UNIVERSITY OF KASDI MERBAH OUARGLA

FACULTY of New Information Technologies and Communication

DEPARTMENT of Electronic and Telecommunication

ACADEMIC MASTER Thesis

Domain : Science and Technology

Faculty : Telecommunication

Specialty : Telecommunication system

Presented by :

❏ Anissa BECHICHE

❏ Mohammed Noureddine DJEZZAR

Topic

Automatic Detection of Parkinson's Disease Using Human Voice and Artificial Intelligence

Submit Date : June 2022

before the jury :

❏ Mr.Azeddine BENLAMOUDI	President	UKM OUARGLA
❏ Ms. Najla BETYAB	Examiner	UKM OUARGLA
❏ Mr.Khaled BENSID	Supervisor	UKM OUARGLA
❏ Mr.Elmoundher HADJAIDJI	Co-Superviser	UKM OUARGLA

University year : 2021/2022

Abstract

The work we present in this Master thesis consists in setting a diagnostic aid system that allows the detection of Parkinson's Disease from human's voice. Since the accurate diagnosis of Parkinson's disease is a challenging task that involves many physical, psychological and neurological examinations. The examinations include investigating a number of signs and symptoms, reviewing the medical history and checking the nervous system conditions of a patient.

The objective of this project is to design an automatic diagnostic support system for the early detection of Parkinson's disease using speech signals. Mainly, it is a question of differentiating between the PD pathologies patients and healthy controls.

The proposed system is based on two main steps: (i) feature extraction and (ii) machine learning classification. The discriminatory features we have chosen encompass of two parameters: (i) Mel-frequency cepstral coefficient (MFCC) and (ii) Excitation Source Parameters. The classification process is based on three machine learning supervised classifiers: (i) K-Nearest Neighbor (KNN) (ii) Support Vector Machine (SVM) and (iii) Decision Tree (DT).

The feature extraction is achieved by a simulation software: MATLAB Mathworks along with the classification process.

A free accessed online (MDVR-KCL) Parkinson's Disease database was used in our experiments. The performance measures used in this study are: accuracy, sensitivity, specificity, F1 Score and the SVM ROC curve. The obtained results were satisfactory.

Keywords : Parkinson's Disease, early detection, Feature Extraction, Machine learning, Classification.

Résumé

Le travail que nous présentons dans ce mémoire de Master consiste à mettre en place un système d'aide au diagnostic qui permet la détection de la maladie de Parkinson à partir de la voix humaine. Vu que le diagnostic précis de la maladie de Parkinson est une tâche difficile qui implique de nombreux examens physiques, psychologiques et neurologiques. Les examens comprennent l'étude d'un certain nombre de signes et de symptômes, l'examen des antécédents médicaux et la vérification de l'état du système nerveux d'un patient.

L'objectif de ce projet est de concevoir un système de soutien diagnostique automatique pour la détection précoce de la maladie de Parkinson à l'aide de signaux vocaux. Il s'agit principalement de différencier les pathologies PD des patients et des témoins sains.

Le système proposé repose sur deux étapes principales: (i) l'extraction des caractéristiques et (ii) la classification de l'apprentissage automatique. Les caractéristiques discriminatoires que nous avons choisies englobent deux paramètres: (i) le coefficient cepstral de Mel-fréquence (MFCC) et (ii) les paramètres de la source d'excitation. Le processus de classification est basé sur trois classificateurs supervisés d'apprentissage automatique: (i) K-Nearest Neighbor (KNN) (ii) Support Vector Machine (SVM) et (iii) Decision Tree (DT).

L'extraction des caractéristiques est réalisée par un logiciel de simulation: MATLAB Mathworks avec le processus de classification.

Une base de données sur la maladie de Parkinson accessible gratuitement en ligne (MDVR-KCL) a été utilisée dans nos expériences. Les mesures de performance utilisées dans cette étude sont: l'accuracy , la sensibilité, la spécificité, le score F1 et la courbe ROC pour SVM. Les résultats obtenus sont satisfaisants.

Mots-clés : Maladie de Parkinson, détection précoce, extraction de caractéristiques, apprentissage automatique, classification.

ملخص

يتمثل العمل الذي نقدمه في هذه الاطروحة في وضع نظام دعم تشخيصي يسمح بتصنيف الامراض الصوتية، اي مرض الباركنسون على وجه التحديد باستخدام الصوت. نظرا لان التشخيص الدقيق لمرض باركنسون يمثل مهمة صعبة تتضمن العديد من الفحوصات الجسدية و النفسية و العصبية. و تشمل الفحوصات التحقيق في عدد من العلامات و الاعراض و مراجعة التاريخ الطبي و فحص حالات الجهاز العصبي للمريض.

الهدف من هذا المشروع هو تصميم نظام دعم تشخيصي للكشف المبكر عن مرض الباركنسون من الصوت. بشكل أساسي إنها مسألة التمييز بين مرضى الباركنسون و الاشخاص الغير مرضى. يستند النظام المقترح على خطوتين رئيسيتين: (أ) استخراج الميزات و (ب) تصنيف التعلم الآلي. تشمل السمات التمييزية التي اخترناها على ثلاثة معايير: (أ) معامل MFCC (ب) قياس درجة الصوت (ج) قياس ارتعاش وميض الصوت. تستند عملية التصنيف على ثلاثة مصنفات خاضعة للاشراف: (أ) ك-اقرب جار (ب) شعاع الدعم الالي (ج) شجرة القرار.

يتم تحقيق استخراج الميزات و عملية التصنيف من خلال برنامج محاكاة Mathworks. Matlab تم استخدام قاعدة بيانات مجانية عبر الانترنت لمرض الباركنسون في تجاربنا.

مقاييس الاداء المستخدمة في هذه الدراسة هي الدقة و الحساسية و النوعية و منحني .ROC النتائج التي تم الحصول عليها مرضية.

الكلمات المفتاحية: مرض الباركنسون، الكشف المبكر، استخراج الميزات، التعلم الالي، التصنيف

ACKNOWLEDGMENTS

First of all, We would like to thank **Allah** the Most Gracious for giving us the patience, courage and willpower to accomplish this work.

We would like to express our deep gratitude to our superviser **Dr. Khaled BENSID** and Co-superviser **Mr. EL-Moundher HADJAIDJI** for having accepted to direct and to supervise this work. It is an honor for us to work with them. They put at our disposal their knowledge and their rich experience. We also thank them for their advice, their documentation and their availability.

We would also like to thank Mr.Azedine BENLAMOUDI, for the honor of chairing the jury of this thesis.

Our thanks also go to Ms.Najla BETYEB for having agreed to judge this work.

We would also like to thank Mr.Mohamed Fouad HASSANI for his support and his precious advice.

We would like to thank Mr. kaddour MOKRANI for his help and documentation.

We also thank the administration of the Scientific Research Center, and in particular the managers of the Electrical Engineering Department for having welcomed us and for having made available to us the favorable conditions for the realization of this work.

We would like to thank our parents who have helped and supported us, we express our gratitude to them, may Allah the Almighty reward you with the prize of his vast paradise.

Finally, a big THANK YOU to all those who, from near or far, participated in the realization of this thesis.

DEDICATIONS

I dedicate this modest work:
To the one who was the bravest of men, and who supported me so much morally and financially and who always wished to see me where i am today: ***My father***
May God have mercy on him and placed him into his havens
To the woman who has always been beside me with her prayers throughout my life, to her for being the strongest woman and stood by us and kept us standing: ***My mother***
To my sisters Manel Meryem and Fatma Zohra who were always there to support me and to all my family and to Imad Eddine KRAMI for his suuport.
To my colleagues who became more like a family: Manel CHAIBI, Nassira BARKA, Mohammed Larbi SENOUSSI, Nour El-houda AOUNI and Ibrahim El-khalil Ghouar who helped me during all my university years, and for all the unforgettable moments I have spent with you. I wish you all success in your life.
To my dear friend Amina BELOUATAR who helped me during the realisation of this project.
To my partner in this thesis Nouraddine DJEZZAR and his family.
To all my colleagues in the Telecommunications speciality and all my friends.

- Anissa

DEDICATIONS

To the one who always showed me support and encouragement throughout my whole life , to my dear ***mother****.*

To the one who showed me his good will and power, ***my father***

To my sisters and brothers, I wish them good luck and happiness

To all my family one by one

To my colleagues and friends : Med Larbi SENOUSSI, Hassen BOUGUEFFA ATEMEN, Anes MEDJOUEL, Med Laid ABI MILOUD, Ibrahim Elkhalil GHOUAR, Taki AMRI, Amina BELOUATAR.

To my partner in this thesis Anissa BECHICHE and her family.

- Nouraddine

Contents

List of tables

List of figures

List Of Abbreviation

AI:	Artificial Intelligence
DCT:	Discrete cosine transform
DT:	Decision Tree
FFT:	Fast Fourier Transform
FN:	False negative
FP:	False positive
FVD:	Functional Voice Disorders
HC:	Healthy Controls
KNN:	K-Nearest Neighbors
MDVR-KCL:	Mobile Device Voice Recordings at King's College London
MFCC:	Mel Frequency Cepstral Coefficients
ML:	Machine Learning
PD:	Parkinson's Disease
RBF:	Radial Basis Function
ROC:	Receiver Operating Characteristic
SVM:	Support Vector Machine
TN:	True Negative
TP:	True Positive

General Introduction

Motivation

Voice is a primitive natural tool for communication exercised by humans. Voice communication is an integral part of our personal and professional life. However, there are always barriers to effective voice communication [2]. Speech impairments, due to Parkinson's Disease (PD), is one of them. Parkinson's disease (PD) is a progressive neurodegenerative disorder of the nervous system that affects our body movements including speech [3]. PD severely affects patients' quality of life, social functions and family relationships, and places heavy economic burdens at individual and society levels [4].

At the advanced stage, PD can be easily and accurately diagnosed, but effective treatment is a challenging task. Also, if treatment is started in advanced stages, it might have less effectiveness in controlling PD progression. This situation necessitates the early and accurate diagnosis of PD, thus helps the patients in maintaining a good quality of life.

To date, no single blood or laboratory test exists that is helpful in the identification of PD and its progression. Diagnosis of PD is commonly based on medical observations and assessment of clinical signs. However, traditional diagnostic approaches may suffer from subjectivity as they rely on the evaluation of movements that are sometimes subtle to human eyes and therefore difficult to classify, leading to possible misclassification [4] Also, in some cases, it would be difficult to distinguish between various neurological disorders because they share the same etiology. Approximately 75% of clinical diagnosis of PD is confirmed to be idiopathic [3].

Therefor, new methods based on machine learning using the speech signals are proposed to improve diagnosis accuracy rate and to help the doctor to make right decisions.

Contribution

We contribute to design a diagnostic aid system that allows the early detection of the Parkinson's Disease(PD).Specifically, we are interested in the distinction between PD subject patients and healthy control voices.

The approach we propose is based on two essential levels: (i) feature extraction

and (ii) classification. The discriminatory features we have chosen encompass two types of parameters: (i) Mel Frequency Cepstral Coefficients (MFCC) and (ii) Pitch, Jitter and Shimmer parameters.

The classification scheme is based on three Machine Learning supervised classification methods: (i) K-Nearest Neighbhor (KNN) (ii) Support Vector Machines (SVM) and (iii) Decision Tree (DT).

Both the feature extraction and the classification process were achieved by the simulation software: MATLAB Mathworks.

The Parkinson's Disease speech Database was used in the experimental part to measure: Accuracy, Sensitivity, Specificity, F1 score and ROC curve.

Thesis Organisation

The thesis is organized as follows:

- **First Chapter:** Introduces the voice production anatomy followed by the voice characteristics. Then proceeded by the voice disorders, their causes and their different types where we present the Parkinson's disease.
- **Second Chapter:** is dedicated to the proposed method.
- **Third chapter:** is dedicated to the experimental results and discussion.
- **Conclusion and perspectives.**

Chapter 1

Voice Disorders Generalities

I. Introduction

Like fingerprints, each person's voice is unique. The voice is the sound made by air passing from our lungs through our larynx, or voice box. In the larynx are our vocal cords, two bands of muscle that vibrate to make a sound. For most of us, our voices play a big part in who we are,what we do, and how we communicate. It is one of the most important professional tools for many functions. As a result, the intensive or abusive use of the voice can cause voice disorders and affect health and social life in general as well as professional activities[5] [6]. Voice disorders may also result from other diseases.

In this chapter we present in a first part the anatomy of the human voice production and its characteristics. then the definition of voice disorders and their general causes. The second part will be about their different types then the disease we focus on in our study is introduced which is Parkinson's Disease. We end this chapter with a conclusion.

II. Voice Production Anatomy

To define voice disorders, the voice and the anatomy of the human voice should be defined first:

1. The Voice

The voice is an auditory perceptual term. That means the audible sound produced by the larynx [7]. The production of the sound is often referred to as "phonation," and the act of producing the words sound is called "speech" or "articulation" [1].

The human vocal system is mainly composed of lungs, larynx and vocal tract, as shown in Figure 1 [2].

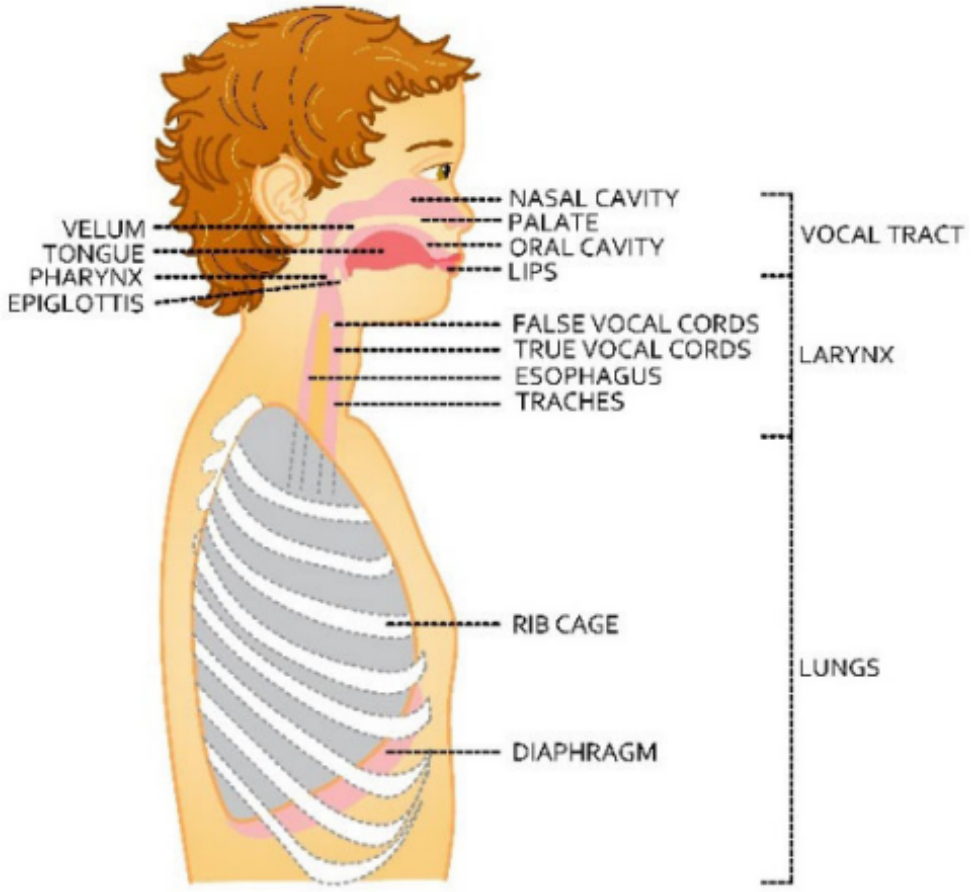


Figure 1: Components of the human vocal system.
[2]

1.1 The Lungs:

are the source of energy for our vocal system. During voice generation, we inhale air by expanding the rib cage that surrounds the lungs and expel air by lowering the diaphragm at the base of the lungs. We maintain a steady flow of air by controlling the muscles around the rib cage based on the length of the sentence or phrase. This action causes air to flow into vocal trachea to the epiglottis [2].

