HERIOT-WATT UNIVERSITY

MASTERS THESIS

Parkinson's Disease Prediction using Machine Learning

Author

Edwin Binu

Supervisor

Dr. Smitha Kumar

MSc in Artificial Intelligence

School of Mathematical and Computer Sciences

August 2025

HERIOT WATT UNIVERSITY

Course code and name:	F21MP, Masters Project and Dissertation
Type of assessment:	**Individual**
Coursework Title:	Parkinson's Disease Prediction using Machine Learning
Student Name:	Edwin Binu
Student ID Number:	H00482595

Declaration of authorship. By signing this form:

- **I declare** that the work I have submitted for individual assessment OR the work I have contributed to a group assessment, is entirely my own. I have NOT taken the ideas, writings or inventions of another person and used these as if they were my own. My submission or my contribution to a group submission is expressed in my own words. Any uses made within this work of the ideas, writings or inventions of others, or of any existing sources of information (books, journals, websites, etc.) are properly acknowledged and listed in the references and/or acknowledgements section.

- I confirm that I have read, understood and followed the University's Regulations on plagiarism as published on the University's website, and that I am aware of the penalties that I will face should I not adhere to the University Regulations.

- I confirm that I have read, understood and avoided the different types of plagiarism explained in the University guidance on Academic Integrity and Plagiarism

Student Signature *(type your name):* *Edwin Binu*

Date: *11/08/2025*

Copy this page and insert it into your coursework file in front of your title page.
For group assessment each group member must sign a separate form and all forms must be included with the group submission.

Your work will not be marked if a signed copy of this form is not included with your submission.

HERIOT WATT UNIVERSITY

Acknowledgments

I thank God Almighty for the blessings which helped me throughout this humble endeavour. I sincerely acknowledge my gratitude to the Program Director of Heriot-Watt, Prof. Neamat Elgayar, for granting me all the resources available in the university to make my project. I also extend my gratitude towards the head of the Department of Artificial Intelligence Associate Prof. Neamat Elgayar, for the valuable support throughout the project. This project would not have been a success if not for the timely advice, guidance, and support from my project supervisor Ms. Smitha Kumar , the Mathematical and Computer Sciences Department. I would also like to express my gratitude to the lab assistants for their valuable support throughout the work on this project.

HERIOT WATT UNIVERSITY

Declaration of Authorship

I, Edwin Binu hereby declare that this research project titled “Parkinson’s Disease prediction using Machine Learning ” is entirely my own work and has not been submitted, in whole or in part, for any other academic degree or professional qualification. All resources of information used in the project have been appropriately acknowledged through citations and references.

Signed: Edwin Binu

Student ID: H00482595

Date: 11 August 2025

HERIOT WATT UNIVERSITY

Contents

List of Figures

HERIOT WATT UNIVERSITY

List of Tables

HERIOT WATT UNIVERSITY

Abstract

Parkinson's Disease (PD) is an age-based degenerative disease whose late diagnosis usually remains unnoticed by conventional clinical tests due to their qualitative nature. With an attempt to overcome this problem, in this research, an approach to Parkinson's Disease diagnosis based on deep learning using vocal biomarkers of the UCI Parkinson's dataset is introduced. A hybrid version of the Convolutional Neural Network and Transformer (CNN–Transformer) network is used in this research to preserve spatial information and temporal relations in structured acoustic features such as jitter, shimmer, and harmonics-to-noise ratios. Normalization, class balancing, and reshaping were utilized in data preprocessing to ready data for learning in sequence. The hybrid approach performed outstandingly showing an accuracy of 78% and an F1 score of 74% along with measures such as precision and recall in distinguishing healthy individuals from people suffering from Parkinson's Disease. This research illustrates the potential of vocal-based deep learning models as non-invasive, low-cost, and scalable medical devices for Parkinson's Disease diagnosis at its onset, enabling timely medical interventions and improved patient care.

HERIOT WATT UNIVERSITY

Chapter 1

Introduction

Parkinson's Disease is an age-dependent neurological disorder that is caused due to the loss of dopaminergic neurons in substantia nigra of brain. This is mainly responsible for causing motor symptoms such as tremors, bradykinesia (slowness of movement), rigidity, and posture instability, as well as non-motor symptoms such as disturbance in sleep, depression, and cognitive dysfunction. [Postuma et al., 2015]. Approximately 10 million people have PD, and its incidence increases exponentially with increase in age. Neurological diseases, of which Parkinson's constitute an included part, cause majority of the disability in the world and its burden is expected to increase as populations get older as stated by the World Health Organization [WHO 2022]. Due to insidious onset of symptomatology and lack of definite biomarkers, diagnosis is difficult in initial stage and tends to postpone intervention after extensive neurological damage has occurred.

PD is complicated because its cause can be multifactorial and include genes, toxins in the environment, and lifestyle. Mutational changes in genes like in LRRK2 or in SNCA cause familial PD, while toxins like pesticides or heavy metals raise risk among those without family history[Kalia & Lang, 2015]. Diagnosis of PD is also so tricky because physicians rely upon noticing symptoms that mimic those of other movement issues, like trembling. Nearly 1 in 5 cases at initial presentation get misdiagnosed, highlighting the urgent need for better tests [Hoops et al., 2009].

The aim of this project is to create an efficient Parkinson's disease predictive system using machine learning. With integration of disparate data sources such as demographic data, genetic markers, and sensor-based monitoring of movements, this system in this project aims to achieve early and accurate diagnosis. Early diagnosis can potentially enable therapeutic interventions such as dopamine replacement therapy or neuroprotective strategies, which can slow disease progression and augment quality of life [Deuschl et al., 2020]

1.1Aim

The aim of this project is to develop a Machine Learning model for early prediction and detection of Parkinson's disease using biomedical data. It aims to analyse the various patterns in the Parkinson's Telemonitoring dataset and apply feature engineering to extract features that will optimize the model performance and reduce noise.

1.2. Objectives

1. To identify an appropriate dataset for Parkinson's Disease classification using vocal biomarkers and pre-process it by handling missing values, normalizing features, encoding categorical variables and to extract and engineer relevant acoustic features that are indicative of dysphonia in Parkinson's patients.

2. To apply dimensionality reduction using Principal Component Analysis (PCA) to retain the most informative features.

3. To design and implement a hybrid Convolutional Neural Network and Transformer (CNN-Transformer) architecture for capturing both local and global patterns in vocal signal features for Parkinson's Disease detection

4. To evaluate performance of the CNN-Transformer model using certain metrices such as accuracy , precision , recall, F1-score, and confusion matrix and also to ensure its effectiveness in diagnosing the Parkinson's Disease.

CHAPTER -2

With the increase in technological advancements there has been a wider and extensive usage of Machine Learning and Deep Learning in the medical industry. These models have provided a reliable and better classification methods and early diagnosis is made with the help of these models providing early intervention of the disease. This chapter provides an overview of the current research and the models used for the detection of the disease. It also presents the related work using various approaches including EEG based methods, voice analysis and hybrid models. This section also provides insights about the challenges, critical analysis and also the research gaps and findings.

Literature Review

With the advancement in technology, there has been a wide usage of Machine Learning(ML) and Deep Learning(DL) techniques in the medical field. Both of them have proved to be a reliable method for the classification and the prediction of several cases and diseases. Early diagnosis is made with the help of these models, thereby preventing the intervention of the disease [Ouhmida et al., 2021]. The major goal is assisting the doctors in fighting these diseases with the help of medical care and infection control suggestions.

2.1. Machine Learning in Medical Diagnosis

Various ML , DL methods have been used to diagnose a patient with the disease . In recent years Deep Learning models have attracted a lot of attention in the medical field, especially with the detection of Parkinson's Diseases (PD). Various CNN and Deep Learning models have shown [Fathima et al., 2024] great success in the analysis of various medical data and are able to diagnose the disease at a very early stage. Moreover, the usage of LSTM network for the detection has also been explored .

Additionally [Naanoue et al., 2023] leveraged the usage of deep learning models for Parkinson's disease detection from speech analysis demonstrating the capability of LSTMS to

handle complex datasets. Diagnostic techniques such as MRI, SPECT only provides supplementary information about the functional alteration related to PD. Acoustic features such as jitter, shimmer, harmonic to noise ratio(HNR), and format frequency are usually sensitive to PD related vocal fold rigidity and respiratory deficits. Models such as CNN and LSTMS are used to analyse these features directly from raw audio signals or spectrograms. Integrating these voice data with deep learning models , the systems are able to detect the voice anomalies even before the motor symptoms thereby enabling the early diagnosis.

Millions of people around the world are usually affected with PD . This is usually characterized by motor symptoms such as tremor, bradykinesia , rigidity, and postural instability as well as other symptoms such as cognitive decline and disturbances in sleep and also emotional dysfunction. The only way for prevention is early diagnosis and that can lead to a better quality of life and improved disease management. However, it is really difficult for early detection due to the variation of symptoms in the prodromal phase. Recently due to the advancement in technology and also due to the emergence of various ML and DL techniques , it has made it possible to for easy diagnosis and also its high accuracy classification and automated feature extraction makes it reliable. This chapter presents an overall review of the methods that can be used for early intervention and detection of the disease focusing mainly on the ML and DL across different data modalities which includes imaging , voice, and motion sensor and also EEG.

2.2. Related Works

2.2.1 Image Based Diagnosis

Single Photon Emission Computed Tomography (SPECT), Photon Emission tomography(PET) and Magnetic Resonance Imaging (MRI) are mainly used for visualizing the structure and functional changes in the brains of PD patients. The study by [Antikainen et al., 2021] uses SPECT images feature for early detection of PD. This study evaluated 23 SPECT image features for early PD diagnosis and used two databases. This included 646 subjects (299 with PD disease) that aged from 30 to 89 from both the databases. The database one included cases from multicentre considering some dementia as Parkinson’s Disease. From these 28

uncertain diagnoses were excluded that resulted in 276 SPECT images. While the database two included 195 healthy people and 175 affected ones. Furthermore, they also used various ML classifiers such as SVM, Logistic Regression, KNN, Random Forest and also decision trees. They achieved an overall accuracy of 94% which indicates that the high discriminative power in distinguishing PD patients from healthy controls. Moreover, the analysis also identified four key features that were most important for the classification, that included the most affected side of the putamen and the length of the most affected striatum.

However, there were various disadvantages to this paper , some of them being , limited feature set. Even though the study explored 23 SPECT image features , it was found that only eight features could achieve the performance . The next was although the study aimed for improved model generalization by fully independent test set , the reliance on specific features derived from the DaTQUANT software limited the applicability of the findings to other datasets. Although multiple ML models were used, they increased the risk of overfitting, especially if the models were complex. While some of the future work included focusing on validating the findings across different dataset and clinical settings to ensure the generalizability of the model. There is also a need to explore additional SPECT image features that may enhance the model's performance. Moreover, it should also aim to improve the robustness of the ML models , so that it prevents overfitting and ensure that it performs well on the unseen data. Finally, it can emphasize the importance of translating these findings into clinical practice.

2.2.2. EEG based Detection Methods

EEG is one of the best and cost-effective methods to find out the brain activity. Due to its high temporal resolution is it highly recommended. There have been various studies done using the EEG signal based . The study A brief review on EEG based diagnosis of Parkinson's Disease using Deep Learning models by [Ghasemi et al. 2024] focuses on one such. Here they have used a hybrid deep learning model incorporating CNN and Long Short-Term Memory (LSTM) networks to classify EEG signals of PD patients. One of the highest advantages is that they gained an accuracy of 96.9%. Other than that, Graph Neural Network (GNNs) were also used . This also suggests various future research where further integration of deep learning methods with standardized clinical assessment tools like UPDRS and MoCA which provides the

evaluation of PD. Furthermore, exploring the GNNs and deep fuzzy classifiers to improve the understanding of neural disruptions in PD. Moreover, continued development of automated tools for PD detection using EEG , that could facilitate the adoption of AI systems.

While all these into consideration the paper also lacked the interpretability. This was one of the major disadvantages of the paper. Even though these models could achieve very high accuracy their complex nature makes it difficult to understand as to how the decisions are made. This also points out that even though EEG is a non-invasive method for diagnosis it may not be able to capture all the relevant aspects of the disease.

A study by [Li et al. 2023] , a comparative analysis where a hybrid deep learning model were applied to the EEG data. The study utilized two datasets University of New Mexico(UNM) and university of Lowa (UL). The UNM datasets included 27 PD patients and 27 healthy ones. The UI dataset consisted of 14 PD patients and 14 healthy ones. The EEG data were recorded on a 64-channel system. The model integrates VGG convolutional blocks and LSTM for temporal pattern integration. The CNN architecture included an input convolutional, ReLu pooling and fully connected layers . The CNN was increased in depth so that it led to better outcomes. Also, VGGs smaller convolution kernel enhance the feature extraction complexity. It used a 10-fold CV method for training and testing. The model further achieved an accuracy of 95.23. The combination of both the models enhanced feature extraction and temporal corelation.

