Early Detection of Parkinson’s Disease Using Deep Learning Models

MSc Research Project

Data Analytics

Abhijeet Nautiyal

Student ID: 22165835

School of Computing

National College of Ireland

Supervisor: Prof. Jaswinder Singh

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Abhijeet Nautiyal		
Student ID:	22165835		
Programme:	MSc in Data Analytics		
		Year:	2024-2025.
Module:	Research Project		
Supervisor:	Prof. Jaswinder Singh		
Submission Due Date:	12-12-2024		
Project Title:	Early Detection of Parkinson’s Disease Using Deep Learning Models		
Word Count: 9700	**Page Count :21**		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Abhijeet Nautiyal
Date:	12-12-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Early Detection of Parkinson’s Disease Using Deep Learning Models

Abhijeet Nautiyal

x22165835

Abstract

The study presents the efficient deep learning models such as CNN-LSTM and CNN-GRU architectures to detect Parkinson's disease (PD) at an early stage through voice data. By employing both intelligent audio as well as nonlinear indicators like jitter, shimmer, and harmonic-to-noise ratio (HNR) using Recurrence Period Density Entropy (RPDE) and Detrended Fluctuation Analysis (DFA). The experiments for evaluating the model took place among default and optimized configurations of both CNN-LSTM and CNN-GRUs, with and without early stopping. Results denote that early stopping off considerably ameliorates all metrics and optimized CNN-GRU models (conv=64, gr=75,100) are the leading ones in the tests across all metrics. The CNN-GRU model was able to accurately predict 76.47% of the test data, reaching a F1 score of 77.78% and a balance of recall of 100% and a precision of 63.64%. These results pin down the CNN-GRU model's capacity to generalize and at the same time correctly identify. This study points out the fact that deep learning in voice diagnostics has the potential to become a scalable, non-invasive, and economical solution for early Parkinson's detection. The models that are built in this study offer a firm base for further integration into telehealth systems and thus making early diagnoses and personalized interventions more effective.

1. Introduction

Parkinson's disease is defined to be a chronic and progressive neurodegenerative disease, which mainly affects the brain's motor control centers. It is caused by the degeneration of dopamine-producing neurons in basal ganglia which is the region that is very important for regulating soft and coordinated movement. Dopamine is a neurotransmitter that facilitates the transmission of nerve impulses so the muscle is able to contract. It means that its shortcoming will affect the body's ability to perform the controlled and automatic movement processes properly. Classical symptoms such as tremors, muscle rigidity, and postural instability are very visible. Beyond motor symptoms, PD can also affect non-motor functions, such as cognition, mood, and sleep. While the definite cause of Parkinson's disease is not known yet, the most recent research implies that genetic predisposition and environmental factors combinations contribute to its beginning and subsequent course of development. PD is the second neurodegenerative disease that is the most widespread one in the world; as of now, it affects over 10 million people all over the world according to WHO (World Health Organization). The disease is more common in elderly people, so it is a problem of public health which the world is facing due to aging of the population. The above-named impacts are enormous from the point of view of the society as a whole. Besides the direct treatment costs and lost productivity, PD causes deep emotional suffering for both the patients and their care givers.

The techniques of voice analysis have appeared as a new method in which determining diseases in an easier way and with fewer invasions can be achieved. Several studies conducted lately examined the employment of acoustic analysis on speech signals and tones in the identification of conditions such as Parkinson's disease (PD). These voice-centric methods make it possible to distinguish between the healthy and the PD-affected persons effectively.

Along with the rise of super-intelligent systems and new diagnostic solutions that are based on voice data, clinicians now have a very powerful counter against the risks posed by the age-old diagnostic techniques. Thus, these methods can ease the process of making decisions for healthcare professionals, and, consequently, minimize the probability of failures and false-positive outputs. Besides, they are the ones that make possible more organized and timely patient medical follow-ups. The inclusion of a plethora of databases and voice data has been a very important factor for the classification-based approaches.

This work is a proposal of using a hybrid model that covers many kinds of technologies to be able to infer the rules and still have very good results. The interpretability of the model served for delivering valuable insights and knowledge from the voice data. These are the only two pieces you can find in speech. Health care providers are thus able to correct the diagnosis of PD with higher precision by means of these new methods, and consequently, patients are treated better and their outcomes are improved.

1.1 Research Question

The research question is how much the voice analysis would be capable of early detection of Parkinson's disease. Specifically, firstly, the study aims to investigate: With respect to voice tremor, pitch, variations, and prosody elements (NHR), *how well does the Parkinson symptom vocal analysis technology identify and measure the disease in its early stages*?

Secondly, *how can hybrid CNN machine learning models effectively process these vocal features to enhance diagnostic accuracy and reliability in real-world applications*?

1.3 Research Objective

The main aim of this research is to explore and prove the practicality of voice analysis for the early detection of Parkinson's disease, using machine learning models, especially hybrid CNN’s. The study is to explore the vocal weaknesses like pitch shifting, low quality speech, and the loss of speech fluency which subsequently leads to diagnosing PD at the early stage. The study of advanced acoustic features and machine learning techniques provides a new platform for the field of non-invasive, accessible, and economical diagnostic tools. These tools potentially can help to timely intervention, improve the patient outcomes, and to rely on the traditional, expensive, and clinical diagnostics less. It also, addresses the problems related to variability in recording conditions and the necessity of, balanced and huge datasets for generalizing over the populace. Overall, the goal is to make way for early detection which will guarantee a better life for patients and a reduction in the disease burden on the society.

2. Related Work:-

2.1 Advances in Acoustic Feature Analysis for Parkinson’s Disease Detection

Voice-based Parkinson’s disease (PD) detection has rocked the world at present, acoustics being the primary research tool for scientists to uncover the features of voice that may be related to the problem of vocalizations due to PD. Recent research has revealed that pitch-related parameters (mdvp_fo_hz), F0_med (mdvp_fhi_hz), and F0_max (mdvp_flo_hz) are key markers for Parkinson’s patients in the quantization of pitch-based impairments. It is these metrics that measure the mobility and range of the voice that suffer from motor impairments (Pah, N.D. et al.,). For example, the research of (Yuan, L., Liu, Y et al.,) revealed that the pitch-related characteristics are the main distinguishing features that help in the differentiation of PD and non-PD.

Therefore, let us dwell on the electrifying discovery of two very useful metrics - Jitter and Shimmer - which are the indicators of pitch variability and amplitude variability correspondingly. Jitter is a measure of frequency perturbations, whereas shimmer is the assessment of amplitude excursions. Both effects are characteristic indicators of PD disease (Sajal, M.S.R et al.). (Pah, N.D. et al.,), argued that although these metrics are used in combination with other acoustic features, they alone are sufficient lending a robust vocal-disability representation. For instance, the experiments conducted in (Lv, C., Fan, L., et al.,) demonstrated the importance of the jitter measures (mdvp_Jitter%, Jitter:DDP), the ones that are responsible for capturing the fine- grained irregularities of the Parkinsonian discourse.

The harmonic-to-noise ratio (HNR) and noise-to-harmonics ratio (NHR) are the indicators that give a measure of the 'breathiness' and 'hoarseness' of the speech. These features, which were represented by MDPI in the year 2021, are very necessary as they indicate the distinction between the patients who have Parkinson's disease and those who are healthy. HNR provides the measure how much harmonic is the main one in the speech while NHR reads vice versa - how loud prominence is the noise. They effectively use the development, for example, of a system of (El-Sayed, R.S et al.,), in almost every industrial application thus the importance of these two measures cannot be overemphasized, to say the least, 'they efficiently diagnose early vocal impairments.

Integration of multiple acoustic features like shimmer, jitter, and MFCCs along with HNR and RPDE, became a major development. Through the use of broad-feature sets, scientists have better the argument and explanation of ML models. The (Ouhmida, A., et al.,) and (Rizvi, D.R., Nissar et al.,) these are two of these features that were brought to bear. The ensemble models were the result of the accuracy data being over 95%. These studies provide the main evidence of the need for diverse feature sets to capture the multidimensionality of shortness of voice in PD.