1.2 Larynx:

The larynx has a complex anatomical structure, as shown in the figure 2. It has the shape of a hollow tube, located between the pharynx and the trachea, it is formed by a cartilaginous osteo skeleton that contains [6]:

- At the top, the hyoid bone;
- Thick and resistant cartilage: thyroid cartilage, cricoid cartilage, arytenoid cartilage

and epiglottic cartilage that forms the lid of the larynx.

These various elements are joined together with other neighbouring organs by ligaments and muscles. The larynx cavity is lined by the laryngeal mucosa which is raised by ligaments. This cavity has:

- An upper orifice communicating with the pharynx and closed by the epiglottis during swallowing;
- An upper stage between the epiglottis and the upper vocal chords;
- A medium stage between the upper and lower vocal cords;
- A lower stage where its opening is in the trachea.

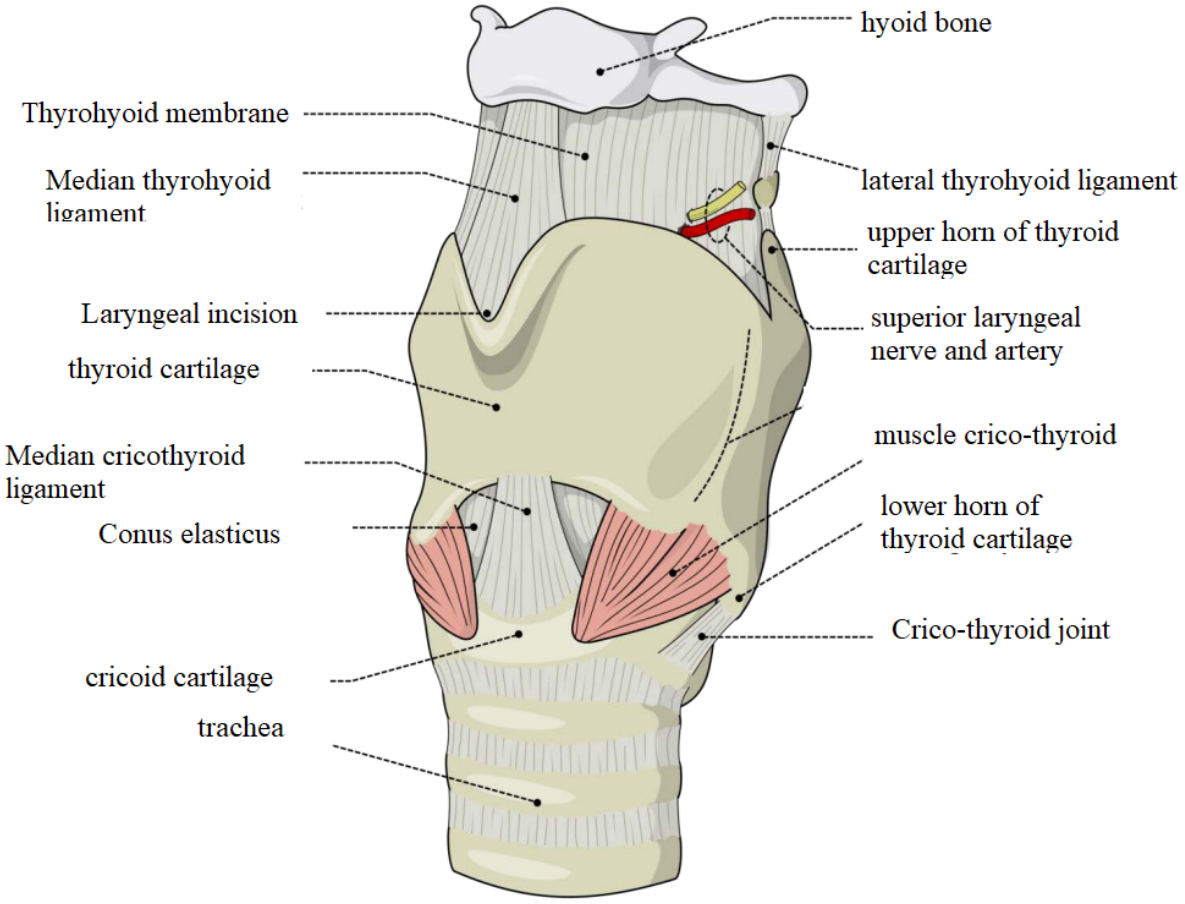


Figure 2: Anatomy of the larynx [6]

1.3 Vocal Tract:

The vocal tract contains the nasal cavity, oral cavity and pharynx. Its length is between 13 cm and 17 cm for women and between 17 cm and 24 cm for men. The location of the articulators (teeth, tongue, jaw, lips, palate veil) allows a variation from 0 to more than 20 cm2 of the area of the vocal tract's right section. The typical length of the nasal tract is 12 cm. The degree of coupling between the nasal cavity and the oral cavity is determined according to the size of the opening of the palate veil. The spectral characteristics of the sounds radiated by the lips can be influenced by the coupling of the nasal cavity. When the palate veil is oriented downwards, acoustically, the nasal cavity is coupled to the oral cavity and the timbre produced by this coupling is nasal. In order to achieve a non-nasal sound, the palate veil must be positioned so as to close the entrance of the nasal cavity by decoupling it from the oral cavity [6].

2. Physiology of Voice Production System:

The lungs provide the energy needed for the production of the voice, which is the air, with a well-determined rate [6]. During normal conditions, the vocal cords are in a state called "breathing." In the breathing state, the vocal cord mass relaxes and the glottis opens. Air from the lungs flows freely through the glottis without the vocal cords vibrating. During vocalization, the vocal cords can be in two states namely unvoiced and voiced. In the unvoiced condition, the vocal folds approach each other and create turbulence on their own. While in the voiced condition (that is during the production of vowels), the vocal folds come closer, become more tensed, and partially close the glottis. The partially closed glottis and increased vocal fold tension cause oscillation of the folds. Air flow from the lungs is interrupted by the vocal cords and produces quasi-periodic pressure waves. The impulses of this pressure wave is called a pitch and the frequency of the pressure is called the pitch frequency. The quality of the larynx adjusts the length and tension of the vocal folds to "fine-tune" pitch and tone. The articulators (i.e the tongue, palate, cheeks, and lips) pronounce and filter the sounds coming from the larynx. Vocal cords and articulators produce highly complex sounds [2].

The figure 3 shows a block diagram of the voice production [8].

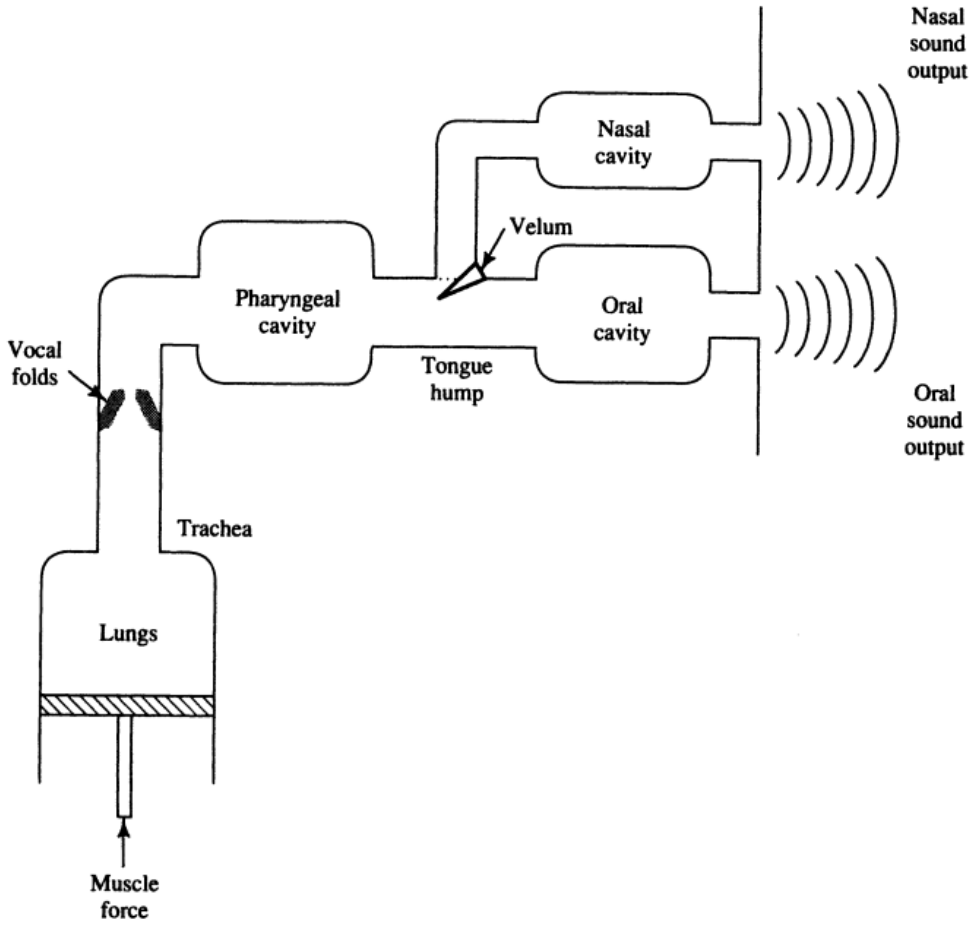


Figure 3: A block diagram of human speech production [8]

3. Voice Characteristics

the voice is characterized by three parameters:

3.1 Pitch:

Pitch or the fundamental frequency F0 [9]. One of the three essential qualities of a sound (with timbre and intensity), depending on the frequency of the vibration, and perceived subjectively as an impression of acuity or gravity more or less. A sound is perceived to be all the more acute as its frequency is high. The frequency is measured in Hertz [10].

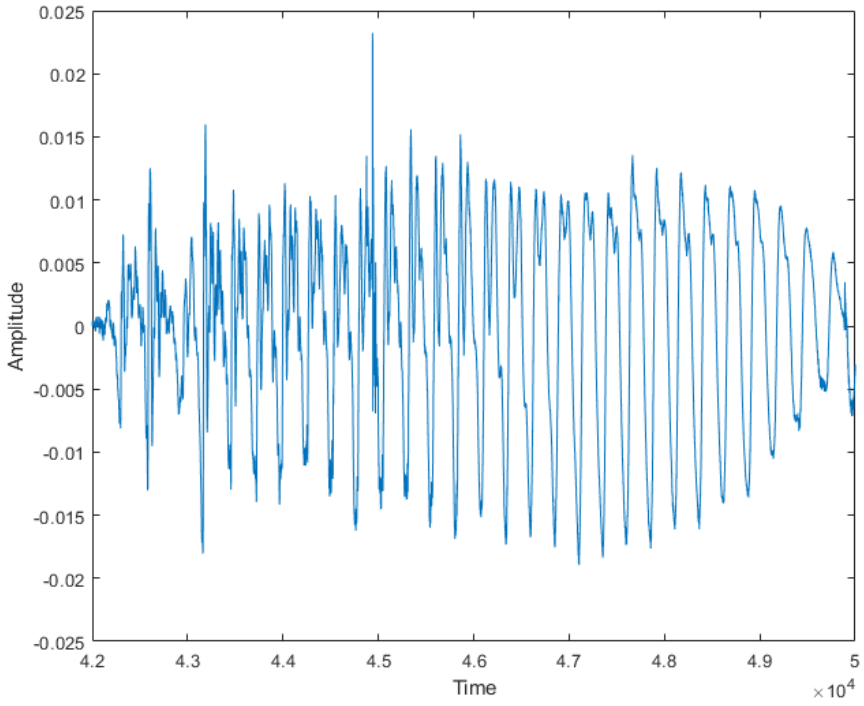


Figure 4: Sound Pitch.

3.2 Intensity:

Sound intensity is the subjective perception of sound pressure. Every person has a different perception of loudness, which means that loudness cannot be measured objectively. The sound intensity of an acoustic sound or signal is therefore relative. Physically measurable, on the other hand, is the sound pressure, which is converted into sound level and further expressed in decibels (dB). The second measurable quantity is hertz, which measures the number of air pressure fluctuations per second

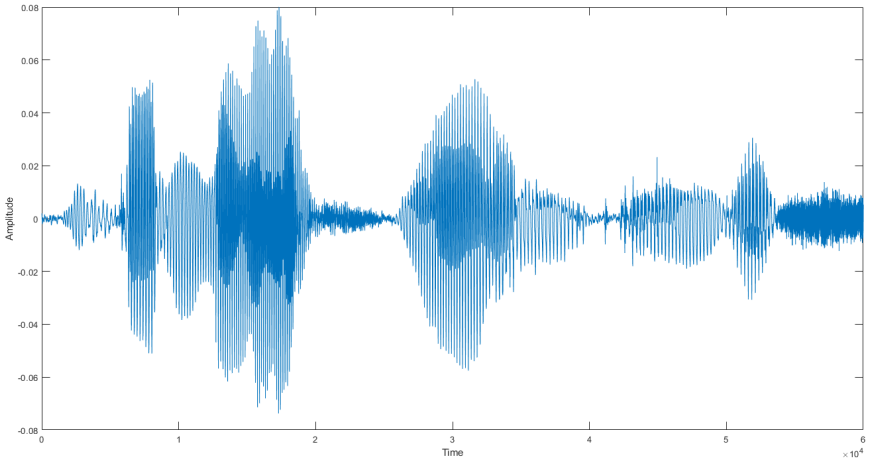


Figure 5: Sound Intensity.

3.3 Timbre:

The specific quality of a sound, independent of pitch and intensity, resulting essentially from the combination of harmonics that accompany the fundamental note played. If two sounds are produced at the same intensity and at the same height but by two different instruments, a violin and a clarinet, it is the timbre of these two sounds that will allow our ear to recognize the instruments. It is also by the timbre that the vowels differ from each other [10].

When there's deviations in vocal quality or in the system mentioned above, disorders appear in the Voice production. So, what are voice disorders?

III. Voice Disorders

1. Definition:

Voice disorders represents an area of communication dysfunction [11]. Which affects a large population of people. They range from complete absence of the voice (aphonia) to varying degrees of vocal impairment (dysphonia) [12]. They are common and can develop at any age suddenly or gradually over time, the voice can be described as hoarse, rough, raspy, strained, weak, breathy, or gravely.

A voice disorder exists when quality, pitch, loudness, or flexibility differs from the

voices of others of similar age, sex, and cultural group. Consequently, a wide range of voices accepted as normal differs based on occupational needs and cultural expectations, and many dysphonic individuals are not even aware they have a voice problem [7].

Individuals with voice disorders range from a simple case of laryngitis that usually resolves spontaneously to more physical or organic conditions such as laryngeal malignancy [11].

2. Causes of voice Disorders:

Clinical conditions of speech disorders can be classified to the phonation mechanism, they are classified as follows:

1. Glottal closure disorder.
2. Stiffness of the affected vocal cords.
3. Vocal cord asymmetry.
4. Respiratory/resonant cavity dysfunction.
5. Psychological factors.

