République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي

University of AbouBekr BELKAID
Faculty of Science
Department of Computer Science

Master's Thesis

For the State Computer Engineering diploma

Option : Artificial intelligence

Early detection of Parkinson's disease using deep learning approaches.

Produced by :
KHELLADI Abdelhamid

Supervised by :
Mr. BENAZZOUZ MOURTADA (UABT)
Pr. MELGANI Farid (UNITN)

Supported on 01/07/2025, Before the jury composed of :

President: Mr. Meziane Abdelfettah (UABT)
Examiner: Mrs. Marouf Rajaa (UABT)

Promotion : 2024/2025

Acknowledgements

*Thanks to **Allah** for giving me the strength and patience to complete this work. Everything I have achieved in my life, I believe, is a gift from Him.*

*I am deeply grateful to my family, especially my **sister** and **parents**, for their efforts, unwavering support, encouragement, and love throughout my journey. They have shaped who I am today. Their guidance and sacrifices have been invaluable to my success.*

*I would like to express my sincere gratitude to my supervisors, **Mr. BENAZZOUZ Mourtada** and **Pr. Melgani Farid**, for their invaluable guidance, support, and encouragement throughout my thesis. Their expertise and insights have been instrumental in shaping my work and helping me achieve my goals. I am truly grateful for their mentorship and the opportunity to learn from them.*

*I would also like to extend my heartfelt thanks to all the members of the **School of AI Algiers** for introducing me to the world of AI and for their invaluable technical support.*

*I would like to thank my brother whom I met at UABT, **HAMMOUMI Tarik**, for all the wonderful moments we shared and for his support and encouragement throughout my project and journey. I am truly grateful for his friendship and wish him all the best in his personal and professional life, and to all the friends group for being a source of joy, motivation, and inspiration.*

*Finally, I would like to express my gratitude to all my friends in Trento for creating such a warm and welcoming environment. I am especially thankful to the Tunisian student group, who made me feel at home with their presence and support during difficult moments. I am deeply grateful for the wonderful and enjoyable memories we shared. I also want to say a massive thank you to **DAOUD Hiba** for all her help and support during the writing of this thesis.*

Contents

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AD Alzheimer's disease
ADLs Activities of Daily Living
ANN Artificial Neural Network
AS All Segments
BN Batch Normalization
CBAM Convolutional Block Attention Module
CNN Convolutional Neural Network
DA Dopaminergic
DLB Dementia with Lewy bodies
FC Fully connected layers
FOG Freezing of Gait
FS First Segment
HC Healthy Controls
HY Hoehn and Yahr
KNN K-Nearest Neighbors
LMS Log-scaled Mel Spectrogram
LSTM Long Short-Term Memory
LSTMs Long Short-Term Memory Networks
MDS-UPDRS III Movement Disorder Society Unified Parkinson's Disease Rating Scale Part III
MLP Multi-Layer Perceptron
MMSE Mini-Mental State Examination
ND Neurodegenerative diseases
NM Neuromelanin
PD Parkinson's disease
PD_Mild Mild Parkinson's Disease
PD_Severe Severe Parkinson's Disease
REM Rapid Eye Movement
RF Random Forest
RNN Recurrent Neural Network
ResNet Residual Network
RT Regression Tree
SE Squeeze-and-Excitation
SNpc Substantia Nigra pars compacta
STFT Short-Time Fourier Transform
SVM Support Vector Machine
VGG Visual Geometry Group
VaD Vascular dementia

Introduction

Context

Parkinsonism-type diseases are a type of neurodegenerative disease: they are caused by damage to certain nerve cells in the brain that help control coordination and precise muscle movement. Parkinsonism-type diseases are a category of neurodegenerative diseases, which are characterised by the degeneration of neurons responsible for regulating motor function. This encompasses Parkinson's disease and other forms of parkinsonism, which refer to conditions that manifest with symptoms analogous to those observed in Parkinson's disease. Early detection of neurodegenerative diseases is crucial, as it enables the implementation of effective therapeutic interventions and facilitates the provision of adequate support to patients and their families, thereby mitigating the progression of the disease and enhancing quality of life.

Problem Statement

The main issue of our study lies in the design and implementation of deep learning models for the early detection of PD. This challenge involves not only the search and processing of medical data but also the development of neural networks capable of efficiently analyzing and interpreting this data to provide rapid and accurate diagnoses. Traditional diagnostic approaches, such as the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and brain imaging, are often time-consuming, expensive, and limited by accessibility issues, clinician subjectivity, and difficulties in tracking disease progression and treatment effectivenes. Speech difficulties are often the first noticeable symptoms of PD, which makes them particularly useful for early diagnosis. Advances in artificial intelligence and the increasing availability of digital health data present a remarkable opportunity to improve diagnostic precision and processes. Our problem statement focuses on the gap concerning automated systems that reliably assist healthcare professionals in detecting signs of PD through speech analysis, enabling intervention and improving patient outcomes.

contribution

In this study, we present a deep learning models for the early detection of Parkinson's disease. The proposed solution involves the development of various convolutional neural network architectures to categorize voice recordings into three distinct classes: healthy, mild, and severe. This classification is achieved following the implementation of a preprocessing stage, feature extraction from recordings and the creation of a mel-spectrogram.

Structure

- **Chapter 1:** Provides a general definition of neurodegenerative diseases and Parkinson's disease.
- **Chapter 2:** Introduces some machine learning and deep learning concepts.
- **Chapter 3:** Provides a comprehensive explanation of the proposed solution and discusses the results achieved.

Chapter 1

Parkinson's Disease Overview

1.1 Introduction

Neurodegenerative diseases (ND) are a group of progressive disorders that can be characterised by progressive loss of damaged neurons.There are several types of ND, one widely recognized type is Alzheimer's disease (AD) is the most prevalent form of dementia and is marked primarily by extracellular amyloid- plaques and intracellular neurofibrillary tangles composed of tau protein, with both familial and sporadic forms influenced by genetic mutations in amyloid precursor protein and presenilin genes as well as environmental risk factors [14].

Vascular dementia (VaD) is another major degenerative cognitive disorder, often coexisting with AD pathology, and is primarily associated with cerebrovascular insults; its pathology may include ischemia-induced amyloid changes, contributing to the overlap observed between these two conditions [14].

Dementia with Lewy bodies (DLB), which is typified by alpha-synuclein-containing Lewy bodies, represents a synucleinopathy frequently overlapping with AD pathology and is distinguished by its cortical and subcortical distribution of abnormal protein aggregates [14]

Parkinsonism-type diseases are type of different types of these disorders they are caused by damage to certain nerve cells in the brain that help control coordination and precise muscle movement.This includes Parkinson's disease and other forms of parkinsonism, which refer to conditions that manifest with symptoms analogous to those observed in Parkinson's disease [15, 16].

1.2 Parkinson's Disease

Parkinson's disease (PD) is one the most common neurodegenerative disorder and one of the fastest-growing neurodegenerative disorders once diagnosed with an emphasized effect on the muscular growth and maintenance of an individual throughout their life [1].

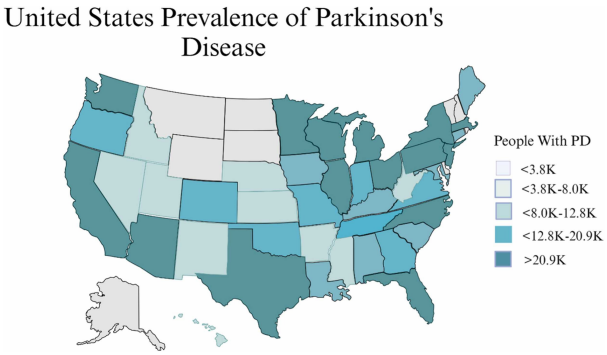


Figure 1.1: Prevalence of Parkinson's Disease in the United States in 2022.[1]

As seen in the figure1.1, PD incidence rates were estimated to be around 40,000–60,000 cases a year. However, the new incidence rates have increased 1.5x to over 90,000 cases a year. The coastal states are seen to be more densely affected by those with PD.[1]

1.2.1 Pathology

As mentioned in the introduction, PD is caused by damage to the group of neurons that are responsible for movement control, and as mentioned in [15] "The pathophysiology of PD primarily includes frontal cortex atrophy and ventricular enlargement. However, the most distinctive morphological alteration observed in the PD brain is the loss of pigmentation in the locus coeruleus and substantia nigra pars compacta (SNpc), which stems from the death of dopaminergic (DA) neuromelanin-containing neurons"

in simpler words the two parts mentioned —the locus coeruleus and the substantia nigra—start to lose their dark color (pigmentation), this color comes from a special neurons that contain neuromelanin (NM) which is absent at birth and naturally increases throughout a person's lifetime.These neurons produce dopamine (dopaminergic),which is important for controlling movement [2]. The loss of these dopamine-producing neurons causes the typical movement problems like shaking, stiffness, and slow movements.

Although the exact cause of this neuronal death is not fully understood, it is believed to involve a combination of genetic, aging-related factors and several other factors, like excessive caffeine intake, smoking, and exposure to environmental toxins [15].

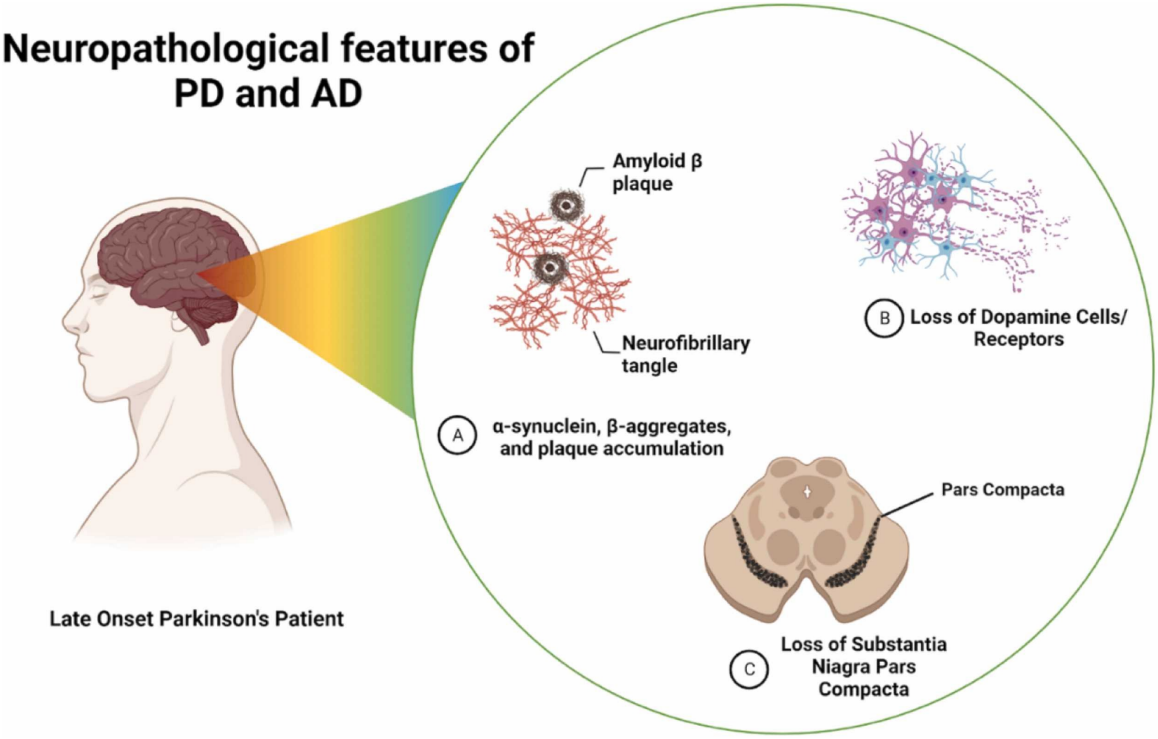


Figure 1.2: neuropathological changes in Parkinson's disease[1]

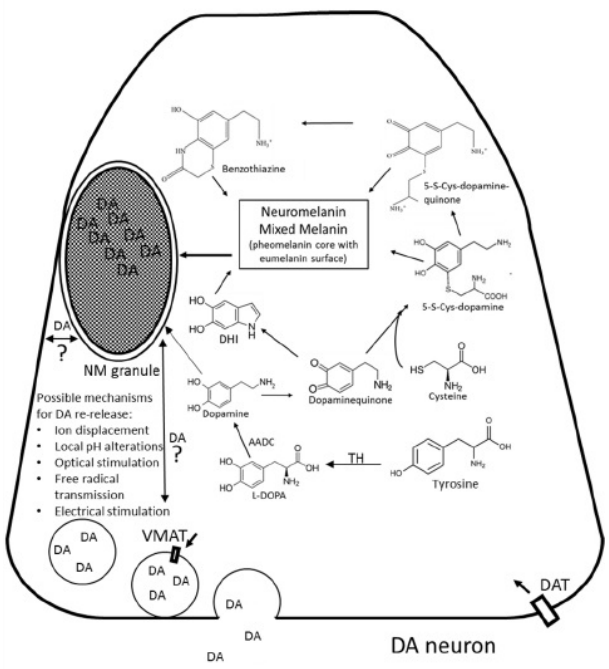


Figure 1.3: Hypothetical role of neuromelanin in dopamin metabolism [2]

Symptoms

PD symptoms can be divided into 2 groups motor(movement) symptoms and nonmotor symptoms(Fig 1.4), and individuals diagnosed with Parkinson's disease typically have gradual development of nonmotor symptoms for years before movement symptoms begin [17].

The figure1.4 depicts the various types of symptoms a patient diagnosed with Parkinson's Disease will undergo throughout their life while living with PD. The image features both the physical and mental symptoms that will affect a late-onset patient. Some of the physical symptoms include Bradykinesia, rigidity, postural instability, and tremors. Some mental symptoms of PD include depression, anxiety, and cognitive impairment. While these symptoms are not necessarily life-threatening, they are debilitating to the patient as they are unable to use their body properly.[1]

Figure 1.4: Parkinson's symptoms : motor and nonmotor effects ou parkinsons disease : motor and cognitive symptoms [1]

1.2.2 Stages of PD

Parkinson's disease (PD) is staged using both clinical and pathological frameworks that capture its complex motor and nonmotor progression. One widely used clinical tool is the Hoehn and Yahr (HY) scale, which classifies patients into five stages based on the severity of motor symptoms and disability [18].

