INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Recognition and overcoming of Dysarthria

Anurag A. S.
sampathanurag3@gmail.com
Sri Jayachamarajendra College of Engineering, Mysore, Karnataka

Dr. M. N. Jayaram
jayarammelkote@sjce.ac.in
Sri Jayachamarajendra College of Engineering, Mysore, Karnataka

Jeevan D'souza
Jeevanjd97@gmail.com
Sri Jayachamarajendra College of Engineering, Mysore, Karnataka

Manoj M. S.
manojsrinath97@gmail.com
Sri Jayachamarajendra College of Engineering, Mysore, Karnataka

Abhishek P.
abhishekaryan13@gmail.com
Sri Jayachamarajendra College of Engineering, Mysore, Karnataka

ABSTRACT

A speech disorder is a condition where a person cannot properly create speech sounds required for communication. About 5% of children in the world suffer from speech disorders. Dysarthria is a motor speech disorder which results in slurred or slow speech that is hard to understand. It is a neurological disorder that affects the muscles which help to produce speech. In this paper, we propose a method where machine learning can be used to predict the actual word spoken by dysarthric patients while the dysarthric speech may not be clearly understood by the common man. We have also proposed a method to predict the gender of the speaker and the severity of dysarthria in patients. We have also built an application which has also the above-mentioned features along with the option of saving the details and also emailing the data.

Keywords— *Dysarthria, Speech disorder, Machine learning, Neural networks.*

1. INTRODUCTION

A speech disorder is a communication disorder that affects the ability to produce normal speech. Dysarthria is a motor speech disorder which is mainly caused by neurological injury. Dysarthria affects the nerves and muscles that help in producing speech required for communication. This disorder often makes it difficult for the speaker to pronounce words. This disorder results in impairments in intelligibility and poor articulation of speech. The main cause of this speech disorder is traumatic brain injury or an embolic stroke.

Machine learning provides systems with the ability to learn and improve from experience rather than explicitly programmed. Machine learning algorithms learn based on the inputs given by the user. In this paper, we propose a method where machine learning can be used to predict the actual word spoken by dysarthric patients while the dysarthric speech may not be clearly understood by the common man.

Our paper on dysarthric speech disorder tries to curb the effects of dysarthria using the latest and in-demand technologies present in the market such as machine learning and deep learning which would help any victim of dysarthria to have a decent conversation and lead a normal life and save him from social embarrassment. We, in our paper try to recognize the speech spoken by the patients using machine learning models and to convert it into a speech that any common man can easily understand we move a step forward and also try to recognize the gender of the patient and intelligibility of the patient which would make our prediction and overcome of the speech disorder more effective.

1.1 Dysarthria

Dysarthria is a neurological disorder where the muscles that help in speech become weak and hence becomes hard to control them. The common causes of dysarthria are a brain injury, stroke, brain tumor or facial paralysis. The common symptoms of dysarthria are slurred or slow speech, uneven speech volume, monotone speech, inability to speak louder or speaking too loudly.

1.2 Machine learning

Machine learning is the science of making computers act based on its experience rather than being explicitly programmed. Machine learning problems can be classified into four methods based on the nature of learning: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Machine learning algorithms produces output based on the previous inputs given. We have used Convolutional Neural Network (CNN) for training our model for word prediction. Initially, for training the neural network we have to collect the appropriate data set. These data sets are fed into the neural network. Based on these data given, the neural network trains itself by changing the weights and biases of each node. After training the network with the datasets, we obtain a model which we use for our application.

© 2019, www.IJARIIT.com All Rights Reserved

In this paper, we have also proposed a method to find the gender of the speaker and the severity of disease in the speaker by using CNN and Support Vector Machines. We have also compared different machine learning algorithms for their efficiency in predicting the gender of the speaker.

1.3 Convolutional Neural Network (CNN)

It is a class of deep neural network which use a variation of multilayer perceptron. Convolutional networks were inspired by the connectivity pattern of neurons. CNN's have application in image and video recognition, medical imaging and natural language processing. A CNN has one input and, an output layer and multiple hidden layers. The hidden layers consist of convolutional layers and RELU layers. The convolutional layers apply convolution operation on the input and pass the result to the next layer.

