International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 05 | May - 2021 ISSN: 2582-3930

Diagnosing Voice Disorder with Machine Learning

Hemal Sawdekar[1] Sharvari Birute[2] Apurva Narkhede[3]
Nandini Sonawane[4] Prof. Namrata Ghuse[5]

hsawdekar@gmail.com[1] birutesharu@gmail.com[2] narkhedeapurva@gmail.com[3]

nandinigsonawane84@gmail.com[4] namrata.ghuse@sitrc.org[5]

Abstract - Edge computing offers useful computing resources at the edge of the network to maintain low latency and real-time computing. The advancement of next-generation network technologies provides a huge improvement in healthcare facilities. Technologies such as 5G, edge computing, cloud computing, and the Internet of Things realize smart healthcare that a client can have anytime, anywhere, and in real time. Edge computing offers useful computing resources at the edge of the network to maintain low-latency and real-time computing. In this System, we propose a smart healthcare framework using edge computing. We develop a voice disorder assessment and treatment system using a deep learning approach. A client provides his or her voice sample captured by smart sensors, and the sample goes to the edge computing for initial processing. Then using SOAP protocol system sends data to a core cloud for further processing. The assessment and management are controlled by a service provider through a cloud manager. Once the automatic assessment is done, the decision is sent to specialists, who prescribe appropriate treatment to the clients. Voice disorders are medical conditions involving abnormal pitch, loudness or quality of the sound produced by the larynx and therefore affecting speech production.

Key Words: Voice disorder diagnosis, Machine Learning, SVM, KNN, Gradient Boosting, Ensemble Learning etc.

I. INTRODUCTION

Voice disorders are medical conditions involving abnormal pitch, loudness or quality of the sound produced by the larynx and therefore affecting speech production. Most of the traditional diagnostic methods on voice disorders rely on the expensive devices and clinician's experience. These methods result not only in considerable cost for the patients and incorrect detection of diagnosis, but also cause delay for the patients in places without the specialists and medical resources. Computer aided medical systems are being used more and more often to help doctors diagnose the pathological status of the voice with lower cost and non-subjectivity. In the sense of data science, the diagnosis of pathological voice is a multiclass classification problem that is dependent on the audio signal from individuals. Technologies such as 5G, edge computing, cloud computing, and the Internet of Thingsrealize smart healthcare that a client can have anytime, anywhere, and in real time.

Edge computing offers useful computing resources at the edge of the network to maintain low-latency and real-time computing. The voice is one ofthe most important factors that we use in communication between people. Voice and speakingskills are the easiest way to reflect the thought and it is the most important component that distinguishes it from other living things. The voice is part of the personality and character of almost every person. We can also understand diseases using voices because some diseases directly affect human voice. Finding diagnoses such as frontal lobe resection, spasmodic dysphonia and cordectomy from patient voice data has become possible with today's technologies. The healthcare industry is booming at a great pace and expects to earna revenue of billions of dollars. The industry is moving toward some new trends, which include meeting the needs of consumers of developed and emerging countries, an aging population, and a growing middle class. According to the World Health Organization (WHO)[11], chronic diseases are expected to rise by more than 50 percent by 2020. Life- threatening diseases, such as SARS, Ebola, MERS, swine flu and H1N1, need coordinated healthcare management. Nowadays, patients demand more sophisticated, flexible, user- friendly, and personalized healthcare services. They want to be monitored wirelessly, sitting at home, rather than visiting doctors. For example, patients have started to consider having hospital-based medical treatment, such as chemotherapy, in their own homes. They also want their check-ups, such as diabetes, voice disorders, and heartbeat irregularity, to be done at home by sending readings or signals wirelessly.