In addition to the HY scale, the neuropathological progression of PD has been conceptualized by the Braak hypothesis. This model describes how Lewy body pathology spreads in the brain over time. It starts in the lower brainstem regions, like the medulla and the olfactory bulb, during the prodromal phase. These early stages are marked by nonmotor symptoms, including a decreased sense of smell (hyposmia) and rapid eye movement sleep behavior disorder [17].

Therefore, the progression of Parkinson's disease can be divided into prodromal, early, mild, moderate, and severe (advanced) phases, with each phase marked by a characteristic constellation of symptoms.

1.2.2.1 Prodromal Phase

During this phase, pathology begins in the medulla and olfactory bulb, resulting in non-motor signs that may appear years before any motor deficits. In this phase, patients often experience rapid eye movement (REM) sleep behavior disorder as well as a decrease

or loss of smell, constipation, and other autonomic disturbances such as orthostatic hypotension or urinary dysfunction, orthostatic hypotension, excessive daytime sleepiness, and depression. These early symptoms are subtle and often go unreported unless specifically asked, because they are not Parkinson's disease specific, but when they co-occur, the risk of a subsequent Parkinson's disease diagnosis is greater [17].

1.2.2.2 Early Stage (HY Stage I–1.5)

In the early stage, typically corresponding to Hoehn and Yahr (HY) stages I or 1.5, motor symptoms are minimal and predominantly unilateral. Patients have largely preserved motor function with only slight tremor, rigidity, or bradykinesia, and no significant impairment in daily activities is evident [18]. The subtle motor signs are often detected only through objective clinical examinations, such as low Unified Parkinson's Disease Rating Scale (UPDRS) III scores, reflecting early brain changes with minimal tissue damage [18].

1.2.2.3 Mild Stage (HY Stage II)

Mild-stage Parkinson's disease is characterized by a combination of subtle motor and non-motor symptoms that typically emerge after significant dopaminergic loss but before major functional impairment occurs. The motor symptoms in this early stage predominantly include the cardinal signs of Parkinson's disease, such as bradykinesia, which is the generalized slowness of movement and tremor, and rigidity, a continuous increased resistance to passive movement.

In addition to these motor features, mild-stage Parkinson's disease patients often exhibit early gait disturbances. These include a reduction in step length, slowed pace, dragging one leg ,slightly bent posture while walking, and decreased arm swing [19].

Although these motor impairments can result in functional limitations, patients remain ambulatory and physically independent, indicating that while their motor symptoms are detectable, they do not severely hinder daily activities.

For the non-motor symptoms in this phase, such as changes in mood, sleep disturbances including issues with sleep quality or abnormal sleep patterns, or subtle cognitive deficits, which can be detected with clinical tools like the Mini-Mental State Examination(MMSE) [17, 19].

1.2.2.4 Moderate Stage (HY Stage III)

Motor symptoms in the moderate stage extend beyond simple slowness and stiffness to include tremors, often presenting as a resting or even postural tremor, though freezing of gait (FOG) becomes increasingly problematic in some patients, suggesting an underlying disruption of the central pattern generators controlling locomotion [19].

In addition to these major motor deficits, moderate-stage Parkinson's disease is often accompanied by complications related to dopaminergic treatment. These include motor fluctuations, where patients cycle between "on" periods of relatively improved motor function and "off" periods when the control over symptoms decreases, which may also contribute to the development of dyskinesias-abnormal, involuntary movements typically occurring during peak medication dosage [18].

Non-motor features also become more pronounced during the moderate stage. Although not as disabling as the motor symptoms, patients may experience mild cognitive

impairment that can affect executive functions and attention, alongside psychiatric disturbances such as depression and apathy. Autonomic dysfunction is commonly observed and adds an additional layer of complexity to patient management.Sleep disturbances, particularly those involving disruptions in REM sleep behavior, are reported as well, and may further contribute to overall functional decline [20].

Furthermore, difficulties in activities of daily living (ADLs) become evident. Patients may struggle with tasks such as dressing, handwriting, and personal hygiene, as well as more complex instrumental ADLs, that impact activities of daily living and quality of life [18].

1.2.2.5 Severe Phase (HY Stages IV–V)

In addition to motor symptoms that appear in the moderate phase getting worse, advanced PD patients often develop secondary motor complications. These include dysarthria (speech difficulties) and dysphagia (swallowing disorders), which compound the disability by limiting communication and increasing the risk of aspiration. Dystonias may also appear – sometimes as a complication of long-term levodopa treatment – further contributing to abnormal posturing and impaired motor function,.Postural instability emerges as a defining feature that predisposes patients to falls and related injuries. In tandem with these, gait disturbances become severe: patients demonstrate a shuffling gait, festination (rapid small steps), and freezing episodes that are resistant to dopaminergic medications. Furthermore, abnormal postures such as stooped posture, bent spine syndrome (camptocormia), and lateral flexion (Pisa syndrome) may develop as the disease progresses [18, 21].

In the advanced stages of Parkinson's disease (PD), non-motor symptoms become highly disabling and deeply affect quality of life. Cognitive issues like memory loss, difficulty making decisions, and problems with visual-spatial understanding are common and can progress to dementia. Emotional and mental health symptoms, including depression, anxiety, apathy, hallucinations, and delusions, also worsen significantly. These problems stem from widespread brain chemical imbalances affecting multiple neurotransmitters beyond just dopamine. In addition to sensory symptoms, notably musculoskeletal and neuropathic pain, also afflict patients in this stage, adding to both the physical and psychological distress. Fatigue is common, and the interplay of these nonmotor symptoms with robust motor deficits such as severe freezing of gait increasingly limits daily activities and independence [22].

1.2.3 Cure

There is currently no cure for Parkinson's disease (PD), and people with Parkinson's disease are hoping for the arrival of disease-modifying treatments. Many ongoing researches focus on ameliorating symptoms and the possibility of slowing the PD progression, or delaying its onset in people with a prodromal phase. [3].

1.2.3.1 Early detection

Early detection of neurodegenerative diseases is crucial, as it enables the implementation of effective therapeutic interventions and facilitates the provision of adequate support to patients and their families, thereby slowing the progression of the disease and improving quality of life.

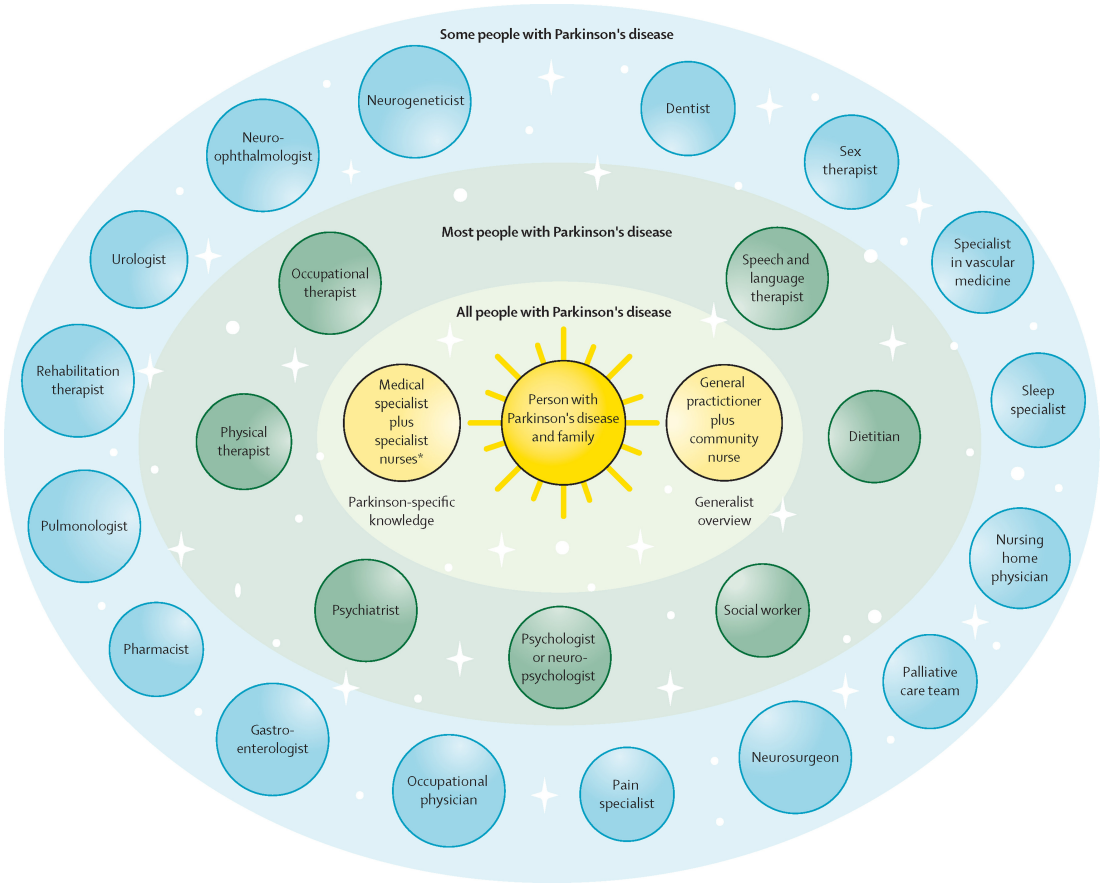


Figure 1.5: Multidisciplinary care of Parkinson's disease[3]

Professional disciplines involved in the multidisciplinary care for people with Parkinson's disease. There are no individual stars within the multidisciplinary team, but the person with Parkinson's disease can be seen as the sun around which the various professionals revolve to deliver their support. Some professionals are always involved in the care of people with Parkinson's disease, which includes the medical specialist (neurologist or geriatrician, depending on the specific setting) supported by a specialist nurse, and the family practitioner who, being a generalist, oversees issues such as comorbidity and polypharmacy. Other professionals are involved in the care for most people with Parkinson's disease, whereas some are involved with only a smaller group of individuals. This model sketches an ideal situation in which each person with Parkinson's disease has access to each of these disciplines, which unfortunately is not the case in most places in the world. *Nurses who care for people with Parkinson's disease. [3]

1.2.4 Artificial intelligence in Parkinson's disease

Artificial Neural Networks or ANN, are the most popular AI technique and are widely used in medicine.In recent years, many new forms of AI have been developed in the form of Machine Learning Algorithms that can be used to help diagnose and treat diseases,As a result of this technology, AI in the form of Artificial Neural Networks has been developed to innovate new treatment methods for certain neurodegenerative diseases such as Parkinson's., and Alzheimer's disease and related disorders. An example of these groundbreaking methods includes a study in the journal Nature Medicine, "Artificial Intelligence-Enabled Disease Using Nocturnal Breathing Signals"[23]. This study developed an AI neural network that could accurately and reliably identify and diagnose those who have PD in their sleep due to their nocturnal breathing patterns. This AI system was able to assess the severity of people's PD and was also able to assess the progression of Parkinson's in the patient's life over time, as shown in (Fig 1.6) [1]

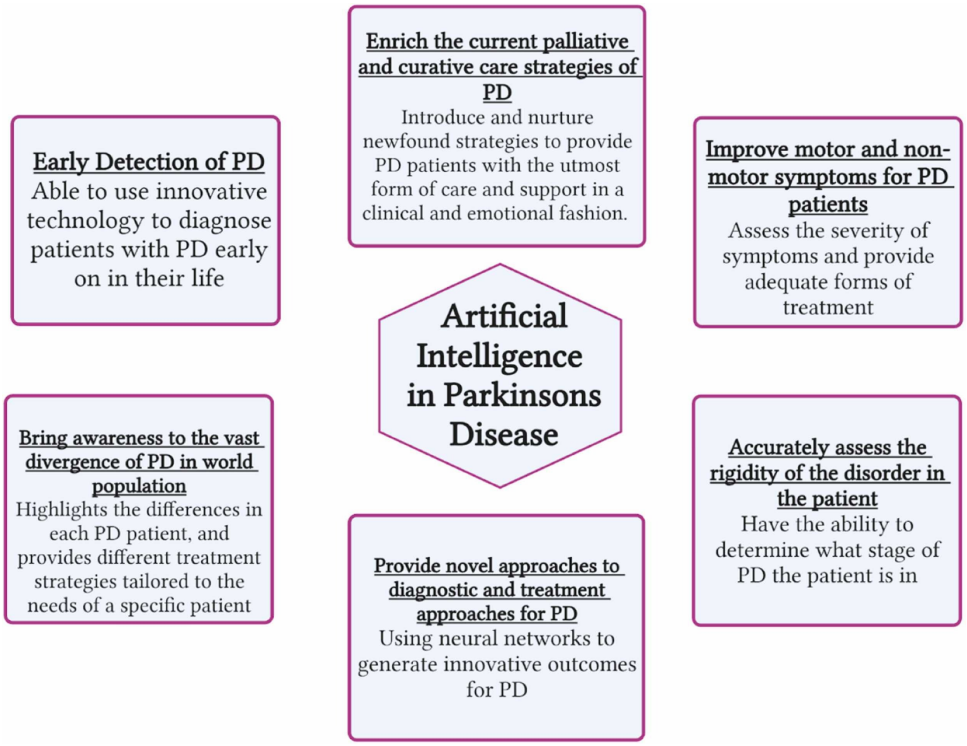


Figure 1.6: AI in PD : from early diagnosis to personalized care [1]

Figure1.6shows the multitude of benefits surrounding the implementation of Artificial Intelligence in Parkinson's Disease. There are six main reasons why Artificial Intelligence should be used in the diagnosis and treatment of PD. Some reasons include decreasing the severity of certain motor and non-motor symptoms of PD through extensive neuroimaging, accurately assessing the rigidity of the disorder in the patient through accurate and accessible diagnostic tests, and ensuring the early detection of PD to provide viable treatment options for the patient, improve their quality of life, and move one step closer towards finding a cure for PD. [1]

1.3 Conclusion

Parkinson's disease is known to be one of the most common and complicated neurodegenerative diseases. The primary symptoms of PD are associated with damage to motor and non-motor systems, along with the progression of the disease and the stages it goes through. The condition stems from the loss of dopamine-producing neurons located in critical areas of the brain such as the substantia nigra and the locus coeruleus. This results in the hallmark movement related impairments along with numerous systemic bodily disorders. The damage proceeds through a preclinical phase, which eventually evolves into more severe stages of the disease that place a huge burden on patients and impair quality of life while greatly diminishing independence. Even though a definitive cure isn't available today, current therapeutic approaches look into symptom alleviation, improving daily activities, along with other research focused on changing the course of the disease. Further progress in the early detection Parkinson's key indicators, neuroprotective strategies and biomarkers are crucial in reshaping modern approaches for managing the disease and improving the conditions for those afflicted with it.