2. LITERATURE SURVEY

G. Vyas, M. K. Dutta, J. Prinosil and P. Harr in their paper [1] make an automatic diagnosis and assessment of the dysarthric speech which is done by selecting speech disorder prosodic features. These features selected are then used for training and testing of SVMs. The features extracted from the speech samples for diagnosis of dysarthric speech are MFCCs, peak amplitude, skewness, kurtosis, fundamental frequency and formats. They have implemented Support Vector Machines (SVM) to diagnose dysarthric speech. An-jali Pahwa and Gaurav Aggarwal have used speech feature extraction for gender recognition in their paper [2]. They have used a combined model of SVM and neural network classification to determine their gender using stacking. MFCC is one of the most dominating features among all the features of speech. MFCC makes a close analogy with the human ear by considering those parameters extracted by the human ear. Lansford, K. L. & Liss, J. M have made an extensive study of vowel acoustics in dysarthria [3]. The goal of this experiment was to identify vowel metrics that differentiate disordered from non-disordered speakers and the dysarthria subtypes. Acoustic metrics that capture production deficits in dysarthria have the potential to be powerful and objective diagnostic and prognostic tools. L. Deng and X. Li has presented a paper about automatic Speech Recognition (ASR) in [4] has historically been a driving force behind many machine learning (ML)techniques, including the ubiquitously used hidden Markov model, discriminative learning, structured sequence learning, Bayesian learning, and adaptive learning.

T. B. Ijitona, J. J. Soraghan, A. Lowit, G. Di-Caterina and H. Yue make a study on the automatic detection of dysarthria using extended speech features called as centroid formants. [5] J. Chhikara and J. Singh, in the paper [6] present an adaptive noise cancellation algorithm which is used for noise reduction in the speech signal. The received signal is corrupted by noise where both received signal and noise signal changes continuously. The two adaptive algorithms used in this paper to reduce noise are the Least Mean Square (LMS) algorithm and the NLMS algorithm. Adaptive filter adjusts their coefficients to minimize the error signal. From the various research papers, journals, textbooks we can infer the following with respect to our project. We can extract the various speech features like MFCC, peak amplitude, skewness etc. by using the python libraries and simple statistical measures. From another paper, we can find a competitive approach of the same topic by using centroid formants which has the advantage of being less prone to noise which increases the capability of our system. Since overfitting is a major concern in machine learning model training, a research paper has given us an idea about dropping some of the neuron units while training. TensorFlow is a major tool which can be used for this project to improve the process of building the neural network [7]. From one of the papers about vowel acoustics in dysarthria, we have found a disadvantage that the vowels of the disordered and non-disordered speakers cannot be differentiated by much while in isolation. This made us change the approach in solving our problem. Long Short-Term Memory (LSTM) and Recurrent Neural Network could be an alternative approach to solve the recognition problem for comparison purposes. From one of the researchers, we learnt crucial information that the speaker traits are not long-term distributional patterns but short time variations which will immensely help us in speech recognition [8]. An improvement of the neural network built can be done by back propagation which is analogous to the feedback system in electronic circuits to improve the accuracy and minimize the error. Since noise could be of concern while recording speech in outdoor environments, an adaptive noise cancellation algorithm using Least Mean Square can be modified and implemented to remove the noise from our speech signals .N. M. Joy and S. Umesh in [9] have explored ways of using Deep Neural Networks and Hidden Markov Models to improve the acoustic models for the TORGO Database. We have used another dataset and these algorithms could be used to enhance the models for our dataset. M. Kim, Y. Kim, J. Yoo, J. Wang and H. Kim in [10] have used a variation of [9] where the Kullback-Leibler divergence-based hidden Markov model is used for Dysarthric Speech Recognition where emission probability of state is parameterized by a categorical distribution using phoneme posterior probabilities obtained from a deep neural network-based acoustic model (KL-HMM).

The organization of the paper is as follows. The paper begins with the introduction of the topic and the problem at hand. Then, the brief methodology is specified in chapter 2. The application build i.e. the Graphical User Interface for this project is spoken about in chapter 3. Following on with the results in chapter 4 and conclusions in chapter 5.

3. METHODOLOGY

The process of building an entire application has several steps such as: collecting data sets, developing a machine learning algorithm and training the model. Then we build the GUI for unique user experience.