Even though the progress is obvious, yet, the constancy of the appropriate feature could still be a problem across the variety of recording conditions. The microphone quality differences and the patients' environment add to the noise of the extracted features, thus making them less reliable (Rizvi, D.R., Nissar et al.). The future developments in this field have to take a leap forward in increasing noise robustness and generalization by applying methods like data augmentation and domain adaptation. Besides, the combination of the acoustic features with the other biomarkers, such as the gait or handwriting analysis, could even more enhance the systems accuracy of PD detection, as suggested (El-Sayed, R.S et al.).

2.2 Machine Learning Techniques in Voice Analysis for Parkinson’s Disease

The utilization of machine learning (ML) in voice analysis for Parkinson's disease (PD) has been largely reshaped to various methods that make use of acoustic features for early diagnosis. Conventional ML classifiers, such as Support Vector Machines (SVMs) and k-nearest neighbors (kNN), have proven to be effective in recognizing the most distinct and distinctive characteristics in the cases of jitter, shimmer, and HNR (harmonics-to-noise ratio). SVMs, in particular, are famous for their capability to manage complicated features spaces, which result in high-level accuracy as seen from a very recent study by (Khaskhoussy, R et al) and (Pah, N.D. et al.,). kNN is efficient in straightforward, normalized data sets, but its reliance on distances greatly restricts its generalizability (AIP Publishing, 2023).

Deep learning tech has really switched up the scene by launching automatic feature extraction abilities using convolutional and recurrent neural networks. For particle analysis and classification of structured matrices like MFCCs, Convolutional Neural Networks (CNNs) find spatial patterns in acoustic feature matrices easily, while recurrent architectures based on Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) can reliably model temporal dependencies. Hybrid models like CNN-GRU that combine CNNs and RNNs, among other architectures, have proven particularly effective in detecting PD-related vocal impairments according to the research report releases by (Vidya, B et al.,) and (Pahuja, G et al.,). These models integrate local and sequential features, which in turn lead to enhanced diagnostic precision.

Ensemble methods are a popular category of techniques in machine learning that typically involve a concept of boosting-based algorithms such as Adaboost, XGBoost, and more. They have been developed to the point where the sum of the benefits of the individual models is even greater than the models themselves. Other notable ensemble methods include random breaths (RF) and stacking methods that, given their robustness in the management of high dimensional data, have wide application. As a case in point, (Sorathiya, A, et al.,) and (Aşuroğlu, T. et al.,) showed through (Parisi, L., et al.,) and (Senturk, Z.K., et al.,) that ensembling can stabilize individual models and improve generalization across the board of datasets.

Feature selection and dimensionality reduction are still key solutions in achieving the best model performance. Methods like Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) have been used for determining the most appropriate set of criteria by removing the irrelevant features and keeping only the ones that are crucial. They can reduce overfitting and calculation bottlenecks which have been shown in (Rizvi, D.R., Nissar et al.,) and (El-Sayed, R.S et al.,) publications.

These advances notwithstanding, some issues remain with respect to the capability of voice recognition models to stay coherent within various recording circumstances and among different patient groups. Dataset variation and imbalances are still the main issues for model robustness which creates a need for huge and diverse datasets along with sophisticated data augmentation methods. Generative models like GANs for synthetic data generation should be studied for this purpose and new architectures to further improve diagnostic accuracy and accessibility should be investigated. This suggestion was made in the comment of (Rizvi, D.R., Nissar et al.,) to the journal (Pahuja, G et al.,).

2.3 Multimodal Approaches and Telemonitoring for Parkinson's Disease Detection

The latest advancements in Parkinson's disease (PD) research have stressed the utilization of multimodal approaches and telemonitoring systems for better diagnostic and patient monitoring processes. These methods integrate different data modalities, such as speech analysis, motor activity, and other bio-signals, to give a comprehensive view of PD's signs. Machine learning (ML) based multimodal approaches, on one hand, improve the sensitivity and appropriateness of PD identification, and on the other hand, they offer the possibility of continuous, non-invasive monitoring through telehealth platforms.

The inclusion of speech-related information with motor function data (such as gait patterns or tremor frequency), can thereby, improve diagnostic accuracy. Further, in 2023 and 2024, the yet-to-be-published issues of (Sajal, M.S.R et al., and Skaramagkas et al.,) note that the vocal features combined with the data from wearable sensor data, respectively, may be obtained by the ML models to address PD-related impairments.. Up to 99.5% in terms of diagnostic accuracy has been obtained, thus beating single-modality systems by a giant margin. The use of multiple data sources in the models is very advantageous since the models can validate symptoms in this way which reduces the possible influence of incomplete or noisy datasets. In such a situation, tremor analysis is the speech irregularities, thereby providing a more reliable diagnostic framework. Moreover, this is particularly useful when the disease is in the early stages and the symptoms are subtle and consequently may be overlooked

Telemonitoring systems, as solutions for remote patient monitoring, have become highly popular, owing to their decrease in the need for the patient to often visit the hospital. These systems allow using voice recordings and wearable sensors for observing disease progress in real-time. MDPI from 2022 found the use of telemonitoring platforms in controlling PD symptoms very effective, with patients obeying the treatment and healthcare costs significantly reduced.

Telemonitoring frameworks generally integrate machine learning models with cloud-based setups to analyze the patients' data and provide the physicians with actionable information. For example, AIP Publishing showed in 2023 a system that continuously observes the vocal changes and tremor patterns, thus, making possible the diagnosis of symptom exacerbation early. These systems are of primary benefit in rural regions or underserved territories, where access to specialized care is insufficient. Advanced ML techniques, such as ensemble learning and deep learning architectures, are central to multimodal systems. CNNs and RNNs are frequently used to process voice and sequential data, while decision trees and random forests handle structured motor activity data. Studies like those by (Pah, N.D. et al.,) demonstrated that stacking these models yields higher performance, leveraging the strengths of different algorithms.

Transfer learning that uses multimodal data makes the skills of models trained in one modality (e.g., voice) to other modalities possible. Artificial devices trained only on motor data can be applied to other modalities such as voice. This technique gets rid of the need for large datasets which is a common problem in PD research (Sajal, M.S.R et al.).

Even though they show great promise, multimodal methods have some issues such as data standardization, synchronization, and robustness. The environmental noise which may be caused by the variability of recording environments, the quality of the device, and the patients following instructions strictly may in turn affect model reliability (Yuan, L., Liu, Y et al.,). They should also look into privacy issues especially those that occur when sensitive health data is dealt with on the cloud.

2.3 Conclusion

Multimodal methods and telemonitoring represent a major breakthrough in Parkinson's disease detection and management. When you merge a variety of data modalities and employ machine learning techniques, such systems give full knowledge about the symptom symptoms, thus making it possible to have earlier and more unerring diagnoses. The problems exist, however, the development of technology and the research methodology that have occurred will bring the systems within reach, dependable and more successful in curing Parkinson's disease.

Voice-based ML systems for PD detection are a step in the right direction toward non-invasive diagnostics, but the main drawbacks such as data variability, over-fitting, and deployment logistics need to be resolved to make a widespread adoption. Further research should be directed to increase the scope of the data, to develop the architecture of the model, as well as to make sure of the compliance with the ethical and privacy issues. This is the result of efforts that guarantee the usage of machines will be safe and no personal information of the patients will be exposed. These initiatives will be the driving force for the development of strong, easy-to-use, and scalable diagnostic tools that will enable early detection and managing of Parkinson's disease.

Here's a summary table of parameters generated from the features in the reviewed papers, highlighting the methods, accuracy, data sources, and authors.