Chapter 2

Deep Learning

2.1 Introduction

Artificial intelligence (AI) is a branch of computer science that focuses on creating systems or machines that can perform tasks typically requiring human intelligence. Such tasks include learning, reasoning, problem solving, perception, understanding language, and making decisions. A key subset of AI is machine learning (ML), which enables systems to learn from data and improve their performance over time without being explicitly programmed. A more advanced subset of ML is deep learning (DL), which uses artificial neural networks to model complex patterns in large datasets. This makes it especially powerful for tasks such as image and speech recognition.

2.2 Machine Learning

According to Ng (2017), "Machine learning is the science of getting computers to learn without being explicitly programmed."[24] and as mentioned in Designing Machine Learning Systems by Huyen (2022), "Machine learning is an approach to learn complex patterns from existing data and use these patterns to make predictions on unseen data". [25]

2.2.1 ML Architectures

2.2.1.1 SVM

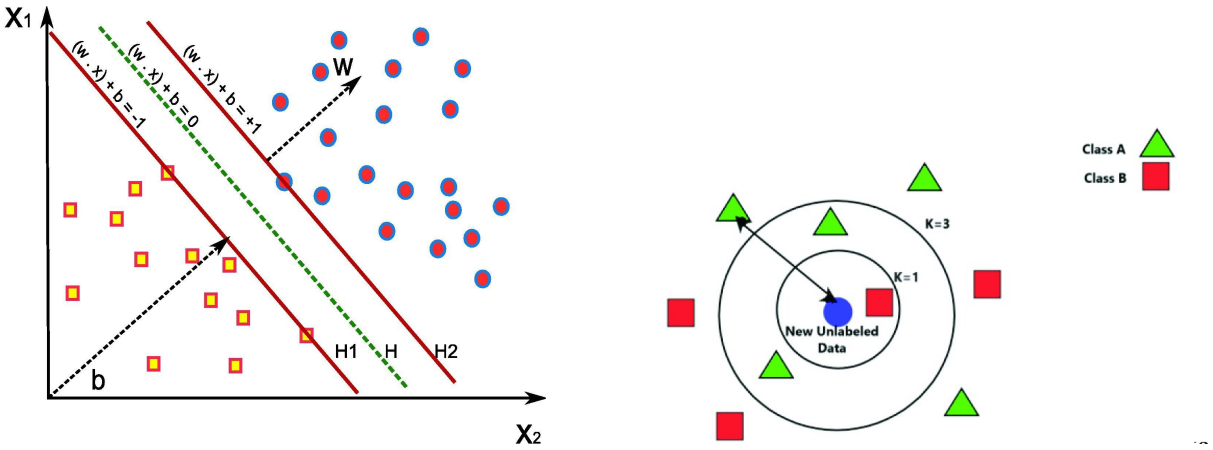
Support Vector Machine (SVM) aims to separate several classes with a surface that maximizes the margin between them, allowing one to maximize the generalization ability of a model. [26]

2.2.1.2 KNN

K-Nearest Neighbors (KNN) is one of the supervised classification techniques that is easy to implement and efficient in yielding positive results. In classifying new, unlabeled data, KNN relies on a labeled training dataset where it finds the closest K labeled examples (neighbors) and then assigns a label based on the majority vote of those neighbors, often employing Euclidean distance. Among all existing algorithms for classification, KNN has found broad acceptance because there are very few prerequisites for its application and it does not require high computational resources; however, its performance heavily depends on the value of K selected.[27]

2.3 Data Types

- **Sequential Data:** Sequential data is any kind of data where the order matters, i.e., a set of sequences. Examples include text streams, audio fragments, video clips, and time-series data.
- **Image or 2D Data:** A digital image is made up of a matrix, which is a rectangular array of numbers, symbols, or expressions arranged in rows and columns. Matrix, pixels, voxels, and bit depth are the four essential characteristics of a digital image.



(a) SVM [26] (b) KNN [27]

Figure 2.1: SVM and KNN models

- **Tabular Data:** A tabular dataset consists primarily of rows and columns. Each column (field) must have a name and may only contain data of the defined type.

Deep learning models can learn efficiently on tabular data and allow us to build data-driven intelligent systems.[28]

2.4 Deep Learning

Deep learning is a subset of machine learning that is based on artificial neural networks inspired by the structure and function of the human brain. DL can be considered as an AI technique that mimics the brain's information processing mechanism.[28]

2.4.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of artificial neural networks widely used in various applications such as image and video recognition, image classification, medical image analysis, and speech recognition.[29]

The popularity and wide range of application domains of deep CNNs can be attributed to the following advantages:

1. CNNs fuse the feature extraction and feature classification processes into a single learning body.
2. CNNs can process large inputs with great computational efficiency compared to conventional fully-connected Multi-Layer Perceptrons (MLP).
3. CNNs are immune to small transformations in the input data, including translation, scaling, skewing, and distortion.
4. CNNs can adapt to different input sizes.

[5]

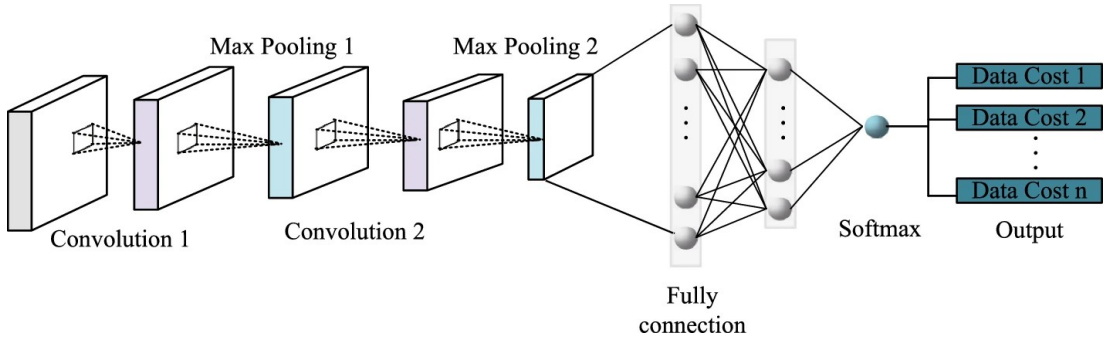


Figure 2.2: Structure of CNN (Suppose this is an n-classification problem. [4]

2.4.1.1 CNN Layers

Convolutional Layers The main and critical layer in CNNs is the convolution layer, which extracts various features from different local regions of the input data without the need for manual definition. Stride and Padding are two parameters that affect the convolution procedure and the output size.[4]

Stride: Stride can be considered as the step of the kernel in images. It is the number of rows and columns that the convolution kernel slides over the input matrix.[4]

Padding: Padding changes the input data size by adding a certain number of pixels to the edges of the input data so that the size of the output data can match the input data.[4]

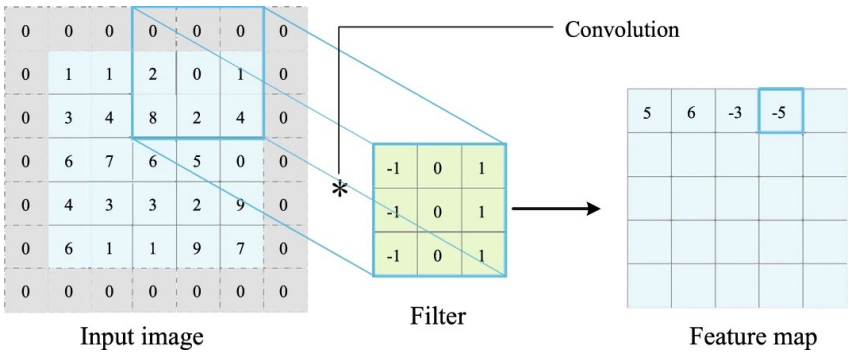
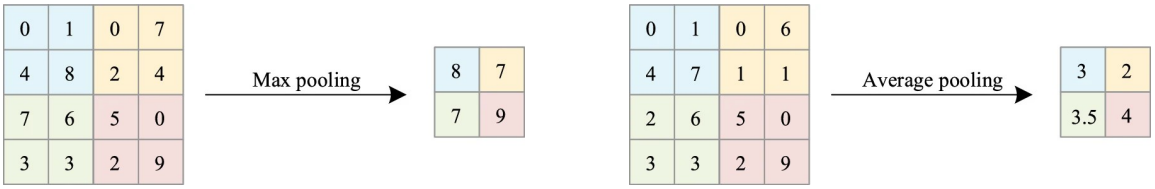


Figure 2.3: Convolution procedure with padding[4]

Pooling Layer The pooling layers are typically placed between consecutive convolution layers to compress the amount of data and parameters, reduce the dimension of the feature map, and minimize overfitting. The most commonly used pooling functions are max pooling and average pooling.[4]

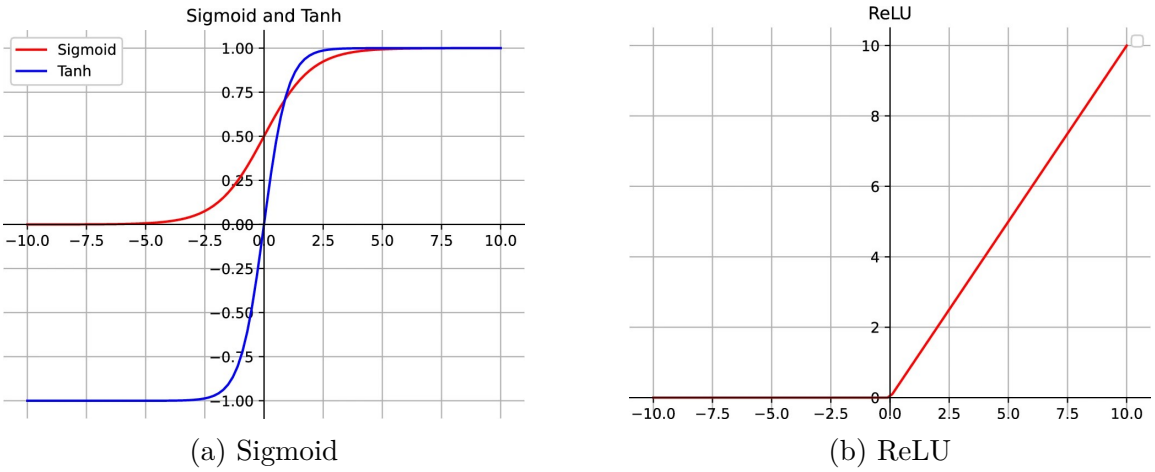


(a) Max pooling (b) Average pooling

Figure 2.4: (a) Max pooling and (b) Average pooling operations.[4]

Activation Functions

Activation functions allow the network to learn nonlinear mappings. The most classical and widely used activation functions are Sigmoid, Tanh, Softmax, ReLU, and Leaky ReLU.[4]



(a) Sigmoid (b) ReLU

(c) Softmax

Figure 2.5: Some of the most used activation functions (a) Sigmoid, (b) ReLU, and (c) Softmax[4]

Batch Normalization Batch normalization (BN) is a regularization technique that can speed up the neural network's training process, improve the network's adaptability, and network's generalization by unifying the distribution of feature-map values by setting them to zero mean and unit variance[4]

2.4.1.2 Fully connected Layers

Fully connected layers (FC) or dense layers are typically employed at the network's conclusion for classification. Once the feature mapping is obtained after several convolu-

tion and pooling operations, these feature mappings will be sent to the fully connected layer as a long vector, followed by the output layer, for classification. [4]

Dropout

Dropout facilitates regularization in the network by randomly disabling some neurons with a specified dropout rate probability, which eventually enhances generalization.[4]

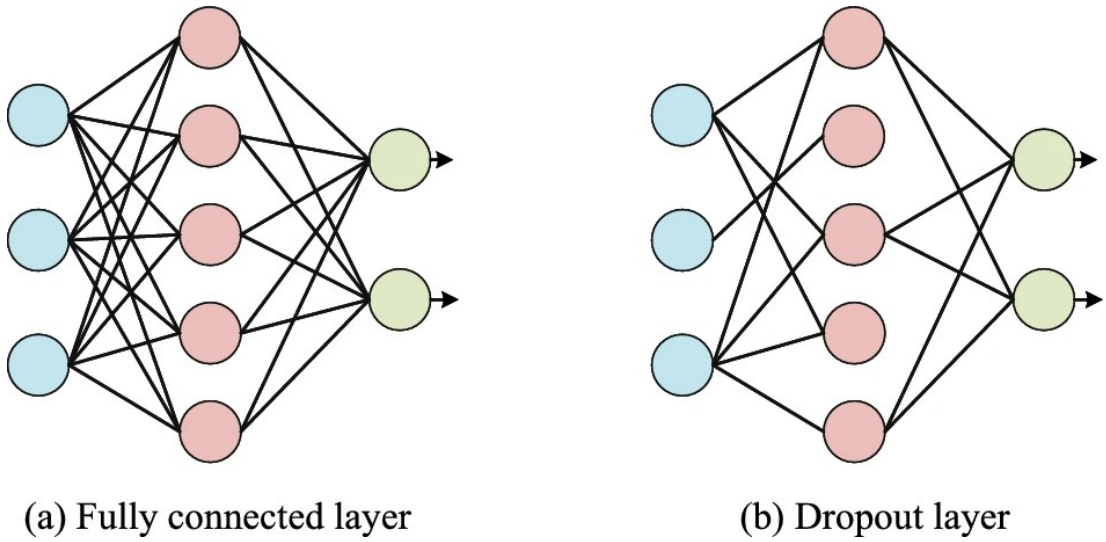


Figure 2.6: Distinction between a fully connected layer and a dropout layer)[4]

2.4.2 2D-Convolution vs 1D-Convolution

The main difference between 1D-conv and 2D-conv is in the type of input data and the features they extract.[5]

- **1D-Convolution:** Designed to handle sequential data, such as time-series and speech.
- **2D-Convolution:** Suited to data arranged into rows and columns, like images.

