REPUBLIC OF TÜRKİYE

ALTINBAS UNIVERSITY
Institute of Graduate Studies
Information Technology

SPEAKER VOICE RECOGNITION USING FEATURE SELECTION AND SVM CLASSIFICATION

Hassaneen Ehsan Kareem AL GHAZI

Master's Thesis

Supervisor

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

Istanbul, 2022

SPEAKER VOICE RECOGNITION USING FEATURE SELECTION AND SVM CLASSIFICATION

Hassaneen Ehsan Kareem AL GHAZI

Information Technology

Master 's Thesis

ALTINBAŞ UNIVERSITY

2022

The thesis titled SPEAKER VOICE RECOGNITION USING FEATURE SELECTION AND SVM CLASSIFICATION prepared by HASSANEEN EHSAN KAREEM AL GHAZI and submitted on 16/12/2022 has been **accepted unanimously** for the degree of Master of Science in Information Technology.

Asst. Prof. Dr. Abdullahi Abdu Ibrahim
Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering. Altınbaş University	_____
Asst. Prof. Dr. Ayça KURNAZ TÜRKBEN	Department of Software. Altınbaş University	_____
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical. Istanbul Arel University	_____

I hereby declare that this thesis meets all format and submission requirements Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: ___/___/___

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Hassaneen Ehsan Kareem AL GHAZI

Signature

.

ACKNOWLEDGEMENTS

To Asst. Prof.Dr. Abdullahi Abdu Ibrahim, one of the most significant instructors in my life; it was from him that I learned to draw my first letter, acquire the talent of pronouncing Arabic, create my first complete sentence, and apply the first of many rules. To my colleagues, whose unwavering support I can always count on, and who provided me all the tools I need to complete my thesis without a second thought.

ABSTRACT

SPEAKER VOICE RECOGNITION USING FEATURE SELECTION AND SVM CLASSIFICATION

AL GHAZI, Hassaneen Ehsan Kareem
M.Sc., Information Technology, Altınbaş University,
Supervisor: Asst. Prof. Dr. Abdullahi Abdu IBRAHIM
Date: 12/2022
Pages: 67

Gender recognition based solely on the speaker's voice is a fairly simple task for any human being, however, it's not as simple as it is for humans compared to any computing systems, the task requires multiple tedious processes of feature recognition and selection, and multiple computational processes to acquire such experience in gender recognition. neural networks (NN) has always been the best choice when it comes to image and audio classification and other pattern analysis tasks, the accuracy and precision of the output results widened the prospect of utilizing the different ANN variations in voice recognition, in this paper we present a system design for using K-nearest neighbor network to further enhance the results of the gender detection results by the voice recognition process.

Keywords: Feature Extraction, KNN, Fuzzy Logic, Gender Detection, Voice Recognition.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

ABBREVIATIONS

KNN	:	K nearest number
DWT	:	Discrete Wavelet Transform
CNN	:	Convolutional neural network
PCA	:	Principal Component Analysis
NN	:	Neural network
GD	:	Gender Detection
SVM	:	Support Vector Machine
PNCC	:	Power Normalized Cepstral Coefficients
VAD	:	Voice Activity Detection
CV	:	Cross Validation
CT	:	Computed Tomography
SCSE	:	Supervised single-Channel Speech Enhancement
SNR	:	Signal-to-Noise Ratio
GLCM	:	Gray Level Co-occurrence Matrix
ROI	:	Regions Of Interest
DT	:	Decision Tree
HU	:	Hounsfield unit
LBP	:	local binary pattern

rVAD	:	Record Voice Activity Detection
ZCR	:	Zero Crossing Rate
API	:	Application Programming Interface
ITU-T	:	International Telecommunication Union Telecommunication Standardization
GLN	:	Gray Level Nonuniformaity
RLN	:	Run Length Non-uniformity
RP	:	Run Percentage
k-NN	:	K nearest Neighbors
ANNs	:	Artificial Neural Networks
UBM	:	Universal Background Model
GLRLM	:	Gray Level Run Length Matrix
MFCC	:	Mel-frequency cepstral coefficient

1. INTRODUCTION

1.1 BACKGROUND

There is little doubt that determining a person's gender is a simple task for the human brain to do, but robots are not able to deal with the same quantity of information or time intervals as humans are. There is little doubt that determining a person's gender is an easy task for the human brain to do. To train a computer to provide results that are acceptable and correct when determining the gender of a speaker solely based on the aural input, a massive amount of secondary input data is required. This is because training the computer requires a lot of data. It is not possible to differentiate between male and female voices based purely on the auditory information available. The reason for this is that male and female voices sound identical. The procedure for identifying a person's gender based on audio samples is depicted in Figure 1.1. Which is an example of a general model that can be used to determine the gender of an audio sample by applying deep learning technique:

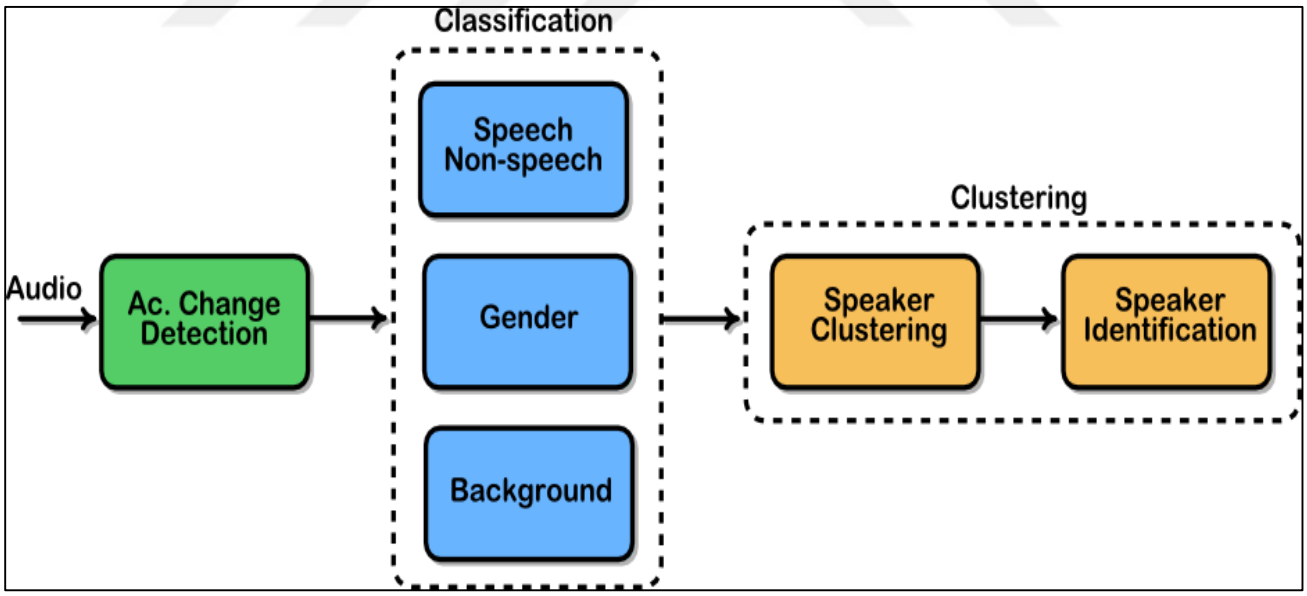

Figure 1.1: Generic model for audio-based gender detection

Three major stages are involved in the process:

i. Filtering off background noise from of the audio sample to avoid using it as input data, which might lead to errors in the final computation.

ii. To be used as training & comparison data, extract the input sample's characteristics.

iii. Compare it to the characteristics retrieved in the second phase, assign a classification to the testing audio sample.

The use of deep learning techniques to audio datasets has lately shown excellent results, as demonstrated by SVM [1], CNN [2], and KNN [3], or artificial neural networks [4] and genetic algorithms [5], as well as other forms of swarm algorithms [6-17]. The graphic below is an example of how deep learning may improve the performance of natural language processing.

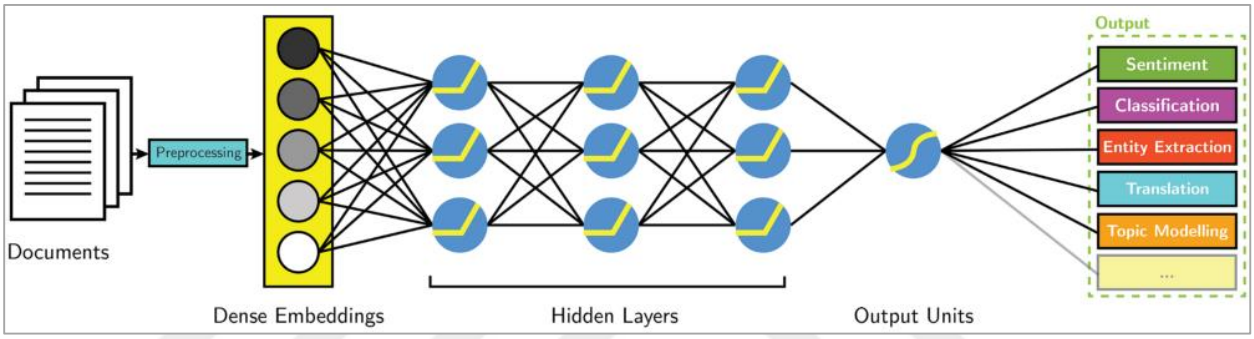

Figure 1.2: CNN as an example of NLP

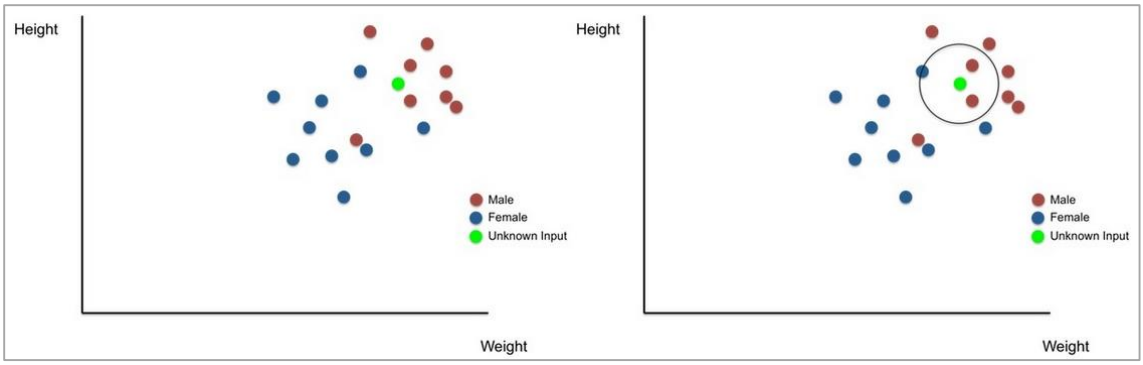

Figure 1. 3: Showing the featured extracted by the KNN algorithm.

KNN is used to solve classification and prediction problems when features from both labels (men and women) interconnect together in such a shared attributes [18], which may cause some difficulties in the classification step while using a hybrid Feature extraction approach like the one shown in the Figure. KNN is used to solve these problems. As a result of this, we decided to use KNN as our classification and prediction method of choice because it is used to tackle problems

of this kind where features from both labels interconnect together in the form of shared attributes.

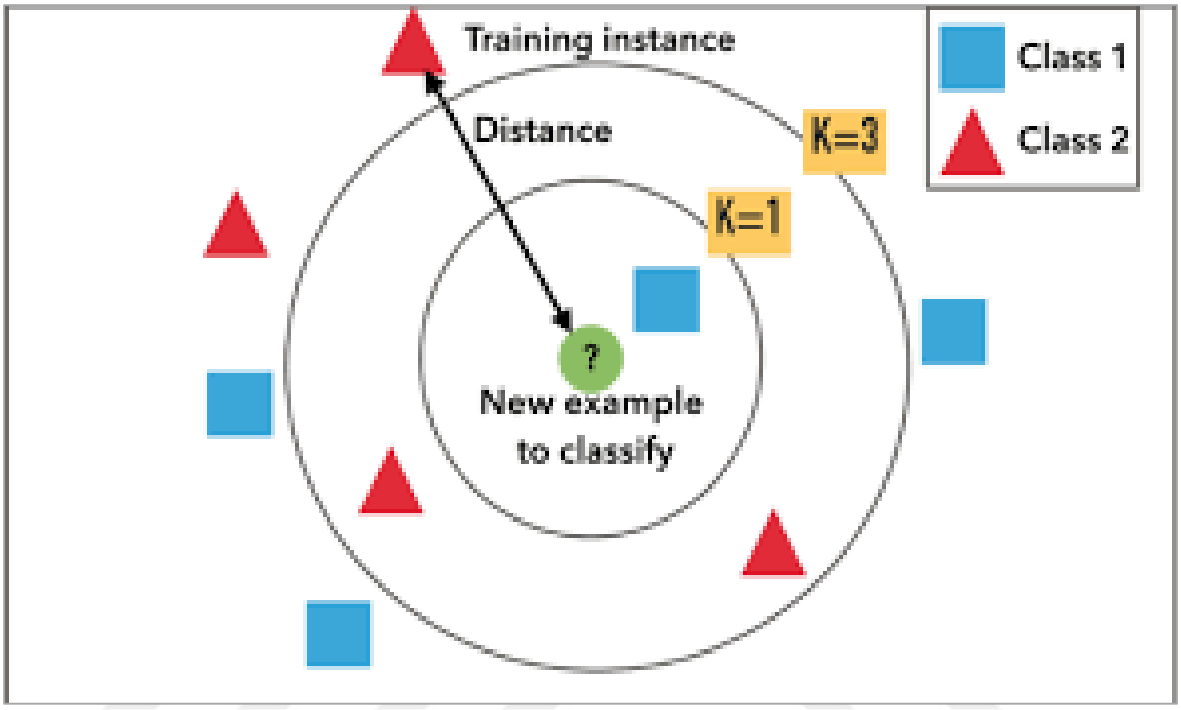

Figure 1.4: The training process of the KNN algorithm.

1.2 PROBLEM STATEMENT

An rise in the amount of interaction that takes place between humans and computers has resulted in the development of algorithms for speaker identification as well as gender detection in biometrics. These improvements have been motivated by the rise in human-computer contact, despite the fact that it is difficult for a computer to discern between male and female voices in an audio sample. When applied to a detection method that makes use of a feature extraction and classification algorithm, the accuracy of the feature extraction and classification processes has a direct bearing on the degree of precision with which gender can be effectively detected. This is because the accuracy of the feature extraction and classification processes is directly related to the accuracy of the detection method.

1.3 CONTRIBUTION

To provide a brief summary, we developed a KNN scheme in order to extract wavelet transform features from an audio file. After that, we used the same algorithm in order to categorize these features into labels. After that, we tested the dependability of the software by importing a fresh test file into the system and performing the same series of operations on it as we had done previously. Because the KNN method is not a feature extraction strategy but rather a classification algorithm, it is typically used in conjunction with a feature extraction strategy in order to improve performance. This is due to the fact that the KNN method is not a feature extraction strategy. This is because the KNN method is not a feature extraction strategy, which is why this result occurred. The KNN is broken down into its component parts in greater depth in the stages that follow this one:

a. Assemble the audio data in the appropriate format.
b. Set the value of k to a negative number.
c. Iterate from one training data point to the whole set of training data points in order to establish the projected classification.
d. Calculate the difference in distance between each row of training data and each row of test data. We will use the Euclidean distance as our distance metric since it is the most often used distance metric. Other metrics, cosine, and others, may also be used.
e. The estimated distances are sorted according to their ascending distance values.
f. To get the first k rows of the sorted array, use the following formula.
g. Get the class that is most often used in the row.
h. Return the class that you were expecting.

1.4 OBJECTIVES

The primary objective of this study is to develop a model for identifying the gender of a speaker based on their voice, which can then be applied in a wide variety of contexts. This project is based on the hypothesis that a fuzzy logic model may be able to outperform a conventional personal computer in terms of the actual performance they are capable of achieving. The parallelism of neural networks as well as the methods that are employed to extract data are to blame for this phenomenon. In conclusion, the goal of this project is to generate a considerable

relative improvement in reaction time, throughput, and energy efficiency so that the model can be operated on an embedded multi purpose system rather than on a fuzzy logic model. This improvement will allow the model to function more efficiently.

In order to provide a clearer picture, the following categories of endeavors will be included within the ambit of the project:

i. An audio signal extracting features system that can extract vocal information that can be used to identify the gender of the speaker should be designed.
ii. Design & train a fuzzy logic system using the feature extractor system.
iii. Use a general-purpose embedded architecture to implement a neural network method utilizing the feature extraction method and the previously trained deep neural network. Working in real-time and optimizing efficiency while also reducing difficulty of the task at hand.
iv. Improving performance by using the architecture's inherent qualities and optimizing reaction time and reducing power usage.
v. Make a comparison between the two designs in terms of reaction time, throughput and power consumption.

1.5 THESIS STRUCTURE

The rest of the thesis is organized as follows: Chapter 2 is where we will review some of the previous works implemented in the gender detection process, in Chapter 3 we will explain the methods and materials used in our proposed method, Chapter 4 is where we will put our proposed method in perspective, in Chapter 5 we will review the results of our method and compare it with other methods used, and in Chapter 6 we will conclude our work and open the door for future work.

2. LITERATURE REVIEW

Some of the earlier approaches for filtering and determining the gender of a speaker are described in this chapter, and some of their findings will be utilized to compare our results with theirs.

2.1 VOICE RECOGNITION

Improved speaker identification was achieved via the use of a range of JITTER & SHIMMER variations in conjunction with speech recognition technology. Moreover, it was discovered [21] that the best results were obtained when the PNCC coefficients were combined with the SHIMMER CV. A hacker classification would categorize the least distorted proposed hostile audio network as goals, but our classifier would label it as such. [22].

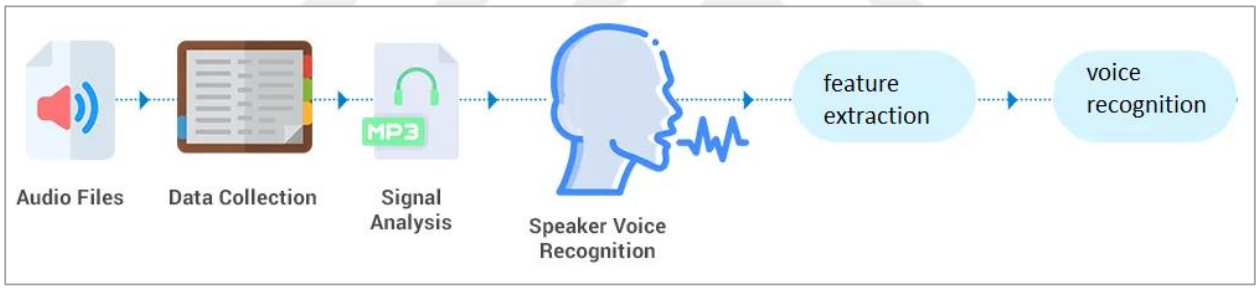

Figure 2.1: Voice recognition steps.