Of these, 1 to 3 abnormalities in the shape and mobility of the larynx are the main causes of the disease. When examining a patient complaining of a voice disorder, first of all they are asked about their main complaints, medical history, degree and quality of hoarseness, past history, occupation, and daily living habits, and the social background associated with phonation and possible causes of the voice disorder will be assessed first [1].

Below, are the main diseases that cause the hoarseness and voice disorders. In disorders (1) or (2) below, the patient must be directed to a specialist physician, if the hoarseness does not improve, urgently in the case of disorder (3) and swiftly in the case of disorders (4) or (5) [1].

1. **Vocal Cord Polyps:** These are the most common organic disorders of the larynx which cause voice disorders. A common site for polyps is from the front third to the center of the membranous portion of the vocal cord, and in many cases the sides are asymmetrical, regardless if the polyp occurs on one side or both sides of the

vocal cords (Fig. 6A). Polyps are thought to be caused by submucosal bleeding of the vocal cords, and contributing factors include voice misuse and smoking. With regard to hoarseness, in many cases the patient is found to have rough hoarseness or breathy hoarseness [1].

2. **Acute Corditis Vocalis:** The larynx becomes inflamed due to a common cold, etc., causing the voice to become whispery and hoarse, in some cases the patient's voice becomes aphonic. There is diffuse reddening and swelling of both vocal cords, and histologically this is regarded as being caused by inflammatory cell invasion of the superficial lamina propria, edema, and vasodilatation. Due to the rapid swelling of the vocal cords, the mucosa is extended excessively, reducing its mobility; mucosal waves are diminished, with asymmetrical vocal cord vibrations [1].

3. **Acute Epiglottitis:** Even when symptoms of acute inflammation such as pharyngeal pain or fever causing a muffled voice are observed, examination of the oral cavity may overlook the inflammation finding, making it extremely important that the larynx also be examined. This is a potentially lethal disorder that causes airway stenosis and must not be overlooked (Fig. 6B). This disorder is frequently observed in men in the prime of their lives; it can easily cause medical disputes as the disease progresses quickly with the patient dyspnea, resulting in hypoxic encephalopathy or death [1].

4. **Recurrent Nerve Paralysis:** After leaving the ambiguous nucleus of the medulla oblongata as the vagus nerve, the recurrent nerve, which controls the movement of the vocal cords, travels a long distance before reaching the larynx, and may become paralyzed due to damage sustained along its various parts. In many cases this is caused by neck or chest disorders, and it may also be caused by tumors such as thyroid cancer, esophageal cancer, and lung cancer; lymph node metastasis, aortic aneurysm, and complications of surgeries performed to cure these; infectious diseases such as viruses; and tracheal intubation. It may also be caused by idiopathic and intracranial/skull base diseases. Symptoms for unilateral recurrent nerve paralysis include hoarseness and misswallowing, while bilateral paralysis causes airway stenosis with respiratory difficulties. The position in which the vocal cords are fixed determines the status of glottal closure and the degree of hoarseness also changes

depending on vocal cord position. (Fig. 6C). Although strictly speaking these differ, terms such as laryngeal paralysis, vocal cord paralysis, and vocal cord fixation are also used [1].

5. **Laryngeal cancer:** Laryngeal cancer is divided into supraglottic, glottic, and subglottic types. The glottic type is the most common, and when the cancer invades the mucosal lamina propria and muscle layer of the vocal cords, it impairs vocal cord vibration, causing hoarseness (Fig. 6D). In the supraglottic and subglottic types, patients are slow to become aware of the hoarseness and detection of the cancer is slower than that for the glottic type. When the cancer invades the cricoarytenoid articulation, the vocal cord becomes fixed, developing incomplete glottal closure, causing air leak and breathy hoarseness during vocalization [1].

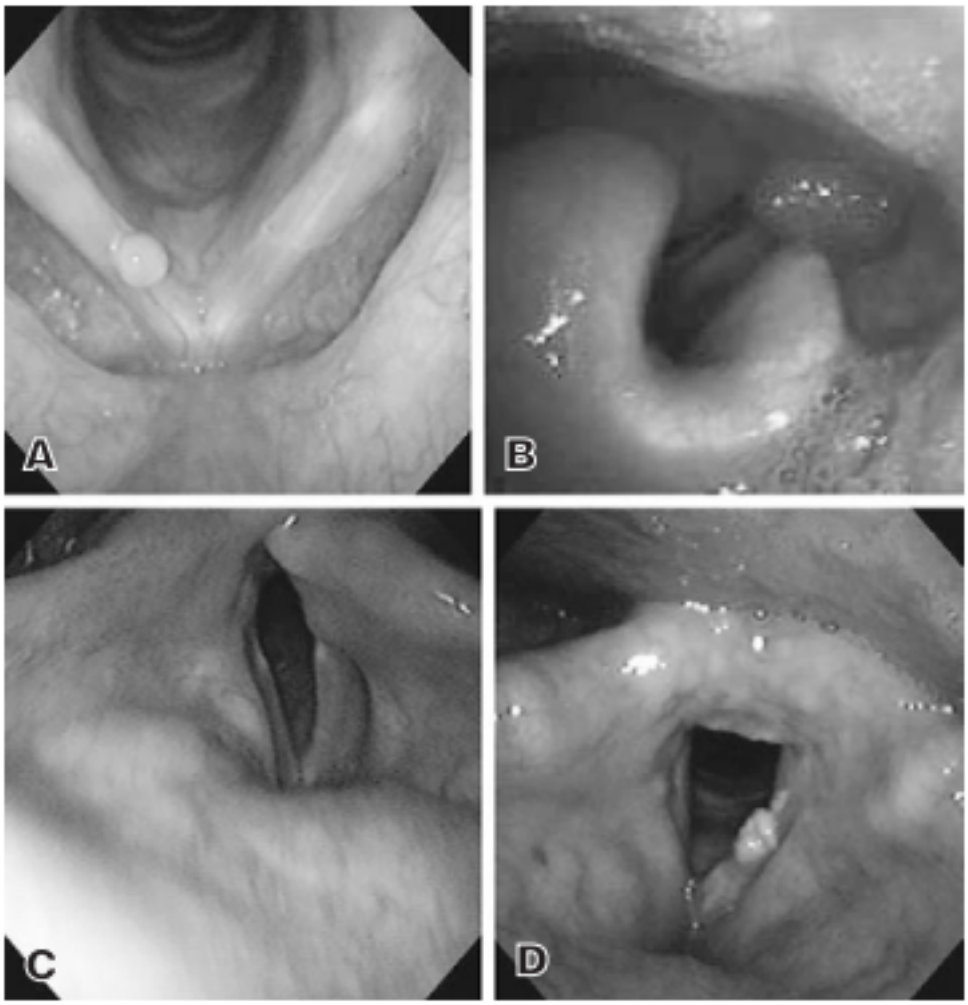


Figure 6: **Main conditions which cause voice disorder**
A: Right vocal cord polyp
B: Acute epiglottitis
C: Left recurrent nerve paralysis
D: Laryngeal cancer of the left glottis [1].

3. Different Types of Voice Disorders:

The larynx may be affected by several types of pathology. These pathologies are classified into two categories: organic and non-organic (functional) [6]:

3.1 Non-Organic or Functional Voice Disorders(FVD):

(FVD) are caused by insufficient or improper use of the larynx and diaphragm without any identifiable physical structural abnormality or neurological dysfunction. These disorders often have a noticeable adverse impact on social and occupational function. Although there is a vast amount of literature relating to the topic, there is no consensus

regarding foundational concepts or methodology of evaluation of voice disorders, as very few authors have rigorously investigated the most effective vocal assessment or video laryngoscopic examination techniques required to produce a consistent and definitive diagnosis of FVD. The most common FVDs are vocal fatigue, muscle tension dysphonia or aphonia, diplophonia, and ventricular phonation. Vocal fatigue is caused by overuse of the voice and resultant tiring of the laryngeal musculature. Muscle tension dysphonia or aphonia is caused by hypertonicity of the laryngeal musculature, which in turn limits the vocal folds' ability to abduct and adduct with coordination and rapidity. Diplophonia is the phenomenon in which two separate fundamental frequencies are being produced during phonation, which may result from waves of different phases passing through the vocal fold mucosal surface or from different oscillatory frequencies occurring in the left and right vocal folds. Lastly, ventricular phonation occurs when the false vocal folds, also known as the ventricular folds, become the primary vibratory surfaces of the larynx due to stiffness of the true vocal folds or maladaptive voicing habits. All of these phenomena tend to present with hoarseness [13].

3.2 Organic Voice Disorders:

In organic voice Disorders, the faulty voice is caused by structural or physical disease of the larynx itself, or by systemic illness that alters laryngeal structure. Voice problems associated with nervous system involvement include spasmodic dysphonia, myasthenia gravis and vocal fold paralysis [11]. Organic voice disorders can be classified as follows:

- **3.2.1 Structural Organic Voice Disorders:** Structural organic voice disorders are a result of physical changes in the voice mechanism such as edema, vocal nodules, due to alteration in vocal fold tissues and age-related structural changes in the larynx. [13].
- **3.2.2 Neurogenic Organic Voice Disorders:** Neurodegenerative diseases are defined as hereditary and sporadic conditions which are characterized by dysfunction of the progressive nervous system [3]. It is usually part of a more general dysarthric disorder which includes motor problems of the respiratory and the supralaryngeal (above the larynx) muscles during speaking. The voice problems encountered in these patients may either reflect direct involvement of the laryngeal motor system or a compensatory reaction to respiratory or articulatory dysfunc-

tions [14].This disorder includes vocal tremors, spasmodic dysphonia, paralysis of the vocal folds, etc [13].

Among the Neurogenic organic voice disorders mentioned above, our study focuses on Parkinson's Disease for being the second most common neurodegenerative disease. In the next subsection the mentioned PD is presented.

4. Parkinson's Disease (PD):

Dr. James Parkinson in 1817 discovered this disease and described the condition which he called it the 'Shaking Palsy' [3].

It is a disorder of the nervous system that affects muscle control. PD is slowly progressive and worsening over time. Out of many neurodegenerative diseases like Alzheimer's disease, Brain Cancer, Degenerative Nerve Diseases and Epilepsy, Parkinson's Disease is considered to be the second most common neurodegenerative disease after the Alzheimer's [3, 15].

Parkinson's disease is one of the most common neurodegenerative disorders of the elderly population with an average age of onset of 60 years of age. However, PD affects 17.4 per 100,000 of individuals between the ages of 50-59 and 93.1 people per 100,000 of individuals between the ages of 70-79 [16], and a rare form of the disease affects teenagers [15].

4.1 PD Causes:

The main cause of the disease is the death of dopamine-generating cells in the parts of the midbrain known as substantia nigra [17].

Yet, after decades of exhaustive study, the causes of PD are still unknown. Many of the researchers think that a combination of genetic and environmental factors; such as exposure to the environmental toxin, head injury, rural living, drinking water, manganese and exposure to pesticides, are responsible for PD. These factors may vary from person to person [3, 15].

4.2 PD Symptoms:

There are many symptoms that appear in PD patients, including posture and balance deficiencies, dysphonia, and slowed movement, etc. Among these symptoms is vocal dysfunction, which results in vocal instability, loudness, and damaged vocal quality [18].

The psychological symptoms include depression, Dementia and emotional problems [17].

Neurological symptoms surpass humans Direct observation and require advanced human nervous system and brain diagnostic techniques. Figure 7 shows a healthy and Parkinson's brain [17].

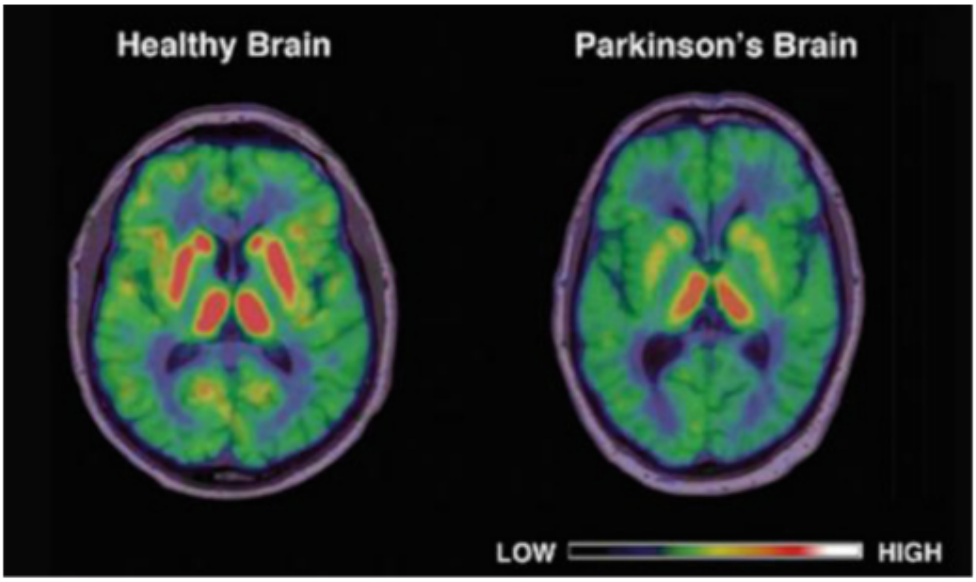


Figure 7: Parkinson's disease symptoms in a human brain.

There are some specific symptoms that an individual experience and each PD patient experience these symptoms differently. [3].

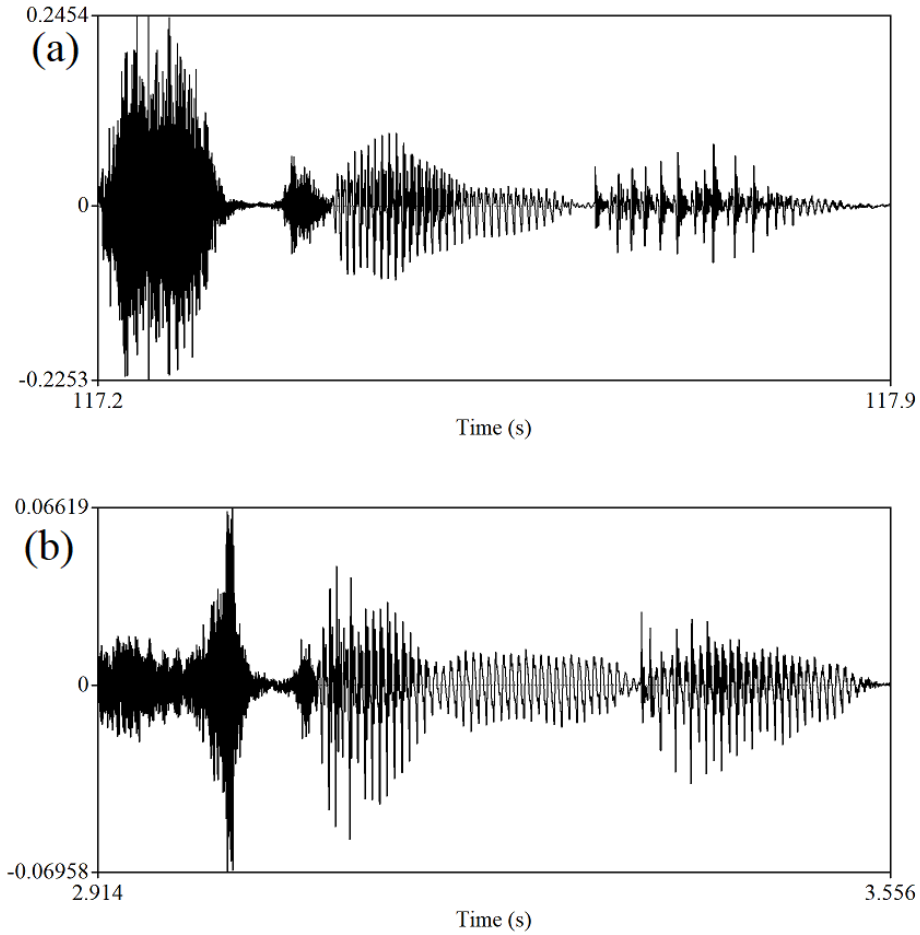


Figure 8: Two selected examples of speech signals: (a) healthy, (b) subject with PD. The horizontal axis is time in seconds, the vertical axis is signal amplitude.

