A

Project Stage-II Report on

SPEECH PROCESSING WITH MACHINE LEARNING FOR THROAT DISEASE DETECTION

Submitted in the partial fulfillment of the requirements of Semester-VIII
For the Award of the Degree of Bachelor of Technology (B.Tech) in
Electronics and Communication Engineering

Submitted by

SHASHANK	PRN:	2014111107
SATYAM SURESH	PRN:	2014111123
ASHWINI ANAND	PRN:	2014111141

Under the Guidance of

PROF. PRASHANT A. CHOUGULE

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY)
COLLEGE OF ENGINEERING, PUNE - 411 043

ACADEMIC YEAR: 2023-2024

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BHARATI VIDYAPEETH (DEEMED TO BE UNIVERSITY)
COLLEGE OF ENGINEERING, PUNE

CERTIFICATE

This is to certify that the Project Stage -II report on **”SPEECH PROCESSING WITH MACHINE LEARNING FOR THROAT DISEASE DETECTION”** submitted by

SHASHANK	**PRN:**	**2014111107**
SATYAM SURESH	**PRN:**	**2014111123**
ASHWINI ANAND	**PRN:**	**2014111141**

in partial fulfillment of the requirements of Semester-VIII (Academic Year: 2023-2024), for the award of degree of Bachelor of Technology (B.Tech) in Electronics and Communication Engineering.

Prof. Prashant A. Chougule
Guide, ECE Dept.,
BV(DU), COE, Pune

Dr. Dhiraj M. Dhane
Project Co-ordinator
BV(DU), COE, Pune

Prof.(Dr.) Arundhati A. Shinde
Head of the Department
BV(DU), COE, Pune

Date:

Place: Pune

Acknowledgements

Apart from our efforts, the success of our project depends largely on the encouragement and guideline of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project. We are gratefully indebted to our esteemed guide Prof. Prashant A. Chougule, Prof.(Dr.)Dhiraj M. Dhane ,Prof.(Dr.) Arundhati A. Shinde for sincere guidance and priceless support which would have been impossible for us to complete this project.

We express our gratitude to the staff members of Bharati Vidyapeeth (Deemed to be university) who directly or indirectly helped .

Contents

Acknowledgements i

List of Figures iv

List of Tables v

Abbreviations vi

Abstract vii

1 Introduction 1

1.1 Introduction 1

1.1.1 Evolution 1

1.1.2 Emerging technologies 4

1.2 Need of the Project 5

1.2.1 Advantages 5

1.2.2 Applications 6

2 Literature Review 8

3 Objectives 14

3.1 Objective of the Project 14

3.2 Aim of the Project 15

4 Methodology 17

4.1 Data Preprocessing 17

4.2 Model Architecture 17

4.3 Model Training 18

4.4 Model Evaluation 19

4.5 Post-Processing and Interpretation 20

5 Design and Implementation 21

5.1 Flowchart of Model 21

5.2 Project tools requirements 22

5.2.1 Software Requirements 22

5.2.2 Hardware Requirements 23

6 Results and Discussion 24

6.1 Exploratory Data Analysis 24

6.2 Simulation Results 28

7 Conclusion and Future Scope 30

7.1 Conclusion 30

7.2 Future Work 31

7.2.1 Data Acquisition and Enhancement 31

7.2.2 Model Development and Improvement 31

7.2.3 System Integration and User Experience 32

7.2.4 Clinical Validation and Applications 32

References 33

List of Figures

1.1 Throat Culture (Credit: Blausen Medical, via Wikimedia Commons) . . 3

1.2 Laryngoscopy (Credit: Tsezer, via iStock) 3

1.3 Nasopharyngoscopy (Credit: National Chemical Laboratory) 4

5.1 Flowchart for Deep Learning Model to classify Throat Disease 21

6.1 Reading dataset and determining its shape 25

6.2 Barplot indicating count of audio samples for each voice type i.e. normal(n) and pathological(p) 25

6.3 Barplot indicating count of audio samples for each pathology type i.e. normal, dysphonia, laryngitis and recurrent palsy 26

6.4 Above cell indicates count and datatype for each column; below cell indicates number of audio samples for each voice type and gender . . . 26

6.5 Barplot indicating count of audio samples for each voice type and gender 27

6.6 Result on voice type with both gender combined 28

6.7 Result on voice type with male gender only 28

6.8 Result on voice type with female gender only 28

6.9 Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with both gender combined 28

6.10 Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with male gender only 28

6.11 Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with female gender only 28

List of Tables

2.1 Overview of Research Studies on SVD Dataset 13

List of Abbreviations

ML	: Machine Learning
RADT	: Rapid Antigen Detection Test
CT	: Computed Tomography
MRI	: Magnetic Resonance Imaging
OCT	: Optical Coherence Tomography
DNA	: Deoxyribonucleic Acid
RNA	: Ribonucleic Acid
SVD	: Saarbrücken Voice Database
KNN	: K-Nearest Neighbor
MFCC	: Mel-frequency Cepstral Coefficients
CNN	: Convolutional Neural Network
SVM	: Support Vector Machine
ANN	: Artificial Neural Network
GMM	: Gaussian Mixture Model
KM	: K-Means Clustering
RF	: Random Forest
ELM	: Extreme Learning Machine
ReLU	: Rectified Linear Unit
CPU	: Central Processing Unit
GPU	: Graphics Processing Unit
XAI	: Explainable AI
RNN	: Recurrent Neural Network
EHR	: Electronic Health Records

Abstract

This work explores a novel application of Artificial Neural Networks (ANNs) for automated throat disease detection. By analyzing Mel-Frequency Cepstral Coefficients (MFCCs) extracted from voice samples, the ANN learns to differentiate healthy voices from those indicative of dysphonia, laryngitis, and recurrent palsy. This non-invasive approach offers a potentially rapid and accessible screening tool for early disease detection, paving the way for improved patient outcomes.

Index Terms- Pathological Voice, Saarbrücken Voice Database (SVD), Voice Pathology Identification, Machine Learning Techniques.

Chapter 1

Introduction

1.1 Introduction

We know that voice quality changes when the speech production system malfunctions due to pathology. Voice pathology can be caused by the presence of tissue infection, systemic changes, mechanical stress, surface irritation, tissue changes, neurological and muscular changes and other factors. In voice disorder the normal voice changes into a weak or tense voice that affects the quality of normal voice. Voice disorders are common among individuals in teaching, music, and related professions. If it is not detected in time then it will lead to critical conditions like permanent loss of voice or facing some other problems related to voice. Prospects for treatment are improved if a pathological condition can be detected early.

1.1.1 Evolution

The detection methods for throat diseases have undergone substantial evolution in recent years, marked by the continuous emergence of innovative technologies. This section provides a overview of the past and current aspects of throat disease detection.

Past methods of throat disease detection

In the past, throat diseases were diagnosed primarily based on a physical examination and the patient's medical history. Doctors would look for symptoms such as white spots or white patches on the back of throat, small red or purple spots inside mouth, signs of inflammation, such as redness and swelling, and would feel for enlarged lymph nodes in the neck. They would also ask the patient about their symptoms and any risk factors they might have.

In some cases, doctors might also order a throat swab to test for infection. A throat swab is a simple procedure in which a doctor or nurse uses a cotton swab to collect a sample of cells from the back of the throat. The sample is then sent to a laboratory to be tested for bacteria and viruses.

Current methods of throat disease detection

Today, there are a number of new technologies that can be used to detect throat diseases. One of the most common is the RADT. RADTs are quick and easy to perform, and they can provide results in as little as 15 minutes. RADTs work by detecting the presence of specific antigens, which are proteins that are produced by bacteria and viruses.

Another common method of throat disease detection is the throat culture. Throat cultures are more sensitive than RADTs, but they take longer to produce results. Throat cultures are typically used when a RADT is negative or when the doctor suspects a more serious infection.

In addition to RADTs and throat cultures, there are a number of other imaging tests that can be used to detect throat diseases. These tests include:

- Laryngoscopy: A laryngoscopy is a procedure in which a doctor uses a thin, flexible tube with a camera on the end to examine the larynx.

- Nasopharyngoscopy: A nasopharyngoscopy is a procedure in which a doctor uses a thin, flexible tube with a camera on the end to examine the nasopharynx, which is the area behind the nose and above the soft palate.

- CT scan: A CT scan is a type of X-ray that creates detailed images of the inside of the body. It is useful for detecting tumors, cysts, blockages and other calcified structures in the throat.
- MRI scan: An MRI scan uses a strong magnetic field and radio waves to create detailed images of the inside of the body. It is useful for detecting soft tissue abnormalities, such as inflammation and infections.

