



OPEN Enhancing parkinson disease detection through feature based deep learning with autoencoders and neural networks

P. Valarmathi^{1✉}, Y. Suganya^{2✉}, K. R. Saranya^{2✉} & S. Shanmuga Priya^{2✉}

Parkinson's disease is a neurodegenerative disorder that is associated with aging, leading to the progressive deterioration of certain regions of the brain. Accurate and timely diagnosis plays a crucial role in facilitating optimal therapy and improving patient outcomes. This study presents an innovative approach to identify Parkinson's disease (PD) through the examination of audio waves using Feature Based - Deep Neural Network (FB-DNN) techniques. Autoencoder, a specific form of Artificial Neural Network (ANN) that is designed to excel in the task of feature extraction, is utilized in our study to effectively capture complex patterns present in audio data. Deep Neural Networks (DNNs) are utilized in the task of classification, using the capabilities of deep learning (DL) to differentiate between audio samples that exhibit Parkinson's disease (PD) and those that do not. The deep neural network (DNN) model is trained using the retrieved data, allowing it to effectively distinguish minor variations in voice characteristics that are linked to Parkinson's disease. The suggested methodology not only enhances the precision of diagnosis but also enables prompt identification, perhaps resulting in more efficacious treatment methodologies. The present study introduces a potentially effective approach for the automated and non-intrusive identification of Parkinson's disease through the analysis of audio data. The integration of Autoencoder-based feature extraction with Deep Neural Networks (DNN) presents a dependable and easily accessible solution for the early detection and continuous monitoring of Parkinson's disease. This approach has promise for significantly improving the quality of life for persons affected by this condition. The implementation in Python was conducted as part of our experimentation. Upon analyzing the accuracy, it became apparent that the Feature-Based Deep Neural Network (FB-DNN) exhibited superior performance compared to the other models. Notably, the FB-DNN achieved the highest accuracy score of 96.15%.

Keywords Parkinson's Disease (PD), Mel-Spectrogram, Deep neural networks, Autoencoder feature extraction

Parkinson's disease (PD) is a neurological condition that causes the substantia nigra's dopamine-producing neurons to die off. Dopamine, a vital neurotransmitter from the catecholamine and phenethylamine families, plays a central role in coordinating physical movement by transmitting messages between the brain and the substantia nigra. The onset of PD occurs when 60–80% of dopamine-producing cells are lost, resulting in insufficient dopamine levels to control motor functions, thus manifesting as PD symptoms. This dopamine deficiency leads to a loss of control over essential motor functions¹. Key motor symptoms include tremors, rigidity, impaired balance, and slowed movements. Beyond motor symptoms, nonmotor features encompass issues like dementia, depression, restless legs, temperature sensitivity, and digestive problems. While PD remains incurable, various treatments have been developed to manage motor and nonmotor symptoms, including both noninvasive pharmaceutical interventions and invasive surgical procedures. Notably, voice disorder testing emerges as a valuable noninvasive tool for early PD detection, given that approximately 90% of PD patients exhibit dysphonia or vocal impairments, distinguishing them from healthy individuals^{2,3}. Thus, diagnosing PD through voice analysis represents a promising and effective avenue.

¹Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, India.

²Department of Computer Science and Engineering, SRM Institute of Science and Technology, Tiruchirappalli, India. ✉email: goodmathi1996@gmail.com; valarmathi@mookambigai.ac.in; suganyasuchithrra@gmail.com; dr.k.r.saranya@gmail.com; priya501@gmail.com

PD, affecting roughly 10 million persons globally, ranks as the second most frequent neurodegenerative ailment behind Alzheimer's disease. The risk of PD increases with age, particularly after 65, with men more susceptible than women. Early symptoms, such as loss of smell, constipation, and sleep disturbances, often precede motor symptoms. Timely interventions in the early stages are critical for slowing or halting disease progression. Nevertheless, clinical symptom-based PD diagnosis remains challenging and intricate. Given that 90% of PD patients experience voice disorders, utilizing voice data for detection has gained prominence in recent times^{4,5}. Audio signals play a pivotal role in early PD diagnosis, even when voice abnormalities are imperceptible to the human ear in the early stages, they can be detected through voice signal analysis.

The diagnostic journey commences with a comprehensive medical history evaluation, a pivotal step that entails meticulous inquiries into the patient's symptoms, their onset, progression, and any pertinent contributing factors. Conversations with the patient's caretakers or family members, who frequently know more about the patient's medical history, are an important part of this process. Following the medical history evaluation, a thorough physical examination takes place, with the primary aim of observing and evaluating the patient's motor functions^{6,7}. This examination encompasses an assessment of the patient's posture, gait, and movements, with a specific focus on identifying signs of muscle stiffness, tremors, and other distinctive features that are indicative of PD which is shown in Fig. 1. The significance of these clinical assessments cannot be overstated, as they serve as the foundation for guiding subsequent diagnostic procedures, treatment planning, and patient care. These traditional diagnostic methods have long been essential in the early identification and management of PD^{8,9}. However, advancements in medical technology and a growing body of research have spurred the exploration of complementary diagnostic approaches that can enhance accuracy and efficiency in the diagnosis of this complex

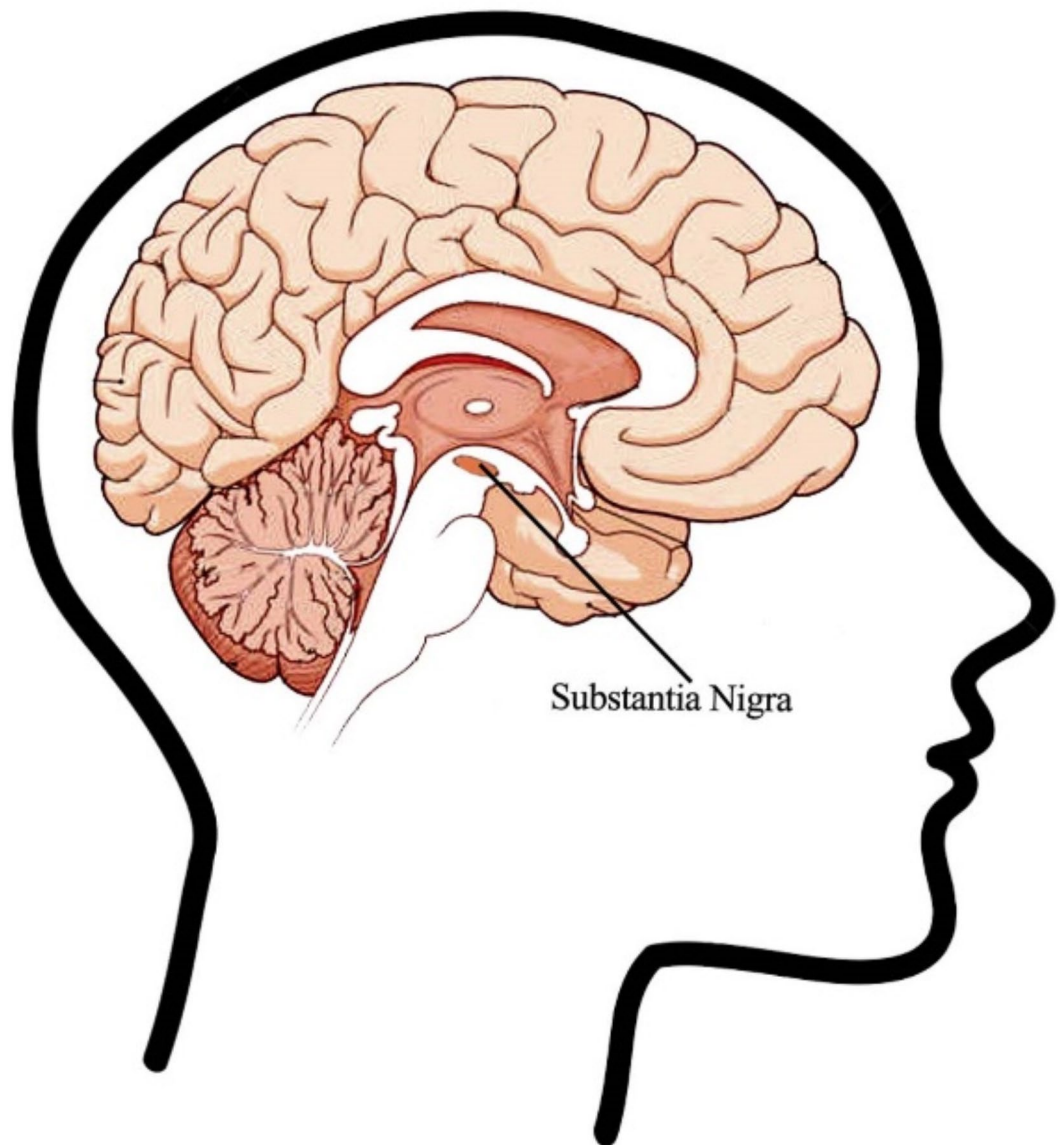


Fig. 1. Diagrammatic Explanation of PD.

neurological disorder. PD and other movement disorders can be categorized into two phases: the preclinical phase, characterized by neurodegeneration without clinical symptoms, and the prodromal phase, marked by clinical symptoms that do not yet provide a definitive diagnosis^{10–12}. Therefore, early detection in both phases is crucial for timely medical intervention. Presently, there are no established biomarkers for efficient early PD detection. Consequently, there is a pressing need to harness artificial intelligence (AI) techniques to facilitate early and accurate PD diagnosis in the healthcare sector. The development of a computer-aided diagnostic system is imperative to analyze voice data and differentiate between voices of people with PD and those without. Recent research has explored noninvasive methods for PD diagnosis using Audio signal analysis.

In the realm of PD diagnosis, the standard protocol involves a comprehensive history and physical examination, which is a cornerstone of clinical assessment. However, to gain a deeper understanding of the patient's condition and to provide a more nuanced evaluation, further diagnostic investigations often become necessary. These supplementary assessments encompass specialized neurological tests designed to assess both motor abilities and cognitive functioning^{13,14}. In pursuit of a precise diagnosis, an array of tests is employed, including evaluations of hand-eye coordination, fine motor skills, and cognitive capabilities. In addition to these examinations, neurologists make use of standardized rating scales such as the Unified PD Rating Scale (UPDRS) to quantify the severity of motor symptoms and provide a comprehensive evaluation of the patient's overall health status. These scales offer a structured framework for assessing various dimensions of PD, encompassing motor function, mood, and daily living activities.

Although these methods are invaluable, they inherently retain a subjective element and may not always exhibit the requisite sensitivity for early diagnosis. Machine Learning (ML) models, which offer an objective and data-driven approach to PD detection^{15,16}. By analyzing multifaceted datasets, which may include clinical information, neuroimaging data (such as MRI or CT scans), and even audio recordings, ML algorithms can discern patterns and biomarkers indicative of PD. These algorithms possess the capacity to swiftly process vast amounts of data and may detect subtle changes that might elude human clinicians. Notably, one of the primary strengths of ML and Deep Learning (DL) models is their ability to extract pertinent features and biomarkers from various data sources. For instance, in voice recordings, changes in speech patterns, such as alterations in pitch, rhythm, and articulation, can serve as invaluable biomarkers for PD detection^{17,18}.

