Contrastive Analysis Of Supervised And Unsupervised Learning Techniques For Voice Pathology Detection And Classification

A PROJECT REPORT

Submitted by

MAYURI M (TKM20MCA-2022)

to

The APJ Abdul Kalam Technological University

In partial fulfillment of the requirements for the award of the degree of

MASTER OF COMPUTER APPLICATIONS

Thangal Kunju Musaliar College of Engineering
Kerala

DEPARTMENT OF COMPUTER APPLICATIONS

JULY 2022

DECLARATION

I undersigned hereby declare that the project report titled **"Contrastive Analysis Of Supervised And Unsupervised Learning Techniques For Voice Pathology Detection And Classification**", submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under the supervision of **Prof.** Jasmin M R**.** This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed as the basis for the award of any degree, diploma or similar title of any other University.

Place: Kollam

Date: 12/07/2022 MAYURI M

DEPARTMENT OF COMPUTER APPLICATIONS

THANGAL KUNJU MUSALIAR COLLEGE OF ENGINEERING

Certificate

This is to certify that, the project report entitled "**Contrastive Analysis Of Supervised And Unsupervised Learning Techniques For Voice Pathology Detection And Classification**", submitted by **MAYURI M (TKM20MCA-2022),** to the **APJ Abdul Kalam Technological University** in partial fulfillment of the requirements for the award of the Degree of **Master of Computer Applications,** is a bonafide record of the project work carried out by her under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor Head of the Department External Examiner

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. First and foremost, I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I am extremely grateful to ***Prof. Dr. Fousia M Shamsudeen***, Assistant Professor and Head of the Department, MCA, TKMCE, for her constant support and encouragement throughout the project work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Prof. Jasmin M R**, Department of Computer Applications, for her expert guidance, co-operation, and immense encouragement. I also extend my thanks to the entire faculty and staff of the Department of Computer Applications, TKMCE, who have encouraged me throughout my course of study.

I also express my thanks to my family and friends, for their support and encouragement in the successful completion of this project work.

MAYURI M

ABSTRACT

The development of technology makes it possible to offer better solutions to the complicated issues that people encounter. The early identification, treatment, and ongoing monitoring provided by today's smart healthcare sectors are crucial in lowering hospital visits, travel expenses, and waiting times.A medical condition known as voice pathology affects the vocal chords and makes it difficult for the patient to speak. As a result of this, the patient may experience difficulty communicating. A study that was only recently presented found that vocal pathology detection systems are capable of accurately diagnosing voice pathologies at an early stage.These systems made use of machine learning strategies, which are regarded as particularly reliable instruments for identifying speech disorders. However, the majority of suggested algorithms for detecting voice disorders used small databases.The low accuracy rate continues to be one of the most difficult problems for these approaches. A technique for identifying voice pathology is described in this research paper.Utilizing the Mel-Frequency Cepstral Coefficient, the voice features are retrieved (MFCC). Vowel /a/ speech samples were equally obtained from the Saarbrücken voice database (SVD). As assessment indices, accuracy is used to compare the effectiveness of various machine learning classifiers.

The voice signals in this work are classified as either healthy or disordered using a CNN architecture.

CONTENTS

List of Figures

List of Abbreviations

- SVD – Saarbrücken voice database
- MFCC - Mel-Frequency Cepstral Coefficient
- OSELM - Online Sequential Extreme Learning Machine
- LDA - Linear Discriminant Analysis
- MEEI - Massachusetts Eye and Ear Infirmary
- DT - Decision Tree
- RF - Random Forest
- GMM - Gaussian Mixture Model
- DCT - Distinct Cosine Transform

Chapter 1

INTRODUCTION

Early disease identification is being advanced through the healthcare industry application of machine learning techniques.A voice pathology problem, often known as throatiness in speech, is a speech organ defect brought on by a mental illness, injury, autism, or other disorders. The typical vibration pattern for the glottis is impacted by voice disorders or pathologies in the vocal cords, which results in throatiness in the voice.Vocal pathology cannot be diagnosed using conventional medical methods. These techniques rely on the inspection of the voice cords, which might lead to misunderstandings and inaccurate assessments. These techniques are also time-consuming, expensive, and need a variety of equipment.

Machine learning algorithms have significantly improved the early proposed methods for detecting speech pathology, and these algorithms have demonstrated their effectiveness.and efficiency in applications for healthcare.Some of these algorithms, nevertheless, still have issues with poor classification accuracy, lengthy execution times, and unbalanced data; they struggle with heavy workloads or huge datasets in speech pathology monitoring systems. This study uses a varied number of voice samples to propose a voice pathology detection method. The feature extraction and classification processes in the suggested system leverage the MFCC and CNN architectures, respectively.

1.1 Problem Statement

Even though hospitals play a major role in voice pathology detection and classification; current approaches are much more difficult, time-consuming, and cost-effective, thus introducing machine learning algorithms. Mainly three stages are included in this process. First off, because it is balanced, the SVD dataset was used for this study. Preprocessing aids in reducing noise.Using three methods CNN, Logistic Regression, and Random Forest, categorize and compare the voice signals as normal or aberrant.

1.2 Objectives

- The main goal is to create an effective speech pathology detection method using feature extraction.
- Preprocessing of voice signals is done which helps to avoid noise.
- Feature extraction techniques help to remove unwanted noise and balance the time consumption.
- Algorithms are used to estimate the accuracy .
- The voice signal is then classified as either healthy or abnormal using the best model.

Chapter 2

Literature Survey

In today's medical world, voice pathology detection is regarded as a very essential field. Considering that voice disorders affect the majority of the global population.As a part of literature review through various papers and presentations on this topic. A quick summary of findings is specified in this chapter.