The block diagram shown in figure 1 explains the working of our project. The application of the project begins with having voice samples. For our voice samples, we have taken it from the UA Speech Database. The voice data is sampled. Since the number of samples is too high, the samples need to be properly segregated and sorted. The segregated samples are clustered with respect to their different use cases. Our use case is a prediction of the word, gender and severity. Hence, we have clustered the data into labels and provided Metadata for the context while training the samples using neural networks and other machine learning algorithms. A simple noise filter is used to remove any noise or interference to provide clarity and understandability to the listener of the sample. Various speech features such as MFCC, Amplitude and Frequency and Statistical parameters are extracted from each of the samples to create a data-frame which will be used for processing. The models are trained by using supervised learning algorithms and hence will be used to predict the results of an unseen sample accurately.

3.1 Requirements

The project places minimal requirements for the successful running of the application. The entire project is built on the programming language Python. There are various open source Python libraries used for the various modules.

© 2019, www.IJARIIT.com All Rights Reserved

The only special hardware requirement would be a microphone for testing the application by recording the different voice samples.

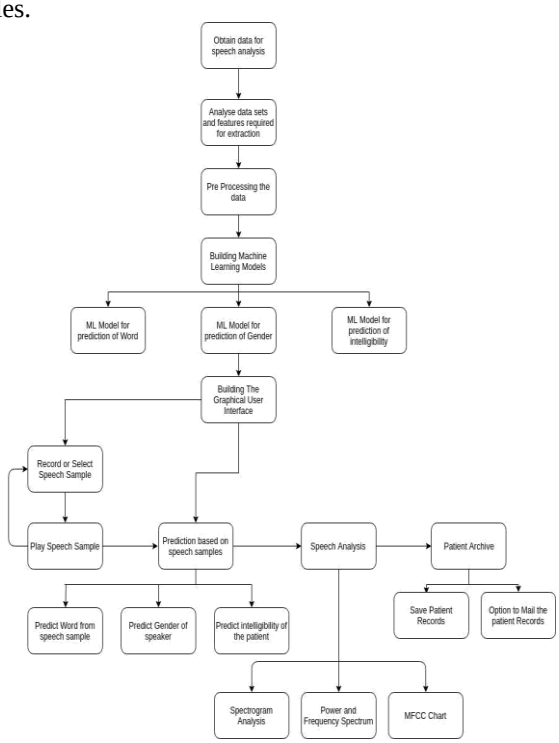

Fig. 1: Block diagram

3.2 Data set

The Machine learning algorithms require data set for building a mathematical model. The data set required to build the model is known as training data. The two suitable data sets with dysarthric speech samples of different speakers found are.

3.3 UA-Speech Database TORGO Database

The second data set is taken from the UA speech database from the University of Illinois. Among these two data sets, the second one is more diverse and suitable for our project requirements. There is more variation in the data samples, the number of samples is considerably larger and this data set incorporates a number of different speakers. Hence, we have chosen the UA-Speech Database. [11] [12]

3.4 Data Description

Digits (10 words X 3 reps):"one, two, three,"

Letters (26 words X 3 reps): the 26 letters of the International Radio Alphabet, "alpha, bravo, Charlie,"

Computer Commands (19 words X 3 reps): word processing commands, e.g., "command, line, paragraph, enter," Common Words (100 words X 3 reps): the most common 100 words in the Brown corpus, e.g., "the, of, and,"

Uncommon Words (300 words X 1 rep): 300 words selected from Project Gutenberg novels using an algorithm that seeks to maximize biphobe diversity, e.g.,"naturalization, faithfulness, frugality,"

3.5 Data Pre-Processing

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. The data from the UA-SPEECH database needed to be extracted from an FTP server hosted by the University. The data was downloaded and streamed into channels for further processing. Data segregation is the first stage of pre-processing. Here the data from the data sets are segregated in two ways:

(a) Separating all the voice samples into different word labels.
(b) Separating all the voice samples into different speakers.
(c) Separating all the voice samples into different severities.

After segregating the data, the data samples are sampled into time-amplitude values and stored in files along with their metadata. The various speech features are extracted from the sampled and segregated audio files for training the different machine learning algorithms. The different speech features extracted for our project is described below.