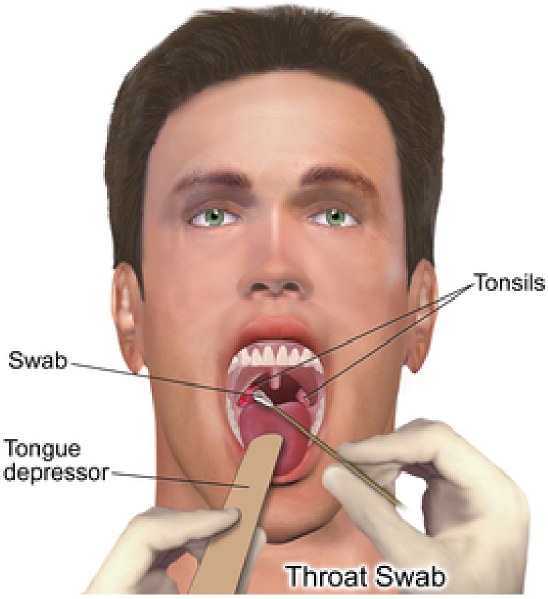


Figure 1.1: Throat Culture (Credit: Blausen Medical, via Wikimedia Commons)

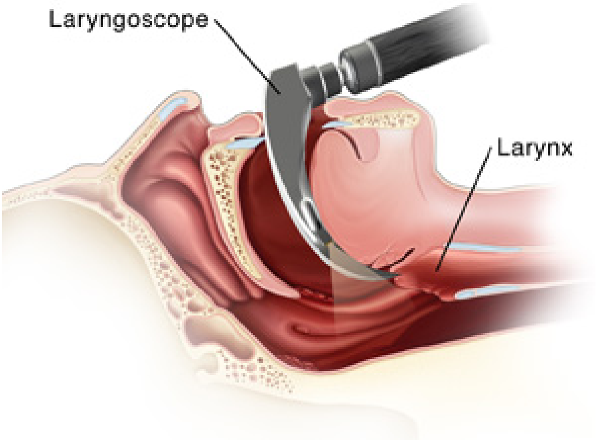


Figure 1.2: Laryngoscopy (Credit: Tsezer, via iStock)

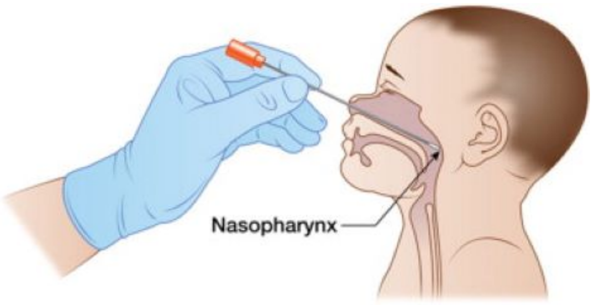


Figure 1.3: Nasopharyngoscopy (Credit: National Chemical Laboratory)

1.1.2 Emerging technologies

These emerging technologies have the potential to revolutionize the way that throat diseases are detected and managed.

- Deep learning voice disorder classification: This technology uses machine learning to analyze voice recordings and identify patterns that are associated with different throat diseases. It is still under development, but it has the potential to revolutionize the way that throat diseases are detected.
- OCT : This non-invasive imaging technique uses light to create high-resolution images of the inside of the body. It can be used to visualize the structures of the throat, such as the larynx, vocal cords, and pharynx. This can help doctors to identify abnormalities that may be indicative of throat disease.
- Raman spectroscopy: This technique uses light to analyze the molecular composition of a substance. It can be used to identify the presence of specific molecules in the throat, such as bacteria and viruses. This can help doctors to diagnose throat infections.
- Salivary diagnostics: Saliva is a promising biofluid for the diagnosis of throat diseases. It contains a variety of biomarkers, including proteins, DNA, and RNA, that can be used to detect the presence of disease. Researchers are developing new salivary diagnostic tests for throat diseases, such as strep throat, tonsillitis, and cancer.
- Wearable devices: Wearable devices, such as smartwatches and smart glasses, are being developed to monitor throat health. These devices can track voice changes,

swallowing patterns, and other indicators of throat disease. This information can be used to identify early signs of throat disease and to track the progression of disease over time.

1.2 Need of the Project

Throat diseases are a common group of medical conditions that can affect the throat, larynx, and pharynx. They can be caused by a variety of factors, including infections, allergies, and irritants. Symptoms of throat diseases can include pain, difficulty swallowing, hoarseness, and a cough. Traditional methods of throat disease detection, such as physical examinations and throat swabs, are not always accurate and timely. Physical examinations can be subjective, and throat swabs can produce false negative results.

Speech processing with ML has the potential to improve the accuracy and timeliness of throat disease detection. ML algorithms can be trained to analyze voice recordings and identify patterns that are associated with different throat diseases.

1.2.1 Advantages

- Accuracy: ML algorithms can be trained to achieve high levels of accuracy in throat disease detection. Studies have shown that ML algorithms can achieve accuracies of over 90% in detecting certain throat diseases, such as laryngeal cancer and vocal cord paralysis.

- Timeliness: ML algorithms can provide real-time or near-real-time results, which is important for the early detection and treatment of throat diseases.

- Non-invasiveness: Speech processing with ML is a non-invasive method of throat disease detection. This is important for patients who are uncomfortable with or unable to undergo traditional throat examinations.

- Cost-effectiveness: Speech processing with ML is a relatively cost-effective method of throat disease detection. This is important for making throat disease detection more accessible to patients.

1.2.2 Applications

- Early Detection of Throat Diseases: The system could be used as a screening tool for early detection of various throat diseases, including laryngeal cancer, vocal cord paralysis, and strep throat. By analyzing speech patterns, potential abnormalities could be identified, prompting individuals to seek further medical evaluation.

- Telemedicine and Remote Monitoring: The technology could be integrated into telemedicine platforms, allowing healthcare professionals to remotely assess patients for potential throat problems during virtual consultations.

- Augmentation of Clinical Diagnosis: The system could serve as a supplementary tool for clinicians by providing objective data alongside a patient's medical history and physical examination findings, potentially leading to more accurate diagnoses.

- Speech Therapy and Rehabilitation: The technology might be adapted to monitor progress during speech therapy for individuals with voice disorders, allowing therapists to personalize treatment plans and track improvements more objectively.

- Public Health Initiatives: The system could be incorporated into public health initiatives aimed at early detection and prevention of throat diseases, such as laryngeal cancer, particularly in areas with limited access to traditional medical facilities.

- Voice-Based Screening Tools: Potential applications extend beyond throat diseases. The technology could be adapted for voice-based screening tools for other conditions that may affect speech patterns, such as Parkinson's disease or neurological disorders.

Specific Applications for Different Throat Diseases:

- Laryngeal cancer: Laryngeal cancer is a type of cancer that develops in the larynx, or voice box. Speech processing with ML can be used to detect laryngeal cancer by analyzing voice recordings for changes in voice quality, such as hoarseness and breathiness.

- Vocal cord paralysis: Vocal cord paralysis is a condition in which one or both vocal cords are unable to move properly. This can cause hoarseness and difficulty swallowing. Speech processing with ML can be used to detect vocal cord paralysis by analyzing voice recordings for changes in voice quality, such as breathiness and weakness.

- Strep throat: Strep throat is a bacterial infection that causes inflammation of the throat. Speech processing with ML can be used to detect strep throat by analyzing voice recordings for changes in voice quality, such as hoarseness and muffled speech.

Chapter 2

Literature Review

Voice disorder is the abnormal tone of voice which is produced by vocal cord infecting the viruses. If it is not detected in time then it will lead to critical conditions like permanent loss of voice or facing some other problems related to voice.The existing studies aim to identify parameters for measuring voice quality and develop new classification systems for detecting voice disorders.

In [1], Fonseca proposed a system for voice pathology identification and detection that uses three different extractions, such as zero-crossing rate, signal entropy and energy. The Saarbrücken Voice Database (SVD) was used, and the maximum accuracy was 95%. Another work extracted glottal signal parameters for voice disorder detection and applied k-NN and SVM to classify the voice signal using SVD. SVM had an accuracy of 98.5% while k-NN had 88.2%, but with a limited set of voice samples.

By evaluating MFCC, shimmer and jitter in [2], El Emary classified speech and voice signals. They used the GMM algorithm to detect neurological voice disorders on a small dataset of 38 pathological and 63 healthy voices from the SVD database. In [3], Fonseca employed an algorithm based on LS-SVM and linear prediction coefficients to detect laryngeal voice disorder and tested it on a private dataset.