Description of different stages of PD is reported as follows [3]:

- **Mildest stage (Stage 1)** In this stage, the PD patients have least interference with routine tasks. Tremors and other symptoms are restricted to one side of the body.
- **Moderate stage (Stage 2):** In this stage, symptoms like stiffness, resting tremors and trembling can be sensed on both sides of the body. Also facial expressions of PD patients may get changed.
- **Mid-stage (Stage 3):** During this stage, major changes like balance loss, decreased flexes in addition with stage 2 symptoms will be observed in PD patients. Occupational therapy combined with medication may help in decreasing the symptoms.

- **Progressive stage (Stage 4):** The condition of PD patient will get worse in this stage and it becomes difficult for the patient to move without some assistive device like a walker.
- **Advanced stage (Stage 5):** this stage is the most advanced and debilitating stage of PD. Stiffness in legs may cause freezing when standing. Patients are frequently unable to stand without falling. They may experience hallucinations and occasional delusions.

Apart from the symptoms mentioned above, PD patient may suffer from different complications.

4.3 PD Complications:

4.3.1 Motor complications:

Motor complications in PD consist of motor fluctuations and dyskinesia, as a result of the pharmacological treatment. These can be either excessive hyperkinesias (e.g. freezing, rigidity, increasing off times, dysphasia, dysarthria, and respiratory compromise) or excessive hyperkinesia (e.g. choreiform and dystonic dyskinesia). Motor complications decrease the patient's quality of life in many ways as they affect emotional health, decrease mobility, decrease independence for activities of daily living, and cause social stigma [16].

1. **Dyskinesia:** Among the major complications of managing PD is the presence of dyskinesia. Dyskinesias consist of abnormal movements (e.g. movement of the head, neck and limbs) that are debilitating, physically tiring, and embarrassing. Several reports show that the rate of this problem vary greatly, ranging from 19 to 80% in PD patients.
2. **Dystonia:** Dystonia as a neurological movement disorder occurs in untreated PD patients. The treatment of dystonia varies based upon clinical presentation. Early morning dystonia, a symptom of overnight wearing off, may respond to nocturnal longacting dopaminergic agents. In contrast, peak-dose dystonia that occurs during the day may respond to reduced dose of dopaminergic medications, given more frequently in smaller doses. [19]

3. **Freezing:** PD patients can experience freezing of mobility through any movement, but it is most prominent and difficult when this freezing involves gait. Freezing is especially frequent when initiating gait (start hesitation) and when passing through tight spaces such as doorways. Freezing can be the result of either too much or too little dopaminergic effect. "Off freezing" may react to changes in the aforementioned medications, while "on freezing" is often associated with end- stage disease and is typically difficult to handle. [19]

4. **Hypokinesia/akinesia:** Among individuals with PD, gait disorders are one of the most common factors that affect independence and quality of life. Debilitating hypokinesia as a type of motor fluctuation is one of the most common signs of end-stage PD. In these individuals, episodes of hypokinesia can occur many times per day and these events are typically associated with either a failure to respond or the "off" phase of dopaminergic treatment. [19]

5. **Rest tremors:** Tremor, or shaking, often in a hand, arm, or leg, occurs when you're awake and sitting or standing still (resting tremor), and it gets better when you move that body part. Tremor is often the first symptom that people with Parkinson's disease or their family members notice. At first the tremor may appear in just one arm or leg or only on one side of the body. The tremor also may affect the chin, lips, and tongue. As the disease progresses, the tremor may spread to both sides of the body. But in some cases the tremor remains on just one side. Emotional and physical stress tends to make the tremor more noticeable. Sleep, complete relaxation, and intentional movement or action usually reduce or stop the tremor. [19]

6. **Rigidity:** Rigidity causes stiffness and inflexibility of the limbs, neck and trunk. Muscles normally stretch when they move, and then relax when they are at rest. In Parkinson's rigidity, the muscle tone of an affected limb is always stiff and does not relax, sometimes contributing to a decreased range of motion. People with PD most commonly experience tightness of the neck, shoulder and leg. A person with rigidity and bradykinesia tends to not swing his or her arms when walking. Rigidity can be uncomfortable or even painful.[19]

7. **Bradykinesia:** Bradykinesia means "slow movement" A defining feature of Parkin-

son's, bradykinesia also describes a general reduction of spontaneous movement, which can give the appearance of abnormal stillness and a decrease in facial expressivity. Bradykinesia causes difficulty with repetitive movements, such as finger tapping. Due to bradykinesia, a person with Parkinson's may have difficulty performing everyday functions, such as buttoning a shirt, cutting food or brushing his or her teeth. People who experience bradykinesia may walk with short, shuffling steps. The reduction in movement and the limited range of movement caused by bradykinesia can affect a person's speech, which may become quieter and less distinct as Parkinson's progresses. [19]

4.3.2 Non-Motor complications:

Although they receive less attention than the motor complications, non-motor complications are also common, resulting both directly from PD and indirectly from the dopamine fluctuations [20]. here are some of the main non-motor complications:

1. **Psychosis:** In most cases, psychosis develops late in PD, often due to underlying dementia and as a result of anti PD medication use. Around 40% of PD patients develop dementia in the late stages of the disease, and in these, psychosis is common. Patients suffering from PD dementia and psychosis are more likely to be placed in a nursing home and are also at an increased mortality risk. Inter current medical conditions like constipation, dehydration, electrolyte abnormalities, pneumonia or urinary tract infection, may be hidden causes to psychosis and should be investigated and treated appropriately before starting antipsychotic treatment. [21]

2. **Depression:** Depression and anxiety occur in up to 40% of all PD patients, possibly higher among end-stage patients with increasing motor complications. Anxiety as well as depression also tends to be more frequent during off periods and often getting better when the dopaminergic treatment is optimized thus having less pronounced and less frequent off-periods. They can both be present throughout the disorder. Depression in the PD population has the most severe negative impact on reported quality of life. In late stage PD, it is essential to closely evaluate patients, with the assistance of family and other caregivers, to identify depression.[21]

3. **Sleep disturbance:** Sleep disturbances affect up to 60% of PD patients. Sleep disorders can result from motor-related aspects as well as from non-motor-related aspects and medications. Anticholinergic medications and dopaminergic drugs can worsen the condition and must be considered to be decreased or stopped. Low-dose nightly clonazepam or melatonin might be useful. Nocturnal dystonia and cramping may cause sleep fragmentation and could be treated by a dopamine agonist at bedtime. [21]

4. **Apathy:** Apathy occurs in 16.5 to 42% of PD patients. It is common and a main feature in end-stage PD and causes problems for clinical management and care. It seems that apathy is associated with cognitive impairment and depression and these should first be properly treated. Based upon current clinical evidence, it is not clear whether apathy is improved by Levadopa treatment. [21]

4.4 PD Epidemiology:

It is estimated around 6.3 million people suffer from PD worldwide. The World Health Organization gives an estimated crude prevalence of 160 per 100,000, and an estimated incidence of 16-19 per 100,000. The incidence of PD varies across the globe. However, this distribution depends on geographical or ethnic factor. It is known that the PD is more prevalent in North America and Europe than in Asia and West Africa. The Figure 9 below shows the statistics of epidemiology of PD worldwide. [16]

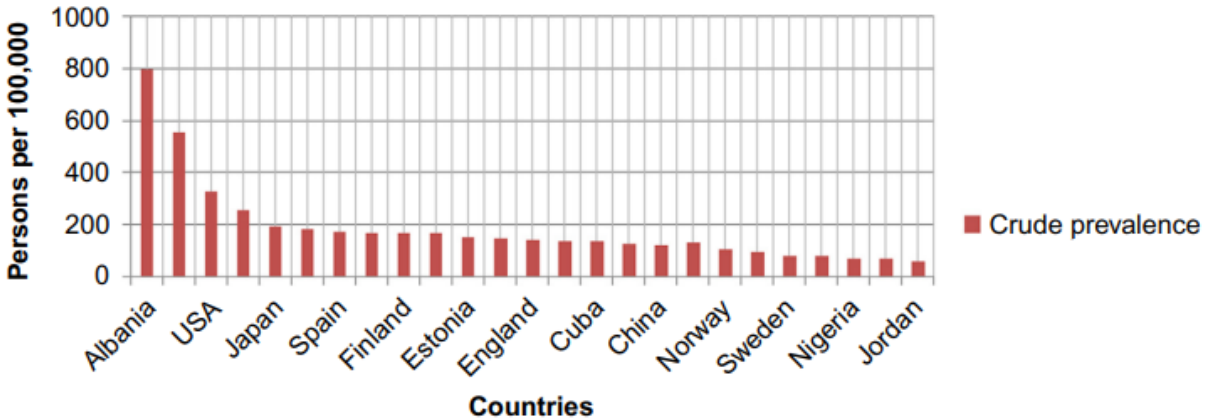


Figure 9: PD epidemiology worldwide.

4.5 PD Current Clinical Diagnosis:

The diagnosis of PD is still largely a clinical one, with the exception of gene testing in a reduced number of cases. PD is a disease combining clinically defined parkinsonism with specific pathological findings, namely, dopaminergic neuron loss in the region of substantia nigra pars compacta, as well as the presence of intraneuronal Lewy bodies.[22]

From a practical perspective, the first step for the diagnosis of PD is careful history taking. Thorough questioning of the patient and family should be performed, trying to define which symptoms emerged and their sequence, as well as perceived anatomical involvement. Inquiry about the presence of pre-motor symptoms including sleep-related REM sleep behavior, loss of smell, and constipationcan be helpful if present. Drug intake history, both past and present, especially concerning drugs able to cause parkinsonian symptoms, is paramount. Likewise, possible exposure to environmental toxics should also be searched for. Past and present medical disorders should be systematically recorded. Family history is also an important stage, and should include neurological disorders in other family members, as well as inquiry about ethnic ancestry as monogenic forms of PD are more prevalent in some (e.g., Ashkenazi Jewish and North African Arabs who have a higher frequency of Mutations in leucine-rich repeat kinase 2 (LRRK2) genetic PD).[22]

In typical circumstances, a restricted number of investigations are necessary to establish the clinical diagnosis of PD. Brain structural imaging, either by computed tomography(CT)or magnetic resonance imaging (MRI) should always be performed; where available the latter is preferred, because some positive findings occasionally reveal other diagnostic entities. CT scan should be used whenever calcium deposits are being searched for. Dopamine functional imaging might be considered to confirm that degenerative parkinsonism is the cause of symptoms. Positron emission tomography (PET) with fluorodopa is one of the technologies available, but the costs and limited accessibility make it difficult to use. In this regard, dopamine transporter (DAT) imaging with single-photon emission CT (DAT-SPECT) is a very useful approach, because it is sensitive for the detection of presynaptic dopaminergic neuron degeneration in the striatum. None of these methods is able to distinguish PD from other causes of degenerative Parkinsonism, but presynaptic dopamine imaging is normal in essential tremor, dystonic tremor,

drug-induced, psychogenic tremor, and psychogenic Parkinsonism. In appropriate circumstances, genetic testing might be considered; whenever pathogenic mutation is found, a definitive diagnosis of PD is achieved in vivo [22].

Among many other symptoms, for diagnosis, speech disorders are manifested in PD patients at the prodromal stages as early as five years before the occurrence of gross motor dysfunctions. Speech disorders caused by PD can be characterized by symptoms such as reduced vocal tract volume and tongue flexibility, inappropriate pauses, impairments in voice quality, and reduction in pitch range and voice intensity [23].

4.6 PD Alternative Diagnosis:

One of the novel approaches to diagnose PD is using the Artificial Intelligence (AI) as a non-invasive technique.

Current advances in the field of AI include lead to the coming out of Decision Support Systems (DSS) and expert systems for medical applications. Furthermore, in the last a small number of decade's computational tools have been designed. to advance the abilities and experiences of doctors and medical specialists in making decision about their patients. With no disbelief the estimation of data taken from patients and decision of expert are still the mainly significant factors in diagnosis. on the other hand, expert systems and different (AI) techniques for classification have the promising of being good helpful tools for the expert. Classification systems can assist in increasing accurateness and reliability of diagnoses and minimizing likely errors, in addition to making the diagnoses further time efficient [24].

In the speech of people with PD, there are some subtle abnormalities which may not be perceived by listeners but can be beneficial to detect PD by employing acoustical analysis. Recently, many techniques have been developed to analyze the speech in order to classify people suffering from PD and healthy people. In these researches, the speech of people has been recorded and some specific features have been extracted to discriminate people with PD from healthy subjects by employing different techniques [25].

IV. Conclusion:

The study of this chapter, allowed us to learn the human sound production anatomy and then we learned their characteristics. In addition, the definition of the voice disorders and their general causes were presented. It also allowed us to discover what are the different types of the voice disorders where we had an idea of Parkinson's disease.

In the next chapter, we will present the machine learning classification methods of the voice pathologies and more precisely Parkinson's Disease diagnosis aid system that we suggest.

Chapter 2
Detection of PD Methods Using Speech Signals

I. Introduction

As the Known neuro-degenerative disorder, Parkinson's disease (PD) affects the nerve cells of the human brain. Early detection and treatment can help to relieve the symptoms of PD [18]. Nowadays, however, the diverse medical techniques that exist for direct examination and diagnostics of pathologies used by medical specialists possess a number of disadvantages. Amongst them; the difficulty to access Human voice pathways for visual examination during the phoning process, which makes it more difficult to identify a pathology. In addition, these diagnostic tools can cause great discomfort to patients and distort the actual signal, which can also lead to an incorrect diagnosis. Hence, researches on the detection of voice disorders and the classification of pathological voices now extend over four decades, as a non-invasive diagnostic technique and a method that does not have any particular disadvantages compared to the methods mentioned above [26, 27].

In this chapter, we present a diagnostic aid method that we have proposed. In a first part we are interested in the feature extraction techniques in general, then we focus on the methods used in our study which are MFCC and the Excitation Source Parameters (Pitch, Jitter and Shimmer). In the second part, Machine Learning Classifiers used in our study are presented. We conclude this chapter with a conclusion.