ML and DL have proven to revolutionize medical data analysis and image processing, which has improved in disease diagnosis, identification, and treatment planning. Besides, ML related to algorithms that can learn from data whereas Deep learning or DL, a subfield of ML that utilizes multilayered neural network showed excellent ability to model complex patterns successfully^{19,20}. In medical image analysis DL, strategies especially CNNs have found to be extremely efficient in tasks related to image classification, segmentation, and anomaly detection. For instance, CNNs have employed in diagnoses of skin lesions at that equal efficiency to dermatologists increasing the rate of early Melanoma detection, interpretation of mammography to assist in the detection of breast cancer enhancing precision while minimizing false positive results. Some investigations have been to reapply more innovative DL models, including YOLO and transformer architecture, for identifying the anomalies like masses present in mammograms and found the functionality of these models in overshadowing the conventional imaging difficulties^{21,22}. DL and ML methods have also been used to analyze diseases in general body areas, CT scans, and MRIs for diseases, including Alzheimer's, lung cancer, and cardiovascular diseases. These models facilitate the accurate identification of a disease as well as timely treatment by revealing small patterns within large sets of medical information. Nevertheless, issues like compute requirement, model interpretability, and requirement for a variety of good quality data are still core^{23,24}. The contemporary studies aim at addressing these limitations to enhance the utilization of ML and DL in healthcare. Furthermore, DL and ML are not only limited to imaging to process structured and unstructured data such as patient records and genomics for the discovery of true personalized medicine.

The proposed study also presents a Modified Band Pass Filter and Stacked AutoEncoders (SAE) in the Feature-Based Deep Neural Network (FB-DNN) for the early detection of Parkinson's disease. The exact preprocessing pipeline is another promising aspect in which the Modified Band Pass Filter increases and decreases specific Parkinsonian voice characteristics, including tremors, jitter, and pitch fluctuations. This adaptive filtering method is much better than conventional blind filtering approaches that have low target selectivity; it thus greatly improves the quality of the input data by increasing the signal-to-noise ratio. The SAE presents a major improvement in terms of feature representation by correctly modelling non-linear hierarchical structures in audio data. In contrast to simple linear dimensionality reduction methods like PCA, which overlooks slight changes in vocal features related to Parkinson's, the SAE extracts such features. This helps to prevent overfitting and guarantees that non-redundant and highly specific biomarkers of the disease will be extracted.

The most important result of this study is because:

- This research presents an innovative approach, the Feature-Based Deep Neural Network (FB-DNN), which is specifically designed for the detection of PD. FB-DNN represents a novel and effective method for addressing the diagnostic challenges posed by this neurological condition.
- The methodology employed in this study leverages the power of audio signal analysis as a means of PD detection. This approach recognizes that audio signals can carry valuable diagnostic information and provides a unique perspective for assessing PD.
- Autoencoders, as a fundamental component of the proposed method, are selected based on their exceptional feature extraction capabilities.
- In parallel, DNNs are recognized for their proficiency in intricate pattern recognition. As such, they serve as the ideal choice for the subsequent classification phase within this novel approach.
- By Ensemble the strengths of Autoencoders and DNNs, this research aspires to elevate the accuracy and effectiveness of PD detection.

The subsequent sections of this work have been structured as follows: Sect. 2 provides an overview of existing approaches to PD classification. Section 3 introduces the proposed technique, FB-DNN, followed by the presentation of research findings in Sect. 4, which are subsequently discussed. Lastly, Sect. 5 presents the study's conclusion.

Related works

In recent years, the utilization of ML models in PD research has gained significant traction, driven by collaborative efforts between the fields of biomedicine and computer science. PD, a condition affecting over 10 million individuals worldwide, holds the distinction of being the second most prevalent neurological ailment, trailing only Alzheimer's disease. Instead, medical experts rely on an in-depth clinical assessment of the patient's medical history for diagnosis. However, a notable challenge emerges in achieving timely and accurate diagnosis, as illustrated by studies conducted by the National Institute of Neurological Disorders. It is revealed that the precision of initial diagnosis, particularly within the first five years following the onset of symptoms, stands at a modest 53%. While this level of accuracy may not represent a significant improvement over random guesswork, early diagnosis remains paramount for effective and timely interventions. In addressing this critical diagnostic challenge, a promising approach is the utilization of feature vectors that are weighted based on their relevance. These feature vectors are harnessed for classification purposes, with methods such as linear discriminant analysis (LDA), K-nearest neighbor analysis (KNN), decision trees (DT), and neural networks (NN) being explored, as demonstrated in a notable study conducted by²⁵. The research incorporates data from the UPDRS, which encompasses motor symptoms used in the diagnosis of PD. A head-to-head comparison with other state-of-the-art approaches underscores the effectiveness of the UPDRS-based parameters in achieving the highest success rate in PD diagnosis.

In contemporary healthcare settings, there is a notable shift towards the adoption of noninvasive tele-monitoring as a viable alternative for disease screening. This approach exhibits promise in delivering reliable and cost-effective diagnostic assessments while reducing the need for frequent visits to medical clinics. As a result, the financial burden on patients is lightened, leading to more precise clinical status evaluations. The data collected through tele-monitoring can be leveraged to accurately predict the severity of PD symptoms. The optimization of model parameters has employed to enhance prediction accuracy, as demonstrated in a study by²⁶, which involved an array of models, including linear regression, decision trees, random forests, gradient boosting, and XGBoost. Notably, the random forest model demonstrated superior overall accuracy, as assessed by R2, and offered the best results in terms of RMSE (Root Mean Square Error). The study also delved into the relationship between the total UPDRS and the motor UPDRS, two dependent outcome measures. Impressively, a correlation of 0.947 was established between these two dependent target variables, underscoring the potential for data-driven insights in PD assessment.

The manifestation of rhythmic shaking or tremors in a limb is indicative of PD, a pathological illness that impacts the functioning of the nervous system. An individual diagnosed with PD has symptoms that are characteristic of PD. The aforementioned ailment hampers motor function, impairs verbal communication, diminishes job execution capabilities, and renders writing a time-intensive endeavor for individuals afflicted by it. The current prevalence of PD is estimated to be between 1 and 2 cases per 1000 individuals. The prevalence of PD increases proportionally with advancing age, exhibiting a notable impact on around 1% of those aged 60 years and above. Despite the significant breakthroughs in technology, the task of early illness diagnosis remains a formidable challenge in contemporary times. Consequently, it is imperative to devise automated techniques grounded on ML to aid doctors in effectively detecting this ailment at its first stages. Existing literature provides evidence for the association among PD and the cardiovascular system. Existing literature highlights the correlation between PD and the cardiovascular system. In line with this, the primary objective of this study is to employ the Extreme Gradient Boosting ML technique²⁷ to investigate the relationship between PD and cardiovascular disease, shedding light on potential links and insights.

Dopaminergic neurons, which are responsible for the production and release of dopamine, a neurotransmitter crucial to intracellular communication in the brain, are disrupted in PD. Dopamine-producing cells in the brain are responsible for the regulation, adaptability, and fluidity of movement. Motor function is severely impaired by PD, a fatal degenerative neurological condition. There is currently no effective treatment for this highly prevalent neurological illness, which leads to functional disability and shorter life expectancy. More than 90% of people with bradykinesia also experience speech difficulties, tremors, rhythmic shaking, and a delay in their motor skills. Several ML techniques, including Support Vector Machines (SVM), KNN, Random Forest (RF), and Logistic Regression (LR) models, can be used to make predictions about PD²⁸. Instead of using KNN Model, a RF Model has been used for predicting PD. Prediction accuracy is crucial, and early on in a patient's recovery, the RF approach shows improved precision. The utilization of ML holds the promise of unlocking more effective diagnostic strategies for PD.

The acoustic analysis of human voice stands out as a cost-effective and non-intrusive method for PD detection, characterized by its reliability and simplicity. The early signs of PD are often marked by alterations in vocal quality, deviating from an individual's typical vocal patterns. Both individuals with and without PD employ nonlinear dynamic approaches to assess the complexity of their sustained speech signals. The integration of these algorithms holds substantial promise for advancing e-healthcare systems for patients, streamlining treatment processes, and mitigating the severity of illnesses. In a comparative study involving classifiers like Adaptive Boost, Gradient Boost, Light Gradient Boost, and XGradient Boost, a wide range of performance metrics was assessed²⁹. The dataset was partitioned, with 30% reserved for evaluation and 70% used for training. Notably, XGradient merged as the top performer, achieving an impressive success rate of 87.39%. Additionally, the study identified key features critical for the classification of individuals with PD through a feature significance analysis.

Recent approaches in Parkinson's disease detection have shown the opportunity of deep learning architectures as well as new data modalities. For example, an algorithm for Deep Learning-Based Prediction of Freezing of Gait in Parkinson's Disease with the Ensemble Channel Selection Approach³⁰ was designed for detecting FoG events using ACC signals from sensors located on the ankle, leg, and trunk. The authors suggested using a new architecture of a bottleneck attention module combined with a bidirectional long short-term memory network (BiLSTM), which was then improved into a convolution-bottleneck attention BiLSTM (CBA-BiLSTM). The work comes close to this realization the realization of the need to find ways of handling real-time applications with least complexity a factor that is crucial when scaling of diagnostic systems in the medical field. Hence, despite the fact that this study was concerned with gait-related features, its focus on feature optimization and real-time utilization corresponds to the goals of the FB-DNN model, especially in terms of functionality enlargement and speed. Freezing of gait prediction specifically considers only movement data from ankle, leg, and trunk sensors and hence cannot be generalized to other symptoms of Parkinson's, including the loss of voice. Furthermore, specialized hardware used for data collection proves to be a hindrance in terms of scalability especially in resource constrained environments. While the group channel selection strategy has reduced computational costs, it does so at the cost of numerous new implementation problems and increased impracticality for real-world applications in open settings.

ElasticNet-Based Vision Transformers for Early Detection of Parkinson's Disease³¹, also presents the use of Vision Transformer models integrated with convolutional neural networks (CNNs) to diagnose Parkinson's disease at an early stage. This approach went for handwriting data, moreover; the features were extracted from the input images using Vision Transformers on ViT_B_16 backbone and then opting for the ElasticNet selection algorithm as a feature selection criterion. The use of Vision Transformers and ElasticNet deeply complements the idea of adopting innovative feature extraction strategies along with effective features selection to achieve high accuracy diagnosis. Although the study was performed on handwriting data and not audio-based, extending from the work to inform other fields is possible because the methodology of feature selection and model training used in this study offers valuable information for boosting the classification pipelines of the FB-DNN framework. For instance, the proposed FB-DNN model eliminates these limitations through audio-based features and helps us detect other biomarkers for Parkinson's disease essential to early diagnosis. It does not require specialized hardware or specific inputs from patients and hence is more flexible and highly applicable in multiple contexts and different from other approaches. With a moderate computational load, this approach should be more effective in real-world applications than Vision Transformers or an ensemble of models.