Dankovicová focused on feature selection (FS) and machine learning methods, such as K-nearest neighbors (KNN), random forests (RF), and support vector machines (SVM). The sustained vowels /a/, /i/, and /u/ generated by normal, high, low, and low-high-low were used. These vowels were selected from the Saarbrucken voice database, and 94 pathological subjects and 100 healthy subjects were chosen. The SVM classifier achieved the highest accuracy by reducing the feature set to 300 using the filter FS method in the original 1560 feature. The overall classification performance based on feature selection was the highest, with 80.3% for mixed samples, 80.6% for female samples, and 86.2% for male samples [4].

A study by Mohammed focused on transfer learning strategies, such as an effective pre-trained ResNet34 model for CNN training. Due to the unequal distribution of samples, this model adjusted the weights of the samples used for the minority groups during training as a means of compensation. A three-part weight product is the weight of the final sample. A class weight, a gender weight, as well as a gender–age weight, each led to a final sample weight. The 300 training samples extracted in the SVD were divided equally into 150 healthy and 150 pathological classes to ensure a balanced training process. An additional 1074 tested samples, divided into 200 healthy and 874 pathological classes, were included in the study. The system achieved a high prediction accuracy result of up to 94.54% accuracy on the training data and 95.41% accuracy on the testing data [5].

Hedge presented surveys of research works conducted on the automatic detection of voice disorders and explored ways to identify different types of voice disorders. They also analyzed different databases, feature extraction techniques, and machine learning approaches used in various studies. The voices were generally categorized as normal and pathological in most of the papers; however, some studies included Alzheimer's disease and Parkinson's disease (PD). Finally, this paper reviewed the performance of some of the significant research work conducted in this area [6].

A study by Hemmerling et al. sought to evaluate the usefulness of various speech signal analysis methods in the detection of voice pathologies. First, the initial vector consisted of 28 parameters extracted from the sustained vowels /a/, /i/, and /u/ at high, low, and normal pitch in time, frequency, and cepstral domains. Subsequently, linear feature extraction techniques (principal component analysis) were used to reduce the number of parameters and select the most effective acoustic features describing speech signals. They also performed nonlinear data transformations that were calculated using kernel principal components. The initial and extracted feature vectors were classified using k-means clustering and random forest classifiers. Using random forest classification for female and male recordings, they obtained accuracies of up to 100% for the classification of healthy versus pathological voices [7].

In Nawal and Cherif study, the ANN is proposed as unconventional approach in addition to the SVM as a new method successfully exploited in speech recognition. The main motivation for conducting this research was to investigate the efficiency of each of those classifiers in the identification of voice disorders. In addition, it was interesting to scrutinize the contribution of the first and second derivatives of the MFCC features for every classifier. The experimental results demonstrate that the effect of these derivative features depends on the classifier. Indeed, when the SVM is used as classifier, the first and second derivatives do not provide any improvement to the system performance comparing to the original MFCC features. However, when the ANN is used as classifier, these derivative features can be considered important since they contribute in the improvement of the system performance. In this case, there is an average improvement about 4% between the combination of the MFCC, MFCC Delta1 and the MFCC Delta2 [8].

In Deepak and Satija study, an automated technique for the classification of pathological speech is presented. The proposed technique has following major stages: Preprocessing, Feature extraction, Feature selection and Classification. Saarbrucken Voice Database (SVD) dataset is taken for our study, comprising both the pathological and healthy speech signals of vowel /a/ at normal pitch. The different features of pathological speech and healthy speech signals are extracted and most relevant features are selected by comparing the feature score. We found that the bark spectrum is the most important feature for the classification of pathological and healthy speech signals. The selected features are given as input to the classifiers with 5-fold cross-validation. K nearest neighbour (KNN), support vector machine (SVM), ensemble model, and neural network (NN) model are discussed in this paper which classifies the pathological and healthy speech with 94.45%, 90.00%, 95.60%, 93.35% accuracy respectively [9].

Alhussein and Muhammad proposed a system for voice pathology detection using a mobile platform and smart healthcare framework that is based on a deep learning model known as Convolutional Neural Network (CNN). Voice signals are recorded on smartphones, processed and analyzed in the cloud, and classified into three different parallel models. Parallel CNNs achieved 95.5% accuracy on the SVD dataset with 686 healthy voice samples and 1342 pathological voice samples, but only for sustained vowel /a/ spoken at normal pitch [10].

The Felipe and Teixeira work consists in a classification problem of four classes of vocal pathologies using one Deep Neural Network. Three groups of features extracted from speech of subjects with Dysphonia, Vocal Fold Paralysis, Laryngitis Chronica and controls were experimented. The best group of features are related with the source: relative jitter, relative shimmer, and HNR. A Deep Neural Network architecture with two levels were experimented. The first level consists in 7 estimators and second level a decision maker. In second level of the Deep Neural Network an accuracy of 39.5% is reached for a diagnosis among the 4 classes under analysis [11].

Laverde proposed a voice pathology detection and identification system using ANN and SVM classifiers and Particle Swarm Optimization for optimal parameters. Three types of features such as noise features, common voice features and acoustic features are extracted from each voice sample including healthy and pathological. The voice dataset (SVD) is divided into three groups (D1, D2, and D3), with each group containing the same number of voice samples. D1 consists of normal-pitched vowel /a/ sounds; D2 contains sentences; and D3 holds recorded sentences. The SVM provides an accuracy of 92.77%, while the ANN has a 93.27% accuracy, both based on group D3 of recorded sentences. However, the system's performance for other vowels such as /u/ and /i/ pronounced with varied intonations was not assessed using the speech database (SVD) [12].

According to[13] analysis, a client provides his or her voice sample and the sample goes for initial processing. Once complete the initial process, this data forwarded to the convolutional filters and max pooling filters for the detection of voice disability and will get the results. In CNN, inconsistency problems can be easily solved through the use of max-pooling. The CNN algorithm works well and identifies the pathology and non-pathology disease. The results have shown that the best accuracy in voice pathology detection is achieved using the Convolutional Neural Network. This technique classifies a voice as pathological or healthy with an accuracy equal to about 97-97.3% using all parameters. All analyses are performed on a wide dataset from the Saarbruecken Voice Database.

There is already a great number of related works in this area of expertise. Summarized information about other papers published on SVD can be found in Table I [7]-[23]. The results vary greatly between the published papers mainly due to differences between sets of data that were used for the experiment.

Table 2.1: Overview of Research Studies on SVD Dataset

Article	Features Used	Classifier Used	Overall Accuracy
[14]	peak, lag, entropy/eight band pass filter	SVM	99.53%
[15]	Maximum peak and lag	SVM	90.98%
[16]	MFCCs	GMM	80.02%
[17]	PCA	K-Means Clustering	100%
[18]	glottal flow features	ANN-SVM	99.27%;98.43%
[19]	Mutual information p/w voice classes (normophonic/dysphonic)	SVM	94.1%
[20]	Glottal source features and MFCCs	SVM	76.19%
[21]	Jitter, shimmer and HNR, MFCCs	SVM	71%
[22]	MFCCs and Temporal Derivatives	SVM	86%
[23]	MFCCs	SVM, GMM	96.5% -95.5%
[2]	MFCCs, jitter and shimmer	GMM	82.37%
[7]	28 parameters extracted from time, frequency and cepstral domain	KM,RF	100%
[24]	energy, entropy, contrast, homogeneity	GMM	99.98%
[25]	MFCCs	GMM	99%
[26]	MPEG-7 low-level audio and IDP	SVM,ELM,GMM	95%
[27]	MFCC first and second derivatives	ANN	87.82%
[28]	MFCC, harmonics-to-noise ratio, normalized noise energy, glottal-to-noise excitation ratio	GMM	79.40%
[29]	Peak value and lag for every frequency band	GMM,SVM	72%

Chapter 3

Objectives

3.1 Objective of the Project

This project aims to develop a system for automatic throat disease detection using speech processing and machine learning techniques. The specific objectives are:

- Feature Extraction: To extract informative features from speech signals that effectively differentiate between healthy and pathological voices. This may involve utilizing established techniques like MFCCs and exploring additional features relevant to voice quality assessment.

- Machine Learning Classification: To train and evaluate machine learning models for classifying speech samples into different categories. This could involve exploring various algorithms such as SVMs, Random Forests, or even deep learning architectures like CNNs.

- Multi-class Classification: To go beyond simply classifying healthy vs. pathological voices and achieve classification of specific throat diseases. This would enable a more precise diagnosis by distinguishing between different types of vocal pathologies (e.g., dysphonia, laryngitis, recurrent laryngeal palsy).

- Improved Accuracy and Generalizability: To develop a model that achieves high accuracy in throat disease detection. This involves using large and diverse datasets encompassing various voice samples and pathologies. Additionally, techniques like feature selection and hyperparameter tuning can be employed to optimize model performance.