II. Feature Extraction

Feature extraction is the first step in any voice disorder detection system. [28]. The feature extraction process aims to extract a compact, efficient set of parameters that represent the acoustic properties observed from input speech signal, for subsequent utilization by acoustic modeling.[29]

The most common voice features are formants, wavelets, Linear Predictive Coding (LPC), perceptual linear prediction (PLP), relative spectral transform (RASTA), The Linear Predictive Cepstral Coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC), Pitch, Jitter and Shimmer. These voice features are briefly described in the following subsections [2]

1. Formants:

The formant frequency or simply formant analysis is an important voice feature investigated by the researchers. The formant frequencies are the resonance frequencies of the vocal tract and they change with different vocal tract configurations. The formant usually refers to the entire spectral contribution of a resonance. The peaks of the spectrum for vocal tract response correspond approximately to its formants. The formants can be plotted with frequency as shown in Figure 10. The formant plot shows distinct peaks at certain frequencies. It also shows that the peaks are separated by some frequency band and are of decreasing magnitudes. The formant plot shows that the pathological voice exhibits very distinct formants compared to normal voice. For example, the first three peaks are closely located and are almost having the same magnitude for pathological voice. On the other hand, normal voice shows peaks that are located at almost equal distances and the peak values decrease in magnitude. Although the first formant of a normal voice carries a power similar to that of pathological voice, the other formants carry low power compared to those of pathological voice.[2]

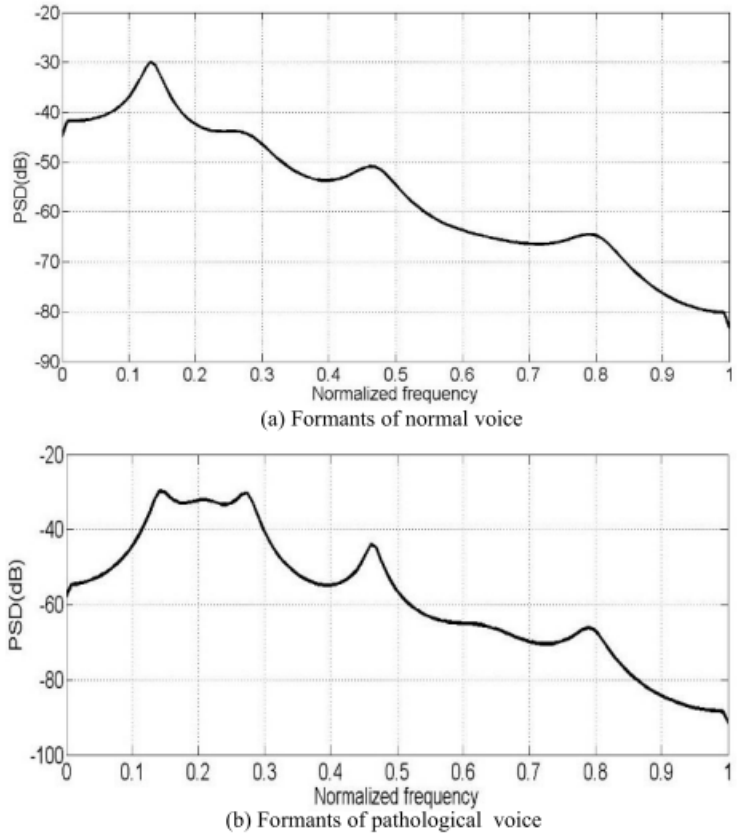


Figure 10: The comparison of the formants

2. Wavelet Analysis:

The wavelet transform is another important tool used in voice disability detection. Its main advantage over the Fourier transform is that wavelet can provide accurate information about the fast fluctuations of signals in the time domain.
It maps a time function into two functions namely scale, a and translation, b. The continuous wavelet transform (CWT) of a signal $f(t)$ is defined as:

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\varphi_{ab}(t)d\varphi \tag{1}$$

A scaled version of the function

$$\varphi(t)$$

with a scale factor of a is defined as $\varphi(\frac{t}{a})$. The wavelet is a useful tool to investigate the discontinuity in pathological voice. [2]

3. Linear Predictive Coding (LPC):

Primarily, LPC has been introduced to compress digital signals for efficient transmission and storage. However, now LPC has become one of the most powerful speech analysis techniques and it has gained popularity as a formant estimator. The LPC method is based on modeling the vocal tract as a linear all-pole infinite impulse response (IIR) filter, which is defined by

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}} \tag{2}$$

where p is the number of poles, G is the filter gain, and $a_p(k)$ are the coefficients. Given a short-time segment of a speech signal sampled, a speech encoder determines proper excitation function, pitch period for voiced speech, gain parameter G, and the coefficients $a_p(k)$. The LPC is computed based on the least mean-squared error approach. In this approach, the speech signal is approximated as a linear combination of its previous samples. [2]

4. The Perceptual Linear Prediction (PLP):

PLP, introduced by Hermansky [30], models the human speech based on the concept of the psychophysics of hearing. The main function of PLP is to discard irrelevant information

contained in the speech. PLP has spectral characteristics that are transformed to match the human auditory system unlike LPC. Hence, PLP is more adapted to human hearin compared to LPC. The other main difference between PLP and LPC is that both use two different types of transfer functions. For example, the LPC model assumes an all-pole transfer function of the vocal tract with a specified number of resonances within the analysis band. On the other hand, the transfer function of PLP is also an all-pole model; however, it approximates the power distribution of equal magnitude at all frequencies of the analysis band. The steps of PLP computation are shown in Figure 11. [2]

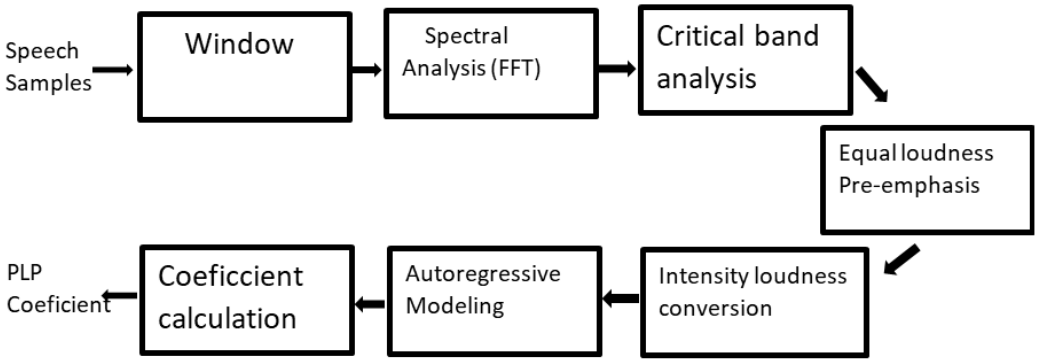


Figure 11: The computation of PLP

5. Relative Spectral Transfrom (RASTA):

This speech enhancement technique was originally introduced with the aim of subsiding the unwanted and additive noise in Automatic speech recognition. RASTA not only alleviates the impact of noise in speech signal but it also enhances the quality of speech with background noise. It also involves linear filtering of trajectory of power spectrum in the case of noisy speech. Thus, RASTA is a modulation frequency band pass filtering technique that can be used either in cepstral domain or log spectral domain. RASTA filter band passes each feature coefficient. The figure 12 block diagram gives the process of RASTA speech processing technique. [31]

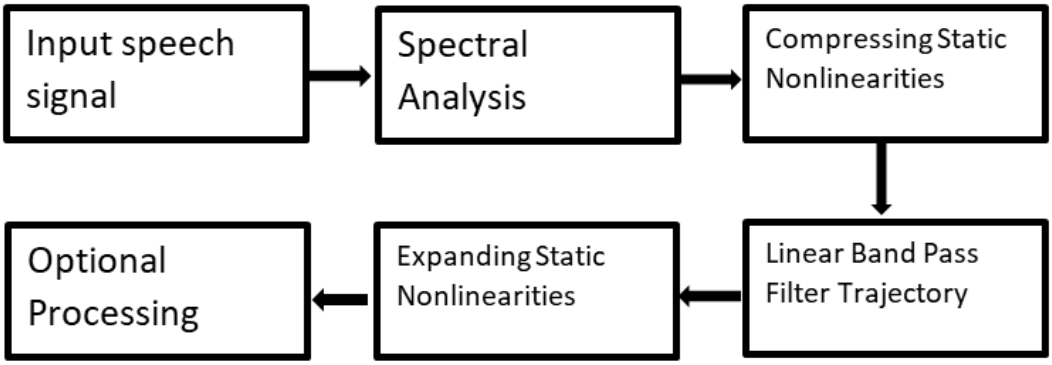


Figure 12: Block diagram of RASTA

6. The Linear Predictive Cepstral Coefficients (LPCC):

The Linear Predictive Cepstral Coefficients (LPCC) technique is an improvised version of LPC to overcome the channel effects by executing Cepstral Mean Subtraction (CMS). The Cepstral is a sequence of power spectral density extracted from periodogram, is utilized in pitch tracking. The cepstrum is obtained by performing Inverse Fourier Transform for the obtained power spectrum of voice signals and may classify as real cepstrum, complex cepstrum, phase cepstrum and power cepstrum where the power cepstrum is employed in speech analysis. The figure 13 shows the process of the LPCC. [32]

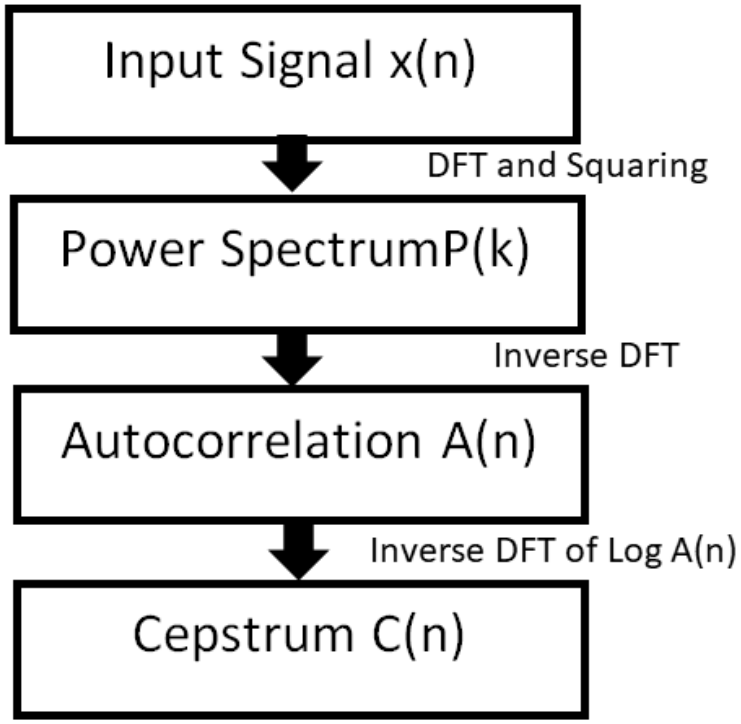


Figure 13: LPCC computation process

For our study we focus on the Mel-frequency cepstral coefficient (MFCC) and the excitation source parameter in the feature extraction process.

6. Mel-frequency cepstral coefficient MFCC:

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behavior [33]. MFCCs are very popular features used for representing speech signals. This feature seeks to emulate human perception of decoding the speech signal to extract the information from the input [34]. MFCC is based on known variation of the human ear's critical bandwidth with frequency. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. The overall process of the MFCC is shown in Figure 14 [33].

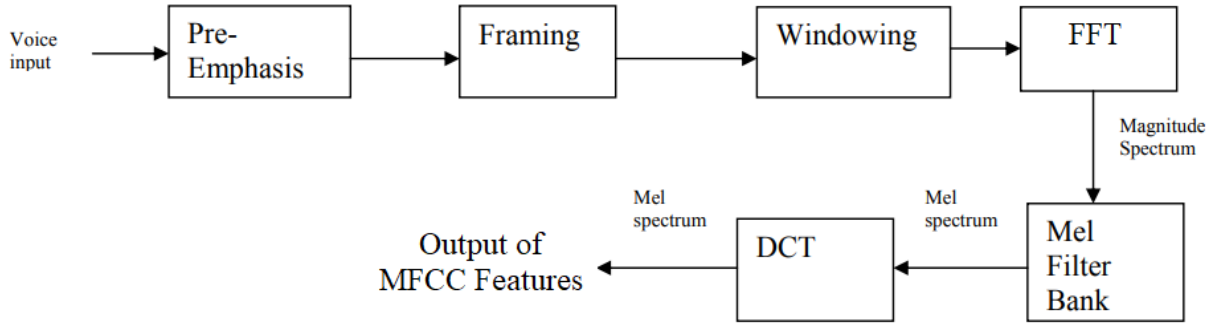


Figure 14: MFCC Block Diagram

6.1 Pre-emphasis:

Pre-emphasis refers to a system process designed to increase, within a band of frequencies, the magnitude of some (usually higher) frequencies with respect to the magnitude of the others(usually lower) frequencies in order to improve the overall SNR. Hence, this step processes the passing of signal through a filter which emphasizes higher frequencies[33].

6.2 Framing:

The process of segmenting the speech samples into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M ($M < N$) [33].

6.3 Hamming window:

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines. The Hamming window is represented as shown in the Eq. 3 window is defined as $W(n)$, $0 \leq n \leq n-1$ where

N = number of samples in each frame

$Y[n]$ = Output signal

$X(n)$ = input signal

$W(n)$ = Hamming window, then the result of windowing signal is shown below [33]:

$$\mathrm{Y[n] = X(n) * W(n)} \tag{3}$$

6.4 Fast fourier transform:

To convert each frame of n samples from time domain into frequency domain FFT is being used. The Fourier Transform is used to convert the convolution of the glottal pulse $U[n]$ and the vocal tract impulse response $H[n]$ in the time domain. This statement supports as shown in Eq.4 below:

$$\mathrm{Y(w) = FFT\,[h(t)^{*}X(t)] = H(w)^{*}X(w)} \tag{4}$$

If $X(w)$, $H(w)$ and $Y(w)$ are the Fourier Transform of $X(t)$, $H(t)$ and $Y(t)$ respectively [33].

6.5 Mel filter bank processing:

The frequencies range in FFT spectrum is very wide and voice signal does not follow the linear scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the Centre frequency and decrease linearly to zero at centre frequency of two adjacent filters. Then, each filter output is the sum of its filtered spectral components. After that the following equation as shown in Eq.5 is used to compute the Mel for given frequency f in HZ [33]:

$$\mathrm{F(Mel) = [2595 * \log 10[1 + f/700]} \tag{5}$$

6.6 Discrete cosine transform:

This is the process to convert the log Mel spectrum into time domain using DCT. The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector [33].

6.7 Mel Frequency Cepstrum Coefficient:

MFCCs are coefficients that represent audio based on perception with their frequency bands logarithmically positioned and mimics the human vocal response [34].

7. Excitation Source Parameter

7.1 Pitch:

Pitch is one of the most important features for identifying speech disorders and has traditionally been considered the most effective tool for detecting sounds. Pitch is mainly related to the fundamental frequency of the vibration of the vocal cords, which enables the tone's frequency [35].

7.2 Jitter and Shimmer:

As with extraction, jitter and shimmer measurements involve the initial segmentation of the audio signal into fundamental periods, and are often used to assess speech function. These two measures are widely used in the literature to classify speaker voice features and diagnose voice pathology [35].

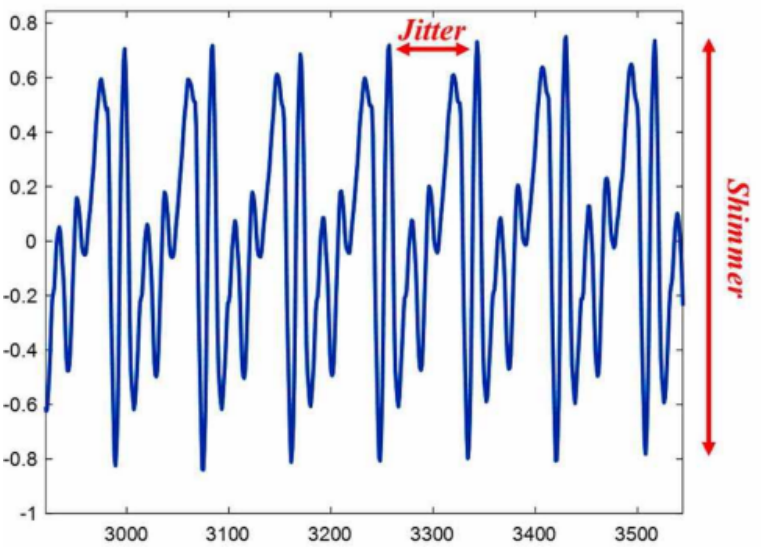


Figure 15: Jitter and Shimmer perturbation measures in speech signal.