While this model had achieved a high accuracy there were also some drawbacks for this model. One of the primary drawbacks of this model was the usage of smaller dataset . A small dataset was used for training and validation. Another one was the lack of severity categorization where the current model is able to distinguish between PD and healthy controls, but it is unable to categorize the degree of the disease. One major drawback was the dependency on the EEG data quality. The models performance highly depended on the quality of the EEG collected. Any noise or impairments in the EEG could severely affect the accuracy of the diagnosis and detection of the disease.

The future works that could be carried on is focusing on the dataset expansion. This would enhance the model's robustness and also improve the generalizability. Also, the model could further aim to incorporate mechanism to categorize PD, which would further help the clinicians. Moreover, since a hybrid model is used it could be further modelled to identify various other neurological diseases.

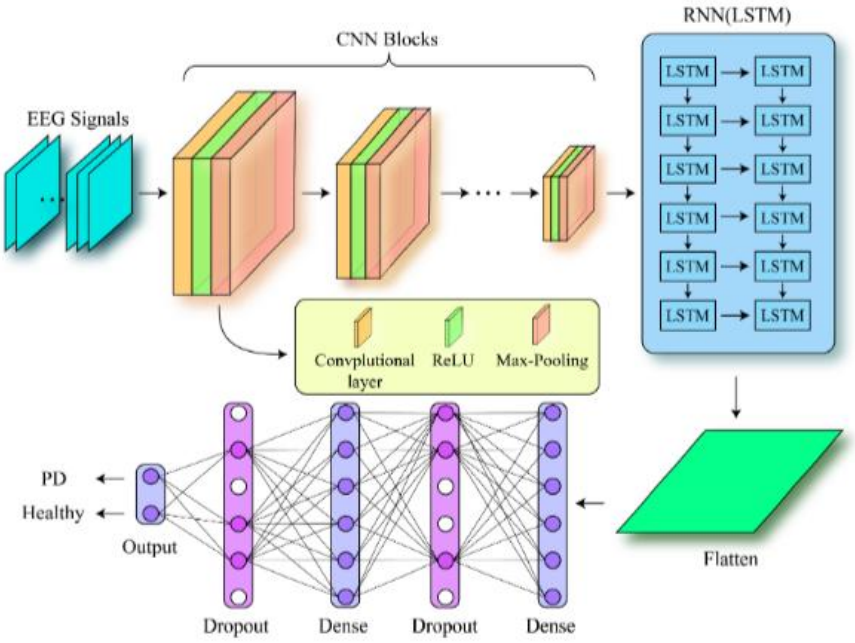


Figure 1:Deep learning Hybrid Model Architecture Parkinson's EEG classification [Li et al. 2023]

2.2.3. Voice and Speech Analysis

One of the earliest symptoms of PD is hypokinetic dysarthria, also known as voice impairment. So, voice analysis can help in early detection of the disease. Various ML techniques have been used in this to identify and diagnose the disease at the earliest stage. Various features such as jitter, shimmer, harmonics to noise ratio and Mel Frequency Cepstral Coefficients (MFCCs). The study by [Ouhmida. et al. 2021], uses Deep Learning approach on voice recordings. For this study two distinct datasets from UCI machine learning repository named as database1 and database2 were utilized which mainly focused on the components and also on the acoustic features. The database one consisted of 195 voice recordings from 31 individuals where 23 were affected with PD while 8 were healthy. The model achieved an accuracy of 93.10 while this dataset was used with a CNN classifier. The second dataset consisted of 45 acoustic features. The CNN classifier was able to achieve an accuracy of 88.89% with this dataset while the ANN was able to secure an accuracy of 72.22%. In summary both the dataset differed in

the terms of the acoustic features. The dataset was also a bit challenging due to its larger feature set.

While this study brought out really good accuracy and results there were also quite few disadvantages . One of the major drawbacks of this model was the utilization of two datasets with varying number of acoustic features. The increase complexity in the database two may lead to the confusion of the model. Another drawback was the limited generalizability where the models trained on the UCI machine learning repository , could limit their applicability to broader population or different dataset. One major disadvantage was that this study focused mainly on the vocal features only.

Keeping aside the limitations some of the future works that can be carried out was the exploration of other advanced deep learning methods. A hybrid system that also combines various techniques and also datasets.

2.2.4 . Smart phone sensor-based detection

With the rapid increase in technologies, now smartphones have emerged as one of the tools for monitoring PD symptoms. The in-built accelerometers , gyroscopes and microphones in smartphones are used to continuously capture data related to motor and vocal behaviour, which are significant in the diagnosis of the disease. The smartphone sensors are able to collect diverse data from large cohorts. Other than these DL methods such as Transformers and also CNNS have shown significance in modelling such data.

The study [Prince & de Vos, 2021] where data from daily activities such as walking and tapping tests are collected by smartphone sensor data by introducing a Deep learning framework. The data is collected from the mPower study using iPhone . Participants finished the walking , memory, and voice tests remotely. Out of the 9,520 participants that were enrolled around 6,333 completed the demographics and the tapping test. This model uses the above mentioned Convolutional Neural Network (CNN) for the classification. The accuracy turned out to be the highest and the F1 scores were the highest compared to all other models . Additionally, a Deep Neural Network (DNN) was also implemented . The DNN also showed strong performance and was able to achieve the highest AUC and accuracy among the classification techniques.

One of the major advantages is that these tests can be performed outside clinical environments which makes the studies and symptom tracking more feasible. The DL models that were trained on this dataset were able to differentiate between patients with PD and also healthy ones .

One of the major disadvantages of this model was the reduced performance of the usual classification techniques when they were applied to the remotely collected dataset . Another drawback was the CNN architecture that was used in the study . The CNN architecture was described to be simple, and it had it limitations to its ability to capture more complex patterns in the data. Furthermore, this also lacked the longitudinal data that provided deeper insights into the progression of PD .

2.2.5. Feedforward Neural Networks and clinical data

Since there are not much diagnostic tests for PD , it is leading to high rate of misclassification and inappropriate treatment , which can be detrimental to patients. The study [Parkinson's disease detection using feed forward neural networks, 2023] proposed a method that uses Feed Forward Mechanism. This approach uses the voice recording dataset from both PD patients and also healthy individuals , that is extracted from UCI machine learning repository.

The dataset includes 26 features that is extracted from the voice recordings with a total of 197. The dataset was imbalanced with 115 records from the PD patients and 41 from the healthy patients that required a model training. The method uses Feed Forward Neural Network(ANN). ANN with two activation function were used in this model. The testing involved 39 voice recording . The study was reported to achieve an accuracy of 100 % that indicated the perfect classification on the dataset. Other than the ANN various other modes such as Random Forest as well as Bidirectional LSTM was also used where their accuracy was reported to be 97.5% and 75.56% respectively. The evaluation metrics that were used were accuracy , precision , recall, F1 score.

Although the classification was accurate and precise there were several drawbacks for this model. One of them was the usage of a smaller dataset. A small dataset led to overfitting as it performed best on the training data, but it performed poorly on the unseen data. Other major limitation was the use of imbalanced dataset. Even though techniques such as SMOTE was used to balance the data ,the effectiveness of the methods in improving the model performance is still a concern.

The future studies could focus more in collecting a larger dataset that could include a variety of patient demographics and also various voice features. It can also include comparing the Feed Forward Neural Network with other advanced ML models such as CNNs or RNNs

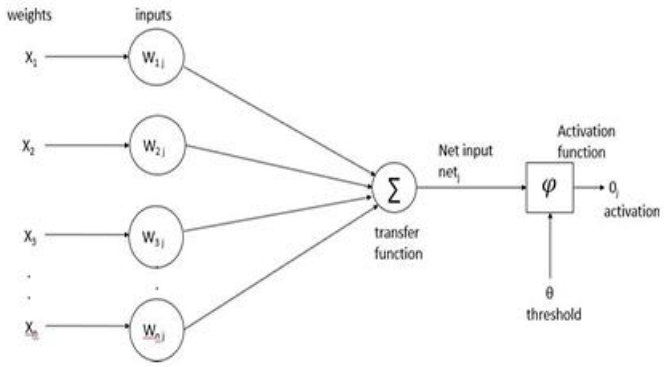


Figure 2 Mathematical model of a neuron[Parkinson's disease detection using feed forward neural networks, 2023]

Another study by [Aliero & Malhotra, 2023] proposed a deep learning-based detection for the Parkinson's Disease. This study mainly focused in using CNN and LSTM for the capture of spatial and temporal features in the clinical data. The various clinical data include voice recordings , motor function data and additional symptoms. These usually include the data from motion sensor , analysed changes in the speech pattern in the voice recordings and other clinical indications. The proposed method achieves an accuracy of 80% when CNNs are used for analysing clinical data. But CNNs may not be able to capture temporal dynamics for diagnosis. Therefore, an LSTM is also used. The LSTM model demonstrated a higher accuracy than the CNN. It produced an accuracy of 99%. Thus, both the models were integrated together in order to get a higher accuracy rate.

While the study provides high accuracy and classification it also has a lot of limitations. Some of the limitations are data heterogeneity but also does not mention how the multimodal data inconsistencies were normalized . Another drawback is the lack of model explainability .The final drawback is the lack of diversity .

In summary this study emphasized the usage of various clinical data for diagnosis and classification and used advanced deep learning mechanisms to improve the detection of the disease

2.2.6. Hybrid Deep Learning Models

One of the Hybrid mechanisms is the integration of the different learning mechanisms and data modalities. This mechanism improves the robustness and generalizability of the model. One of the approaches was combining the CNN and Random Forest networks. The CNN serves as an automatic feature extractor , that learns hierarchal and spatially significant patterns from the input data whereas Random Forest acts as a powerful and interpretable classifier that makes the predictions based on the extracted deep features. The integration usually makes it effective for biomedical data including voice recordings in Parkinson's disease where extracting subtle acoustic cues and classifying them is critical for early-stage detection.

A study by [Mehta& Kumar 2024] , where initially a hybrid model was developed that combines a Convolutional Neural Network(CNN) and Random Forest (RF) algorithm to enhance early diagnosis of the disease. The model combines the feature extraction capabilities for intricate data for instance MRI scans and sensor-based movement data and the patterns classifying abilities of Random Forest. The CNN structure is designed to perform the extraction quickly through multiple convolutions , pool and dropout layers which is beneficial in identifying pattern unique to PD. The RF model enhances the precision of the model using multiple decision tree and reducing the likelihood of overfitting data processing. The research also uses various evaluation measures such as ROC curve , confusion matrices for the model's performance. The data set used were from an extensive dataset merging various medica imaging and physiological patient measurements for Parkinson's Disease. The dataset was an instrumental part in enabling controlled and accurate tests of the model performance. This model gained an accuracy at different stage of PD

Stages	**Accuracy**
Stage 1	92%
Stage 2	88%
Stage 3	93%

Stage 4	93%
Stage 5	91%

Table 1: Accuracy [Mehta& Kumar 2024]

Overall, the model achieved an accuracy of 92% and precision rates of 85% to 94% and recall rates of 87% to 93%

Despite the model's high accuracy , its performance at different stages of PD was somewhat inconsistent and its precision was least in stage2, in comparison to the other stages. Inconsistency can be an indication of the inability to classify some disease stages properly. The other drawback was the dataset imbalance and overfitting of models.

The future works that can be carried out can include strengthening the model by overcoming said weakness for instance by improving classification for some phases of PD. In addition, larger sized and diverse dataset experiment can help to future enhance generalization capacity of the model and reduce the dataset imbalance. The use of advance feature selection and extraction algorithm can parameter tuning of models can also be useful in increasing the performance of the model.

2.2.7 Residual Network-Based Deep Learning for Parkinson's Disease Classification Using Vocal Datasets

The usage of Deep Learning and other ML methods to vocal datasets has been used extensively for the detection and classification of Parkinson's Detection (PD). The study by [Ogawa and Yang 2021] proposes to develop a non-invasive method for diagnosing and early detection of Parkinson's disease by utilizing vocal feature dataset. This approach uses a 10-layered one -dimensional CNN for the diagnosis of Parkinson's disease from the vocal features. The classification was tested using leave – one -patient -out cross validation (LOPO CV) because of the smaller number of examples within the data. This involves training on all but one patient and testing on excluded patient and then repeating for each patient. The study uses the dataset from the UCI machine Learning repository which included 188 patients

diagnosed with Parkinson's disease(107 males and 81 females) out of which 64 were healthy individuals. The vocal data was collected through continuous production of vowel "a", with each subject three repetitions were taken . Various other methods were utilized from vocal feature analysis. That is Mel Frequency Cepstral Coefficients(MFCCs), Wavelet Transform based Features, vocal Fold features and Tuneable Q-factor wavelet transform (TQWT). This model had an accuracy of 0.888% and an F measure of 0.928 % and the Mathews Corelation Coefficient were 0.692% that indicated the strong performance. It was also found out that residual type deep neural network outperformed the traditional neural networks.