2.1 Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine

The system that was used in this research examined several different speech quality measures.For categorization, it additionally employs the OSELM algorithm. The method seeks to identify and categorize cyst, polyp, and paralysis—three common vocal pathologies—from voice recordings. The MFCC method is used to pull features from the speaker's voice in order to analyze it.In this particular instance, the SVD database is queried to obtain voice samples. The OSELM classifier is trained and tested using a huge number of voice samples from this database. Vowel sound is used to categorize voice types. OSELM displays an accuracy of roughly 75%. The OSELM algorithm has some drawbacks; for example, because its input weights are produced at random, the accuracy is inconsistent.

2.2 Voice Pathology Surveillance Systems Based on Internet of Things and Machine Learning Algorithms

Using a feature selection approach of Fisher discrimination criterion, the study in [8] provided a system of speech pathology detection and multi-classification method. This is done to filter out less useful characteristics and keep the more valuable ones.In this study, aberrant speech was identified and classified using the machine learning techniques of SVM,DT, and RF. On the basis of accuracy, specificity, and sensitivity, the suggested system is assessed.It shows an accuracy of 77.5 percent. By concluding all the data, the problem of voice pathology detection can be done using various techniques. And among them, one of the most effective ways is to do a speech analysis, followed by the development of a machine learning model that can categorize individuals' voices as healthy or unhealthy. And learned that the outcome of this model will be very encouraging, while the amount of error that can be expected from it will be kept to a minimum.

This algorithm has some limitations; they are Unbalanced data. It does not perform well on a large dataset with noise and It requires extensive training time.

2.3 Voice Pathology Detection and Multi-classification Using Machine Learning

Another approach is by using the Support Vector Machine algorithm to classify the voices. For feature extraction, it uses the MFCC technique. Additionally, a method known as Linear Discriminant Analysis (LDA) is employed to lessen MFCC dimensionality. [5] elaborates on an autonomous voice evaluation system that uses four separate feature selection techniques. The techniques mRMR, PCA, Relief algorithm, and others are frequently used to analyze sounds and determine whether they are healthy or unhealthy.also LDA .These strategies are used to get rid of the features that appear repeatedly, which in turn helps to lower the size of the initial feature. In addition, various feature extraction methods, such as MFCC, are used in this approach.Within this system, SVM algorithms are utilized rather extensively for the purpose of categorizing voice signals. The People's Liberation Army General Hospital in China, which may be found in China, is where the voice samples were obtained from..

There are a total of 605 disturbed samples and 200 healthy samples at that location. The accuracy of SVM is 70%.

The limitations of this algorithm; Data imbalance, poor performance on a sizable dataset with noise, and lengthy training time

2.4 Voice pathology detection using artificial neural networks and support vector machines

In [9], a four-feature selection algorithm automated voice evaluation system is described. The mRMR, PCA, Relief algorithm, and LDA algorithms are frequently used to assess and classify sounds as healthy or abnormal.In order to reduce the size of the initial feature, these techniques are employed to eliminate the frequently occurring features. Additionally, this method uses other various feature extraction techniques, such as MFCC. In this system, voice signals are often classified using SVM algorithms.At the People's Liberation Army General Hospital in China, a total of 605 samples of disturbed voice were collected, along with 200 samples of healthy voice. Accuracy-wise, SVM is rated at a 70.. Some restrictions apply to this algorithm; Inaccurate and missing values for an attribute

Chapter 3

Methodology

3.1 Proposed System

Three phases can be used to summarize the proposed methodology for the identification and categorization of voice pathology using supervised and unsupervised learning techniques in this study.There are three phases to this process: first, SVD database voice samples are employed; second, The extraction of MFCC features from speech signals is followed by the identification and categorization of voice samples.The recommended system's architecture is depicted in Fig.3.1.

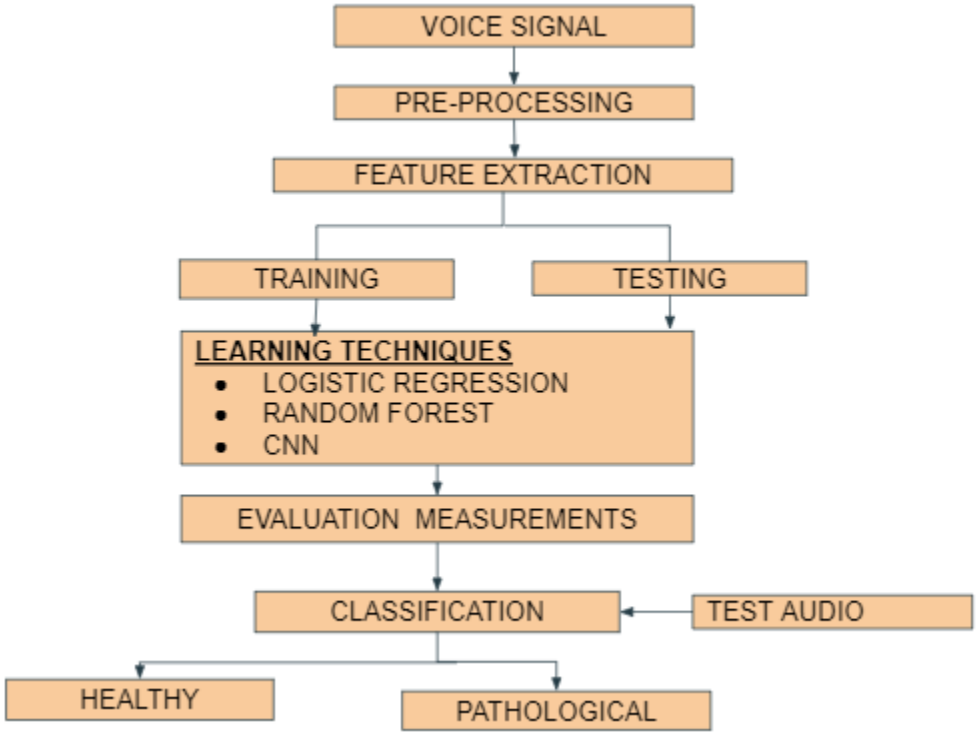


Fig.3.1 The Architecture of the recommended system.