- Advancement in Speech Processing for Medical Applications: To contribute to the field of speech processing for medical applications. This project's findings can pave the way for developing more robust and clinically relevant tools for throat disease detection using speech analysis.

3.2 Aim of the Project

Voice disorders are surprisingly common, affecting millions of people worldwide. These disorders can significantly impact communication and quality of life. Early detection is crucial for effective treatment. This research explores the potential of speech processing and machine learning for accurate throat disease detection. Traditionally, throat disease detection relies on clinical examinations and acoustic analysis by trained professionals. While effective, these methods can be time-consuming and subjective. Machine learning offers a promising alternative, with the potential for objective and automated voice analysis.

This project explores the potential of speech processing with machine learning to address these limitations and contribute to a more accessible and efficient approach to throat disease detection. We propose a novel system that leverages the power of machine learning to analyze speech patterns and identify potential voice pathologies directly from raw audio signals.

This research builds upon the growing body of work investigating the application of machine learning in voice pathology detection. Existing studies have demonstrated the effectiveness of using machine learning models trained on extracted acoustic features, such as Mel-frequency cepstral coefficients, to classify pathological and healthy voices. However, this approach often requires feature engineering expertise and can be susceptible to variations in feature extraction methods.

Our proposed system aims to bypass manual feature engineering, a process that traditionally requires expertise and can be sensitive to feature extraction methods. Instead, the system employs ANNs to learn discriminative features directly from the raw speech data. We trained the ANN model on a comprehensive dataset encompassing healthy and pathological voice samples. The dataset includes recordings from publicly available databases and patient recordings from hospitals. This diversity helps ensure the model's generalizability. Our approach goes beyond simply detecting the presence of a throat disease. We aim to identify specific types of diseases, such as dysphonia, laryngitis, and recurrent laryngeal palsy. This additional information can be invaluable for guiding treatment decisions.The key contributions of this research are:

- Development of an ANN-based system for classifying specific throat diseases based on speech features.
- Utilization of a diverse dataset incorporating healthy and pathological voice samples from multiple sources.
- Evaluation of the system's effectiveness using various metrics beyond just accuracy.

Furthermore, we recognize the challenge of real-world deployment. Deep learning models often require significant computational resources, which can limit their suitability for use on resource-constrained devices. To address this challenge, we explore techniques to design more efficient ANN architectures that can achieve good performance on embedded devices. Another important consideration for real-world application is the impact of environmental noise. Everyday environments can introduce significant background noise that can degrade the performance of voice analysis systems. We investigate strategies to mitigate the effects of noise, potentially through domain adaptation techniques that train the model to be more robust in noisy conditions.

Chapter 4

Methodology

This section outlines the methodological approach for developing the automatic throat disease detection system. The process will be divided into distinct stages:

4.1 Data Preprocessing

- Feature Extraction: Audio recordings will be converted into numerical representations suitable for machine learning models. Techniques like MFCCs or spectrograms will be employed to capture relevant characteristics of voice quality.
- Data Splitting: The preprocessed data will be divided into three sets: training, validation, and testing. The training set will be used to train the machine learning model, the validation set will be used to fine-tune hyperparameters and prevent overfitting, and the testing set will be used for final performance evaluation.
- Normalization: The extracted features will be normalized to a specific range (e.g., 0-1 or -1 to 1) to ensure consistent input data for the machine learning model. This improves training efficiency and can enhance model performance.

4.2 Model Architecture

- ANN: Design a multi-layer ANN architecture suitable for classifying speech data. The network will consist of:

1. Input Layer: Receives the extracted features i.e. MFCCs.
2. Hidden Layers (1 or more): These layers perform the core computations and learning. Experiment with different numbers of hidden layers and neurons per layer to find the optimal architecture for our dataset. Consider using activation functions like ReLU in these layers to introduce non-linearity and improve the model's ability to learn complex relationships between features and class labels.
3. Output Layer: Generates class probabilities. Depending on the number of throat diseases we aim to identify, use a softmax activation function in the output layer for multi-class classification (e.g., healthy, dysphonia, laryngitis, recurrent laryngeal palsy).

- Experimentation: Explore variations in the ANN architecture by adjusting the number of hidden layers, the number of neurons in each layer, and the choice of activation functions. The goal is to find the most effective architecture that balances accuracy, computational efficiency, and generalizability for our specific dataset.Consider techniques like grid search or random search to systematically explore different hyperparameter combinations.

4.3 Model Training

- Training Process: Train the designed ANN model using the training dataset. A backpropagation algorithm will be employed to iteratively adjust the weights and biases within the network to minimize the error between the model's predictions and the actual labels in the training data. Choose an appropriate learning rate to control the speed and convergence of the training process. A too high learning rate can lead to instability and divergence, while a too low learning rate can result in slow convergence.
- Validation Set: Continuously monitor the model's performance on the validation set during training. Early stopping can be used to halt training when the

validation performance stops improving, preventing overfitting to the training data.

- Regularization Techniques: Techniques like weight decay, dropout, and L1/L2 regularization can be employed to enhance model generalization and prevent overfitting. Weight decay penalizes large weights, discouraging the network from overfitting to the training data.

  Dropout randomly drops out a certain percentage of neurons during training, preventing them from co-adapting too strongly and promoting learning of independent features. L1/L2 regularization penalizes the magnitude of the weights, encouraging the model to learn sparse representations and reduce complexity.

4.4 Model Evaluation

- Testing Set: The trained model's performance will be evaluated on the unseen testing set. This provides an objective assessment of how well the model generalizes to new data and performs in a real-world scenario.

- Evaluation Metrics: Various metrics such as accuracy, precision, recall, and F1-score will be used to evaluate the model's effectiveness in classifying different throat disease categories.

  Accuracy reflects the overall percentage of correctly classified samples. Precision measures the proportion of positive predictions that are actually correct, while recall indicates the proportion of actual positives that are correctly identified. F1-score provides a harmonic mean of precision and recall, offering a balanced view of model performance.

- Error Analysis: Misclassifications will be analyzed to identify potential areas for improvement. By understanding the types of errors the model makes, we can refine the model architecture, training process, or feature extraction techniques to address these shortcomings.

4.5 Post-Processing and Interpretation

- Output Generation: The model's raw predictions will be post-processed to generate a meaningful output for the user. This may involve converting the predicted class probabilities into human-readable labels (e.g., healthy, dysphonia, specific voice complaint).

- Result Interpretation: The results will be interpreted in the context of throat disease detection. The model's performance metrics, along with insights from the error analysis, will be used to evaluate the system's effectiveness and identify potential areas for further development.

Chapter 5

Design and Implementation

5.1 Flowchart of Model

Flowchart for building deep learning model to classify throat disease is shown in Figure 5.1.

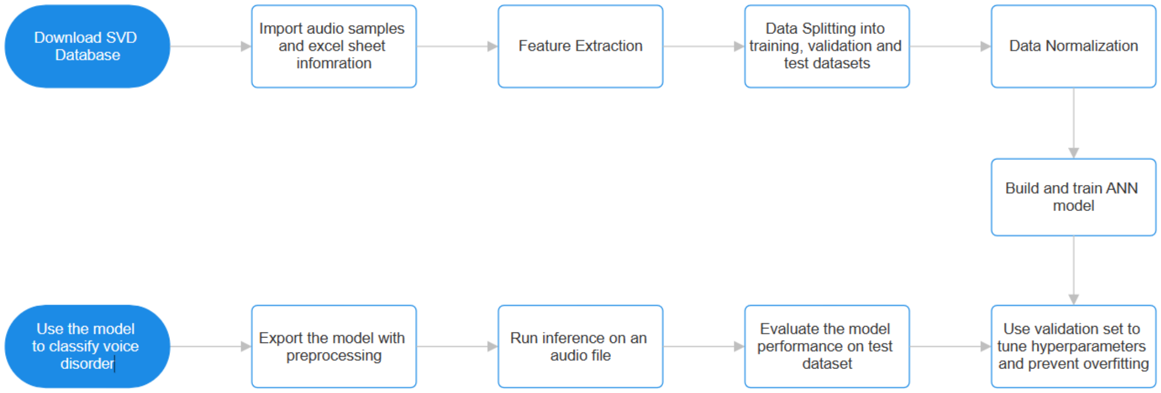

Download SVD Database
Import audio samples and excel sheet infomration
Feature Extraction
Data Splitting into training, validation and test datasets
Data Normalization
Build and train ANN model
Use validation set to tune hyperparameters and prevent overfitting
Evaluate the model performance on test dataset
Run inference on an audio file
Export the model with preprocessing
Use the model to classify voice disorder

Figure 5.1: Flowchart for Deep Learning Model to classify Throat Disease

5.2 Project tools requirements

5.2.1 Software Requirements

- Python: Python is a general-purpose programming language that is widely used for machine learning and data science.
- Jupyter Notebook: Jupyter Notebook is an interactive environment for creating and sharing documents that contain live code, equations, visualizations, and narrative text.
- Librosa: Librosa is a Python library for music and audio analysis. It provides a variety of functions for audio processing, such as feature extraction and signal processing.
- Pandas: Pandas is a Python library for data analysis. It provides a variety of functions for data manipulation and analysis.
- Seaborn: Seaborn is a Python library for statistical data visualization.
- Scikit-learn: Scikit-learn is a Python library for machine learning. It provides a variety of machine learning algorithms, including classification and regression algorithms.
- TensorFlow: TensorFlow is an open-source software library for numerical computation using data flow graphs. It is used to train deep learning models, including CNN models.