Besides this there were several drawbacks of this model there were difficulties in the hyperparameter tuning and the overfitting due to the small size of the data. In addition to this even though the results were good , vocal features can't alone capture all the dimensions of the disease suggesting the need for other models.

2.3. Limitations and challenges of existing models

Despite several advancements, the actual implementations of Deep Learning based PD detection system still continuous to have several limitations. Some of them are:

2.3.1. Data Quality and Scarcity

1. Small or imbalanced dataset :
   Several studies rely upon data sets of either unbalanced class distributions or small sample size . For example, the data set of EEGS based research by [Li et al., 2023] used 27 PD patients and 27 healthy controls, with increased risk of overfitting and decreased generality. Speech models trained using the UCI machine learning repository data [Ouhmida et al., 2021] suffered from unbalanced data (115 PD vs 41 healthy individuals) necessitating the synthetic oversampling strategies such as SMOTE which can introduce bias [Aliero & Malhotra, 2023].

2. Dataset Heterogeneity :

   Study comparison s were limited due to difference in data acquisition protocol and feature extraction software SPECT- based models [Antikainen et al., 2021], for example were created from the features extracted through DaTQUANT software and were less compatible with other software data sets. Smartphone sensor studies [Prince & De vos, 2021] were also impacted by the inconsistencies in remotely

2.3.2 Model Overfitting and Complexity

1. Overfitting in Deep Models: Deep models such as CNN-LSTM models for EEG [Ghasemi et al., 2024] and speech analysis [Ouhmida et al.,2021] were highly accurate and yet overfit due to the complexity and the insufficient training data. Feedforward neural networks [Aliero & Malhotra, 2023] overfitted when trained using few recordings of speech despite achieving high accuracy for test recording

2. Lack of interpretability : Models such as CNN, although powerful , lack transparency in decision making. EEG based research [Ghasemi et al., 2024] noted this as one of the key obstacles.

2.3.3 Limited Clinical Relevance

1. Inability to capture Disease Progression : Majority of the models usually focused on binary classification and failed to categorize the disease severity. For example, EEG models [Li et al., 2023] and hybrid CNN-RF models [Mehta & Kumar, 2024] could not stratify PD stages.

2. Single Modality Focus : Studies that focus mainly on vocal features [Ogawa & Yang, 2021] or SPECT images [Antikainen et al., 2021] misses the multisystem manifestation of PD e.g., motor, and cognitive symptoms.

2.3.4. Technical and Practical Problems

1. Dependency on Data Quality : EEG based models[Li et al., 2023] are highly sensitive to noise and artifacts which lowers the performance. Smartphone senor data [Prince & De Vos, 2021] suffer from variability due to uncontrolled environment in similar fashion.

2. Hardware and Software Limitations: SPECT based solutions[Antikainen et al., 2021] use commercial software (DaTQUANT) that limits the reproducibility.

2.3.5. Generalizability Issues

Ethnic and Demographic Bias: Most dataset lack diversity in age , ethnicity, and comorbidity in the dataset. The SPECT study [Antikainen et al., 2021] excluded uncertain diagnoses, potentially skewing results toward clearer cases and reducing real-world applicability.

2.4. Critical Analysis

Paper -Title	Author	Model	Data Type	Accuracy
SPECT – image features for early Detection of Parkinson's disease using ML Methods	[Antikainen et al., 2021]	SVM, K-NN, Random Forest	SPECT Image features	94 %
A brief review on EEG based diagnosis of Parkinson's Disease using Deep Learning models	[Ghasemi et al. 2024]	CNN, LSTM , Hybrid Models	EEG signals	Upto 96.2 %

A Deep Learning Hybrid Model for Parkinson's Disease Diagnosis Based on Electroencephalogram Signals	[Li et al. 2023]	VGG+ LSTM	EEG signals(UNM and Ul dataset)	95.23%
Voice Based Deep Learning Medical Diagnosis System for parkinsons disease prediction	[Ouhmid et al., 2021]	CNN, DNN	Voice recordings	93.10% DB1 88.89 DB2
Deep Learning framework for remote detection of Parkinsons disease using smartphone sensor data	[prince & De Vos 2021]	CNN, DNN	Smartphone sensor data(mPower study)	Highest AUC
Parkinsons disease detection using feed foreward neural networks	[Unnamed 2023]	Feedforward Neural Network(ANN), RF	Voice recordings	100% ANN 97.5% RF
Deep Learning based detection for Parkinson's Disease using clinical data	[Aliero& Malhotra 2023]	CNN, LSTM	Clinical data	80% CNN 99% LSTM
Hybrid CNN-RF model for early diagnosis of PD	[Mehta & Kumar 2024]	CNN Random Forest Hybrid	MRI Scans , sensor-based movement data	92%
Residual Network based Deep Learning for PD classification	[ogwa & yang 2021]	1D CNN	Vocal features	88.8%(F measure 92.8%)

Table 2: Critical Analysis

2.5. CNN- Transformer Architecture

The novel hybrid CNN–Transformer model developed intends to identify Parkinson's Disease (PD) from organized voice-based data derived from UCI Parkinson's Disease dataset. The approach takes advantage of both Convolutional Neural Networks (CNNs) and Transformer networks—two popular deep learning algorithms whose combination offers powerful abilities in terms of recognizing spatial patterns as well as temporal sequence learning and also offers robust capabilities.

Features used in this work include 22 numeric vocal features . Features include measures such as jitter, shimmer, fundamental frequency, Harmonics-to-Noise Ratio (HNR), and non-linear signal complexity measures of Recurrence Period Density Entropy (RPDE) and Detrended Fluctuation Analysis (DFA). All have been found in earlier research to be significantly altered in Parkinson's Disease speakers due to hypokinetic dysarthria, an initial sign of PD.

In this approach, CNN is used as front-end feature extractor. It processes structured voice features and captures local relations and relations among corresponding parameters—like its batch of jitter- or shimmer-related features. We use 1D convolution layer, batch normalization, and max pooling for extracting those spatial relations and reducing dimensions as well as accelerating training. CNN can learn feature combinations more indicative of PD automatically without manual feature selection and transformation. The outputs of CNN layers serve as input to a Transformer layer. Although data is not in classical time-series data form, reshaped input feature simulates sequential structure within the model. The self-attention layer of the Transformer models global dependencies , that helps in capturing the relationships that reflects the Parkinson's Disease. In biomedical datasets, such dependencies can reflect changes over time or disease advancement.

The output of Transformer is fed into one or multiple dense layers consisting of 64 neurons along with ReLU as the activation functions, and an overfitting prevention layer using dropout. The dropout layer has a rate of 0.3 in order to prevent the overfitting. Classification is performed using one sigmoid-activated output neuron to give a probability score for input voice belonging to either a healthy subject or Parkinson's Disease subject. Overall, training is

performed using binary cross-entropy loss and optimization is usually done using a standard optimizer like Adam.

The CNN-Transformer structure utilized in this work is simple yet potent. It facilitates learning in an end-to-end manner for raw structured features to labels yet takes advantage of both sequence learning and local pattern identification. This makes it highly suitable for structured biomedical data as in UCI Parkinson's dataset. The model is versatile and scalable as well and can in the future be used for larger speech datasets or real-time speech monitoring devices for speech-based input collected using mobile/wearable devices.

2.6. Summary and Research Gaps

To summarize this literature review provides a landscape of Parkinson's detection using deep learning and machine learning methods. From EEG and voice data to gait analysis, smartphone sensors various methodologies have been investigating that were used for diagnosis. Deep Learning models , such as CNN, LSTM, Transformers, and hybrid architectures like CNN-Transformer performed better when compared to Machine learning algorithms.

Despite this there are several challenges that persist. Limited data availability, inter patient variability , lack of model interpretability and ethical concerns remains significant barriers to clinical deployment. Additionally, models have trouble generalizing beyond demographic and dataset boundaries. Additionally, the majority of research focuses on binary classification (PD vs Non-PD) and lacks the multi-class diagnostic tools necessary to distinguish between closely related illnesses or PD subtypes.

Furthermore, the project addresses these gaps by using a novel, hybrid CNN-Transformer architecture for the detection. It uses both the CNN and Transformer model in order to capture local patterns and global dependencies respectively improving standalone methods like CNN, RNN, LSTM , basic Machine Learning or Deep Learning techniques. A Streamlit dashboard is also developed in order to make real time disease prediction that can be used for clinical usage. However, there are still small challenges like small dataset size that limits the generalizability.

HERIOT WATT UNIVERSITY

CHAPTER -3

This chapter provides an overall view of the methodology used along with the dataset selection, pre-processing, feature extraction and the appropriate model used in capturing the voice data that can be used to detect the Parkinson's disease. In general, it gives an overview of detailed guidelines that were conducted for the selection of the model.

Methodology

3.1. Dataset

For this research, the Parkinson's Telemonitoring Voice Dataset from the UCI Machine Learning Repository will be utilized [Tsanas et al.,2009]. This dataset contains biomedical voice measurements from 42 subjects, that were diagnosed with Parkinson's disease (PD). The dataset contains 5,875 voice recordings and each row in the dataset is labelled with several attributes. The dataset contains various scores that can be used to detect the severity of the disease that the patient has.

3.1.1. Data Source and Attributes

The dataset comprises of 5,875 voice recordings from 42 subjects with 22 acoustic features extracted from the voice signals, along with clinical scores such as *total_UPDRS* and *motor_UPDRS* to indicate the severity of the disease. It also shows some class imbalance in these scores that is assessed to prevent the biased results. Following are the dataset attributes :

S.No	Feature Name	Type	Description
1	Subject#	Identifier	A unique identifier for each subject in study
2	Age	Numeric	The age of subject in years
3	Sex	Categorical	The biological sex of the subject

4	test_time	Numeric	The time of the measurement in days from baseline recording
5	motor_UPDRS	Numeric	The motor components of the UPDRS score , assessing the motor symptoms.
6	total_UPDRS	Numeric	The total UPDRS score assessing both the motor and non motor symptoms
7	Jitter(%)	Numeric	Measure of frequency variation in voice
8	Jitter(Abs)	Numeric	Measure of frequency variation in the voice in absolute terms.
9	Jitter:RAP	Numeric	Relative amplitude perturbation , a measure of jitter.
10	Jitter:PPQ5	Numeric	Five Point period perturbation, a measure of jitter
11	Jitter:DDP	Numeric	Average absolute differences of differences of consecutive jitter periods.
12	Shimmer	Numeric	A measure of amplitude variation in the voice
13	Shimmer(dB)	Numeric	A measure of amplitude variation in the voice
14	Shimmer:APQ3	Numeric	Three-point amplitude perturbation quotient.
15	Shimmer:APQ5	Numeric	Five-point amplitude perturbation quotient.
16	Shimmer:APQ11	Numeric	Eleven-point amplitude perturbation quotient
17	Shimmer:DDA	Numeric	Average absolute difference between amplitudes of consecutive periods.
18	NHR	Numeric	Noise-to-harmonics ratio
19	HNR	Numeric	Harmonics-to-noise ratio.
20	RPDE	Numeric	Recurrence Period Density Entropy, a nonlinear measure of signal complexity.

21	DFA	Numeric	Detrended fluctuation analysis, used to quantify long-range temporal correlations.
22	PPE	Numeric	Pitch Period Entropy, measuring signal unpredictability.

Table 3:Data Set Attributes

3.2. Data Pre-processing and Feature Engineering

Data pre-processing is one of the most important steps for deep learning models. Since Parkinson's disease detection uses voice signal data , the missing values, temporal inconsistencies, and scaling dependencies should be handled . Since hybrid CNN -Transformer model is used the pre-processing should also include the steps for temporal and spatial representation of features. Data quality should be optimized , noise should be reduced.

3.2.1. Feature Extraction

The dataset is pre-processed and gathered via prolonged vowel phonation's; initial feature extraction being performed by means of digital signal processing(DSP) methods.

