VOICE BASED CLASSIFICATION OF PATIENTS USING VARIOUS MACHINE LEARNING & DEEP LEARNING TECHNIQUES

Satyajit Pattnaik, Kalash Jindal
India
ORCID: 0009-0009-7573-1143

Abstract
Different types of research and diagnosis methods are now introduced for patients using voice-based classifications. Many people have voice/speech-related pathology abnormalities and experience a variety of voice and speech-related issues. The major goal is to give a wealth of information and analysis on the identification of pathological conditions caused by speech disorders. To do so, a voice examination must be performed, and the data from different healthy and unhealthy persons are used for the evaluation of the diseases; systematic differences, tonsillitis, or other vocal disorders are examples, and the main use of this proposed method is to get aware of the disease by early detection method. The eventual purpose of this project is to build a strong model for the study based on the speech of healthy and ill classification and to establish an accurate research model. In order to analyse the speech data in this study report, different methodologies are used. The classification of the patients will be done using samples from the obtained dataset for the vowels to recognise the voice behaviour of the healthy and unhealthy persons, and three types of approaches will be followed, one is 1 dimensional CNN, second is the conversion of the audio into Fourier transform, and the third is using CNN 2D.

I. INTRODUCTION

The main global issue arises due to the vocal diseases having an impact on life, and it is affecting newborn children. The disease of the vocal infection has an impact on the social, physical and professional life of the patient. Dysphonia impairs interaction, which has a negative influence on one's social, economic, and physical well-being [1]. When the research about this disease was conducted then, it was found that it is increasing with a rate of 2.5 percent for people having age over 70 years, and the rate of infection in a normal person is 2.5 percent. As far as the disease is concerned, the main effect it has on the old age peoples and on the newborn children [2]. Individuals frequently miss dysphonia, resulting in undetected cases.

Several recent research articles on the subject of picture classification using deep neural networks, notably Convolutional neural networks, have shown some excellent discoveries [3]. CNN is often used in non-image datasets for the incorporation of diverse datasets for pictures. The voice data produced when we speak actually a graph of amplitude of our voice with respect to the time, and this simple 1D data, as explained, can be converted into the 2D model, which can be a spectrogram; this spectrogram can help the algorithm to classify the disease according to the analysis. 2D CNN models are commonly used for dealing with picture categorisation problems. The main aim of this project is to classify the patients according to their voice data received by comparing the data from the classification model using different techniques of machine learning and deep learning.

The quality, performance of the data and the type of information of the data can be used to analyse the effectiveness and the productivity of various data, and the approach can be any; it can be a deep learning approach or a machine learning approach. Machine learning is so organised and intelligent that for every task or the analysis of any data, there is a unique algorithm, so it cannot disturb the other one. Every learning algorithm has unique work, aim, and results and cannot be related to another algorithm. In this paper, some of the algorithms will be used for the classification, and this algorithm can be Fourier transformation. 1D and 2D CNN model algorithm. A range of machine learning-based classifiers capable of identifying disordered voices have been used in a number of recent studies to discriminate between normal and abnormal sounds. To date, a network of deep neural networks has achieved the highest abnormal speech detection accuracy of 99.32 percent [4].

For better results and the evaluation of the performance of the dataset, a single approach is not so much, so different approaches of the machine as well as deep learning are utilised. Finding a decent learning algorithm from these two approaches for a certain domain target application is not straightforward, but it is more difficult. The design which is proposed in this paper will benefit the healthcare and insurance companies as well by early detecting and examining the patients. This paper offers a new approach for identifying and classifying sick people based on their speech. Once placed on a real-time device, the health score for the healthy and unhealthy persons will be provided by analysing the patterns of speech. The higher the person's healthiness score, the healthier they are, and vice versa.

Different attributes of various persons through the dataset can be analysed through these techniques, which will help in improving the main model's accuracy and efficiency. In this work, we offer a new approach for identifying and classifying sick people based on their speech. Fourier transformations are often used to extract time, frequency, Mel-frequency cepstral coefficients (MFCCs), and a variety of frequency domain-related spectrum properties from voice data. The main goal of this research and development is to conduct extensive research and analysis on a variety of techniques, including the basic and old techniques of converting raw data into Fourier transform, CNN 1D, CNN 2D, and others, which will help the unit of healthcare in recognising the healthy and unhealthy persons based on the analysis of their speech. The main reason for choosing this topic is to design a new strategy for the analysis and detection of different diseases from the speech by analysing the plot by using the dataset and then by converting the speech into a spectrogram or other model from which anyone can check the vulnerability related to the disease.