In the realm of PD diagnosis, several limitations characterize existing models. First, a significant constraint lies in the limited usage of speech data, with many current models relying on a restricted volume of such data. This constraint can compromise the accuracy and reliability of the diagnostic process. Additionally, certain models fail to adequately preprocess or enhance input audio data, which, in turn, can lead to suboptimal results and reduced diagnostic accuracy. Small datasets are another impediment, as some models operate with limited data, potentially failing to capture the full spectrum of PD symptoms and variations, thereby restricting the generalizability and robustness of the diagnostic model. Consequently, these limitations can result in reduced model accuracy, both during the training phase and in practical diagnostic applications. To overcome these constraints, it is imperative to address them by harnessing larger and more diverse speech datasets, enhancing audio data preprocessing techniques, and refining model training processes. This concerted effort can significantly enhance the accuracy and reliability of PD diagnosis models. In recent years, there is a growing interest in applying machine learning and deep learning techniques in detecting Parkinson's disease. For instance, PCA was used for feature reduction, while SVM, KNN, and Random Forest models have also been used for classification. However, such methods are poor at identifying more complex patterns and correlation structures that can be characteristic of Parkinson's-specific audio features. Moreover, while a large number of existing, more traditional papers include problems that require complex feature extraction by hand, their applicability is often doubtful at best. These challenges are surmounted in this study by the extension of AutoEncoders (AE) into Stacked AutoEncoders (SAE) that allow both hierarchical and nonlinear feature extraction. While constrained to linear transformations, the SAE extracts feature such as tremors, jitter, and pitch that are important for PD diagnosis. Likewise, Deep Neural Networks (DNNs) are relatively new in the learning paradigm due to their capability of modeling high level abstractions in the data. Unfortunately, previous works on detecting Parkinson's disease using DNNs use standard data pre-processing that does not accentuate specific disease characteristics. The work also presents a Modified Band Pass Filter which is a new filter appropriate for use as a preprocessing for signals derived from Parkinson's related audio characteristics. However, more complicated architectures such as CNNs, RNNs and Transformer-based models have been used in other fields and may improve Parkinson's detection on the basis of audio but are computationally heavy and require extensive training data which is scarce in this case. The FB-DNN can thus be useful in real life applications like telemedicine or remote exams, while not being overly complicated to perform a great deal of computation to be used in robust academic applications.

Methodology

This study introduces an innovative approach to detect PD using audio signal analysis in conjunction FB-DNN techniques. The preprocessing phase involves the utilization of the Modified Band Pass Filter, a crucial signal processing technique that enhances the quality and specificity of the audio signals under investigation. To extract meaningful information from the audio data, we leverage Autoencoder, an artificial neural network specialized in feature extraction. This neural network is adept at capturing intricate patterns within the audio signals, effectively transforming high-dimensional data into a more concise representation. This process aids in noise reduction while highlighting relevant information within the audio signals. These extracted features serve as the foundational elements for subsequent classification tasks. An ensemble method is used to achieve this classification between audio samples indicative of PD and those from individuals without PD.

Data Collection

The Italian Parkinson's Voice and Speech (IPVS)³² serves as the fundamentals of this study because it provides a rich database containing recordings of voice and speech samples from patients with PD. The decision to focus on audio-based data for PD detection grounded in several advantages that audio signals offer:

- **Non-invasive Nature:** Audio data can collect in a way that is not invasive to the patient, thus allowing convenient repeated measurements to privileging discomfort to the patient.
- **Cost-effectiveness:** Sound signal recording and analyzing costs considerably less than other diagnostic techniques, including neuroimaging.
- **Early Detection Potential:** Variation in voice features predates motor signs, giving a chance for early identification because of audible differences.
- **Wide Accessibility:** Diagnosing tools based on audio can performed from a distance, allowing practitioners to use telemedicine preferences as convenient for the patients, regardless of their location.

However, we understand that a huge weakness of this study stems from the data available to us in this case. The chief drawback of the IPVS dataset is that it only consists of recordings of Parkinson's patients in Italy. This geographic and demographic restriction may reduce the generalizability of the finding and reduces the number of diverse audio features that can captured. Further improvement of the results might achieve in further studies by using larger datasets containing patients of different demographics in order to increase the generalizability of the presented method. The dataset does not mean that people are in the early, middle, or late stage of Parkinson's disease.

Data preprocessing

In the preprocessing phase of our research, a carefully orchestrated series of operations is put into motion, centered on the application of the Modified Band Pass Filter to the incoming audio data. The Modified Band Pass Filter was preferred due to the specific features it has in analyzing audio signals for PD. This filter outperforms other signal processing techniques due to the following features:

- **Targeted Frequency Isolation:** Unlike other filters called standard filters, the modified band pass filter actually expected to filter out the frequencies related to Parkinson disease related voice features such as tremor, jitter and pitch perturbation. Because of this, it is more suited for the identification of disease relevant features due to its specificity.
- **Improved Signal Clarity:** This technique reduces the impact of noise and non-salient features that could be included while applying conventional filter techniques such as simple band-pass or low-pass filters.
- **Dynamic Adaptability:** The Modified Band Pass Filter presented as the most suitable because it modifies the parameter of the filter according to the characteristics of the sample, and the fixed-parameter filters cannot match it.
- **Efficiency in Feature Extraction:** The filter enhances the audio signal in a way that is computationally efficient for following analysis stages which makes the technique more efficient for instance using wavelet transforms that normally entail computations that are more elaborate.

Following this, we delve into the intricacies of the signal by calculating its half size, a critical metric that facilitates the subsequent filtering processes. To delve even deeper into the signal's characteristics, we harness the power of the Fast Fourier Transform (FFT), a mathematical tool indispensable for unlocking the frequency domain details hidden within the audio data. After the FFT analysis, we introduce the band pass filter function, thoughtfully crafted to sift and extract specific frequency components from the signal, isolating the vital information required for further analysis. In the subsequent step, we embark on a precise calculation of the low and high cutoff frequencies, following well-defined formulas. The low cutoff frequency, known as "low wt," is meticulously derived, while the high cutoff frequency, referred to as "High wt," is carefully computed. Lastly, the band pass filter is applied to the signal, culminating in a meticulously preprocessed audio dataset. This refined dataset is now poised for comprehensive analysis and the extraction of features that will lay the groundwork for the subsequent stages of our research.

$$low\ wt = \frac{1}{(2 \times \pi \times f_s \times S \times minvalue\ of\ signal)} \quad (1)$$

$$High\ wt = \frac{(2 \times \pi \times f_s \times S \times maxvalue\ of\ signal)}{(f_s * 100)} \quad (2)$$

Where, *low wt* represents the low cutoff frequency, *High wt* represents the high cutoff frequency, signal's sampling frequency f_s , Order S .

Stacked Auto Encoder.

The Stacked Autoencoder (SAE) has chosen for feature extraction owing to its superior performance in identifying latent features and in general, nonlinear patterns in the audio data. Unlike traditional dimensionality reduction techniques like Principal Component Analysis (PCA), the SAE provides several key advantages:

- **Nonlinear Feature Extraction:** Although analyses using PCA and other similar methods may only provide linear transformations, the SAE is able to capture nonlinear relationships within the data, where they are essential for defining the essential subtle voice features linked to Parkinson's disease (PD).

- **Hierarchical Representation:** The multi-layer structure of the SAE makes it possible to extract the features of different layers, thus leading to a deeper representation of the data with comparative to the single layer transforms such as the PCA.
- **Data-Specific Optimization:** In contrast to other methods, designed for abusing the specific data source, the SAE adapts the transformations corresponding to the structure of the Parkinson's voice dataset during the training phase.
- **Noise Reduction:** The SAE itself already involves denoising as one of its steps in extracting features from a given signal, and those results in clear signals without the need to undergo any additional steps for signal denoising.

Autoencoders (AEs) were preferred for feature extraction over other unsupervised learning algorithms because of its ability to learn hierarchical and non-linear interactions in the audio data UT for Parkinson's disease. In contrast to other approaches, for example PCA, which can only perform linear transformations, AEs are capable of discovering such features as tremors and jitter, which are important for diagnosing Parkinson's. Moreover, AEs inherently denoise in the sense that they shed off useless noise while preserving explicitly important information a perfect feature for clinical datasets. While compared with some other architectures such as CNNs or VAEs, AEs mainly target at reconstructing the input data and learning concise disease-specific feature while without the need of further generative or spatial processing. Because AEs are flexible and efficient in learning and extracting features, which are important in this context, they are more suitable for this task.

The Stacked Encoder shown in Fig. 2, is a fundamental component in DL and neural network architectures, often used in tasks such as feature extraction and dimensionality reduction³³. In this particular context, the Stacked Encoder is applied to the input signal, which is initially sized appropriately at the beginning of the process and maintained throughout its completion. The Stacked Encoder's operation involves a series of transformations applied to the input signal. Firstly, the input is processed through a linear transformation represented by y .

$$y = f_e(w_x + b_1) \quad (3)$$

The sigmoid function as well as the Rectified Linear Unit (ReLU) are two examples of activation functions frequently utilized in this setting. As indicated by the notation $f_e(x)$, the sigmoid activation function makes the alteration process non-linear. Specifically, helpful for applications like binary categorization as well as probability estimation, it transfers the input to values that range from 0 to 1. The sigmoid function's curve is often S-shaped.

$$f_e(x) = \frac{1}{1 + e^{-x}} \text{ (Sigmoid)} \quad (4)$$

On the other hand, the ReLU activation function, represented as $f_e(x)$, introduces piecewise linearity into the transformation. It solves the issue of disappearing gradients that arises when employing sigmoid or hyperbolic tangent (tanh) activation functions in DNN as well as is extremely computationally effective.

$$f_e(x) = \max(0, x) \text{ (ReLU)} \quad (5)$$

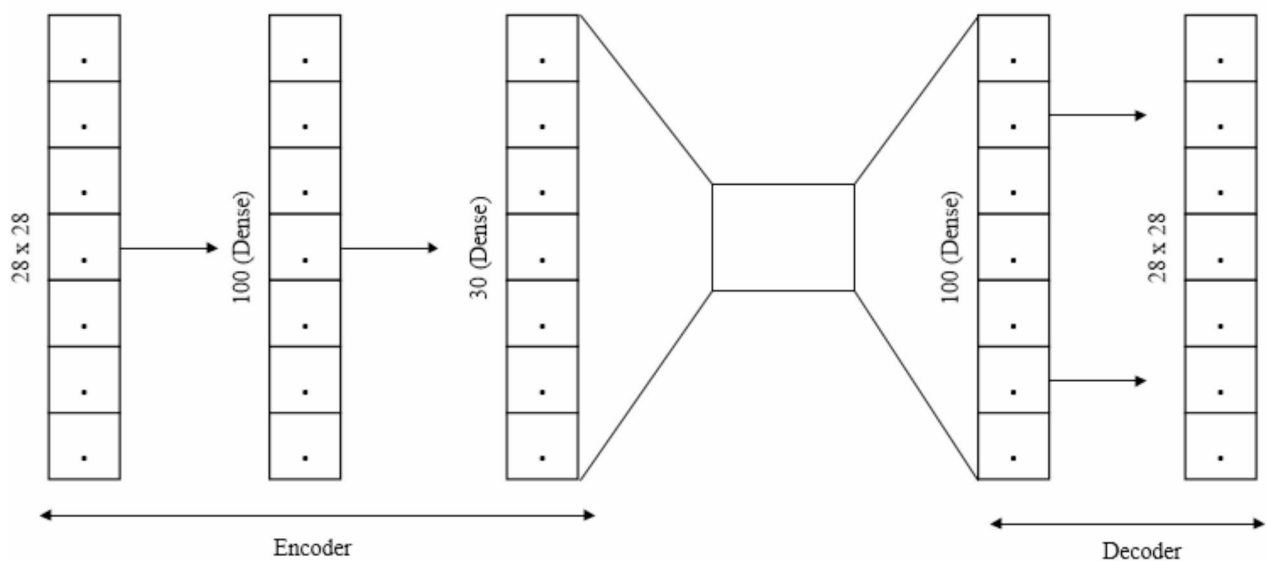


Fig. 2. Stacked Auto Encoder.