3.6 Speech features extracted

Mel-Frequency Cepstrum (MFC) is the representation of the power spectrum of sound, based on linear cosine transform of a log power spectrum on non-linear mel-scale of frequencies.

The Coefficients that collectively makeup MFC is known as Mel-Frequency Cepstrum Coefficients (MFCCs). Peak Amplitude Peak amplitude is the maximum absolute value of the signal. Peak amplitude can also be defined as the maximum value of the signal from zero or equilibrium point.

- **Mean Amplitude:** Mean amplitude is the average amplitude of the signal. The RMS of a speech signal is defined as the square root of the mean over time of the square of the amplitude signal.
- **Kurtosis of Amplitude:** Kurtosis is the measure of the sharpness of the peak amplitude distribution curve. It is a unitless parameter that quantifies the distribution shape of the signal.
- **Standard Deviation of amplitude:** Standard deviation of amplitude gives the measurement of the dispersion of amplitude values of the speech signal.
- **Spectral Centroid:** Spectral centroid indicates where the centre of mass of the speech signal is located. It indicates where the maximum amplitude distribution is located in the speech signal.
- **Entropy:** The spectral entropy gives the measure of power distribution of the spectrum. The entropy is a measure of disorganization and it can be used to measure the peakiness of a distribution.
- **Pitch:** The relative highness or lowness of the tone in a speech signal is known as pitch. The pitch of the speech signal can be used in speech detection.
- **Linear Predictive Coding:** Linear Predictive Coding (LPC) is a tool used to represent the spectral envelope of the digital speech signal in compressed form, using the information of the linear predictive model.
- **Minimum Frequency:** The minimum frequency of the speech signal is the least frequency found in the speech signal.
- **Quartile Ranges of Frequencies:** Quartile range is the mid-spread of the frequency distribution of a speech signal. It is equal to the difference between 75th and 25th percentile, or between upper and lower quartiles.

3.7 Training and testing machine learning algorithms

With the extraction of the different features and their segregation and clustering, the next step is the training of the machine learning models. The machine learning models are trained using the different supervised learning algorithms like decision trees, random forests, gradient boosting, SVM and CNN's.

With the trial of the different algorithms [13] for the predictions on dysarthric speech samples, it was found the CNN gave the

© 2019, www.IJARIIT.com All Rights Reserved

highest accuracy and best prediction among the different algorithms used. Hence, we will discuss the CNN architecture used for the prediction purpose.

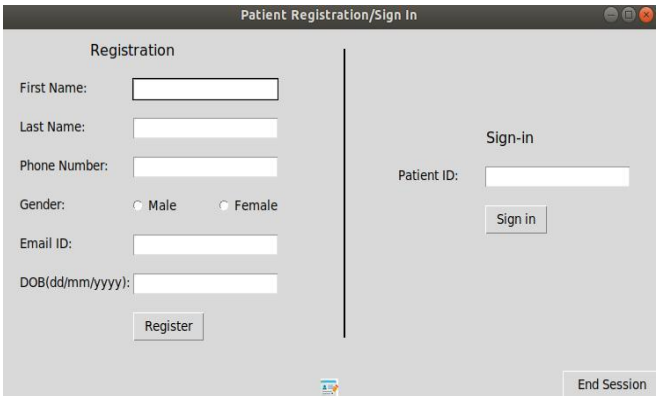

Fig. 2: Patient Registration/ Sign in

3.8 Dashboard

The comprehensive application is a dashboard which features everything from the audio signal recording till archiving and mailing the data. The GUI shown in figure 3 is quite self-explanatory. The dashboard is mainly divided into 5 sections.

(a) Recording or selecting the audio file.
(b) Playing the audio file and making predictions.
(c) Generating different waveforms.
(d) Archiving patient data.
(e) Mailing the results of the session.

4. APPLICATION

4.1 GUI in Python

Python offers multiple options for developing GUI Out of all the GUI methods, tkinter is the most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with Tkinter outputs the fastest and easiest way to create GUI applications.