The remainder of the paper is arranged as follows: The next part provides an outline of the report's major body, as the II section discusses a literature review of deep learning and machine learning for voice-based patient categorisation. Section III explains the project's approach, including how the research began and what the many avenues for the study are, as well as how to make the necessary development. Section IV discusses the project's analysis, including its categorisation and kinds and descriptions of data sets, as well as an outline of the suggested procedure. The outcomes, as well as other debates on the understanding and discoveries, are discussed in Section V. The study's general findings and future recommendations are discussed in Section VI.

II. RELATED WORK

Artificial intelligence has risen in popularity in recent years, as technology strives to mimic human intelligence. Machine learning has piqued the interest of most computer scientists since the 1950s [5]. As a result, the machines' expectations are rising, and deep learning initiatives in this area are gaining steam in the aftermath of the COVID-19 outbreak. In changing the world toward the global village using different techniques of artificial intelligence, the main growth in the sector is actually due to the model of machine learning and deep learning, and one of the main advantages of artificial intelligence is also that it can analyse and compute a large dataset, allowing products to function intelligently [6]. One of the most popular algorithms in the fourth industrial revolution is actually the machine learning model. "Industry 4.0" is actually a futuristic approach of different systems when every system is so intelligent that it can be tuned, configured, troubleshoot from anywhere, and it also allows other systems to learn from the experience of their datasets. The main types of artificial intelligence involve modern smart systems, such as smart locks, homes, industries and any information processing. It is necessary to evaluate the data and design the algorithms effectively.

When we talk about the description of sound pathology in the sense of data science, then it can be seen that it is actually a classification issue detector that extracts information from audio signals. To detect sound pathology, a variety of machine classifiers are frequently utilised. Several methods for better distinguishing healthy and sick sounds have been created in order to improve the efficacy of various procedures. These techniques focus on identifying sound quality standards and creating new tools for diagnosing voice disorders [7]. The decision-making system is based on neural networks. Laura Verde et al. [8] are examining and evaluating the performance of a variety of machine learning algorithms for detecting voice disorder. An SVD database is used in this scenario for the analysis of a sample of voices from healthy and unhealthy persons. The authors describe a unique Convolutional Neural Network (CNN)-based method for identifying aberrant speech [9]. This system makes use of parts of the SVD database. To diagnose speech diseases by combining audio recordings. A high level of the authentication system is used in a world everywhere at every institute, office for providing contactless security to the users is known as face recognition using artificial intelligence [10]. The main problems that arise nowadays are due to the need of the method of analysis due to the irregular behaviour of the spike of the pandemic, so the sector of contactless doors, locks, banks, security systems are widely increasing in the world as the need of the evolution of the world. The constant effort for the research community is actually starting new, future subjects with the existing ones for the proper identification and analysis of early detection of various types of disease[11]. The graphic below shows the relationship between AI, machine learning, and deep learning.

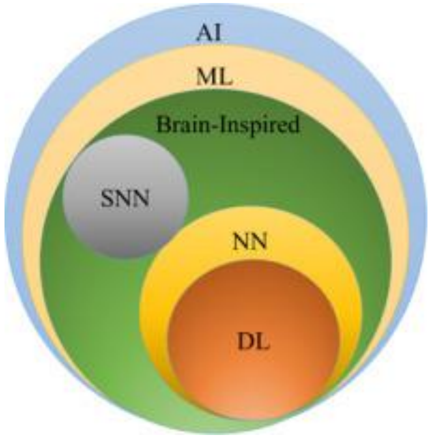

Figure 1. Relation between AI, ML and DL *[12]*

This study uses NN (Neural Network) and SVM (Support Vector Machine) to diagnose voice disorders by combining the American and German databases MEEI and SVD [13]. Classification is a key way for analysing various machine learning approaches that relate to prediction-based modelling in which it acts as the classification method that can also be utilised. A classifier is used to transfer data from the source to the destination and can be used in the destination, identifier medium, and the transport function whenever there is a need to transfer input to the output variable. Spam screening, such as "spam" and "not spam," in the service providers of email might be a categorisation issue. Binary classification refers to the model of classifications that can only use two states of the identifiers [14], such as "true or false" or "yes or no".