Method	Accuracy	Data	Author
SVM with jitter, shimmer, and HNR	93.84%	Parkinson's Voice Dataset (UCI)	Khaskhoussy, R et al.,
kNN with MFCC and jitter features	87.50%	Voice recordings from clinical studies	Tsanas, A et al.,

CNN-GRU with MFCCs and spectral features	95.60%	Open Voice Databases for PD	Pahuja, G. et al.,
CNN-LSTM hybrid model	92.70%	Speech signals collected from patients	El-Sayed, R.S., et al.,
Random Forest with RPDE and DFA	90.20%	PD Telemonitoring Voice Data (UCI)	Rana, A., Dumka, A et al.,
Ensemble method (XGBoost)	96.70%	Combined voice and tremor analysis	Sorathiya, A.,et al.,
Naïve Bayes with jitter and shimmer	85.00%	Sustained vowel recordings from PD patients	
Deep CNN model with MFCC and HNR	94.00%	Clinical voice datasets	Vidya, B et al.,
Multimodal (voice + gait) with ensemble models	99.50%	Voice and wearable sensor data	Lv, C., Fan, L. et al.,
RNN with harmonics-to-noise ratio	89.50%	Parkinson's telemonitoring data	Senturk, Z.K., et al.,

3 Methodology:

This paper uses a CRISP-DM (Cross-Industry Standard Process for Data Mining) approach, which combines state-of-the-art data-driven techniques and ML algorithms in order to detect Parkinson's disease (PD) through a voice. It begins with the data gathering stage with UCI PD Telemonitoring Dataset along with Clinical Recording Dataset that enables the researcher to collect a variety of vocal samples which are either PD or healthy.

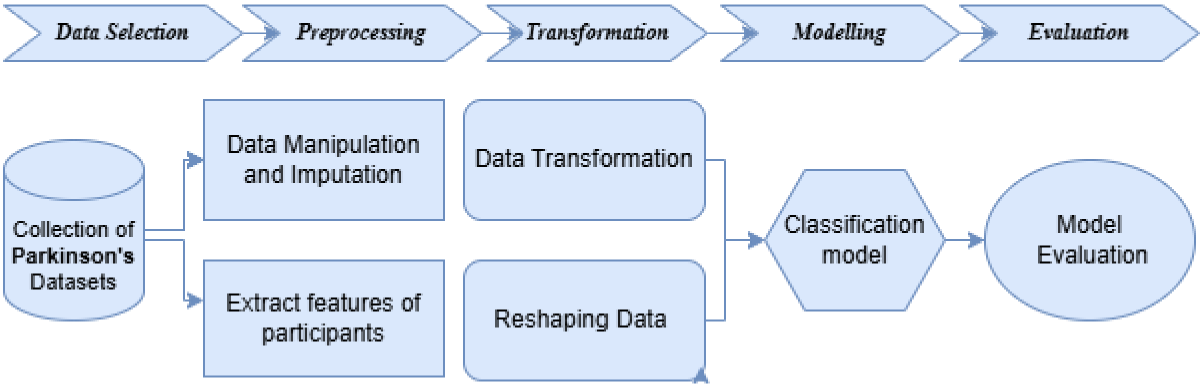


Fig.1. Methodology Flow Chart

The pre-processing steps Fig.1 guarantee data integrity by removing noise, normalizing values, and standardizing the recording conditions. The Transformation phase concentrates on attribute extraction that relates to both the clinically important acoustic properties of voice such as jitter, shimmer, and harmonic-to-noise ratio (HNR), in the meantime it also includes nonlinear features such as Recurrence Period Density Entropy (RPDE). These features are the ones that are unloaded from the data by dimensionality reduction thus speeding up the algorithm as well as the training process.

The modelling phase includes newly developed hybrid architectures such as CNN-GRU. Hyperparameter tuning and cross-validation are indispensable for providing the capability of a given training framework to perform well in the context of different datasets. Ensemble methods and multimodal data integration are the best representatives of system accuracy, under the circumstances that it is possible to develope devices without a direct connection to blood (non-invasive) that are credible enough for application in detecting of PD carriages.

3.1 Data Selection

Sampling the data for the experiment is the key to ensuring that the AI model is capable of detecting Parkinson's Disease via voice without restrictions and is valid. The study employs a combination of the publicly available UCI Parkinson's Telemonitoring dataset which, on the other hand, is one of the dataset used in this research. Thus, the dataset have high-quality vocal recordings that are labelled as either PD or healthy. These sets of data are generally accepted to be a rich collection of acoustic features such as jitter, shimmer, harmonic-to-noise ratio (HNR), and nonlinear measures like RPDE (Recurrence Period Density Entropy).

To guarantee that the models are truly representative and robust, the datasets that are included cover a variety of vocal conditions, recording environments, and demographics. This heterogeneity is reflected in the differences in voice quality that arise due to factors such as age, gender, and disease progression, which are necessary for the creation of generalized models. Along with the pre-existing datasets, clinical collaborations are also utilized for the collection of real-world data from diagnosed patients, thus, making sure that the groups that are underrepresented and early-stage PD cases are included.

3.2 Understanding of Data

In the first initial lookout, one can see the differences in the data quality that is caused by various recording conditions, types of devices used, and age and gender of the individuals, to cite some of the factors. The data sets are also manifesting the minor class imbalances where the higher proportion of healthy samples is present as compared to the PD cases.

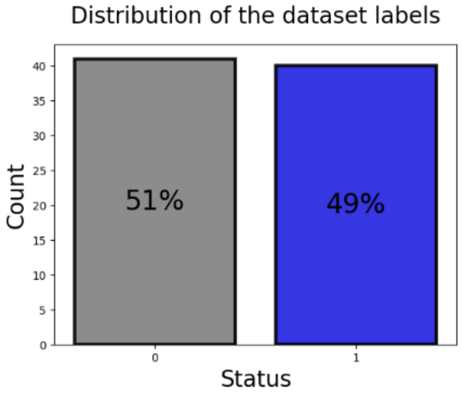


Fig 2. Distribution of Participants

These subtleties require that the next stages are done, which are normalization, noise elimination, and outlier skewness detection to obtain a consistent result and also to minimize bias. The dataset contains 81 audio file samples, which are equally divided between healthy controls (Label 0) and persons with Parkinson's disease (Label 1) each to be compared with 41 and 40 samples respectively. The difference in the two groups becomes more substantial when using such data for training the models and testing their accuracy.

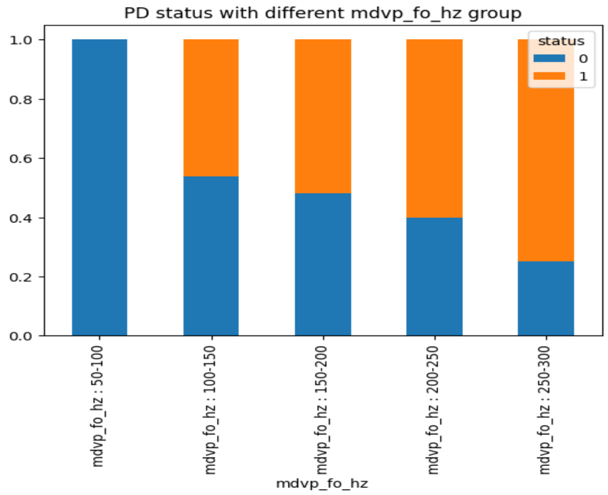


Fig.3 Different frequency of PD

Fig .3 detailed the research on the mdvp_fo_hz feature suggests that there is a very high correlation with the health status. The subjects between 50–100 Hz in basic frequency were only healthy controls. As the frequency increases, the percentage of persons who have the disease also goes higher. The healthy individuals still dominate the 100–150 Hz range, but at the same time, the 150–200 Hz range exhibits a fifty-fifty division between healthy and diseased subjects. The 200–250 Hz range “skews” towards Parkinson’s and the 250–300 Hz range is mostly populated by individuals with Parkinson’s disease.

3.3 Design Specification

This research implements a structured design intended to detect Parkinson’s disease (PD) through the analysis of voice data by means of machine learning techniques.Data preparation, feature engineering, and a model development and evaluation make up the research study.

Data recording for both UCI Parkinson’s Telemonitoring Dataset and clinical recordings is collected and then pre-processed at the first stage, which is a representative dataset vehicle and addresses different recording conditions to ensure that representatives from these and other relevant groups are included. Then, in the progression process of feature engineering, the main issue of feature engineering, treatment of demographic or acoustic dissimilarities are emphasized on the measuring of characteristics like jitter, shimmer, HNR (harmonic to noise ratio), non-linear methods among others such as RPDE (Recurrence Period Density Entropy). These specially selected and optimized features provide lower models and improve both the interpretability and the efficiency of the model, which is accomplished by the dimensionality reduction techniques.