3.2 SYSTEM ARCHITECTURE

3.2.1 The Voice Database

Over 2,000 people's voicemails have been captured and are available for free download in the Saarbruecken Voice Database (SVD). Both sick and healthy voice samples were present in the SVD database. The database includes vowels that make the sounds /a/, /i/, and /u/ when pitched at normal, high, low, and low-high-low levels. A distinct tone was employed for each sustained vowel. At 50 kHz and 16-bit resolution, the SVD evaluated each utterance that was recorded. The pathologic voice samples had a neutral or normal intonation for the vowel /a/. This was carried out to discern between typical voices and strange noises. The disorders from the SVD database that were selected include Dysphonie, Frontolateral Teil Resection, Vox senilis, Reinkedem, Spasmodic Dysphonia, and Leukoplakie.

3.2.2 PREPROCESSING

The speech signal processing step should be completed before the other operations in the preparation section. It involves sampling, quantization, and the conversion of an accompanying analogue signal to a digital signal.

A continuous-time signal is reduced to a discrete-time signal through sampling. Similarly, the discretization of the signal with numerous quantization levels completes the quantization of the corresponding analogue signal. A continuous-amplitude sample is quantized into a discrete-time signal by multiplying the sampled amplitude values by a finite number of levels.

3.2.3 Mel-Frequency Cepstral Coefficient (MFCC)

The properties of the voice are derived using the Mel-Frequency Cepstral coefficient (MFCC). The main processes in the MFCC feature extraction approach are windowing the signal, using the DFT, calculating the magnitude's log, warping the frequencies on a Mel scale, and using the inverse DCT.

3.2.3.1 Pre-emphasis:

The speech stream is then run through a filter that stresses higher frequencies to complete the pre-emphasis process. To put it another way, the following equation states that this action will enhance the signal energy at the subsequent frequency:

$$S0m = Sm - 95 * Sm - 1$$

where: S0m is the new sample value, S is the sample value, and m refers to the sample variance.

3.2.3.2 Framing `

In the frame stage, frames are created from auditory communication. Each frame has a size, Tk = 25 ms, and a frameshift, Tv = 10 ms, which is the interval in milliseconds between the left borders of succeeding windows.The following formula is used to get the sample size, Nw:

$$point = Tk \times rate$$

As a consequence of this, the frame size has been calculated to be 1103 samples, and the frame rate has been set to 44100 samples per second. During this time, the frameshift in samples, which is denoted by Ns, is calculated as follows:

$$Ns = Tv \times rate$$

Consequently, Ns is provided as 441 for this task. All frames that contain only ordinal bits are discarded when auditory transmission is divided into frames.

3.2.3.3 Windowing

A playacting window is applied to each frame of the auditory transmission by taking into account the next block in the feature extraction process chain and merging all of the highest frequency lines. The playacting window equation is represented by the following expression:

$$Wn = 0.54 - 0.46\cos(2\pi n \text{ point})$$

where n is the sample variety and point is the sample's frame size.

3.2.3.4 FFT

The Fast Fourier Transform is the instrument that is used in order to transform each frame of data from the time domain into the frequency domain (FFT). Because the input length for FFT is the next power of two after a point, its length is denoted by the symbol nfft = 2048.

3.2.3.5 Mel Filter Bank Processing

The voice signal is made up of tones at various frequencies. Every tone has a genuine frequency, f, expressed in rate as well as a subjective pitch, evaluated on the Mel scale, as was previously mentioned.Each filter has three frequencies that are considered to be cut-off, each of which has the shape of a triangle, adequate unity in the middle, and a linear reduction to zero in the middle of two adjacent filters.When all of the filters have been applied, the final product will be the sum of the spectral components that each filter has removed. A linear frequency spacing on the Mel-frequency scale can be lower than 1000 Hz, while an exponent frequency spacing can be higher than 1000 Hz.By applying the following equation, you may convert frequencies from mel to hertz:

$$fmel = 2595 \times \log 10 (1 + fhz/ 700)$$

3.2.3.6 DCT

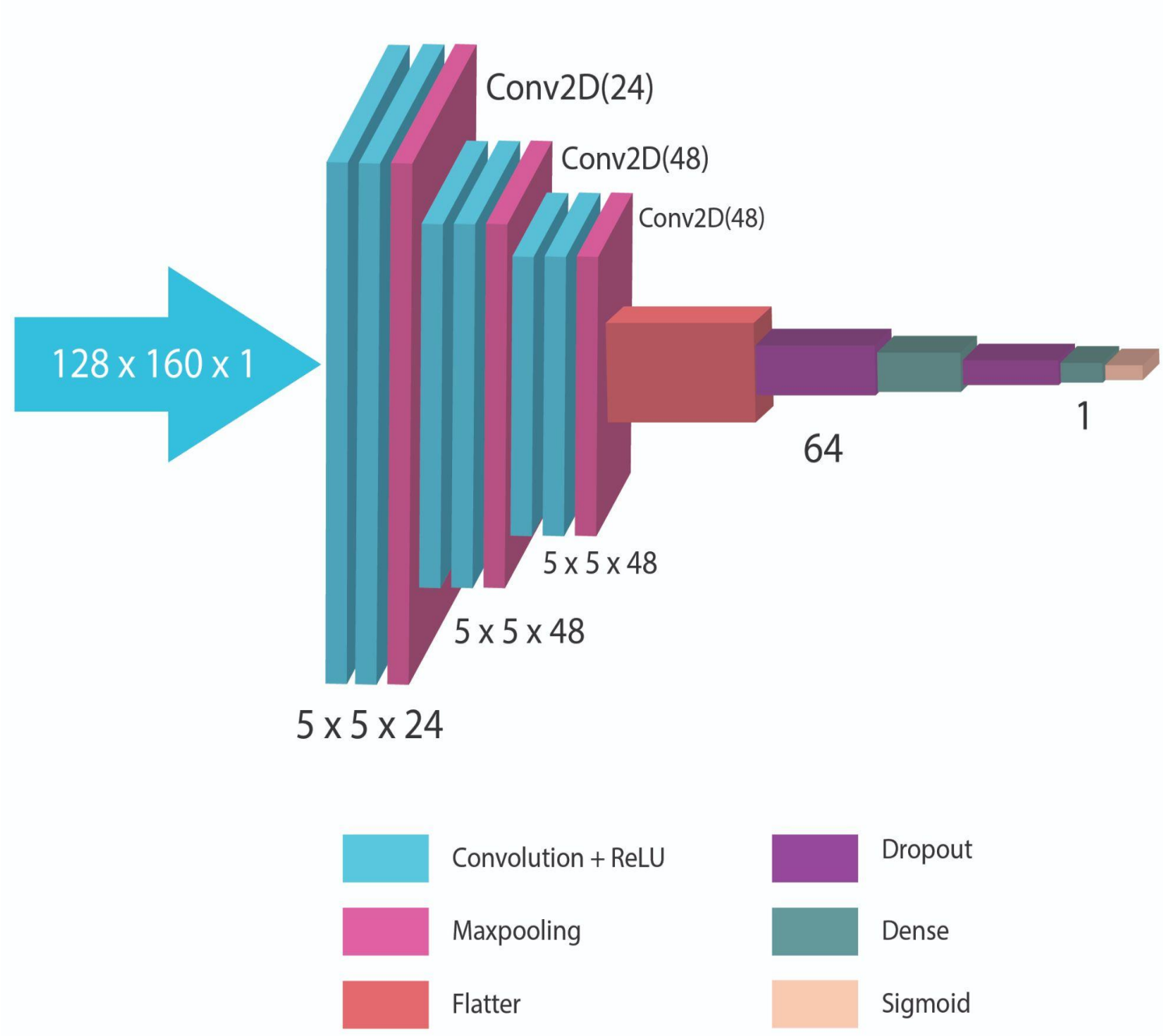
In order to complete the process of transforming the log Mel spectrum into the time domain, the distinct cosine transform (DCT) method is utilised as the very last step.This is done so that the procedure can be completed successfully. Cepstral constants of mel frequencies are utilized in order to provide a description of the results of the conversion.Acoustic vectors are the name given to the set of coefficients. Each input auditory communication is finally transformed into an array of acoustic vectors.

3.2.4. Classification

3.2.4.1 VOICE NET ARCHITECTURE

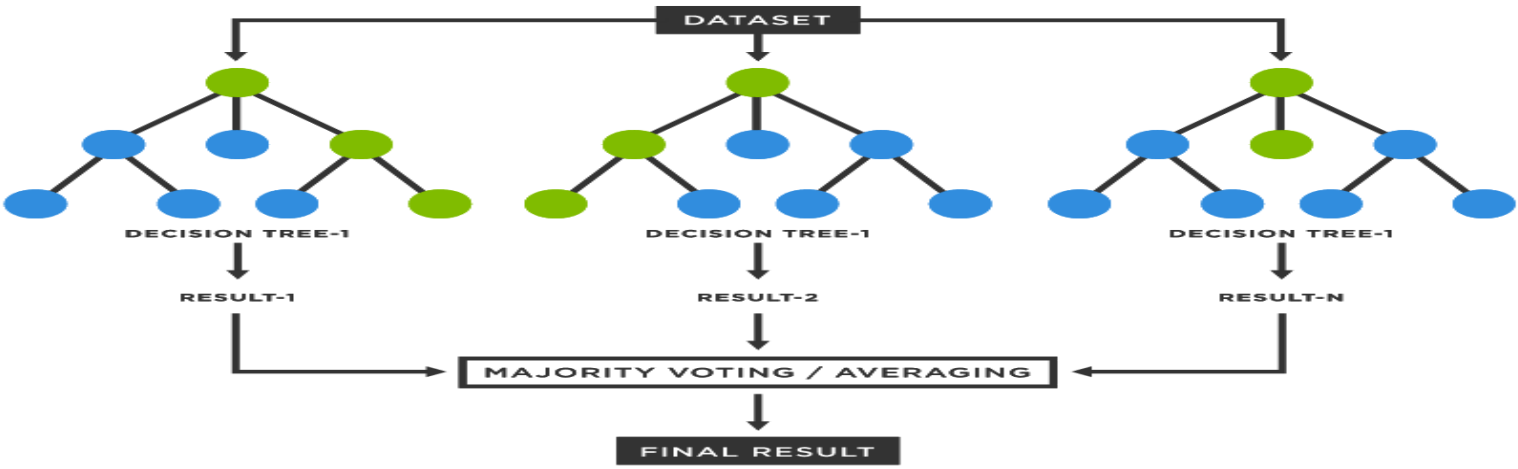
The routing layers and grouping layers that make up the CNN's scale circumstances are specified by a wide range of maps.Similar to this, the convolutional position in the VOICE NET structure's initial stage accepts input position data.The first convolutional block includes a convolutional position " Conv2D " and a " MaxPooling2D" position. The convolutional position uses 24, 5 by 5 pixel kernel size that are applied to each part of the image, returning 24 arrays of activation values called point maps that show where certain features are located in the image.In order to lessen the complexity of the point mappings, the grids of 4 by 2 pixels that span the image have been shrunk down to a single pixel, which shows the maximum activation value threatgrid. The outside pooling position is where this reduction took place.Adding fresh convolutional layers that allow the model to identify further fine-scale patterns.RULE is used as the activation function except the last position. In the third block use a flatten position for converting the 3D image intoonedimensionalarray.For avoiding overfitting and underfitting use powerhouse position of rate 0.5. Add two thick layers with 64 bumps and one knot in the fourth block for converting the dimension into 1. originally, we 'll add the thick position

that makes prognostications for the two classes(mannish or womanish). Because it's a double outgrowth, use a sigmoid activation function rather than softmax that numerous are familiar with.