- Fundamental frequency measurements (Fo, Fhi, Flo): They are derived through autocorrelation and cepstral methods
- Measures of jitter and shimmer: Derived from cycle-to-cycle variability in the amplitude as well as in the time domains.
- Harmonics-to-Noise (HNR) and Noise-to-Harmonics (NHR) ratios: Derived using frequency – domain analysis such as Fourier Transform or Linear Predictive Coding (LPC)

3.2.2. Handling Missing Values

One of the preliminary tasks in the pre-processing pipeline is the handling of the missing values within the dataset since, without handling them, they lead to biased learning as well as biased performance of the model. It was found that there are no missing values in the dataset. This is due to the dataset from controlled voice recordings which in turn allows the model to train completely and much efficiently thereby reducing the bias or accuracy loss.

3.2.3. Feature Encoding

Even though the data is primarily numeric features derived from voice signals, any categorical metadata such as gender, language or recording conditions will require encoding. Techniques used are:

- Label encoding for binary or ordinal features
- One Hot encoding in case categorical features having multiple classes is thereafter incorporated

It is usually compatible with deep learning models, in particular CNN layers that only take numerical inputs

3.2.4. Normalization and Standardization

For both features to have an equal contribution towards the learning of the model , feature scaling is really necessary. Following are the methods used:

- Z score normalization is used to normalize the features to have zero mean and unit variance. This then enable the CNN layers to learn quicker.
- Min – Max scaling is attempted as the alternative , especially for those features where ranges are considered like jitter and shimmer.

3.2.5. Dimensionality Reduction

To improve generalization overall and minimise the overfitting dimension reduction is investigated:

- It is done to retain the 95% variance using principal Component Analysis (PCA)

3.3. Model Selection

Choosing an appropriate model is an important task in designing an efficient Parkinson's Disease (PD) detection system. It relies on the nature of the dataset the complexity of the underlying structure as well as the need for predictive capability as well as for interpretability. In view of the data in this research comprising voice recordings along with their corresponding acoustic features, both conventional machine learning models as well as deep learning models are considered in this regard. Such preference for Deep Learning is strengthened further via proposals for hybrid models, i.e. combination of Convolutional Neural Networks(CNN) as well as a Transformer whose performance in biomedical signal processing tasks has been increasingly promising.

Below are the models that are used :

1. Convolutional Neural Network(CNN):

   CNNs can automatically learn spatial hierarchies of features and are inherently qualified for processing structured feature maps of acoustic features. In this CNNs are applied in order to automatically uncover the local patterns in reshaped features of voice as 2D matrices. This is drawing on effective experience in research on speech emotion recognition as well as pathology detection, in both of these applications where CNNs were successful in separating significant frequency domain distinctions.

2. Transformer

Transformers are very powerful models. They can be used for capturing long range dependencies in sequential data with the self-attention that allows them to prioritize the relevance of various features or time points. Using the self-attention layer, the model is able to emphasize the important acoustic features across the input sequence , such as fluctuations in jitter or HNR that can indicate the disease. The transformer architecture consists of feed – forward layers that allows the model to capture various patterns in the voice dynamics. It also has multiple attention heads and feed forward layers that allows it to offer detailed interaction among the characteristics. Transformers are one of the best when it comes to biomedical domains such as EEG analysis due to their ability to handle sequential data without the issues of vanishing gradients

3. Hybrid CNN-Transformer Model:

A hybrid CNN – transformer network is the focal point of this work and exploits the benefits of both structures

CNNs are used for extracting local as well as spatial corelations between acoustic feature metrics. Then there are Transformers , that model global dependencies that makes it much better to detect Parkinson's related vocal impairments.

The CNN layers are used to extract the local spatial correlations in the CNN layers using the input acoustic features. The convolutional filter is used to detect the patters between the features such as *jitter, shimmer* etc. that are some of the major indicators of the disease.

For the Transformer component the output from the CNN is passed into the layers. The transformer uses a self-attention mechanism in order to capture long range relationships in data. This in turn allows the model to learn how various features like *HNR* and *PPE* vary over time.

Finally, the output from the transformer is then fed into a fully connected layer which in turn has sigmoid as an activation function. This helps in classifying whether the patient has mild to moderate symptoms based on the labels like *total_UPDRS*.

This hybridization allows the model to train on local acoustic features as well as temporal patterns of voice, both of which are important for detecting subtle effects of Parkinson's disease. Hybrid models have performed superiorly in related applications in speech pathology, EEG processing and gait analysis in neurodegenerative disease.

3.4. Programming Environment

Category	**Library**	**Description**
Programming Language	Python 3.10+	Used for development
Integrated Development Environment (IDE)	Jupyter notebook	Used for development , visualization and testing
ML/DL frameworks	TensorFlow, KEras, Scikit-learn	Used for training ML and DL models
Data Manipulation	Pandas , Numpy	Handling dataset and numerical computation
Visualization Libraries	Matplotlib, seaborn	Visualize data distributions
Preprocessing	Scikit-learn	Used for normalization, encoding
Dimensionality reduction	PCA, t-SNE	To reduce feature space

Table 4: programming environment

3.5. Evaluation strategy

For estimating the effectiveness of the hybrid CNN- Transformer methods for PD classification from audio data , a systematic assessment planning is used. Model performance is estimated using classical measuring such as precision , recall, accuracy and F1 score in order to provide an overall balanced estimate of its predictability. A confusion matrix will also be developed in order to provide an overall data of true positive , true negative , false positive and false negative. A stratified sampling-based data splitting of 10-fold cross validation is to be used in order to further estimate the strength of the model. Other than that, an ROC -AUC (Receiver Operating Characteristic – Area Under the Curve) curve is also done for examining of the sensitivity and also specificity trade-off. It also estimates the differentiation among the PD and also healthy individuals , not only that but will also check for consistency in its operation . Hyperparameter tuning is another major evaluation strategy that is used in the PD detection that is used to prevent overfitting.

CHAPTER -4

The implementation process for the hybrid Parkinson's model is discussed in this chapter. It also includes exploratory data analysis, selection, transformation of the data along with outlier capping and other methods. This chapter also includes the architecture of the model used along with the training and testing of the model to analyze the voice features to classify the disease.

Implementation

4.1. Libraries Used

S.No	Library	Usage
1	Pandas	Data Loading, Manipulation, and pre-processing
2	NumPy	Numerical Computations
3	Seaborn	Statistical data visualization for exploratory data analysis
4	Matplotlib	Plotting and visualization
5	Scikit-learn	ML utilities, data splitting, scaling, and encoding
6	TensorFlow	DL framework for building and training models
7	Sklearn	Splitting data, validation, and normalization

Table 5: Libraries Used

4.2. Exploratory Data Analysis

The building block of any Machine Learning or Deep learning project is the quality of data and for the Parkinson's disease detection the principal is crucial. As explained in chapter 3.1 the data set used is the *parkinsons_UPDRS* dataset that contains a variety of biomedical voice measurements that are obtained from various voice recordings of Parkinson's patients .

This dataset includes various features such as *Jitter, Shimmer, HNR* and other various frequency based metrices. The panda's data frames are used to load the dataset and extract the data from the dataset. The *read_csv()* function is used to load the dataset.

```
df = pd.read_csv('/content/parkinsons_updrs.csv')
```

After which the initial data inspection was carried out and the data set was completely explored by using various functions such as *head()* – which was used to view the first five rows of the dataset. A brief overview of the dataset is presented below:

subject#	age	sex	test_time	motor_UPDRS	total_UPDRS	Jitter(%)	Jitter(Abs)	Jitter:RAP	Jitter:PPQ5	Jitter:DDP	Shimmer	Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	Shimmer:APQ11	Shimmer:DDA	NHR	HNR	RPDE
1	72	0	5.6431	28.199	34.398	0.00662	3.38e-05	0.00401	0.00317	0.01204	0.02565	0.23	0.01438	0.01309	0.01662	0.04314	0.01429	21.64	0.41888
1	72	0	12.666	28.447	34.894	0.003	1.68e-05	0.00132	0.0015	0.00395	0.02024	0.179	0.00994	0.01072	0.01689	0.02982	0.011112	27.183	0.43493
1	72	0	19.681	28.695	35.389	0.00481	2.462e-05	0.00205	0.00208	0.00616	0.01675	0.181	0.00734	0.00844	0.01458	0.02202	0.02022	23.047	0.46222
1	72	0	25.647	28.905	35.81	0.00528	2.657e-05	0.00191	0.00264	0.00573	0.02309	0.327	0.01106	0.01265	0.01963	0.03317	0.027837	24.445	0.4873
1	72	0	33.642	29.187	36.375	0.00335	2.014e-05	0.00093	0.0013	0.00278	0.01703	0.176	0.00679	0.00929	0.01819	0.02036	0.011625	26.126	0.47188

Then the function *info()* which was used to understand the values and shapes of the dataset also the dimensions of the data frame is as follows:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5875 entries, 0 to 5874
Data columns (total 22 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   subject#        5875 non-null   int64  
 1   age             5875 non-null   int64  
 2   sex             5875 non-null   int64  
 3   test_time       5875 non-null   float64
 4   motor_UPDRS     5875 non-null   float64
 5   total_UPDRS     5875 non-null   float64
 6   Jitter(%)       5875 non-null   float64
 7   Jitter(Abs)     5875 non-null   float64
 8   Jitter:RAP      5875 non-null   float64
 9   Jitter:PPQ5     5875 non-null   float64
 10  Jitter:DDP      5875 non-null   float64
 11  Shimmer         5875 non-null   float64
 12  Shimmer(dB)     5875 non-null   float64
 13  Shimmer:APQ3    5875 non-null   float64
 14  Shimmer:APQ5    5875 non-null   float64
 15  Shimmer:APQ11   5875 non-null   float64
 16  Shimmer:DDA     5875 non-null   float64
 17  NHR             5875 non-null   float64
 18  HNR             5875 non-null   float64
 19  RPDE            5875 non-null   float64
 20  DFA             5875 non-null   float64
 21  PPE             5875 non-null   float64
dtypes: float64(19), int64(3)
memory usage: 1009.9 KB
```

Initial analysis revealed that the dataset structure consisted of *5,875* entries and *22* columns all of which contain numeric values with no missing data present.