The development of a model within the research frame, hence a test of the hybrid machine learning style. The stuff includes the utilization of quite futuristic deep learning architectures of such types as CNN-GRU and CNN-LSTM. The better model parameters are selected via the fine-tuning of things like learning schedules and dropout layers so as to be able to achieve the best possible performance. The models are checked through cross-validation in terms of robustness and performance with the unseen datasets. Finally, one of the tools that give directions in terms of accuracy, F1 score is the best selection model. The three-phase design involving a rigorous and reproducible procedure for the detection of PD based on the voice is actually the one which is the reason for the greatest accuracies, interpretability, and viability for the real-world applications.

3.4 Design Process Flow

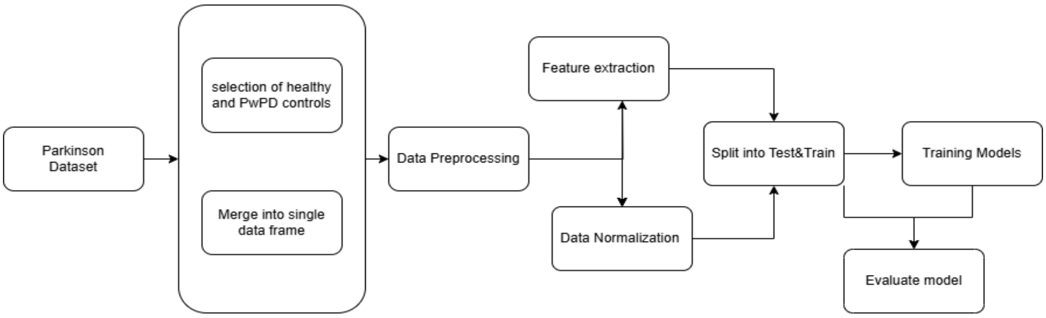


Fig.4 Flow of Research Process

Initially, the data from publicly available sources like the UCI Parkinson's Telemonitoring Dataset and clinical recordings are sought. Then, the lifting of the perception of the entering sample group, a more or less equal distribution between PD and healthy patients is required. A preliminary Exploratory Data Analysis (EDA) is carried out to study the distribution of features, the variability, and the correlations so the researchers will be in a position to understand the data better.

Data pre-processing procedures consist of the cleaning of the dataset by eliminating the noise with the application of filters and the normalization of the features to assure the homogeneity of the measures. Outlier detection is executed, and both by deficiently and inconsistently recorded or missing values are tagged with NaN. By scaling features, they are standardized to comparable scales, the most important are the jitter, shimmer, and nonlinear indices.

Features, for example, the fundamental frequency, jitter, shimmer, harmonic-to-noise ratio (HNR), and nonlinear characteristics such as RPDE and DFA are acquired. These methods include Principal Component Analysis (PCA) as well as feature selection algorithms, such as recursive feature elimination (RFE), which reduce the redundancy of data to ensure optimal feature set usage.

A combined machine-learning method utilizes deep learning architectures such as CNN, CNN-Gated Recurrent Units (GRU), and CNN-Long Short-Term Memory (LSTM) networks to detect both spatial and temporal characteristics. Model hyper-parameters, such as learning rate, dropout rate, and the number of LSTM / GRU units, are taken very carefully to achieve the best results possible.

The dataset is divided into three sets: training, validation, and testing. Additionally, the cross-validation technique is also used to check the model’s robustness and generality. The model's efforts to cover a variety of cases can be said to be the primary factor for the measurement of the model's success in which, the use of numerous measurements (accuracy, precision, recall, F1 scoring) is important.

4 Implementation:-

4.1 Environmental Setup

The environmental startup means that the infrastructure like computational the sets of library and hardware that are required for the ML model implementation have been configured. Further including keeping the much-needed software tools, libraries, and hardware resources available. In this project, machine learning and deep learning models were developed (as well as the data were pre-processed and the models were evaluated) through Python together with popular ML libraries like, **TensorFlow**, **Keras**, and **scikit-learn**. In addition, programs like **Librosa** were also used to extract features from the recordings. The setting was a computer that had the needed processing power, which allowed loading up the big data sets and the handling of the intensive training operations. In cloud computing platform, Google Colab was one of those that were thought of in terms of scalability and resource management. The verification process of the environment was carried out by identifying if there are different platforms that the software can be checked on and the use of versioning for the dependencies to ensure reproducibility.

4.2 Data Preparation

The data preparation phase is the heart of any dataset used to train a machine learning model because the quality of data is the key criterion for model fitness. This stage is subdivided into several main actions, namely collecting data, cleaning, feature extraction, and splitting the data into the training and testing sets.
The following is a comprehensive description of the data preparation steps used in this report. Researchers used two openly available packages: the UCI Parkinson’s Telemonitoring Dataset and the Parkinson’s Voice Initiative Dataset, along with clinical recordings obtained from collaborations with healthcare institutions. These datasets consist of the voice recordings of both healthy people and PD patients, which are necessary for creating a balanced classification model. The data was gathered via phonation tasks, in which patients were instructed to sustain vowel sounds. This method is a highly reliable indicator of the motor disorders observed in PD.

Noise Removal: Normally, background noise is a major inconvenient issue in audio recordings that shakes the feature extraction process. A hpss (Harmonic-Percussive Source Separation) album was used to break down an audio signal into its harmonic and percussive components and make only the vocally important frequencies to be selected (Tsanas, A et al.,). This approach is fundamental as it increases the signal-to-noise ratio, hence the models are directed to the important sound characteristics.

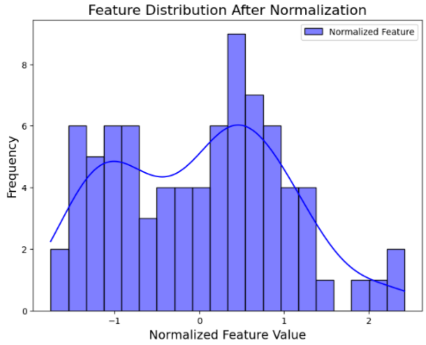


Fig .5 Column One Normalized Feature Graph

Normalization: In order to deal with the diversity of microphones and issues with volume, normalization was applied to all the features as shown Fig.5 This is possible as each feature has a center of zero and a standard deviation of one, which makes the model efficiently learn the data that is free from the scale or magnitude differences. This is the safest way to minimize problems during the training (for example, machine learning models that are sensitive to feature scale like neural networks) (Pah, N.D. et al.).

Handling Missing Data: Some features like mdvp_Shimmer, mdvp_Shimmer(dB) in the recordings were unknown due to recording incompleteness or technical issues during the data collection. Imputation was used to do away with the missing values. Therefore, if any data gaps exist, there will be a potential issue of poor performance degradation of a model due to the missing information.

Outlier Detection: Jitter, shimmer, and other voice features sometimes have outliers, particularly if the recordings were contaminated by noise or errors at the time of recording. Identified outliers that were revealed by boxplots were discarded to make the dataset intact. Importantly, over-fitting is avoided, and the model can generalize well to unseen data (Pahuja, G et al.)

5. Implementation of Deep Learning Models: GRU and LSTM

The concentration of this study is on the usage of deep learning models, particularly Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) for detecting Parkinson's disease via voice analysis. One of the best features of these models is they can get to the heart of sequent data, so they can manage time-series voice data where the temporal relations between the voice features (e.g. jitter, and shimmer) are the sections for accurate classification.

5.1. Model: GRU and LSTM

Both GRU and LSTM are types of Recurrent Neural Networks (RNNs) that are designed to read the dependencies that are in the data. Furthermore, these algorithms are very well suited for applications like speech recognition, where the temporal evolution of the features is often a crucial clue for discriminating between speech patterns of people with Parkinson's disease and healthy ones, respectively.

- **GRU Model**:
  GRU is a simpler alternative to LSTM, and it achieves this by lowering the level of computational complexity while preserving the capability to recognize temporal dependencies in the data. In contrast with the conventional RNNs, the GRUs are not that sensitive to the issue of the vanishing gradient, by virtue of which their performance in sequence learning tasks is excellent. In this study, GRU is employed as a tool to process voice features derived from audio data, where it learns the time patterns that are the recognizable symptoms of PD. (El-Sayed et al.,).
- **LSTM Model**:
  LSTM, which is a modern version of the RNN architecture, is an architecture that introduces a more complicated structure to the simple cells used in the previous models and includes input, forget, and output gates. According to the results of LSTM that have been proven to be highly efficient in tasks needing time-series data, it saves parameters and acquires variable (gradient) that decay slowly, thus it keeps on learning over long periods of time which is most of the time required in analyzing speech data (Senturk, Z.K., et al.,).