The study [15] considers the use of a laryngophone as an additional treatment approach for the identification and classification of speech data into distinct subunits using phonetics techniques. The method depends on throat correlation acoustic coefficient, and utilisation of different methods for automatic speech signal segmentation can also be analysed through this technique, and it can be simply used for speech segment classification. The most significant method in identification jobs is voice pre-processing. When acoustic data is transferred to the frequency domain, speech sounds fluctuate swiftly and dramatically in the time domain; nevertheless, the spectrum that is employed is plainly observable. These strategies will assist in separating the signals to these frames and their usage in the window functions in order to enhance the integrity of the diverse voice signals [16]. This method is actually used for the conversion of different spectral energy data into data units so that the Mel filter can easily analyse it and the features of the Mel filter involves the frequency of 16 cepstral, and the range also includes 300-800Hz.

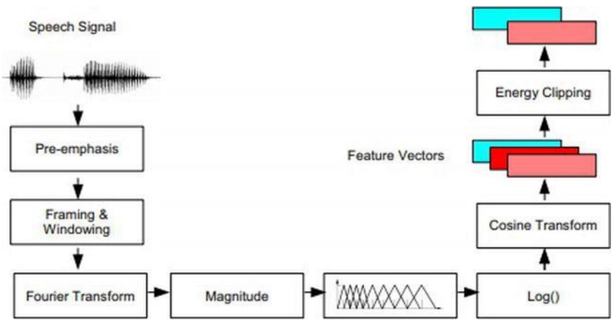

Figure 2. Voice Classification

Toleu and his colleagues [17] recommended MLP and LSTM models with characters for the separation of the credential at the same time and for recognising different sentences. These proposed models projected the characteristics' embeddings into a directly adjacent dimension to provide facilitate growth characteristics that could help with the speech signal. These combinations were tested in three languages: English, Italian, and Kazakh. According to the studies, individual LSTM and MLP types of algorithms can be used for phrase and token categorisation in different economic and professional systems. The researcher [18] has used the SVD dataset for the analysis of the diseases for the early detection method. The main part consists of this type of dataset is actually the data for the healthy and unhealthy persons.

III. METHODOLOGY

When we discuss sound pathology in the context of data science, we can see that this is essentially a categorisation problem detector that extracts data from audio signals. To detect the issues related to the sound, a variety of machine classifiers are frequently utilised. There are numerous methods for distinguishing between healthy and sick sounds that have been created in recent years and assessed based on performance. These strategies focus on creating new instruments for identifying voice problems and defining measures for measuring sound quality. For the evaluation of the audio waveform of the speaker, a technique called speaker recognition is used, and the approach actually depends on the speaker, and the recognition of characteristics involves the behaviour and phycological information of the person speaking. The process of working and the relation of the input/output is also the same for the machine as well as deep learning models. Machine learning techniques including k-nearest node, support vector machines, decision trees, random forests, naive Bayes, linear regression, association rules, and k-means clustering are widely utilised in a variety of applications.

The dataset to be used (Saarbrucken Voice Database) comprises a range of data for the voice that also includes the recording of voice from healthy as well as unhealthy persons, and it also includes ECG signals. For the extraction of different characteristics, just as density and centroid of the spectral from the audio data and then save that data into the excel file can also be utilised by the use of Fourier transform. Following that, we'll experiment with various exploratory data analysis, feature selection, and extraction strategies before training a variety of standard learning algorithms of machine models such as Decision Tree, SVM, or ensemble approaches such as Random Forest Classifier. The Saarbrucken Voice Database is used, which was recorded and made public by the Phonetic Institute of the University of Saarland in Germany. The Saarbrucken Voice Database is utilised, which was captured and made public by the University of Saarland's Phonetic Institute. The SVD is a collection of healthy and disordered people's voices. There are 687 healthy voice recordings and 1354 ill voice recordings accessible, representing 71 diseases. The vowels, as well as the sentence "Good morning, how are you doing?" and the database actually depend on this. (How is your day going for you?) a German expression to categorise healthy and ill sounds, we used only the /a/ vowel and all of the entries accessible in the SVD database. The voice ID contains information such as age, gender, and category for each individual.

