Voice Pathology Detection Based on Deep Neural Network Approach

Jeberson Retna Raj, J Jabez, S Senduru Srinivasulu, S Gowri, and J S Vimali
Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai, India

Abstract. The advancement of technology offers solution to the complex problems faced by the society and brings the wellbeing of the individuals. Smart healthcare is prominent nowadays for diagnosis, treatment and constant monitoring which reduces visitation of hospital, transport cost and waiting time. Voice pathology is a decease which affects the person vocal cord so that one who facing difficult in speech. If the decease not identified in time, it leads to permanent loss of voice for an individual. Traditionally, the decease is identified through oral examination or manual procedures. Due to the advent of smart phone, one can record the voice and send it to the cloud server for processing. Our system classifies the voice data and provides the decision to the user. This greatly reduces the transport cost, and waiting time for oral examination at medical center. The mobile phone recorded the patient voice data and it will be stored into the cloud. The voice data is synthesized to signals and with the help of deep Neural network the voice pathology can be identified. The system has been tested with the data set and the test result shows promising.

1. Introduction
Integration of wireless network with cloud technology allows seamless interaction of applications and message exchange. Mobile healthcare is famous due to its application helps in diagnosis of deceases, training the medical practitioners, testing the deceases with mobile based applications. The health industry and the products are reaching out the customer based on the credibility and accuracy of solving the patient problem. Smart health care is imperative for countries and smart cities integrate the patient, medical attendant and physician. This will highly reduce the patient overhead by visitation of hospital manually. Early diagnosis, preventive advices can be remotely handled with the available technology. Voice pathology is one of the decease affected worldwide. Survey shows that over 7 million people affected in US for voice pathology decease. The consequence of the decease is broken voice, change in voice pattern, and degradation of voice. The decease affected by the small growth of fuss, tumor in the path of vocal cord. Prompt and timely identification and treatment is required. Otherwise, one who lose the voice and in the worst case operation is necessitated. The decease may common for those who involves in teaching, music and related profession. Furthermore, there is less muscle movement at the vocal cord is also victims of the decease. In essence, the voice pathology can be identified with two procedures namely subjective and objective methods. The subjective method deals with oral examination of doctor with the patient. The voice test will be examined physically and the doctor can identify the level or stages of the decease. The objective method is based on supported or assistive tools with the help of software for identifying the decease. The software based applications employed with the different algorithms to identify the decease. However, the accuracy if the objective
based methods for detecting voice pathology is a challenge due to its accuracy. Oral examination sometimes better for conclude the decease and its pattern. The objective procedure works based on the algorithm given and the dataset taken for training and testing. The objective method provides better accuracy if better the algorithm has taken. In our work, a CNN algorithm based voice pathology detection model has been introduced to identify the disease in the voice samples. With the help of smart phones, the voice data is recorded and stored into the cloud. The voice data is preprocessed before applying the CNN algorithm. The training and testing data is given to the algorithm and the algorithm efficiently classifies the test data with positive or negative remarks.