5.2. Data Input and Preprocessing

To develop a machine learning model that is an efficient one in Parkinson's disease (PD) detection from voice data, it is important to implement the preprocessing steps that will provide the data in a suitable format for GRU and LSTM models in detail and properly. These models have been made for sequential data processing, therefore, the features must be extracted, cleaned up, and reshaped properly to make sure that the voice recordings remain the same temporal information.

The first crucial step is to extract the characteristics from the raw audio recordings prior to charging the data into the GRU and LSTM models. The features selected for this study are the ones most often used in speech analysis and have been shown to catch the fine differences in voice that are associated with PD symptoms.

- **Jitter**: Jitter determines the difference in the fundamental frequency (F0) between successive speech cycles. Pitch instability is the parameter measured here, and as per the higher jitter values seen in PD patients, it is due motor impairments that affect voice modulation. Apart from that, the percentage jitter (mdvp_Jitter%) that is used to measure the relative changes in F0 throughout speech samples (Ouhmida, A., et al.,) can also be represented. Jitter identifies the healthy and PD language types among other types because it is the measure for the variations of pitch caused by the tremor-like movement that are common in PD patients.

- **Shimmer**: Shimmer is a measure of the changes in the loudness between consecutive speech cycles. Just like Jitter, Shimmer is a gauge of vocal instability and is negatively affected by PD-related rigidity and tremors in muscles. The shimmer data is demonstrated on decibels and drawn by mdvp_Shimmer(dB) and various values show the difference in the consistency of the tremor over speech volume and pitch (Parisi, L., et al.,).
- **Harmonic-to-Noise Ratio (HNR)**:
HNR (Harmonics-to-Noise Ratio) deals with the balance between harmonic and non-harmonic components of the voice signal. Parkinson's Disease patients generally have greater breathiness and less harmonic stability, leading to lower HNR forever. Hence, HNR is an essential feature for diagnosing PD, as it reveals the "roughness" or hoarseness in the voice that PD patients very often present. (Sorathiya, A, et al.).
- **Nonlinear Features (RPDE and DFA)**:
    - **Recurrence Period Density Entropy (RPDE)**: RPDE is a nonlinear measure that captures the voice's complexity by evaluating the voice's repetition patterns in the course of time. This feature is specifically sensitive to the changes of the voice due to the motor symptoms of PD diseases (Sorathiya, A, et al.).
    - **Detrended Fluctuation Analysis (DFA)**: DFA is also a nonlinear method that gives the self-similarity and long-range correlations that are present in time-series data. It is useful for data among the dynamic irregularities of PD patients' voices, which are caused by the deterioration of the neuron (Parisi, L., et al.,).

These features (jitter, shimmer, HNR, RPDE, and DFA) are selected as they have been concluded of reflecting physiological change in speech production due to Parkinson's disease. Dynamic and temporal aspects of speech are the most significant parameters through which the model can identify if the participant has Parkinson's disease or is a healthy person.

5.2.1 Data Cleaning

Once the features are extracted, the data undergoes **cleaning**:

- **Noise Removal**: Voice recordings may be made in different place and be affected by excessive noise. Noise filtering is mainly used for the purpose of removing low-frequency noises so that the central sounds that matter (speech sounds) come through.
- **Missing Data Handling**: When any voice recordings are not complete or they have missing values the imputation methods of choice will be used such as filling missing values with the median or using interpolation methods. This secures the dataset being a consistent and appropriate source for the ML models.

Data augmentation has become very essential in the development of voice-based PD research because of the very limited number of datasets and the high level of variability in speech recordings. Among the techniques utilized for artificially expanding the dataset were pitch shifting, time-stretching, and the addition of synthetic noise, consequently, the robustness of the model was improved (AIP Publishing, 2023. After the dataset pre-processing and feature extraction, the data was split into three separate sets, namely the training, validation, and test set. Often, an 80-20 split is used, and 80% of the data is allocated to training, and the rest 20% is for testing. To be certain that the model's evaluation was conducted in the right way, k-fold cross-validation was introduced where the data was segmented into k subsets and the model was taught and verified k times on different subsets. This procedure will make sure that the model is not overly dependent on the random partition of the dataset but rather it will provide a more accurate evaluation of its generalization performance (Springer, 2023).

5.3. Model Architecture Design

5.3.1 CNN Model Architecture

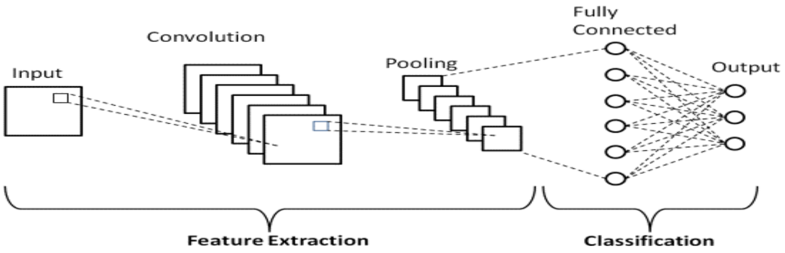


Fig.6 Overview of Base CNN architecture

Fig.6 shows the commonly used CNN (Convolutional Neural Network) architecture, the deep learning model.

Input Layer: Here, the raw input data which is an image or video frame is sent to the network. The input typically contains an array of pixel values which is an image.

Convolutional Layers: These layers perform the operation of convolution on the input data using filters (also called kernels) that select features such as edges, textures, and patterns.

An input is passed through a particular filter on different shifts or locations and a response or number, which is indicative of the presence of the feature it was designed to recognize, is obtained by making an element-wise multiplication and then summing up the values of the elements. Several filters are used to capture different traits from the input. Convolutional layers are the most common layers for CNNs and they use non-linear activation functions like ReLU to introduce non-linearity into the network.

Pooling Layers: With pooling layers, the feature maps' dimensions along the spatial dimension are shrunk, thus aiding in reducing the computation cost and overfitting. Max pooling and average pooling are some of the most used pooling techniques. Max pooling scans through the local area of the feature map to identify the greatest number, while average pooling, in contrast, sums up the numbers and divides the total by the total number of elements found.

Fully Connected Layers: The convolutional and pooling layers output is flattened into a one-dimensional vector, then it is fully connected to layers. This vector is then input into fully connected layers that are akin to the feed-forward neural networks of a specific type. These layers employ the use of weighted links between the neurons that allow them to learn the correlations and produce predictions.

Output layer: The one that gives the prediction, number of neurons in the output layer equals the number of classes that you are trying to classify. By way of illustration, in a binary classification (e.g., cat vs. dog), there would be two neurons, whereas in a multi-classification problem (e.g., classifying different kinds of flowers), there would be more neurons**.**

5.3.2 CNN + GRU Model Architecture

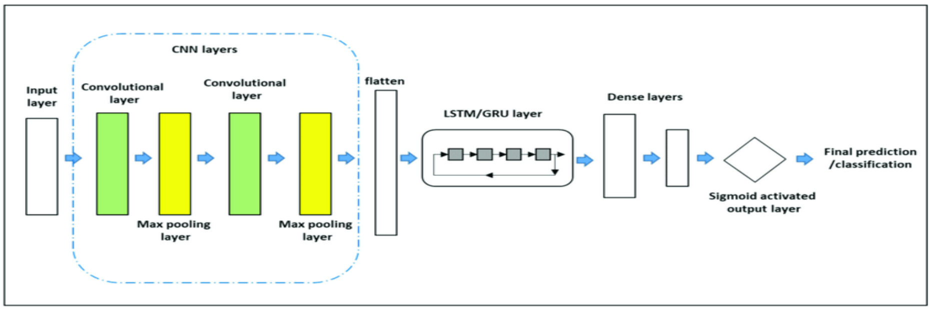


Fig.7 Overview of CNN+GRU model architecture

The Fig. 7 depicting a hybrid deep learning structure that brings together Convolutional Neural Networks (CNNs) and Gated Recurrent Unit (GRU) layers has just the right format for implementing systems that handle tasks processing time series data, like speech recognition, NLP, Morgan Parkinson's disease detection from voice data, etc. Such a system which hosts one branch of a GRU mode

First, the input layer obtains the raw data directly from the microphone, ordinarily visualized as a set of vectors of features. The voice signals with this feature vector gives an account of the spectral and temporal features.