The following sections describe the specific software requirements for each stage of the project :-

1. Data collection and preparation
    - Librosa: Librosa will be used to load and preprocess the voice recordings.
    - Pandas: Pandas will be used to create and manipulate the dataframes containing the features extracted from the voice recordings.

   - Tensorflow : Tensorflow will be used to convert audio data into spectrograms.

2. Model development

   - Scikit-learn: Scikit-learn will be used to implement the ANN model.
   - Tensorflow: Tensorflow will be used to build and train the ANN model.

3. Model evaluation

   - Scikit-learn: Scikit-learn library will be used to evaluate the performance of the trained ANN model on a held-out test set.

4. Model deployment

   - Tensorflow: Tensorflow will be used to export the ANN model to a production environment.

5.2.2 Hardware Requirements

- A computer with a powerful CPU and GPU: A powerful CPU and GPU are required to train the ANN model efficiently.
- A microphone: A microphone is required to record the voice recordings.
- A headset (optional): A headset is recommended to reduce background noise during recording.

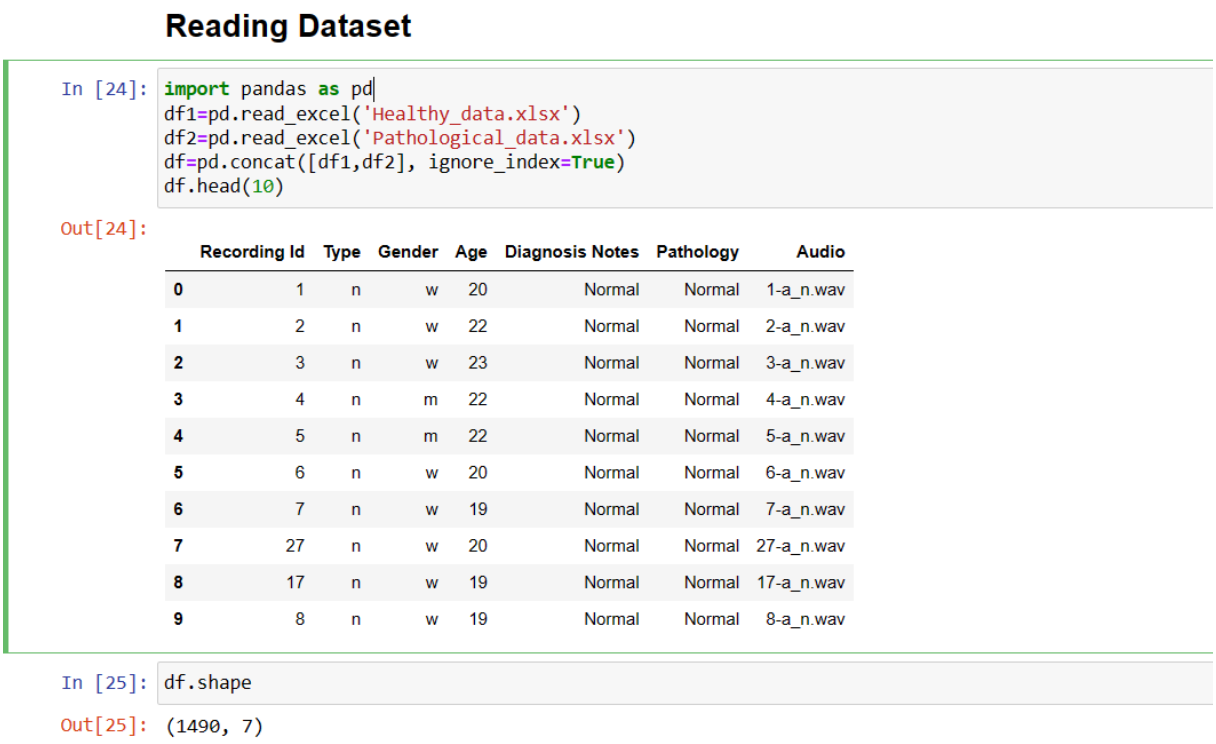
Chapter 6

Results and Discussion

6.1 Exploratory Data Analysis

It involves analyzing your data to understand its characteristics, identify patterns, and uncover potential issues before diving into model building. Following are the analyses obtained during this research :-

- Figure 6.1 indicates top 10 rows of dataset used for buildng model and its shape, which means number of rows and columns in dataset.
- Figure 6.2,6.3 and 6.5 indicate barplots for distribution of data sample in different categories of voice disorders.
- Figure 6.4 indicates the non-null count and data type for different columns in the dataset.



Reading Dataset

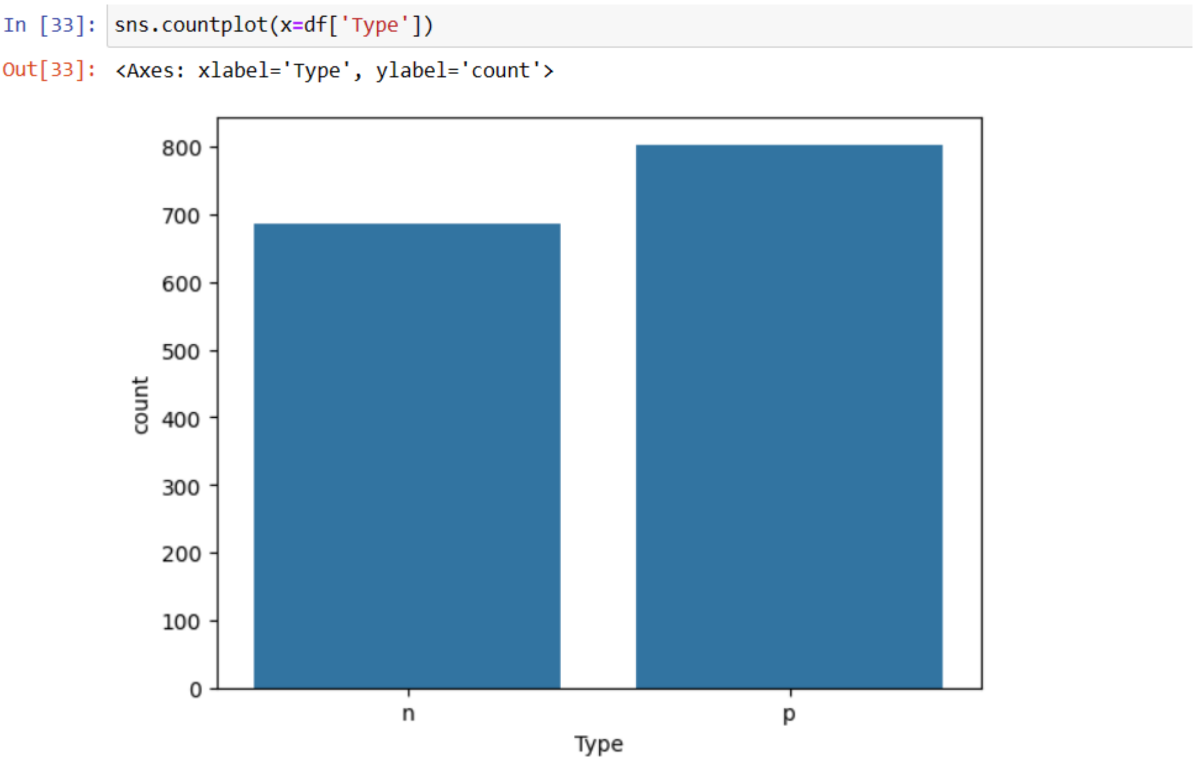
```
In [24]: import pandas as pd
         df1=pd.read_excel('Healthy_data.xlsx')
         df2=pd.read_excel('Pathological_data.xlsx')
         df=pd.concat([df1,df2], ignore_index=True)
         df.head(10)
```