- **7.2.1 Jitter**

  Many metrics can be calculated using the Jitter, including absolute jitter (jitta), local or relative jitter (jitt), relative average perturbation (rap), and five-sample jitter (ppq5) [35].

  Jitter (absolute): is the change in fundamental frequency from cycle to cycle, which is the average absolute deviation between successive cycles, measured in seconds or μs

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}| \tag{6}$$

  where T_i is the duration of each period and N denotes the total number of cycles.

  Jitter (relative): the mean absolute difference between two successive periods divided by the mean period. It can be determined by the following:

$$jitt = \frac{jitta}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{7}$$

  Jitter (rap): It represents the mean disorder, which is defined as the mean absolute difference between a period and the mean of the period with its two neighbors divided by the mean period:

$$\text{rap} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{8}$$

Jitter (ppq5): The mean absolute difference between a period and the mean including its four nearest neighbor periods,the two previous and two subsequent periods, divided by the mean period, is known as jitter (ppq5):

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-2}\left|T_i - \left(\frac{1}{5}\sum_{n=-2}^{i+2} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{9}$$

- **7.2.2 Shimmer**

The shimmer is another voice feature that is widely used in voice disability detection. Unlike Jitter, Shimmer focuses on the peak values of a signal. To determine Shimmer parameters, the algorithm begins by determining the onset time of glottal pulses of a signal and the respective magnitude of the signal at that sample. The algorithm is then applied to determine the values of each parameter of Shimmer similarly as for Jitter [2]. There are several Shimmer parameters as follows [35]: **Shimmer (local):** the mean absolute difference between the amplitudes of two consecutive periods divided by the mean amplitude:

$$\text{Shim} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \tag{10}$$

Shimmer (local, dB): It represents the mean absolute difference of the base 10 logarithm of the difference between two consecutive periods, called ShdB. It is given by:

$$\text{ShdB} = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20^{*}\log\left(\frac{A_{i+1}}{A_i}\right)\right| \tag{11}$$

Shimmer (apq3): It represents the quotient of the three-point amplitude perturbation, in other words, the mean absolute difference between the amplitude of a period and the mean amplitudes of its two neighbors, divided by the average amplitude. It is given by:

$$\text{apq}\,3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \tag{12}$$

Shimmer (apq5): It is calculated as the mean absolute difference between the amplitude of a period and the mean amplitude of its four nearest neighbors divided

by the mean amplitude of five periods. apq5 is given by:

$$\text{apq5} = \frac{\frac{1}{N-1}\sum_{-2}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n-1}^{i+2} A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \times 100 \tag{13}$$

III. Classification

After feature extraction, the next stage is classification. It is an instance of supervised learning and can be defined as the problem of identifying the class to which a new observation belongs [3].

Research works in the domain of pathological voice detection and classification started in early 1980s when pattern recognition and Machine Learning (ML) techniques were not much advanced and known. However, in recent days, ML algorithms are popularly used for detection of pathological voice based on the computed acoustic features of the input signal [28].

Considered a subset of Artificial Intelligence (AI), ML exhibits the experiential "learning" associated with human intelligence, while also having the capacity to learn and improve its analyses through the use of computational algorithms. These algorithms use large sets of data inputs and outputs to recognize patterns and effectively "learn" in order to train the machine to make autonomous recommendations or decisions [36].

ML techniques including Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Random Forest (RF) and Decision Tree (DT) have been generally used in PD classification.
In our study we focus on KNN, SVM and DT.

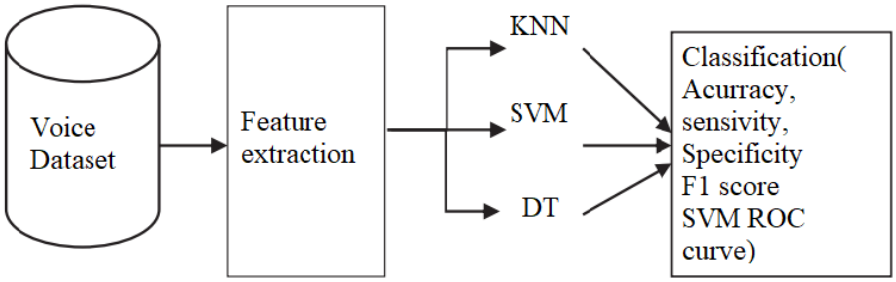


Figure 16: Methods applied for PD classification in our study

1. K-Nearest Neighbors KNN:

K-nearest neighbor (KNN) is one of the simplest ML algorithms, also called as lazy learner [28]. The KNN is a basic simple method in pattern recognition and when the information about the data distribution is not enough the method is amongst the first classification methods chosen. This method includes two parts:

1. determining K close neighbors.
2. determining class type using these close neighbors.

Suppose the training data space D described as:

$$D = \{X_1, X_2, \ldots, X_n\} \tag{14}$$

Which includes n samples and each sample Xi is defined by f features as:

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{if}) \tag{15}$$

And the whole data include C different classes. For determining the X' data class, first of all, its distance from all the data in space D is measured and then K data are determined from space D which are in the nearest neighborhood of X'. Class type of all K data is determined and so the X' belongs to the class which has had the most iteration between all K data. There are various criteria to evaluate the distance between X' and Xi but the most common one is the Euclidean distance criteria. Euclidean distance between X' and Xi along with the feature dimension, f, is defined as [37] [38] :

$$d = \sqrt{(x_{i1} - x'_1)^2 + (x_{i2} - x'_2)^2 + \ldots + \left(x_{if} - x'_f\right)^2} \tag{16}$$

2. Support Vector Machine SVM:

A Support Vector Machine (SVM) is a supervised algorithm that classifies data and is also well-suited for regression purposes [18].

SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors.The basic SVM takes a set of input data and, for each given input, predicts which of two possible classes forms the output, making it a

non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm creates a model that assigns new examples into one class or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a class based on which side of the gap they fall on.

Figure 17 support vector machine constructs a hyperplane or a set of hyperplanes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (also functional margin), since in general the larger the margin the lower the generalization error of the classifier. A linear support vector machine consists of a set of given support vectors z and a set of weights w. The calculation for the output of a given SVM with N support vectors $z_1, z_2..., z_N$ and weights $w_1, w_2, ..., w_N$ is then given by [39]:

$$F(x) = \sum_{i=1}^{N} \text{Wik } (Zi, x) + b \tag{17}$$

Where $k(Zi, x)$ is the Radial Basis Function (RBF) kernel function, adopted in this work [35]

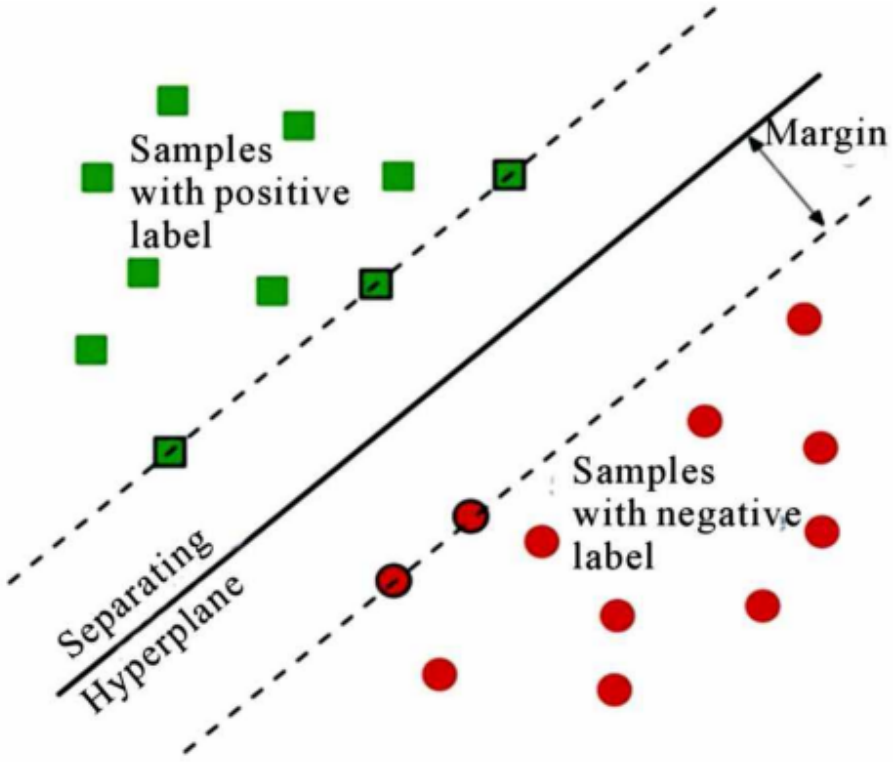


Figure 17: SVM Concept

3. Decision Tree

The decision tree algorithm is a flowchart-like tree structure. In decision algorithms, an internal node represents feature, the branch represents a decision rule, and each leaf node represents a result. The top-level node, in a decision tree, is known as the root node. The decision tree algorithms learn to partition data based on the feature value. It recursively divides the decision tree. This flowchart-like structure facilitates decision making process , similar to human-level thinking. Unlike other neural networks, decision tree algorithms share internal decision logic. The main advantage of decision tree algorithms is that they are faster than other neural networks. The complexity of decision trees lies in the number of datasets and the number of attributes in a given dataset. The decision tree algorithms do not depend on probability distribution assumptions. Therefore, they can handle high-dimensional data with good accuracy [2].

IV. Conclusion

In this chapter, In a first part in the feature extraction techniques were defined in general. After that, the techniques used in our study are mentioned and focused on, which are MFCC, Pitch, Jitter and Shimmer. In the second part, Machine Learning Classifiers used in our study KNN, SVM,DT are presented and defined.

In the next chapter, we will share the experimental results and discuss it.

Chapter 3

Experimentation, Results and Interpretation

I. Introduction

Recent studies have extracted the features from vocal disorders as an early warning sign for PD detection, as patients experience voice changes and impairments in the early stages of PD. At present, the application of Machine Learning (ML) is thriving in the field of prediction. therefor, it is widely used in PD [18].

In this chapter we explain the results obtained by the application of the proposed method for Parkinson's disease classification. First, we present the system design, then the state of art. After that, we introduce the experimental database, setup and metrics. Eventually the experimental results are shown and discussed. We end this chapter with a conclusion.

II. PD Detection Using Speech Signals

The diagnosis of Parkinson's disease (PD) is usually based on medical observation and evaluation of clinical signs, including the characterization of various motor symptoms. However, traditional diagnostic methods can suffer from subjectivity, as they rely on the assessment of movements of the human eye that are sometimes subtle and therefore difficult to classify, leading to possible misclassification. Meanwhile, the early non-motor symptoms of PD can be mild and can be caused by many other conditions. As a result, these symptoms are often overlooked, making early diagnosis of PD difficult. To address these difficulties and improve diagnostic and assessment procedures for PD, machine learning approaches for classification of PD and healthy controls or patients with similar clinical presentations have been implemented [4].

Therefore, the early detection of PD can be done by analyzing speech signals, as 90% of PD patients face vocal problems in the incipient stage of the disease. Consequently, the recent focus of PD detection research emphasizes the vocal disorders of patients.

In current studies, clinical features from the speech of PD patients were extracted and fed into a classification model with the help of various speech signal processing algorithms. Based on speech recordings, this telemonitoring study mapped the vocal features of PD to a clinical evaluation system that predicted the possibility of PD in patients. Moreover, the collection of speech data was a non-invasive process, which made

the data easy to collect and subsequently provide as the input of the telediagnosis system [18].

Many ML methods including Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Decision Tree (DT), and Support Vector Machine (SVM) have been used in PD classification based on patient vocal disorders in many studies. Anyhow, the success rate of accurate detection depends on the quality of data, on the relevance of the features extracted from them, and on the associated ML models. Various recent studies have been conducted on a publicly available dataset that consists of the sound measurements of 8 healthy and 23 PD-affected instances, aggregating 195 data samples [40]. Another publicly available dataset includes the data of 20 PD patients and 20 healthy individuals [41]. Both datasets consist of some common features extracted from the speech signals, including vocal fundamental frequency, measures of the ratio of the noise-to-tonal components, measures of variation in amplitude, measures of variation in fundamental frequency, etc. Since a good number of the studies regarding PD classification from vocal features are conducted with these datasets, the features extracted from these datasets are referred to as the baseline features. Other features are also used to detect PD, such as Mel-frequency Cepstral Coefficients (MFCC) and Signal-to-Noise Ratio (SNR) [18].

Our study adopted two different feature extraction techniques, which are the MFCC and excitation source parameter. MFCC is a very simplified model of auditory processing; it is easy and relatively fast to compute. Moreover, MFCC can imitate the characteristics of the human ear and have been used in different speech recognition tasks [42].

For the choice of Pitch , Jitter and Shimmer extraction, Speech and voice can be researched through voice analysis and determination of some parameters of speech and language, such as subtle changes in voice frequencies (jitter), voice cycle-to-cycle magnitude difference (shimmer), volume (amplitude). In terms of speech analysis, people with Parkinson's have higher jitter and shimmer, decreased pitch range and increased phonation threshold pressure [43].

To classify between a PD patient and a healthy individual KNN, SVM, DT classifiers are employed independently after the feature extraction method to get the highest

performance.

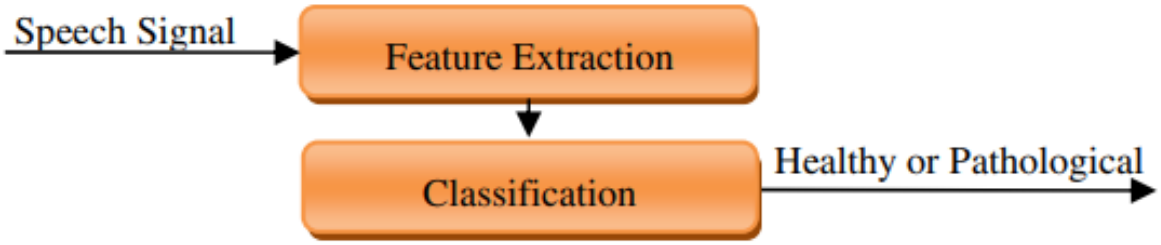


Figure 18: general scheme of vocal fold pathology diagnosis

III. State of Art:

In this section, Some of the different recent studies on PD diagnosis and classification using machine learning algorithms have been summarized.

Karimi Rouzbahani and Daliri [15], used voice signals for PD detection. Measurements like pitch (F0), jitter and shimmer. Using a combination of the three available classifiers and the seven most important features allowed them to have the best performance . The selected features were used to train the SVM, KNN, and discrimination-function-based (DBF) classifiers. KNN performed the best with an accuracy rate of 93.82%.

In Ma et al.[44] a new hybrid process is proposed, an ensemble of Subtractive Clustering Features Weighting (SCFW) and Kernel-based Extreme Learning Machine (KELM) for the diagnosis of PD patients. SCFW, a data-preprocessing tool, is used to reduce the variance in the dataset. The suggested method outperformed the SVM-based, ELM-based, KNN-based, and other methods with a classification accuracy of 99.49%.