This was accomplished by directly linking the class labels of a audio input to a mixed Gaussian prior, that was then applied to a supervised i-vector model [23], as described above. A novel hybrid feature extraction strategy was developed for the first time by merging the model with binary and ternary patterns, and this was the first time this had been done. [24] When creating an Arabic ASR model, it was important to distinguish between the several dialects of Arabic that were heard. In order to classify illnesses, we created a voice-based classifier that used LBP & CNN to features extracted from a variety of datasets [26].

2.2 GENDER DETECTION

Gender-based activities need a precise characterization of the gender of the individuals involved. The notion of gender definition is not a new one. However, during the past many years, the employment of computer systems to carry out this operation has undergone steady improvement and development. Based on the speech characteristics extracted from the sample and processed by long short-term memory extractors, an accuracy of 98.4 percent was reached when attempting to estimate the gender. With the help of this innovative technique [27].

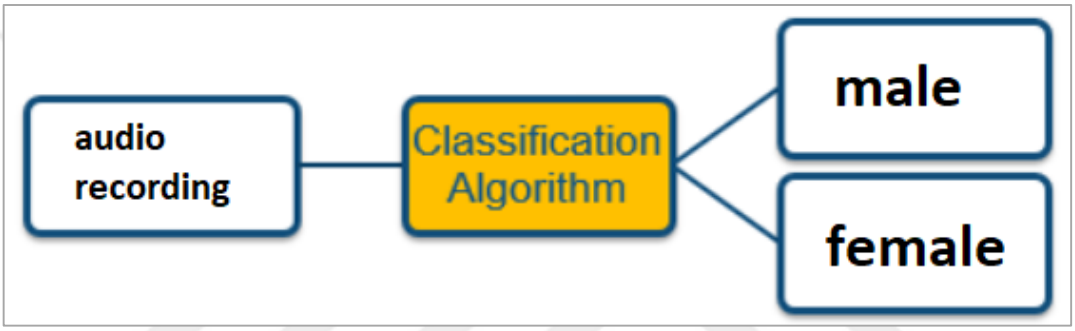

Figure 2.2: Gender detection process.

In order to determine if extensive training of various speech categorization algorithms is beneficial for testing or not, a comparative study was carried out [28]. The model was built on 22 deferment Amazigh characters, and it was utilized to discern between genders of characters according on their voice features [29].

2.3 VOICE FILTERING

The method of speech filtering consists on de-emphasizing and filtering incoming audio in order to remove excess or unwanted information. It has been discovered that it the two-stage rVAD performs best in an environment with high noise and low frequency [30].

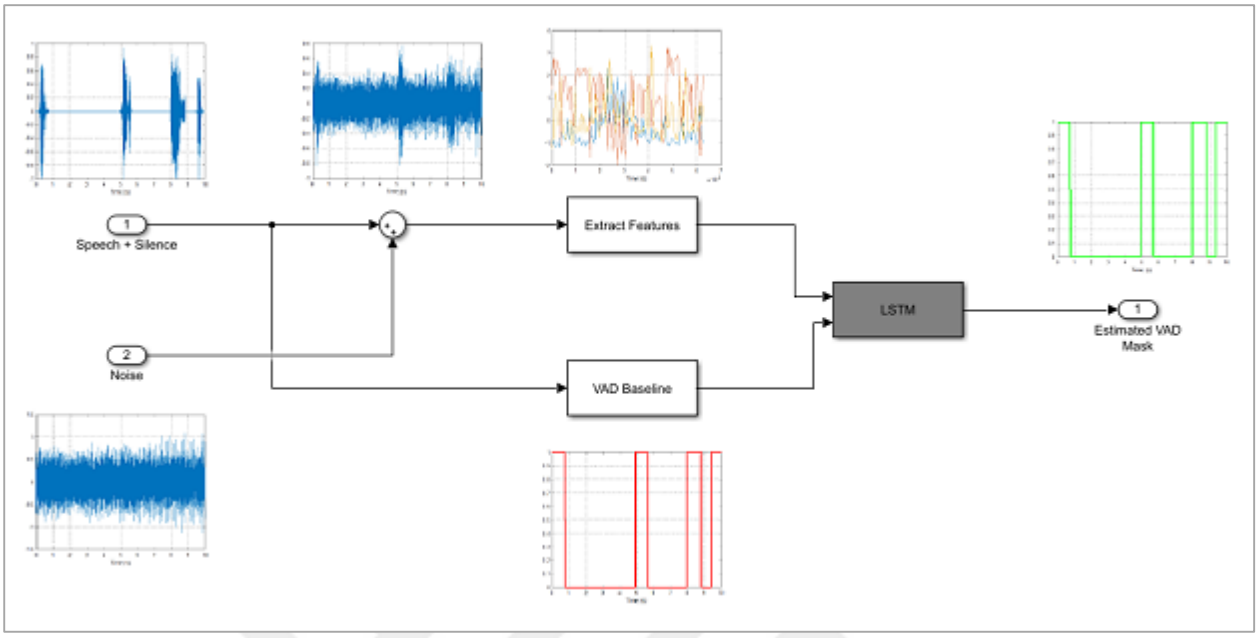

Figure 2.3: VAD and LSTM based model for gender detection.

Developed a novel way for using the parallel processing capabilities of a Bidirectional Recurrent CNN to filter out loud voice samples using a Bidirectional Recurrent CNN A large number of audio samples were surveyed and then turned back into audio samples so that they could be utilized as training data for a deep learning speech filtering algorithm [32].

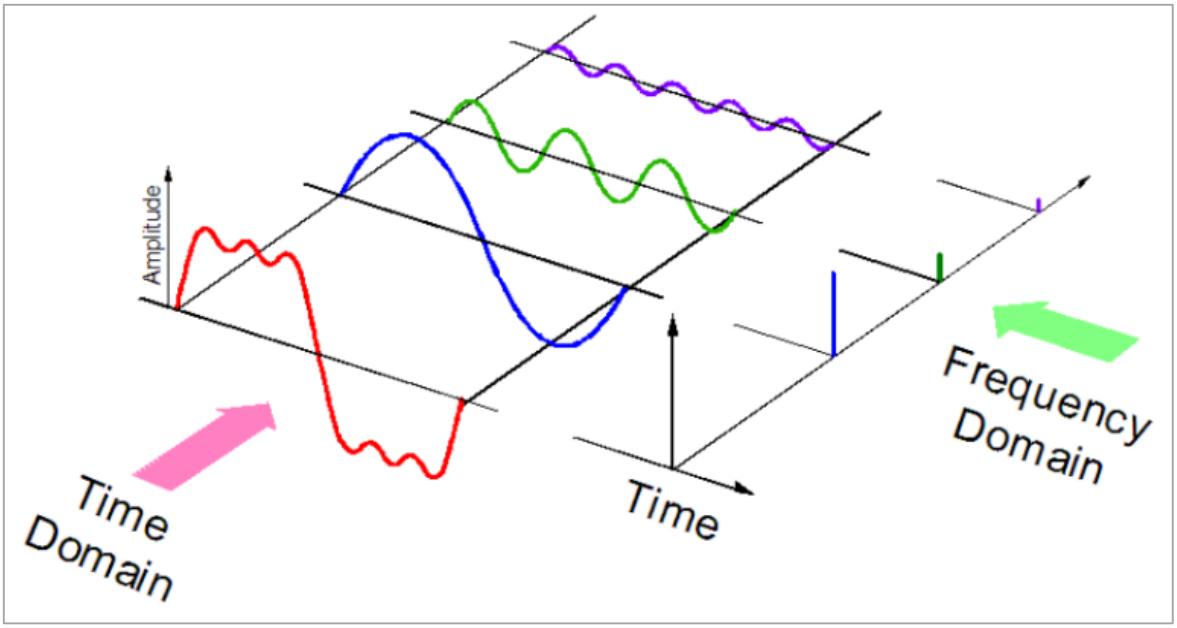

Figure 2.4: Domains of audio filtering.

In order to do Gaussian filtering of audio samples, we investigated and developed an adversarial learning inference (ALI) model. A skin-attached hardware device was developed in order to test multiple deep learning voice filtering algorithms in real time. SCSE approaches [34] for supervised single-channel speech enhancement have been studied. It is necessary to isolate useful information from background noise in order to obtain the most valuable voice recordings [35] from audio samples.

.

3. MATERIALS AND METHODS

In this chapter we give a brief background on some of the techniques implemented in our model and go through some of the reasons why we chose these techniques over similar other ones:

3.1 VOICE FILTERING USING FUZZY LOGIC

Audio filtering is a necessary step for preprocessing any input audio as any noise in the audio file may yield a miscalculation in the final classification output of the method [37], we experimented with many audio filters such as adaptive median filter and low pass filter etc. but the linearity of those filters lead to us searching for a non-linear approach in denoising the audio file, other filters had a hard time deciding if the feature of an audio sample is noise or not [38], we wanted a method that works on that gray area of deciding whether the feature is a noise or an actual audio feature, Fuzzy logic algorithm works on that gray area and showed to give great results when it comes to denoising the audio file [39]. The figures below show the workflow of the fuzzy logic

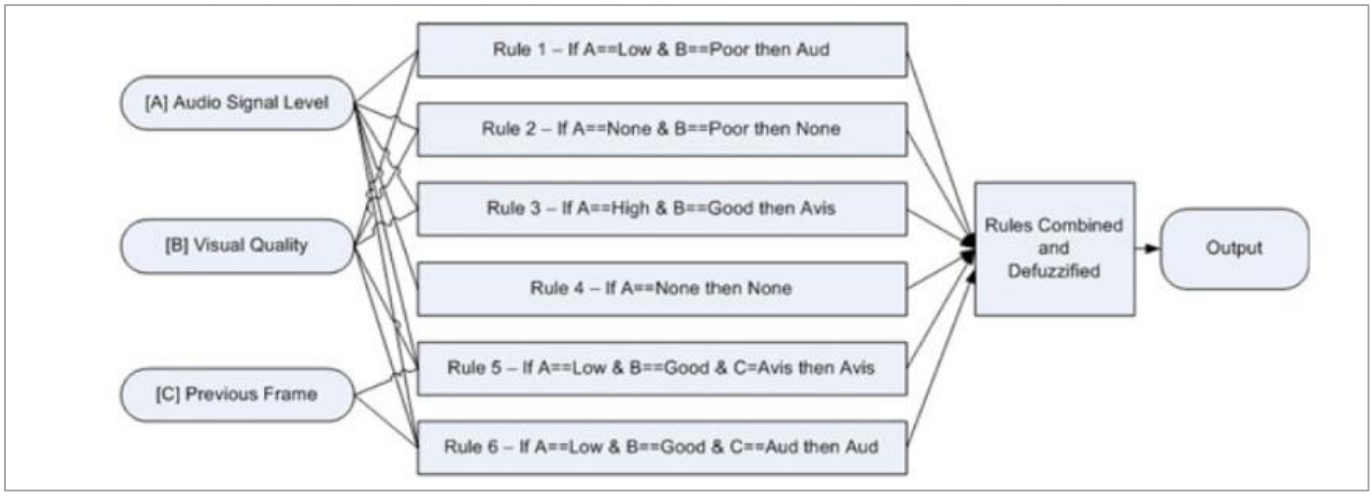

Figure 3.1: General model of the Fuzzy logic audio filtering [40].

audio filtering.

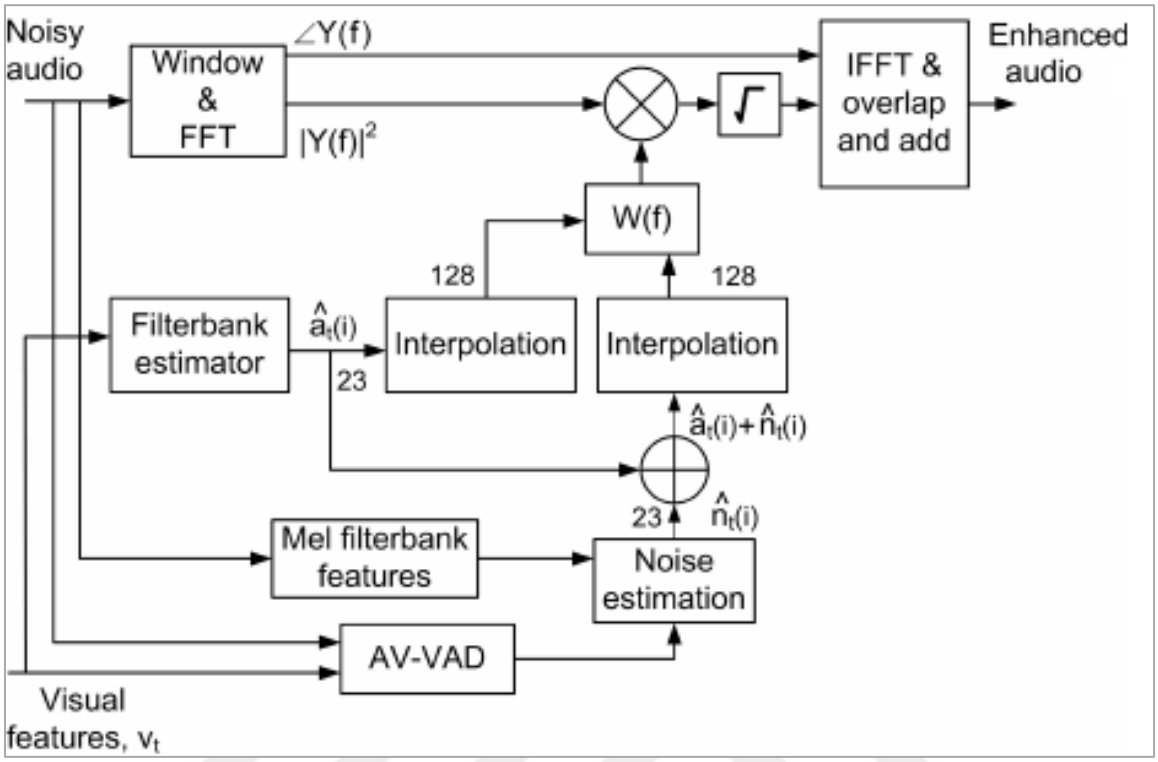

Figure 3.2: MATLAB implementation of the Fuzzy logic audio filtering.

We proposed a new speech interval detection algorithm based on fuzzy c-means clustering to detect speech intervals from noise-intensive speech signals. Figure 3.1 shows the proposed voice interval filtering and enhancement algorithm. It is a volumetric flow chart. First, when a voice signal with mixed noise is input, calculate the entropy of the call. Then, the initial speech section is detected using the calculated entropy. After the signal determined as noise through the detection of the initial speech section is subjected to the fuzzy membership transition technique, the proposed fuzzy membership Transient c-means clustering is used to detect the main negative section. Through this process, a voice interval detection signal with separated speech and noise is obtained.

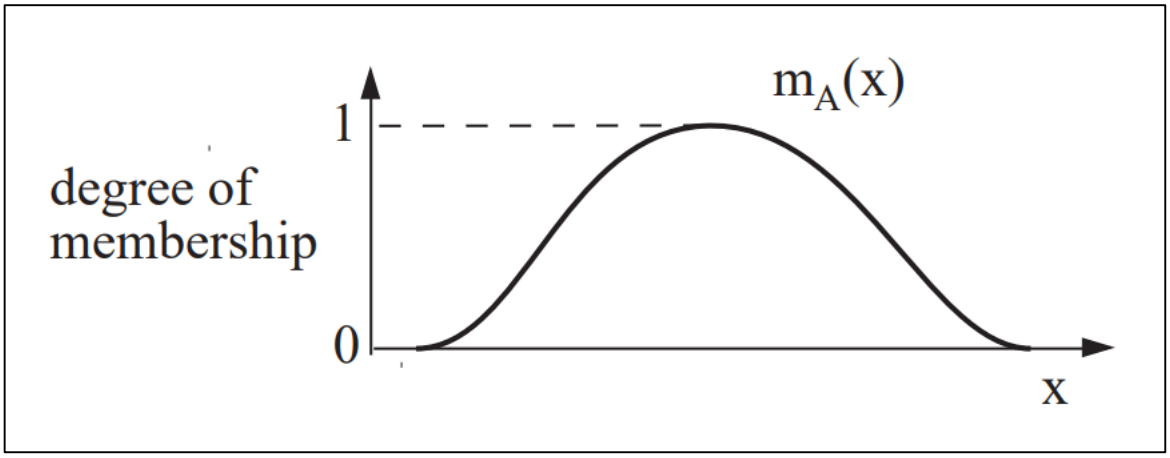

Figure 3.3: Degree of membership in Fuzzy logic audio filtering.

Figure 3.2 presents a breathtaking illustration of a set that is available for your inspection. When applied to the fields of information and communication, in addition to signal processing, one is able to conceptualize entropy as an inverse thermostat. This is a viable way to think about entropy. Simply put, "statistical chaos" is what its name means when taken in its literal form. The Brock line, which is a component of the algorithm that is suggested for detecting speech intervals, is also used as a measure of the amount of information that is already present in the data [10]. This is done by comparing the length of the line to the total amount of information that is already there. To accomplish this, a comparison must be made between the amount of information that is currently present and the amount of information that has previously been there. It is possible to extract the statistical features that are inherently present in the data that have been gathered from the audio signal if the degree to which entropy has changed over the course of time is quantified and measured. These data were acquired from the audio signal. An explanation of the following expression for the probability p [x I of the signal x I at a specific point I in the time domain is provided as follows: Within the context of the overall speech signal, the likelihood is stated as the relative magnitude of the signal multiplied by I. This gives an idea of how likely something is. Here, max (x) refers to the value that has increased by the greatest amount across the entirety of the frame, and the entropy H (k) in a certain frame k of a given speech signal is defined in accordance with Equation (2) [10].

Where " signifies the whole length of the frame that was used in the computation of the entropy, and " stands for the symbol ". In addition, the entropy of the time, which is calculated by using

equation (2), is utilized to not only represent the entropy of the signal for continuous data but also to calculate the change over time in the entropy in relation to the standard deviation of the mean entropy and the entropy in the frame. Both of these functions are dependent on the continuous nature of the data. This is achieved by employing the entropy of the time as a means of not only representing the entropy of the signal but also calculating the change in the entropy that has occurred over the course of time. In order to carry out these computations, the entropy of the signal is first applied to the continuous data in order to determine its value. The expression E (t) can be written out in a variety of ways, one of which is demonstrated in the following example [11].