```
df.shape
```

```
(5875, 22)
```

Following the initial phase , the exploratory data analysis (EDA) was conducted to detect various patterns and relationship between the data. The *total_updrs* is the key indicator of Parkinson's severity and it can be selected as the target variable as it provides the more evaluation of the patient's condition. Since *motor_updrs* focuses only on motor symptoms such as tremors, rigidity, and bradykinesia , *total_updrs* includes both motor and non-motor components that is the cognitive function , speech and autonomic issues that are also important for the detection of the disease.

Since the data set itself had a yield of extracted acoustic features from the patient's voice samples , which contains both motor and non-motor symptoms , selecting the *total_updrs* would provide with the best use of the available data.

Moreover, in terms of clinical point of view selection of *total_UPDRS* would align properly with the objective of developing a better system for the diagnosis. Training the model with *total_UPDRS* guarantees the models prediction to hold the entire set of patient's current signs and provides better results

Furthermore, the *total_UPDRS* was examined using *Value_Counts()* which was used to see the frequency of each score . This helped to understand the relation between different features , and a correlation matrix was also computed using the *corr()* function . This was in turn represented as a heatmap, the graphical representation of the relationships between the variables and the direction of strength of the relationships being much clearer. This is an important step as it assists in detecting highly corelated features that may be redundant and could be discarded to enhance the model performance . It was found that the features such as *motor_UPDRS* and other features of *jitter* and *shimmer* were also highly corelated. Also features with very high corelation were indicated as redundant(corelation coefficient > 0.9)

The image below illustrates the correlation relationship among the variables:

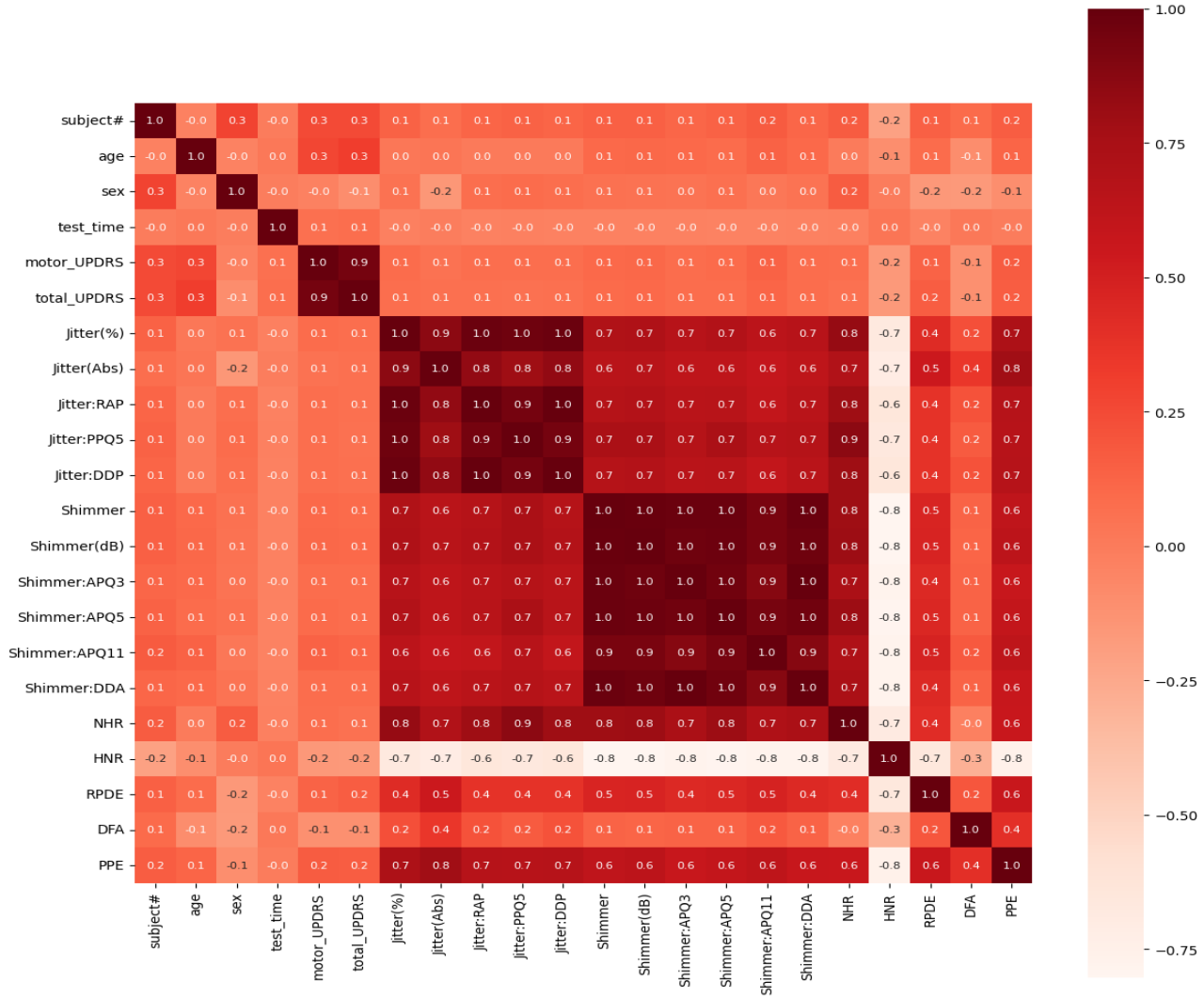


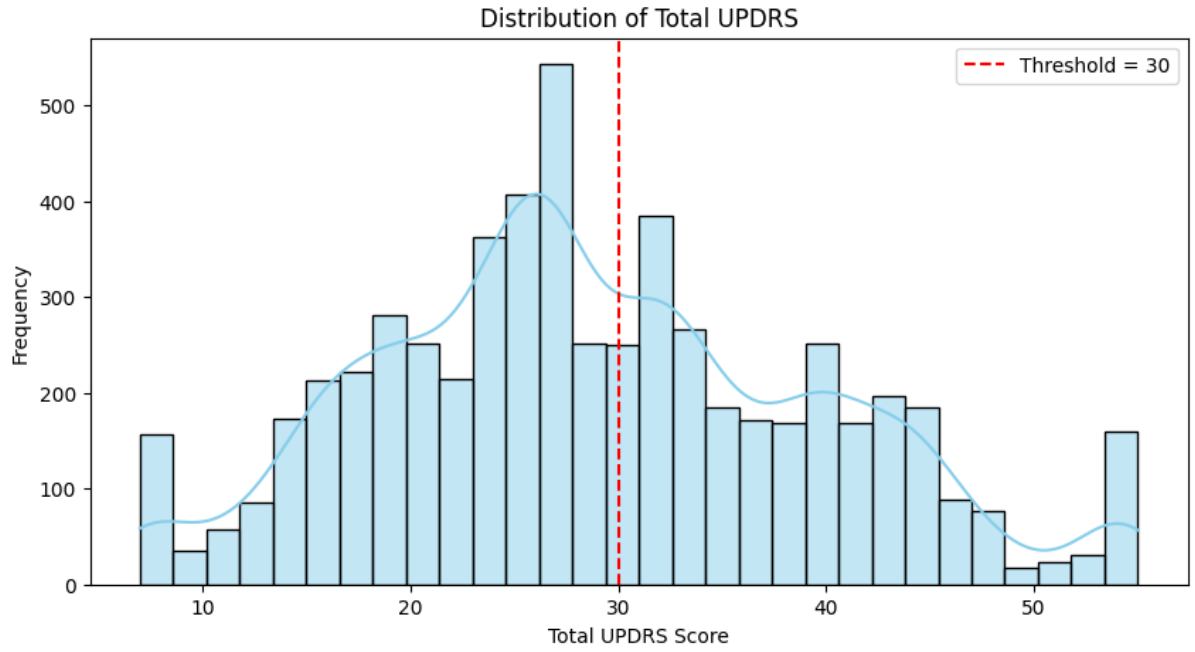
Figure 3: Correlation Heatmap

4.3. Target Variable Selection and Transformation

In comparison to *motor_UPDRS* which only focused on the motor symptoms , the *total_UPDRS* consists of both motor and non-motor components such as speech and cognitive impairments and also it is extremely clinically comprehensive for detection in early stage .

To prepare the set of data for classification, the *total_UPDRS* was treated as categorical variable , reclassifying the result as a series of categorical labels. This helps the model to assign the patient conditions to discrete severity classes which make it easy for interpretation and

clinically valuable. Moreover, a histogram was plotted to obtain the visual representation of the pattern of the scores and after making inferences about the pattern of the scores and the density , binning was adopted. The histogram was also used to find the threshold of the *total_UPDRS* which was later used to convert into binary.



The red threshold was line was marked at a score of 30. This was chosen as a clinical heuristic to separate the patients with milder symptoms (<= 30) and from those with more severe conditions (>30)

NHR	HNR	RPDE	DFA	PPE	updrs_group
0.014290	21.640	0.41888	0.54842	0.16006	Mild
0.011112	27.183	0.43493	0.56477	0.10810	Mild
0.020220	23.047	0.46222	0.54405	0.21014	Mild
0.027837	24.445	0.48730	0.57794	0.33277	Mild
0.011625	26.126	0.47188	0.56122	0.19361	Mild

Furthermore, a binary transformation was used to accommodate the subtype of clinical score to classification. An additional column was generated to differentiate between two categories . Patients with *total_UPDRS* =30 and below 30 were included into Class 0 with low severity and that with above 30 were included into Class 1 with high severity. This helps with the easy detection and classification of the disease.

Shimmer:APQ5	Shimmer:APQ11	Shimmer:DDA	NHR	HNR	RPDE	DFA	PPE	updrs_group	label
0.01309	0.01662	0.04314	0.014290	21.640	0.41888	0.54842	0.160060	Mild	1
0.01072	0.01689	0.02982	0.011112	27.183	0.43493	0.56477	0.108100	Mild	1
0.00844	0.01458	0.02202	0.020220	23.047	0.46222	0.54405	0.210140	Mild	1
0.01265	0.01963	0.03317	0.027837	24.445	0.48730	0.57794	0.332770	Mild	1
0.00929	0.01819	0.02036	0.011625	26.126	0.47188	0.56122	0.193610	Mild	1
...	...	...	...	...	...	...	...	...	...
0.02386	0.02937	0.05855	0.058288	20.301	0.62138	0.56444	0.258520	Mild	0
0.01913	0.02807	0.04779	0.055323	19.261	0.62820	0.58027	0.426715	Mild	0
0.02197	0.02714	0.05065	0.048822	20.339	0.57962	0.57977	0.272590	Mild	0
0.01185	0.01597	0.02722	0.009344	23.746	0.47321	0.57912	0.182230	Mild	0
0.02088	0.03510	0.05049	0.045591	18.553	0.65461	0.57819	0.335140	Mild	0

There were various factors that were taken into consideration in order to derive to the threshold they were:

1. The visual analysis of the *total_UPDRS* showed a separation around the score of 30, that served as an indication of division between mild and more severe cases.
2. Clinical benchmarks also suggested that the *total_UPDRS* scores in the range of 30 and above are associated with noticeable functional impairment and increasing severity .
3. Balanced class distribution was another factor . the provided threshold is evenly split between the classes which is crucial for prevention of class imbalance and to ensure the effective training of the model.

With the conversion into binary classification , the model was able to learn the distinctions and avoid overfitting and also generalize the unseen data. The conversion ultimately supports the objective of early-stage detection .

```
# Convert total_UPDRS to binary label with threshold 30
df['label'] = (df['total_UPDRS'] > 30).astype(int)

# Check class distribution
print(df['label'].value_counts())
print(df['label'].value_counts(normalize=True) * 100)
```

```
label
0    3341
1    2534
Name: count, dtype: int64
label
0    56.868085
1    43.131915
Name: proportion, dtype: float64
```

4.4. Data Pre-processing & Feature Extraction

Data Pre-processing is an important step in both Machine Learning as well as Deep Learning as is ensures that raw data is made ready to be properly trained on effectively and efficiently. The data usually contains noise and missing values , outliers and inconsistencies that can affect the performance of the algorithms. Pre-processing helps in cleaning and transforming and normalizing data that can ensure that the models are learning meaningful patterns. This is even more essential in the context of medical data.

One of the major issues that were considered in the course of pre-processing was the existence of outliers. Outliers are unusual values that deviate significantly from the rest of the data . this is caused due to the measurement errors , data entry mistakes or natural variability in patient responses. Although some of the outliers may indicate actual difference in the conditions of patients most of them are just noises and may make it difficult to interpret such statistical measurements as the average , standard deviation, and gradient of the model. Thus, the task of identifying them is such a necessity.

In order to detect and handle outliers , an Interquartile Range (IQR) method is used , which is one of the best techniques for identifying extreme values. This method usually uses the differences between the third quartile (Q3) and the first quartile(Q1) of a dataset. These quartiles usually represent 75^{th} and 25^{th} percentiles respectively.

The IQR can be calculated as IQR= Q3-Q1. Any data points that lie below Q1-1.5*IQR or above Q3+1.5*IQR is considered as an outlier. This method is widely used in medical and biological datasets where the data variance is high.

Excluding the target variable , IQR based capping was applied to all the numerical features in the dataset. Upper bounds and lower bounds were calculated based on the IQR and then the values that fell below the bound were replaced with the lower bound value and that above upper bound were capped at upper threshold. This helps in maintaining the overall spread of data and the shape of the data Excluding the outlier rows would lead to loss of data especially in smaller data set sizes and capping retains the instances but restricts leverages.