2.4.3 CNN Architectures

2.4.3.1 VGG Network

The Visual Geometry Group (VGG) model, developed in 2014 by Oxford University, brought forth the potency of stacking simple 3x3 convolutional layers.[6]

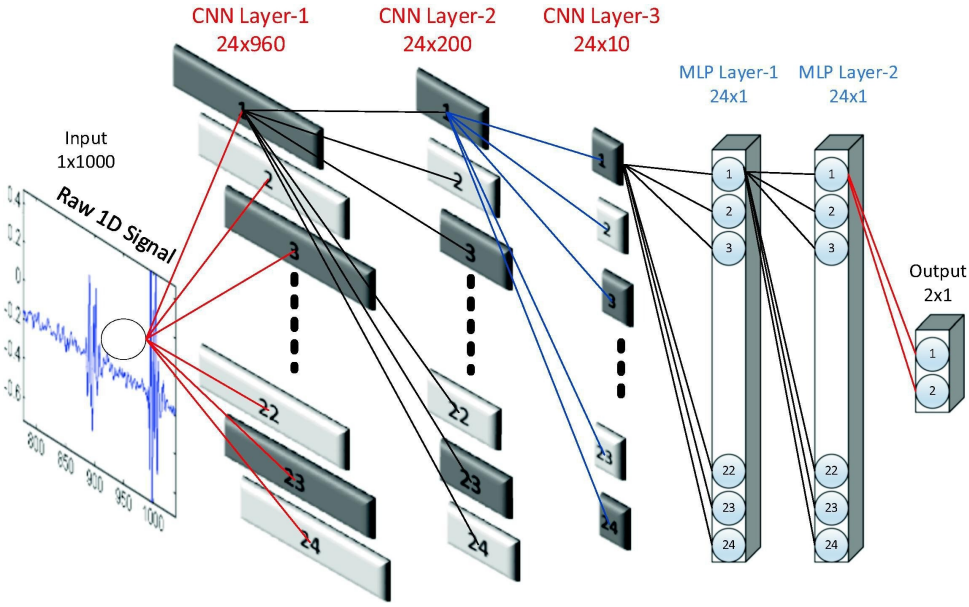


Figure 2.7: A sample 1D CNN configuration with 3 CNN and 2 MLP (FC) layers.)[5]

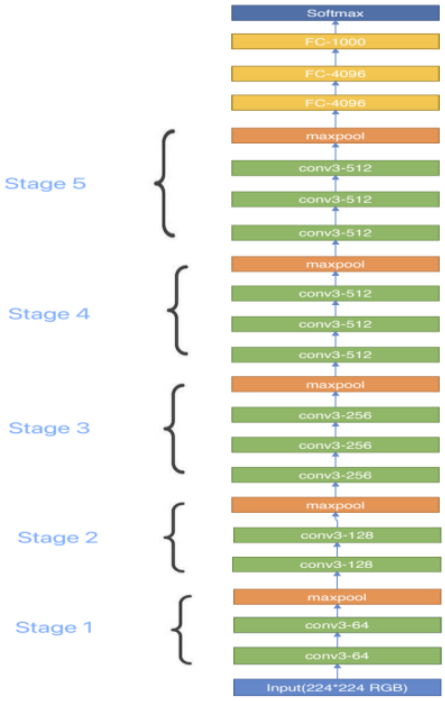
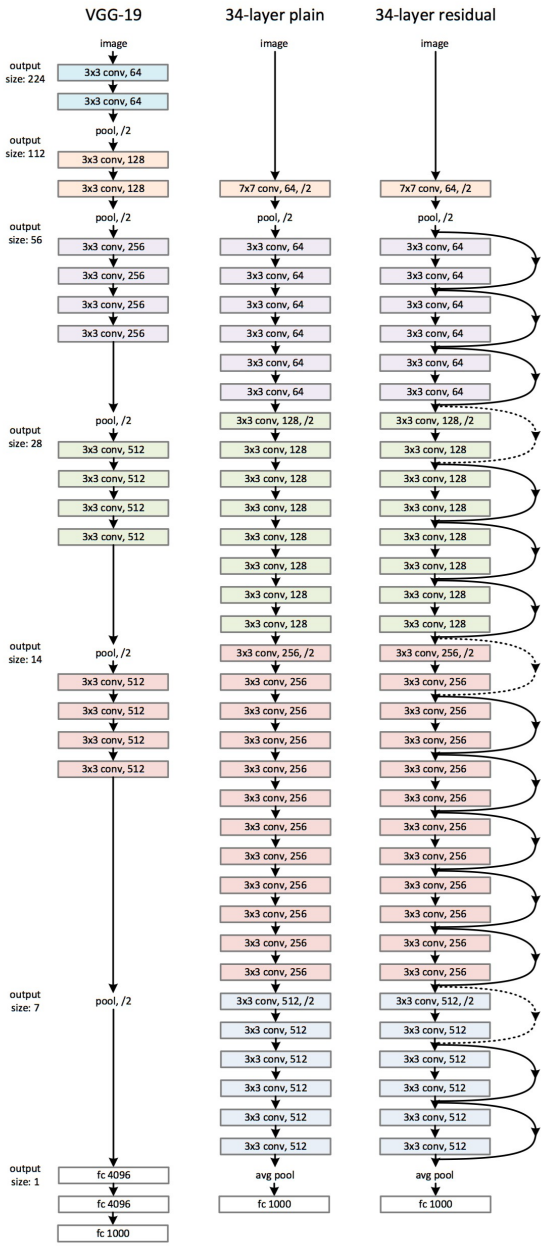


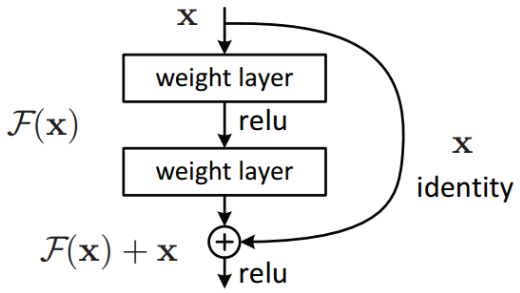
Figure 2.8: Structure of VGG. [6]

2.4.3.2 ResNet

The Residual Network (ResNet) is a network structure based on the VGG19 network with the insertion of residual shortcuts, which ensures that the weights learned from the previous layers do not vanish during backpropagation.[7]



(a) network architectures for ImageNet.Left: the VGG-19 model, Middle: a plain network with 34 parameter layers, and Right: a residual network with 34 parameter layers.

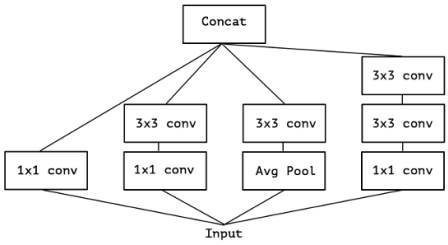


(b) Residual learning: a building block.

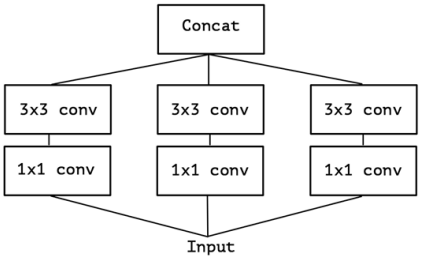
Figure 2.9: (a)comparison between vgg19,plain and resedual network and (b) Residual block [7]

2.4.3.3 Xception

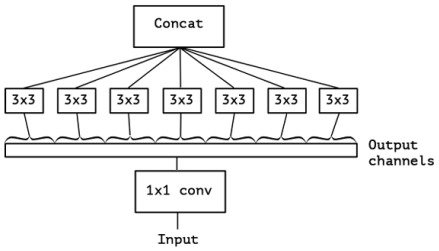
The Xception network, which stands for "Extreme Inception," is inspired by the InceptionV3 model with the replacement of the inception module with depthwise separable convolution.[8]



(a) A canonical Inception module (Inception V3).



(b) A simplified Inception module.



(c) An "extreme" version of our Inception module, with one spatial convolution per output channel of the 1x1 convolution.

Figure 2.10: Comparing xception module (c) to inception v3(a) and simplified inception module(b). [8]

2.4.4 Transformers

Transformers marked a revolutionary step by introducing the attention mechanism.[30] It is a prominent deep learning model widely adopted in various fields, such as natural language processing (NLP), computer vision (CV), and speech processing.[31]

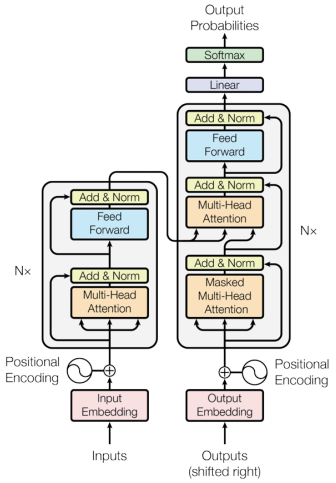


Figure 2.11: The Transformer- model architecture[9]

2.4.4.1 LSTM

Long Short Term Memory (LSTM)[9], an extension of recurrent neural networks (RNN), was introduced to manage situations in which RNNs fail and to solve the vanishing gradient problem.[32]

2.4.5 Transfer Learning

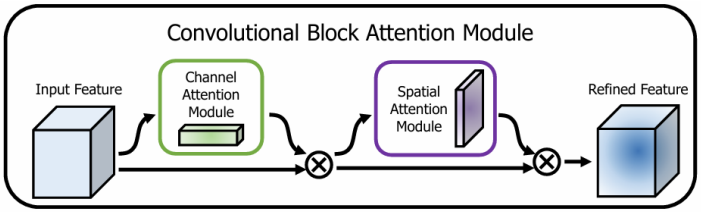
Transfer learning refers to the family of methods where a model developed for a task is reused as the starting point for a model on a second task.[25]

2.4.6 Fine-Tuning

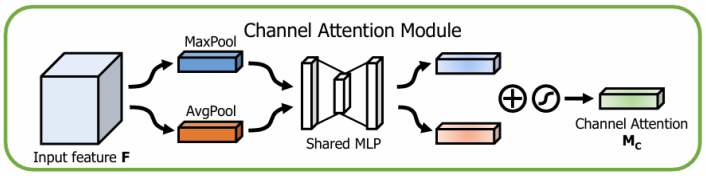
Fine-tuning means making small changes to the base model, such as continuing to train the base model or a part of the base model on data from a given downstream task.[25]

2.4.7 Convolutional Block Attention Module

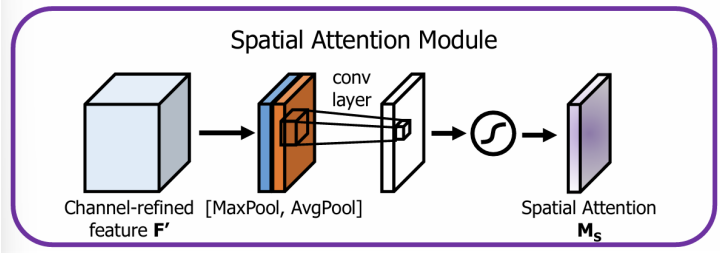
Convolutional Block Attention Module(CBAM)(Fig. 2.12a) is a simple and effective attention module for feed-forward CNNs. CBAM operates by sequentially generating attention maps along two distinct dimensions, channel (Fig. 2.12b) and spatial (Fig. 2.12c). .[10]



(a) CBAM Block



(b) Channel Attention Module



(c) Spatial Attention Module

Figure 2.12: CBAM: Channel and Spatial Attention Modules[10]

2.4.8 Squeeze-and-Excitation

The Squeeze-and-Excitation" (SE) block is a specialized architectural unit designed to improve the representational power of a network by enabling it to perform dynamic channel-wise feature recalibration. This mechanism improves the network's ability to distinguish useful features from less important ones. [11]

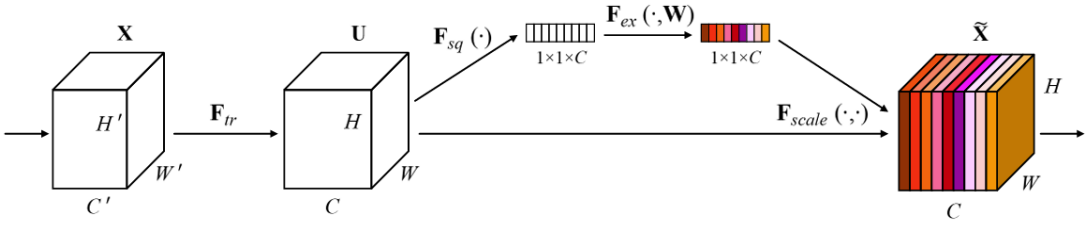


Figure 2.13: A Squeeze-and-Excitation block [11]

2.5 DL vs ML in Parkinson's Disease detection

As mentioned in the previous chapter and on [12] there are many of criteria for diagnosing PD, such as brain scans or Parkinson's Disease Rating Scale (MDS-UPDRS). However, these criteria often come with challenges such as cost, accessibility, clinician bias, and difficulty monitoring progression and treatment effectiveness. One of the initial

and most serious signs of Parkinson's disease is speech difficulties, while being objective, cost-effective, and accessible, making them an alternative diagnostic approach for neurological examination methods.[12]

Most of Machine learning approches that used speach data involves hand-crafting acoustic features, including certain variants of the jitter, shimmer, and harmonic-to-noise ratio that are indicative of PD speech impairments with Machine learning models such as support vector machines (SVM), random forests (RF), k-nearest neighbors (KNN), and regression trees (RT).[12]

The other approach focuses on using deep learning to automatically learn features directly from speech data. Several neural network architectures have been designed and tested, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNN) like Long Short-Term Memory Networks (LSTMs), a combination of them, and more recently, transformer-based models.[12]

Author	**Model**	**Accuracy** [%]
Aversano et al. [33]	LSTM	97.1
Klempíř et al. [34]	Wav2Vec	95.0
Hireš et al. [35]	Xception	97.8
Toye et al. [36]	SVM	98.9[1]
Malekroodi et al. [12]	Swin_s	98.5 ± 2.50
Malekroodi et al. [12]	VGG16	98.1 ± 3.23

[1] Using hand-crafted features.

Table 2.1: Accuracy comparison with previous works that utilized the Italian-speaking Parkinson's dataset.

2.6 Conclusion

In this chapter we explored several Machine Learning and Deep Learning methods with a focus on the CNN architectures. We also mentioned the application of these techniques in PD detection that showed that DL approaches used automatic feature extraction, as shown in (Table 2.1), the VGG16 (CNN Model) has better results compared to most of the models and lower model complexity than transformer-based models. For this reason, we will focus on CNN architectures in our implementation in the next chapter.