The overall average of H is denoted by the symbol M (H) in this equation, while the overall standard deviation of H is denoted by the symbol S (H). Figure 2 displays the entropy of the speech signal that was obtained by applying equation (3) to it, together with the voice signal that was used in the experiment. Figure 2 also displays the voice signal that was used. Figure 2 shows depicts the voice signal that was utilized in the experiment.

The speech signal is depicted in Figure 2a along with an increase in the amount of white noise equal to 5 dB SNR, and the entropy of the speech signal is depicted in Figure 2b. Both telephone numbers are provided below for your reference and convenience. When (a) is contrasted with (b), it is clear that in (b), which is symbolized by the entropy of the signal, the speech component is greatly increased, whereas the noise element is attenuated. This is because the entropy of the signal represents the signal's degree of disorder. This is because the entropy of the signal is a representation of the value b. This point can be driven home by drawing comparisons and contrasts between the two. This feature was used to its full potential in order to successfully complete the initial speech section detection using entropy, which is covered in this body of work. The corresponding equation is applied at various points along the duration of carrying out this activity (4). In this particular instance, the threshold for the voice section detection was set to 0, allowing the first stage of the process for detecting voice sections to be carried out successfully. The following diagram, which may be seen further down on this page, illustrates this point:

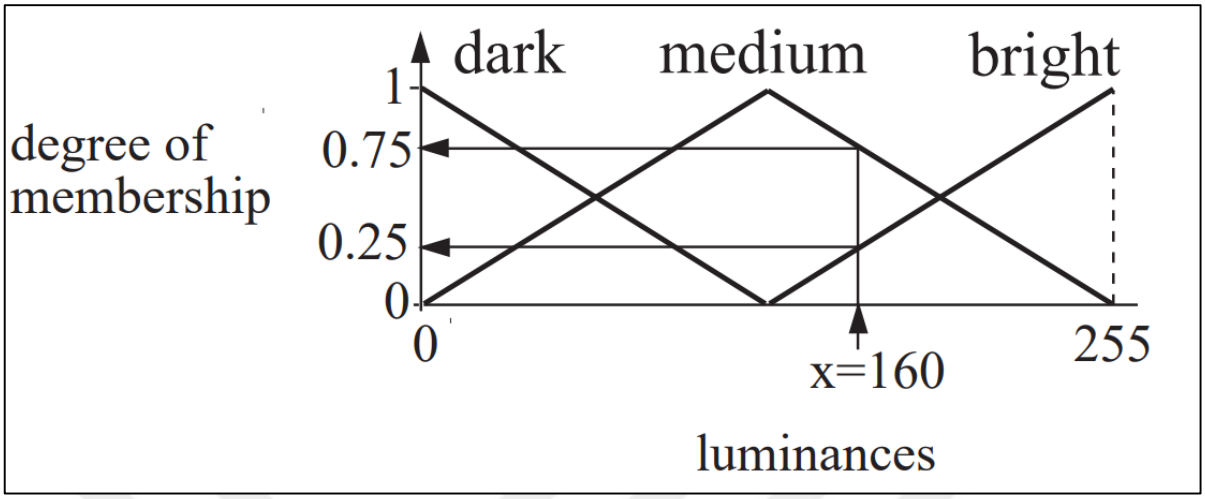

Figure 3.4: Degree of membership grades in Fuzzy logic audio filtering.

In the voice section (VAD, voice activity detection), the fuzzy algorithm can be applied to various voice processing systems such as voice recognition and ventilator, which is an important feature that contributes to the overall performance of the system [15]. In addition, the fuzzy algorithm can be used to detect voice activity in the VAD section. It is possible to make use of the most cutting-edge technological equipment, speech recognition software, and a range of other functions between the texts, which is the place where the presentation material has emerged. Utilize the Wiener filter in addition to the Kalman filter, both of which are from the class of voice recognizers that are currently seeing extensive application. The quinceanera is traditionally performed in the traditional (stop) fashion. This is the most common method. The irregularity elimination stage of the pre-processing stage of speech recognition has finally been achieved, although the process is moving along rather slowly at the moment. The recognition of sounds happens regardless of whether or not there is abnormal noise present. It improves the zero crossing rate (ZCR) and energy efficiency of the current method for feature extraction, while at the same time decreasing the amount of computation and energy that is required for speech recognition performance. Despite the fact that it is no longer employed, the unvoiced sound that is used in Korean should be kept separate from other sounds [18]. There are alternative approaches to feature extraction, like as likelihood ratio (LR) and entropy, that are regarded as being more trustworthy than [4]. These additional approaches include: When listening through statistical features, in point of fact, the extraction of performance features takes place.

Exceptionally high levels of performance are displayed by it [7-8]. In addition, Asgari [9] has strong performance in sentence units in the SNR environment when the speech section of the speech section syllable employing entropy is good in the frequency domain. This is the case when the speech section of the speech section syllable utilizing entropy is good. In an environment with a high signal-to-noise ratio (SNR), this will be the case. The standard voice codec that the ITU-T uses for its commercial communication product, G729, features a VAD algorithm that delivers good performance. It is a performance that does not even come close to meeting the bar. For the purposes of this particular piece of research, the signal entropy is taken into consideration. Suggest. The entropy in the transitional c-average standing time is used as an additional dividing factor in the suggested fuzzy connection. This results in an additional division between the speech segment and the zone. In that location, the temperature tick sound segment and the nasal sound section are distinguished as being the two various types of zones that can be found there. Experiments under strict supervision devised with the purpose of determining whether or not the proposed algorithm is useful. An SNR setting was used throughout the experiment, and the speech signal that was being used was formatted in the form of sentences. In order to perform a comparison between the outcomes of the experiment and the algorithm, a number of graphs and objective values of the voice interval were utilized. The strategy that was advised as a direct consequence of this has been fairly successful as a direct consequence.

3.2 FEATURE EXTRACTION USING KNN

The Fast KNN methodology makes the KNN-based feature extraction functionality available to users. It creates new features with a multiplicity that is equal to k times c, where c is the maximum number of class labels that can be generated. Based on the distance between the visual and its nearest neighbors under each category, the following formula is employed to determine whether or not new features have been added to the visual: The first test feature consists of the distances between each test sample and the sample that is geographically closest to it in the first stage. Within the confines of the first stage, these distances are measured. The second part of the examination involves making a tally of the distances that separate each test sample from the two test samples that are located directly next to the first stage of the examination. The third part of the examination entails making a count of the distances that separate each test specimen from its

three immediate neighbors close to the beginning of the process. And this continues to happen. During this procedure, the label of each class is repeated, which results in the development of new features that are k times c in size. After then, the CV approach was used to incorporate new training features in order to stop an excessive amount of utilization of the older ones. Matches are available. You have the option of setting the number of threads that will be used through the use of the thread parameter. The ideas that were presented in the answer that was regarded to be effective for the classification issue served as the basis for the development of the method of feature extraction that has been suggested in this article.

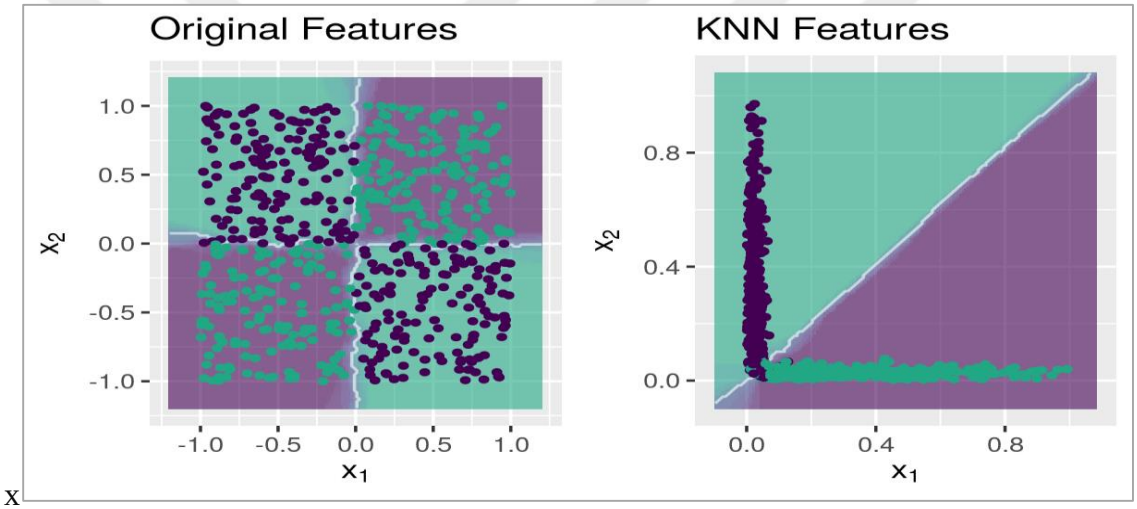

Figure 3.5: Extracted features in from a clear input.

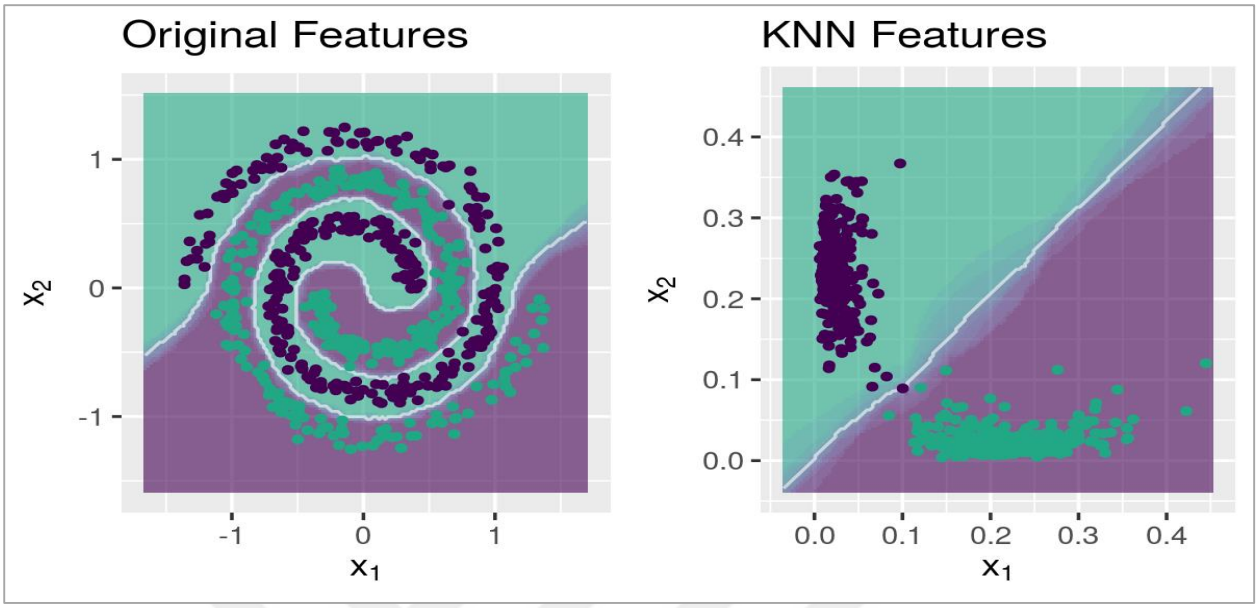

Figure 3.6: Extracted features in from a noisy input.

The newly developed zone of interest served as the foundation for establishing the category of liver free shadow nodule that was present. It is necessary to conduct separate research on each potential topic of interest in order to discover the characteristics that share commonalities with one another. The histogram makes it possible to extract features including those that have a value that is higher than the threshold, the percentile CT attenuation value, and textural attributes. The information pertaining to these four distinct categories of participants is presented in Table 1. The following is a list of characteristics that were determined to be present in The major information that is used by the histogram function comes from the histogram that is associated with the area of interest.

Both the numerical value that was acquired and the brightness level that was considered to be the bare essential need for the liver free shade tubercle were found to be same.

This computation also takes into account the brightness level that is capable of reaching its maximum. One of the qualities of the value that composes the threshold is the fact that The term "highlight pixel" refers to a pixel that, in relation to the brightness values of the other pixels, possesses a brightness value that is more than a particular brightness value. In order to arrive at this particular result, a portion of the total number of cells has been applied as a factor in the calculation. There is not a lot of complexity in the characteristics that make up the range for the limit that has been exceeded.

Both the solid component and the glass shade component both come with a variety of different brightness levels thanks to this feature.

It is measured using a scale that ranges from a value of -1 000 (HU) all the way up to a number of 100. (HU). The amount of brightness, as measured by the percentile CT scale, is as follows: The brightness features correspond to a particular percentile, such as 75% or 97.5%, while the value features correspond to 2.5%, 25%, and 50% of all of the pixels that are located in the region of interest. A region of interest typically has elements of value that can be discovered there.

Both Contrast Road and the contrast matrix (gray)level co-occurrence matrix; GLCM) can be found woven into the fabric of the material. In the process of computing the gray level run-length matrix, also known as GLRLM for its abbreviation, this matrix will be utilized. The computation takes into account both the distance that separates each individual pixel and the direction that they face relative to one another. The morphological and textural properties of the nodule will be covered in this discussion [6,7]. Because of this, there are 286 features in the region of interest that is represented in 2D, 3D857 features in the region of interest that is represented in 2.5D, and 253 features in the multi-view region of interest that is shown in 2.5D. There are a total of 2529 features, and 25D of them can be found inside a 2.5D multi-view-slab zone of interest. There is access to the capability of extracting features. Within the confines of a region of interest that is just two-dimensional, there can be found a total of eight histograms. The following is a list of some of the characteristics that have been retrieved from the data: characteristics, twelve excess thresholds, five percentile CT brightness values, sixteen levels, and one hundred eighty-six different factors to investigate. There were three distances and four directions considered in the analysis of the level 32 KNN features, as well as the same number of directions for both the level 16 and level 3288 KNN features. a zone of interest that is three-dimensional (ROI) In total, the histogram contains 8 features, there are 12 criteria for excess, and there are 5 bags.

In the event that it is desired to do so, the Quartile CT brightness value characteristic can be recovered, along with three distances measured at 16 and 32 levels. There are a total of 546 GLCM features that account for 13 directions across 16 levels, and there are a total of 286 GLRLM features that account for 13 directions over 32 levels. A total of 72 histogram features with brightness values that were higher than 108Threshold and in the 45th percentile of CT were found in the 2.5D multi-view area of interest and the 2.5D multi-view slab region of interest. These features were found in both the area of interest and the slab region. Both level 16 and level 32 were responsible for the discovery of these aspects of the histogram. 15 12 GLCM traits that account for three distances in addition to four directions; 792 GLRLM traits that account for four directions at levels sixteen and thirty-two The retrieval process includes each and every one of the qualities. Relevant for the purposes of classification on the basis of characteristics collected from each area of interest When deciding which characteristics are the most important, selecting them needs very little work but produces a huge number of irrelevant background noise. It was decided that the optimal solution would involve a Relief feature selection method that was less sensitive. C [8]. The day Data that represent a number of different classes are broken down into the respective categories that they are a part of as part of the KNN algorithm. Find the data that has the characteristics that are most similar to the data that was chosen. The data that is regarded to be of the same class is referred to as the "k-nearest hits," and it is called the data that is in the class that is k the closest to it. And the k data that are the most comparable one to another across all of the other categories.

After that, we refer to it as the "k-nearest misses," which is an abbreviation. The efficiency of the characteristics that were chosen for selection all during the process of classification In order to make the calculation of the weights that the attribute represents easier to perform, the entire attribute has been assigned a weight. After the initialization of the teeth, if the k-nearest misses option was selected, the weight of the eater will be dropped if their distance is closer than the k-nearest hits. This will take place only if the k-nearest hits option was also selected. This will take place only if the option to find the k-nearest hits was also selected. It is possible for the weights to fall anywhere on the scale from -1 to 1, and they can. When the value is higher, the categorization is considered to have a higher level of relevance to the topic at hand. Main

We are going to assign a particular amount of weight to each region of interest in order to calculate the total number of features that are present in the dataset. Variability in the relative weights of the qualities that are responsible for the majority of the graph's contributions When determining the 2D region of interest, it is necessary to take into consideration the total number of features that were derived from this. The ten most fascinating aspects of, in addition to the thirty-five most fascinating three-dimensional points of interest Characteristics, one hundred of the most significant features discovered within of the 2.5D multi-view-slab region of interest The top one hundred features in the 2.5D multi-view region of interest are referred to as "features." The top one hundred features in the 2.5D multi-view region of interest are referred to as "features." Displayed for consideration during the screenings Currently, the GLCM feature, which is one of the cross-tipped characteristics, consists of skewness and kurtosis, as well as maximum intensity, the 97.5th of percentile CT brightness value, technical Entropy and its inverse, and the 97.5th of percentile CT brightness value. In addition, it also includes the 97.5th of percentile CT brightness value. In addition to this, one of the cross-tipped characteristics is the maximum intensity.

GLRLM, on the other hand, possesses characteristics that are also present in SRE, GLN, RLN, and RP.

They are chosen, in the vast majority of instances, to represent one of the four main fields of academic research.

Histograms of mixed liver free shade nodules are a few examples of the attributes that have been picked.

The operation of this solid component has a direct impact on the state of health of the pure liver free from tubercle, which allows it to return to its normal level.

According to the findings of the tomography test, the distribution of mixed liver free shadow tubercles appears to be different from patient to patient.