4.2 Registration/ Sign in

There are typically two scenarios. The patient may be visiting the doctor for the first time or he may be an already diagnosed patient at the same hospital. Therefore the GUI provides the feature for existing patients and newcomers to sign into the application. The patient will be given a unique patient ID which can be used to sign in. If he/she is a first timer, then after entering the general details, a new ID will be generated. The screen displayed can be seen in figure 2. [14]

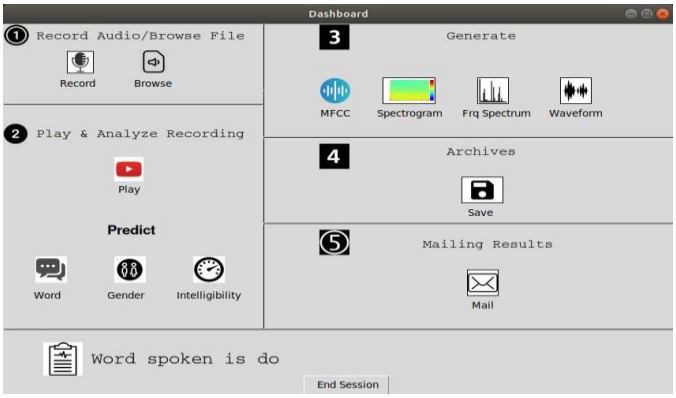

Fig. 3: Dashboard

5. RESULTS

5.1 Gender prediction

Different machine learning algorithms were tested on the data generated and processed. The results of the training and testing of the data are summarized in the table below.

Figure 4 shows the different features used for gender prediction and their distribution with the data samples. From the above, we can see the difference in the features of male and female speech samples. Some features are more clearly distinguishing both the genders than the other features. Features like entropy and skewness of amplitude clearly demarcated than the others and will have relatively higher feature importance as shown in the figure from figure 7 we can conclude that both the genders and have more feature importance than the other features. This can be seen in figure 5.

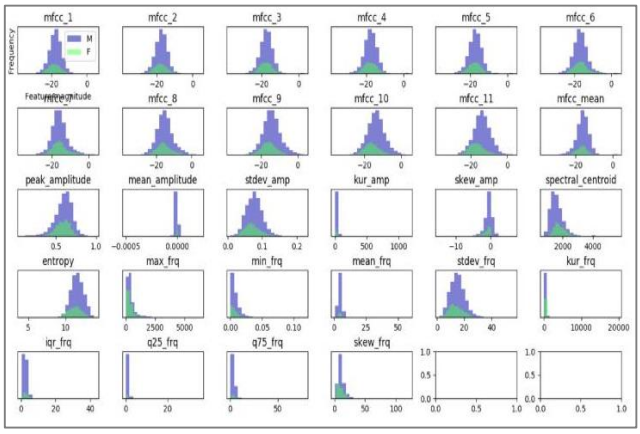

Fig. 4: Feature for gender prediction

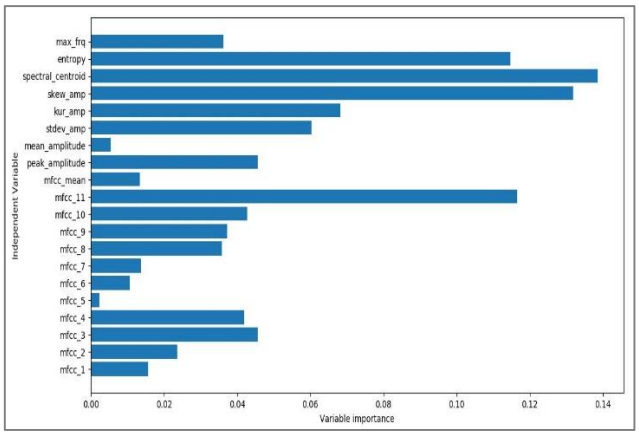

Fig. 5: Feature importance for gender prediction

The below table 1 shows the accuracy of training and testing for the different algorithm. Compared with this algorithm Convolutional Neural Networks (CNN) has better accuracy on both training and test set. As we can see from the above table CNN has the best accuracy. Hence, CNN was chosen.

Table 1: Accuracies for gender prediction

Algorithm	Accuracy-training set	Accuracy-test set
Decision Tree	1.000	0.807
Random Forests	0.985	0.841
Gradient Boosting	0.834	0.829
Support Vector Machines	0.891	0.885
Multilayer Perceptron	0.915	0.897
CNN	0.929	0.903

5.2 Intelligibility prediction

Various supervised machine learning algorithms were used for the prediction of intelligibility of dysarthria. The results and observations are as shown below.