Out[24]:

	Recording Id	Type	Gender	Age	Diagnosis Notes	Pathology	Audio
0	1	n	w	20	Normal	Normal	1-a_n.wav
1	2	n	w	22	Normal	Normal	2-a_n.wav
2	3	n	w	23	Normal	Normal	3-a_n.wav
3	4	n	m	22	Normal	Normal	4-a_n.wav
4	5	n	m	22	Normal	Normal	5-a_n.wav
5	6	n	w	20	Normal	Normal	6-a_n.wav
6	7	n	w	19	Normal	Normal	7-a_n.wav
7	27	n	w	20	Normal	Normal	27-a_n.wav
8	17	n	w	19	Normal	Normal	17-a_n.wav
9	8	n	w	19	Normal	Normal	8-a_n.wav

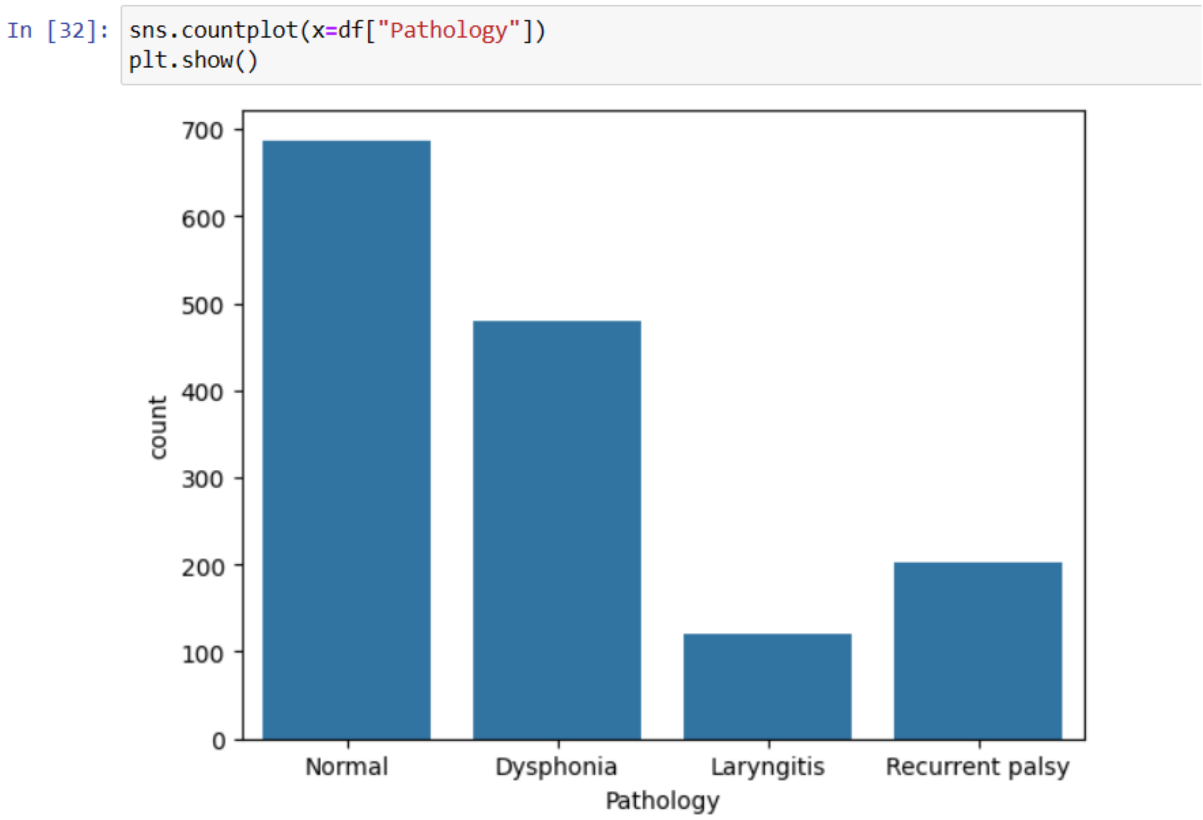
```
In [25]: df.shape
Out[25]: (1490, 7)
```

Figure 6.1: Reading dataset and determining its shape



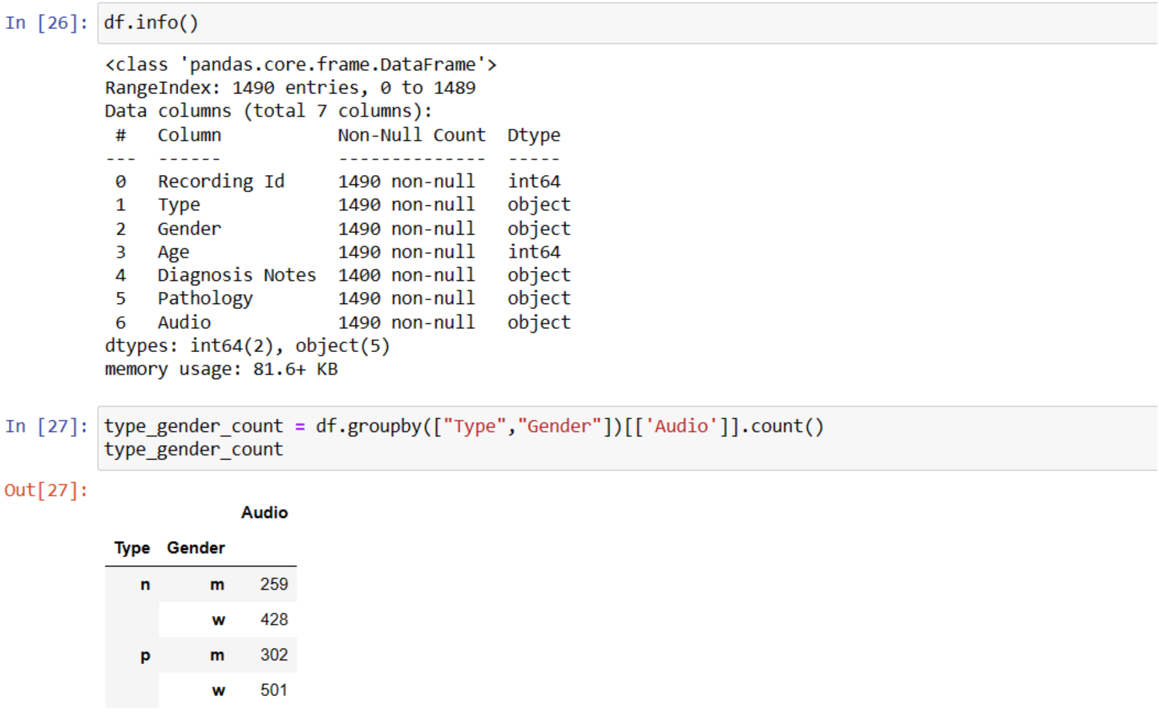
```
In [33]: sns.countplot(x=df['Type'])
Out[33]: <Axes: xlabel='Type', ylabel='count'>
```

Figure 6.2: Barplot indicating count of audio samples for each voice type i.e. normal(n) and pathological(p)



```
In [32]: sns.countplot(x=df["Pathology"])
         plt.show()
```

Figure 6.3: Barplot indicating count of audio samples for each pathology type i.e. normal, dysphonia, laryngitis and recurrent palsy