It is evidence that the consistency is more varied than that of a color that is completely free of liver. This can be seen by comparing the two,

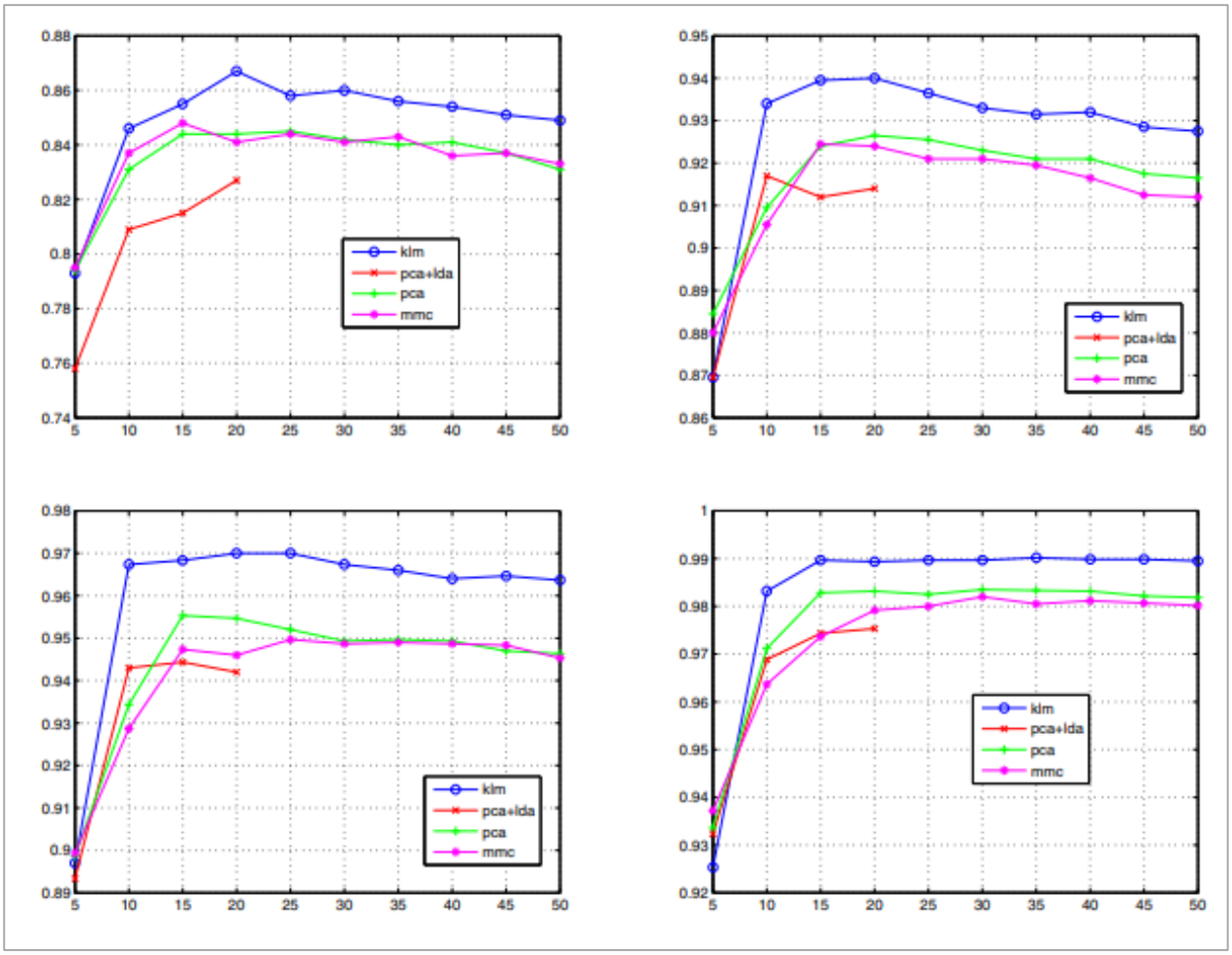

Figure 3.7: Comparing feature extraction accuracy in KNN and other method.

3.3 AUDIO CLASSIFICATION USING KNN

There has been many attempts and approaches used to classify an audio sample based solely on the features of the audio, some of them show good results while others fail to determine the label of the audio sample, this failure is attributed to the lack of parallel processing and the linearity of classifying the similar features of the same audio samples, but KNN algorithm proved to reach better results when it comes to finding similarity of multiple audio sample features in the form of wavelets when its converted to Discrete wavelet transform or DWT, the KNN classification and feature extraction processes is shown in the figures below:

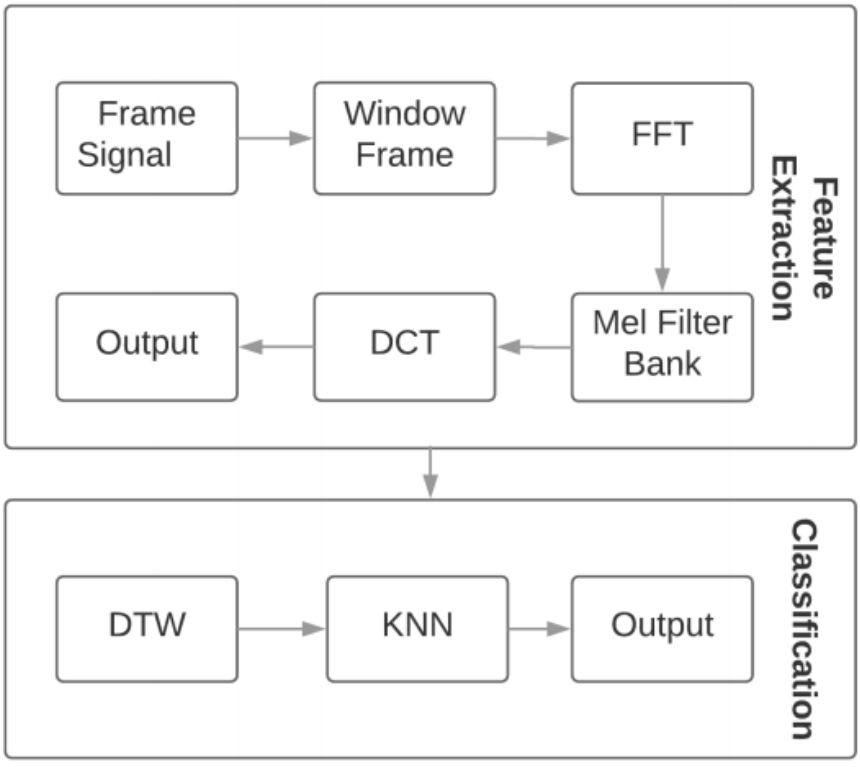

Figure 3.8: Feature extraction and classification process of the KNN.

We can calculate the distance between two scenarios using some distance function, where they are conditions that are compounded by factors, such as. The two functions of distance are discussed in this summary:

Total distance:

$$\boldsymbol{d_{\mathrm{A}}(x,y) = \sum_{i=0}^{N} |xi - yi|} \qquad (3.1)$$

Euclidean distance estimation:

$$\boldsymbol{d_{\mathrm{E}}(\mathrm{x},\mathrm{y}) = \sum_{\mathrm{i}=0}^{N} \sqrt{\mathrm{xi}^2 - \mathrm{yi}^2}} \qquad (3.2)$$

Because the distance between the two conditions depends on the times, it is recommended that the resulting ranges are calculated by the fact that the mean of the statistics in all data is 0 and normal

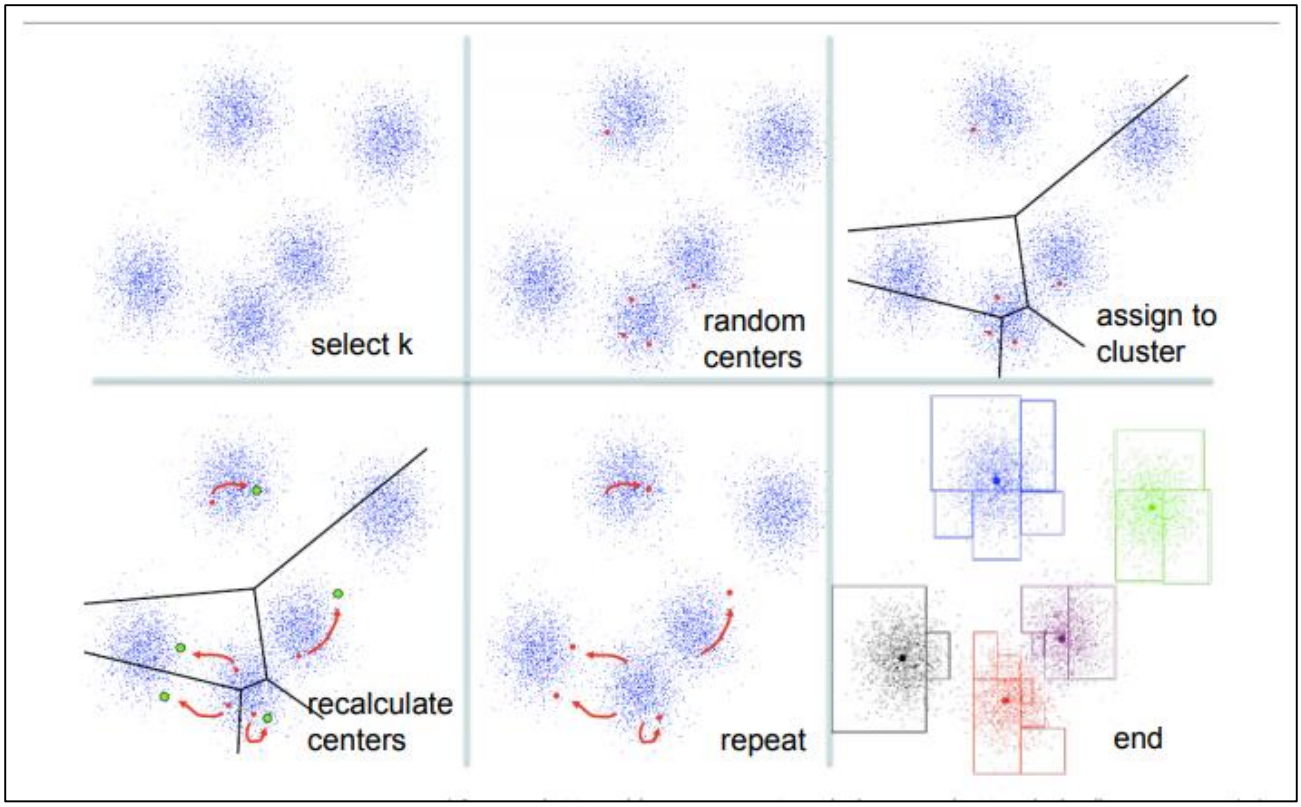

Figure 3.9: Selecting the K in KNN

Deviation 1. This can be done by replacing the color according to the Next activity:

$$x' = \frac{x - x^-}{\sigma(x)} \qquad (3.3)$$

Where the value is excluded, refer to feature statistics for the entire data set, its standard deviations also the sum of the obtained values. Arithmetic mean is defined as:

$$\bar{\mathbf{x}} = \frac{\mathbf{1}}{N}\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{xi} \qquad (3.4)$$

Then we can calculate the standard deviation as follows:

$$\boldsymbol{\sigma}(x) = \sqrt{\frac{\mathbf{1}}{N}\sum_{i=1}^{N}(xi - x^-)^2} \qquad (3.5)$$

Remote operations As mentioned earlier, we only discuss full and Euclidean grade functions. However, we may choose to offer real, hidden values or uses convert them to using the equation function

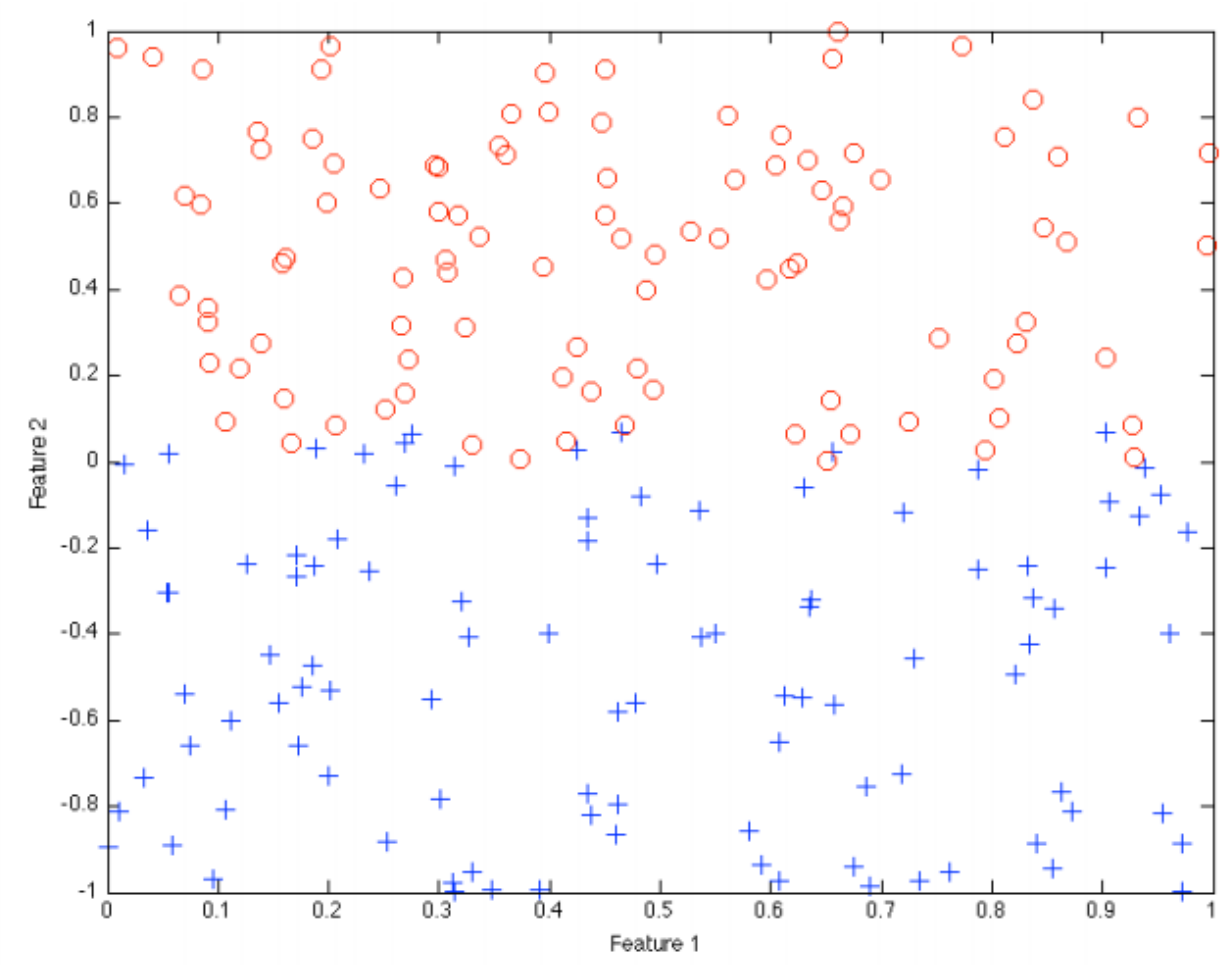

Figure 3.10: Classifying the K in KNN

This technique is applied for the purpose of doing data analysis throughout the course of the supervised learning process, and more specifically when dealing with challenges pertaining to categorization. The categorization problem refers to the challenge of determining which category an existing collection of data currently belongs to when new data is introduced. This can be a tough task when new data is introduced. Because of this difficulty, it may be difficult to determine which category an existing piece of data currently belongs to. The k-NN approach is a forward-thinking strategy, and "since is the group whose data it matches closest, then is the group," says that it needs to be applied. It is feasible to classify it as a specific kind of algorithm.

[Case in point:] [Case in point:] The value of k is a number that reflects the number of times that one has to evaluate the data that is geographically nearest. This number may be thought of as representing the number of times that one must evaluate the data. Within the parameters of this discussion, the value of k is denoted by the number. Today, shall we have a look at an example that is a little bit more concrete? Why don't we go ahead and try that? Assume that there is one more piece of data N in addition to the six current pieces of data A to F, as shown in figure (3.11) below. This would mean that there are a total of seven pieces of data. Because of this, there will now be a total of seven different bits of data. This would suggest that there are a total of seven distinct pieces of information that need to be taken into consideration at this time. In the situation in which k is equal to one, the newly obtained data will be categorized only on the basis of the C that is most similar to the other three options. This will occur in the case in which k is equal to one. As a consequence of this, the letter N is positioned under the heading designated as, which is the same heading as the letter C. This is a direct result of the circumstances that have arisen. When the value of k is equal to three, the newly acquired information is organized into categories by taking into account the letters C, D, and E because these are the ones that are located in the location that is closest to the third distance. When the value of k is equal to two, the information is organized into categories by taking into account the letters A, B, and D because these are the ones that are located in the Everyone is on board with the concept of voting in accordance with the path that receives the majority of votes in the event that there is dispute within the group right at this very minute. At this very moment, it will transform into the ratio 1: 2, and the symbol that will appear next to it will denote the variable N.

If k is equal to 5, then the most accurate way to classify newly obtained data is to search up to the fifth nearest C, D, E, and B accordingly. This is because this approach searches up to the fifth nearest neighboring value. This will guarantee that the facts are categorized in the most logical and useful way possible. At this particular moment, the ratio shifts to three to two, and N is classified as an a. N can be understood in a number of different ways, each of which is different depending on the amount of k that are present. These methods are dependent on one another. Despite the fact that nothing has changed in regard to the facts, this is still the case. The phase in this method that involves picking the proper value for k is one that is extremely necessary. (2) The difficulty caused by the several units of measurement that are used for determining distance - Work needs to be completed before the standardization k (Euclidean

distance) can be computed. This work must be done before the distance can be measured. Standardization. When attempting to calculate the Euclidean distance, it is of the utmost importance to choose the unit that is most appropriate for the calculation. This is due to the fact that several units each have their own unique relationship to the computation. The proximity of two points to a k-NN reference frame is used to determine the Euclidean distance between those points and any other two points.

Take for example the data that was displayed earlier on the y-coordinate; the symbol for the dollar ($) was used as the unit of measurement when referring to that specific instance of the data.

As of right now, the Euclidean distance between point A and point N is 3.162, but the Euclidean distance between point B and point N is only 2.828. As a consequence of this, point B is located in a location that is situated in a location that is closer to point N than point A is.

On the other hand, let's assume that we come to the realization that the circle ought to serve as the new standard for the measurement of y-coordinates. In the event that one dollar is equivalent to one thousand won, the graph will be altered so that it more accurately depicts the situation. This will ensure that the data presented is always accurate.

It is possible to draw the conclusion that point A is located in a position that is more advantageous to point N given that the Euclidean distance between point A and point N is currently 1000.004, whereas the Euclidean distance between point B and point N is 2000.001.

In this way, the distance varies depending on what the unit of each variable is, and as a consequence of this, the order in which things are determined to be closest to one another also shifts around. In other words, the order in which things are determined to be closest to one another is also shifted around. In other words, the order in which things are determined to be closest to one another is also subject to change. This means that the order in which things are closest to one another can change.

Because this indicates that the order is not the same, the classification result for k-NN will be different because the order is not the same in this case. It is possible to draw the conclusion that the order is different given that the fact that it is different indicates that the order is different. As

a consequence of this fact, the k-NN method necessitates that the data be normalized in advance of its application in order for it to function in an appropriate manner. This is an essential requirement for the successful operation of the technique. members of the K who are considered to be of the highest prestige. You are strongly encouraged to return to this page in order to acquire knowledge regarding how to determine the value of k that is the most precise. This is as a direct result of the fact that. To put it another way, you are able to check and determine the quantity of k that is required for an accurate classification of validation data based on train data. This can be done by using the train data. Using the data from trains is one way to accomplish this goal. It is possible to achieve this objective by making use of the information that was gathered in the past.