Chapter 3

Contribution and Implementation

3.1 Introduction

In the previous chapters, we provided an overview of Parkinson's disease, highlighting its clinical characteristics, diagnostic challenges, and the significance of early detection through voice and speech analysis. We also introduced the core concepts of deep learning, with a particular emphasis on Convolutional Neural Networks (CNNs), attention mechanisms, and the role of feature extraction in audio processing. In this chapter, we present our main contributions and findings, describing the datasets utilized, the preprocessing steps, the architectures and techniques implemented, as well as a comprehensive evaluation of our models for the classification of Parkinson's disease from voice recordings.

3.2 Tools used

3.2.0.1 Work Environments

- **Remote access to workstation:**

  1. **Hardware Specifications:**

     - CPU: Intel(R) Core(TM) i9-10940X CPU @ 3.30GHz
     - RAM: 251 GiB
     - Disk used: SSDé 1.8T
     - Gpu RTX3090 24G

  2. **Software Specifications:**

     - Operating system: Ubuntu 22.04
     - Python Libraries/Frameworks: PyTorch, NumPy, pandas, scikit-learn, Matplotlib, Seaborn, OpenCV, Librosa, Parselmouth, tqdm

Software and Libraries	Description
PyTorch	Deep learning framework developed by Facebook, offering dynamic computation graphs and flexible APIs. Widely used in research and production for training neural networks.
NumPy	Core library for numerical computing in Python, providing support for large multidimensional arrays, matrices, and a collection of mathematical functions.
pandas	Data analysis and manipulation library for Python, offering data structures like DataFrames for handling structured data.
scikit-learn	Machine learning library in Python that provides simple and efficient tools for data mining and analysis, supporting both supervised and unsupervised learning.
Matplotlib	2D plotting library for Python that allows for the creation of static, animated, and interactive visualizations.
Seaborn	Statistical data visualization library based on matplotlib, providing a high-level interface for drawing attractive and informative graphics.
OpenCV	Open-source computer vision and image processing library with tools for real-time image and video analysis.
Librosa	Python package for music and audio analysis, providing tools for feature extraction, signal processing, and visualization of audio data.
Parselmouth	Python interface to the Praat software, enabling phonetic analysis and manipulation of speech and voice signals.
tqdm	Lightweight Python library for displaying smart progress bars in loops and command-line interfaces. Useful for tracking the progress of long-running operations.
VS Code (Visual Studio Code)	Lightweight but powerful source code editor developed by Microsoft, offering built-in debugging, Git support, syntax highlighting, and extensions for Python and data science workflows.

Table 3.1: Common Python Libraries for Machine Learning, Audio Analysis, and Data Science

3.3 Dataset

The dataset used in our work is the Italian Parkinson's voice and speech database. It comprises speech recordings in .wav format from Italian individuals diagnosed with Parkin-

son's disease and healthy control subjects. This database was collected through the efforts of Dimauro et al., as referenced in [37, 38], and as implemented and mentioned in [12]. It has been shown that vowels are more predictive of Parkinson's diagnosis compared to words or sentences. We follow the same approach and use vowel recordings only (/a/, /e/, /i/, /o/, and /u/).

As shown in Table 3.2, we have recordings from 22 healthy controls (12 female, 10 male) and 28 PD patients, who were classified based on their score on Part III of the MDS-UPDRS.

Table 3.2: Demographic information, including gender and age ranges of the dataset [12].

Class	**MDS-UPDRS III**	**Subjects**		**Age**	
		Male	Female	Male	Female
Healthy	∼	10	12	60–72	60–77
PD_Mild	1–10	7	3	50–77	40–63
PD_Severe	11–24	12	6	65–75	54–80

3.3.1 Preprocessing

In this study, we used this dataset in various formats and implemented a range of preprocessing methods. All recordings were resampled to 16 kHz. Recordings with excessive background noise were removed (2 healthy participants excluded). We trimmed leading silence and segmented the recordings into fixed-length overlapping segments (50% overlap). Two dataset versions were created per segment length:

First Segments (FS): Only the first segment of each recording.

All Segments (AS): All segments from each recording.

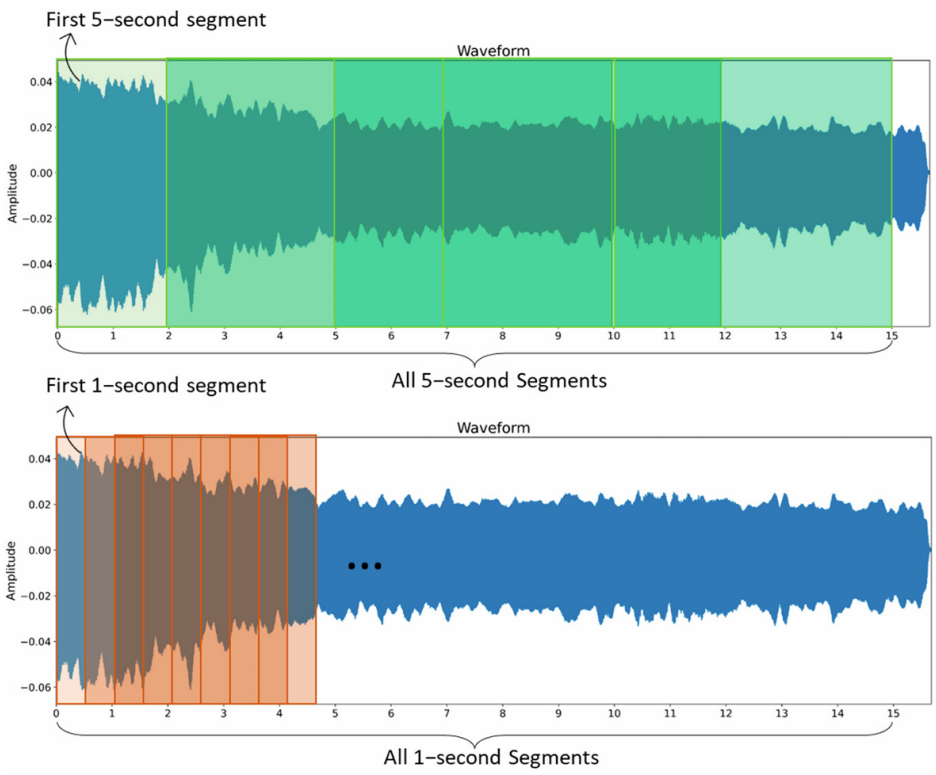


Figure 3.1: Overview of the process used to construct distinct datasets from the original dataset [12].

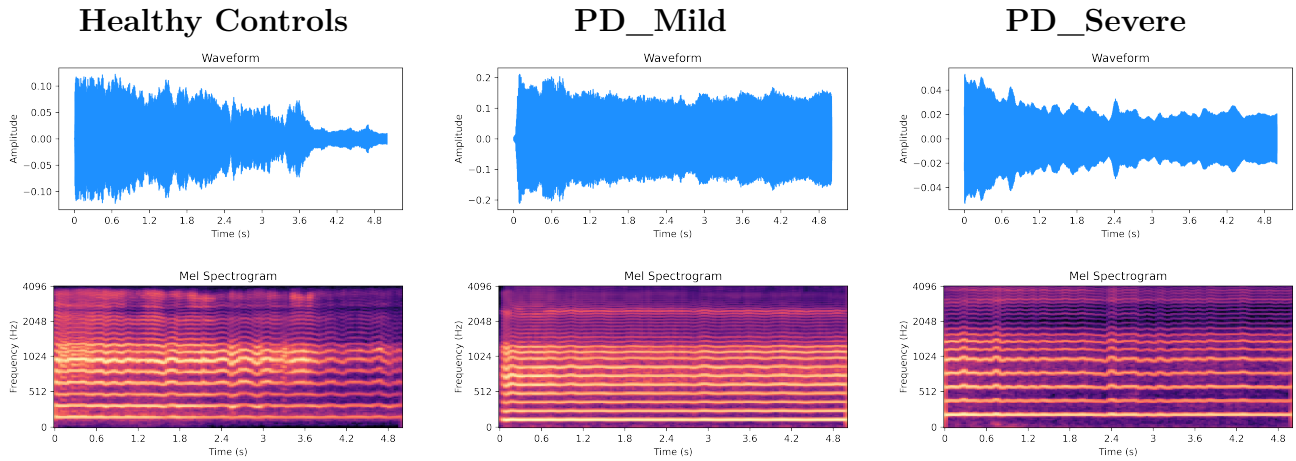
3.3.1.1 2D Images

We adopted the approach from [12], where voice recordings are transformed into LMS-based images. LMS is obtained by computing a spectrogram using STFT, then converting the frequency axis to the mel scale.

STFT parameters:

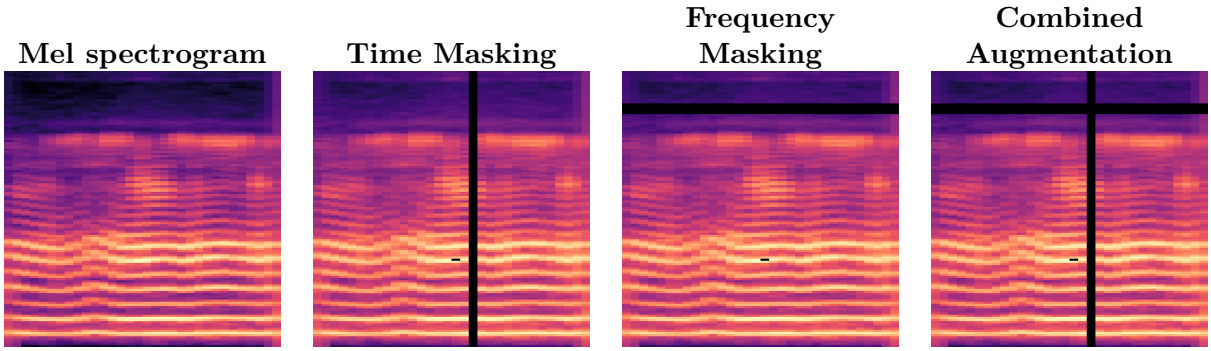
- Window length: 128 ms (2048 samples)
- Hop length: 32 ms (512 samples)

To improve generalizability and reduce overfitting, we applied audio augmentation before the LMS transformation. Augmentations included time masking, frequency masking, and a combination of both as shown in Fig 3.3.



(a) (b) (c)

Figure 3.2: Speech sound examples. The upper panel shows the waveform; the lower panel shows the log mel spectrogram (128 mel bands).



(a) (b) (c) (d)

Figure 3.3: The effects of data augmentations on LMSs. (a) Original, (b) Time masking, (c) Frequency masking, (d) Combined.

3.3.1.2 Tabular Data

The tabular data used are the features extracted from voice recordings.

Acoustic Features: Jitter (absolute, relative, rap, PPQ5), Shimmer (dB, relative, APQ3, APQ5), Fundamental Frequency, Harmonics-to-Noise Ratio (HNR), and Pitch.

MFCC Features: For each segment, a matrix of MFCC values (frames $\times$ 13) was computed. We averaged across frames to obtain a 13-dimensional vector.

As presented in [36], features were normalized: Min-Max for acoustic, StandardScaler for MFCC.

3.3.1.3 Raw Audio

The raw audio followed the same preprocessing: resampling, silence trimming, and segmentation.

3.3.2 Data Support

3.3.2.1 Raw Audio

The raw audio data has been used for training without any data augmentation mentioned before; the class distribution among each dataset is shown in table 3.3 and table 3.4

	1AS	**2AS**	**3AS**	**4AS**	**5AS**	**6AS**	**7AS**
healthy	3736	1715	1045	718	508	319	279
mild	2738	1294	815	578	425	334	277
severe	3076	1396	837	557	419	347	295
total	9550	4405	2697	1853	1352	1000	851

Table 3.3: Sample distribution across different severity levels and anatomical segments (AS).

	FS
healthy	200
mild	100
severe	175
total	475

Table 3.4: Sample distribution for FS category.

3.3.2.2 Augmented Data

The augmentation mentioned earlier was applied to each audio segment before generating the spectrograms (as in Figure 3.3). The distribution of each class for each dataset is shown in Table 3.5

	1AS	**1FS**	**5AS**	**5FS**
healthy	13852	800	1872	800
mild	10128	400	1572	400
severe	11964	700	1604	700
total	35944	1900	5048	1900

Table 3.5: Sample counts across conditions and categories.

3.4 Training Process

The training process used in this study is the same as the one proposed in [12], where the data was split into three folds with no patient overlap across folds to avoid any data leakage. Each model was trained in two folds and evaluated in the remaining fold, and this was repeated three times so that each fold served as an evaluation set once.For the training parameters we used the ones proposed in [12](Table **??**) since it gave good results

Parameter	**Values**
Image size	224×224 pixels
# Epochs	100
# Batch-size	64
Initial Learning Rate	3×10^{-4}
Optimizer	AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$, Weight decay $= 0.01$)
Loss	Cross entropy

Table 3.6: Training Parameters

To resolve the problem of class imbalance in our datasets, we attempted applying class weighting during training. This approach shifts the loss function for infrequent classes by scaling down the contribution from frequent ones, and conversely. This reduces bias towards dominant categories. We applied class weights to improve fairness and performance for minority classes, while also increasing the model's overall generalisation.

3.5 Approaches and Techniques

We have tested several architectures with different data types during this study. All the models that will be mentioned later share the same classifier, where the original classification layers were removed and replaced with two dense layers before the final classification layer (Output layer). The first dense layer has 256 neurons, and the second one has 128 neurons. After each dense layer, a dropout with a probability of 0.5 was applied. This classifier architecture was proposed in [12].