For the evaluation of the trials, the entire dataset with a balanced system has a complete dataset of 685 males and females. The basic requirement for this analysis is the need for comprehensive domain knowledge. The main reason for this dataset chosen is due to the fact that this dataset has a large number of databases from where every activity of the individual can be analysed improved as well. The data collected from the database can be easily utilised for the complete system of voice classification. Using Fourier transforms, the raw voice data would be turned into tabular data, and the dataset may then be cleaned and preprocessed using different approaches. Missing value imputation, incorrect data type elimination, and other data cleaning processes are all available. We can employ ANN or CNN 1D models, which can handle audio signals/data if this strategy fails to give adequate accuracy. Another approach for the system can be achieved by converting the audio waves into a 2D image of the spectrogram, and if a large number of images of spectrograms are created, then it can be helpful in the development of a classification model. The same dataset has been used previously in different researches, and different projects have also been completed due to this dataset, such as the Hidden Markov model, Gaussian mixing model, and so on. The main logical flow of the model is shown below:

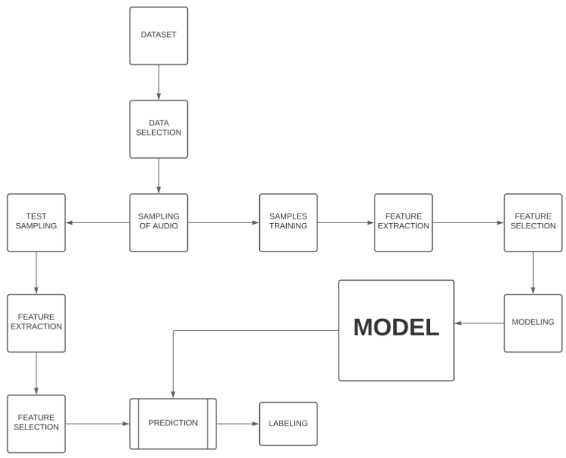

Figure 3. Main Flow

The extraction stage is usually accompanied by decision-making procedures like detection, categorisation, and knowledge fusion in a traditional audio processing strategy. For the analysis of the different views of sick and healthy persons using the classification model and for checking the abnormal vibrations or any spikes of audio waves can be used for the classification purpose, and this is the basic purpose of this project. This system can extract spectrum parameters such as spectral centroid and spectral density from the audio input and analyse them, then export the results as a CSV/excel file. Following that, analysis of the data, feature selection and extraction, and training of classic machine learning models such as Decision Trees and SVMs, as well as ensemble approaches such as Random Forest Classifiers, could be used to classify the mode.

CNN 1D model can be used for extracting information from audio signals if the first method of the learning algorithm fails or it does not give a better result. The difficulty of discerning unique sequences of data from presented data is the problem of identifying human bodily health. Examining and then using data that is simply based on fixed-sized windows and the data presented is time-series data, as well as using such learning models to enable decision trees, are traditional ways for solving the problem. For the feature extraction of data from the CNN 2D model, the main need is for the proper technical knowledge of the CNN model. Converting audio to spectrograms, recording multiple spectrogram images, and using CNN 2D to develop an image classification model are the final steps in analysing the speech. For the visualisation of the intensity of various voice date signals in a waveform across time and at various frequencies, a design is used called a spectrogram. Not only can you check because certain frequencies, such as 2 Hz vs 10 Hz, have more or less energy, the fluctuation of the energy level over time can also be checked by the spectrogram analysis. A heat map, which is a picture in which the intensity is portrayed by altering the colour or brightness, is a typical depiction of a spectrogram.

IV. ANALYSIS

In the German Saarbrucken Voice database, there is data of 2000 people having details of ECG data and different recordings of voice, including 687 healthy adults (259 men and 428 women) and 1356 sick patients and the complete study actually depend on this dataset. The recorded length of these vowels are approximately 1-2 seconds long, and these are recorded at a frequency of 50KHz with a resolution of 16-bit. In the dataset, each participant has recorded the following files:

- Neutral, low, high, and medium pitches of the A, I, and u vowels.
- The German phrase "Good morning, how are you doing?" translates to "How are you doing?" in English.

After manually selecting the recording sessions on the database website, different datasets of the specimens of voice and it also includes a signal of ECG which can be used in the system for the proper classification of the model and the data for the speakers are also shown below in the tabular form.