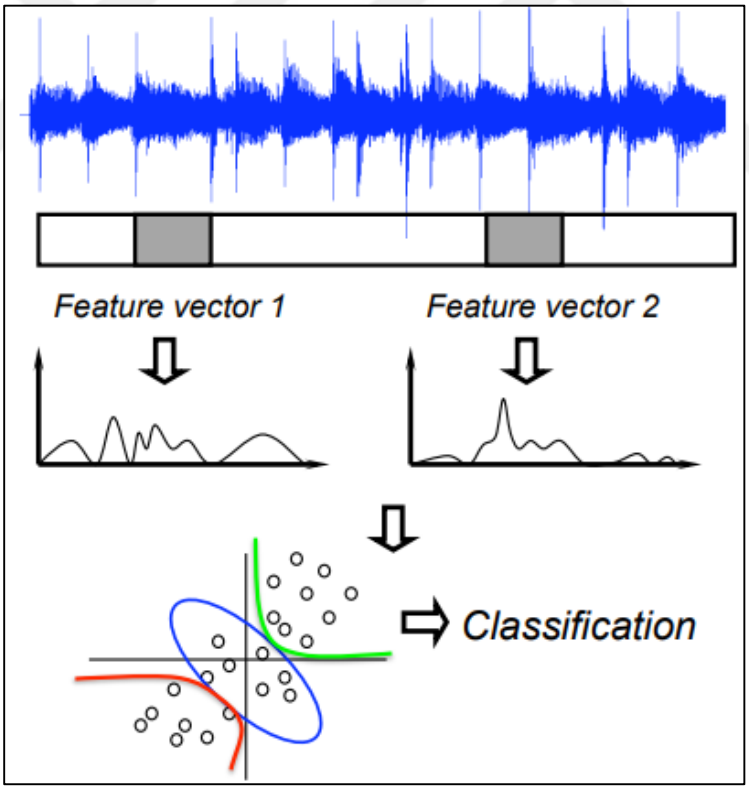

Figure 3.11: Proposed KNN model.

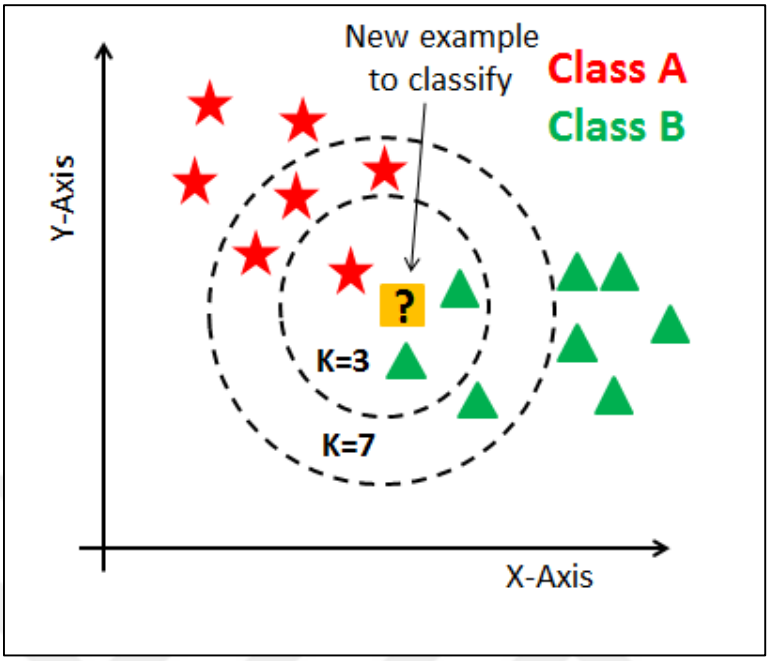

Figure 3.12: Proposed KNN workflow.

The k-nearest neighbors algorithm is a supervised learning classifier that is non-parametric and makes use of proximity in order to produce classifications or predictions about the grouping of an individual data point. It does this by taking into account how closely related data points are located to one another. It is able to accomplish this by taking into consideration the proximity of the data points that are associated to one another. This is made possible by taking into account the geographical proximity of the data points that are connected to one another in some way. This can be accomplished by conducting an investigation into the data point in issue in terms of the context of its immediate surroundings. The amount of data points that are believed to be located in the most immediate vicinity to one another is where the expression "nearest proximity" got its start. This is also where the name of the methodology derives from, so it's fitting. It is also sometimes referred to as KNN or k-NN, the latter of which is the more frequent abbreviation, especially in certain circles. In certain communities, it is also commonly referred to as k-NN. The categorizing process is when this method shines the brightest and is used the most frequently. This is as a result of the fact that it is dependent on the idea that points that share similar qualities may be located in close proximity to one another. The reason for this is that it is predicated on the assumption that comparable features can be found. This is because putting the idea into practice is the part of the process that presents the fewest challenges to the average

person. This presumption serves as the foundation upon which the method is built and is therefore referred to as the cornerstone. It is used for classification the vast most of the time, despite the fact that it may be used to solve problems involving regression or classification. However, the vast majority of the time, it is used as a method for classification.

In classification problems, a class label is issued on the basis of a majority vote, which means that the label that is utilized is the one that is most frequently represented around a specific data point. This means that the vote that determines which label is used is based on the frequency of representation. This indicates that the vote that decides which label will be used will be dependent on the amount of times each representation is cast in the vote. This suggests that the vote that determines which label will be used will be depending on the number of times each representation is cast in the vote. Specifically, the vote will decide which label will be used based on how many times each representation is cast. This leads one to believe that the vote that decides which label will be used will be dependent on the number of times each representation is cast in the vote. To be more specific, the number of times each representation receives a vote will determine the winner of the vote, which will then determine which label will be utilized. Because of this, one could be led to believe that the vote that determines which label will be used will be based on the number of times each representation is cast in the vote. To be more exact, the winner of the vote will be determined by the number of times each representation obtains a vote. The winner of the vote will then choose which label will be used. Even though the process at hand is more accurately referred to as "plurality voting," the phrase "majority vote" is the one that is used the most commonly when referring to voting outcomes in written works. This is despite the fact that "majority vote" is the one that is used the most frequently. Majority vote This is due to the fact that the method in question is more accurately described by the term "plurality voting," which is a more contemporary expression. These two interpretations of "majority voting" are distinct from one another due to the fact that "majority voting" in its purest form requires a majority of more than 50%, and because it functions at its most efficient level when there are only two distinct categories from which voters can choose, the purest form of "majority voting" requires a majority of more than 50%. It is possible that you will not require fifty percent of the votes in order to come to a decision on a class when there are multiple classes, such as when there are four different categories from which to choose. This is the case when there are a number of classes. Because there are so many classes, this scenario is not

impossible to realize. You are allowed to affix a label to a category so long as there is support for doing so from more than twenty-five percent of the votes that have been cast. This is the minimum threshold necessary. After that, you will have the opportunity to give the category a name of your choosing when you proceed to the following step. We are indebted to the University of Wisconsin–Madison for providing us with a link to the aforementioned document, for which we are really thankful (PDF, 1.2 MB). This offers a highly convincing explanation of the subject matter that was covered in the sentence that came before this one (link resides outside of ibm.com).

Both regression and classification problems make use of a similar concept; however, in the case of regression problems, the average of the k nearest neighbors is used to generate a prediction regarding a classification. In classification problems, on the other hand, the average of the nearest neighbors is used. In contrast to this, categorization difficulties employ an entirely new notion to solve the problem at hand. When trying to solve problems involving classification, you have to apply an altogether other concept. In contrast to this, classification problems always make use of an entirely new and different concept throughout the entire process. In contrast to this, the solution to problems with categorization involves thinking about the problem that has to be handled in a manner that is wholly original to itself. In order to find a solution to the problem, it is required to do this first. The statistical research approaches known as classification and regression are fundamentally distinct from one another due to the fact that classification is utilized for the analysis of discrete data while regression is utilized for the analysis of continuous data. This is the fundamental distinction that can be drawn between the two different statistical research methodologies. This is the most important distinction that can be made between the two methods of statistical research. This is the single most important distinction that can be made between the two approaches to statistical analysis. Both approaches are important in their own right. However, in order to construct any kind of categorization at all, the distance needs to be measured first in order to make certain that the results are accurate. Only then can any sort of categorization even be constructed. Then and only then is it possible to establish any kind of categorization at all. The computation of the Euclidean distance is by far the most common approach, and in the paragraphs that are going to follow, you will have the opportunity to gain more information regarding to this topic area.

It is essential to keep in mind that the KNN algorithm is a member of a family of models that are collectively referred to as "lazy learning," and it is equally essential to keep this fact in mind. It is essential to keep in mind that the KNN algorithm is a member of a family of models that are collectively referred to as "lazy learning." It is essential to keep in mind that the KNN algorithm is a member of a family of models that are collectively referred to as "lazy learning," and it is essential to keep in mind that this family of models was developed by Google. It is very important to keep in mind that the KNN algorithm is a member of the same model family as the other models. Keeping this in mind will help ensure that you don't make any mistakes. The phrase "store and forward" is where the term "store and forward" originates from. The process in which an algorithm does not go through a training step but rather merely saves a training dataset is called "store and forward," and the term comes from the phrase "store and forward." This is how the term that is being questioned should be understood. It is essential that, as part of your analysis, you take into account this facet of the situation that has arisen. Because of the aforementioned fact, each and every piece of computation is finished at the exact same time that a classification or a prediction is being made. This allows for a seamless transition between the two processes. This guarantees that the reliability of the results is not jeopardized in any way. This ensures that the results obtained from each of these processes will be comparable to those obtained from the others. This is the conclusion that one must reach in order to make sense of the information that was presented earlier in the conversation. This method is also known as a memory-based learning approach and an instance-based learning method from time to time. Both of these names are also used interchangeably. Both of these terms adequately describe the action that is taking place. This is as a result of the fact that it keeps all of its training data in memory and places a significant amount of reliance on memory to accomplish this. This method is also known as "learning through instance," which is another term for it. [Note: This is because it places a significant emphasis on the utilization of one's memory, which is the reason for the result that was mentioned above.] The reason for this is because the utilization of one's memory is the reason for the result. The underlying reason of the issue is the manner in which an individual's system makes use of their memories, which is why this occurs

4. PROPOSED METHODOLOGY

We implemented a model to preprocess the input audio by filtering the unwanted and redundant noise using the Fuzzy logic audio filtering and the output of that filtering is an input to the next step in our model, then we implement a KNN based scheme to extract the features from the filtered audio, then classify the sample features into 2 classes either male or female the model of these processes is shown in the diagram below:

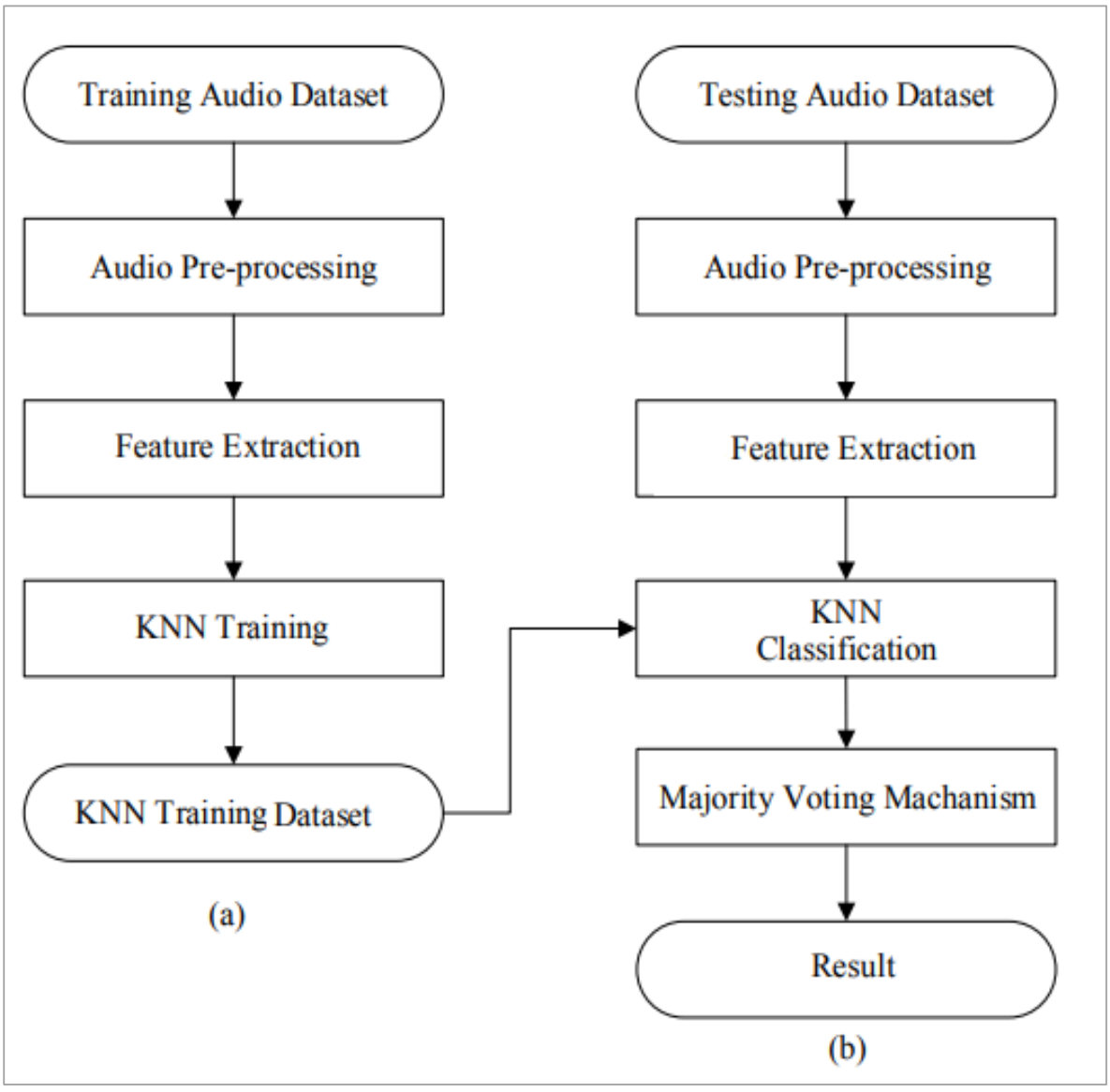

Figure 4.1: Feature extraction and classification process of the KNN where (a) shows the training process and (b) showing the testing process.

The below steps further explain the flow work of our model

A. first, we filter out any unwanted background noise that may affect our results by confusing the algorithm into thinking that they are features of the input audio using the Fuzzy logic filter as show in the figure below:

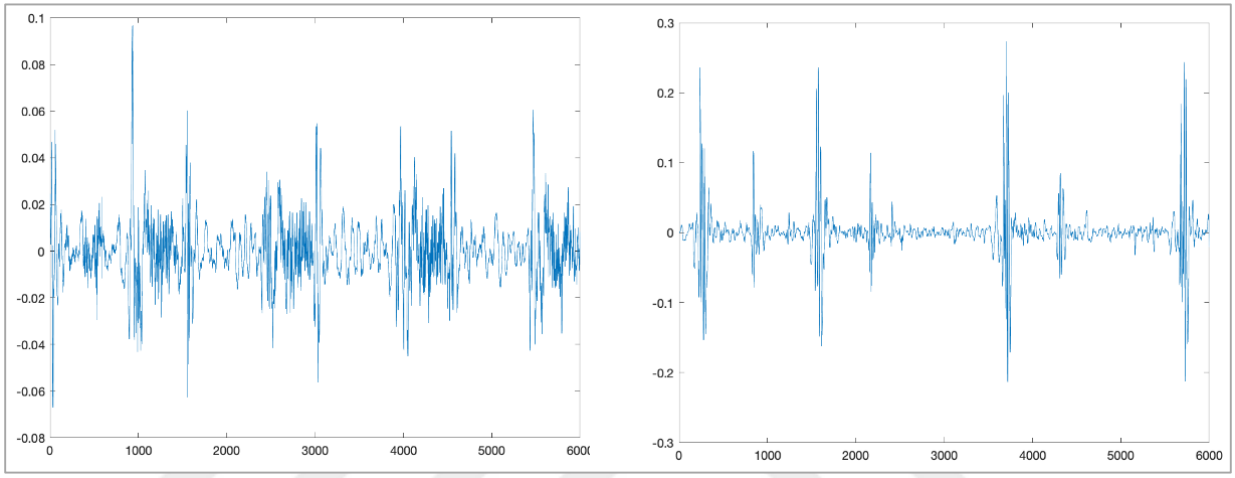

Figure 4.2: Input audio before and after the noise filtering.

```
subplot(4,1,1);
plot(sample_data);          % Original Signal
title('Original Signal');
xlabel('Time (s)'); ylabel ('Amplitude');
subplot(4,1,2);
plot(filtered_sound2);        % Filtered output
title('Filtered Signal Double');
xlabel('Time (s)'); ylabel ('Amplitude');
subplot(4,1,3);
plot(fs2_d);       % Filtered output
title('Filtered Signal Double');
xlabel('Time (s)'); ylabel ('Amplitude');
subplot(4,1,4);
plot(Clean);               % Expected output
title('Expected Signal');
xlabel('Time (s)'); ylabel ('Amplitude');
```

```
[Clean, Fs2] = audioread('expected.m4a');
[sample_data, sample_rate] = audioread('male female.m4a');
signal = medfilt1(sample_data,90);     % Applying median filter
%_____________________________________________________________
_____%
Fs = sample_rate;                  % Sampling Frequency (Hz)
Fn = Fs/2;                         % Nyquist Frequency (Hz)
Wp = 1000/Fn;                      % Passband Frequency (Normalised)
Ws = 1010/Fn;                      % Stopband Frequency (Normalised)
Rp =  1;                           % Passband Ripple (dB)
Rs = 150;                          % Stopband Ripple (dB)
[n,Ws] = cheb2ord(Wp,Ws,Rp,Rs);        % Filter Order
[z,p,k] = cheby2(n,Rs,Ws,'low');       % Filter Design
[soslp,glp] = zp2sos(z,p,k);   % Convert To Second-Order-Section For Stability
filtered_sound2 = filtfilt(soslp, glp, signal);
%_____________________________________________________________
_____%
fs2_d = [filtered_sound2,filtered_sound2];
sound(fs2_d, sample_rate)
```

Figure 4.3: The code for the fuzzy audio filter

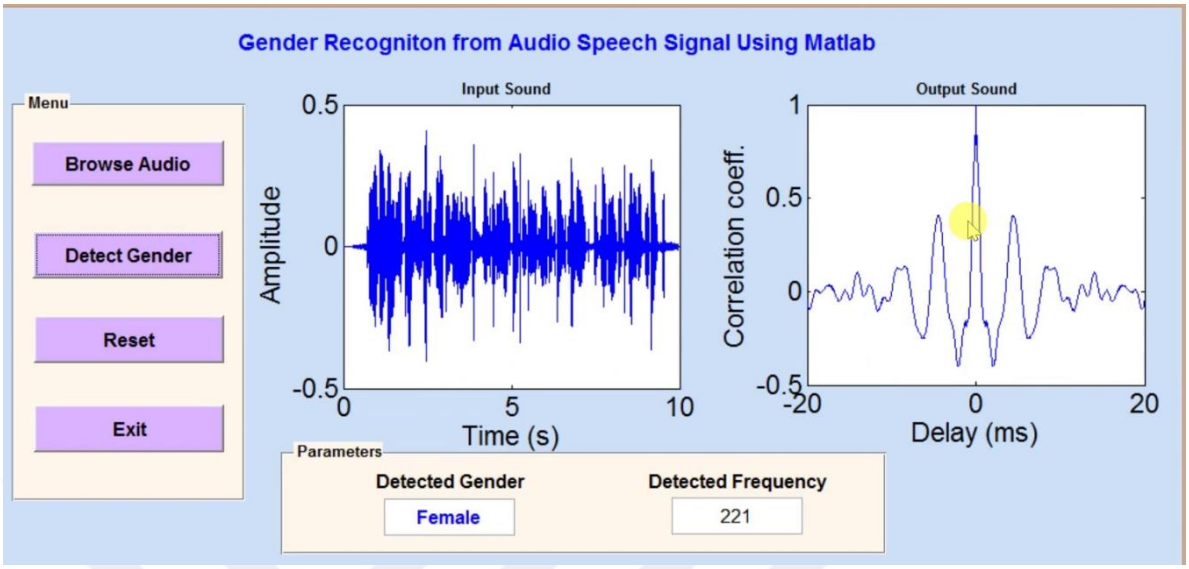

Figure 4.4: MATLAB UI of the implementation of the proposed method.