3.2.4.2 RANDOM FOREST

- An ensemble literacy system for bracket, retrogression, and other tasks called random timbers or arbitrary decision timbers works by building a large number of decision trees during training. For bracket duties, the class named by utmost trees is the affair of the random timber. Instead of relying solely on one decision tree, the arbitrary timber takes the advice from each tree and bases its prediction of the outcome on the maturity votes of prognostications. It's likely that some decision trees will predict the correct outcome while others won't because the arbitrary timber blends numerous trees to predict the dataset's class. But when viewed together, every tree makes the right prediction. Following are two theories for an improved Random wood classifier.
- There should be some factual values in the point variable of the dataset so that the classifier can prognosticate accurate results rather than a guessed result.
- The prognostications from each tree must have veritably low correlations.



III . LOGISTIC REGRESSION

One of the programmes that is employed the most commonly is logistic retrogression, which is one of the ML methods used in the field of supervised literacy.It makes use of a set of independent factors that have already been determined in order to make a prediction about a categorical dependent variable.The outcome of a categorical dependent variable can be predicted using the method of logistic regression. Therefore, the outcome needs to be a value that can be categorized or differentiated. Similar to linear retrogression, but with a different application, is logistic retrogression. Retrogression problems are solved using linear retrogression, whereas bracket problems are solved using logistic retrogression. Examining the link between one or more independent factors and a dependent data variable is the main purpose of a logistic retrogression model.This model is used to create predictions about the dependent data variable.A logistic retrogression may be used, for instance, to predict whether a candidate for public office would succeed or fail in an election or if a high school student would be accepted into a specific council or not. Both of these scenarios involve attempting to forecast the future. Because of these two issues, they are able to have an open conversation with one another.

$$f(x) = \frac{1}{1 + e^{-(x)}}$$

3.3 Software requirements and specification

- Develop a user interface
- connect the UI with the model.

3.3.1 Hardware Requirements

- Core i3 processor from Intel
- 512 GB of hard disc space for storage
- RAM memory: 4 GB

3.3.2 Software Requirements

- Operating Systems: Windows and Linux
- Platform: Google Collaboration and Python
- Librosa, pandas, matplotlib, numpy, and sklearn were the libraries used.

➔ **Platform Used**

3.3.2.1. PYTHON

Python, an object-oriented programming language, was initially developed by Guido Rossum in the year 1989. It is an excellent choice for rapidly prototyping complicated applications due to its suitability. It can be compiled in either C or C++ and provides interfaces to a variety of operating system calls as well as libraries. Python is a widely used programming language, and notable users include the National Aeronautics and Space Administration (NASA), Google, YouTube, BitTorrent, and others. Programming in Python is utilized extensively in a variety of cutting-edge areas of computer science, including artificial intelligence, natural language processing, neural networks, and many more. In the late 1980s, Guido van Rossum developed the high-level, open-source programming

language known as Python. Python is the language that is used to operate the system at this time.The basis for software construction.It originated with the ABC language, which he co-created early in his career. Python is a potent language that you may use to create GUIs, create web applications, and create games. It is a sophisticated language. In Python, reading and writing statements differ greatly from reading and writing statements in standard English. Python programmes must first be processed because they are not written in a language that is machine readable. An understood language is Python. This indicates that each time a programme is executed, its interpreter reads the program's code and translates it into byte code that can be read by a computer. Python is an object-oriented language that allows programmers to create and execute programmes by allowing users to manage data structures or objects. In reality, Python has top-notch everything. In Python, all classes, data types, functions, and methods are treated equally. Programming languages are developed to meet the needs of users and programmers for an effective tool to construct programmes that have an impact on people's lives, way of life, economy, and society. By boosting communication, productivity, and potency, they contribute to the improvement of life. When a language falls short of expectations, it dies and is replaced by a more powerful language, which makes it obsolete. Artificial language used in Python programming has withstood the test of time and has remained popular among programmers, consumers, and users of all backgrounds. It is highly suggested as a major programming language for individuals who wish to get started and learn programming because it is a living, thriving, and incredibly beneficial language.

3.3.2.2. Google Colaboratory

Google Collab was created to give anyone who requires access to GPUs and TPUs for building a machine learning or deep learning model free access to them. Google Collab is a more effective alternative to Jupyter Notebook. A programme called Jupyter Notebook enables editing and use of Notebook documents. Both a web browser and an Integrated Development Environment can access it (IDE).It will function with Notebooks rather than files. Along with text, photos, figures, tables, graphs, equations, and a variety of other graphical data, notebook pages can also contain executable lines of code. Simply said, writing executable documents that are human-readable is possible with Notebook documents. A Notebook's building blocks are called cells. A notebook's whole contents are made up of cells. There are two categories:

• Text box: Text, photos, links, and a lot more can all be included in a text cell. Double-clicking a text cell will allow you to change what is inside. Markdown is supported in the text cell.

• Code cell: The executable code is contained in a code cell. The run button to the left of a code cell allows you to run the cell's code. The results of a cell run are shown underneath the cell.

3.3.2.3 GRADIO

Using the open-source Gradio Python module, you can rapidly develop flexible and easy-to-use user interface (UI) elements for your machine learning (ML) model, any API, or other subjective capability with only a few lines of code.Additionally, you can share the URL with anyone or embed the GUI right into the Python notebook.

3.4 Functional Requirements

All of the actions or procedures that the suggested system must carry out are included in the functional requirements.It includes

• sklearn: sk learn (formerly sci-kit learn and sometimes called sk learn) is a machine learning library that can be used in python programming language. By using this library, we can implement various regression,

classification and clustering algorithms such as random forest, support vector machine, k-means and DBSCAN. The sklearn library is built in a way that it can work with various scientific and numeric libraries of python such as scipy and numpy.