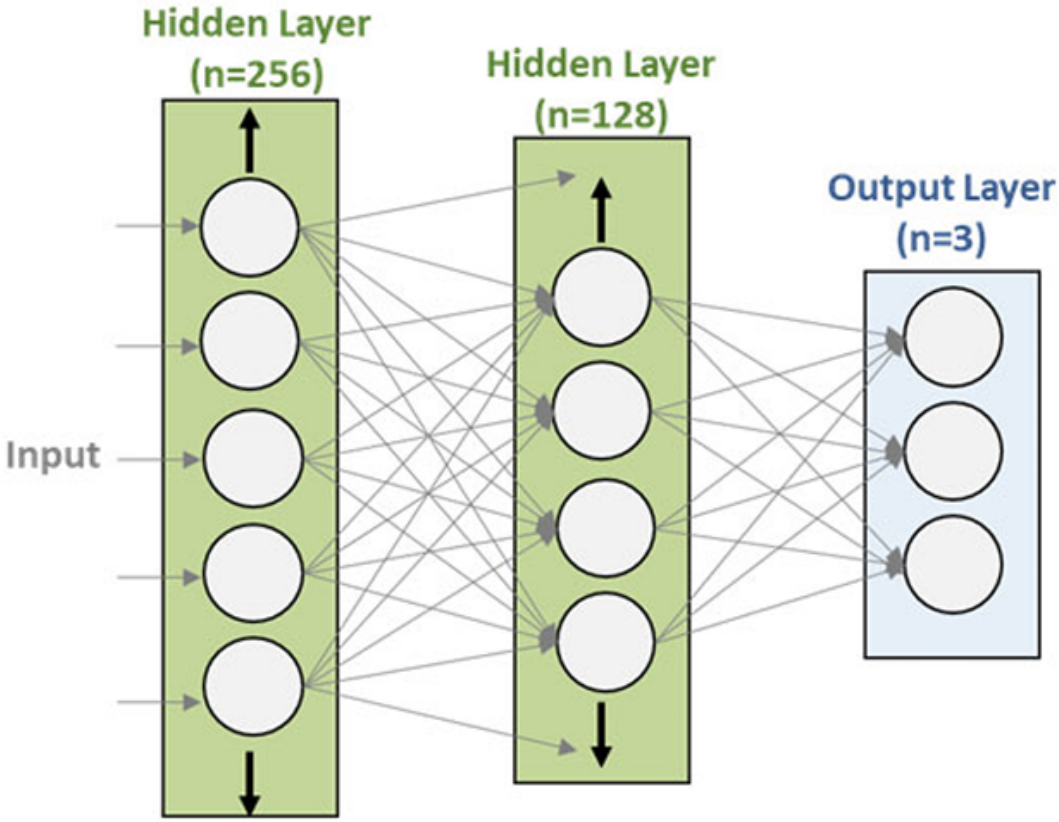


Figure 3.4: Classifier architecture used in this study. [12]

3.5.1 VGG16+CBAM

As mentioned in the previous chapter(Table 2.1 VGG16 did well on the previous study[12] while keeping a simpler architecture compared to other models like the Swin_s model. We added attention blocks (CBAM) to the original model architecture as shown in Fig 3.5, We inserted CBAM blocks after each MaxPooling layer. This idea was inspired from [39], where they used CBAM with a Resnet model by adding CBAM blocks for each residual block.For the base VGG16 model, we used the pretrained one on the Imagenet dataset, and we fine-tuned our classifier layers.

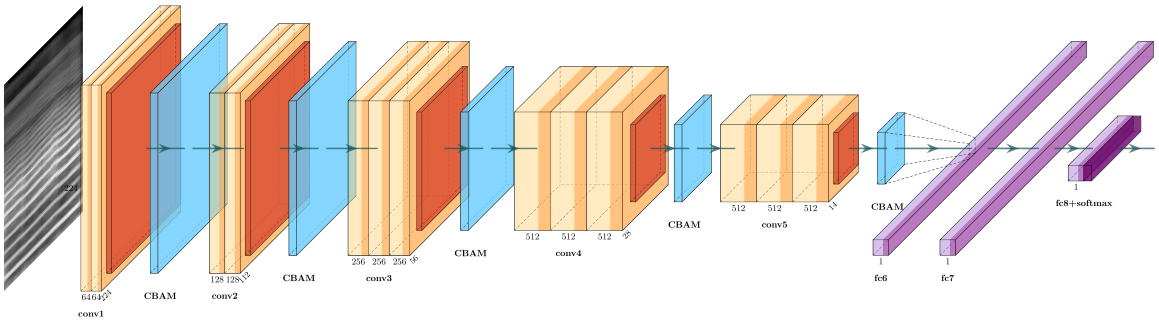


Figure 3.5: VGG16 with CBAM blocks

3.5.2 ResNet18+CBAM

Since ResNet18 has a simple architecture and was been tested on [12]. We tested the idea proposed on [39] by adding a CBAM block to ResNet18 base blocks. Channel attention is applied post the last convolution (before adding the shortcut), then Spatial attention is applied next, at the end, the attention-augmented result is then added to the shortcut connection[39] as shown in Figure3.6. We used the pre-trained ResNet18 on the ImageNet dataset as a base model, then we applied fine-tuning.

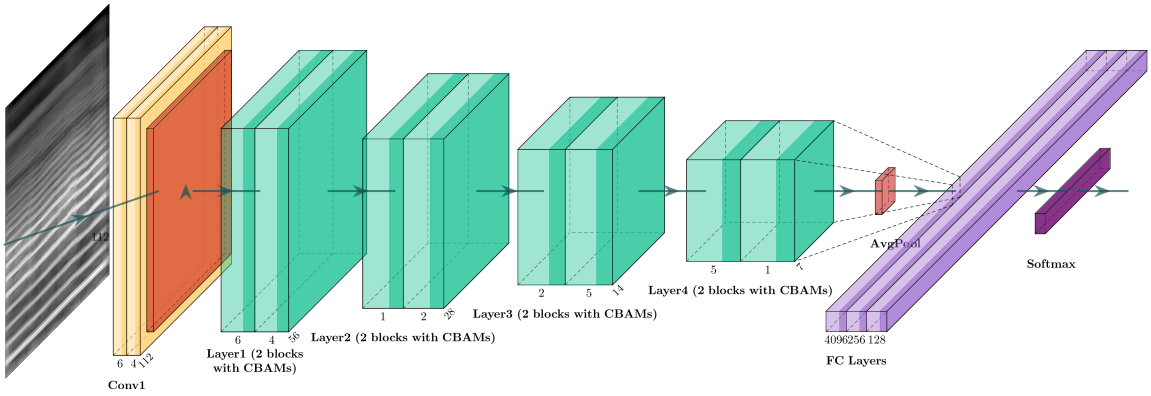


Figure 3.6: ResNet18 with CBAM blocks

3.5.3 1D CNN

For the raw audio data, we first attempted to perform classification using a simple 1D-CNN model.The model architecture was proposed on [40], the architecture proposed by the author is shown on Table3.7

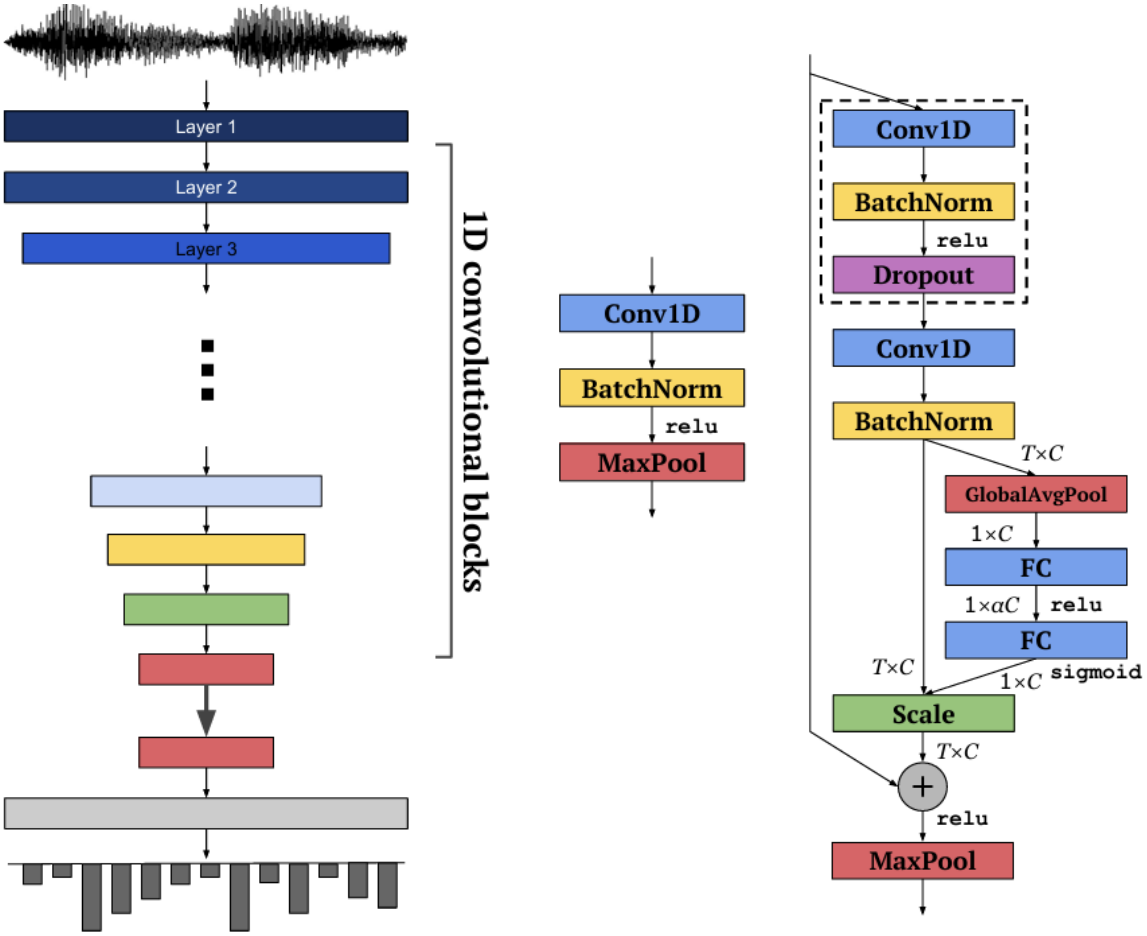
layer	stride	output	# of params
conv 3-128	3	19683×128	512
conv 3-128	1	19683×128	49280
maxpool 3	3	6561×128	
conv 3-128	1	6561×128	49280
maxpool 3	3	2187×128	
conv 3-256	1	2187×256	98560
maxpool 3	3	729×256	
conv 3-256	1	729×256	196864
maxpool 3	3	243×256	
conv 3-256	1	243×256	196864
maxpool 3	3	81×256	
conv 3-256	1	81×256	196864
maxpool 3	3	27×256	
conv 3-256	1	27×256	196864
maxpool 3	3	9×256	
conv 3-256	1	9×256	196864
maxpool 3	3	3×256	
conv 3-512	1	3×512	393728
maxpool 3	3	1×512	
conv 1-512	1	1×512	262656
dropout 0.5	–	1×512	
sigmoid	–	50	25650
		Total params	1.9×10^6

Table 3.7: 3^9 model, 19683 frames
59049 samples (2678 ms) as input
[40]

We took the same base architecture and replaced the classification layer with our classifier (figure 3.4) to adapt it to our task.

3.5.4 ReSE-2-Multi

ReSE-2-Multi is a term proposed in [13] that refers to an enhanced SampleCNN where they added residual connections, squeeze-and-excitation modules and multi-level feature concatenation. in our study, we used 8 of the ReSE blocks(Figure 3.7b) and multi-level feature concatenation of the last 3 blocks, the output of the levels was concatenated after computing the global max pooling for each, the vector we got was passed to our classifier



(a) ReSE model (b) ReSE block

Figure 3.7: ReSE-2-Multi Architecture[13]

3.5.5 ReSE-2-Multi + Extracted Features

We kept the same architecture explained in the previous title 3.5.4, and added a small neural network to extract patterns from audio features extracted before **??**. As shown in (Figure3.8), the neural network is composed of 2 hidden layers of 128 neurons, one hidden layer of 256 neurons, and the last layer of 23 neurons. The output of this neural network is concatenated with ReSE-2-Multi convolutional layers, then passed to our classifier

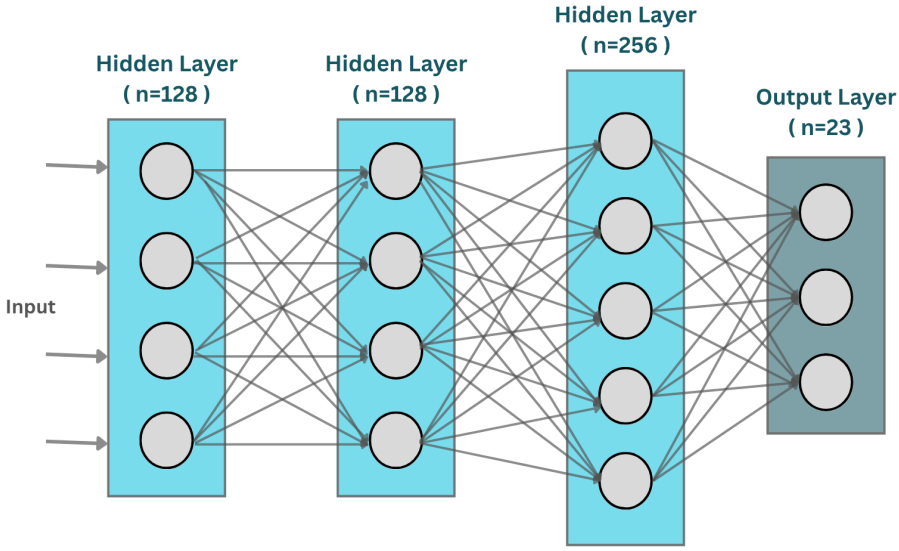


Figure 3.8: Neural network to process extracted features

3.5.6 ReSE-2-Multi (Frozen) + Extracted Features

We used the same model explained in the previous title 3.5.5, and we froze the ReSE-2-Multi convolutional layers and used weights from trained ReSE3.5.4 while training the rest of the layers (Features block + last classifier)

3.6 Results

the values presented in table 3.8 and table 3.9 represent the average of values from each validation fold and the standard deviation.The metrics used to evaluate are precision, recall, and F1 score per classand overall accuracy.