II. LITERATURE SURVEY

Nowadays, patients demand more sophisticated, flexible, user-friendly, and personalized healthcare services. They want to be monitored wirelessly, sitting at home, rather than visiting doctors. For example, patients have started to consider having hospital-based medical treatment, such as chemotherapy, in their own homes. They also want their check-ups, such as diabetes, voice disorders, and heartbeat irregularity, to be done at home by sending readings or signals wirelessly. Next generation network technologies are a boost to advance healthcare services. Technologies such as fifth generation (5G) networks, edge computing (EC), cloud computing, and the Internet of Things (IoT) bring on- demand, fast, ubiquitous, seamless, and uninterrupted communications that help in many applications, including healthcare service. These technologies provide huge computing resources, storage, access from anywhere anytime, distributive and parallel computing, quick access, and sensing devices with Wi-Fi or Bluetooth communication. There are several studies on healthcare service using these technologies. Muhammad et al. proposed a healthcare framework using IoT and the cloud [2]. They realized the framework for a case study of voice pathology classification. The voice signals were captured by IoT sensors, and the processing was done in the cloud. A software defined network for healthcare services was proposed in [3]. An e-healthcare framework was suggested for elderly people in [4]. The authors mainly investigated aging patients requiring special needs using multisensory data, including image, audio, and Kinect. A system was developed to detect finger movement within the framework. A ubiquitous healthcare framework was developed using cloud computing in [5]; however, this framework did not include the latest technologies such as edge or fog computing, or IoT. Last year, Shih-Hau Fang, et al. compared the Deep Neural Network (DNN) with Gaussian mixture model (GMM) and support vector machine (SVM) and concluded that 3 layers DNN classifier outperformed others [6]. This year, Huiyi Wu, et al. performed the convolutional neural network (CNN) to the Saarbrucken voice data, which contains 71 different pathologies with speech recording from 2000 individuals. It turned out that CNN effectively extracts features from diagnosis voice disorders and also makes the system more robust [1]. Voice disorders in schoolchildren may occur due to functional, structural, or neurologic processes. Voice disorders caused by a functional process (e.g., vocal behaviors resulting in tissue damage) include vocal fold edema, nodules, and polyps. Fluency or articulatory disorders are also caused by functional processes such as increased effort in phonation.

III. PROPOSED SYSTEM

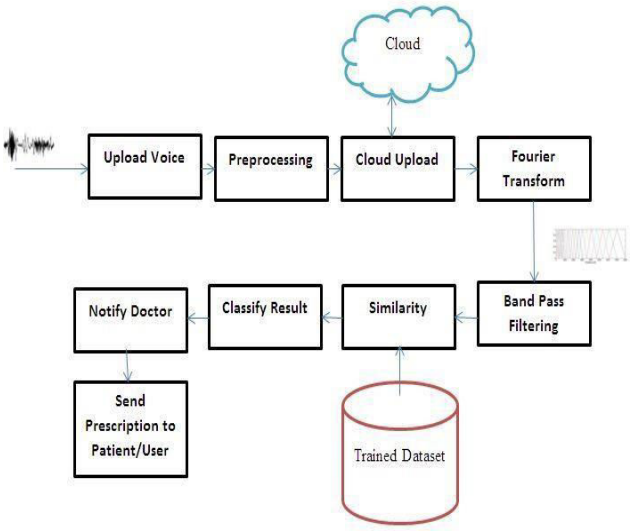

Fig. Block Diagram

We propose a smart healthcare framework using EC and cloud computing. As a case study, we propose a voice disorder detection and classification system in the framework. The voice disorder assessment includes the detection and classification of voice samples. A voice disorder can cause an irregular movement of the vocal folds during phonation. Many people, including children and the elderly, suffer from voice disorders. Some professionals, such as singers, teachers, and lawyers, suffer more from voice disorders because these people extensively use their voice. Although a significant number of people have voice disorders, it is very rare that all these people turn to specialized clinics for several reasons, including transportation problems, scarcity of specialists, and lack of proper diagnostic machines. One of the affected groups is school children. Voice disorders due to a structural process include papilloma, asthma, laryngopharyngeal reflux caused by a cough, and granuloma. These types of voice disorders can be congenital or acquired. Voice disorders from a neuromuscular process may include cerebral palsy and vocal fold muscular dystrophy. There are other common voice disorders such as vocal fold(s) paralysis, cysts, paradoxical vocal fold motion, and laryngoceles. The proposed voice disorder assessment and treatment framework involves several main components: clients, assessors, SLPs, service providers, and network structure. The framework can work in a smart city. The smart city has smart homes, smart schools, a smart transportation system, and smart shopping. The clients can be citizens of the city or schools, where the students are actually the recipients. The assessors are specialists such as laryngologists. The assessors are affiliated with designated clinics, which are registered in the framework. The SLPs can move between the clinic and the clients if necessary, or they can provide therapy remotely. The traffic is controlled by a central system,