• matplotlib: It's used for the visualization of data in python programming language. It's implemented to work with the wider scipy stack and it's built on numpy arrays. It's a multi platform data visualization technique. It was developed in 2002 by John Hunter. Visualization is the most efficient way to understand the data. Using this library, we can represent our data in various plots such as line, bar, histogram, scatter etc.

3.5 Non Functional Requirements

3.5.1 Performance Requirements

- Accuracy: The system should retain accuracy in its operation and the degree of user-friendliness.
- Speed: The system must have the ability to provide speed.
- Low cost: This system is easy to use and relatively inexpensive to implement.
- Less Time Consuming: It takes far less time than the current system does.
- User-Friendly: The proposed system is very user-friendly and makes it possible to produce high-quality environment

Chapter 4

Result And Discussion

The suggested study made use of all words and vowels containing the sounds /a/, /i/, and /u/ that were created in various pitch intonations from the SVD database, as well as voice samples from healthy and ill classes (i.e., normal, low, and high). The characteristics of these vowels, which are created without the assistance of oral organs, clearly demonstrate how the human larynx functions. More than 71 different ailments were represented in the voices. The frame lasts 25 ms, however the audio signals last one second. The MFCC technique is used to extract the features of voice samples. The CNN classifier will then receive these features and categorize the voice stream as a result. All tests are done on a machine with a 2.50 GHz processor, 6 GB of RAM, and a 1 TB hard drive known as the Google Colaboratory server. The correctness of the proposed method is assessed.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Whereas TN (True Negative) denotes that the suggested method denotes the true diseased sample as pathological, and TP (True Positive) denotes that the suggested system designates the true healthy sample as abnormal; False Positive (FP) indicates that the suggested system misidentifies the true healthy sample as pathological, while False Negative (FN) suggests that the suggested system misidentifies the true diseased sample as healthy.

The results of the research suggest that the proposed method using the CNN classifier had the highest accuracy when compared to other algorithms like logistic regression and random forest.The accuracy of the other two algorithms, logistic regression and random forest, is 0.56 and 0.74, respectively, whereas CNN has the maximum accuracy of 80%. These findings show that the suggested CNN approach can distinguish between sick and healthy voices. In terms of accuracy in speech pathology identification, Fig. 2 shows a bar graph comparison of the suggested method with CNN, LOGISTIC REGRESSION, and RANDOM FOREST methodologies. The proposed method, which uses the CNN algorithm, outperformed them all and had the highest level of success in detecting speech pathology.

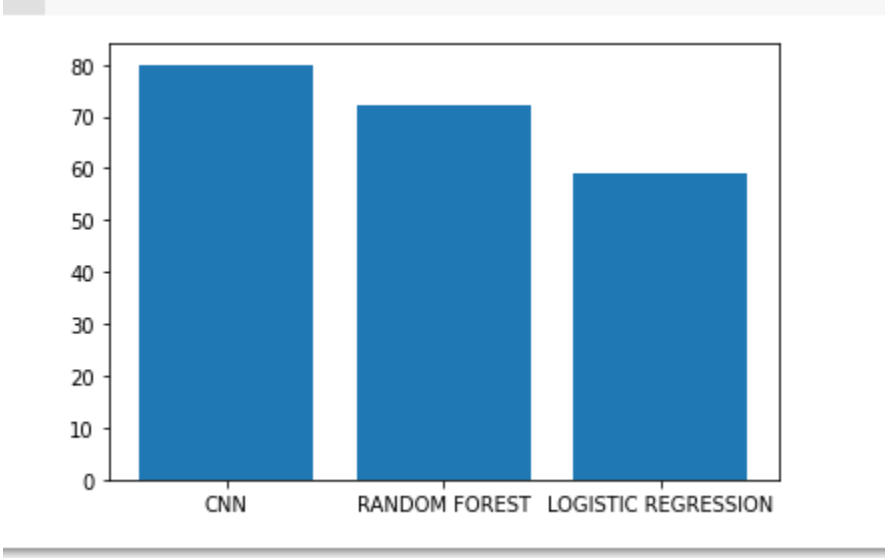


Fig .4.1 illustrates the bar graph comparison between the proposed methods

CNN

```
10/10 [==============================] - 0s 29ms/step
              precision    recall  f1-score   support

           0       0.76      0.72      0.74       121
           1       0.82      0.85      0.83       178

    accuracy                           0.80       299
   macro avg       0.79      0.78      0.79       299
weighted avg       0.79      0.80      0.79       299
```

Fig.4.2 Accuracy obtained using CNN

LOGISTIC REGRESSION

```
              precision    recall  f1-score   support

           0       0.36      0.12      0.17       121
           1       0.59      0.86      0.70       178

    accuracy                           0.56       299
   macro avg       0.47      0.49      0.44       299
weighted avg       0.50      0.56      0.49       299
```

Fig 4.3 Accuracy obtained using Logistic Regression

RANDOM FOREST

```
              precision    recall  f1-score   support

           0       0.66      0.72      0.69       120
           1       0.80      0.75      0.78       179

    accuracy                           0.74       299
   macro avg       0.73      0.74      0.73       299
weighted avg       0.74      0.74      0.74       299
```

Fig 4.4 Accuracy obtained using Random Forest

Chapter 5

Conclusion

The system of speech pathology discovery proposed in this project was based on a variable quantity of voice signals. The MFCC method is used in the suggested system to uproot the voice signal characteristics.The CNN algorithm may make use of either of these traits to assess whether the voice signal is healthy or abnormal. In addition, all phrases and vowels that were articulated with distinct pitch intonations, such as /a/, /i/, and /u/, were incorporated into the SVD database (i.e., normal, low, and high).. We employed a large number of speech samples from both healthy and sick students in order to estimate the Speech Net Architecture of CNN in this work. It is also possible to estimate the correctness of the proposed system. In contrast to other algorithms such as logistic regression and random forest, the trial results have demonstrated that the CNN classifier is appropriate for the task of differentiating pathological samples from healthy samples, with an accuracy of 80 percent being the highest that could be achieved. The accuracy of the other two models is as follows: Random Forest has a score of 74, while Logistic Regression receives a score of 56.

