, the research findings offer significant novel insights into the categorization of voice disorders. Interestingly, Random Forest's 100% accuracy rate demonstrated how well it could differentiate between normal and pathological voice samples. While KNN showed amazing accuracy at 99.8968%, Naïve Bayes and SVM showed respectable accuracy levels of 66.419% and 65.8364%, respectively. These findings highlight the effectiveness of machine learning methods in the classification of vocal pathologies, with Random Forest and KNN showing great promise. A comparison with previous research is necessary to put our findings into perspective. Although the approaches and datasets used in speech pathology classification studies differ, our results are consistent with newer studies showing the value of machine learning in this field. To avoid discrepancies in stated accuracies, it is necessary to recognize variances in datasets, preprocessing methods, and assessment measures. Standardized benchmarks should be the main focus of future research in order to enable more insightful comparisons. Our study's findings add to the body of knowledge by highlighting the advantages and disadvantages of various machine learning techniques for classifying voice pathologies. The increasing use of ensemble methods and non-parametric techniques in healthcare-related applications is consistent with the performance of Random Forest and KNN. These findings highlight the potential of various machine learning approaches to improve voice disorder diagnostic tools. It is critical to recognize the constraints this research has to work within. The Saarbrucken Voice Database is the basis for the study, and its unique features may have an impact on how broadly applicable the results are. In addition, the robustness of the models is challenged by possible imbalances and biases in the dataset in addition to differences in recording settings. To improve the results' external validity, more varied datasets should be investigated in future studies. It is suggested for other researchers to investigate machine learning algorithms and feature extraction approaches in order to further progress this discipline. In addition, to improve the generalizability of models, efforts should be undertaken to collect larger and more diverse datasets. Researchers from different universities working together can make it easier to create benchmarks and standardized datasets, which will further our understanding of voice pathology classification.

4. CONCLUSION

In conclusion, this study has significantly advanced the field of voice pathology categorization and detection. Mel Frequency Cepstral Coefficients (MFCC) were the main focus of the study, and a variety of machine learning techniques were utilized to obtain more accuracy than previous efforts, especially when using the Saarbrucken Voice Database (SVD). The study makes a contribution by employing MFCC characteristics and demonstrating how well they work to improve classification accuracy for pathlogical voices. The outcomes outperform a large number of earlier research that made use of the SVD database with MFCC features. In particular, the Random Forest method showed outstanding effectiveness in identifying the four listed diseases with 100% accuracy. This highlights the effectiveness and promise of ensemble approaches for the categorization of voice pathologies. By increasing the scope to include the identification of more than four disorders within the field of vocal pathology, future research initiatives can build upon this work. The study provides the groundwork for more extensive research into the use of machine learning in various pathological voice problems.The real-world implications of this research's findings extend to the creation of precise diagnostic instruments for vocal disorder. The suggested model offers a viable method for practical uses in healthcare and medical diagnostics by combining MFCC features with appropriate machine learning techniques.It is possible to make improvements to the study by acknowledging its inherent limitations, which include biases particular to the dataset and possible variations in the recording settings. These constraints might be overcome in later research by adding more varied datasets and improving the models correspondingly. In order to improve the accuracy and resilience of speech pathology detection models, it is advised that future research look into new features and machine learning methods. Standardized benchmarks and collaboration could advance the field.

BIBLIOGRAPHY

[1] R. Islam, M. Tarique, and E. A. Raheem, "A survey on signal processing based pathological voice detection techniques," *IEEE Access*, vol. 8, pp. 99, 2020. DOI: 10.1109/ACCESS.2020.2985280

[2] "Acute Asthma, Prognosis and Treatment," World Allergy Organization, July 2021. [Online]. Available: https://www.worldallergy.org/education-and-programs/education/allergic-disease-resource-center/professionals/acute-asthma [Accessed March 2023].

[3] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray and E. M. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech, Lang., Hearing Res.,* vol. 47, pp. 281-293, 2004. DOI: 10.1044/1092-4388(2004/023)

[4] A. Rehman, S. Arif, H. M. Hayat, A. Kamran and S. Shakeel "Prevalence and Risk Factors for Occupational Voice problems in Teachers," *Asian Journal of Allied Health Sciences (AJAHS)*, vol.2(2), pp. 33-36, 2020. DOI: 10.1159/000089610

[5] N. Shokouhi and J. Hansen, "Teager–kaiser energy operators for overlapped speech detection," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 25(5), pp. 1035-1047, 2017. DOI: 10.1109/TASLP.2017.2678684

[6] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K.H. Malki, A. T. Mesallam and M. F. Ibrahim, "Voice pathology detection and classification using auto correlation and entropy features in different frequency regions," *IEEE Access*, vol. 6, pp. 6961-6974, 2018. DOI: 10.1109/ACCESS.2017.2696056

[7] G. Muhammad, G. Altuwaijri, M. Alsulaiman, T. Mesallam, K. Malki, G. Altuwaijri, Z. Ali M. Farhat and A. Al-nasheri "Automatic voice pathology detection and classification using vocal tract area irregularity," *Journal of Applied Biomedicine*, vol. 36(2), 2016. DOI: 10.1016/j.bbe.2016.01.004

[8] L. Cai and J Zhao, "Speech quality evaluation: A new application of digital watermarking," *IEEE Instrumentation and Measurement Technology Conference Proceedings,* vol. 56(1), pp. 45-55, 2007. DOI: 10.1109/IMTC.2005.1604213