COLUMNS	DESCRIPTION	DATATYPE
ID	Unique identity	Numeric value
T	A healthy or unhealthy person	Characters Sting
D	Date of recording	Time and Date
S	The ID of the speaker	Numeric value
G	Gender	Characters Sting
A	Age	Numeric value
Disease	Type of the disease	Characters Sting
Comments	Remarks of disease	Characters Sting

In the sense of the data analysis for the classification model of the dataset, the feature extraction proves to be the best in every way as it provides the complete overview and the features of different spectral issues. Speech classification methods of machine learning are also utilised here for this extraction. After determination of the raw features in terms of MFCC, the quantitative metrics such as kurtosis and some other variables are also computed across vowels. The original data is then used to produce transformed characteristics. The following seem to be the specifics:

- The age of each topic.
- Gender of each subject
- Raw feature means and standard variations computed throughout each record's epochs
- Labels that define each record as healthy or abnormal are used as labels throughout trials.

This study's database has a significant number of ailments with brief records. As we identified in the related articles, no other research has applied deep learning algorithms for speech pathology diagnosis, despite past efforts. On the basis of the feature extraction of different voice databases, this research is actually decided by taking into the concern of this model. We use a simple method of converting into CSV format for speech pathology detection and identification, then work with the approach of the CN1D model, then use the CNN2D approach to provide further reasoning.

The extraction step is usually followed by decision-making in a conventional audio processing method; techniques such as detection, classification, and knowledge fusion are used (Hussain M, 2018). The actual work involves the comparison of different results of the healthy and unhealthy person spectrograms, Fourier transform plots and checking some other properties for early detection of diseases. Basic features are determined in terms of MFCC after conversion to an excel file, and statistically, factors like kurtosis, swekness, and other properties are then evaluated over vowels using different pitch values. After that, processed features were developed based on the raw data.

For verifying and assessing the appropriate working of the dataset, we first try classification of the KNN and then use the classification of the decision tree, random forest, XGBoost, and Adaboost can also be utilised. These all-classification approaches are applied to the dataset and produce results based on their characteristics. For the feature extraction of data from the dataset, we used all three usage approaches. The first approach is a simple method in which the voice is changed into spectral characteristics in the same way as a CSV file is transformed, and after the conversion, the classification algorithm will be placed on it, which will work then and provide the result based on the feature selection. To conduct classifications, we follow the first technique and employ the dataset's CNN 1D and 2D feature extraction algorithms. Each model outperformed the preceding one, and one of them delivers the best solution that may be used as the end outcome, as stated in the following section.

V. RESULTS

Since there are three main models chosen from the information by treating all values of pitches as high, low, and medium pitch values, the model is built using different values of pitch and different vowels. The first model, A, is generated by picking the highest pitch value for the voice-based classification of healthy and sick patients from the dataset. The provided voice will be transformed to an excel file at first, following which a filter would be used to the information, followed by several classification techniques, as well as the findings would be evaluated appropriately. The model is also created with the vowel E in mind and a high pitch value, as well as the vowel A in mind and a low pitch value. The dataset's age distribution for distinct people is first examined and then displayed.

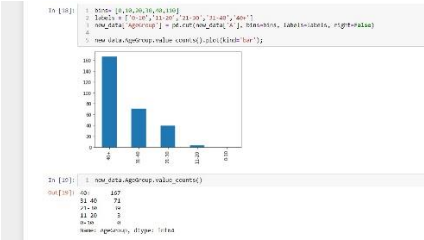

Figure 4. Age distribution

Below is a graph of time vs frequency for healthy patients, with vowel A and High pitch is taken into account. By checking from the figure below, the accuracy of the model can be analysed, and the accuracy is directly proportional to the pitch, and if the pitch increases, the accuracy increase. The values at zero on the y axis are normally ignored in the spectrum display, and level 0 frequencies can be filtered out of the final design.