1- second, we define the 26 features to our input audio and define the names of our classes and train the algorithm to extract those features from our samples and increase the number of training iteration to achieve more accurate feature predictions as shown in the figures below:

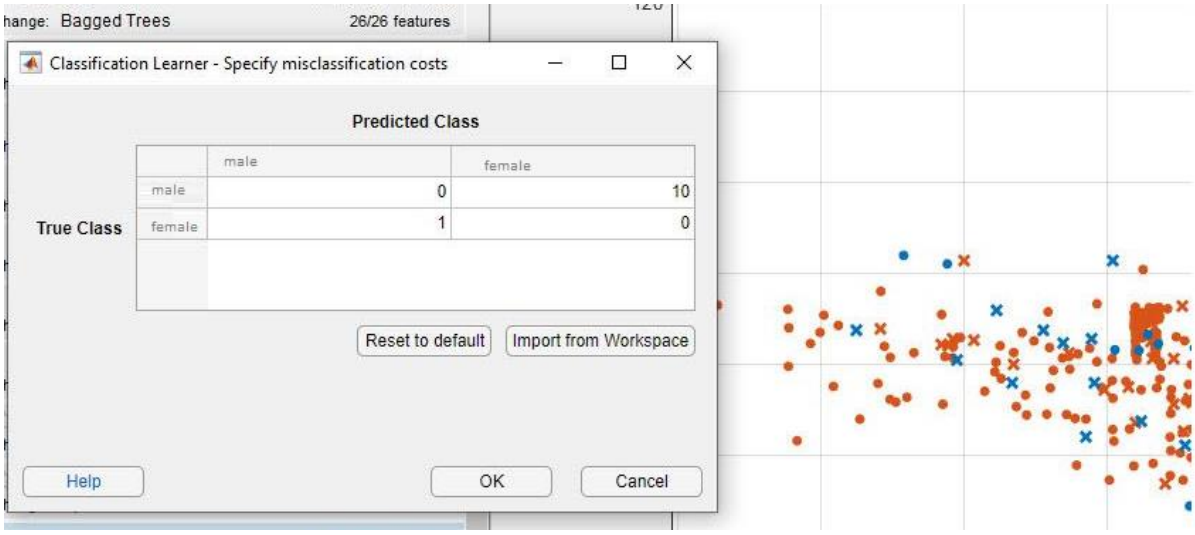

Figure 4.5: Predicted features of a single audio sample with 26 features.

The number of predicted features in the training step is shown in the figure below:

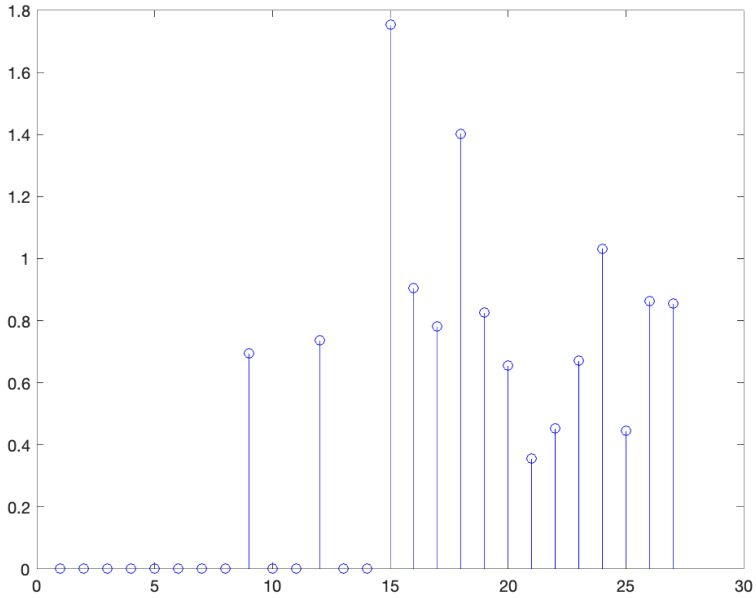

Figure 4.6: Number of features extracted by the KNN.

```
feature_table_all = table();
labelMap = containers.Map('KeyType','int32','ValueType','char');
keySet = {-1, 1};
valueSet = {'male','female'};
labelMap = containers.Map(keySet,valueSet);
while hasdata(fds)
   PCG = read(fds);
   current_class = reference_table(strcmp(reference_table.record_name, audio.filename),
:).record_label;
   N = length(signal);       % Bernhard: to keep track of #samples in files we process
   number_of_windows = floor( (N - overlap_length*fs) / (fs * step_length));
   feature_table = table();
   for iwin = 1:number_of_windows
      current_start_sample = (iwin - 1) * fs * step_length + 1;
      current_end_sample = current_start_sample + window_length * fs
```

Figure 4.7: The feature extracted are listed in a matrix named Feature Matrix.

```
% Extract features from the power spectrum
    [maxfreq, maxval, maxratio] = dominant_frequency_features(current_signal, fs, 256, 0);
    feature_table.dominantFrequencyValue(iwin, 1) = maxfreq;
    feature_table.dominantFrequencyMagnitude(iwin, 1) = maxval;
    feature_table.dominantFrequencyRatio(iwin, 1) = maxr
    % Extract features
    % REMOVED because didn't contribute much to final model (only 1 of
    % them was selected by NCA among the "important" features)
```

Figure 4.8: The percentages of the acquired features in the training process.

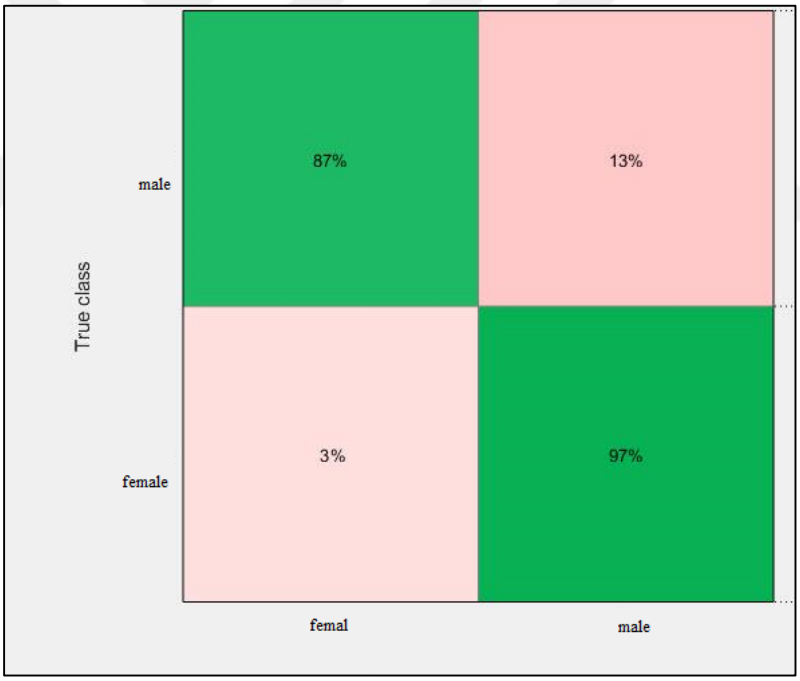

Figure 4.9: The success rate in predicting the classes,

The successfully predicted class based on the audio sample features is shown in the figure below:

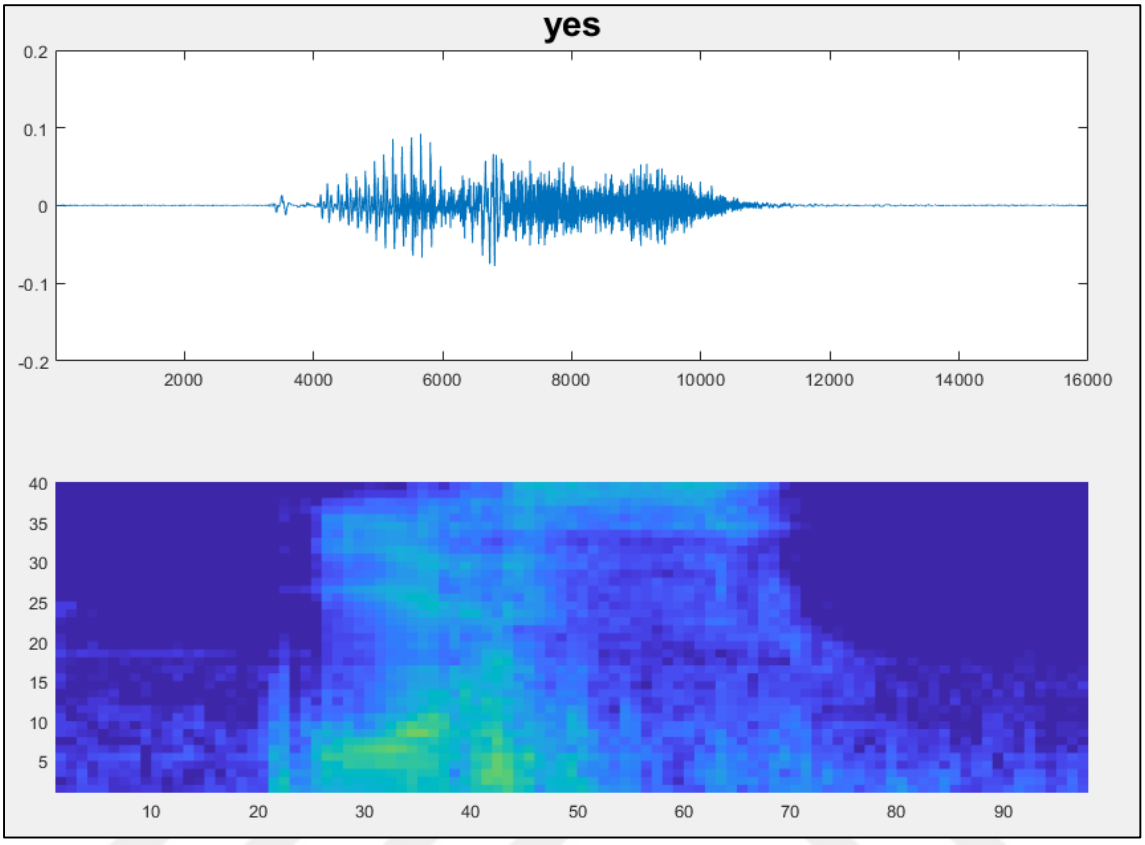

Figure 4.10: Feature extracted by the KNN from the audio sample.

While some unknown features may appear, we define the 26 features to our input audio and define the names of our classes and train the algorithm to extract those features from our samples and increase the number of training iteration to achieve more accurate feature predictions as it is best to classify as unknown rather than give wrong predictions we filter out any unwanted background noise that may affect our results by confusing the algorithm into thinking that they are features of the input audio using the Fuzzy logic filter as shown in the figure below:

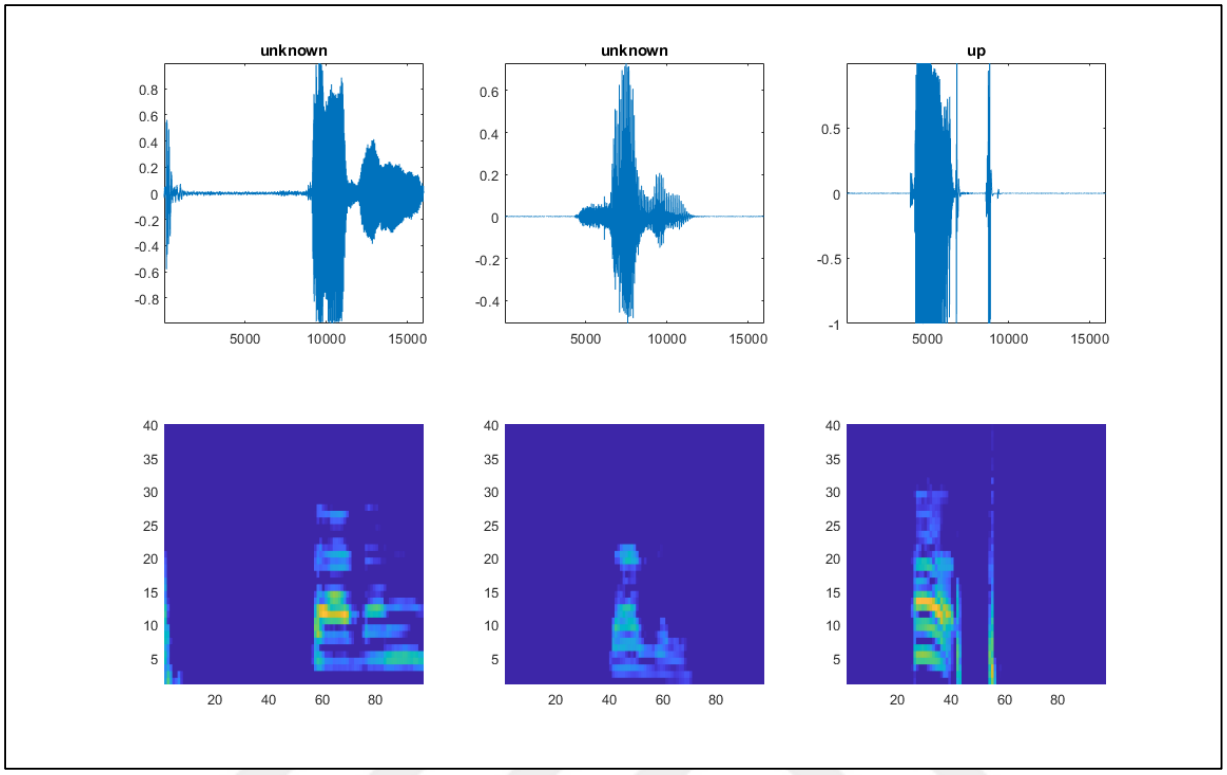

Figure 4.11: Unknown features extracted by the KNN.

The unknown features of the audio samples are not discarded rather classified as unknown to show the user that some errors may occurred in the audio filtering process, the preprocessing of the audio sample is a step that occurs before the feature extraction and classification method as shown in the figure below [41]:

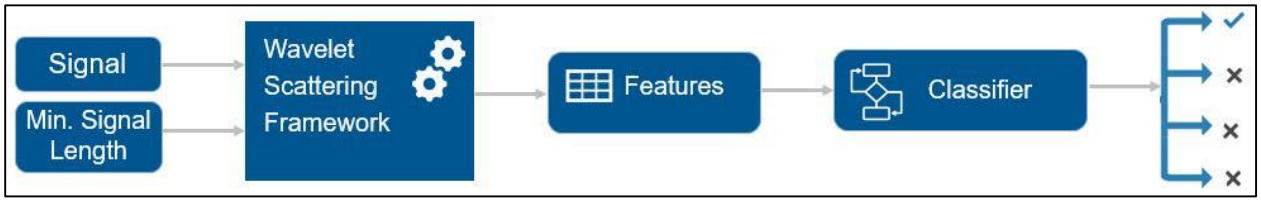

Figure 4.12: Building the dataset of the features extracted by the KNN.

We compare our results with the built-in algorithm functions inside R2018a and see if our result accuracy increased or decreased by the number of training iterations as shown in the figure below:

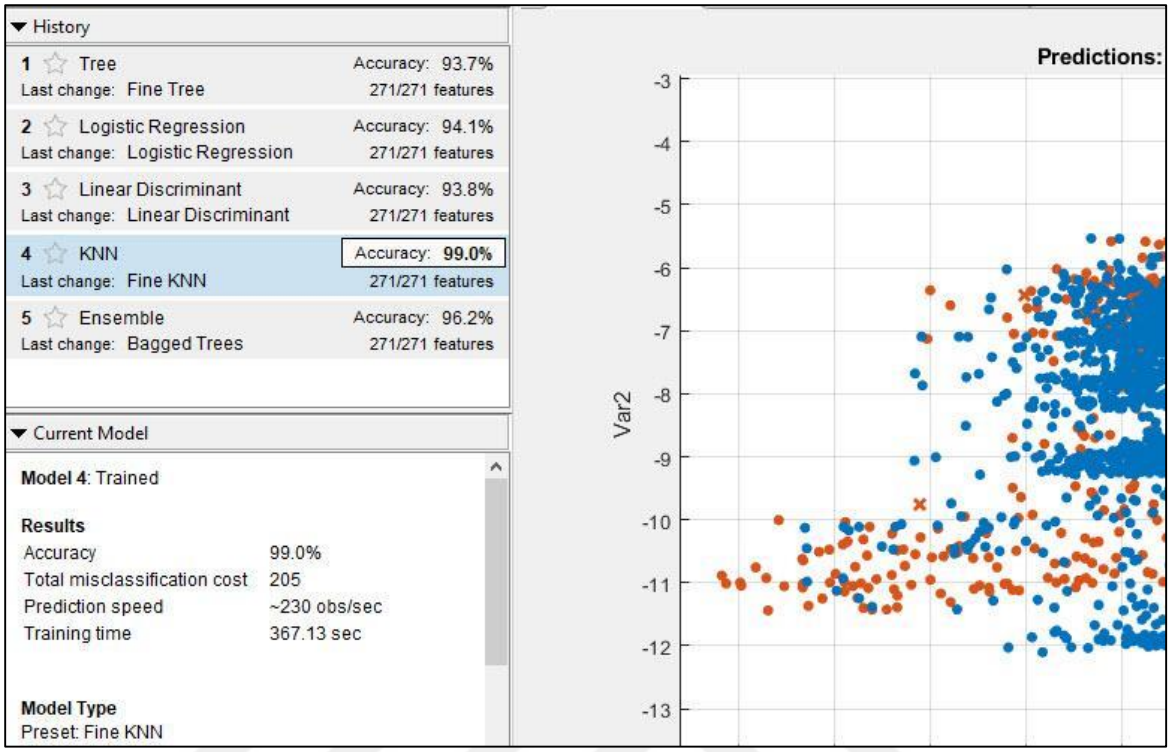

Figure 4.13: The high accuracy of our algorithm.

```
function label = classifyGender(sound_signal, sampling_frequency) %#codegen
%sound signal and sampling frequency as input and produces a classification
%of 'Male' or 'Female'.
% Window length for feature extraction in seconds
win_len = 5;
% Overlap between adjacent windows in percentage
win_overlap = 0;
% Extract features
features = extractFeaturesCodegen(sound_signal, sampling_frequency, win_len, win_overlap);
% Load saved model
Mdl = loadCompactModel('GenderClassificationModel');
% Predict classification for all windows
predicted_labels = predict(Mdl,features);
% Predict male if even one window sounds Male
if find(strcmp(predicted_labels, 'Abnormal'))
```

```
label = 'Male';
else
   label = 'Female';
end
end
```

Figure 4.14: The classification process is explained in the code below.