The first two models in table 3.8 and table 3.9 were trained on 2D augmented data while the other models were trained and test on 1D data

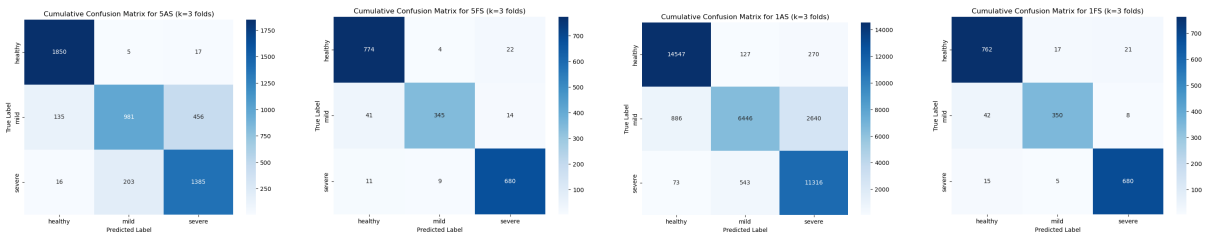
Table 3.8: Comparison of model performance on FS datasets. Boldfaced values indicate the best performance for each metric.

FS Datasets		**Models**					
		Metric (%)	**VGG16+CBAM**	**ResNet18+CBAM**	**Rese**	**Rese_Acoustic**	**Rese_Acoustic(Frozen)**
5s	HC	Precision	93.39 ± 6.56	85.83 ± 6.47	88.55 ± 9.97	**96.60 ± 3.20**	92.44 ± 6.12
		Recall	**96.79 ± 1.10**	89.70 ± 3.72	95.65 ± 3.32	85.45 ± 10.95	96.44 ± 2.82
		F1 score	**94.90 ± 3.03**	87.57 ± 4.14	91.46 ± 4.23	90.15 ± 4.96	94.23 ± 2.81
	PD_Mild	Precision	**95.66 ± 2.99**	77.90 ± 12.90	56.76 ± 9.47	51.95 ± 8.46	42.45 ± 10.52
		Recall	**86.87 ± 11.40**	66.55 ± 2.94	56.54 ± 15.17	62.47 ± 10.66	59.84 ± 13.39
		F1 score	**90.66 ± 6.97**	71.49 ± 7.40	55.27 ± 8.79	55.19 ± 1.46	49.24 ± 10.65
	PD_Severe	Precision	**95.02 ± 0.47**	89.56 ± 6.48	82.86 ± 3.10	73.63 ± 10.77	70.27 ± 13.64
		Recall	**97.06 ± 1.53**	93.44 ± 2.06	76.84 ± 6.53	73.84 ± 6.48	53.43 ± 18.790
		F1 score	**96.02 ± 0.59**	91.38 ± 4.08	79.52 ± 3.54	73.19 ± 6.48	60.12 ± 17.62
		Accuracy	**94.60 ± 2.88**	86.12 ± 1.73	80.22 ± 2.20	75.57 ± 4.05	71.46 ± 11.96
1s	HC	Precision	**92.59 ± 5.07**	83.90 ± 7.17	79.26 ± 13.78	**97.15 ± 2.05**	95.46 ± 2.59
		Recall	**94.98 ± 2.95**	92.44 ± 1.15	95.27 ± 3.73	87.25 ± 15.44	94.63 ± 2.63
		F1 score	**93.64 ± 2.52**	87.79 ± 3.92	85.65 ± 8.21	90.98 ± 8.37	95.01 ± 1.84
	PD_Mild	Precision	**94.53 ± 2.92**	85.26 ± 4.82	31.21 ± 22.97	50.30 ± 5.47	48.87 ± 3.51
		Recall	**87.81 ± 8.93**	73.86 ± 15.91	28.12 ± 29.66	58.05 ± 5.84	60.31 ± 5.64
		F1 score	**90.64 ± 3.68**	77.70 ± 7.74	28.66 ± 26.93	53.32 ± 1.19	53.64 ± 0.15
	PD_Severe	Precision	**95.95 ± 0.57**	90.84 ± 3.15	77.89 ± 3.93	67.88 ± 5.58	73.54 ± 3.89
		Recall	**97.09 ± 0.74**	89.38 ± 3.41	72.23 ± 4.32	74.37 ± 9.72	65.88 ± 1.92
		F1 score	**96.51 ± 0.09**	90.02 ± 1.84	74.78 ± 2.25	70.20 ± 2.19	69.44 ± 2.122
		Accuracy	**94.27 ± 1.74**	86.83 ± 2.86	73.28 ± 7.91	74.22 ± 2.44	76.48 ± 3.81

Table 3.9: Comparison of model performance on AS datasets. Boldfaced values indicate the best performance for each metric.

AS Datasets		**Models**					
		Metric (%)	**VGG16+CBAM**	**ResNet18+CBAM**	**Rese**	**Rese_Acoustic**	**Rese_Acoustic(Frozen)**
5s	HC	Precision	**90.83 ± 10.86**	86.06 ± 5.19	75.58 ± 8.55	88.93 ± 7.15	88.20 ± 8.73
		Recall	**98.92 ± 0.35**	91.56 ± 5.96	99.08 ± 0.72	98.10 ± 1.01	95.84 ± 2.50
		F1 score	**94.34 ± 6.15**	88.38 ± 1.18	85.44 ± 5.47	93.14 ± 3.86	91.54 ± 3.86
	PD_Mild	Precision	81.70 ± 14.30	**83.35 ± 7.44**	54.28 ± 6.63	48.63 ± 3.53	49.41 ± 2.38
		Recall	62.72 ± 10.53	**64.94 ± 16.86**	37.07 ± 15.87	27.32 ± 3.82	32.59 ± 4.41
		F1 score	70.84 ± 11.84	**72.27 ± 12.38**	41.12 ± 12.13	34.66 ± 2.76	39.01 ± 2.58
	PD_Severe	Precision	**74.74 ± 1.22**	72.37 ± 3.06	70.41 ± 30.61	64.00 ± 20.67	65.36 ± 21.41
		Recall	**87.78 ± 6.74**	86.95 ± 3.98	56.11 ± 10.66	72.49 ± 3.51	69.97 ± 6.24
		F1 score	**80.66 ± 3.52**	78.98 ± 3.28	60.74 ± 19.06	66.22 ± 13.04	66.35 ± 15.06
		Accuracy	**83.35 ± 6.50**	81.15 ± 2.40	67.38 ± 13.34	69.78 ± 9.13	69.69 ± 9.70
1s	HC	Precision	**92.60 ± 8.36**	88.61 ± 6.25	75.39 ± 6.82	88.32 ± 8.31	86.23 ± 6.25
		Recall	**97.46 ± 0.43**	96.35 ± 0.31	94.42 ± 2.09	87.23 ± 7.21	91.58 ± 4.40
		F1 score	94.76 ± 4.46	92.21 ± 3.61	83.73 ± 5.04	87.10 ± 1.97	88.50 ± 1.38
	PD_Mild	Precision	**90.39 ± 1.56**	82.44 ± 2.83	42.04 ± 13.92	44.22 ± 26.93	41.92 ± 23.67
		Recall	**65.49 ± 4.13**	61.54 ± 3.76	17.00 ± 5.27	39.70 ± 26.10	26.23 ± 5.38
		F1 score	**75.84 ± 2.46**	70.42 ± 3.12	22.68 ± 4.29	35.53 ± 14.48	31.05 ± 10.03
	PD_Severe	Precision	**78.86 ± 4.46**	77.06 ± 3.73	71.09 ± 22.11	64.14 ± 20.11	59.93 ± 17.37
		Recall	**94.85 ± 0.25**	86.55 ± 3.58	81.61 ± 6.09	64.59 ± 18.78	67.88 ± 13.63
		F1 score	**86.05 ± 2.67**	81.52 ± 3.61	72.94 ± 12.27	58.44 ± 5.26	59.76 ± 6.58
		Accuracy	**87.64 ± 3.14**	83.82 ± 3.02	68.52 ± 10.87	65.68 ± 4.08	66.84 ± 5.38

Since the VGG16+CBAM model outperformed the rest of the models in all datasets, we will will plot its cumilitive confusion matrix for each dataset to see the model's performance in more details (figure3.9)



(a) (b) (c) (d)

Figure 3.9: VGG16+CBAM model performance on different datasets. (a) 5AS dataset, (b) 5FS dataset, (c) 1AS dataset, (d) 1FS dataset.

3.6.0.1 2D data:

For 2D data, VGG16 with CBAM outperformed Resnet18 with CBAM in all datasets used in this study.

We can observe that the results of 2D CNN models of FS datasets were close to the results of AS datasets,with the AS ones just coming out a little better.

Due to class distrebution imbalance the results indicate that the models performed better on the Healthy and severe classes compared to the mild class, which is represented with the lowest amount of samples. Even with class weighting during training, the model still found difficulties in classifying the mild class as as shown in figure 3.9.

3.6.0.2 Raw audio

: Due to the huge difference in the size of the datasets, we can notice that the difference in results compared to 2D CNNS, since the pd classes are represented with very few samples, the best accuracy reached is 76% by ReSE-2-multi(frozen) with extracted features

3.6.1 Binary classification

In order to compare our results with the previous studies that used binary classification for the Italian-speaking Parkinson's speech dataset, we categorized HC as negative and all PD cases as positive, as implemented in [12] results are presented in Table 3.10

Model	Accuracy (%)
VGG16+CBAM	**95.92 ± 0.61**
ResNet18+CBAM	91.04 ± 1.85
1D CNN	85.47 ± 5.22
ResSE (Raw Audio)	87.79 ± 3.05
ResSE+Acoustic Features+MFCC	91.47 ± 2.18
ResSE(Frozen)+Acoustic Features+MFCC	**93.15 ± 2.78**

Table 3.10: Accuracy comparison of different models

3.6.1.1 2D Data

For 2D data, nothing changed as in 3-class classification, VGG16+CBAM was the model that performed the best compared to other models, where it got 96% accuracy, and it stable if with a standard deviation of 0.61

3.6.1.2 Raw audio

1D CNNs did pretty well if we consider the amount of data used in training, where the largest dataset is 1AS, which has 9550 samples before splitting into folds

We can notice that the ReSE-2-multi model performed better than the Simple 1D CNN, which shows the importance of adding the SE module in improving the model's performance, where it got 87.79% accuracy, which is better than the 1D CNN model with 85.47% accuracy.

The best model for raw audio is the ReSE-2-multi+Extracted Features with freezing the convolution layers as explained in 3.5.6 where it outperformed the rest of the 1D CNNs, where it takes advantage of the weights of the pre-trained ResE-2-multi, which leave it with fewer parameters to train compared to 3.5.5

The obtained results show that the integration of extracted features improved performance, which leads us to wonder: if we use only the extracted features, will the results be better or worse?

3.6.2 Do we need Raw audio?

Model	**Accuracy (%)**
svm	90.92 ± 4.54
knn	95.71 ± 3.24
our network	**99.55 ± 0.20**

Table 3.11: Accuracy comparison of different models using only extracted features

The network presented in (Figure 3.8) was tested on the extracted features from 1-second and 5-second datasets (1AS and 5AS); we just added a last output layer of 2 classes. We also tested KNN and SVM.

Our network outperformed the other models (SVM and kNN). Not only these models, but also all other models used in this study.

by comparing our network when it was used alone (only with extracted features), to the same networks when it was used with ReSE-2-multi, we notice a drop in its performance, where it goes from 99.29% accuracy to 93.15%, while being more stable (lower standard deviation) than every other 1D-CNN model

In conclusion, we can say that using raw audio is not necessary, as the extracted features alone yielded significantly better results compared to using raw data alone or a combination of both raw audio and extracted features. The extracted features provided a more stable and higher accuracy, demonstrating their effectiveness in the classification task.

3.6.3 Window Size Importance

3.6.3.1 Raw Audio

Since the best performed on raw audio datasets is ReSE-2multi with extracted features (frozen)3.5.6

We tried to see if the window size has an impact on the model's performance; for that reason, we tested ReSE-2-Multi(frozen) with the extracted features on several window lengths from 1 second to 7 seconds as mentioned in Table 3.3

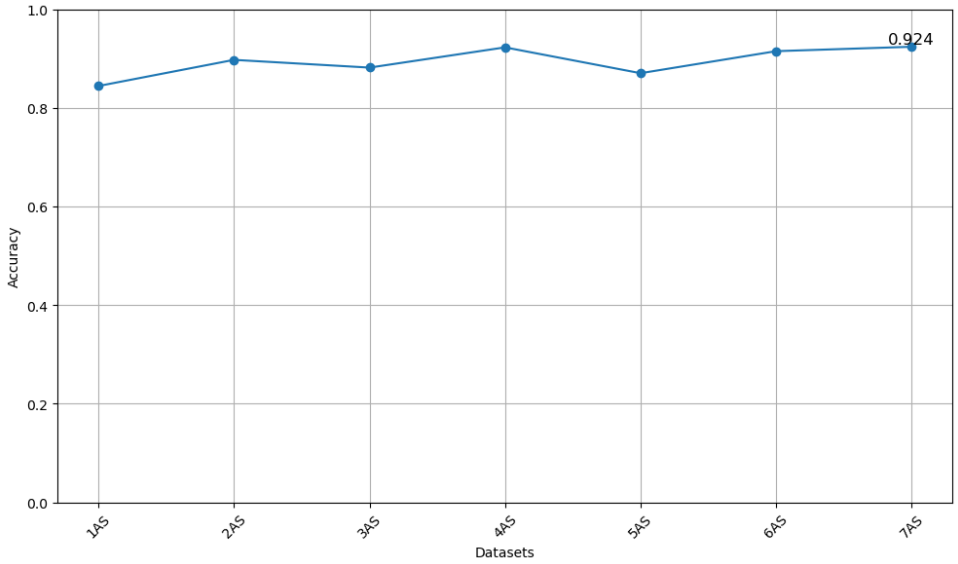


Figure 3.10: Model Accuracies Across AS Datasets for Binary Classification

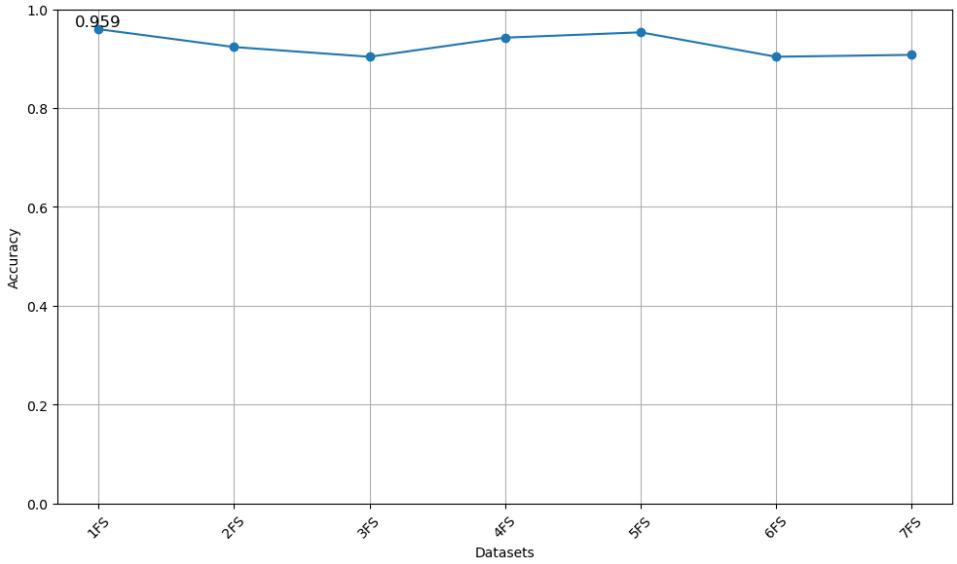


Figure 3.11: Model Accuracies Across FS Datasets for Binary Classification

We can notice that there is a small difference in accuracy when we change the length of the window. For the AS datasets, the best accuracy was obtained from the longest window length dataset (7AS). In the other hand we have the best result for the FS dataset was obtained from the shortest window lenght dataset (1FS)

3.6.3.2 Extracted Features

We tried to test the best-performing model (our network) as shown in Table 3.11, the same way we did for the ReSE-2-Multi model. We computed several metrics for each dataset; the metrics used are accuracy, F1 score, Precision, and recall.The values in the table 3.12 are computed the same way as before, where they represent the average of the three folds ± the standard deviation.