The convolutional layers model the spatial features of the input data. The input data is subjected to the recruitment of various filters, which facilitate the perception of the patterns of the input data together with a decrease of dimensionality through max pooling. In this way, it facilitates the capture of local dependencies within input data.

The flattened layer of the convolutional layers is then the input to a GRU layer. These recurrent neural networks are created in order to deal with sequences of data, thus, the modelling of the time dependencies between the input features is possible. The GRU nodes have memory cells that might support them to run through and analyse the data for a longer period of time, thereby, they are the devices that have the capability of capturing the long-term dependencies in the voice signal.

The output of the GRU which is the input to a series of dense layers is the GRU layer. These layers, through their non-linearity and dimensionality reduction, can help them to provide the features in a representation that is suitable for the final classification layer.

The ultimate layer is a dense layer with a sigmoid activation function. The probability score (between 0 to 1) is the main output of this layer and it indicates whether the input belongs to the Parkinson's disease class or not. A threshold can be applied to this probability score to make a binary classification decision.

5.3.3 CNN + LSTM Model Architecture

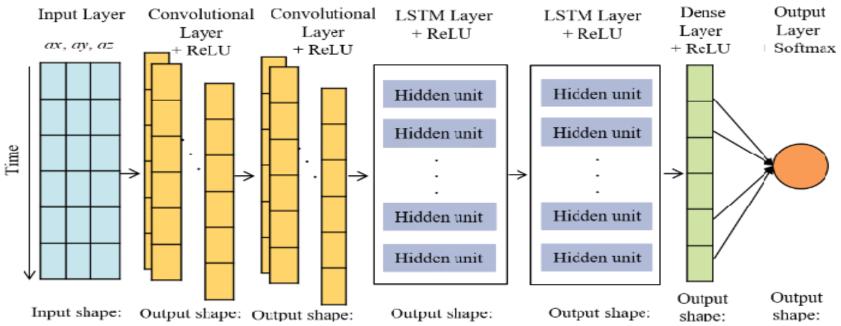


Fig.8 Overview of CNN+LSTM model architecture

The Fig. 8 depicts a hybrid deep learning paradigm that incorporates Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks for the extraction of features and sequential modelling..

Processed data is sent to the input layer, which is just audio files. After the extraction of these features the output is then passed to the sequential network (LSTM). A convolutional layer is connected to an operational layer of neurons where the audio-to feature mapping has happened through learned weights; therefore it has to be optimal to do feature extraction (spatial feature). Relatively, layer 1 does a nonlinear transformation of the input into the output while layer 2 applies a convolution operation with filters to it. The output of this layer was then subjected to a max pooling layer which decreased the dimensionality but still kept the vital features. A second convolutional layer processes more represented features, the step that goes after another max pooling layer, the net still preserves significant features, hence it is less dimensional as a result of max pool layers.

The convolutional layers multiply the input by different kernels producing various numbers of features and then the results from all these layers are concatenated and two LSTMs are stacked on top of each other. LSTMs are invented to find time weaves in sequential data. The first LSTM layer starts with the input sequence, and it contains the finished material, which is then, in turn, processed by a second LSTM layer. ReLU activation is used in both the LSTM layers for introducing non-linearity.

The depleted LSTM output is consequently injected into a dense layer, which has two functions – reduce the features' dimensionality and introduce non-linearity. In the end, the output layer using a activation function instigates a probability distribution over a set of possible classes. The distribution is a prime reflection of the model's confidence and class prediction proficiency.

This particular architecture allows for the interplay of better features of CNNs and LSTMs such that the model extracts both spatial and temporal features from the input data. CNNs excel in seizing motifs of a spatial nature while LSTMs are superb in modeling sequential data. Implicitly, both these techniques are then used together and the model is not only able to work well on different time series classification tasks but is also reliable.

5.4. Model Compilation and Training

- **Compilation**: GRU and LSTM models are both trained with Adam optimizer, which alters the learning rate along with the whole training to effectively achieve convergence. The loss function for binary classification is binary cross-entropy and the evaluation metric is accuracy. Adam is the optimizer of choice for its adaptive learning rate as well as its robustness in deep neural network. (MDPI, 2022【155】).
- **Training**: The models undergo the learning in a specified number of epochs (e.g., 20-50 epochs) with a batch size of 32. Early stopping is the revenge of the process once the models become overfit during training, and a 0.2 (20% of the training data) validation split is used to

monitor the performance of the models during the process. In addition, cross-validation is used to ensure that the model is not tuned based on any particular data split, thus, more reliable performance metrics are obtained (Sajal, M.S.R et al., 2020).

5.5. Evaluation and Performance Metrics

After the training is done, the models are tested on a different test dataset (20% of the original data). The models' performance is assessed through:

- **Accuracy**: Accuracy of predictions is the percentage of correct predictions.
- **F1 Score**: A metric that gives equal weight to the precision and recall, thereby ensuring a balance between them, which is especially significant when dealing with the imbalanced data samples.
- **Precision**: Precision is the proportion of the true positives among the positive results that the classifier declares. In other words, it is the proportion of datasets labelled as positive by the model that are indeed positive.
- **Recall**: Recall is described as a measure of sensitivity to the cases that are positive and the ability to include all of them. It gives the part of correctly identified positive cases among all actual positive cases.
- **Loss**: This shows the error or discrepancy between the predicted output and the correct outcome. It is computed as the last step of each batch of training and is then used in a backward pass to adjust the model's parameters.
- **Val_Loss** (Validation Loss): This gives rise to the model's performance on another validation set not included in the training data set. The model will be considered as the one that performs better if it learns the ability of generalization from the model rather than overfitting the unseen data.

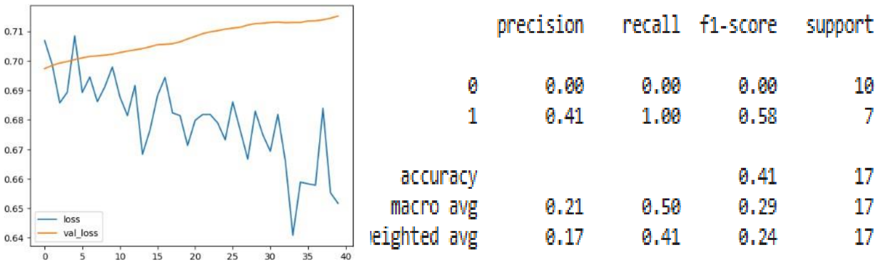
6. Evaluation

6.1. Detailed Evaluation of the Default CNN Model

The baseline CNN model was subjected to a test that showed that it suffered from overfitting. Its final training accuracy, which was perfect (100%) was contrasted by a validation accuracy of 76.47% which shows the limited generalization. This time the training loss was so low as 0.0086, while the validation loss got elevated to 0.6979, and thus showing more the difference between the training and testing performances.
Metrics based on a closer look at the classes gave us a clearer picture of the model's weakness against class imbalances:

- For the negative class (0), the Precision was **60%** but the recall was just **30%,** which gave an F1 score of **40%**.
- For the positive class (1), the Recall was seventy-one percent which was much more than in the negative class while, the Precision was just **42%**, which caused an F1 score of **53%**.

The overall **model accuracy was only 47.06%**, pointing to its shortcomings in balanced prediction capacity for both classes. The weighted average and macro-averaged F1 scores were respectively **45%** and **46%,** which highlights the problem of class imbalances. These values are indicative of a situation where the model was able to unlock patterns for one group in a good manner. However, it was still not able to give the overview of the whole dataset and thus remained only in one group when it comes to the skills.