5.1 Advantages

The main merits of proposed model are:

- It classifies the voice more accurately.
- The Human resource needed for the classification can be saved.
- Need less execution time to classify the voice.
- It is an ensemble model so that it shows the performance comparison of different algorithms.
- Since the process of classification is automatic, the user can check their result soon after submitting the voice.

5.2 Future Enhancement

The feature scope of this particular machine learning model can be extended to multiple dimensions. Future development on this topic may lead to the diagnosis of which voice diseases. In future, more technology can be brought into this topic along with power full back-end. Also, it will be designed as a mobile application. So that it can be used by medical institutions. Future work also includes the improvement of model accuracy.

References

[1] Voice Pathology Detection and Classification by Adopting Online Sequential Extreme Learning Machine,Conference on Communications (APCC),IEEE Access,Fahad Taha Al-Dhie**f**, 2021

[2] M. A. A. Albadr et al., "Mel-Frequency Cepstral Coefficient Features Based on Standard Deviation and Principal Component Analysis for Language Identification Systems," Cognitive Computation, pp. 1-18, July 2021.

[3] Voice Pathology Detection and Multi-classification Using Machine Learning Classifiers,International Conference on Sensing, Measurement &amp**,**IEEE Access,
Wu Yuanbo; Zhou Changwei; Fan Ziqi; Zhang Yihua; 2020

[4] T. Drugman, T. Dubuisson, and T. Dutoit, ''On the mutual information between source and filter contributions for voice pathology detection,'' 2020, arXiv:2001.00583. [Online]. Available: http://arxiv.org/abs/2001.00583

[5] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, ''Detection of pathological voice using cepstrum vectors: A deep learning approach,'' J. Voice, vol. 33, no. 5, pp. 634–641, Sep. 2019.

[6] M. Alhussein and G. Muhammad, ''Automatic voice pathology monitoring using parallel deep models for smart healthcare,'' IEEE Access, vol. 7, pp. 46474–46479, 2019, doi: 10.1109/ACCESS.2019.2905597.

[7] Voice pathology detection using machine learning,Aerospace and Electronic Systems Magazine,*IEEE Access*, Laura Verde, Giuseppe De Pietro Member, 2018

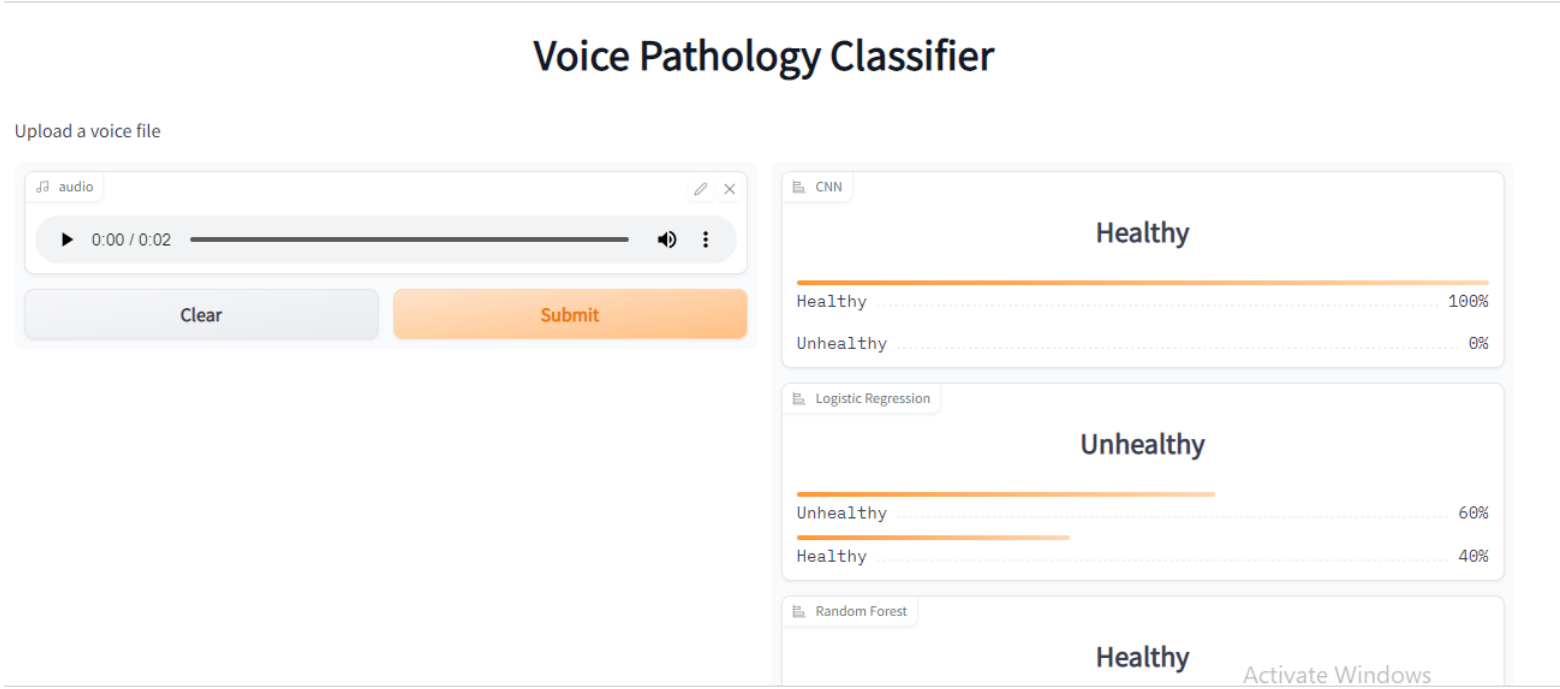
[8] H. J. A. Laverde, A. E. C. Ospina, and D. H. P. Ordóñez, ''Voice pathology detection using artificial neural networks and support vector machines powered by a multicriteria optimization algorithm,'' in Proc. Workshop Eng. Appl., Cham, Switzerland: Springer, vol. 915, 2018, pp. 148–159