Dataset	Accuracy	F1 Score	Precision	Recall
1AS	99.84 ± 0.08	99.85 ± 0.07	99.89 ± 0.09	99.81 ± 0.05
2AS	99.50 ± 0.10	99.55 ± 0.09	99.56 ± 0.31	99.55 ± 0.38
3AS	99.22 ± 0.32	99.33 ± 0.26	99.11 ± 0.63	99.55 ± 0.32
4AS	98.93 ± 0.58	99.08 ± 0.50	99.23 ± 0.23	98.93 ± 0.78
5AS	99.45 ± 0.16	99.54 ± 0.13	99.26 ± 0.25	99.81 ± 0.26
6AS	99.23 ± 0.31	99.35 ± 0.26	98.93 ± 0.60	99.78 ± 0.31
7AS	98.54 ± 0.21	98.77 ± 0.14	98.31 ± 0.84	99.26 ± 0.58

Table 3.12: Performance metrics (mean ± std) across 7 datasets

It can be observed that the model did we on every window lenght with a negligible difference. Also results show that the model is stable for every dataset used. We can also notice that the recall values are excellent, which is good for our medical classification task, since recall is close to 1 for every dataset.

3.7 Conclusion

In this chapter, we explored several models used on the Italian Parkinson's voice and speech database. We saw the performance of CNNs with CBAM block. We also saw the improvement in 1D-CNN performance after adding the SE block to it.

We also compared the results of using extracted features and raw audio, where we noticed that using the extracted features only gave the best results.

The results shows that the feature extracted give the best results even being the simplest way to classify the audio data.

Conclusion & Challenges & Future Perspectives

3.7.1 Conclusion

Throughout this Master's thesis, we focused on using deep neural networks for the early detection of Parkinson's disease through speech analysis. A key challenge was handling class imbalance and variability in voice recordings. To address this, we applied techniques such as data augmentation and fine-tuning to improve model robustness.

We developed and evaluated several CNNs and ANN on the Italian Parkinson's voice and speech database, achieving promising results, particularly in detecting early PD symptoms. This work highlights the potential of using voice recordings and Deep learning in supporting non-invasive and accurate diagnosis of Parkinson's disease.

3.7.2 Challenges

One of the most well-known challenges in medical-related tasks is that it the model predictions must be validated by professionals due to the sensitivity of the problem. In addition to that, another limitation is interpretability, because medical decisions have to be explained to say why that decision was made, which is not possible for every model in ML and DL, this leaves us with a limited models with easier interpretation, in order to deploy these models for real-world data.

3.7.3 Future Perspectives

We are planning to deploy the trained models to test them on new data. The plan is to deploy the best model from each data type, in addition to creating an ensemble model to explore the performance when models are used together.

Focus more on XAI(explainable artificial intelligence) to create more interpretable results, alongside our search for neuroscientists and neurologists to support interdisciplinary collaboration and to validate results obtained.

Bibliography

[1] Aananya Reddy, Ruhananhad P. Reddy, Aryan Kia Roghani, Ricardo Isaiah Garcia, Sachi Khemka, Vasanthkumar Pattoor, Michael Jacob, P. Hemachandra Reddy, and Ujala Sehar. Artificial intelligence in parkinson's disease: Early detection and diagnostic advancements. *Ageing Research Reviews*, 99:102410, 2024.

[2] Robert L. Haining and Cindy Achat-Mendes. Neuromelanin, one of the most overlooked molecules in modern medicine, is not a spectator. *Neural Regeneration Research*, 12(3):372–375, March 2017.

[3] Eduardo Tolosa, Alicia Garrido, Sonja W Scholz, and Werner Poewe. Challenges in the diagnosis of parkinson's disease. *The Lancet Neurology*, 20(5):385–397, 2021.

[4] X. Zhao, L. Wang, Y. Zhang, et al. A review of convolutional neural networks in computer vision. *Artificial Intelligence Review*, 57:99, 2024.

[5] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 2021.

[6] Xiao Zhang, Ningning Han, and Jiaming Zhang. Comparative analysis of vgg, resnet, and googlenet architectures evaluating performance, computational efficiency, and convergence rates. *Applied and Computational Engineering*, 44:172–181, 2024.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[8] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258, 2017.

[9] L. Aversano, M. Bernardi, M. Cimitile, M. Iammarino, D. Montano, and C. Verdone. A machine learning approach for early detection of parkinson's disease using acoustic traces. In *2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–8, 2022.

[10] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In-So Kweon. Cbam: Convolutional block attention module. *ArXiv*, abs/1807.06521, 2018.

[11] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2017.

[12] Hadi Sedigh Malekroodi, Nuwan Madusanka, Byeong-il Lee, and Myunggi Yi. Leveraging deep learning for fine-grained categorization of parkinson's disease progression levels through analysis of vocal acoustic patterns. *Bioengineering*, 11(3), 2024.

[13] Jongpil Lee, Taejun Kim, Jiyoung Park, and Juhan Nam. Raw waveform-based audio classification using sample-level cnn architectures, 2017.

[14] Richard Armstrong. What causes neurodegenerative disease? *Folia Neuropathologica*, 58(2):93–112, 2020.

[15] Richard N. L. Lamptey, Bivek Chaulagain, Riddhi Trivedi, Avinash Gothwal, Buddhadev Layek, and Jagdish Singh. A review of the common neurodegenerative disorders: Current therapeutic approaches and the potential role of nanotherapeutics. *International Journal of Molecular Sciences*, 23(3), 2022.

[16] Brittany N. Dugger and Dennis W. Dickson. Pathology of neurodegenerative diseases. *Cold Spring Harbor Perspectives in Biology*, 9(7):a028035, 2017.

[17] Werner Poewe. Clinical measures of progression in parkinson's disease. *Movement Disorders*, 24(S2):S671–S676, 2009.

[18] Melissa J. Armstrong and Michael S. Okun. Diagnosis and treatment of parkinson disease: A review. *JAMA*, 323(6):548–560, 02 2020.

[19] Eduardo Tolosa, Alicia Garrido, Sonja W Scholz, and Werner Poewe. Challenges in the diagnosis of parkinson's disease. *The Lancet Neurology*, 20(5):385–397, 2021.

[20] Csaba Váradi. Clinical features of parkinson's disease: The evolution of critical symptoms. *Biology*, 9(5), 2020.

[21] Francesca Magrinelli, Alessandro Picelli, Pierluigi Tocco, Angela Federico, Laura Roncari, Nicola Smania, Giampietro Zanette, and Stefano Tamburin. Pathophysiology of motor dysfunction in parkinson's disease as the rationale for drug treatment and rehabilitation. *Parkinson's Disease*, 2016(1):9832839, 2016.

[22] Anthony H. V. Schapira, K. Ray Chaudhuri, and Peter Jenner. Non-motor features of parkinson disease. *Nature Reviews Neuroscience*, 18(7):435–450, July 2017.

[23] Yuzhe Yang, Yuan Yuan, Guo Zhang, Hao Wang, Ying-Cong Chen, Yingcheng Liu, Christopher G. Tarolli, Daniel Crepeau, Jan Bukartyk, Mithri R. Junna, Aleksandar Videnovic, Terry D. Ellis, Melissa C. Lipford, Ray Dorsey, and Dina Katabi. Artificial intelligence-enabled detection and assessment of parkinson's disease using nocturnal breathing signals. *Nature Medicine*, 28(10):2207–2215, 2022.

[24] Andrew Ng. Machine learning [online course]. `https://www.coursera.org/learn/machine-learning`, 2017. Coursera, Stanford University.

[25] Chip Huyen. *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*. O'Reilly Media, 2022.

[26] J. Cervantes, F. Garcia-lamont, L. Rodríguez-mazahua, and A. Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215, 2020.

[27] Kashvi Taunk, Sanjukta De, Srishti Verma, and Aleena Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. In *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, pages 1255–1260, 2019.

[28] I. H. Sarker. Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6):420, 2021.

[29] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2021.

[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NIPS)*, volume 30, pages 5998–6008, 2017.

[31] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah. A survey of transformers. *Artificial Intelligence Review*, 55(3):1977–2017, 2022.

[32] Ralf C. Staudemeyer and Eric Rothstein Morris. Understanding lstm–a tutorial into long short-term memory recurrent neural networks. *arXiv preprint arXiv:1909.09586*, 2019.

[33] L. Aversano, Mario Bernardi, Marta Cimitile, Martina Iammarino, Debora Montano, and Chiara Verdone. A machine learning approach for early detection of parkinson's disease using acoustic traces. pages 1–8, 05 2022.

[34] Ondřej Klempíř, David Příhoda, and Radim Krupička. Evaluating the performance of wav2vec embedding for parkinson's disease detection. *Measurement Science Review*, 23(6):260–267, 2023.

[35] Máté Hireš, Peter Drotár, Nemuel Daniel Pah, Quoc Cuong Ngo, and Dinesh Kant Kumar. On the inter-dataset generalization of machine learning approaches to parkinson's disease detection from voice. *International Journal of Medical Informatics*, 179:105237, 2023.

[36] Adedolapo Aishat Toye and Suryaprakash Kompalli. Comparative study of speech analysis methods to predict parkinson's disease, 2021.

[37] Giovanni Dimauro, Vincenzo Di Nicola, Vitoantonio Bevilacqua, Danilo Caivano, and Francesco Girardi. Assessment of speech intelligibility in parkinson's disease using a speech-to-text system. *IEEE Access*, 5:22199–22208, 2017.

[38] Giovanni Dimauro and Francesco Girardi. Italian parkinson's voice and speech, 2019.

[39] Zijun Li, Jinpeng Yu, Weixuan Kong, Na Liu, Xuefeng Li, and Hui Xiao. Usage of resnet18 with cbam attention mechanisms in facial emotion recognition. In *2023 International Conference on Sensing, Measurement Data Analytics in the era of Artificial Intelligence (ICSMD)*, pages 1–6, 2023.

[40] Jongpil Lee, Jiyoung Park, Keunhyoung Luke Kim, and Juhan Nam. Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms, 2017.

Abstract

Parkinson's disease (PD), a progressive neurodegenerative disorder, often presents subtle early symptoms such as speech impairments that are challenging to detect with traditional methods. This Master's thesis proposes deep learning models to assist healthcare professionals in the early diagnosis of Parkinson's disease through speech analysis. Our approach combines voice records processing and artificial intelligence, utilizing advanced deep learning models. Various techniques were implemented to enhance model robustness and generalization. Multiple models were evaluated, yielding promising results. This system aims to improve the accuracy and speed of PD diagnosis, offering a valuable tool for early intervention and better patient care.

Keywords: neurodegenerative disorders, Parkinoson's disease, deepl learning, early detection,speech analysis.

الملخص

مرض باركنسون (PD) هو اضطراب عصبي تنكسي تدريجي، وغالبًا ما تظهر أعراضه المبكرة بشكل خفي، مثل اضطرابات الكلام التي يصعب اكتشافها بالطرق التقليدية. تقترح أطروحة الماجستير هذه نماذج تعلم عميق لمساعدة المهنيين الصحيين في التشخيص المبكر لمرض باركنسون من خلال تحليل الكلام. يجمع نهجنا بين معالجة التسجيلات الصوتية والذكاء الاصطناعي، باستخدام نماذج تعلم عميق متقدمة. تم تنفيذ تقنيات مختلفة لتعزيز متانة النموذج وتعميمه. تم تقييم نماذج متعددة، مما أدى إلى نتائج واعدة. يهدف هذا النظام إلى تحسين دقة وسرعة تشخيص مرض باركنسون، مما يوفر أداة قيمة للتدخل المبكر وتقديم رعاية أفضل للمرضى.

الكلمات المفتاحية: الاضطرابات العصبية، مرض باركنسون، التعلم العميق، الكشف المبكر، تحليل الكلام.

Résumé

La maladie de parkinson (MP), une maladie neurodégénérative, se manifeste souvent par des symptômes précoces subtiles tels que des troubles de la parole, difficile à détecter par des méthodes traditionnels. Ce mémoire de Master propose des modèles d'apprentissage profond pour aider les professionnels de santé dans le diagnostic précoces de la MP à partir de l'analyse de la parole. Notre approche combine le traitement d'enregistrement vocaux et l'intelligence artificielle en s'appuyant sur des modèles avancés de l'apprentissage profond. Diverses techniques ont été mise en œuvre pour améliorer la robustesse et la généralisation des modèles. De multiples modèles ont été évalués, donnant des résultats prometteurs. Ce système vise à améliorer la précision et la rapidité du diagnostic de la MP, en offrant un outil précieux pour une intervention précoces et une meilleure prise en charge des patients.

Mots-clés : troubles neurodégénératifs, maladie de Parkinson, apprentissage profond, détection précoce, analyse de la parole.