The SAE transforms input data x into a compressed latent representation z and reconstructs the input as \hat{x} . The encoding and decoding functions are:

$$z = \sigma(W_c x + b_c) \quad (6)$$

$$\hat{x} = \sigma(W_d z + b_d) \quad (7)$$

Where W_c and b_c are the weights and biases of the encoder, W_d and b_d are the weights and biases of the decoder, and σ is the activation function. The reconstruction loss L_{recon} is minimized during training:

$$L_{recon} = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (8)$$

Where N is the number of samples.

In the context of feature extraction from audio data using a Stacked Autoencoder (SAE), the process involves several steps that transform the raw audio signals into meaningful and compact representations, including features like MFCC (Mel-Frequency Cepstral Coefficients), harmonic information, and percussive information. The process begins with a set of audio files as input. These audio files can contain various speech of PD and Non-PD Patients. These files are typically represented as time-domain waveforms, where the amplitude of the signal is sampled at discrete time intervals. Before feeding the audio data into the SAE, Modified Band Pass Filter has been applied. This includes tasks like resampling the audio to a consistent sample rate, removing noise, and ensuring a uniform format for analysis. The heart of this process is the Stacked Autoencoder, which is a type of artificial neural network. The SAE is trained to automatically learn a hierarchy of features from the raw audio data. It consists of multiple layers of encoding and decoding, where each layer captures progressively abstract representations of the input audio. One of the key features extracted by the SAE is MFCC (Mel-Frequency Cepstral Coefficients). These coefficients represent the spectral characteristics of the audio signal. They are particularly useful for tasks like speech and audio classification. MFCCs capture information about the distribution of spectral energy in different frequency bands.

In addition to MFCCs, the SAE also extract harmonic features. Harmonic analysis is essential for tasks related to music and speech processing. These features can capture characteristics related to pitch, harmonics, and tonal aspects of the audio. Percussive features represent elements of an audio signal associated with rhythmic and percussive components. These are important for music analysis, beat tracking, and rhythm-related tasks. Depending on the application, the extracted features may be aggregated over time to create a feature vector that summarizes the characteristics of the entire audio clip. Aggregation methods may include mean, variance, or higher-order statistical moments. These extracted features are often more informative and less computationally intensive than working directly with raw audio waveforms. By using a Stacked Autoencoder for feature extraction from audio files allows for the automatic generation of compact and discriminative representations, such as MFCCs, harmonic features, and percussive features, which are crucial for audio processing tasks. This process enhances the ability to work with audio data efficiently and effectively.

AEs have incorporated in the proposed FB-DNN because of their capability to extract expressive, concise, and high-quality features from large and diverse data streams including the audio streams. While methods like the Principle Component Analysis (PCA) provides only for a linear transformation of features, the AEs can model non-linear relationships very effectively because identifying subtleties like tremors, jitter, or irregular pitch that are traditionally indicative of Parkinson's disease, requires non-linear transformations. Moreover, AEs naturally denoise since they discover latent features of the input data, minimizing unnecessary fluctuations, which is important when identifying disease characteristics in noisy audio signals. The stacked architecture enables them to learn hierarchical structures, meaning that they are capable of extracting features at multiple layers with the complexity and relevance of features increased towards higher levels. Furthermore, AEs learn feature extraction in data-driven mode to generate their outputs according to the dataset characteristics and decrease computational resource demand due to the transformation of raw data into compact latent space. This efficiency would be specifically helpful to the classification stage of the FB-DNN where the non-Parkinsonian and Parkinsonian cases can be well distinguished. Since the proposed FB-DNN can overcome the shortages of the traditional methods and better suit the requirement of Parkinson's disease identification, the integration of AutoEncoders will further improve the capabilities and versatility of the FB-DNN.

Flow Diagram.

The goal of an FB - DNN, a special form of ANN, is to learn and extract features from data. In contrast to more typical ML designs, which rely on expert-designed feature engineering, feature-based DNN may learn meaningful patterns and features autonomously, straight from the input data.

These networks often include several layers, such as convolutional layers for pictures or recurrent layers for sequential data that extract more and more abstract as well as hierarchical information as they go along. Activation functions allow the model to be non-linear, therefore allowing it to capture intricate connections within the information being captured. These networks excel at tasks like image recognition as well as natural language comprehending where feature extraction is crucial for precise forecasts and categorizations by iteratively adjusting internal variables during training, frequently utilizing methods like backpropagation and gradient descent. Figure 3 show the flow chart of proposed system.

Deep Neural Network - DNN.

The DNN is an ideal and highly effective strategy of ML capable of applying in various fields³⁴. In this specific ANNs structure, the information, which input to the network, undergoes a series of transformations throughout layers of interconnected neurons. Data fed in a DNN at the input layer and goes through several hidden layers

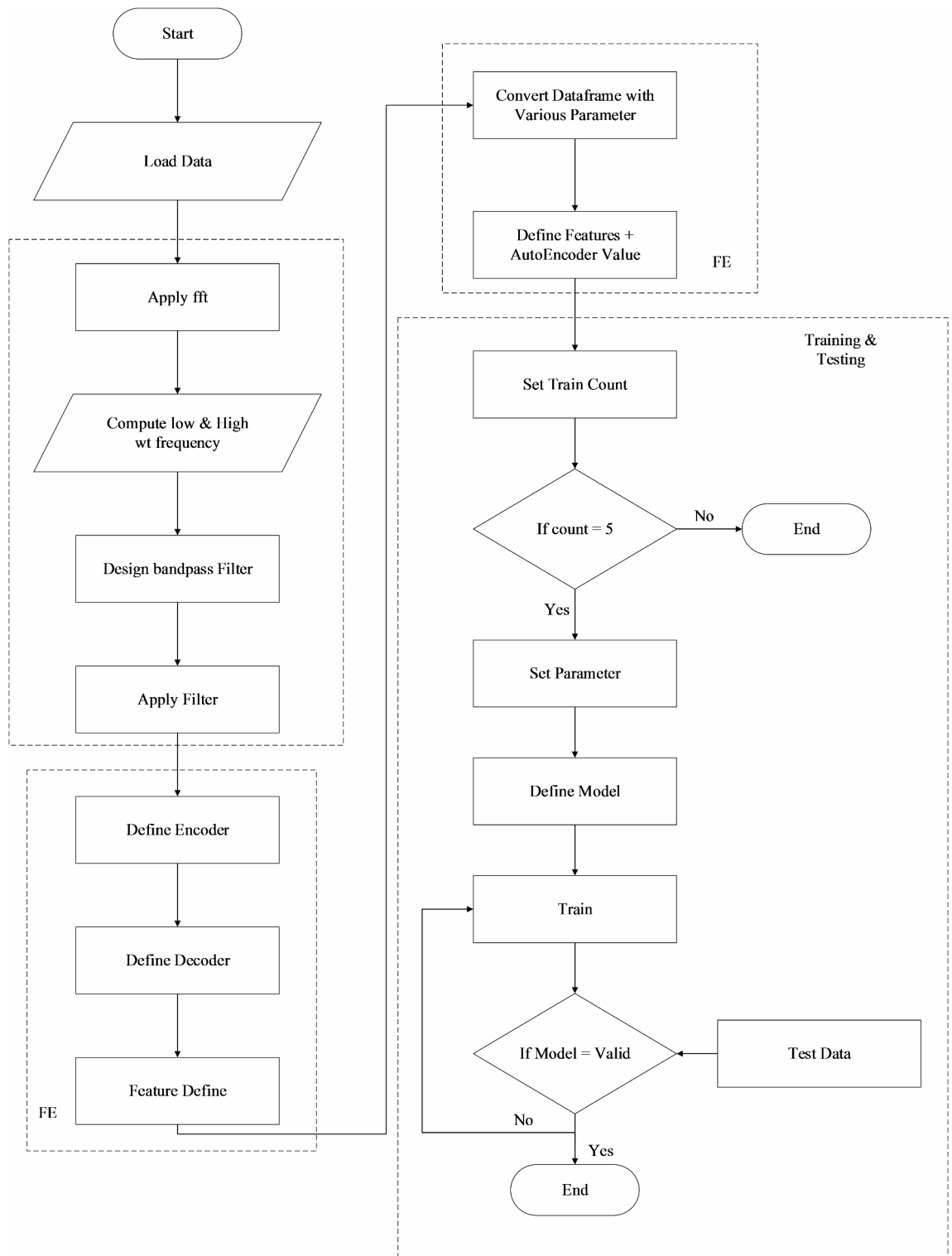


Fig. 3. Flow Chart of Proposed System.

before going through the last layer where forecasts and classifications made. Every layer contains many neurons that perform algebraic manipulations on the input data such as sum of inputs multiplied by weight and then an activation function is applied. The element of non-linearity also arises out of these activation functions, which provide the model a great potential to interpret elaborate data structures and representation. A majority of the common uses of ML will require performing operations on high-dimensional inputs, which form the DNNs' strengths. With their assistance, critical features might be extracted directly from raw data—primarily no need for feature extraction.

In training, the neuronal weights and biases of a DNN learnt to minimize a loss or error function of interest through backpropagation. When the network is in the midst of this optimization process, it is able to identify patterns and correlations inherent within the training data set. The type and quantity of DNN architectures is vast and depends on a number of factors; the job to be done, number of hidden layers, number of neurons in each hidden layer, or even the type of activation function to be used. The learning rate, the size of the batch and the optimization method are the three major hyperparameters that dictate the best model to use. Overall, DNNs are a revolutionary shift in ML and artificial intelligence, which allows for the automated feature extraction process. They have contributed greatly over the year in enhancing the various fields and still in the frontline in promoting the modern research and applications in the big data and DL solving complex problems.

The choice of Deep Neural Networks (DNNs) as the most suitable model for this type of study based on the facts that DNNs possess the unique ability of being able to learn nonlinear hierarchical features from the input data and of being a very efficient and adaptable model for the identification of Parkinson's disease. It is unlike ensemble models such as the Random Forest or the Gradient Boosting where a lot of importance placed on the selection of the feature and manual injection of features into the model during model construction. It reduces the complexity of data preprocessing because the model will search for disease patterns directly in the distribution of sound waves. Although, models sharing aspects of multiple models like ensemble models combined with feature extractors are good, they bring in more difficulty in model developing and large resources needed in training and optimizing. These models may also not possess the integrated nature of DNNs which feed features to classifiers and which is an end-to-end scheme of things. In addition, ensemble models are often more proficient in rectangular numerical data than DNNs, and they can distort complex form and informative characteristics such as audio signals. The use of DNNs also suits such a model because of scalability and flexibility resulting from differences in data sets. Compared with other models, the modification of DNNs does not require large changes when extending such as adding more data or solving issues that are more difficult. Furthermore, a Stacked Autoencoder (SAE) has also introduced within the framework of the developed FB-DNN to improve the accuracy of the model, particularly in working with non-linear and hierarchical structures inherent to the data. For this particular application, therefore, factors such as the following make DNNs a better option than, for instance, ensemble models or other hybrid architectures.

The DNN processes the SAE features z to classify the input into Parkinson's or non-Parkinson's categories. For each layer l , the output $a^{(l)}$ is computed as:

$$a^{(l)} = \sigma \left(W^{(l)} a^{(l-1)} + b^{(l)} \right) \quad (9)$$

Where $W^{(l)}$ and $b^{(l)}$ are the weights and biases of layer l , and σ is the activation function. The final output \hat{y} is computed using a softmax activation:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (10)$$

Where z_i is the score for class i and C is the total number of classes.