[9] M. Ghulam, S. Rahman, A. Alelaiwi and A. Alamri, "Smart health solution integrating iot and cloud: A case study of voice pathology monitoring," *IEEE Commun Mag,* vol. 55(1), pp. 69-73, 2017. DOI: 10.1109/ MCOM.2017. 1600425CM

[10] M.S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: a voice pathology detection paradigm for smart cities," *Multimed Syst*, vol. 25(5), pp. 565-575, 2019. DOI: 10.1007/s00530-017-0561-x

[11] A. Lovato, M. R. Barillari, L. Giacomelli, L. Gamberini and C. D. Filippis "Predicting the outcome of unilateral vocal fold paralysis: a multivariate discriminating model including grade of dysphonia, jitter, shimmer, and voice handicap index-10," *Ann Otol Rhinol Laryngol*, vol. 128(5), pp. 447-452, 2019. DOI: 10.1177/0003489419826597

[12] A. Castellana, A. Carullo, S. Corbellini and A. Astolfi, "Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel," *IEEE Trans Instrum Meas*, vol. 67(3), pp. 646-654, 2018. DOI: 10.1109/TIM.2017. 2781958

[13] Z. Changwei, Z. Lili, Z. Xiaojun, W. Yuanbo, W. Di and T. Zhi "Classification of normal and pathological voices using convolutional neural network.," *2020 International conference on sensing, measurement & data analytics in the era of artificial intelligence (ICSMD)*, pp.325-329, 2020. DOI: 10.1109/ICSMD50554.2020.9261730

[14] S.R. Kadiri, and P. Alku, "Analysis and detection of pathological voice using glottal source features," *IEEE J Select Top Signal Process*, vol. 14(2), pp. 367-379, 2020. DOI: 10.1109/JSTSP.2019.2957988

[15] K. Yokota, Y. Koba, S. Ishikawa and S.Kijimoto, "Inverse analysis of vocal sound source using an analytical model of the vocal tract," *Appl Acoust*, vol. 150, pp. 89-103, 2019. DOI: 10.1016/j.apacoust.2019.02.005

[16] T. Tuncer, S. Dogan and F. Ertam "Automatic voice-based disease detection method using one dimensional local binary pattern feature extraction network," *Appl Acous*, vol. 155, pp. 500-506, 2019. DOI: 10.1016/j.apacoust. 2019.05.023

[17] S. Souli, R. Amami and S. B. Yahia "A robust pathological voices recognition system based on dcnn and scattering transform," *Appl Acoust*, vol. 177, p. Article 107854, 2021. DOI: 10.1016/j.apacoust.2020.107854

[18] S. Fujimura, T. Kojima, Y. Okanoue, K. Shoji, M. Inoue, K. Omori, and R. Hori, "Classification of voice disorders using a one-dimensional convolutional neural network," *J Voice*, vol. 36(1), pp. 15-20, 2022. DOI: 10.1016/j.jvoice.2020.02.009

[19] S. Hidaka, K. Wakamiya, Y. Lee and T. Nakagawa "Automatic estimation of pathological voice quality based on recurrent neural network using amplitude and phase spectrogram," *Proc. Interspeech* 2020, pp. 3880-3884, 2020. DOI: 10.21437/Interspeech.2020-3228

[20] H. Wu, J. J. Soraghan, A. Lowit and G. D. Caterina "A deep learning method for pathological voice detection using convolutional deep belief networks.," *Interspeech 2018*, 2018. DOI: 10.21437/Interspeech.2018-1351

[21] Z. Ali, M. Alsulaiman, G. Muhammad and I. Elamvazuthi "Vocal fold disorder detection based on continuous speech by using MFCC and GMM," *2013 7TH IEEE GCC conference and Exhibition (GCC)*. IEEE, 2013. DOI: 10.1109/IEEEGCC.2013.6705792

[22] G. Muhammad, G. Altuwaijri, M. Alsulaiman and Z. Ali "Automatic voice pathology detection and classification using vocal tract area irregularity.," *Biocybern Biomed Eng.*, vol. 36, 2016. DOI: 10.1016/j.bbe.2016.01.004

[23] M. Dahmani, and M. Guerti, "Vocal fold Pathologies classification using Naïve Bayesian Networks.," *6th International Conference on system and control (ICSC)*. IEEE, 2017. DOI: 10.1109/ICoSC.2017.7958686

[24] G. Muhammad, M. Alsulaiman, Z. Ali and T. Mesallam, "Voice Pathology Detection using interlaced derivative pattern on global source excitation," *Biomed Signal Process Control*, 2017. DOI: 10.1016/j.bspc.2016.08.002

[25] J. D. Arias-Londoño, J. I. Godino-Llorente, M. Markaki and Y. Stylianou "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices," *Logoped Phoniatr Vocol.*, 2011. DOI: 10.3109/14015439.2010.528788

[26] A. Mahmood, "A Solution to the Security Authentication Problem in Smart Houses Based on Speech," *PhD Thesis, King Saud University, Riyadh*, 2019. DOI: 10.1016/j.procs.2019.08.085

[27] L. Verde, G. D. Pietro, and G. Sannino, "Voice Disorder Identification by Using Machine Learning Techniques," I*EEE Access*, vol. 6(1), pp. 6246-16255., 2018. DOI: 10.1109/ACCESS.2018.2816338

[28] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K.H. Malki, A. T. Mesallam and M. F. Ibrahim, "Voice pathology detection and classification using auto correlation and entropy features in different frequency regions," *IEEE Access*, vol. 6, pp. 6961-6974, 2018. DOI: 10.1109/ACCESS.2017.2696056