The sample audio prediction implementation is shown in the figures below:

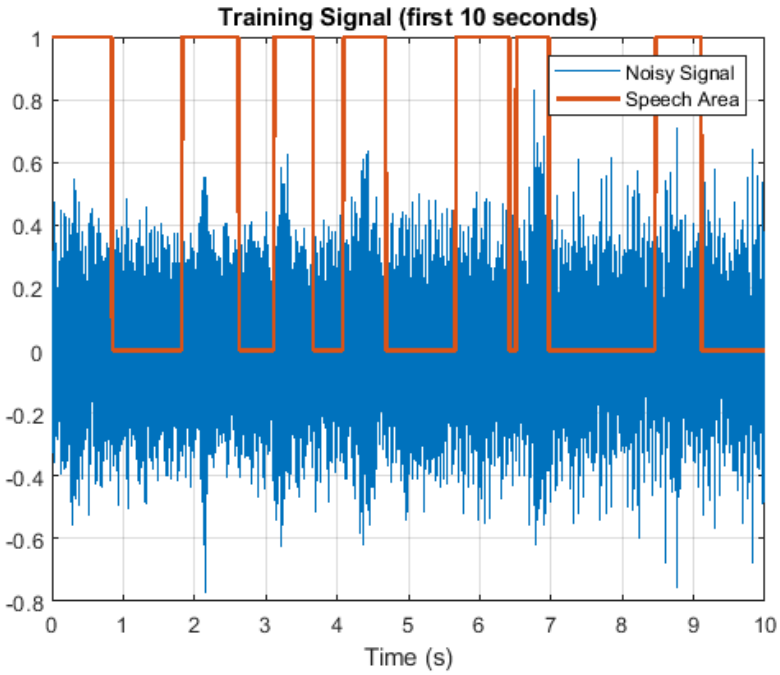

Figure 4.15: Determining the noise and the feature set of the audio sample in the training stage.

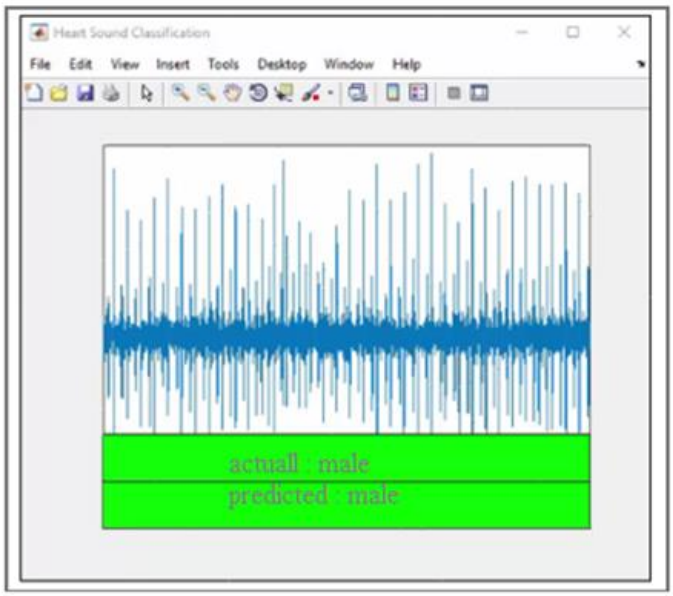

Figure 4.16: Audio sample predication by KNN based on the training.

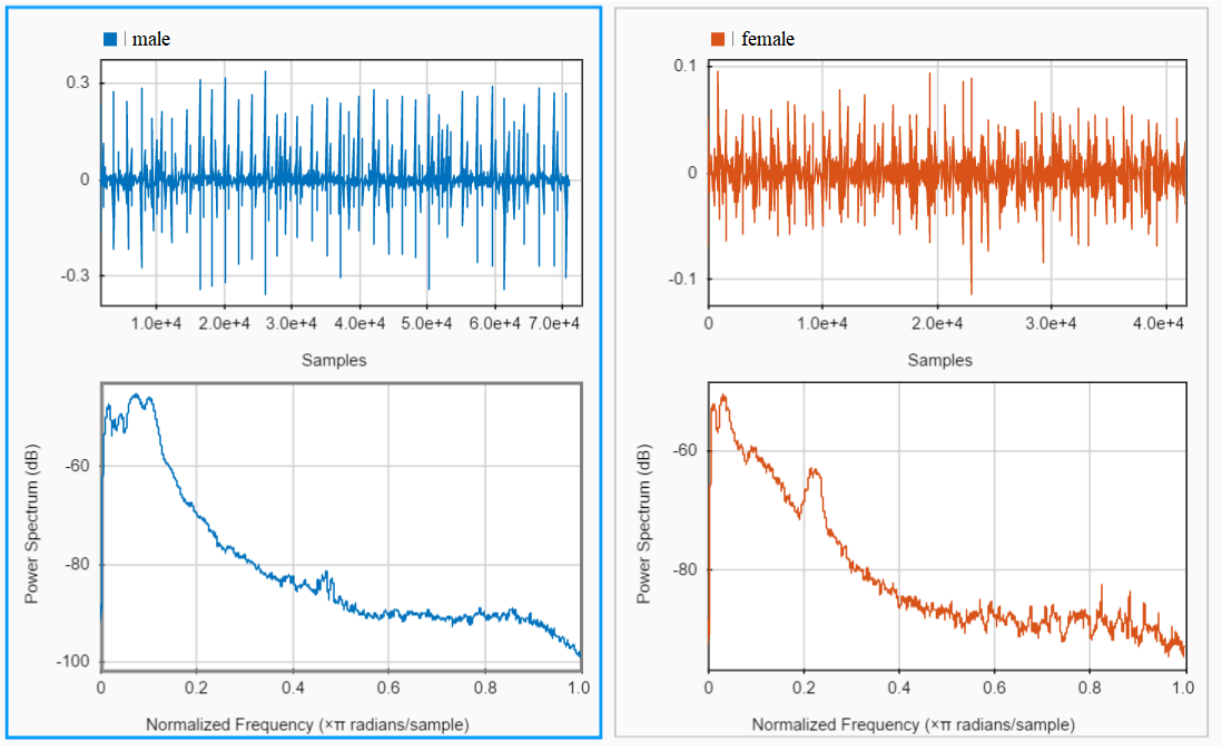

Figure 4.17: Audio sample predication by KNN based on the validation.

5. RESULTS AND DISCUSSION

This research work is aimed at enhancing the input audio quality while simultaneously recognizing the speakers voice and whether the speaker is a male or a female based on the audio features extracted by our KNN model, the model showed high accuracy better than previous methods which are using different variations of ANNs such as SVM or DT as illustrated in the table below:

Table 5.1: Comparing our method to previous methods

Name of the algorithm used	Percentage of accuracy	Number of features
DT	90.4%	271/271
SVM	86.8%	271/271
Logistic regression	85.9%	271/271
Our method (KNN)	94.6%	271/271

furthermore, the below figure Figure5.1 showing the increase of accuracy by the increase of the iteration times needed in our training process while Figure5.2 showing the number of features successfully extracted from our audio samples:

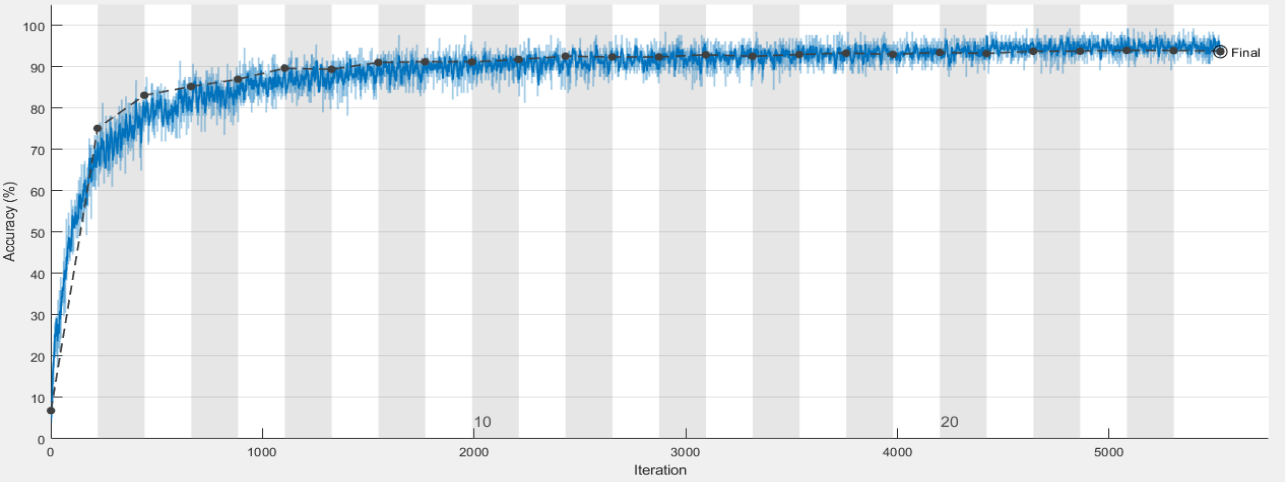

Figure 5.1: Accuracy by iteration number.

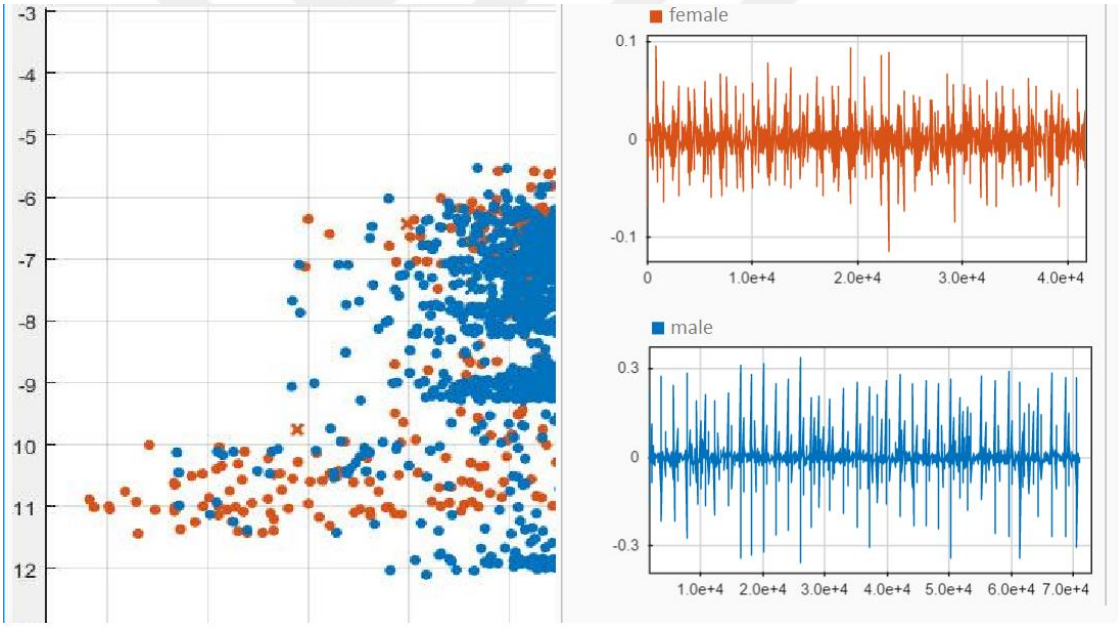

Figure 5.2: The number of samples successfully extracted from our audio samples.

6. CONCLUSION AND FUTURE WORK

Because doing so requires a greater level of processing power than the human system is able to supply, any sort of computing technology will have a tough time establishing the gender of an audio sample. This is because doing so requires a higher level of processing capacity. As a direct result of this, it is strongly recommended that we put into practice strategies that are founded on artificial intelligence (machine learning), so that we can achieve greater degrees of accuracy. This is due to the fact that these tactics will make it possible for us to: To classify the recordings into the appropriate categories, we first extract features from our audio samples via the K-nearest neighbor method, and then we use those features to classify the recordings into the relevant categories. In the next parts, a more in-depth description of this work is presented. Before we start the process of categorization, we use a speech filtering approach to get rid of any background noise that might be present.

Because of this, we are able to improve the precision of our classifications while simultaneously increasing the speed with which we can carry them out. It has been demonstrated that the outcomes of utilizing our method are superior to those of utilizing any other way that has been utilized in this field, and the precision of the outcomes improved as the level of difficulty increased. The results have also been shown to be superior to those of utilizing any other way that has been utilized in this field. By employing a feature scheme that included 271 iterations and 5000 iterations, our method was able to achieve an accuracy of 94.6% during the training iteration. This was made possible by the use of a feature set. In the future, we propose making use of deep learning techniques in order to preprocess the incoming audio and segment the audio based on the regions of interest (ROI).

This will be done in order to improve accuracy (ROI). The presentation of a high-level recognition system that is capable of being implemented in significant areas for the purpose of establishing a learning system of proposal for standards, including a description of the available working tools and the provision of example codes that can be used in subsequent endeavors. The goal of this presentation is to establish a learning system of proposal for standards. This talk will focus on establishing a learning system of proposal for standards so as to fulfill its intended goal.

This discussion will center on the major purpose of developing a learning system that proposes standards as its primary objective.

This project will study the potential of building a speaker recognition system based on a selection of applications that make use of a text-independent modality. The apps will be chosen based on their ability to recognize speakers. This location is currently acting as the research hub for the project. It is impossible for the concepts of machine learning and speech modeling to coexist in any way. The FUZZY LOGIC method has been shown to perform well in text-independent speaker selection applications, which is why the outcomes of the study that has been carried out suggest that it will be the option that is given the highest priority when making a selection. In addition to that, we will present you with a comprehensive review of the resources that are available to you at this time.

When it comes to the many different sources of information at our disposal, we now consider the reader to be trustworthy and impartial. In order to put into practice the strategy that is outlined in this article, we are interested in finding a free program that is compatible with MATLAB and that explains machine learning in a way that can be comprehended by individuals who have not received any type of formal education on the subject. Both of these requirements can be satisfied by the application. Free programming tools that are geared toward speech recognition include Alize and Spear, two examples of which are provided here. However, there is not a lot of potential for customization when it comes to the animals that are used to build the models because they do not provide much leeway. In addition, there are premium, pre-built solutions that can be utilized, such as the Speaker Recognition API that is made available by Microsoft Azure. These solutions can be purchased separately.

These remedies are available for purchase independently. The process of machine learning is totally abstracted by this solution, which offers an API that can be used for both training and testing purposes. This application programming interface is available for usage in either of these scenarios. The Scikit-learn toolbox was chosen as the solution because of the fact that it is open-source software and that the machine learning that it employs can be customized. This was a major factor to take into mind throughout the entire selection process.

It is intended that a description of the numerous tools that have been stated, in addition to a comparative study of the system that has been constructed, will be provided. Figure 10 depicts the system model that we intend to employ for the FUZZY LOGIC algorithm, and the method itself will serve as a starting point for the building of the other machine learning animals. Figure 10 depicts the system model that we intend to work with in order to implement the FUZZY LOGIC algorithm. We are able to utilize a modeling technique that classifies the target according to the environment in which it is located since it is not necessary to construct a model before making use of the fuzzy logic. This is possible because of the nature of the fuzzy logic itself.

A backdrop model is used in order to determine whether or not an input signal was produced by a faker. This can be accomplished by comparing the signal to the model. This model can either be tailored specifically to the needs of each registered speaker, or it can be constructed in a more generalized form. This degree of probability is typically expressed as a percentage of the total. The Universal Background Model is the name of the tactic that is currently enjoying the greatest level of popularity among adopters (UBM). A standardized version of the background is modeled using this approach, and that model is used for all of the speakers. A randomized selection of audios from the data that was utilized for training is applied in order to populate the model. This selection comes from the data.

The steps involved in putting together this model are quite similar to the steps involved in putting together the speaker model. It is going to be selected by random selection which gender will belong to each half of the voices, thus half of the halves are going to have male voices and the other half are going to have female voices. This will be done to ensure that the UBM system cannot be used in any way, shape, or form in order to meet the requirements necessary to ensure that it is inaccessible to people of both sexes. Both of these solutions, once the voice signal has been initially taken in, process it at the front end of the device where it was initially taken in. It is possible that speaker recognition would suffer as a consequence if unused speech signal resources were removed (MFCC), as would be the case with end processing.

This investigation will be built on the foundation of the English Language Speech Database for Speaker Recognition, which is a database that focuses primarily on speaker recognition. This

database will serve as the foundation upon which this investigation will be constructed. Both audio recordings of people reading aloud and a substantial amount of sample audio that can be used for academic study are included in this collection. Within this collection, you will find examples of both types of recordings.

REFERENCES

[1] V. M. Sardar and S. D. Shirbahadurkar, “Speaker identification of whispering speech: an investigation on selected timbrel features and KNN distance measures,” Int. J. Speech Technol., vol. 21, no. 3, pp. 545–553, 2018, doi: 10.1007/s10772-018-9527-4.

[2] K. M. Bellisarioet al., “Contributions of MIR to soundscape ecology. Part 3: Tagging and classifying audio features using a multi-labeling k-nearest neighbor approach,” Ecol. Inform., vol. 51, no. July 2018, pp. 103–111, 2019, doi: 10.1016/j.ecoinf.2019.02.010.