In the realm of PD classification, the integration of DNNs with feature extraction through Autoencoder models has emerged as a powerful approach. This method harnesses the capacity of Autoencoders to capture intricate patterns and salient features within complex medical datasets. It begins with meticulous data preprocessing, ensuring data cleanliness and appropriate labeling. The Autoencoder model is then constructed and trained on this data. The input data is first mapped by the encoder into a lower-dimensional latent space, and then the decoder uses this representation to try to recreate the original data. Subsequently, the encoder portion of the trained Autoencoder is employed for feature extraction. These features represent the most crucial information distilled from the initial dataset. Once the feature extraction is complete, a separate DNN classifier is constructed, taking these extracted features as input and providing the final classification outcome, distinguishing between individuals with PD and those without. The design of this DNN classifier shown in Fig. 4, involves the selection of layers, units, activation functions, and regularization techniques, with subsequent training employing binary cross-entropy loss and optimization algorithms. Rigorous monitoring on validation data guides the fine-tuning of hyperparameters. The effectiveness of this model is evaluated using various classification metrics on a dedicated test dataset. This holistic approach facilitates a more nuanced understanding of PD by distilling complex data into meaningful features, leading to enhanced classification accuracy and potentially advancing the field of diagnosis and treatment. Ethical considerations and collaboration with healthcare professionals are essential throughout the process, ensuring compliance with healthcare data standards and the ethical principles of medical research.

The categorical cross-entropy loss L_{class} is minimized for classification:

$$L_{class} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (11)$$

Where $y_{i,c}$ is the true label for sample i and class c , and $\hat{y}_{i,c}$ is the predicted probability for sample i and class c .

Novelty in Proposed Work.

The proposed Feature-Based Deep Neural Network (FB-DNN) introduces a novel and effective approach to Parkinson's disease detection by integrating advanced techniques for feature extraction and classification. Unlike traditional models that rely on linear dimensionality reduction techniques such as Principal Component Analysis (PCA), the FB-DNN employs a Stacked Autoencoder (SAE) to extract nonlinear and hierarchical

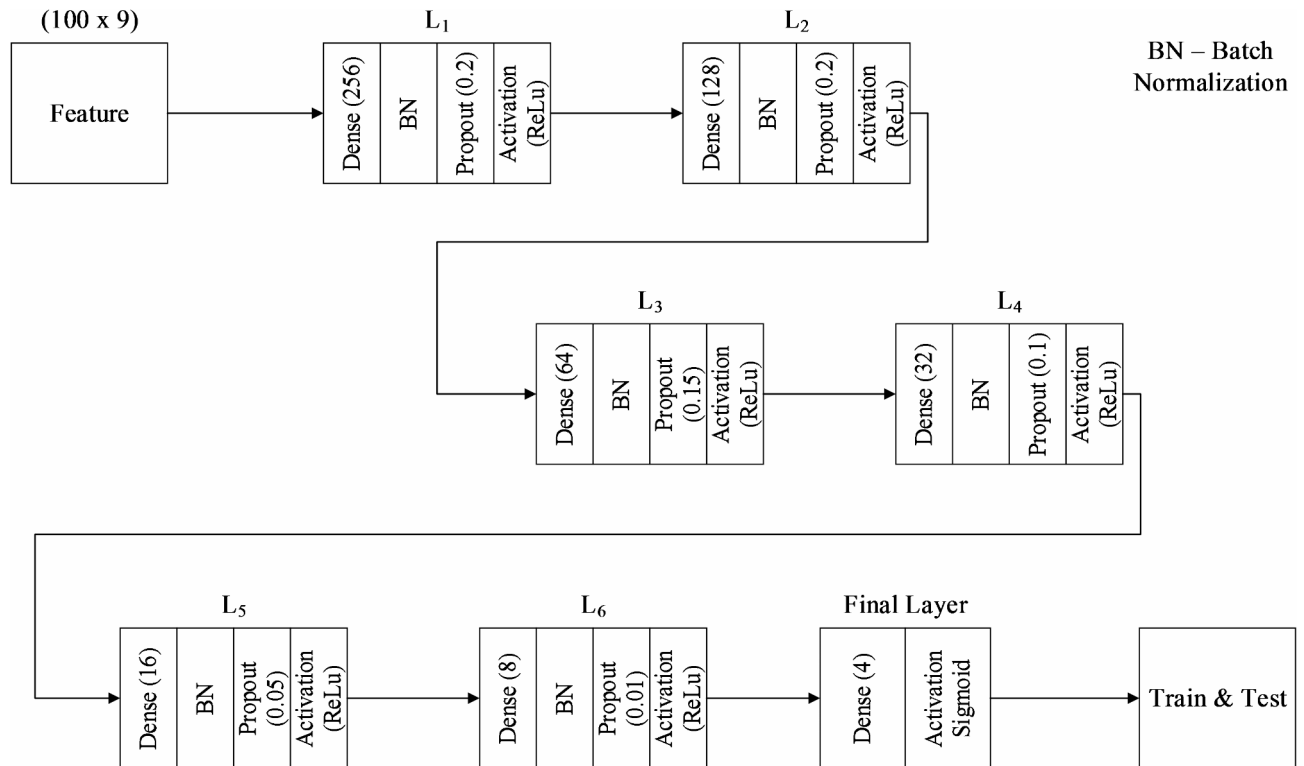


Fig. 4. DNN Diagram.

features from audio data. This enables the model to capture intricate patterns and subtle variations in voice characteristics, which are often missed by conventional methods. Additionally, the model includes a customized audio preprocessing pipeline using a Modified Band Pass Filter, uniquely designed to isolate frequency ranges associated with Parkinson's-specific vocal impairments. This preprocessing step enhances the signal-to-noise ratio and highlights critical disease-specific audio features, ensuring robust feature representation. The hybrid architecture of the FB-DNN combines the powerful feature extraction capabilities of the SAE with the classification strength of deep neural networks, providing a highly efficient framework for identifying Parkinsonian vocal characteristics. Its ability to focus on disease-specific features such as tremor, jitter, and pitch irregularities makes it particularly well-suited for early detection and diagnosis. Furthermore, the modularity of the FB-DNN allows it to adapt seamlessly to diverse datasets and clinical settings, demonstrating scalability and practical applicability. These advancements highlight the FB-DNN's potential as a transformative tool in audio-based medical diagnosis, offering significant advantages over conventional models in terms of flexibility, robustness, and precision in feature extraction and classification.

Result & discussion

The implementation of the proposed method was conducted in the Jupyter Notebook integrated development environment (IDE). The system configuration consists of a Windows 10 operating system running on a machine with 16GB of RAM and a Core i5 processor. Additionally, there is a substantial 1 TB hard disk drive, offering ample storage capacity. This setup appears to be well-suited for a range of tasks, including data analysis and ML applications. To facilitate these tasks, a selection of essential Python packages is installed. Among these are sklearn for ML, pandas for information visualization, seaborn for manipulating information, numpy for numerical calculations, scipy for scientific computing, python's standard library, and Keras for DL. With this combination of hardware and software resources, we have a robust environment capable of handling a variety of data-driven projects, from data preprocessing and analysis to model development and evaluation, making it a versatile setup for data scientists and ML practitioners. This experimental environment provided the necessary computational resources to develop, test, and evaluate the proposed method for its intended purpose. Python, with its extensive libraries and packages, offers a versatile platform for various ML and data analysis tasks, making it a suitable choice for conducting research and experiments. The Jupyter Notebook IDE, with its interactive and document-oriented interface, facilitates code development, experimentation, and result visualization, enhancing the efficiency of the research process. The specific hardware configuration mentioned ensures that the experiments were conducted on a reasonably capable system capable of handling the computational demands of the proposed method.

Loading a dataset from the IPVS Dataset, which contains audio recordings of both patients with PD and healthy individuals, represents a crucial initial step in leveraging ML for the classification or analysis of PD. The dataset's audio format encapsulates valuable vocal characteristics and patterns that can be indicative of

the disease's presence or severity. However, before any meaningful analysis or modeling can take place, it is essential to involve in preprocessing, a critical phase that often includes techniques like the Modified Band Pass Filter. Despite the fact that the study does not consider the research obstacle of recording audio data, the following measures have been taken to prevent the vagueness of noises, inconsistency in quality of microphones, and patients' behaviors during the recording. The FB-DNN model described in this paper uses configurable implementations of noise reduction filters and normalization for noise removal as part of the data preprocessing step. Such steps make it possible to extract the features that are disease related and not skewed by recording disorderliness. Furthermore, the data used in this research is a corpus built from restricted recordings that followed protocol to limit the variance from variations in microphone quality and recording backgrounds. The instructions for recording were given to patients and made an attempt was made to regulate the speed, loudness and clarity of speech among participants. Moreover, robust feature extraction is applied to the model and can handle minor inconsistencies in audio quality, using for instance dynamic filters, and feature normalization. Figure 5 shows the input audio.

The Modified Band Pass Filter is a signal processing method used to enhance specific frequency components in audio signals. In the context of PD analysis, it can be employed to isolate relevant vocal features and suppress noise. By applying this filter to the raw audio data, unwanted frequency components outside a predefined range are attenuated or eliminated. This process can improve the signal-to-noise ratio, making it easier to identify patterns and characteristics in the vocal recordings that are associated with the disease. Overall, the combination of loading audio data from the IPVS Dataset and preprocessing it with the Modified Band Pass Filter sets the stage for further analysis and modeling. It allows for the extraction of pertinent audio features that can subsequently be used in ML algorithms to classify or diagnose PD. This multi-step approach demonstrates the power of combining data acquisition from specialized sources with advanced signal processing techniques to drive advancements in healthcare and medical diagnostics. Figures 6 and 7 shows the preprocessed and various features of dataset.