Figure 6 shows the different speech features used for severity prediction. There are 3 classes now instead of 2 earlier. The three classes are low severity, medium severity and high severity

© 2019, www.IJARIIT.com All Rights Reserved

which are marked by L, M and H respectively. Like the earlier case, some features are more clearly demarcated.

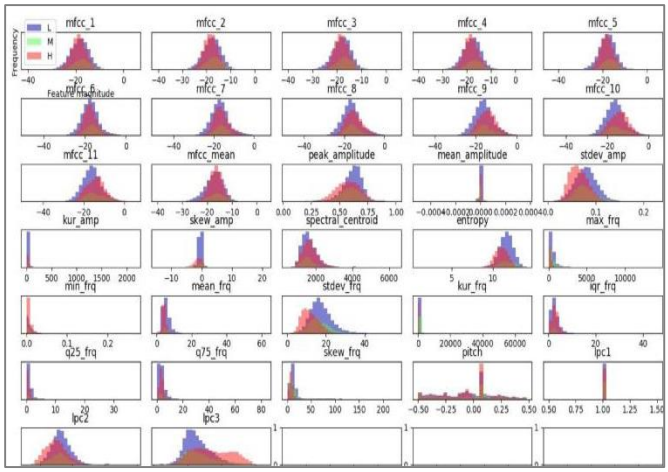

Fig. 6: Speech features for intelligibility prediction

LPC features and kurtosis of frequency, the standard deviation of amplitude are relatively more important than the other features used for severity prediction. This varies with data. The features that were important in gender prediction may be irrelevant or less important in severity prediction and vice versa.

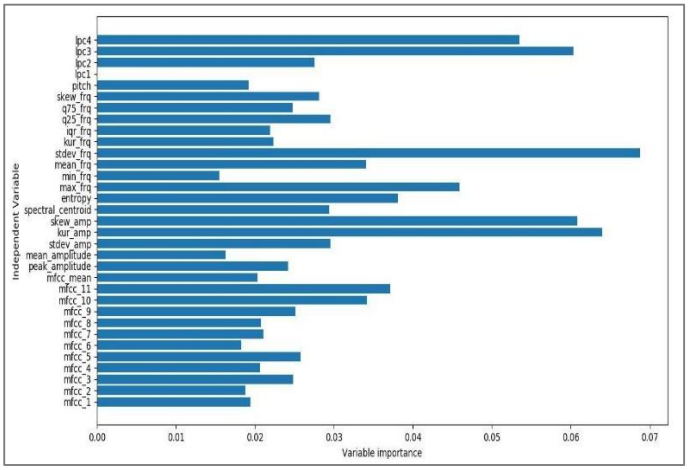

Fig. 7: Speech features importance for intelligibility prediction

The below table 2 shows the accuracy of training and testing for the different algorithm. As we can see that compared to other algorithms CNN has the best training and testing accuracy in finding the intelligibility level of the dysarthric speech signal. As a result, we have used CNN algorithm for finding the intelligibility level of dysarthric speech.

Table 2: Accuracies for intelligibility prediction

Algorithm	Accuracy-training set	Accuracy-test set
Decision Tree	1.000	0.701
Random Forests	0.973	0.749
Gradient Boosting	0.767	0.758
Support Vector Machines	0.829	0.825
Multilayer Perceptron	0.863	0.844
CNN	0.878	0.867

5.3 Word prediction

Word prediction is a multi-class classification problem. The number of words that need to be predicted from the number of classes. Hence it becomes increasingly difficult to classify a large number of words with a limited amount of data.

The accuracy results were obtained as follows. This data shown in table 3 is for the classification of 50 words. With a greater number of samples, the accuracy will improve considerably.