[3] H. F. C. Chuctaya, R. N. M. Mercado, and J. J. G. Gaona, “Isolated Automatic Speech recognition of Quechua numbers using MFCC, DTW and KNN,” Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 10, pp. 24–29, 2018, doi: 10.14569/IJACSA.2018.091003.

[4] V. Mohtasham-zadeh and M. Mosleh, “Audio Steganalysis based on collaboration of fractal dimensions and convolutional neural networks,” Multimed. Tools Appl., vol. 78, no. 9, pp. 11369–11386, 2019, doi: 10.1007/s11042-018-6702-1.

[5] Z. Chen and X. Yuan, “Audio amplitude-level quantification vector for identification of audio postprocessing operation,” Proc. - 2018 Int. Conf. Sens. Networks Signal Process. SNSP 2018, pp. 226–230, 2019, doi: 10.1109/SNSP.2018.00050.

[6] Sheikh, B. Garlapati, S. Chalamala, and S. K. Kopparapu, “A fuzzy approach to mute sensitive information in noisy audio conversations,” Comput. y Sist., vol. 23, no. 3, pp. 675–682, 2019, doi: 10.13053/CyS-23-3-3260.

[7] C. Kaur and R. Kumar, “A fuzzy hierarchy-based pattern matching technique for melody classification,” Soft Comput., vol. 23, no. 16, pp. 7375–7392, 2019, doi: 10.1007/s00500-018-3383-7.

[8] L. Deng et al., “IEEE Signal Processing Magazine - November 2007,” IEEE Signal Process. Mag., vol. 24, no. 99, pp. c1–c1, 2008, doi: 10.1109/msp.2007.4317458.

[9] B. Widrow and M. A. Lehr, “30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation,” Proc. IEEE, vol. 78, no. 9, pp. 1415–1442, 1990, doi: 10.1109/5.58323.

[10] J. Chen, Y. Wang, and D. Wang, “A feature study for classification-based speech separation at low signal-to-noise ratios,” IEEE/ACM Trans. Audio Speech Lang. Process., vol. 22, no. 12, pp. 1993–2002, 2014, doi: 10.1109/TASLP.2014.2359159.

[11] Farzan, S. B. Mashohor, A. Nourmohammadi, and S. Sadra, “Novel voice activity detection based on cepstrum moments,” 2010 2nd Int. Conf. Comput. Autom. Eng. ICCAE 2010, vol. 5, pp. 768–770, 2010, doi: 10.1109/ICCAE.2010.5451357.

[12] P. K. Ghosh, A. Tsiartas, and S. Narayanan, “Robust Voice Activity Detection Using Long-Term,” Language (Baltim)., vol. 19, no. 3, pp. 600–613, 2011.

[13] W. Ying, L. Zhang, and H. Deng, “Sichuan dialect speech recognition with deep LSTM network,” Front. Comput. Sci., vol. 14, no. 2, pp. 378–387, 2020, doi: 10.1007/s11704-018-8030-z.

[14] T. Zhang, Y. Shao, Y. Wu, Y. Geng, and L. Fan, “An overview of speech endpoint detection algorithms,” Appl. Acoust., vol. 160, p. 107133, 2020, doi: 10.1016/j.apacoust.2019.107133.

[15] G. ThimmarajaYadava and H. S. Jayanna, “Enhancements in automatic Kannada speech recognition system by background noise elimination and alternate acoustic modelling,” Int. J. Speech Technol., no. 0123456789, 2020, doi: 10.1007/s10772-020-09671-5.

[16] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, “Front-end factor analysis for speaker verification,” IEEE Trans. Audio, Speech Lang. Process., vol. 19, no. 4, pp. 788–798, 2011, doi: 10.1109/TASL.2010.2064307.

[17] J. K. Bizley and Y. E. Cohen, “The what, where and how of auditory-object perception,” Nat. Rev. Neurosci., vol. 14, no. 10, pp. 693–707, 2013, doi: 10.1038/nrn3565.

[18] M. Huzaifah bin MdShahrin and L. Wyse, “Applying visual domain style transfer and texture synthesis techniques to audio: insights and challenges,” Neural Comput. Appl., vol. 32, no. 4, pp. 1051–1065, 2020, doi: 10.1007/s00521-019-04053-8.

[19] S. Waldekar and G. Saha, “Analysis and classification of acoustic scenes with wavelet transform-based mel-scaled features,” Multimed. Tools Appl., 2020, doi: 10.1007/s11042-019-08279-5.

[20] T. A. Mohammed, S. Alhayali, O. Bayat, and O. N. Uçan, “Feature reduction based on hybrid efficient weighted gene genetic algorithms with artificial neural network for machine learning problems in the big data,” Sci. Program., vol. 2018, 2018, doi: 10.1155/2018/2691759.

[21] M. Noor, “PV PENETRATION EFFECT FOR REGULATING VOLTAGE USING REACTIVE POWER,” vol. 2, no. 1, pp. 17–25, 2018.

[22] T. A. Mohammed, O. Bayat, O. N. Uçan, and S. Alhayali, “Hybrid Efficient Genetic Algorithm for Big Data Feature Selection Problems,” Found. Sci., no. 0123456789, 2019, doi: 10.1007/s10699-019-09588-6.

[23] M. Hamidi, H. Satori, O. Zealouk, and K. Satori, “Amazigh digits through interactive speech recognition system in noisy environment,” Int. J. Speech Technol., no. 2009, 2019, doi: 10.1007/s10772-019-09661-2.

[24] J. F. Bonastre, T. Kinnunen, A. Larcher, and J. Yamagishi, “Introduction to the special issue ‘Speaker and language characterization and recognition: Voice modeling, conversion, synthesis and ethical aspects,’” Comput. Speech Lang., vol. 60, pp. 2019–2021, 2020, doi: 10.1016/j.csl.2019.101021.

[25] R. Jennane, A. Almhdie-Imjabber, R. Hambli, O. N. Ucan, and C. L. Benhamou, “Genetic algorithm and image processing for osteoporosis diagnosis,” 2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC’10, pp. 5597–5600, 2010, doi: 10.1109/IEMBS.2010.5626804.

[26] S. Al-Hayali, O. Ucan, and O. Bayat, “Genetic algorithm for finding shortest paths problem,” ACM Int. Conf. Proceeding Ser., pp. 2–7, 2018, doi: 10.1145/3234698.3234725.

[27] S. Ramoji and S. Ganapathy, “Supervised I-vector modeling for language and accent recognition,” Comput. Speech Lang., vol. 60, p. 101030, 2020, doi: 10.1016/j.csl.2019.101030.

[28] F. Ertam, “An effective gender recognition approach using voice data via deeper LSTM networks,” Appl. Acoust., vol. 156, pp. 351–358, 2019, doi: 10.1016/j.apacoust.2019.07.033.

[29] W. H. Kang and N. S. Kim, “Adversarially learned total variability embedding for speaker recognition with random digit strings,” Sensors (Switzerland), vol. 19, no. 21, 2019, doi: 10.3390/s19214709.

[30] T. Tuncer and S. Dogan, “Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition,” Appl. Acoust., vol. 156, pp. 176–185, 2019, doi: 10.1016/j.apacoust.2019.06.029.

[31] Z. H. Tan, A. kr Sarkar, and N. Dehak, “rVAD: An unsupervised segment-based robust voice activity detection method,” Comput. Speech Lang., vol. 59, pp. 1–21, 2020, doi: 10.1016/j.csl.2019.06.005.

[32] J. W. Johnson, “Adapting Mask-RCNN for Automatic Nucleus Segmentation,” vol. 1, pp. 500–510, 2018, doi: 10.1007/978-3-030-17798-0.

[33] H. Kwon, Y. Kim, H. Yoon, and D. Choi, “Selective Audio Adversarial Example in Evasion Attack on Speech Recognition System,” IEEE Trans. Inf. Forensics Secur., vol. 15, no. c, pp. 526–538, 2020, doi: 10.1109/TIFS.2019.2925452.

[34] T. S. Dinh Le et al., “Ultrasensitive Anti-Interference Voice Recognition by Bio-Inspired Skin-Attachable Self-Cleaning Acoustic Sensors,” ACS Nano, 2019, doi: 10.1021/acsnano.9b06354.

[35] T. Tuncer, S. Dogan, and F. Ertam, “Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network,” Appl. Acoust., vol. 155, pp. 500–506, 2019, doi: 10.1016/j.apacoust.2019.05.023.

[36] N. Saleem and M. I. Khattak, “A review of supervised learning algorithms for single channel speech enhancement,” Int. J. Speech Technol., vol. 22, no. 4, pp. 1051–1075, 2019, doi: 10.1007/s10772-019-09645-2.

[37] N. Lavan, S. Knight, V. Hazan, and C. McGettigan, “The effects of high variability training on voice identity learning,” Cognition, vol. 193, no. May, p. 104026, 2019, doi: 10.1016/j.cognition.2019.104026.

[38] G. Sharma, K. Umapathy, and S. Krishnan, “Trends in audio signal feature extraction methods,” Appl. Acoust., vol. 158, p. 107020, 2020, doi: 10.1016/j.apacoust.2019.107020.

[39] Y. Yu, S. Luo, S. Liu, H. Qiao, Y. Liu, and L. Feng, “Deep attention based music genre classification,” Neurocomputing, vol. 372, no. xxxx, pp. 84–91, 2020, doi: 10.1016/j.neucom.2019.09.054.

[40] S. Abed, M. Alshayeji, and S. Sultan, “Diacritics Effect on Arabic Speech Recognition,” Arab. J. Sci. Eng., vol. 44, no. 11, pp. 9043–9056, 2019, doi: 10.1007/s13369-019-04024-0.

[41] T. A. Mohammed, O. Bayat, O. N. Uçan, and S. Alhayali, “Hybrid Efficient Genetic Algorithm for Big Data Feature Selection Problems,” Found. Sci., no. 0123456789, 2019, doi: 10.1007/s10699-019-09588-6.

[42] M. Hamidi, H. Satori, O. Zealouk, and K. Satori, “Amazigh digits through interactive speech recognition system in noisy environment,” Int. J. Speech Technol., no. 2009, 2019, doi: 10.1007/s10772-019-09661-2.

[43] J. F. Bonastre, T. Kinnunen, A. Larcher, and J. Yamagishi, “Introduction to the special issue ‘Speaker and language characterization and recognition: Voice modeling, conversion, synthesis and ethical aspects,’” Comput. Speech Lang., vol. 60, pp. 2019–2021, 2020, doi: 10.1016/j.csl.2019.101021.

[44] R. Jennane, A. Almhdie-Imjabber, R. Hambli, O. N. Ucan, and C. L. Benhamou, “Genetic algorithm and image processing for osteoporosis diagnosis,” 2010 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC’10, pp. 5597–5600, 2010, doi: 10.1109/IEMBS.2010.5626804.

[45] S. Al-Hayali, O. Ucan, and O. Bayat, “Genetic algorithm for finding shortest paths problem,” ACM Int. Conf. Proceeding Ser., pp. 2–7, 2018, doi: 10.1145/3234698.3234725.

[46] S. Ramoji and S. Ganapathy, “Supervised I-vector modeling for language and accent recognition,” Comput. Speech Lang., vol. 60, p. 101030, 2020, doi: 10.1016/j.csl.2019.101030.

[47] F. Ertam, “An effective gender recognition approach using voice data via deeper LSTM networks,” Appl. Acoust., vol. 156, pp. 351–358, 2019, doi: 10.1016/j.apacoust.2019.07.033.

[48] W. H. Kang and N. S. Kim, “Adversarially learned total variability embedding for speaker recognition with random digit strings,” Sensors (Switzerland), vol. 19, no. 21, 2019, doi: 10.3390/s19214709.

[49] T. Tuncer and S. Dogan, “Novel dynamic center based binary and ternary pattern network using M4 pooling for real world voice recognition,” Appl. Acoust., vol. 156, pp. 176–185, 2019, doi: 10.1016/j.apacoust.2019.06.029.

[50] Z. H. Tan, A. kr Sarkar, and N. Dehak, “rVAD: An unsupervised segment-based robust voice activity detection method,” Comput. Speech Lang., vol. 59, pp. 1–21, 2020, doi: 10.1016/j.csl.2019.06.005.

[51] J. W. Johnson, “Adapting Mask-RCNN for Automatic Nucleus Segmentation,” vol. 1, pp. 500–510, 2018, doi: 10.1007/978-3-030-17798-0.

[52] H. Kwon, Y. Kim, H. Yoon, and D. Choi, “Selective Audio Adversarial Example in Evasion Attack on Speech Recognition System,” IEEE Trans. Inf. Forensics Secur., vol. 15, no. c, pp. 526–538, 2020, doi: 10.1109/TIFS.2019.2925452.

[53] T. S. Dinh Le et al., “Ultrasensitive Anti-Interference Voice Recognition by Bio-Inspired Skin-Attachable Self-Cleaning Acoustic Sensors,” ACS Nano, 2019, doi: 10.1021/acsnano.9b06354.

[54] T. Tuncer, S. Dogan, and F. Ertam, “Automatic voice based disease detection method using one dimensional local binary pattern feature extraction network,” Appl. Acoust., vol. 155, pp. 500–506, 2019, doi: 10.1016/j.apacoust.2019.05.023.

[55] N. Saleem and M. I. Khattak, “A review of supervised learning algorithms for single channel speech enhancement,” Int. J. Speech Technol., vol. 22, no. 4, pp. 1051–1075, 2019, doi: 10.1007/s10772-019-09645-2.

[56] N. Lavan, S. Knight, V. Hazan, and C. McGettigan, “The effects of high variability training on voice identity learning,” Cognition, vol. 193, no. May, p. 104026, 2019, doi: 10.1016/j.cognition.2019.104026.

[57] G. Sharma, K. Umapathy, and S. Krishnan, “Trends in audio signal feature extraction methods,” Appl. Acoust., vol. 158, p. 107020, 2020, doi: 10.1016/j.apacoust.2019.107020.

[58] Y. Yu, S. Luo, S. Liu, H. Qiao, Y. Liu, and L. Feng, “Deep attention based music genre classification,” Neurocomputing, vol. 372, no. xxxx, pp. 84–91, 2020, doi: 10.1016/j.neucom.2019.09.054.

[59] S. Abed, M. Alshayeji, and S. Sultan, “Diacritics Effect on Arabic Speech Recognition,” Arab. J. Sci. Eng., vol. 44, no. 11, pp. 9043–9056, 2019, doi: 10.1007/s13369-019-04024-0.

[60] ABDULLA, W. Auditory based feature vectors for speech recognition systems. p. 231–236, 01 2002.

[61] ADAMOVIC, S.; MILOSAVLJEVIC, M. Information analysis of iris biometrics for the needs of crypto logy key extraction. In: Serbian Journal of Eletrical Engineering.

[62] BOLES, A.; RAD, P. Voice biometrics: Deep learning-based voiceprint authentication system. In: 2017 12th System of Systems Engineering Conference (SoSE). [S.l.: s.n.], 2017.

[63] BONASTRE, J. F.; WILS, F.; MEIGNIER, S. Alize, a free toolkit for speaker recognition. In: Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. [S.l.: s.n.], 2005. v. 1, p. 737–740. ISSN 1520-6149.

[64] CHEN, K.; WANG, L.; CHI, H. Methods of combining multiple classifiers with different features and their applications to text-independent speaker identification. International Journal of Pattern Recognition and Artificial Intelligence, World Scientific, v. 11, n. 03, p. 417–445, 1997.

[65] . CLOUD, G. What is machine learning? 2017

[66] DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological), JSTOR, p. 1–38, 1977.

[67] . FAUNDEZ-ZANUY, M.; MONTE-MORENO, E. State-of-the-art in speaker recognition. IEEE Aerospace and Electronic Systems Magazine, v. 20, n. 5, p. 7–12, 2005.

[68] HASAN, M. R. et al. Speaker identification using mel frequency cepstral coefficients. 12 2004.

[69] HONG, L. Automatic Personal Identification Using Fingerprints. Tese (Doutorado), East Lansing, MI, USA, 1998. AAI9909316.

[70] KHOURY, E.; SHAFEY, L. E.; MARCEL, S. Spear: An open source toolbox for speaker recognition based on Bob. In: IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP). [s.n.], 2014.

[71] KINNUNEN, T.; HAIZHOU, L. An overview of text-independent speaker recognition: From features to supervectors. Speech communication, v. 52, n. 1, p. 12–40, 2010.

[72] LATHI, B. P. Linear Systems and Signals. 2nd. ed. Oxford, UK: Oxford University Press, 2009. ISBN 0195392566, 9780195392562.

[73] LIU, Y. et al. Investigation of frame alignments for gmm-based text-prompted speaker verification. CoRR, abs/1710.10436, 2017.

[74] MARKOV, K. P.; NAKAGAWA, S. Integrating pitch and lpc-residual information with lpc-cepstrum for text-independent speaker recognition. Journal of the Acoustical Society of Japan (E), Acoustical Society of Japan, v. 20, n. 4, p. 281–291, 1999.

[75] MOHAMED, Z.; HAMIDIA, M.; AMROUCHE, A. Investigation of the Use of PLP Coefficients for Speaker Recognition over IP Network. 2013.

[76] MUDA, L.; BEGAM, M.; ELAMVAZUTHI, I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. CoRR, abs/1003.4083, 2010.

[77] O'SHAUGHNESSY, D. Speech communication: human and machine. Universities Press (India) Pvt. Limited, 1987. (Addison-Wesley series in electrical engineering: digital signal processing). ISBN 9788173713743.

[78] PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011.

[79] RAMGIRE, J. B.; JAGDALE, S. M. A survey on speaker recognition with various feature extraction and classification techniques. International Research Journal of Engineering and Technology, v. 3, n. 4, p. 709–712, 2016.

[80] Citadonapágina 10. REYNOLDS, D. A. Speaker identification and verification using gaussian mixture speaker models. Speech Commun., Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 17, n. 1-2, p. 91–108, ago. 1995. ISSN 0167-6393.

[81] RUSSELL, S. J.; NORVIG, P. Artificial Intelligence: A Modern Approach. 2. ed. [S.l.]: Pearson Education, 2003. ISBN 0137903952. Citadonapágina 23. S.D., B. M. D. Automatic speech and speaker recognition by mfcc, hmm and vector quantization. v. 3, p. 93–98, 07 2013.

[82] Citadonapágina 22. SUNNY, S.; S, D. P.; K, P. J. Combined feature extraction techniques and naive bayes classifier for speech recognition. p. 155–163, 07 2013.

[83] Citadonapágina 25. TIWARI, V. Mfcc and its applications in speaker recognition. v. 1, 01 2010. Citadonapágina 22.