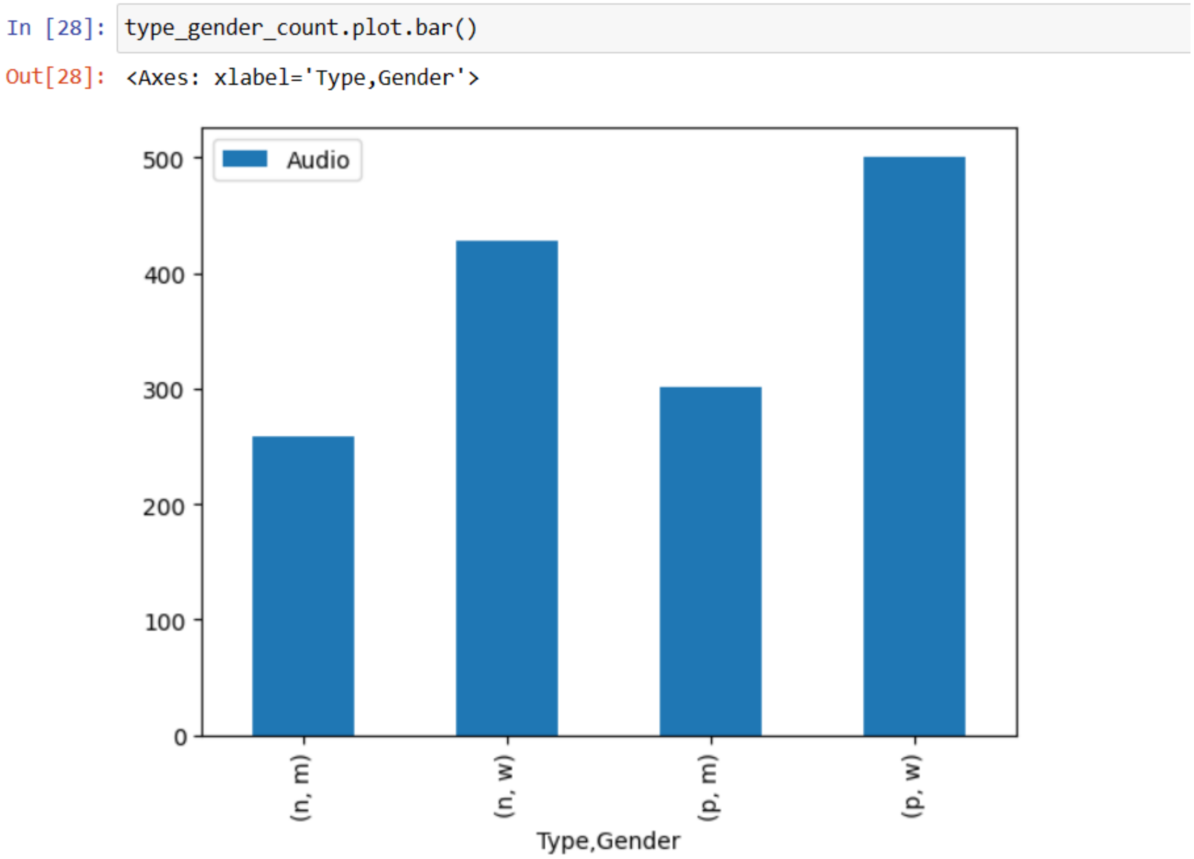
```
In [26]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1490 entries, 0 to 1489
Data columns (total 7 columns):
 #   Column           Non-Null Count  Dtype 
---  ------           --------------  ----- 
 0   Recording Id     1490 non-null   int64 
 1   Type             1490 non-null   object
 2   Gender           1490 non-null   object
 3   Age              1490 non-null   int64 
 4   Diagnosis Notes  1400 non-null   object
 5   Pathology        1490 non-null   object
 6   Audio            1490 non-null   object
dtypes: int64(2), object(5)
memory usage: 81.6+ KB
```

```
In [27]: type_gender_count = df.groupby(["Type","Gender"])[['Audio']].count()
         type_gender_count
```

Out[27]:

		Audio
Type	**Gender**	
n	**m**	259
	w	428
p	**m**	302
	w	501

Figure 6.4: Above cell indicates count and datatype for each column; below cell indicates number of audio samples for each voice type and gender

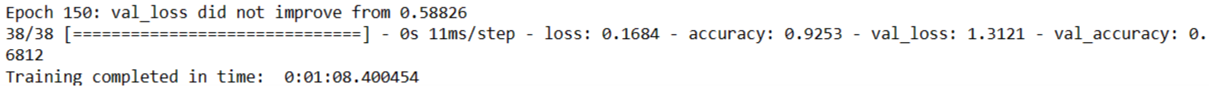


```
In [28]: type_gender_count.plot.bar()

Out[28]: <Axes: xlabel='Type,Gender'>
```

Figure 6.5: Barplot indicating count of audio samples for each voice type and gender

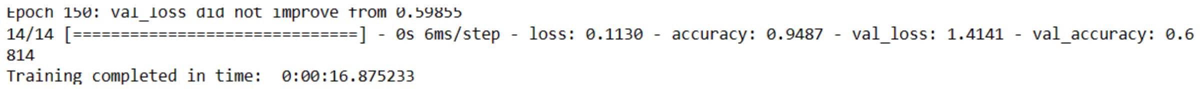
6.2 Simulation Results

Results including training set accuracy, training set loss, validation set loss and validation set accuracy of different subsets of dataset used for model training is shown in Figure 6.6,6.7,6.8,6.9,6.10 and 6.11.



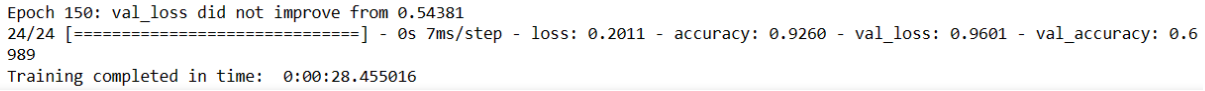
```
Epoch 150: val_loss did not improve from 0.58826
38/38 [==============================] - 0s 11ms/step - loss: 0.1684 - accuracy: 0.9253 - val_loss: 1.3121 - val_accuracy: 0.
6812
Training completed in time:  0:01:08.400454
```

Figure 6.6: Result on voice type with both gender combined



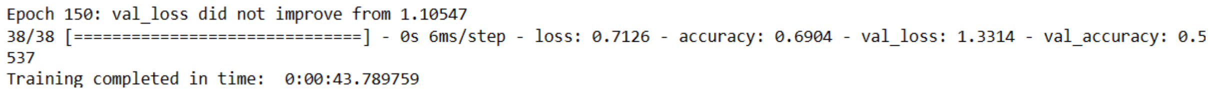
```
Epoch 150: val_loss did not improve from 0.59855
14/14 [==============================] - 0s 6ms/step - loss: 0.1130 - accuracy: 0.9487 - val_loss: 1.4141 - val_accuracy: 0.6
814
Training completed in time:  0:00:16.875233
```

Figure 6.7: Result on voice type with male gender only



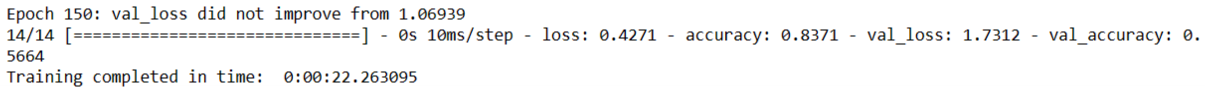
```
Epoch 150: val_loss did not improve from 0.54381
24/24 [==============================] - 0s 7ms/step - loss: 0.2011 - accuracy: 0.9260 - val_loss: 0.9601 - val_accuracy: 0.6
989
Training completed in time:  0:00:28.455016
```

Figure 6.8: Result on voice type with female gender only



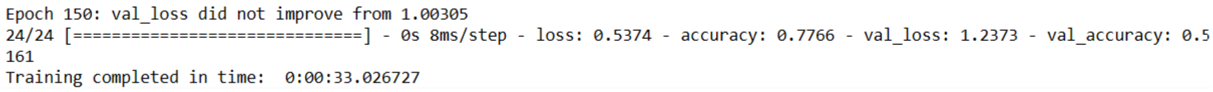
```
Epoch 150: val_loss did not improve from 1.10547
38/38 [==============================] - 0s 6ms/step - loss: 0.7126 - accuracy: 0.6904 - val_loss: 1.3314 - val_accuracy: 0.5
537
Training completed in time:  0:00:43.789759
```

Figure 6.9: Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with both gender combined



```
Epoch 150: val_loss did not improve from 1.06939
14/14 [==============================] - 0s 10ms/step - loss: 0.4271 - accuracy: 0.8371 - val_loss: 1.7312 - val_accuracy: 0.
5664
Training completed in time:  0:00:22.263095
```

Figure 6.10: Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with male gender only



```
Epoch 150: val_loss did not improve from 1.00305
24/24 [==============================] - 0s 8ms/step - loss: 0.5374 - accuracy: 0.7766 - val_loss: 1.2373 - val_accuracy: 0.5
161
Training completed in time:  0:00:33.026727
```

Figure 6.11: Result on pathology type(including normal, dysphonia, laryngitis and recurrent palsy) with female gender only

- These result indicates that separate model each for male and female predicts voice disorder with more accuracy than a model trained on combined data for both gender.

- These result also indicates that a model trained only on voice type(i.e. normal and pathological) has much better accuracy than the model trained with different pathological types(i.e. normal, dysphonia, laryngitis and recurrent palsy)

Chapter 7

Conclusion and Future Scope

7.1 Conclusion

A healthcare framework using machine learning techniques was proposed. This paper provided an overview of the techniques used by machine learning methods in the voice disorder detection. We can conclude the pathological voices into two forms of observation. One using medical methods by using expensive equipment to check and Second is by computer based system check by using Deep learning approach. In the framework, we develop a voice disorder assessment and treatment system using a machine learning approach. A client provides his or her voice sample and the sample goes for initial processing. Then MFCCs were extracted from voice samples to capture the spectral characteristics of speech. Once complete the initial process this data forwarded to the different layers of ANN for the detection of voice disability and will get the results. An ANN model was trained on a dataset labeled with different throat conditions, including healthy, dysphonia, laryngitis, and recurrent palsy. The model learned the relationships between MFCC features and disease states, achieving promising results in classifying new voice samples.