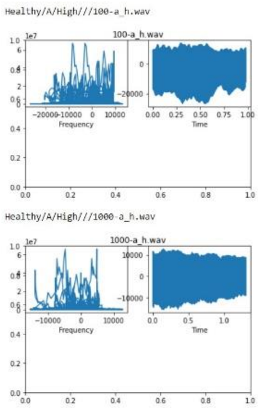

Figure 5. Healthy Person Model-1

The graph below demonstrates the relationship between time and frequency in sick people using a technique that incorporates vowel A and low pitch. The model's accuracy reduces as the pitch drops, as illustrated in the figure below. The values at 0 just on the frequency display's y-axis are often low, and all these values can indeed be processed out of the finished product, rendering the approach useless. As can be seen below, the time graph has a quick transition from a high to a low value, making it unsuitable for analysis in the lower plot, while the top plot has abrupt shifts, making it unsuitable for classification.

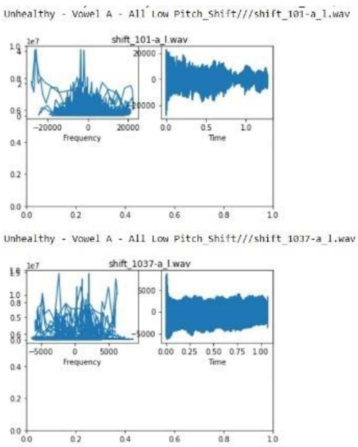

Figure 6. Unhealthy Patient time vs frequency Model-1

The next stage is to put these models through their paces with CNN2D, a deep learning algorithm, after putting them through their paces with machine learning methods. To create a spectrogram from the processed sound stream, Mel frequency cepstral coefficients (MFCCs) is used. Investigators used pre-emphasis, windowing, fast Fourier transform, filtering using Mel, nonlinear treatment, and discrete cosine transform to extract features from MFCCs. Below is a spectrogram for healthy persons that use the vowel A with high pitch paradigm; The graph below is highly smooth, with zero fluctuations inside the spectrogram, as could be observed in the figure, and the brightness is undoubtedly the key premise in verifying the consistency of spectrogram operation, trying to make this method ineffective for this categorisation.

```
[1]:
import matplotlib.pyplot as plt
from scipy import signal
from scipy.io import wavfile

sample_rate, samples = wavfile.read('Healthy/A/High/1-a_h.wav')
frequencies, times, spectrogram = signal.spectrogram(samples, sample_rate)

plt.pcolormesh(times, frequencies, spectrogram)
plt.imshow(spectrogram)
plt.ylabel('Frequency [Hz]')
plt.xlabel('Time [sec]')
plt.show()
```

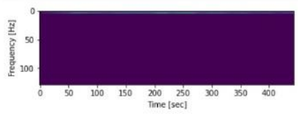

Figure 7. Spectrogram for a healthy person

It can be inferred that the pitch is highly essential in every situation based on the responses from all three models mentioned above for healthy and unwell people. Afterwards when, certain operations on the incoming voice are conducted, such as volume 0 filtering of frequency, normalisation of the signals, and ultimately implantation in the high pitch model for final testing, after which these three models are averaged out by taking one vowel and three pitches. The graph below illustrates that the method is substantially more exact irrespective of if the voice's averaged value is 0 or 1. The final model's output is shown in the diagram below.

	precision	recall	f1-score	support
0	0.95	0.94	0.95	192
1	0.89	0.91	0.90	102
accuracy			0.93	294
macro avg	0.92	0.93	0.93	294
weighted avg	0.93	0.93	0.93	294

Figure 8. The output of the final model

VI. CONCLUSIONS

As indicated in the preceding sections, this study examines three models using different models from the dataset, and different features of the dataset will also be extracted from this. The highest achievement of this system is due to the voice disease identification process of this system which was achieved by using the average and balanced dataset with the XGBoost classification model. This is a better result than the previous research and experiments' best result from the literature review. Once all samples are considered, the classification technique yields the best accuracy and sensitivity results; however, when classifiers use gradient boosting machine-generated features, SVM yields the greatest specificity results. When feature selection was employed in prior techniques, accuracy, sensitivity, and specificity increased, but the final and right result provided by XGBoost did not. To identify illness in speech sounds, a range of machine and deep learning methods are applied in this work. Using the SVD dataset for the continuous speech classification by using normal pitch gives the result according to the details it has, just as the variety of diseases etc., and all models are created and evaluated according to this approach. The dataset of healthy and unhealthy persons is analysed through different techniques. A balanced average model is also created using these datasets by representing people of different ages and diseases. For asymmetrical and complete data set, the performance of several sound pathology diagnostic techniques, such as Support Vector Machine, Decision Tree, Bayes Classification, Logistic Model Tree, and KNN, is compared, and it is decided that XGBoost seems to be the best choice.