Table 3: Accuracies for word prediction using CNN

Training Accuracy	Testing Accuracy	Error
0.9886	0.7540	1.4478

6. CONCLUSION

The project Recognition and Overcoming of Dysarthric Speech was motivated by the fact that it would help the patients lead a better life and make the work of a speech-language pathologist easier. Having said that, the project began with fetching the data from the dataset courtesy. The data fetched was reprocessed (segregated and clustered). Machine learning models were applied to the data to predict the word spoken, the gender and the severity level of the patient. As an addition, a GUI dashboard with all the required functionality from recording the voice sample to analyzing and getting insights was created. The dashboard also provides more insight into the speech sample, archiving of patient data and mailing the session details. Speech disorders can affect the way a person creates a sound to form words. Communication is very crucial in life, especially in education. Speech and language disorders as with any learning disability will call social embarrassment along with setbacks. Not many people are working to solve the problem to an extent by using machine learning, neural networks and deep learning. Hence speech recognition of Dysarthria could be of help to the patients in effectively communicating with everyone around them.

7. ACKNOWLEDGEMENT

The satisfaction that accomplices the successful completion of any task would be incomplete without the mention of people who made it possible. First and Foremost, we ought to pay due regards to our institution Sri Jayachamarajendra College of Engineering Mysuru, which provided us with a great opportunity for carrying out this project work. We are very much thankful to Dr T. N. Nagabhushan, Principal, Sri Jayachamarajendra College of Engineering Mysuru. We are extremely grateful, Dr N. Shankaraiah, Professor and Head of the Department of Electronics and Communication Engineering, Sri Jayachama-Rajendra College of Engineering Mysuru.

We would also like to thank Dr Mark Hasegawa Johnson who helped us with acquiring the speech dataset. We sincerely thank our family members for all the support and strength they have given us to finish this project.

8. REFERENCES

[1] G. Vyas, M. K. Dutta, J. Prinosil and P. Harr, ”An automatic diagnosis and assessment of dysarthric speech using speech disorder-specific prosodic features,” 2016 39th International Conference on Telecommunications and Signal Processing (TSP), Vienna, 2016, pp. 515-518.

[2] Anjali Pahwa and Gaurav Aggarwal, ”Speech Feature Extraction for Gender Recognition,” I.J. Image, Graphics and Signal Processing, 2016, 9, 17-25

[3] Lansford, K. L. & Liss, J. M. Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. J. Speech Lang. Hear. Res. 57, 5767 (2014).

[4] L. Deng and X. Li, ”Machine Learning Paradigms for Speech Recognition: An Overview,” in IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 5, pp. 1060-1089, May 2013.

© 2019, www.IJARIIT.com All Rights Reserved

[5] T. B. Ijitona, J. J. Soraghan, A. Lowit, G. Di-Caterina and H. Yue, "Automatic detection of speech disorder in dysarthria using extended speech feature extraction and neural networks classification," IET 3rd International Conference on Intelligent Signal Processing (ISP 2017), London, 2017, pp. 1-6.

[6] J. Chhikara and J. Singh, Noise Cancellation using Adaptive Algorithms, International Journal of Modern Engineering Research, Vol. 2, No. 3, pp. 792-795, 2012.

[7] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. TensorFlow: A System for Large-Scale Machine Learning. In OSDI, Vol. 16. 265283.

[8] Wang, Dong & Li, Lantian & Shi, Ying & Chen, Yixiang & Tang, Zhiyuan. (2017). Deep Factorization for Speech Signal. [Online] https://arxiv.org/abs/1706.01777

[9] N. M. Joy and S. Umesh, "Improving Acoustic Models in TORGO Dysarthric Speech Database," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 26, no. 3, pp. 637-645, March 2018.

[10] M. Kim, Y. Kim, J. Yoo, J. Wang and H. Kim, "Regularized Speaker Adaptation of KL-HMM for Dysarthric Speech Recognition," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 9, pp. 1581-1591, Sept. 2017.

[11] Rudzicz, F., Namasivayam, A.K., Wolff, T. (2012) The TORGO database of acoustic and articulatory speech from speakers with dysarthria. Language Resources and Evaluation, 46(4), pages 523–541

[12] Kim, H., Hasegawa-Johnson, M. A., Perlman, A., Gunderson, J., Huang, T. S., Watkin, K., & Frame, S. (2008). Dysarthric speech database for universal access research. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 1741-1744.

[13] Sharma, Diksha & Kumar, Neeraj. (2017). A Review of Machine Learning Algorithms, Tasks and Applications. 6. 2278-1323.

[14] Python GUI: "https://python-textbok.readthedocs.io/en/1.0/Introduction to GUI Programming.html"

© 2019, www.IJARIIT.com All Rights Reserved