```
cap_features=[col for col in df.select_dtypes(include='number').columns if col not in ['total_UPDRS', 'index']]
for col in cap_features:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    # Cap the values
    df[col] = np.where(df[col] < lower_bound, lower_bound,
                np.where(df[col] > upper_bound, upper_bound, df[col]))

print("Outliers capped successfully using IQR method.")
```

The major reason for IQR capping over other methods like Z score was non gaussian distribution . In EDA with the visual inspection using boxplots and violin plots it was observed that majority of the variables were non gaussian distribution with skew in the sense. Identification on the basis of the z score in such situation can be highly misleading. The IQR method is a non-parametric distribution free method and thus best applicable to the dataset.

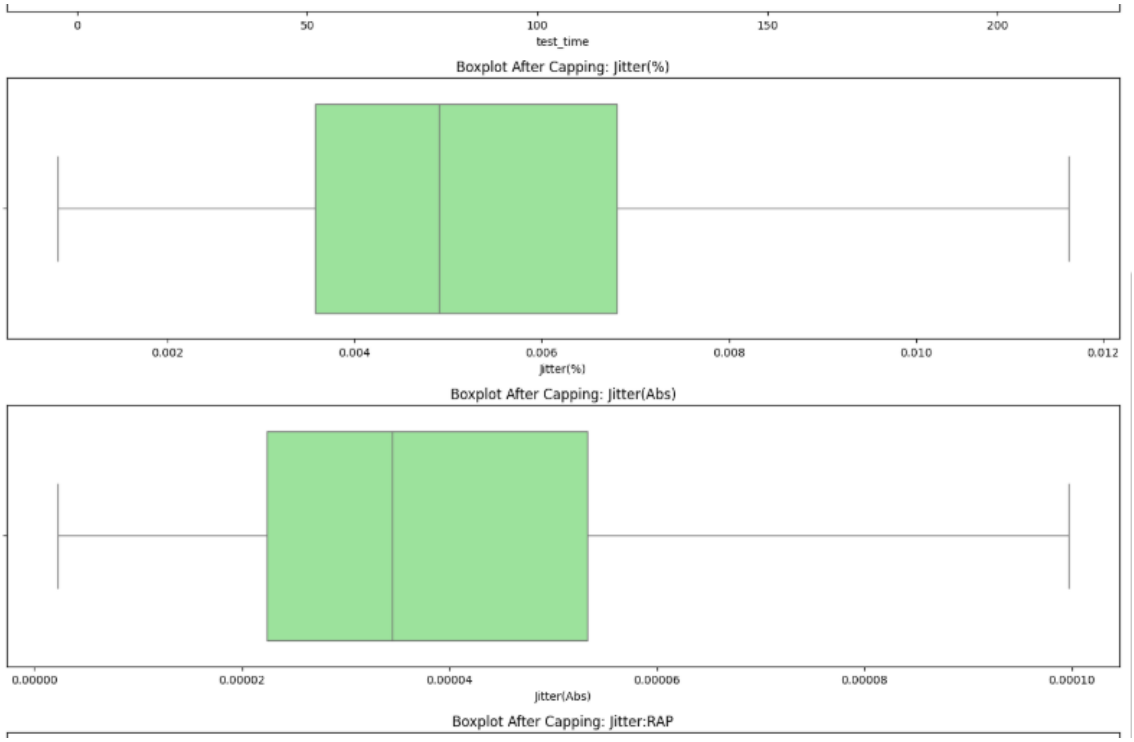


Figure 4: Outlier Capping

Moreover, since the dataset was used for binary classification task ,the pre-processing did not distort the class boundaries or the semantic meaning of the target variable. It was also ensured that *total_UPDRS* was excluded. This maintained the clinical validity of the labels while ensuring cleaner , more consistent input features for the model. Additionally, this was performed after the initial feature reduction and EDA to avoid unnecessary operations that were already discarded.

4.5. Feature Scaling Using Standardization

Feature scaling is an essential step in any Machine learning that ensures that all the numeric inputs are of the same scale. While working with a Convolutional Neural Network (CNN) this is necessary as the model's performance depends on the magnitude of feature values. Since the dataset contains a lot of acoustic features such as shimmer, jitter, and HNR , scaling plays an important role in accelerating the training process.

Since the dataset contains a lot of acoustic features, they have significantly different value scales . Without scaling features with larger ranges could dominate in the learning process and could also cause bias in the models learning. In order to prevent this from happening we use standardization technique that centres the data around zero with standard deviation of one.

This technique transforms each feature 'x' using formula

$$x' = \frac{x - \mu}{\sigma}$$

Equation 1

Where

μ- is the mean of the feature

σ - is the standard deviation of the feature

Standardization ensures that each features distribution becomes approximately normal standards which is suitable for the models. One of the major reasons for choosing the Standard Scaler is its robustness, that preserves the gaussian data. Since features were well distributed after the outlier removal the usage of standard scaler was thought to be the most appropriate one

```
scaler=StandardScaler()
scaler.fit(X_train)
X_train=scaler.transform(X_train)
X_test=scaler.transform(X_test)
```

4.6. Train test split

In order to prevent overfitting , the dataset was divided into training and test set. This ensures that the model learns the patterns from one subset of data and is evaluated on the test set. Scikit lean's *train_test_split* function was used in order to split it into train test subsets. The dataset was divided into 80:20 ratio where 80 % was allocated to the training set and 20% was to test set. It was also ensured that the data leakage was avoided as no information in the test set was used to train the model .This ensures that scaler is only fitted on the training data in order to avoid any information from the test set leaking into the model during training.

Columns such as *total_UPDRS, label, updrs_group* were excluded from the input features so that the model trains and prevents any data leakage from test set . Moreover, since *total_UPDRS and updrs_groups* were highly correlated, and their inclusion could provide the model with information about the outcome, they were dropped.

```
X=df.drop(columns=['total_UPDRS','label','updrs_group'],axis=1)
y=df['label']
```

```
X_train,X_test,y_train,y_test=train_test_split(
    X,y,test_size=0.2,stratify=y,random_state=2)
```

4.7. Model Architecture and Model Training

The Model that is developed is a hybrid deep learning architecture that combines CNN with Transformer . This was selected as it extracts both local patterns and long-range dependencies in biomedical data such as voice signals.

This is designed for binary classification, where the severity of the patient is classified which is derived from the UPDRS scores . The model is provided with an input shape of (19 1) where 19 features per sample across 5,875 from the dataset .

19 features were takes as input from 24(including *label* and *UPDRS_group*) where subject#, age and *sex, motor_UPDRS* and *total_UPDRS* were dropped. The retained features were *jitter*, shimmer, *NHR* and other acoustic features

These features were reshaped to (samples 19,1) to be compatible to the 1D CNN , which would enable the model to learn over feature space. Each sample then becomes a 1D signal with 19-time steps and single channel.

Model: "functional"

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 19, 1)	0	-
conv1d (Conv1D)	(None, 17, 64)	256	input_layer[0][0]
max_pooling1d (MaxPooling1D)	(None, 8, 64)	0	conv1d[0][0]
dropout (Dropout)	(None, 8, 64)	0	max_pooling1d[0]…
flatten (Flatten)	(None, 512)	0	dropout[0][0]
reshape (Reshape)	(None, 1, 512)	0	flatten[0][0]
multi_head_attenti… (MultiHeadAttentio…	(None, 1, 512)	131,776	reshape[0][0], reshape[0][0]
add (Add)	(None, 1, 512)	0	reshape[0][0], multi_head_atten…
layer_normalization (LayerNormalizatio…	(None, 1, 512)	1,024	add[0][0]
flatten_1 (Flatten)	(None, 512)	0	layer_normalizat…
dense (Dense)	(None, 64)	32,832	flatten_1[0][0]
dropout_2 (Dropout)	(None, 64)	0	dense[0][0]
dense_1 (Dense)	(None, 1)	65	dropout_2[0][0]

Total params: 165,953 (648.25 KB)
Trainable params: 165,953 (648.25 KB)
Non-trainable params: 0 (0.00 B)

The model consists of a single input layer which accepts voice data. The input shape is (19,1), that represents 19 consecutive voice features each as a scaler value. It also has seven hidden layers that is divided into various blocks. The first processing stage consists of the CNN and

pooling layers, that is designed to capture the temporal patterns in the voice signals. The **Conv1d** layer consists of 64 filters with a kernel size of 3. Each of the 64 filters acts as a neuron that is specialized for detecting specific small-scale temporal features in the signal. This is also the first hidden layer as it applies 64 1D convolutional filters across the input voice sequence . This layer extracts local temporal patterns from the raw data and it also learns various frequency changes and pitch variations. It also acts as a primary feature extractor by detecting small variations that are relevant for the disease detection.

Following the Conv1D layer is the **Max-Pooling 1 D** layer. It reduces the sequence length by half from 17 to 8 by preserving the number of channels at 64. This layer mainly down samples the output of the convolutional layer by selecting a maximum value in non-overlapping window of size 2. One of the major functions of the layer is that it reduces the temporal dimensions from 17 to 8 and also decreases the computational cost and helps model to focus on various features. A **Dropout laye**r of rate 0.3 will follow and randomly disable 30 % of the training neurons, reducing overfitting and thus it will make the network to learn more powerful feature representations. This layer randomly sets 30% of the activations to zero during training and also prevents overfitting by making the model less reliant and helping it to learn more features.

The CNNs output is then passed to a **Transformer block** of a 1D vector of the size 512 , that is reshaped into a sequence with the length of 1 and a feature size of 512. It consists of a **Multi head Attention Layer** which applies self-attention on the flattened and reshaped feature vector from the CNN block. The main function of this layer is to capture contextual dependencies and relationship between the features learned by CNN. This is main for understanding complex pattens in voice data that can indicate Parkinson's disease. The Multi-Head Attention Layer consists of 2 attention heads and has key dimension of 32 that allows the model to focus on different representation subspaces in parallel. This layer helps in computing the contextual relationships within the features by weighing their importance relative to each other. The output then remains (1,512) in order to preserve the dimensionality. In order to preserve the important information from the earlier stages , an **Add Layer** is used. This layer adds the original input features to the output of the Multi head attention layer. This helps in reducing the vanishing gradient problem and ensures better gradient flow. As this combines the original input features with the output from the attention mechanism , it helps in mitigating the vanishing gradient descent problem.

This is then normalized by using a Normalization Layer, that standardizes activation for faster convergence and also greater model stability. The **Normalization Layer** standardizes activation for faster convergence and also grater model stability is applied to normalize the combined output along the feature dimensions. This layer normalizes the summed feature from the add layer along the feature dimension. This also speeds up the training.

```
# Transformer Block
attention_output = MultiHeadAttention(num_heads=2, key_dim=32)(x_seq, x_seq)
attention_output = Add()([x_seq, attention_output])
x_trans = LayerNormalization()(attention_output)
x_trans = Flatten()(x_trans)
```

Finally, the output is then flattened back to a 512 vector to feed into the dense layers. Following that is a **Dense Layer** that consists of ReLu as the activation function and takes the normalized output of the transformer block and learns high-level nonlinear combination of the features. A **Dropout Layer** that randomly disables 30% of the neurons during the training to reduce overfitting and increases generalization is then added. It is followed by an output layer. This ends with a single **output Layer** that consists of one neuron with a sigmoid as the activation function .It then predicts the output of as representing the likelihood of the disease .

CHAPTER -5

The results of the hybrid CNN-Transformer are presented in this chapter for classifying the severity of the disease. It is evaluated using various measures such as accuracy, F1 score, precision, recall and AUC along with training and validation curves. The chapter also includes an overview of the Streamlit dashboard for demonstrating the model's applicability and predictions.

Results

5.1 Model Performances and Metrices

The study used hybrid CNN-Transformer model to classify the Parkinson's disease based on its severity using the acoustic features. The CNN block was used for feature extraction to extract the features out of the voice data and the Transformer layer was used to learn the temporal dependencies.

Adam optimizer along with cross entropy as the loss function was used to train the model. To improve the generalization and to prevent overfitting techniques such as dropout and layer normalization was used. The training was done for 20 epochs and the results in terms if accuracy and F1 score was noted.

5.2. Training and validation Performance

The training and validation of the model was assessed by accuracy and the loss metrics was recorded over various epochs. The accuracy curve showed the models ability to classify while the loss curve showed the models optimization nature. The models training progression can be inferred from both the model accuracy and model loss graph. They help in providing much

more insights into the models learning behaviour. Tracking these curves allowed the evaluation of models learning stability and generalization performance.

5.2.1. Accuracy Curves

The model learning was monitored on the basis of both the training and validation datasets across various epochs. The models accuracy initially began at 60% and it showed an increasing trend in moving forward through various epochs and over subsequent epochs , the accuracy reached around 75%.

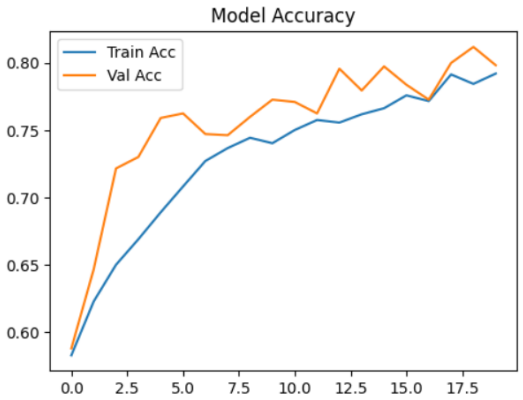


Figure 5: Model Accuracy

The figure, model accuracy shows both the training and validation accuracy through the entire training process. Some of the major findings are:

It is observed that during the initial epochs the training accuracy rises from 0.60 to 0.75. The validation accuracy also follows the same from 0.65 to 0.75 thereby indicating that the model is learning the informative features from the dataset.

It can also be noticed that the validation accuracy also improves, even though it shows a small fluctuation due to the small dataset it rises nearly to 80 and after which it plateaus with minor dips and recoveries. In the final stages it can be noticed that the accuracy rises, and the validation accuracy stabilizes. There is a gap between both, and it is moderate suggesting that there is minimal overfitting. Overall, the curve depicts a good pattern, and the model has captured the data without much divergence between training and validation performance

5.2.2. Loss Curves

In parallel to the accuracy the binary cross entropy was also calculated. The loss plot provides much more added perspective to the accuracy curve as it traces the competitive consistency of the model to minimize the binary cross entropy loss during the training phase. They are very important indicators of the model optimization and convergence.

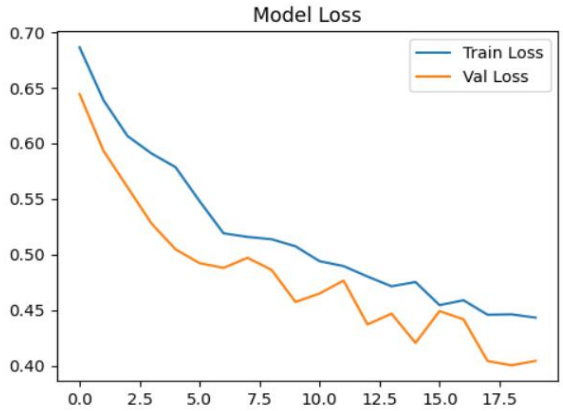


Figure 6: Model Loss

In the initial training phase, the loss decreased with a pleasing rate, during the middle training phase both the training and the validation losses continued to decline gradually. The validation loss at times fell lower than the training loss and this could be explained by good regularization methods such as dropout and layer normalization that ensures generalization. During the final phase the loss curve started to tend towards each other as the training reaches the last epochs. This also ensured that there is no serious overfitting of the model.

5.2.3. ROC curve and AUC Score

The ROC curve provides the view of the ability of the model to classify the diseases based on the patient's condition to mild or severe. This is done by plotting the true positive rates against the false positive rates and this in turn helps to evaluate the model performance across a range of decisions.

The AUC curve was found out to be 0.90 that indicated the excellent diagnostic capability. This is an indication that the model will have 90 % likelihood of giving a higher severity score on a randomly chosen sample of a patient with symptoms of Parkinson's disease than a healthy person or one with mild symptoms.

In the beginning the ROC curve has a very steep increase, that indicates that at a low false positive rate the model is the one with hight sensitivity rate. The curve is in the upper left quadrant indicating good performances and it slowly moves towards the top limit where the sensitivity is close to 1.0. This strength is observed in the features of the model to keep the two classes separated and performing equally towards each class.

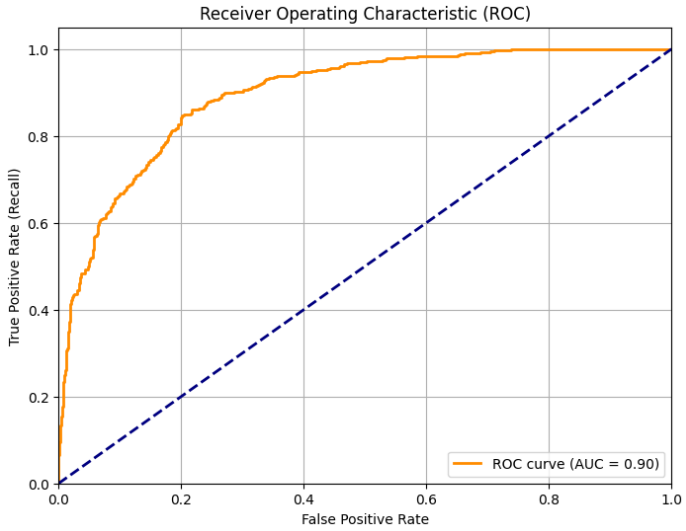


Figure 7: ROC Curve

5.3. Final Test evaluation

The final test evaluation was performed on the test data to provide an unbiased estimation of model's real-world performance. The findings depicted that the model had 78.5 % classification ratio and other performances measures such as precision, recall and F1 score were also calculated to describe the trade-off between detection of positive cases and false ones.

5.3.1. Accuracy

The accuracy of the model was evaluated using various standards and other classification metrices and it was 78.55%, which indicated that the model correctly classified 78.55% of the test samples into their respective categories. This also indicated that 78 out of all the 100 test samples were correctly classified into their respective categories. Moreover, high accuracy also suggested that the model could effectively capture the patterns in the voice data that is relevant to the disease detection.