The correlation matrix shown in Fig. 8, for a dataset containing 19 features from the IPVS Dataset offers a comprehensive insight into the interrelationships among these variables. This matrix quantifies the degree and direction of linear associations between pairs of features, which can be crucial for understanding the dataset's underlying patterns. The matrix's cells are every occupied by a correlation coefficient that indicates the degree

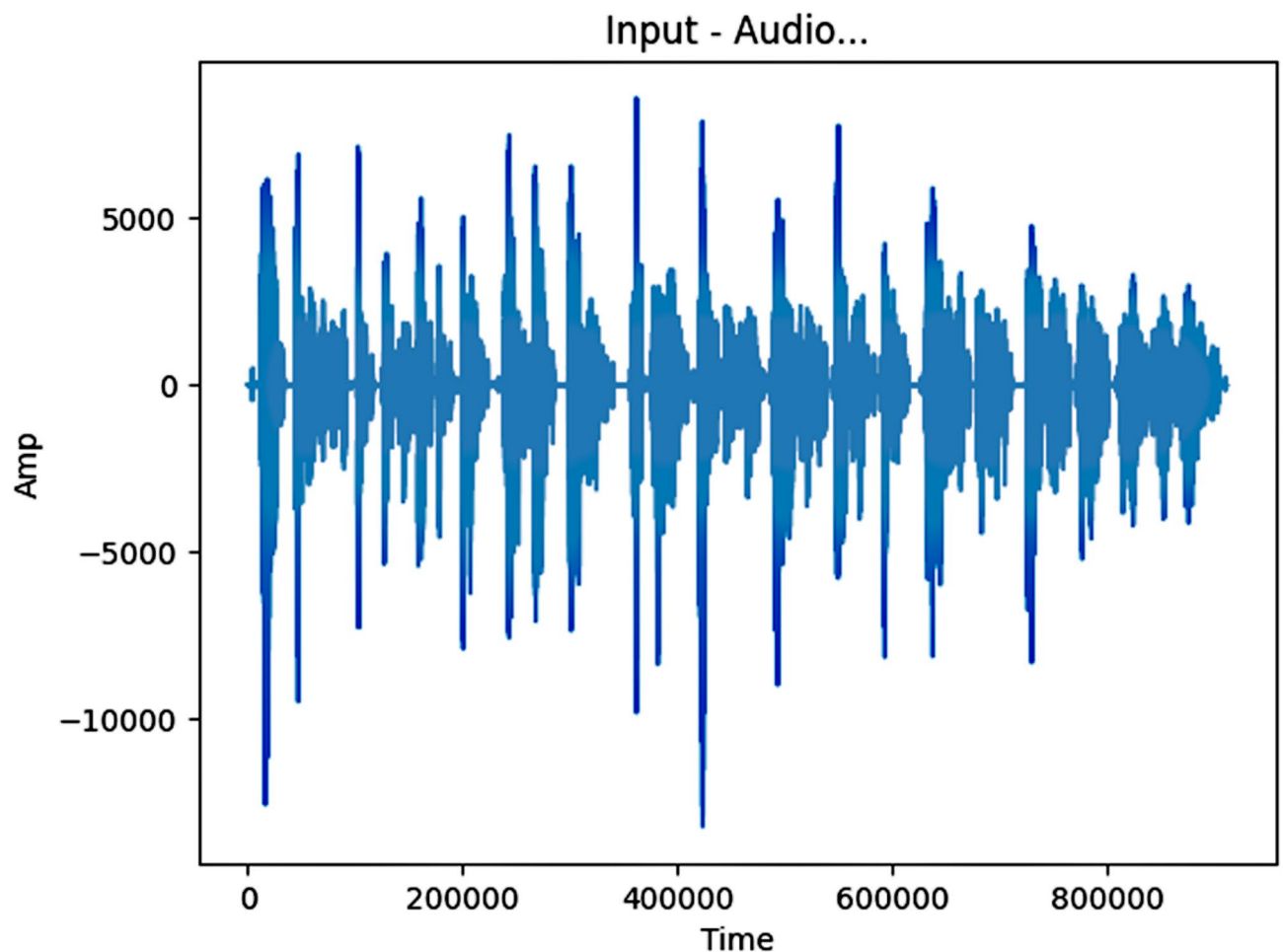


Fig. 5. Input Audio.

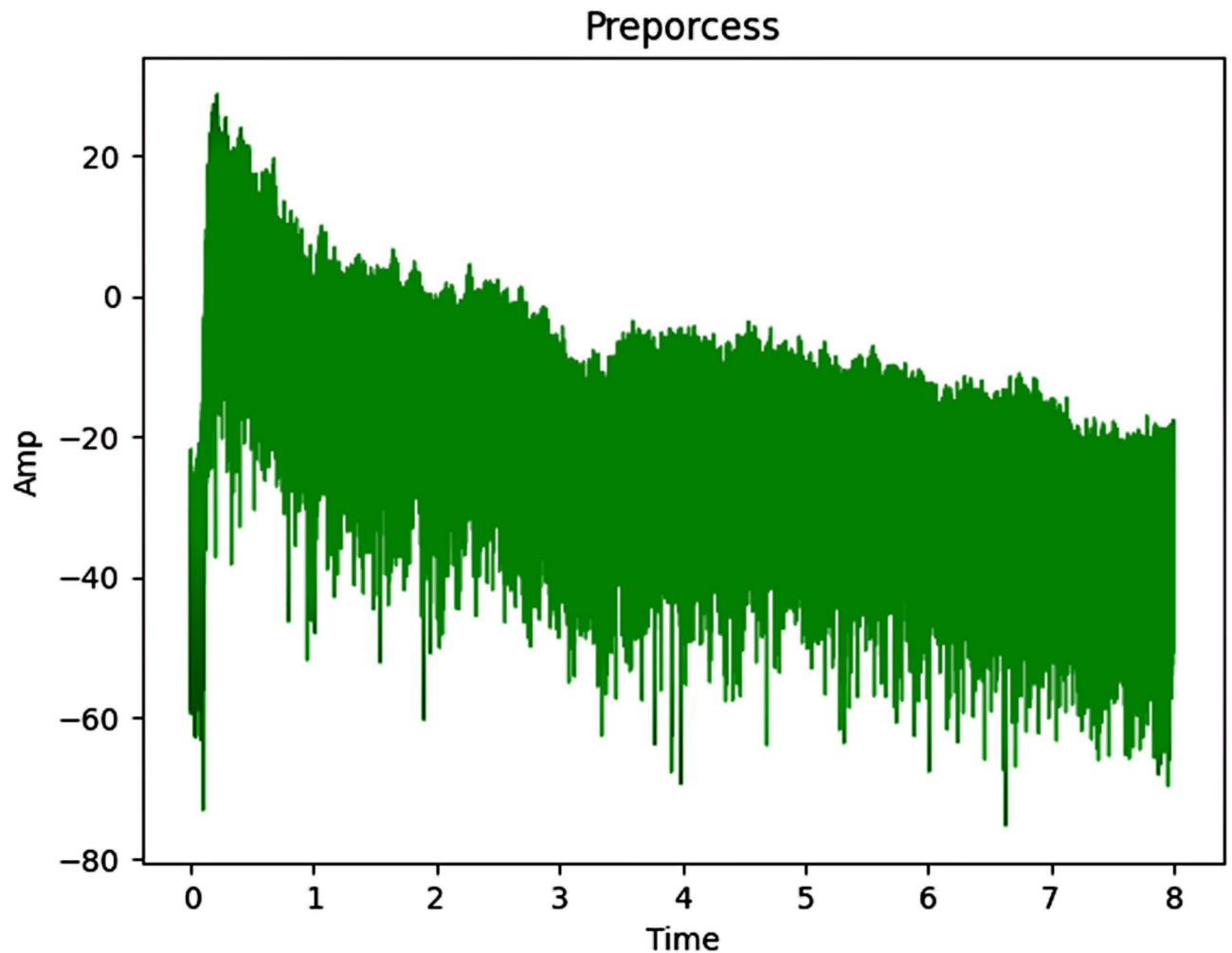


Fig. 6. Preprocess.

and direction of the association among two characteristics; these coefficients are commonly calculated using Pearson's correlation. Correlation coefficients near to 1 indicate a highly favorable relationship between the two variables. In contrast, negative correlations with values around -1 show that one trait tends to improve while the other deteriorates. Close to zero indicates almost little linearity. When analyzing correlation coefficients of a variable, or a system of variables, it becomes easier for researchers and data analysts to discover interdependence as well as patterns within the data. For example, they may find the correlation between features and learn they are highly related which outcomes feature selection and dimensionality reduction techniques. On the other hand, they may discover other variables that have low association suggesting that the variables are orthogonal. Such information is useful in further analysis like, feature extraction, selecting the best model, and or developing hypotheses, which will be tested. Altogether, the correlation matrix of these 19 features becomes an initial and efficient method for analyzing the concrete characteristics of the IPVS Dataset as well as reveals many potentially significant and promising pathways to investigate issues related to PD and voice aspects. To make the understanding of Feature-Based Deep Neural Network (FB-DNN) decision formation easier and to increase the model explainability SHAP values or, Shapley Additive explanations used. SHAP values provide the assessment at feature level that shows how the FB-DNN detects Parkinsonian vocal impairments. The features found to be most important included jitter, shimmer and the extent of the fundamental frequency related with the clinical understanding of the disease, as is manifest in patients' speech. In order to obtain an overall feature importance, a feature importance in the global scale produced, and it shows that features such as jitter improve the classification of Parkinsonian cases whereas features such as stable pitch intervals deteriorate such classification. Moreover, individual force plots extracted for illustrating the effect of the features on the respective prediction for proving that the model learned to lock onto specific patterns in each case.

Within the context of our research, we introduced a novel and powerful model named the Feature-Based DNN, or FB-DNN. This innovative approach represents a significant advancement in the realm of DL and feature extraction for complex datasets. Our approach revolves around the use of an Autoencoder, a well-known DL framework that can learn on its own and automatically extract useful characteristics from raw data. Prior to feeding the data into the DNN, we harnessed the power of the Autoencoder to perform feature extraction. This

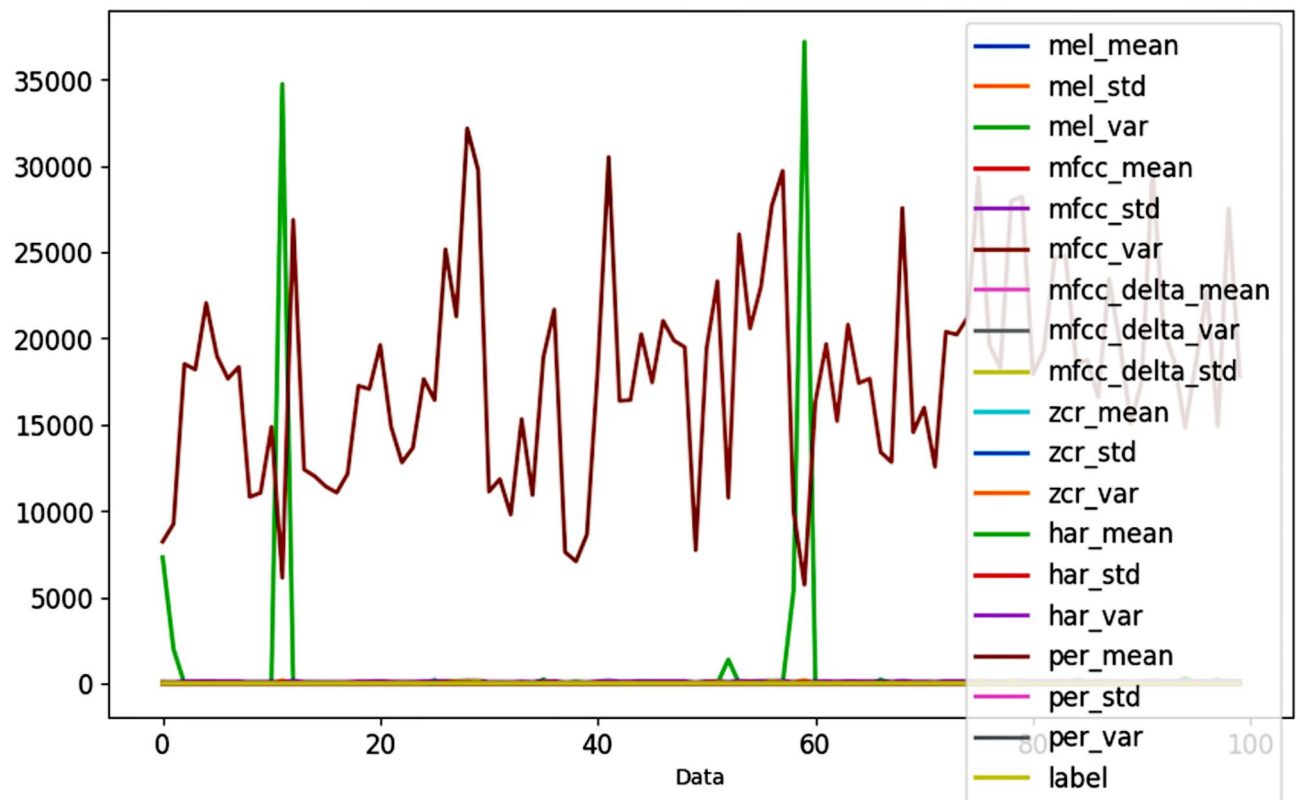


Fig. 7. Various Features of Dataset.

strategic step aimed to enhance the quality and representativeness of the feature set by focusing on the most salient and discriminative attributes within the dataset.