which is synchronized by vehicular networks. The network structure consists of EC and core cloud computing. A software-defined network (SDN) will provide high-bandwidth flexible and programmable communication. Edge computing offers the computing resources at the edge of the network to perform smooth and near-real-time operation.

A. *Data Cleaning*

Sound consists of audible variation in air pressure. Microphones convert this variation into a series of varying voltage. Therefore, a reading of a sound file provides us with a time series. Since the data has sampling rate of 44,100 Hz, a sequence 44,100 points in a time series represent one second of a person's voice.

After quickly skimming through the voice samples, we realize that most of them have silence periods at the beginning and the end of the sound files. Therefore, we first remove the silence parts at the beginning and the end the time series that is created from each of the 200 voice samples. Let Xt be the time series that is created from a voice sample and $Yt = |Xt|$. We define $Y0.25$ as the 25th percentile of all the elements of Yt . we then find the smallest value of t such that $Yt > Y0.25$ and denote this value t = a. Then, we find the largest value of t such that $Yt > Y0.25$ and denote this value t = b. The silence parts of the time series is determined as $\{Xt : t < a \text{ or } t > b\}$ and we removed these two parts from the time series.

This cleaning step is necessary because we do not want the silence period to affect the features extracted from the voice samples. Even if some non-silence points are removed, not much information is lost because the voice samples consist of a prolonged vowel sound /a:/.

B. *Feature Extraction*

For each time series created from a voice sample, we used the process described by Logan [7] to compute the Mel Frequency Cepstral Coefficients (MFCC):

1. First, frame the signal into short frames of length 25milliseconds.
2. For each frame, calculate the periodogram estimateof the power spectrum
3. Apply the meld filter bank with 26 filters to the power spectra, sum the energy in each filter.
4. Take the logarithm of all filter bank energies.
5. Take the Discrete cosine transform of the log filterbank energies, we obtain the MFCC's.

We then come up with a 26×n matrix, with n being the number of frames that is dependent on the length of the cleaned voice sample, and 26 being the number of MFCC's. For example, if a cleaned voice sample is three-second long, the time series has 3*44,100 = 132,300 points, and there are n = 3/0.025 = 120 frames (0.025 second is the frame length). Finally, we calculated the minimum, first quartile, median, third quartile, and maximum of each row (n data point) of the resulted matrix. This gives us the five-number summary for each MFCC of a voice sample. The resulted data set is a matrix of shape200×130. There are 130 columns because there arefive values for each MFCC feature (26*5 = 130).

C. *Machine Learning*

At this point, we have a 200×130 matrix and an array of 200 labels for 200 voice samples. We applied various machine learning and deep learning models to the data and measured their performance.

C1. Performance measurement

The performance measurement that we used is accuracy, i.e., the number of correct classifications divided by the total number of predictions. We decided to use repeated sub-sampling to estimate algorithms' accuracies [8]. The process can be summarized as follow.

1. We first divide the 200 samples into two sets, a set of 160 samples for training and a set of 40 samples for testing. We use the stratified sampling method to ensure that each set has a balanced number of labels. That means that we randomly select 40 out of 50 Normal samples, 32 out of 40 Neoplasm samples, 48 out of 60 Phonotrauma samples, and 40 out of 50 Vocal Palsy samples for the training set.

2. We then train the models on the training set, make predictions on the testing sets, and calculate the accuracy score.

3. Repeat the first two steps 100 times and take the average of the 100 accuracy scores. This average number is our estimate of the performance of the models.

We decided to use repeated sub-sampling because our data is too small for k-fold cross-validation [9].