1.1. Literature Survey

Numerous work has been done with voice pathology identification has been discussed in literature. In [1], an LBP and ELM based classifier has been used to classify the voice samples. The Local Binary Pattern (LBP) used to extract the feature description and Extreme Learning Machine used for binary classification. In [2], a detail survey of voice pathology detection system has been discussed. The comparison of the machine learning algorithm is shown with their accuracy. Furthermore, the author presented the issues and challenges pertained in detecting voice pathology. In [3], a voice pathology detection system has been introduced. They are using transfer learning based CNN algorithm is used to detect the decease. The method is tested with SVD database and the accuracy is shown. Al-Nasheri et al developed a voice pathology detection system for detecting early stages of decease. The system divides the voice frame and detect the peak signal valued frame is extracted for classification. The SVM classifier is used to classify the voice samples [4]. Hossain et al proposed big data base voice pathology detection system for classifying the decease. The voice samples of user from hospital has been collected in the MPEG-7 audio features and IDP features are normalized and fusion. With the help of SVM, ELM and GMM algorithms, the voice samples are classified as either pathological or non-pathological voice [5]. In [6], a Dysphonia Detection Index (DDI) calculation for testing the quality of voice has been introduced. The input voice data is taken and the four acoustic parameters are defined. With the help of tree regression model the relationship between these four parameters measured and a threshold has been fixed. If the quality of the voice exceeds the threshold, then the system reported as DDI which means pathological voice otherwise it will be a healthy voice. In [7], a pathology voice detection system has been introduced to distinguish between the defect voice and healthy voice. A mobile health care unit is framework of modules such as voice acquisition, authentication, signal processing, classification, storage and reporting modules. The Boost tree algorithm is combination of decision tree is used to distinguish between pathology voice and healthy voice. Verde et al introduced the machine learning classifiers applied with the voice signals and the performance is compared. The voice signal main parameters such as fundamental frequency, jitter, shimmer, harmonic to noise ratio and cepstral coefficients are used for clinical evaluation. The algorithms which includes decision tree, logistic model tree, Bayesian classifier, instance-based learning are applied with the SVD data set and the performance is compared [8]. In [9], a deep neural network based algorithm for identifying multi-vowel has been discussed. In [10], a multidimensional voice program (MDVP) featuring with 33 voice parameters experimented with the SVM algorithm and voice pathology is detected. In [11], characterization of pathological voice signal is presented. Three different approaches which include classification, divergence criterion and entropy based computation for detecting pathological voice decease. All these existing systems are limited in accuracy of source estimation in time domain, high estimation error, and low estimation accuracy. Furthermore, some of the existing systems provide theoretical way of solution to the voice pathological decease and not in practical. The existing algorithm incurs high in computational cost and complexity in processing.

The proposed method equipped with a smart healthcare framework which integrates patient, doctor and medical attendants with cloud and wireless technologies. The patient smart phone records the voice data as an input and it will be stored in a cloud server. The voice data is preprocessed by removing silence and make it as window file format. The voice signal parameter like jitter, shimmer,
fundamental frequency is extracted for evaluation. The convolutional neural network algorithm is applied with the jitter parameter and classified with the given input is pathology voice or healthy voice.

2. Voice pathology Detection process
Detection of voice pathology is a challenging process as the voice data is complex in nature. It is very difficult to differentiate the healthy voice and pathological voice. The survey shows that the pathological voice decease may happen at least once in a person’s lifetime. There are 15 percent people affected in Arab countries regarding voice related deceases. It may be slur voice due to paralysis or other related deceases. The growth of cist, gland or small tissue in the voice cord leads to the decease. If the decease not identified at the right time, one can lose the speech permanently. In order to detect the voice pathology, the software based model to distinguish the healthy and patient voice. The proposed system employed with CNN algorithm for efficiently detecting the voice pathology. The complete framework integrated with smartphone, wireless network, and cloud computing technologies. The detailed architecture is shown in figure 1.

3. Working Methodology
The patient recorded the voice through the smartphone and the audio data is stored into the private cloud server. The cloud server is connected with the hospital, doctor and medical staff. The hospital equipped with our voice pathological detection system for detecting voice pathology. The doctor feed the input voice data to the system and the system reports the given voice is pathological voice or healthy one. The proposed voice pathological detection system employs with the following modules:

- Data acquisition
- Extraction of parameters
- Data pre-processing

![Figure 1 Voice pathology detection process](image)
3.1. **Data acquisition**

The voice data is captured through the smart mobile and store it in the cloud server database. The collected data will be converted into the required format for further processing. The captured data will be in .wmi file format and it will be converted into vector representation for the application of algorithm.

3.2. **Extraction of parameters**

The important parameters such as fundamental frequency, jitter, and shimmer, harmonic to noise ratio and cepstral coefficients are extracted for process the data. Jitter is an important parameter due to its pitch and frequency helps in accurately classifies the data.

3.3. **Data preprocessing**

Data preprocessing is important in removing the noise and cleaning the data. The silence in between voice signals will be removed. This help in increase the accuracy.

3.4. **Application of CNN**

Convolutional neural network algorithm is efficient for classifying the complex data. Here, the binary classification labeled with voice pathology and healthy voice is the two classified data. The system reports the decision and it will be stored into the cloud server and eventually the doctor conveyed to the patient. This will greatly reduces the visitation time by the patient, transport cost, waiting time etc.