In the pursuit of model refinement and optimization, we embarked on an extensive training process spanning a substantial 1000 epochs. With more time to hone its internal visualizations, our FB-DNN was better able to pick up on subtleties and nuances in the training data. This extended training period was crucial for the model to reach a state of convergence, where its predictive accuracy could be maximized. During this comprehensive training regimen, we meticulously tracked and documented the model's performance, yielding two key visualization plots: "Accuracy vs. Epoch" and "Loss vs. Epoch" shown in Figs. 9 and 10. These plots served as indispensable tools for monitoring and assessing the model's progress over the course of training. The "Accuracy vs. Epoch" plot provided a clear representation of how the model's classification accuracy evolved as the epochs advanced. Meanwhile, the "Loss vs. Epoch" plot depicted the trajectory of the model's loss function over time, offering insights into its ability to minimize prediction errors. Our research represents a groundbreaking fusion of Autoencoder-driven feature extraction and DNN, culminating in the FB-DNN model. The model's effectiveness and dynamics have been better understood after 1000 epochs of training and careful investigation of accuracy and loss patterns. These findings highlight the opportunity for enhanced feature-based categorization tasks and pave the way for future developments in DL and artificial intelligence.

To control overfitting during the DNN training (performed over 1000 epochs) and improve the model's ability to generalize to unseen data some regularization and validation techniques applied. During training, figuring out overfitting, a dropout of 0.3 applied to the hidden layers randomly shift the neurons off the active network. Weight regularization was done through appropriate addition of a small penalty term ($\lambda = 0.001$) in the loss function to avoid large weights. Batch Normalization applied after each hidden layer to enhance learning and increase the rate of convergence and it helps to avoid overfitting. The training further supervised by a validation set and training stopped for a proper validation in case of extreme overtraining (when the validation loss stopped improving for 50 epochs). During training, the model validated on another dataset of documents specifically for this purpose. The values of validation loss and accuracy tracked for analysis of the trends of all training and validation sets.

Performance Evaluation.

In the context of assessing the performance of a diagnostic or classification model for PD, several key metrics play a crucial role in evaluating its effectiveness. Measures such as these shed light on how well the model can distinguish between people with and without PD. Here is a detailed description of these metrics:

Classification models, such as those used to categorize cases of PD, may be evaluated with the use of a confusion matrix, as illustrated in Table 1. It helps evaluate the accuracy of the model by detailing the forecasts and the actual class labels. The confusion matrices of many models are displayed in Fig. 11. PD categorization is an example of a binary designation issue, and the confusion matrix for such a task normally has four values:

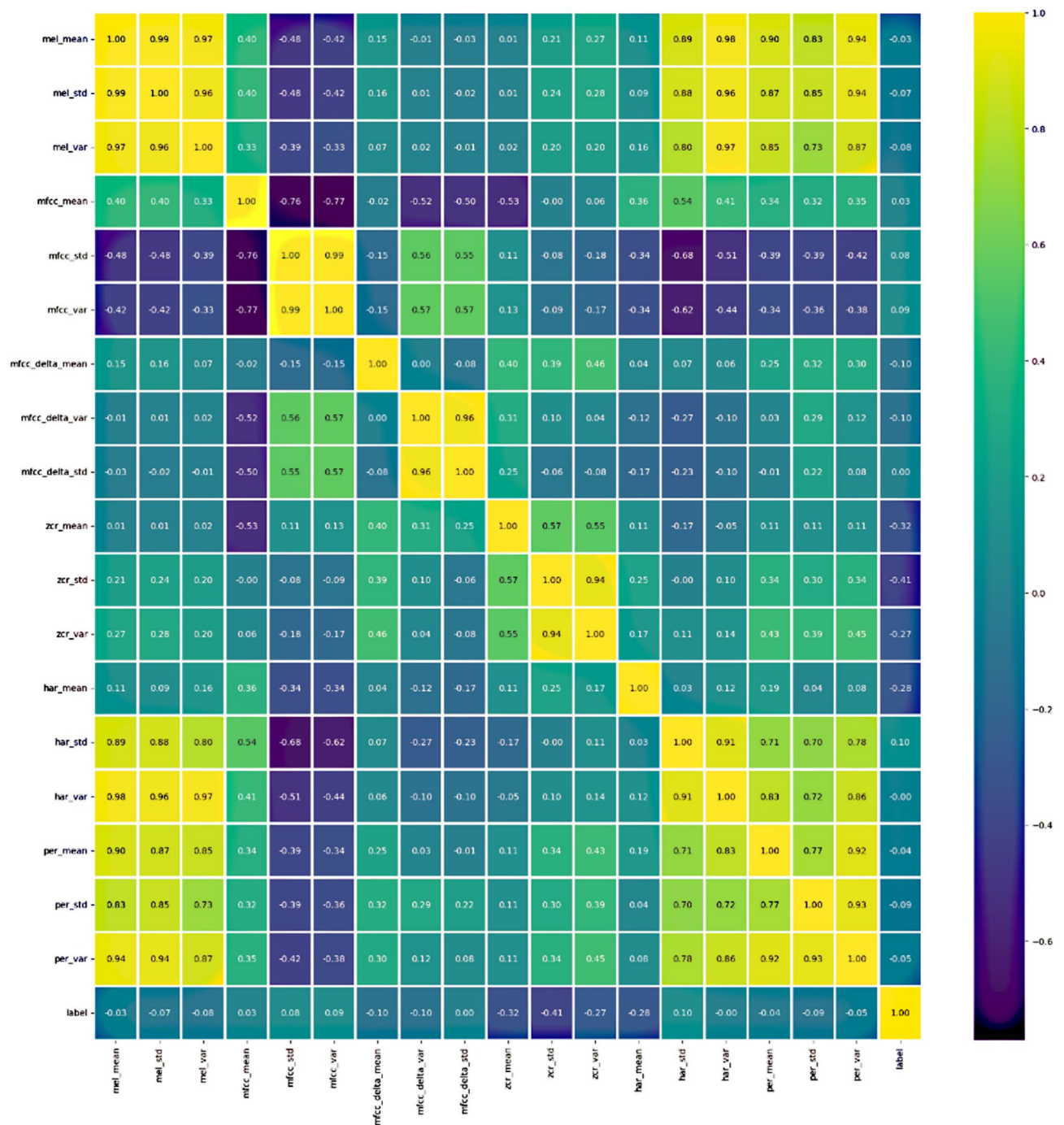


Fig. 8. Correlations Matrix.

True Positives (TrPos): These are instances where the model's diagnosis of PD was spot-on.

True Negatives (TrNeg): Such instances represent successes for the model in ruling out PD.

False Positives (FaPos): These are examples of Type I errors, in which the model mistakenly projected that a patient had PD.

False Negatives (FaNeg): This is an example of a Type II mistake, where the model wrongly indicated that a patient did not have PD.

Accuracy:

For categorization tasks like PD diagnosis, accuracy is a simple and widely employed statistic. It is a metric that evaluates the percentage of cases (positive and negative) that were correctly categorized. When the model's accuracy is high, it usually means that its forecasts are spot on.

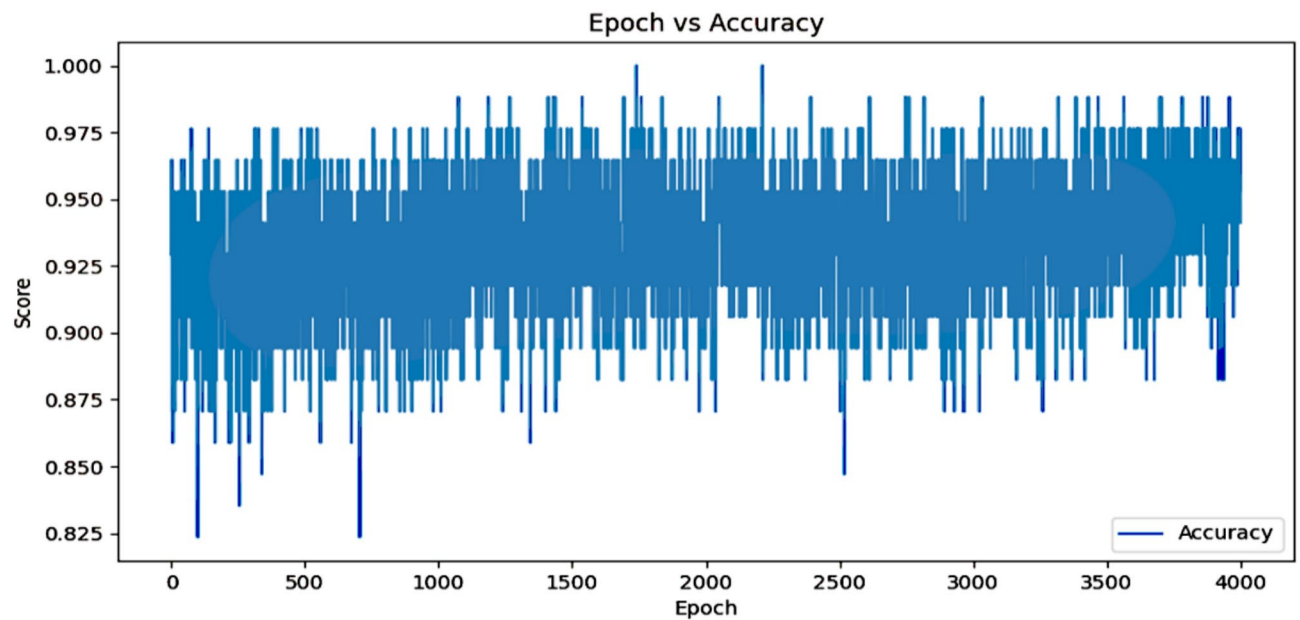


Fig. 9. Epoch vs. Accuracy.