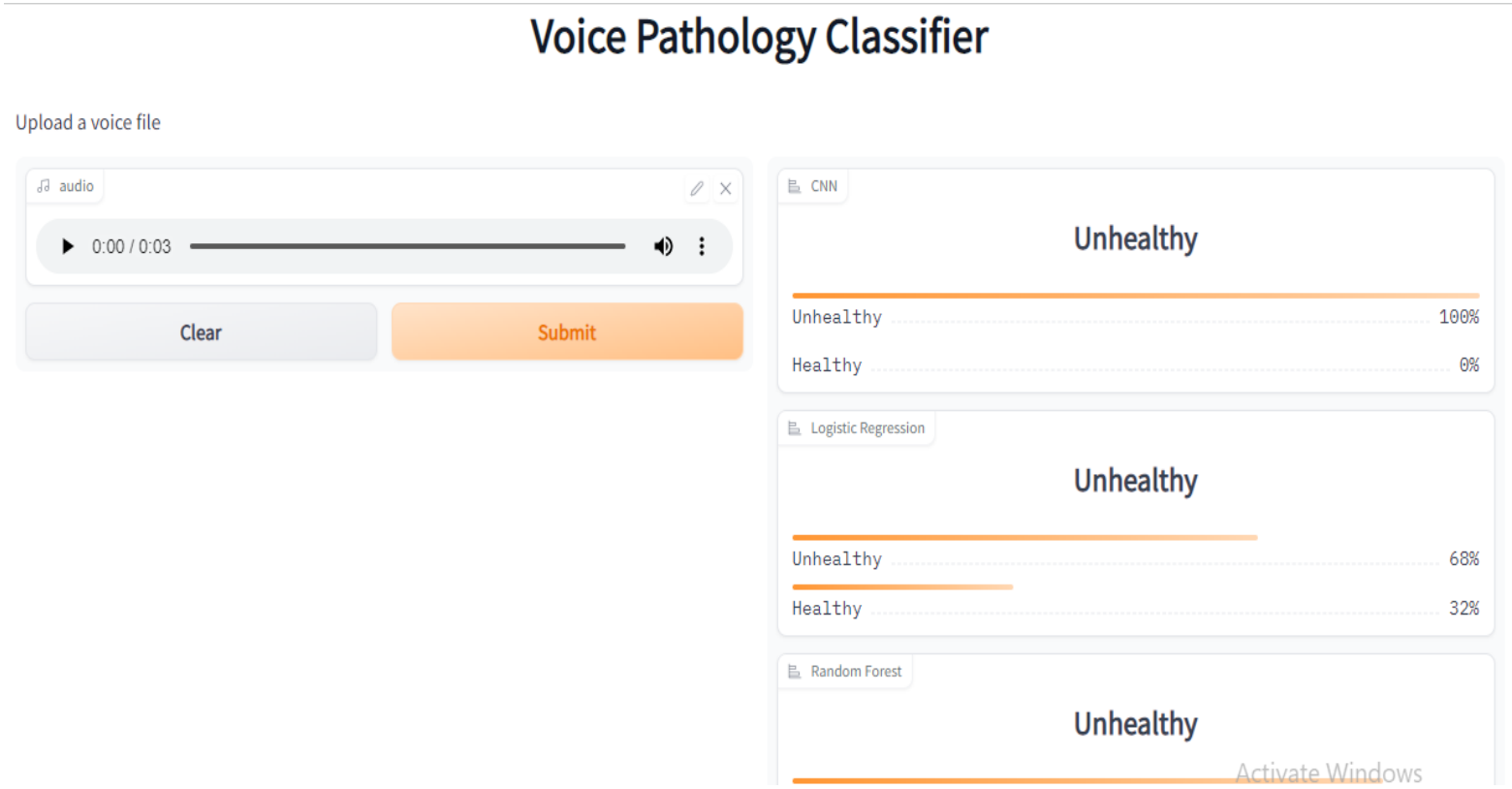
[9] L. Verde, G. De Pietro, and G. Sannino, ''Voice disorder identification by using machine learning techniques,'' IEEE Access, vol. 6, pp. 16246–16255, 2018, doi: 10.1109/ACCESS.2018.2816338.

[10] A Survey of Voice Pathology Surveillance Systems Based on Internet of Things and Machine Learning Algorithms,conference on Internet of Things and Health Care in Pandemic COVID-19: System Requirements Evaluation*,IEEE Access,*Fahad Taha Al-Dhief, 2018

[11] F. Teixeira, J. Fernandes, V. Guedes, A. Junior, and J. P. Teixeira, "Classification of control/pathologic subjects with support vector machines," *Procedia Computer Science*, vol. 138, pp, 2018.Google scholar.

APPENDICES


Voice Pathology Classifier
Upload a voice file
audio
0:00 / 0:02
Clear
Submit
CNN
Healthy
Healthy 100%
Unhealthy 0%
Logistic Regression
Unhealthy
Unhealthy 60%
Healthy 40%
Random Forest
Healthy
Activate Windows


Voice Pathology Classifier
Upload a voice file
audio
0:00 / 0:03
Clear
Submit
CNN
Unhealthy
Unhealthy 100%
Healthy 0%
Logistic Regression
Unhealthy
Unhealthy 68%
Healthy 32%
Random Forest
Unhealthy
Activate Windows