```
Test Accuracy: 0.7855
```

5.3.2. F1 Score

The precision and recall are balanced by the F1 score. The harmonic mean of precision and recall is known as F1 score and it was found to be 74.32%. The F1 score is an important metric as it as it is applicable when the class is skewed because the algorithm works well for two classes especially the minority class. The higher the value of F1 score the better the recall and precision is, thus confirming the model's good capability of capturing the positive cases well.

```
Test F1 Score: 0.7432
```

5.3.3. Precision and recall

The F1 score is calculated using both the precision and the Recall.

- Precision: This shows the proportion of predicted positives that are actually positive
- Recall: The recall shows the proportion of actual positives correctly identified by the model. These values were obtained from the confusion matrix that is explained in the below subtopic.

5.3.4. Confusion Matrix

In order to obtain a much better understanding of how the model performs a confusion matrix is computed on the test set. This usually breaks down the predicted classification vs the actual classification into the categories of

	Predicted Mild (0)	Predicted Severe(1)
Actual Mild (0)	True Negative (TN)	False Positive (FP)
Actual Severe(1)	False Negative(FN)	True Positive(TP)

Table 6: Confusion Matrix

- **True Negative (TN)** : This is where the patient actually has mild symptoms and the model correctly predicted them as mild . The model correctly identified the patients with mild symptoms(class 0) and predicted them as mild A higher TN count reflects the model's strength

- **True Positive (TP)** : This is when the model correctly identifies the cases of moderate to severe . This is when the patients in class 1 were correctly identified by the models correctly .

- **False Positive(FP)** : This is when the patients who belonged to class 0 ; who were actually mild, and the model incorrectly classified them in as moderate or severe.

- **False Negative(FN)** : This is when the patients who actually belonged to group of class 1 that had moderate to severe symptoms were mistakenly predicted as mild . This error is very crucial and can also cause delays in the intervention and better treatment.

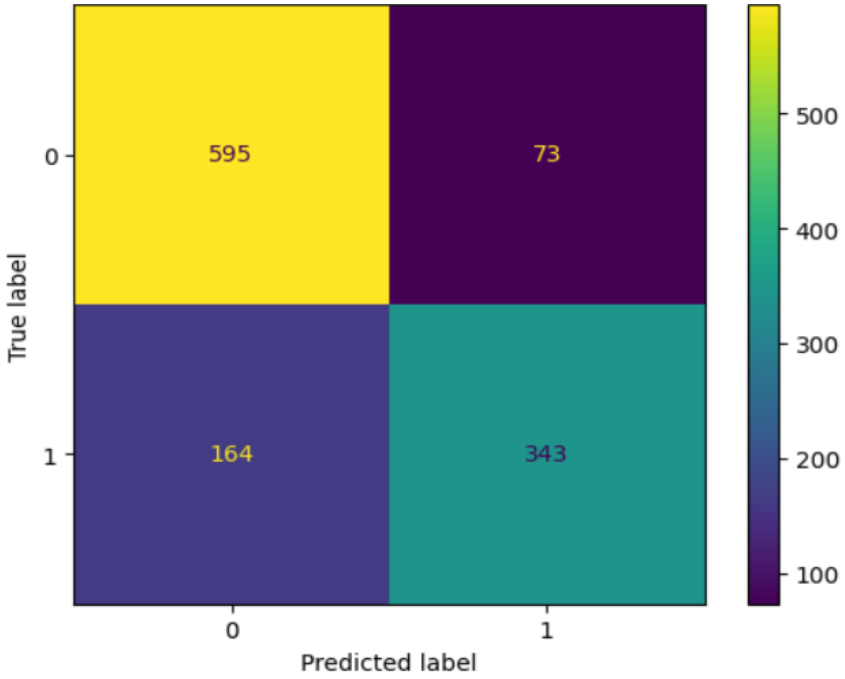


Figure 8: Confusion Matrix

5.4. Streamlit Dashboard

To increase the accessibility and real -world applications, a Streamlit dashboard was developed, this acts as an interface of the real-time prediction of severity of the disease. The idea is to have a user-friendly platform that the clinicals and the researchers could use the tool to make predictions without becoming technically knowledgeable. This interface takes the acoustic measurements of the values of patients as the input and then uses the model to predict it and classify it into categories more over user can also upload the dataset that contains the input features and then the model predicts the output.

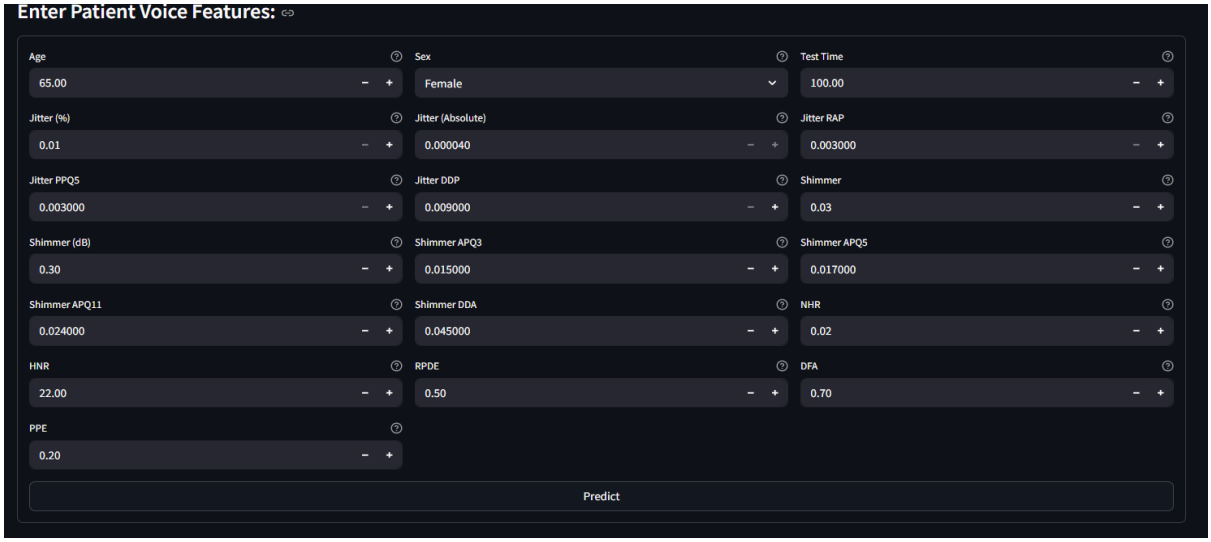


Figure 9: User Interface 1

The dashboard was developed with keras model that is *parkinsons.keras* and a standardscaler object *(scaler.pkl)*. The trained model was first saved in the *.keras* format using the *model.save ("parkinsons.keras")* by preserving the architecture, learned weights for deployment. This then stores the complete model architecture , weights and also the optimizer configuration , for seamless deployment. The dashboard allows the user to input all the voice features directly into the UI. In addition to that it also allows the user to upload an existing data set and the model can detect the data set and give the prediction based on it. The dataset used is an already existing dataset and the model provides the score based on the results of it.

Figure 10: User Interface 2

As soon as the user submits the inputs the scaler transforms the data to the same scale as feature training. The model then predicts the Parkinson's severity score. The results are shown in the

form of the chart that depicts the severity score of the disease. This implementation ensures seamless user experience and better prediction functionality.

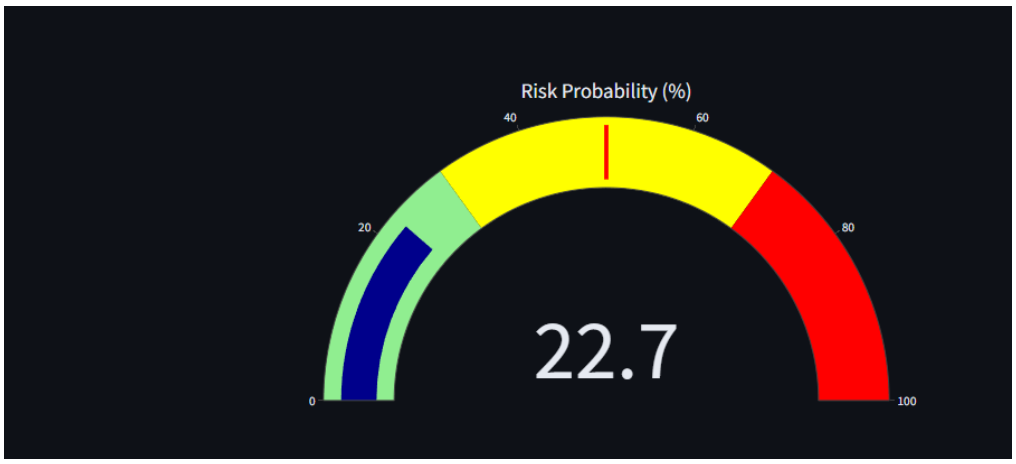


Figure 11: Prediction-Low Risk

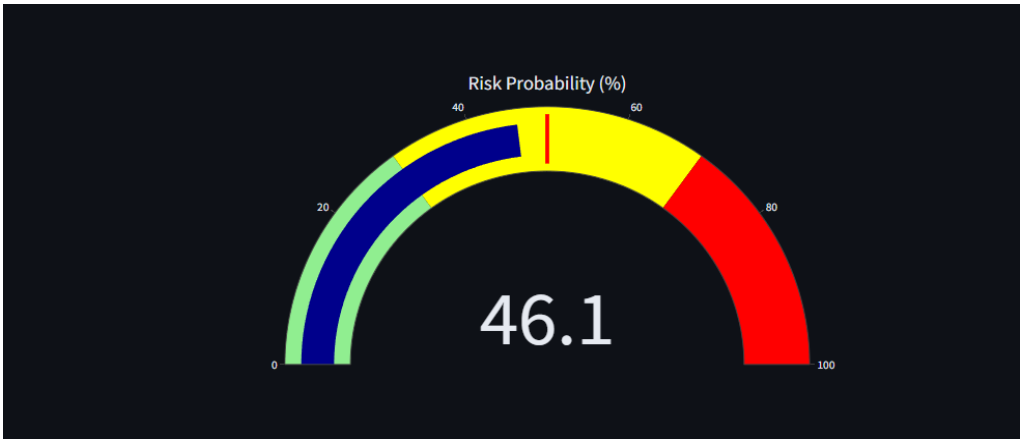


Figure 12: Prediction - High Risk

5.5. Discussion of findings

The hybrid model that was trained on the classification based on the voice records showed really good predictive ability. The performance metrices that showed the accuracy and the F1 score was 78.55 and 74.32 respectively which also showed a good trade-off between the mild and severe cases compared a baseline model of Logistic Regression which was 61.7% and

51.0% respectively. The accuracy and F1 of hybrid model were compared to the Logistic Regression model and it turned out to be the better performer than the LR model.