The pitch is crucial in every case, according to the findings. The analysis is carried out for all values of pitches, for medium-low and high as well, and their results are then analysed after using every single pitch and finding out that the high pitch is giving the best results and the low and medium pitches are not so good for the model. Regardless of the normalised value of the signal data, it can be 0 or 1; the algorithm which converts the raw data into the signal graph is actually the best approach among all of the approaches with high pitch giving the best result. The suggested model's weighted and maco average is also 93 percent accurate, as is the final model that was developed and tested. Precision, recall, and F1-score are similar in that their findings are great for sickness classification.

Acknowledgement

This research could not be possible without the personal and professional help of the supervisor. I'd want to thank Professor in particular for his help. His acts have shown what a good scientist should be. I'd want to thank my parents for their never-ending support and advice throughout my life. They are fantastic role models. My family is deserving of particular mention. My appreciation for all the sacrifices my mum and father has made on my behalf is unfathomable. Up until now, your prayers for me have kept me going. Finally, I thank God, my loving Father, for aiding me in pushing ahead despite the challenges.

References

[1] K. J. R. N. A. C. C. M. Cohen SM, "Prevalence and causes of dysphonia in a large treatment-seeking population," *Laryngoscope,* 2012.

[2] K. A. J. M. Davids T, "Current dysphonia trends in patients over the age of 65: is vocal atrophy becoming more prevalent?," *Laryngoscope,* 2012.

[3] H. T. Y. T. Fujiyoshi H, "Deep learning-based image recognition for autonomous driving," *IATSS,* 2019.

[4] G.-V. P. B.-V. M. Godino-Llorente JI, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans Biomed Eng ,* 2006.

[5] E. B. A. K. S. P. M. Arun, "Detection and monitoring of the asymptotic COVID-19 patients using IoT devices and sensors," *Int. J. Pervasive Comput. Commun,* 2020.

[6] K. A. Sarker IH, "Abc-ruleminer: user behavioral rule-based machine learning method for context-aware intelligent services," *J Netw Comput Appl,* vol. 168, 2020.

[7] A. N. O. M. P. Hussein M. A. Mohammed, "Voice Pathology Classification Using Machine Learning," *ISASE,* 2021.

[8] L. G. D. P. a. G. S. Verde, "Voice disorder identification by using machine learning techniques," *IEEE Access,* vol. 6, pp. 16246-16255, 2018.

[9] H. e. a. Wu, "A Deep Learning Method for Pathological Voice Detection Using Convolutional Deep Belief Networks," *Interspeech,* pp. 446-450, 2018.

[10] B. V. F. J. P. V. K. J. S. Prabu, "Mobile technologies for contact tracing and prevention of COVID-19 positive cases: a cross-sectional study," *Int. J. Pervasive Comput. Commun,* vol. 8, no. 10, 2020.

[11] K. S. N. B. M. Visweswaraiah, "Test mode power computation and IR drop analysis of application specific integrated circuits implementing face detection algorithms," *4th International Conference on Advanced Computing and Communication Systems (ICACCS),* 2017.

[12] S. S. Shinde P. P., "A review of machine learning and deep learning applications," *Fourth International Conference on Computing Communication Control and Automation (ICCUBEA),* pp. 1-6, 2018.

[13] K. e. a. Wu, "Joint learning for voice based disease detection," *Pattern Recognition,* vol. 87, pp. 130-139, 2019.

[14] O. S. T. Y.-W. Hinton GE, "fast learning algorithm for deep belief nets," 2006.

[15] M. V. P. G. R. R. Chaudhari S, "An attentive survey of attention models," *ACM Transactions on Intelligent Systems and Technology,* vol. 12, no. 5, pp. 1-32, 2021.

[16] G. C. C. K. B. Y. Chung J, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[17] W. X. Bu F, "A smart agriculture iot system based on deep reinforcement learning," *Futur Gener Comput System,* vol. 99, pp. 500-507, 2019.

[18] D. Hemmerling, "Voice pathology distinction using autoassociative neural networks," *25th European signal processing conference (EUSIPCO),* 2017.