4. **Experimental Results**

The voice pathology system is tested with the sample voice data. The dataset SVD and voice data from Gaggle is considered for processing. The pathology detection process implemented using python framework and the voice samples are tested successfully. The open source dataset is considered and stored it in a database which can be in the .csv format. In our work, the dataset of a collection of 3,168 voice samples of both pathological and non-pathological voice which are equally present in its Datasets. In the data acquisition phase, the images of this data set are taken for processing. The data is preprocessed and the noise free data is in the form of an image. The CNN algorithm is applied with the preprocessed data and the pathology voice is classified. The split function used to feed the train and test voice data for validating and predicting the answer (i.e...Pathology or not). In x train some values are given to predict either the parameter is pathology or not, for that we use y train. Similarly x test is tested with some more different voice samples and predict it, for that we used y test. Therefore, the predicted values are both of pathology and non-pathology. Numpy is used for vector representation. The label encoder by fit transforms to change the Pathology Data as 1 and non-pathology Data as 0 which are binary values. Then we uses pipeline algorithm, here we are taking an array of n number of features (voice samples) in x test. so, from that we do predict the answers of the pipe_svc__predict (X_test). As we are use an label Encoder by fit transform , the result will be like array of both 1’s and 0’s format which are pathology and non patholgy respectively. For validating the work, we used a cross validation check by taking a dataset from voice sample, we checked it ten times and from that we are getting the result as 97% of accuracy on an average. Figure 2 shows the plotted graph blue labelled as healthy voice and yellow labeled as pathological voice.
Two models are implemented for predicting the pathology and non-pathology disease. They are Pipeline SVC model and Grid search Model. Both models give a 97.6% Accuracy Results. Both are quite efficient only when we have taken in thousand numbers of Datasets. The difference we observed between these models when we have given a million of datasets. Figure 3 shows the classification accuracy of the algorithm

5. Conclusion

Voice pathology can be concluded in two observations. First using medical way by using highly expensive equipment to check and another is by system check by using Deep learning approach. The system successfully distinguished between pathology and normal voice with the given dataset. The voice data of the user is considered as an input and the system extracts the features shimmer, jitter, harmonic noise ratio and cepstrum coefficient. The CNN algorithm works well in the given data set and identifies the pathology and non-pathology disease. The presented model can work well against low band speech analysis for pathology identification with continuous signals. This model is one kind of observation for identifying the pathology disease. However, additional observation is required to confirm the voice pathology disease.
References

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

[2] F. T. Al-Dhief et al., "A Survey of Voice Pathology Surveillance Systems Based on Internet of Things and Machine Learning Algorithms," in IEEE Access.

[3] M. Alhussein and G. Muhammad, "Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework," in IEEE Access, vol. 6, pp. 41034-41041, 2018.

[4] A. Al-Nasheri et al., "Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions," in IEEE Access, vol. 6, pp. 6961-6974, 2018.

[5] M. S. Hossain and G. Muhammad, "Healthcare Big Data Voice Pathology Assessment Framework," in IEEE Access, vol. 4, pp. 7806-7815, 2016.

[6] L. Verde, G. De Pietro, M. Alrashoud, A. Ghoneim, K. N. Al-Mutib and G. Sannino, "Dysphonia Detection Index (DDI): A New Multi-Parametric Marker to Evaluate Voice Quality," in IEEE Access, vol. 7, pp. 55689-55697, 2019.

[7] L. Verde, G. De Pietro, M. Alrashoud, A. Ghoneim, K. N. Al-Mutib and G. Sannino, "Leveraging Artificial Intelligence to Improve Voice Disorder Identification Through the Use of a Reliable Mobile App," in IEEE Access, vol. 7, pp. 124048-124054, 2019.

[8] L. Verde, G. De Pietro and G. Sannino, "Voice Disorder Identification by Using Machine Learning Techniques," in IEEE Access, vol. 6, pp. 16246-16255, 2018.

[9] T. Zhang, Y. Wu, Y. Shao, M. Shi, Y. Geng and G. Liu, "A Pathological Multi-Vowels Recognition Algorithm Based on LSP Feature," in IEEE Access, vol. 7, pp. 58866-58875, 2019.

[10] A. Al-nasheri et al., "Voice pathology detection with MDVP parameters using Arabic voice pathology database," 2015 5th National Symposium on Information Technology: Towards New Smart World (NSITNSW), Riyadh, 2015, pp. 1-5.

[11] C. D. Maciel, J. C. Pereira and D. Stewart, "Identifying healthy and pathologically affected voice signals [Lecture notes]," in IEEE Signal Processing Magazine, vol. 27, no. 1, pp. 120-123, Jan. 2010. Computing, 104 (2020) 104046