```
Accuracy of LR:  0.6170212765957447
F1 score of LR:  0.5108695652173914
```

The accuracy vs training and validation curves of hybrid CNN-Transformer model indicated that there was consistency during learning. The model was performing better with training and was performing very minimal overfitting with the validation accuracy close behind the training curve as the epochs were progressed to 20. Similarly, the loss curve also demonstrated the decreasing tendency of training and validation loss , confirming that model was predicting the significant patterns .

One of the advantages of the model was that the ROC curve yielded to 0.90. This meant that the model performed excellently to classify the different classes. The shape of ROC curve with its rise in dominance and the upper left quadrant shows the models strong sensitivity and specificity trade-offs.

Finally, the Streamlit dashboard also ensured that it allowed real time data prediction through a dashboard, it’s also an interactive tool to assess the severity from the patient’s voice data.

CHAPTER – 6

The limitations, future scope and performance of the hybrid model is discussed in this chapter. This chapter gives a detailed overview of the future work that could be carried on with the hybrid model and about the discussions and conclusion on the model.

Discussion and Conclusion

The hybrid CNN-Transformer was proposed to categorize the severity of the Parkinson's disease based on the acoustic feature into mild or severe. The model depicted an accuracy of 78% and the F1 score of 74% and performed much better in terms of accuracy and other metrices compared to the LR model. These implied that the model performed the best in classification of the disease into its respective categories. Other than these, several other metrices were also used such as training and validation curves, confusion matrix and ROC analysis. In addition to this a user friendly Streamlit dashboard was also developed.

While the study achieved good output and results, there were certain factors that should be considered while looking into its findings. Although the dataset was sufficient for approach, it was limited in size and diversity especially in representing the moderate to severe cases. Additionally recent works focused mainly on the acoustic features rather than providing the additional information such as demographic details, medical history, or any other neurological assessments. Moreover, the evaluation was also conducted on the data that was obtained from a single source and the usage of techniques such as SHAP or Grad-CAM was also not included.

The future developments could include the aspect of expanding the dataset into a much wider range of participants across various demographics, languages, and various stages. Multi source and multi modal data combined with voice features including Electroencephalography (EEG) signals, gait patterns brain imaging could also be used to develop a more complete and useful diagnostic system. Rather than just binary classification, multiple class labelling system could also be created as it differentiates between similar neurological disorder. Real world deployment would also require interpretability and clinical integration. Along with technical considerations, clinical validation research and patient educational initiatives and even ethics will be of importance in order to ensure a safe and responsible application of the model.

Finally, the study shows that hybrid CNN-Transformer could also serve as an effective approach for assessing the Parkinson's disease from the acoustic features. With the integration of deep learning, the model achieved a strong accuracy and other performance metrices and exhibited a good learning behavior. Finally, this study also suggests that AI based detection methods not only have technical potential but also have the ability to be implemented as a part of earlier intervention.

HERIOT WATT UNIVERSITY

REFERENCES

1. Tsanas, A. & Little, M. (2009). Parkinsons Telemonitoring [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5ZS3N.

2. Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., Halliday, G., Goetz, C.G., Gasser, T., Dubois, B., Chan, P., Bloem, B.R., Adler, C.H. and Deuschl, G. (2015). MDS clinical diagnostic criteria for Parkinson's disease. *Movement Disorders*, 30(12), pp.1591–1601. doi:https://doi.org/10.1002/mds.26424.

3. Kalia, L.V. and Lang, A.E. (2015). Parkinson's disease. *The Lancet*, 386(9996), pp.896–912. doi:https://doi.org/10.1016/s0140-6736(14)61393-3.

4. Deuschl, G., Antonini, A., Costa, J., Śmiłowska, K., Becker-Bense, S., & Berardelli, A. (2020). European Academy of Neurology/MDS-ES recommendations on the treatment of Parkinson's disease: 2020 update. Movement Disorders, 35(10), 1709-1719. https://doi.org/10.1002/mds.28215

5. A. Ouhmida, O. Terrada, A. Raihani, B. Cherradi and S. Hamida, "Voice-Based Deep Learning Medical Diagnosis System for Parkinson's Disease Prediction," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 2021, pp. 1-5, doi: 10.1109/ICOTEN52080.2021.9493456. keywords: {Deep learning;Neurological diseases;Databases;Parkinson's disease;Signal processing algorithms;Predictive models;Sensitivity and specificity;Parkinson's disease;Deep Learning;Acoustic dataset;CNN;ANN},

6. J. Naanoue, R. Ayoub, F. E. Sayyadi, L. Hamawy, A. Hage-Diab and F. Sbeity, "Parkinson's disease detection from speech analysis using deep learning," 2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME), Beirut, Lebanon, 2023, pp. 102-105, doi: 10.1109/ICABME59496.2023.10293142. keywords: {Deep learning;Training;Speech

analysis;Parkinson's disease;Feature extraction;Reliability engineering;Data models;long short-term memory (LSTM);Parkinson's disease;speech signals},

7. B. Fathima, L. N. S S, R. M and B. M, "Exploring the Efficacy of CNN Models of Diagnostic Modalities for Parkinson's Disease," 2024 International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2024, pp. 1-5, doi: 10.1109/ICERCS63125.2024.10895289. keywords: {Deep learning;Accuracy;Sensitivity;Scientific computing;Computational modeling;Functional magnetic resonance imaging;Resonance;Convolutional neural networks;Single photon emission computed tomography;Diseases;Convolutional Neural Networks (CNN);Deep Learning;Functional MRI (fMRI);Magnetic Resonance Imaging (MRI);Parkinson's Disease and Single Photon Emission Computed Tomography Dopamine Transporter (SPECT DAT)},

8. Hoops, S., Nazem, S., Siderowf, A. D., Duda, J. E., Xie, S. X., Stern, M. B., & Weintraub, D. (2009). Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson's disease. Neurology, 73(21), 1738-1745. https://doi.org/10.1212/WNL.0b013e3181c34b47

9. World Health Organization (WHO). (2022). Neurological disorders: Public health challenges. World Health Organization. https://www.who.int/publications/i/item/9789241565509

10. J. Li, X. Cao, B. Hou, X. Li and Q. -A. Ding, "A Deep Learning Hybrid Model for Parkinson's Disease Diagnosis Based on Electroencephalogram Signals," 2023 17th International Conference on Complex Medical Engineering (CME), Suzhou, China, 2023, pp. 55-58, doi: 10.1109/CME60059.2023.10565623. keywords: {Deep learning;Training;Accuracy;Recurrent neural networks;Parkinson's disease;Brain modeling;Feature extraction;Parkinson's disease;Deep learning;Convolutional neural network;Electroencephalogram},

11. E. Antikainen, P. Cella, A. Tolonen and M. van Gils, "SPECT Image Features for Early Detection of Parkinson's Disease using Machine Learning Methods," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 2773-2777, doi: 10.1109/EMBC46164.2021.9630272. keywords: {Computational modeling;Machine learning;Feature extraction;Software;Data models;Biology;Single photon emission computed tomography},

12. A. S. Aliero and N. Malhotra, "A Survey on Detection of Parkinson's Disease Through Clinical Data Using Deep Learning Approach," 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN), Indore, India, 2024, pp. 173-177, doi: 10.1109/CICN63059.2024.10847535. keywords: {Deep learning;Surveys;Accuracy;Parkinson's disease;Motors;Real-time systems;Data models;Recording;Long short term memory;Overfitting;Parkinson's disease;deep learning;early detection;convolutional neural networks;clinical data;machine learning},

13. "Parkinson Disease Detection using Feed Forward Neural Networks," 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RAEEUCCI57140.2023.10134449. keywords: {Deep learning;Machine learning algorithms;Parkinson's disease;Computational modeling;Feature extraction;Rigidity;Recording;Parkinson's Disease;Feed Forward Neural Networks},

14. M. Wodzinski, A. Skalski, D. Hemmerling, J. R. Orozco-Arroyave and E. Nöth, "Deep Learning Approach to Parkinson's Disease Detection Using Voice Recordings and Convolutional Neural Network Dedicated to Image Classification," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 717-720, doi:

10.1109/EMBC.2019.8856972. keywords: {Parkinson's disease;Databases;Deep learning;Feature extraction;Spectrogram;Biological neural networks;Audio recording},

15. M. Ogawa and Y. Yang, "Residual-Network -Based Deep Learning for Parkinson's Disease Classification using Vocal Datasets," 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), Nara, Japan, 2021, pp. 275-277, doi: 10.1109/LifeTech52111.2021.9391925. keywords: {Deep learning;Parkinson's disease;Conferences;Life sciences;Convolutional neural networks;Residual neural networks;Parkinson's disease;deep learning;residual network},

16. "Parkinson Disease Detection using Feed Forward Neural Networks," 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RAEEUCCI57140.2023.10134449. keywords: {Deep learning;Machine learning algorithms;Parkinson's disease;Computational modeling;Feature extraction;Rigidity;Recording;Parkinson's Disease;Feed Forward Neural Networks},

17. J. Prince and M. de Vos, "A Deep Learning Framework for the Remote Detection of Parkinson'S Disease Using Smart-Phone Sensor Data," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 3144-3147, doi: 10.1109/EMBC.2018.8512972. keywords: {Feature extraction;Accelerometers;Convolution;Kernel;Parkinson's disease},

18. N. SrinivasaRao, M. B. A. Khan, P. Haneesha and P. Meghana, "Pioneering Early Parkinsons Disease Detection with Deep Learning and Image Processing Innovations," 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2024, pp. 639-643, doi:

10.1109/ICACCS60874.2024.10716990. keywords: {Deep learning;Spirals;Accuracy;Parkinson's disease;Image processing;Transforms;Classification algorithms;Diseases;Residual neural networks;Testing;Residual Neural Network;Convolutional Neural Network;Deep Learning},

19. N. G and M. V. Karki, "Deep Learning Techniques for Parkinson's Detection Using EEG Signals Analysis," 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMITCON58196.2023.10275844. keywords: {Deep learning;Support vector machines;Time-frequency analysis;Parkinson's disease;Feature extraction;Electroencephalography;Convolutional neural networks;Parkinson's Disease;Electroencephalogram;Mel Spectrogram;Support-Vector machines;K-Nearest Neighbors;Random Forest;Convolutional Neural Network},

20. M. Ghasemi, R. Boostani and A. J. Lak, "A Brief Review on EEG-Based Diagnosis of Parkinson's Disease Using Deep Learning Models," 2024 E-Health and Bioengineering Conference (EHB), IASI, Romania, 2024, pp. 1-4, doi: 10.1109/EHB64556.2024.10805584. keywords: {Deep learning;Time-frequency analysis;Accuracy;Parkinson's disease;Reviews;Signal processing;Brain modeling;Feature extraction;Electroencephalography;Context modeling;Parkinson's Disease diagnosis;Deep Learning;Machine Learning;EEG},