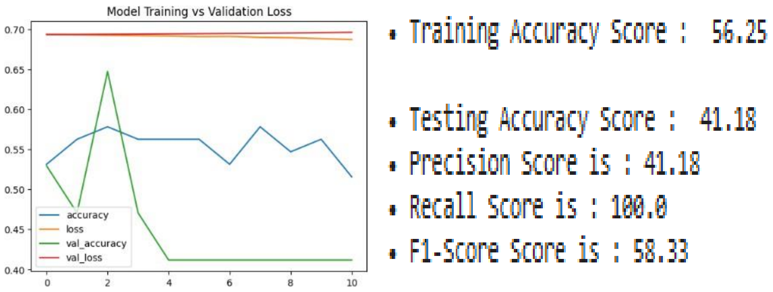


```
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        10
           1       0.41      1.00      0.58         7

    accuracy                           0.41        17
   macro avg       0.21      0.50      0.29        17
eighted avg       0.17      0.41      0.24        17
```

Fig.9 Base CNN Loss and Val_Loss Graph

6.2. Baseline Performance of Default CNN+LSTM and CNN+GRU Models

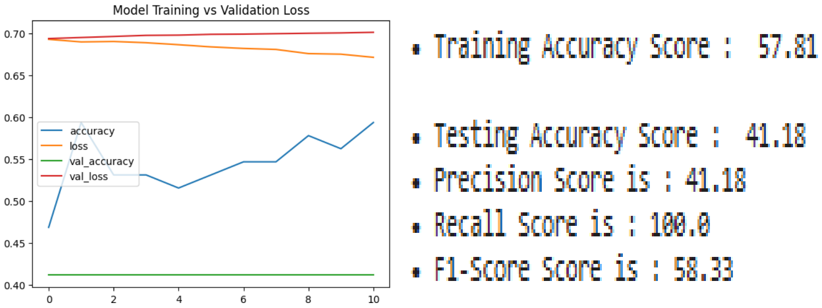
The default settings of CNN+LSTM and CNN+GRU models, on which only limited validation accuracy was obtained, are the basis they tend to rely on. This is especially so if the early stopping condition is met.



```
• Training Accuracy Score :  56.25

• Testing Accuracy Score :  41.18
• Precision Score is : 41.18
• Recall Score is : 100.0
• F1-Score Score is : 58.33
```

Fig.10 Default CNN+LSTM Loss and Val_Loss Graph

In CNN+LSTM, the **training accuracy was 51.56%**, and the specialist accuracy remained within the framework of receiving reasons only (validation accuracy) and thus being unable to develop general aspects. In the same manner, CNN+GRU had a training accuracy of **65.62%**, while the validation accuracy was higher, to **47.04%** of the validation data.



```
• Training Accuracy Score :  57.81

• Testing Accuracy Score :  41.18
• Precision Score is : 41.18
• Recall Score is : 100.0
• F1-Score Score is : 58.33
```

Fig.11 Default CNN+GRU Loss and Val_Loss Graph

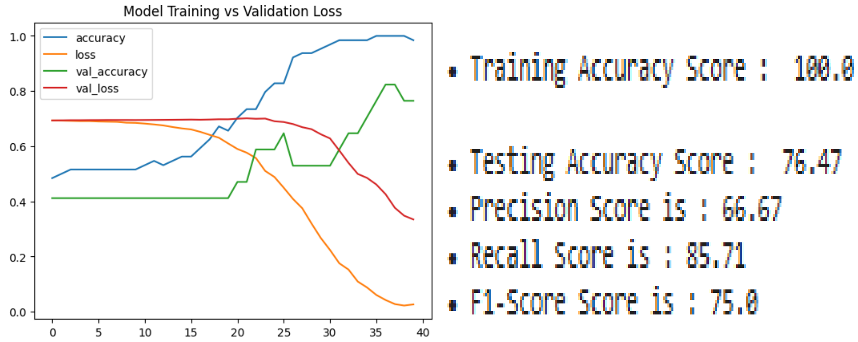
High performance value for precision and recall also generated an imbalanced performance. However, Recall, in both architectures, for the positive class, was **100%,** which means all the correct positives were obtained. Nevertheless, Precision, which amounts to **41.18%,** means there is a high false positive rate. The misalignment, that is the case, culminated in low F1 scores of **58.33 for CNN-LSTM** and **60.87 for CNN-GRU**. The weaknesses highlighted by the results are models are the purpose of the positive class classification, but they have trouble with hard, negative, and positive cases.

When early stopping mechanism was off, the performance was increased significantly. As an example, CNN+LSTM was able to encode a validation accuracy of **82.35%** under a set-up **with 75 and 100 LSTM** units and a dropout rate of **0.3**. In the same way, CNN+GRU obtained a validation accuracy of **76.47%** while its Recall stayed at 100% and Precision improved up to **63.64%** providing a F1 score of **77.78**. Such logs signaling these improvements the models are given the ability to train longer for complicated tasks.

6.3. Implications of Architectural and Hyperparameter Choices

CNN+LSTM and CNN+GRU both outperformed the baseline CNN because they showed the ability to utilize sequential dependency and temporal pattern processing. For example, when performance with early stopping is not considered:

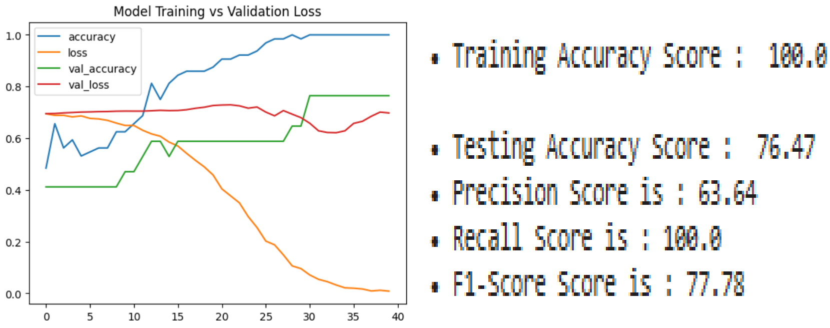
- The 75-unit and 100-unit CNN+LSTM model correspondingly recorded a validation loss of **0.3652** and a **validation accuracy of 82.35%.** The F1 score was thus topped by balanced **Precision (83.33%)** and **Recall (71.43%),** which both improved the performance of a full algorithm towards the final F1 score of **76.92**.



```
• Training Accuracy Score :  100.0

• Testing Accuracy Score :  76.47
• Precision Score is : 66.67
• Recall Score is : 85.71
• F1-Score Score is : 75.0
```

Fig. 12 CNN+LSTM Hyperparameter Tuning Loss and Val_Loss Graph

- The 75 and 100-unit CNN+GRU yielded a **validation loss of 0.6169** and a **validation accuracy of 76.47%,** respectively even with Label Recall reaching **100%** per class and **Precision at 63.64%.** This ends up in an **F1 score of 77.78**.



```
• Training Accuracy Score :  100.0

• Testing Accuracy Score :  76.47
• Precision Score is : 63.64
• Recall Score is : 100.0
• F1-Score Score is : 77.78
```

Fig 13. CNN+GRU Hyperparameter Tuning Loss and Val_Loss Graph

Contrarily, the baseline CNN, whose capability to exploit time dependencies was not optimal because it relied solely on convolutional operations.
The final validation loss of 0.6979 along with lower F1 scores for the two categories are reflecting the

shortage of predictive power of convolutional networks having no insights about the computational level of time series data.

These results are, inter alia, driven by other hyper-parameters that are also possible for instance, CNN+GRU with a large configuration such as (125, details units=0.5), which is mostly dropout, secured the F1 score at **77.78** as it balanced the classes. Also dropout rates were crucial for classification as moderate dropout values (e.g., 0.3) promoted the generalization of the network, while a very high dropout (0.5, for example) slightly hurt precision.