This research contributes to the development of non-invasive and accessible tools for early detection of throat diseases. It has the potential to improve healthcare delivery by enabling convenient screening and promoting timely intervention for patients. Furthermore, this project successfully achieved its primary objective of developing a machine learning model for classifying throat diseases based on speech analysis.

7.2 Future Work

Some of the future research and improvements possible to the project are:

7.2.1 Data Acquisition and Enhancement

- Data Augmentation: Explore techniques like noise injection, pitch shifting, and speed perturbation to artificially expand the training dataset and improve model robustness to real-world variations.
- Multilingual Support: Develop the system to handle speech recordings in various languages, increasing its accessibility and global reach.
- Real-World Data Collection: Collect data from diverse populations in real-world settings (e.g., hospitals, clinics) to capture the full spectrum of voice variations and enhance model generalizability.

7.2.2 Model Development and Improvement

- Ensemble Learning: Investigate combining multiple machine learning models (e.g., SVM, CNN) through ensemble learning techniques like bagging or boosting to potentially improve overall accuracy and robustness.
- XAI: Implement techniques to make the model's decision-making process more transparent and interpretable for clinicians, fostering trust and understanding in its results.
- Deep Learning Architectures: Explore advanced deep learning architectures like RNNs or transformers, which may be particularly suited for capturing sequential information in speech data.

7.2.3 System Integration and User Experience

- Mobile App Integration: Develop a mobile application that allows users to easily record and submit speech samples for analysis, potentially promoting early detection and self-screening.

- Real-time Processing: Investigate real-time speech processing capabilities, enabling immediate feedback on potential voice abnormalities during consultations or speech therapy sessions.

- Customization for Different User Groups: Consider tailoring the system's interface and sensitivity based on user profiles (e.g., age, profession) to provide more relevant and informative results.

7.2.4 Clinical Validation and Applications

- Clinical Trials: Conduct controlled clinical trials to evaluate the system's effectiveness in a real-world clinical setting alongside traditional diagnostic methods.

- Integration with EHR: Develop interfaces to integrate the system's findings with electronic health records, streamlining data collection and improving healthcare workflow.

- Collaboration with Speech-Language Pathologists: Partner with speech-language pathologists to explore how the system can be used for personalized therapy plans and progress monitoring for patients with voice disorders.

By exploring these future directions, we can contribute to the ongoing development of a robust, user-friendly, and clinically relevant system for automatic throat disease detection.

References

[1] E. S. Fonseca, R. C. Guido, S. B. Junior, H. Dezani, R. R. Gati, and D. C. M. Pereira, "Acoustic investigation of speech pathologies based on the discriminative paraconsistent machine (dpm)," *Biomedical Signal Processing and Control*, vol. 55, p. 101615, 2020.

[2] I. El Emary, M. Fezari, and F. Amara, "Towards developing a voice pathologies detection system," *Journal of Communications Technology and Electronics*, vol. 59, pp. 1280–1288, 2014.

[3] E. S. Fonseca, R. C. Guido, P. R. Scalassara, C. D. Maciel, and J. C. Pereira, "Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders," *Computers in Biology and Medicine*, vol. 37, no. 4, pp. 571–578, 2007.

[4] Z. Dankovičová, D. Sovák, P. Drotár, and L. Vokorokos, "Machine learning approach to dysphonia detection," *Applied Sciences*, vol. 8, no. 10, p. 1927, 2018.

[5] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.

[6] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A survey on machine learning approaches for automatic detection of voice disorders," *Journal of Voice*, vol. 33, no. 6, pp. 947–e11, 2019.

[7] D. Hemmerling, A. Skalski, and J. Gajda, "Voice data mining for laryngeal pathology assessment," *Computers in biology and medicine*, vol. 69, pp. 270–276, 2016.

[8] N. Souissi and A. Cherif, "Artificial neural networks and support vector machine for voice disorders identification," *International Journal of Advanced Computer Science and Applications*, vol. 7, no. 5, 2016.

[9] D. Kumar, U. Satija, and P. Kumar, "Automated classification of pathological speech signals," in *2022 IEEE 19th India Council International Conference (INDICON)*. IEEE, 2022, pp. 1–5.

[10] M. Alhussein and G. Muhammad, "Automatic voice pathology monitoring using parallel deep models for smart healthcare," *Ieee Access*, vol. 7, pp. 46 474–46 479, 2019.

[11] F. Teixeira and J. P. Teixeira, "Deep-learning in identification of vocal pathologies," in *13th International Joint Conference on Biomedical Engineering Systems and Technologies*, vol. 4, 2020, pp. 288–295.

[12] F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. A. A. Albader, N. M. Mahyuddin, and M. A. Mohammed, "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77 293–77 306, 2021.

[13] D. U. G. MS.HARSHITA BHAGWAT, "Learning cnn strategy for voice disorder classification and detection," vol. 9, 2021.

[14] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, and M. F. Ibrahim, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *Ieee Access*, vol. 6, pp. 6961–6974, 2017.

[15] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, and M. A. Bencherif, "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification," *Journal of Voice*, vol. 31, no. 1, pp. 113–e9, 2017.

[16] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, A. Al-Nasheri, T. A. Mesallam, M. Farahat, and K. H. Malki, "Intra-and inter-database study for arabic, english, and german databases: do conventional speech features detect voice pathology?" *Journal of Voice*, vol. 31, no. 3, pp. 386–e1, 2017.

[17] V. Guedes, F. Teixeira, A. Oliveira, J. Fernandes, L. Silva, A. Junior, and J. P. Teixeira, "Transfer learning with audioset to voice pathologies identification in continuous speech," *Procedia Computer Science*, vol. 164, pp. 662–669, 2019.

[18] K. Ezzine and M. Frikha, "Investigation of glottal flow parameters for voice pathology detection on svd and meei databases," in *2018 4th International conference on advanced technologies for signal and image processing (ATSIP)*. IEEE, 2018, pp. 1–6.

[19] M. Markaki and Y. Stylianou, "Voice pathology detection and discrimination based on modulation spectral features," *IEEE Transactions on audio, speech, and language processing*, vol. 19, no. 7, pp. 1938–1948, 2011.

[20] S. R. Kadiri and P. Alku, "Analysis and detection of pathological voice using glottal source features," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 367–379, 2019.

[21] 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. 272–279, 2018.

[22] N. Souissi and A. Cherif, "Dimensionality reduction for voice disorders identification system based on mel frequency cepstral coefficients and support vector machine," in *2015 7th international conference on modelling, identification and control (ICMIC)*. IEEE, 2015, pp. 1–6.

[23] F. Amara, M. Fezari, and H. Bourouba, "An improved gmm-svm system based on distance metric for voice pathology detection," *Appl. Math*, vol. 10, no. 3, pp. 1061–1070, 2016.

[24] G. Muhammad, M. F. Alhamid, M. S. Hossain, A. S. Almogren, and A. V. Vasilakos, "Enhanced living by assessing voice pathology using a co-occurrence matrix," *Sensors*, vol. 17, no. 2, p. 267, 2017.

[25] Ö. Eskidere, A. Gürhanlı *et al.*, "Voice disorder classification based on multitaper mel frequency cepstral coefficients features," *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[26] M. S. Hossain and G. Muhammad, "Healthcare big data voice pathology assessment framework," *iEEE Access*, vol. 4, pp. 7806–7815, 2016.

[27] N. Souissi and A. Cherif, "Speech recognition system based on short-term cepstral parameters, feature reduction method and artificial neural networks," in *2016 2nd international conference on advanced technologies for signal and image processing (ATSIP)*. IEEE, 2016, pp. 667–671.

[28] D. Martínez, E. Lleida, A. Ortega, A. Miguel, and J. Villalba, "Voice pathology detection on the saarbrücken voice database with calibration and fusion of scores using multifocal toolkit," in *Advances in Speech and Language Technologies for Iberian Languages: IberSPEECH 2012 Conference, Madrid, Spain, November 21-23, 2012. Proceedings*. Springer, 2012, pp. 99–109.

[29] A. Al-Nasheri, Z. Ali, G. Muhammad, and M. Alsulaiman, "Voice pathology detection using autocorrelation of different filters bank," in *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2014, pp. 50–55.