Zuo et al. [45] proposed an effective method for the diagnosis of PD patients using Particle Swarm Optimization (PSO) that supported Fuzzy K-nearest Neighbor (FKNN). This PSO-FKNN model was justified in terms of the accuracy, specificity, sensitivity, and the AUC of the ROC curve. The proposed model achieved a mean accuracy rate of 97.47%.

Sharma and Giri [24] implemented three methods of classification based on MLP, KNN, and SVM to diagnose PD patients. In these classifiers, SVM with an RBF kernel

showed the best result with an accuracy rate of 85.294%.

Apart from these, various other studies have been performed, that extracts features from different datasets and various ML classifiers where used to diagnose PD patients.

Table 1 summarizes the studies mentioned above and some of the studies available in the literature for PD diagnosis and classification using machine learning algorithms.

Table 1: Summary of available studies in the literature for PD diagnosis with ML classifiers.

Name	feature	Classifier	Accuarcy
Karimi Rouzbahani and Daliri [15]	HNR	KNN	93.82%
Ma et al[44]	SCFW	SVM, ELM, KNN	99.49%
Zuo et al[45]	PSO	PSO-FKNN	97.47%
Sharma and Giri[24]	MLP	SVM	85.294%
Parisi et al[46]	MLP	LSVM	100%
Wroge et al[47]	AVEC, MFCC	DNN	85%
Song Pan et al.[48]	LFP	SVM	81.14%
G.Sateesh Babu and S.Suresh[49]	MCN	ICA	95.55%
R.Armananzas et al[50]	EDA	KNN	80.06%
A. Benba et al[51]	PCA	SVM	87.50%

System Design

The Proposed system to help PD diagnosis is represented by the following flowchart:

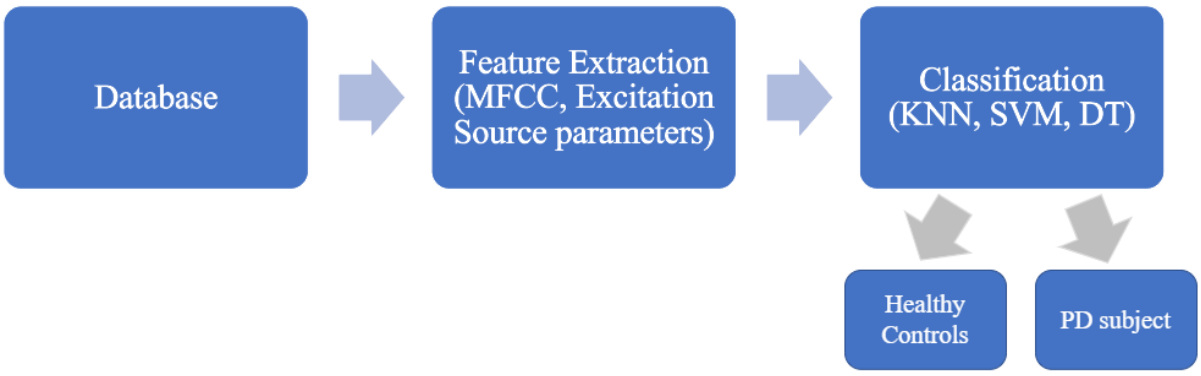


Figure 19: Proposed system to help PD diagnosis.

IV. Experimental database

The MDVR-KCL dataset was recorded at King's College London (KCL) Hospital, Denmark Hill, Brixton, London SE5 9RS. This dataset is freely available online and can be easily downloaded.

For the recording procedure, a Motorola Moto G4 Smartphone is used as recording device. Due to the fact, that they directly record the microphone signal, they end up with high quality recordings with a sample rate of 44.1 kHz and a bit depth of 16 Bit (audio CD quality). The speech were saved in .WAV formats.

In this dataset they tend to Ask the participant to read out a specific text, or tend to start a spontaneous dialogue with him, the test executor starts asking random questions about places of interest, local traffic, or personal interests if acceptable [52].

The voice recordings are labeled with; first, IDNN, with $N \in [0, 9]$ which refers to the subject identification. Next to it is the health status label (hc or pd accordingly).

The provided dataset, has a total of 42 healthy control (21 Reading Text,21 Spontaneous Dialogue) and 31 of PD (16 Reading Text, 15 Spontaneous Dialogue).

V. Experimental Setup

1. Development Software:

1.1 MATLAB

MATLAB is an interactive numerical computing system widely used in teaching and research in industry and academia [53].The 2019 version was used (Matlab R2019a).

The choice of MATLAB was made because of its speed, its rich library,it provides a modern programming language and problem-solving environment with powerful data structures, customizable graphics, and easy-to-use editing and debugging tools.

1.2 PRAAT:

Praat is a computer program for analyzing, synthesizing and manipulating speech and other sound and create publication-quality graphics. It is open source and freely avail-

able for all major computing platforms (MacOS, Windows, Linux) on 32-bit and 64-bit operating systems. It can be downloaded from praat.org [54]. These features make it a complete tool especially for speech study.

In our study we used PRAAT as an easy tool to apply a manual segmentation process that will be clarified in the next section.

2. Development Hardware:

- Desktop HP Compaq Elite 8300 MT
- CPU: Intel(R) Core (TM) i5-3470 CPU @ 3.20GHz (4 CPUs), 3.2GHZ
- RAM: 8192MB RAM

3. Method setup

3.1 Database segmentation:

Our system is based on supervised learning machines requiring the use of labeled databases to extract relevant attributes and to estimate classification models. In our case, we used the PD database mentioned above, and to be more specific the Pd subject and Healthy control speech.

We used PRAAT as an easy tool to manually segment the provided database. Where we took out shorter clips from the speech targetting the concerned person only, whether the healthy person or the PD patient and labeled them with PD or HC.

3.2 Feature Extraction process:

The basic principle here is to extract a sequence of discriminating features from the input signal. The efficiency of this phase is important for the next phase since it affects the behavior of classification process.

Our work at this section is divided into two parts; first, MFCC features, then the Excitation Source Parameters. Both parts of the feature extraction consists of using the MATLAB software.

3.2.1 MFCC process

We used the MFCC MATLAB function to extract MFCC coefficients and log energy. Although there is no hard standard for calculating the coefficients, the basic steps are outlined by the diagram in figure 20. We adjusted in the window length and overlap values.

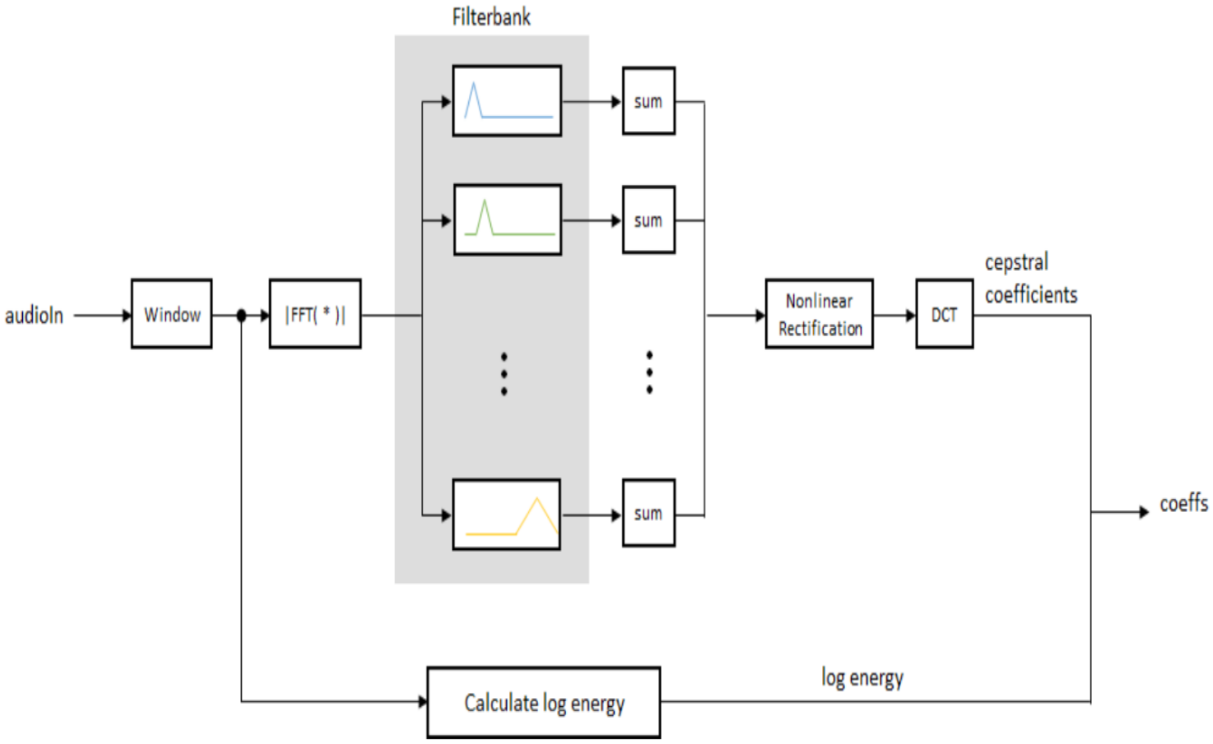


Figure 20: MFCC diagram

After the process we obtained 13 coefficients plus the log energy forming a 14 coefficients in total. This latter is the replacement of the information contained in the zeroth Mel frequency cepstral coefficient and its calculation depends on the input domain (time-domain signal or frequency-domain signal), in our case it is the time-domain signal and it is calculated using the equation:

$$\log E = \log\left(\text{sum}\left(x^2\right)\right) \tag{18}$$

3.2.2 Excitation Source Parameters

Pitch was also applied by the MATLAB function, The parameters we set are window length and overlap length, then we extracted the mean pitch, max pitch and min pitch.

Jitter and Shimmer were calculated by their mathematical equations given in chapter 2 using MATLAB. Where we extracted from our signals : Jitter(absolute), Jitter (relative), Jitter (rap), Jitter (ppq5), Shimmer (Shimmer (local), Shimmer (Shimmer (local, dB)), Shimmer (apq3), Shimmer (apq5), with the pitch parameters extracted previously, forming a total of 11 features, which will form our feature matrix.

3.3 Classification process:

To classify sounds in two classes:

- Class (1) presents patients with PD,
- Class (0) presents Healthy controls

We used the KNN, SVM, DT classifiers algorithms using MATLAB after the construction of the features matrix.

The steps followed for classification are:

- For the KNN we set the Number of neighbors to the default value, in other words, the default KNN classifier uses a single nearest neighbor only.
- For the SVM the Kernel Radial Basis Function (RBF) was used to build our classification model. Where in SVM classification, the (RBF) is the most popular kernel function. It is given by [55]:

$$k(x, y) = \exp\left(\frac{\|x - y\|^2}{2\sigma^2}\right) \tag{19}$$

  Where σ represents the width of the basis function, and $\|x - y\|^2$ is the similarity function.

- The DT was achieved using the TreeBagger function. This latter grows the decision trees in the ensemble using bootstrap samples of the data. Also, TreeBagger selects a random subset of predictors to use at each decision split.
- Construction of the prediction model for each classifier.

VII. Experimental metrics:

The system proposed was used to detect PD patients using voice. For the experiment, the MDVR-KCL dataset mentioned above was used to assess the model performance, then a comparison of the KNN, SVM, and DT classifiers is presented. The results of the classifiers were measured in terms of the accuracy, sensitivity, specificity, F1 score and the SVM ROC curve.

These performance indicators are widely used to assess the effectiveness of various medical decision systems. We can define them as follows [56] :

- **Accuracy:** as a measure of the capability of classifying the samples correctly, which is expressed as follows:

$$\text{Accuracy}(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{20}$$

- **Sensitivity:** the probability that abnormal samples will be diagnosed as positive.

$$\text{Sensitivity } (\%) = \frac{TP}{TP + FN} \times 100 \tag{21}$$

- **Specificity:** the probability of normal samples being incorrectly identified.

$$\text{Specificity}(\%) = \frac{TN}{TN + FP} \times 100 \tag{22}$$

Where TP,TN,FP and FN are :

- **True positive (TP):** When pathologic voice was present, the classifier identified it.
- **True negative (TN):** When the healthy voice was present, the classifier identified it.
- **False positive (FP):** When the healthy voice was present, the classifier identified disease (false acceptance).
- **False negative (FN):** When disease was present, the classifier recognized the healthy (false rejection).

- **F1 Score:** F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'.

$$F1 \text{ Score } = 2 \times \frac{\text{Precision } \times \text{ sensitivity}}{\text{precision } + \text{ sensitivity}} [57]. \tag{23}$$

where the Precision is given by :

$$\text{Precision} = \frac{TP}{TP + FP} \tag{24}$$

- **ROC curve:** ROC stands fro Receiver Operating Characteristic. To produce a ROC curve, the sensitivities and specificities for different values of a continuous test measure are first tabulated. This results, essentially, in a list of a various test values and the corresponding sensitivity and specificity of the test at that value. Then, the graphical ROC curve is produced by plotting sensitivity (true positive rate) on the y-axis against 1– specificity (false positive rate) on the x-axis for the various values tabulated [58].

VIII. Results and discussions

The proposed system was used to classify PD patients and healthy controls. For the experiment, the MDVR-KCL dataset mentioned previously was used to evaluate the model performance to find out the best one, and comparisons with KNN, SVM, and DT were exhibited.

First, we started with the MFCC and implemented the three ML classifiers independently. The settings used are:

- We had a 65% from the dataset as a training data.
- Window Length=0.031s, Overlap Length=0.018s, these parameters were used for all the classifiers with the MFCC technique.

The table 2 exhibit the highest results obtained for the three classifiers SVM, KNN, and DT using the MFCC parameter, in terms of the four adopted indexes (accuracy, sensitivity, specificity, F1 Score and SVM ROC curve).

Table 2: Results of the KNN,SVM,DT classifiers with MFCC features

Classifier	Accuracy%	Sensitivity%	specificity%	F1 score%
SVM	97.36	95	100	97.44
KNN	94.73	90.47	100	95
DT	94.73	90.47	100	95

Table 2 shows that SVM achieved the highest accuracy using the MFCC features where the accuracy reached 97.36%. For the KNN and DT, these two classifiers showed the same results and their accuracy reached 94.73% both equally.

Next, the table 3 exhibits the achieved results for the three classifiers KNN, SVM and DT using the excitation source parameter. The settings used in this method are:

- 65% training data
- For the KNN; the window length=0.032s, overlap length=.009s.
- For the SVM; window length= 0.028s, overlap length=0.009s.
- For the DT; the window length=0.032s, overlap length=0.009s.