C2. Machine Learning models

We applied for Machine Learning models to the data, namely Support Vector Machine, Random Forest, K-Nearest Neighbour, and Gradient Boosting. In addition,

we also used Ensemble Learning to combine the strength of all the four mentioned algorithms. Support vector Machine, K-Nearest Neighbours, and Gradient Boosting all require tuning hyper parameters. This step is done before the performance measurement step described above, by using all the 200 samples in a repeated sub-sampling process to estimate the accuracy of each set of hyper-parameters. Summaries of the models and tuning of hyper-parameters are described as follows.

Support Vector Machine (SVM) is an algorithm that finds classification boundaries so that categories are divided by a clear gap that is as wide as possible [10]. There are three commonly used kernels, include linear kernel, Gamma kernel, and polynomial kernel. To satisfy the assumption of SVM, we need to standardize the data before we apply the SVM. The reason we use this method is to have a generally knowledge of classifier's type, such as linear or non-linear. This model requires tuning of the kernel, the Gamma parameter, and the Cost parameter. We tuned the model on the Linear, Polynomial, and Radial Basis kernels, with the grids of Gamma in {0.001, 0.01, 0.1, 1}, and Cost in {0.001, 0.01, 0.1, 1, 10 }. For Polynomial kernel, we also have aset of polynomial degree {2, 3, 4, 5}.

IV. CONCLUSION

This paper has presented, a novel system that uses pedometric and indoor mobile augmented reality, to recommend the best exit path with the shortest time to users needing to evacuate a building in emergency situations. Through an evaluation of this application, we have shown how our system leverages the sensors on a smartphone, in conjunction with personalized daily walking stride length estimation and emergency information, to support a timely evacuation. Our practical pedometric algorithm with personalized stride estimation provides high positioning accuracy.

V. REFERENCES

[1] WHO, Nutrition Report;http://www.who.int/nutrition/top-ices/2_background/an/, accessed10 Aug. 2017?

[2] G. Muhammad et al., “Smart Health Solution Integrating IoT and Cloud: A Case Study of Voice Pathology Monitoring,” IEEE Common. Mag., vol. 55,no. 1, Jan.2017, pp. 69–73.

[3] L. Hu et al., “Software Defined Healthcare Networks,” IEEE Wireless Common., vol.22, no. 6, Dec. 2015.

[4] M. S. Hossain, M. A. Rahman, and G. Muhammad, “Cyber Physical Cloud-Oriented Multi- Sensory Smart Home Frame- work for Elderly People: An Energy Efficiency Perspective,”

[5] J. Parallel and Distributed Computing, vol. 103, May 2017, pp. 11–21.

[6] C. He, X. Fan, and Y. Li, “Toward Ubiquitous Healthcare Services with a Novel Efficient Cloud Platform,” IEEE Trans. Biomedical Engineering, vol. 60, no. 1, 2013, pp. 230–34.

[7] B. H. Ruddy, V. Lewis, and C. M. Sapienza, “The Role of the Speech-Language Pathologist in the Schools for the Treatment of Voice Disorders: Working within the Framework of the Individuals with Disabilities Education Improvement Act, “Seminars in Speech and Language, vol. 34, no. 2, 2013, pp. 055–062.

[8] N.P. Connor et al., “Attitudes of Children with Dysphonia,” J. Voice, vol. 22, pp. 197-209, 2008.

[9] M. S. Hossain and G. Muhammad, “Healthcare Big Data Voice Pathology Assessment Framework,” IEEE Access, vol. 4, no. 1, 2016, pp. 7806–15.

[10] J. O. Fajardo, I. Taboada, and F. Liberal, “Radio-Aware Ser- vice-Level Scheduling to Minimize Downlink Traffic Delay through Mobile Edge Computing,” Mobile Networks and Management, Springer, 2015, pp. 121–34.

[11] G. Muhammad et al., “Voice Pathology Detection Using Interlaced Derivative Patterson Glottal Source Excitation,” Biomedical Signal Processing and Control, vol. 31, Jan. 2017,

[12] Logan, Beth. "Mel Frequency Cepstral Coefficients forMusic Modeling." ISMI R. Vol. 270. 2000.

[13] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.