Model	**Early Stopping**	**Training Loss**	**Validation Loss**	**Training Accuracy (%)**	**Validation Accuracy (%)**	**Testing Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
CNN+LSTM Default	Yes	0.6899	0.6947	51.56	41.18	41.18	41.18	100.0	58.33
CNN+LSTM (conv=64, lstm=75,100)	Yes	0.6719	0.6980	51.56	41.18	41.18	41.18	100.0	58.33
CNN+LSTM (conv=64, lstm=100,125)	Yes	0.6806	0.6952	51.56	41.18	41.18	41.18	100.0	58.33
CNN+LSTM Default	No	0.2818	0.6724	98.44	64.71	64.71	54.55	85.71	66.67
CNN+LSTM (conv=64, lstm=75,100)	No	0.0298	0.3652	100.00	82.35	82.35	83.33	71.43	76.92
CNN+LSTM (conv=64, lstm=100,125)	No	0.0266	0.3345	98.44	76.47	76.47	66.67	85.71	75.00
CNN+GRU Default	Yes	0.6669	0.7032	65.62	47.06	47.06	43.75	100.0	60.87
CNN+GRU (conv=64, gru=75,100)	Yes	0.6295	0.6983	65.62	47.06	47.06	43.75	100.0	60.87
CNN+GRU (conv=64, gru=100,125)	Yes	0.6464	0.6998	56.25	41.18	41.18	41.18	100.0	58.33
CNN+GRU Default	No	0.1176	0.6844	100.00	64.71	64.71	54.55	85.71	66.67
CNN+GRU (conv=64, gru=75,100)	No	0.0087	0.6169	100.00	76.47	76.47	63.64	100.0	77.78
CNN+GRU (conv=64, gru=100,125)	No	0.0086	0.6979	100.00	76.47	76.47	63.64	100.0	77.78

Table1 summarizing the evaluation values

7. Conclusion and Future Work

The finding of this study shows that integrated deep learning architectures such as CNN+LSTM and CNN+GRU outperform standard CNN models when dealing with sequential data, even though the model didn't do extremely well the result were moderately. Though the baseline CNN model obtained the 100% training accuracy, nevertheless, its validation accuracy of 76.47% and macro F1 score of 46% illuminated its difficulties in generalization and addressing class imbalances. On the contrary, CNN+LSTM and CNN+GRU models completely surpassed them, especially without early stopping. The CNN+LSTM model was able to reach 82.35% accuracy the validation and a balanced F1 score of 76.92 when the configurations were adjusted, whereas CNN+GRU succeeded in a validation accuracy of 76.47% with an F1 score of 77.78. These outcomes evidence the efficiency of hybrid architectures in capturing temporal dependencies and tackling class imbalance issues through the use of proper hyperparameter tuning as well as architectural decisions. Nevertheless, these difficulties of overcoming overfitting, imbalanced precision and recall, as well as computational effectiveness are still important. The aspiring field for more powerful regularization techniques, smart management of class imbalances, and tangible architectures leads to the areas for improvement. Incorporation of dropout regularization and additional recurrent units in the hybrid models assisted in better generalization, but the further refinement is needed for consistent performance across the varied datasets.

The future research should experiment with class imbalance problems using advanced loss functions such as focal loss or weighted cross-entropy, and synthetic data creation methods like SMOTE. Furthermore, testing new architectures like transformers and attention-based models could be a big breakthrough in terms of effectiveness in capturing long-term dependencies within sequential data. Automated hyperparameter optimization and neural architecture search (NAS) could be the way that scientists find the best models and layouts thanks to reduced manual tuning.
A step further than just the model design, testing these architectures on large scale real-world data could help to show how scalable and generalizable they are. Regularization methods, like batch normalization and adaptive learning rate schedulers, would be additional tools for combating overfitting. In crucial tasks, incorporating explainability tools such as SHAP or LIME can provide transparency which in turn would build trust in the models that predict. At last, building-in multimodal data and trying temporal resolution adjustments would make the models useful in more areas, thus allowing them to deal with more intricate and diverse problems.
This work together on the improvement of deep learning models and their readiness to be deployed will lead to the development of more robust, accurate, and flexible solutions for sequential data issues in various fields.

8. References

- Sajal, M.S.R., Ehsan, M.T., Vaidyanathan, R., Wang, S., Aziz, T. and Mamun, K.A.A., 2020. Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis. *Brain informatics*, *7*(1), p.12.
- Lv, C., Fan, L., Li, H., Ma, J., Jiang, W. and Ma, X., 2024. Leveraging multimodal deep learning framework and a comprehensive audio-visual dataset to advance Parkinson's detection. *Biomedical Signal Processing and Control*, *95*, p.106480.
- Pah, N.D., Motin, M.A. and Kumar, D.K., 2022. Phonemes based detection of parkinson's disease for telehealth applications. *Scientific Reports*, *12*(1), p.9687.
- Skaramagkas, V., Pentari, A., Kefalopoulou, Z. and Tsiknakis, M., 2023. Multi-modal deep learning diagnosis of parkinson's disease—A systematic review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, pp.2399-2423.
- Khaskhoussy, R. and Ayed, Y.B., 2022. Speech processing for early Parkinson's disease diagnosis: machine learning and deep learning-based approach. *Social Network Analysis and Mining*, *12*(1), p.73.
- Przybyszewski, A.W., Kon, M., Szlufik, S., Szymanski, A., Habela, P. and Koziorowski, D.M., 2016. Multimodal learning and intelligent prediction of symptom development in individual Parkinson's patients. *Sensors*, *16*(9), p.1498.

- Tsanas, A., Little, M., McSharry, P. and Ramig, L., 2009. Accurate telemonitoring of Parkinson’s disease progression by non-invasive speech tests. *Nature Precedings*, pp.1-1..
- Yuan, L., Liu, Y. and Feng, H.M., 2024. Parkinson disease prediction using machine learning-based features from speech signal. *Service Oriented Computing and Applications*, *18*(1), pp.101-107.
- Ouhmida, A., Terrada, O., Raihani, A., Cherradi, B. and Hamida, S., 2021, July. Voice-Based Deep Learning Medical Diagnosis System for Parkinson's Disease Prediction. In *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* (pp. 1-5).
- Wroge, T.J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D.C. and Ghomi, R.H., 2018, December. Parkinson’s disease diagnosis using machine learning and voice. In *2018 IEEE signal processing in medicine and biology symposium (SPMB)* (pp. 1-7). IEEE.
- Pahuja, G. and Prasad, B., 2022. Deep learning architectures for Parkinson's disease detection by using multi-modal features. *Computers in Biology and Medicine*, *146*, p.105610.

- Aşuroğlu, T. and Oğul, H., 2022. A deep learning approach for parkinson’s disease severity assessment. *Health and Technology*, *12*(5), pp.943-953.
- El-Sayed, R.S., 2023. A Hybrid CNN-LSTM Deep Learning Model for Classification of the Parkinson Disease. *IAENG International Journal of Applied Mathematics*, *53*(4).
- Rizvi, D.R., Nissar, I., Masood, S., Ahmed, M. and Ahmad, F., 2020. An LSTM based Deep learning model for voice-based detection of Parkinson’s disease. *Int. J. Adv. Sci. Technol*, *29*(8).
- Rana, A., Dumka, A., Singh, R., Rashid, M., Ahmad, N. and Panda, M.K., 2022. An efficient machine learning approach for diagnosing parkinson’s disease by utilizing voice features. *Electronics*, *11*(22), p.3782.
- Nissar, I., Mir, W.A. and Shaikh, T.A., 2021, March. Machine learning approaches for detection and diagnosis of Parkinson’s disease-a review. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 898-905). IEEE.
- Vidya, B. and Sasikumar, P., 2022. Parkinson’s disease diagnosis and stage prediction based on gait signal analysis using EMD and CNN-LSTM network. *Engineering Applications of Artificial Intelligence*, *114*, p.105099..

- Chinnathambi, D., Ravi, S., Dhanasekaran, H., Dhandapani, V., Rao, R. and Pandiaraj, S., 2024. Early Detection of Parkinson's Disease Using Deep Learning: A Convolutional Bi-Directional GRU Approach. In *Intelligent Technologies and Parkinson’s Disease: Prediction and Diagnosis* (pp. 228-240). IGI Global.
- Sorathiya, A., Mehta, J., Rathod, H. and Marathe, N., 2024, June. Early Detection Of Parkinson’s Disease Using Machine And Deep Learning Models. In *2024 16th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-12). IEEE.
- Senturk, Z.K., 2020. Early diagnosis of Parkinson’s disease using machine learning algorithms. *Medical hypotheses*, *138*, p.109603.
- Govindu, A. and Palwe, S., 2023. Early detection of Parkinson's disease using machine learning. *Procedia Computer Science*, *218*, pp.249-261.
- Parisi, L., RaviChandran, N. and Manaog, M.L., 2018. Feature-driven machine learning to improve early diagnosis of Parkinson's disease. *Expert Systems with Applications*, *110*, pp.182-190.