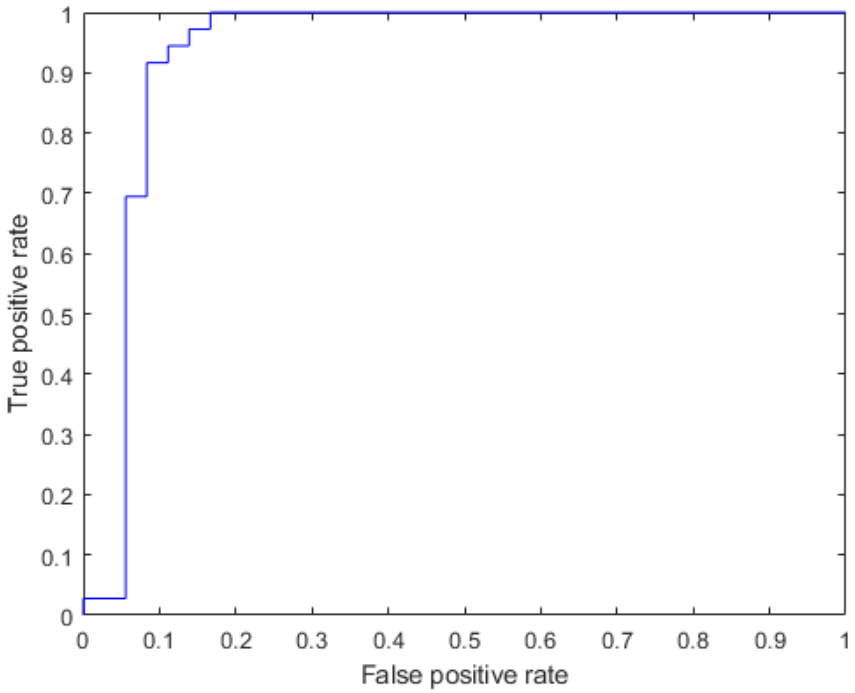
Table 3: Results of KNN ,SVM, DT with excitation source parameter

Classifier	Accuracy%	Sensitivity%	specificity%	F1 score%
SVM	92.10	90	94.44	92.31
KNN	89.47	82.60	100	90.48
DT	94.73	90.47	100	95

Table 3 Shows that the DT classifier obtained the highest accuracy using the excitation source parameter where the accuracy reached 94.73% which is the same result when using the MFCC features in the previous table. Secondly, the SVM has a 92.10% accuracy after the DT. Meanwhile, the KNN showed the lowest results using these parameters with an accuracy rate of 89.73%.

About the sensitivity, specificity and F1 score, they were relative to the best accuracy. SVM with the MFCC parameter had the highest sensitivity and specificity of 95%, 100% and 97.44% respectively. Meanwhile, DT with MFCC features and with the Pitch, Jitter and Shimmer had 90.47%, 100% and 95% respectively.

Finally, SVM with the MFCC parameters achieved the best results compared to all the methods applied with an accuracy rate of 97.36%, sensitivity 95%, 100% and F1 score 97.44%. Therefore, the figure 21 shows the ROC curve of this result.



p

Figure 21: ROC curve of the SVM Classifier Results

Figure 21 shows a plot of SVM with MFCC ROC curve, from this figure the true positive rate vs false positive rate can be easily identified. SVM gave a curve closer to the top-left corner which indicates a better performance. As a baseline, a random classifier is expected to give points lying along the diagonal (FP = TP).

IX. Conclusion

We have proposed a method for identifying Parkinson's Disease voices based on Feature Extraction Techniques and Machine Learning Classifiers.

In this work, we use the MDVR-KCL Parkinson's dataset to experimentally evaluate the performance of three Machine learning classification methods using two different feature extraction techniques independently. The machine learning classification methods are a K-Nearest Neighbor, Support Vector Machine, and Decision Tree, and the feature extracted are the Mel-Frequency Cepstral Coefficient and excitation source parameter.

We implement 65% of a training data for each of the three methods and obtain the classification accuracy results. The classification results show that the Support Vector Machine with the MFCC feature produces higher accuracy rate of 97.36%, followed by the Decision Tree with the MFCC features and the excitation source parameter equally with an accuracy rate of 94.73%. Then the lowest accuracy resulted from the KNN with the excitation source parameter with a rate of 89.47%.

General Conclusion and Future Work

Conclusion

The initial objective of this work is to design a diagnostic support system for the early detection of Parkinson's Disease (PD) from the voice. Mainly, it is about differentiating between PD subject and healthy control. The proposed method consists of two parts:

The first part: is the feature extraction using the Maltab Mathworks software. We worked with the MFCC features then the parameters(pitch, Jitter (local), Jitter (absolute), Jitter (rap), jitter (ppq5), Jitter (ddp), Shimmer (local), Shimmer (dB), Shimmer (apq3) Shimmer (apq5), Shimmer (dda)) independently.

The second part: concerns the classification between the two categories: PD subject or healthy controls. We used the three machine learning classifiers KNN, SVM and DT independently. The MATLAB Mathworks software was also used for this simulation. The speech used was extracted from the MDVR-KCL database.

The highest accuracy was achieved using the SVM classifier with the MFCC features with an accuracy rate of 97.36%.

Future Work

Future work will be focusing on:

- Apply different feature extraction techniques.
- We will try to develop deep learning methods.

- We will try to use a new database to evaluate our model performance.
- The method proposed in our study may, in the future, be the subject of an Android application that would be accessible and widely used for the identification of pathological voices.

Bibliography

[1] K. Omori, "Diagnosis of voice disorders," *JMAJ*, vol. 54, no. 4, pp. 248–253, 2011.

[2] R. Islam, M. Tarique, and E. Abdel-Raheem, "A survey on signal processing based pathological voice detection techniques," *IEEE Access*, vol. 8, pp. 66 749–66 776, 2020.

[3] G. Pahuja and T. Nagabhushan, "A comparative study of existing machine learning approaches for parkinson's disease detection," *IETE Journal of Research*, vol. 67, no. 1, pp. 4–14, 2021.

[4] J. Mei, C. Desrosiers, and J. Frasnelli, "Machine learning for the diagnosis of parkinson's disease: A review of literature," *Frontiers in aging neuroscience*, vol. 13, p. 184, 2021.

[5] H. Arakelyan, "Voice disorders," 02 2020.

[6] M. Boudjerda and A. R. Kacha, "Analyse du signal de parole pour l'évaluation automatique des voix pathologiques," Ph.D. dissertation, 2018.

[7] A. Aronson and D. Bless, *Clinical Voice Disorders*. Thieme, 2011. [Online]. Available: https://books.google.dz/books?id=kmugBjlqGBkC

[8] J. R. Deller Jr, "Discrete-time processing of speech signals," in *Discrete-time processing of speech signals*, 1993.

[9] S. Mrad, "Analyse non linéaire de la parole pour la détection des voix pathologiques," Ph.D. dissertation, Ecole Nationale d'Ingénieurs de Tunis, 2013.

[10] C. D. C. D'ORTHOPHONISTE, "L'intensite: Parametre acoustique en pathologies de la communication."

[11] *Cambridge Handbook of Psychology, Health and Medicine*. CAMBRIDGE UNIV

Press, 2018.

[12] L. Mathieson, *Greene and Mathieson's The Voice and its disorders,chapter 6*. Whurr, 2013.

[13] Y. Naqvi and V. Gupta, "Functional voice disorders," 2022.

[14] C. Dastolfo-Hromack and E. Walsh, "Evaluation of neurogenic voice disorders," in *Neurologic and neurodegenerative diseases of the larynx*. Springer, 2020.

[15] H. K. Rouzbahani and M. R. Daliri, "Diagnosis of parkinson's disease in human using voice signals," *Basic and Clinical Neuroscience*, 2011.

[16] A. Saikia, M. Hussain, A. R. Barua, and S. Paul, "An insight into parkinson's disease: researches and its complexities," in *Smart Healthcare for Disease Diagnosis and Prevention*. Elsevier, 2020, pp. 59–80.

[17] S. A. Mostafa, A. Mustapha, S. H. Khaleefah, M. S. Ahmad, and M. A. Mohammed, "Evaluating the performance of three classification methods in diagnosis of parkinson's disease," in *International Conference on Soft Computing and Data Mining*. Springer, 2018, pp. 43–52.

[18] M. Hoq, M. N. Uddin, and S.-B. Park, "Vocal feature extraction-based artificial intelligent model for parkinson's disease detection," *Diagnostics*, vol. 11, no. 6, p. 1076, 2021.

[19] S. Chapuis, L. Ouchchane, O. Metz, L. Gerbaud, and F. Durif, "Impact of the motor complications of parkinson's disease on the quality of life," *Movement disorders: official journal of the Movement Disorder Society*, vol. 20, no. 2, pp. 224–230, 2005.

[20] T. A. Zesiewicz and M. L. Evatt, "Potential influences of complementary therapy on motor and non-motor complications in parkinson's disease," *CNS drugs*, vol. 23, no. 10, pp. 817–835, 2009.

[21] C. Hinnell, C. S. Hurt, S. Landau, R. G. Brown, M. Samuel, and P.-P. S. Group, "Nonmotor versus motor symptoms: how much do they matter to health status in parkinson's disease?" *Movement Disorders*, vol. 27, no. 2, pp. 236–241, 2012.

[22] J. Massano and K. P. Bhatia, "Clinical approach to parkinson's disease: features, diagnosis, and principles of management," *Cold Spring Harbor perspectives in medicine*,

vol. 2, no. 6, p. a008870, 2012.

[23] N. Narendra, B. Schuller, and P. Alku, "The detection of parkinson's disease from speech using voice source information," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1925–1936, 2021.

[24] A. Sharma and R. N. Giri, "Automatic recognition of parkinson's disease via artificial neural network and support vector machine," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 4, no. 3, pp. 2278–3075, 2014.

[25] N. Fayyazifar and N. Samadiani, "Parkinson's disease detection using ensemble techniques and genetic algorithm," in *2017 Artificial Intelligence and Signal Processing Conference (AISP)*. IEEE, 2017, pp. 162–165.

[26] V. Majidnezhad and I. Kheidorov, "A novel method for feature extraction in vocal fold pathology diagnosis," in *International Conference on Wireless Mobile Communication and Healthcare*. Springer, 2012, pp. 96–105.

[27] P. Murugesapandian, S. Yaacob, and M. Hariharan, "Feature extraction based on mel-scaled wavelet packet transform for the diagnosis of voice disorders," in *4th Kuala Lumpur International Conference on Biomedical Engineering 2008*. Springer, 2008, pp. 790–793.

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

[29] S. C. Joshi and A. Cheeran, "Matlab based feature extraction using mel frequency cepstrum coefficients for automatic speech recognition," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 3, no. 6, p. 1820, 2014.

[30] D. O'Shaughnessy, "Linear predictive coding," *IEEE potentials*, vol. 7, no. 1, pp. 29–32, 1988.

[31] K. Gupta and D. Gupta, "An analysis on lpc, rasta and mfcc techniques in automatic speech recognition system," in *2016 6th international conference-cloud system and big data engineering (confluence)*. IEEE, 2016, pp. 493–497.

[32] D. Prabakaran and R. Shyamala, "A review on performance of voice feature extraction techniques," in *2019 3rd International Conference on Computing and Communications Technologies (ICCCT)*. IEEE, 2019, pp. 221–231.

[33] A. Bala, A. Kumar, and N. Birla, "Voice command recognition system based on mfcc and dtw," *International Journal of Engineering Science and Technology*, vol. 2, no. 12, pp. 7335–7342, 2010.

[34] S. Shetty, S. Hegde, T. Dodderi *et al.*, "Classification of healthy and pathological voices using mfcc and ann," in *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC)*. IEEE, 2018, pp. 1–5.

[35] E. Hadjaidji, M. C. A. Korba, and K. Khelil, "Spasmodic dysphonia detection using machine learning classifiers," in *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*. IEEE, 2021, pp. 1–5.

[36] J. M. Helm, A. M. Swiergosz, H. S. Haeberle, J. M. Karnuta, J. L. Schaffer, V. E. Krebs, A. I. Spitzer, and P. N. Ramkumar, "Machine learning and artificial intelligence: definitions, applications, and future directions," *Current reviews in musculoskeletal medicine*, vol. 13, no. 1, pp. 69–76, 2020.

[37] R. A. Shirvan and E. Tahami, "Voice analysis for detecting parkinson's disease using genetic algorithm and knn classification method," in *2011 18th Iranian Conference of Biomedical Engineering (ICBME)*. IEEE, 2011, pp. 278–283.

[38] P. A. Devijver and J. Kittler, *Pattern recognition: A statistical approach*. Prentice hall, 1982.

[39] V. Sellam and J. Jagadeesan, "Classification of normal and pathological voice using svm and rbfnn," *Journal of signal and Information Processing*, vol. 2014, 2014.

[40] M. Little, P. Mcsharry, S. Roberts, D. Costello, and I. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Nature Precedings*, pp. 1–1, 2007.

[41] B. Erdogdu Sakar, G. Serbes, and C. O. Sakar, "Analyzing the effectiveness of vocal features in early telediagnosis of parkinson's disease," *PloS one*, vol. 12, no. 8, p. e0182428, 2017.

[42] U. Shrawankar and V. M. Thakare, "Techniques for feature extraction in speech recognition system: A comparative study," *arXiv preprint arXiv:1305.1145*, 2013.

[43] L. Berus, S. Klancnik, M. Brezocnik, and M. Ficko, "Classifying parkinson's disease based on acoustic measures using artificial neural networks," *Sensors*, vol. 19, no. 1, p. 16, 2018.

[44] C. Ma, J. Ouyang, H.-L. Chen, and X.-H. Zhao, "An efficient diagnosis system for parkinson's disease using kernel-based extreme learning machine with subtractive clustering features weighting approach," *Computational and mathematical methods in medicine*, vol. 2014, 2014.

[45] W.-L. Zuo, Z.-Y. Wang, T. Liu, and H.-L. Chen, "Effective detection of parkinson's disease using an adaptive fuzzy k-nearest neighbor approach," *Biomedical Signal Processing and Control*, vol. 8, no. 4, pp. 364–373, 2013.

[46] L. Parisi, N. RaviChandran, and M. L. Manaog, "Feature-driven machine learning to improve early diagnosis of parkinson's disease," *Expert Systems with Applications*, vol. 110, pp. 182–190, 2018.

[47] T. J. Wroge, Y. Özkanca, C. Demiroglu, D. Si, D. C. Atkins, and R. H. Ghomi, "Parkinson's disease diagnosis using machine learning and voice," in *2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2018, pp. 1–7.

[48] S. Pan, S. Iplikci, K. Warwick, and T. Z. Aziz, "Parkinson's disease tremor classification–a comparison between support vector machines and neural networks," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 764–10 771, 2012.

[49] G. S. Babu and S. Suresh, "Parkinson's disease prediction using gene expression–a projection based learning meta-cognitive neural classifier approach," *Expert Systems with Applications*, vol. 40, no. 5, pp. 1519–1529, 2013.

[50] R. Armañanzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin, and P. Larrañaga, "Unveiling relevant non-motor parkinson's disease severity symptoms using a machine learning approach," *Artificial intelligence in medicine*, vol. 58, no. 3, pp. 195–202, 2013.

[51] A. Benba, A. Jilbab, and A. Hammouch, "Voice assessments for detecting patients

with parkinson's diseases using pca and npca," *International Journal of Speech Technology*, vol. 19, no. 4, pp. 743–754, 2016.

[52] H. Jaeger, D. Trivedi, and M. Stadtschnitzer, "Mobile device voice recordings at king's college london (mdvr-kcl) from both early and advanced parkinson's disease patients and healthy controls," 2019.

[53] D. J. Higham and N. J. Higham, *MATLAB guide*. SIAM, 2016.

[54] P. Boersma *et al.*, "The use of praat in corpus research," *The Oxford handbook of corpus phonology*, pp. 342–360, 2014.

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

[56] E. Hadjaidji, M. C. A. Korba, and K. Khelil, "Detection of precancerous laryngeal leukoplakia using sub-band based cepstral," in *2021 International Conference on Networking and Advanced Systems (ICNAS)*. IEEE, 2021, pp. 1–5.

[57] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Deep feature based rice leaf disease identification using support vector machine," *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.

[58] Z. H. Hoo, J. Candlish, and D. Teare, "What is an roc curve?" pp. 357–359, 2017.