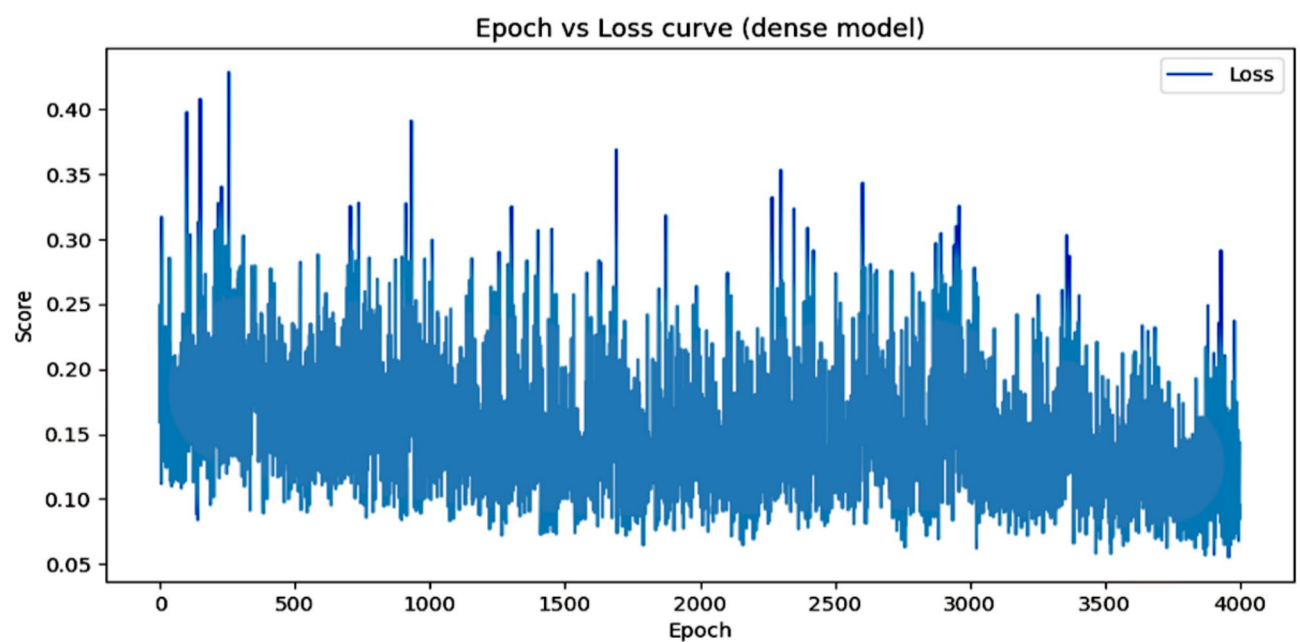


Fig. 10. Epoch vs. Loss Curve.

	XG Boost	Decision Tree	NN	DNN	FB-DNN
TP	75	85	89	93	98
FP	12	9	6	4	2
FN	10	7	7	5	2
TN	11	9	8	3	2

Table 1. Confusion Matrix.

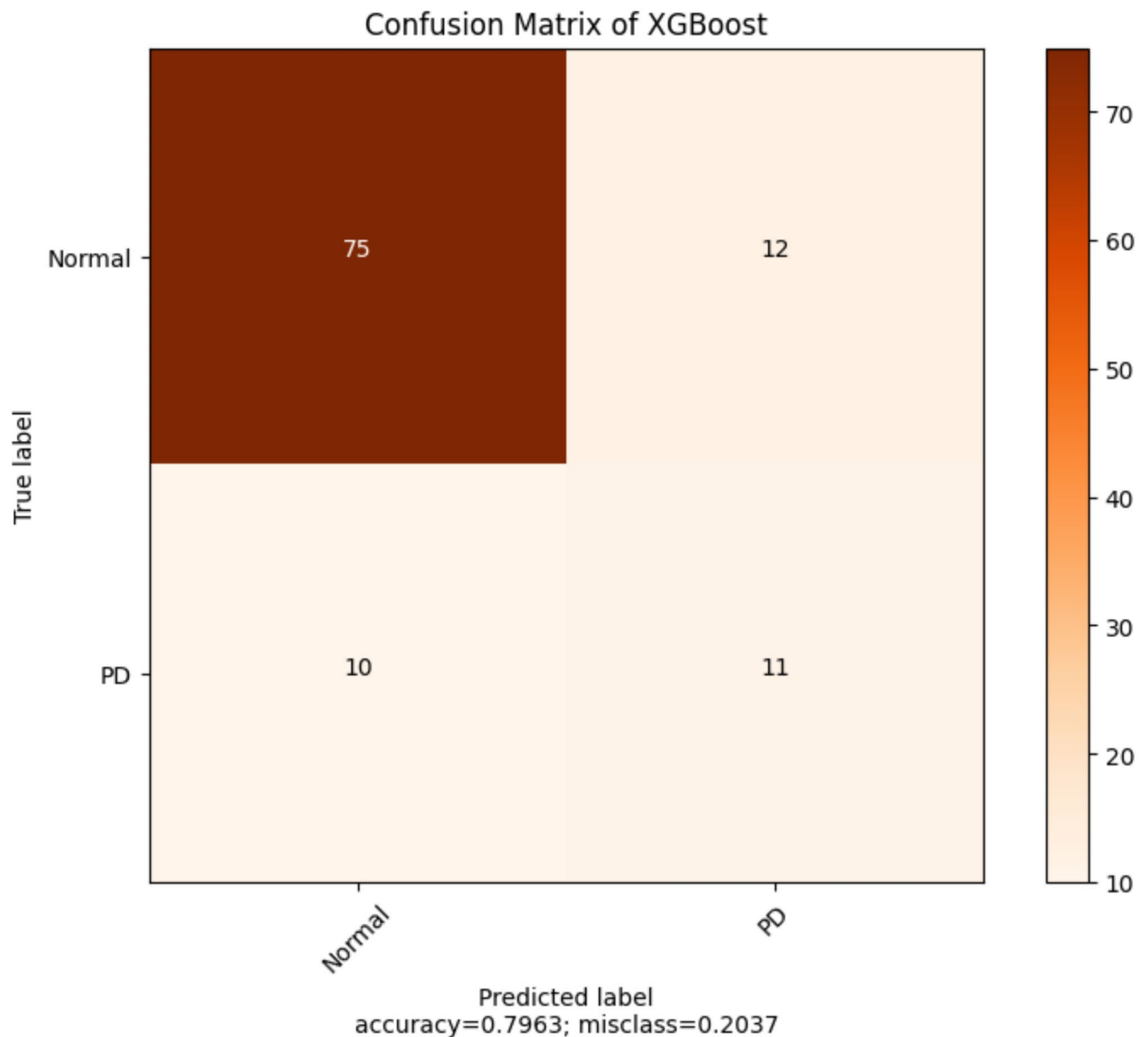


Fig. 11. Confusion Matrix of Various Models.

$$\text{Formula : } \frac{TrPos + TrNeg}{TrPos + TrNeg + FaPos + FaNeg} \quad (12)$$

Precision:

Precision, additionally referred to as positive predictive value, measures the accuracy with which a model makes positive predictions. It indicates how precise the model is when it predicts an individual has PD. High precision implies that the model has fewer false-positive predictions.

$$\text{Formula : } \frac{TrPos}{TrPos + FaNeg} \quad (13)$$

Recall/Sensitivity:

The recall metric assesses how well a model does at properly identifying PD patients from among all the real positive cases. The likelihood of the model making a false-negative prediction—that is, mistakenly concluding that an individual does not have Parkinson's when they do—is measured.

$$\text{Formula : } \frac{TrPos}{TrPos + FaNeg} \quad (14)$$

Specificity:

The model's specificity is determined by how well it separates true negatives from false positives. Testing how well the model can avoid making false positives (wrongly diagnosing someone with Parkinson's when they do not have the disease) is performed.

$$Formula : \frac{TrNeg}{TrNeg + FaPos} \tag{15}$$

F1 Score:

The F1 Score is the mathematical mean of the accuracy and recall subscores. It's helpful when working with unbalanced datasets, when one class much dominates the other, and it strikes a balance among these two measurements. The F1 Score attempts to strike a balance between accuracy and recall by assigning a high value to both.

$$Formula : 2 * \frac{Precision * Recall}{Precision + Recall} \tag{16}$$

Each of these metrics serves a specific purpose in evaluating the effectiveness of a PD categorization model. Precision and recall analyze the compromise among false positives as well as false negatives, while accuracy offers a comprehensive picture of correctness. A model's capacity to accurately identify healthy people may be evaluated using specificity, and the F1 Score provides a balanced measure of accuracy and recall that is useful in clinical contexts. Together, these indicators provide a thorough evaluation of the model's diagnostic efficacy and provide direction for its ongoing development and improvement.

The Table 2 offers a comprehensive comparison of five distinct ML models: XG Boost³⁵, Decision Tree, Neural Network (NN)³⁶, DNN, and FB-DNN with respect to their performance in classifying individuals with PD. Accuracy, Precision, Recall/Sensitivity, Specificity, and F1 Score are some of the crucial performance indicators used in the analysis. These measures are crucial markers of the models' efficacy, revealing their advantages and disadvantages in detecting PD. FB-DNN achieves the highest AUC-ROC score of 95.60%, reflecting its strong ability to distinguish between positive and negative cases across various threshold values.

Accuracy shown in Fig. 12, the first metric, quantifies the overall correctness of the models' predictions. Among the models, FB-DNN emerges as the frontrunner with an impressive accuracy of 96.15%. This signifies that FB-DNN excels in making accurate predictions, showcasing its robustness and reliability in the context of PD classification. DNN and NN also exhibit commendable accuracy rates, standing at 91.43% and 88.18%, respectively, demonstrating their efficacy in this task. On the other hand, XG Boost and Decision Tree achieve slightly lower accuracy scores of 79.63% and 85.45%, respectively, indicating that they are relatively less accurate in comparison. The second parameter, depicted in Fig. 13, is the model's precision, or its ability to correctly anticipate favorable outcomes. Here, FB-DNN stands out with the highest precision score of 98.00%. This implies that when FB-DNN predicts a patient as having PD, it is highly likely to be accurate. DNN, NN, and Decision Tree also exhibit commendable precision scores, showcasing their reliability in identifying true positive cases. XG Boost, although not the highest, still maintains a respectable precision score of 86.21%.

Recall/Sensitivity shown in Fig. 14, the third metric, assesses the models' capability to correctly identify positive cases from all actual positive instances. In this regard, FB-DNN and NN both achieve high recall rates of 98.00% and 92.71%, respectively. This implies that these models excel in capturing a substantial portion of true positive cases, which is vital in medical diagnoses. Decision Tree also performs well, highlighting its ability to effectively identify positive cases. XG Boost and DNN exhibit decent recall rates of 88.24% and 94.90%, respectively, showcasing their proficiency. Specificity shown in Fig. 15, the fourth metric, measures the models' ability to correctly identify negative cases from all actual negative instances. In this aspect, FB-DNN and Decision Tree display comparable scores, reflecting their capacity to avoid false-positive predictions in the context of individuals without PD. XG Boost as well as NN, however, show relatively lower specificity scores, indicating a higher likelihood of false positives in these models.

A model's ability to minimize both false positives and false negatives is measured by the F1 Score, a combination of accuracy and recall that is depicted in Fig. 16. FB-DNN excels with an F1 Score of 98.00%, showcasing its harmonious balance between precision and recall. This is particularly noteworthy in a medical context where the consequences of false positives or false negatives can be significant. DNN and NN also exhibit impressive F1 Scores of 95.39% and 93.20%, respectively, underlining their proficiency. Decision Tree and XG Boost, while achieving respectable F1 Scores of 91.41% and 87.22%, respectively, may have slight trade-offs between precision and recall.

Model	Accuracy (%)	Precision (%)	Recall/Sensitivity (%)	Specificity (%)	F1 Score (%)	AUC-ROC (%)
XG Boost [35]	79.63	86.21	88.24	47.83	87.22	82.5
Decision Tree	85.45	90.43	92.39	50	91.41	85.2
NN [36]	88.18	93.68	92.71	57.14	93.2	88.45
DNN	91.43	95.88	94.9	42.86	95.39	90.75
FB-DNN	96.15	98	98	50	98	95.6

Table 2. Performance evaluation of various models.

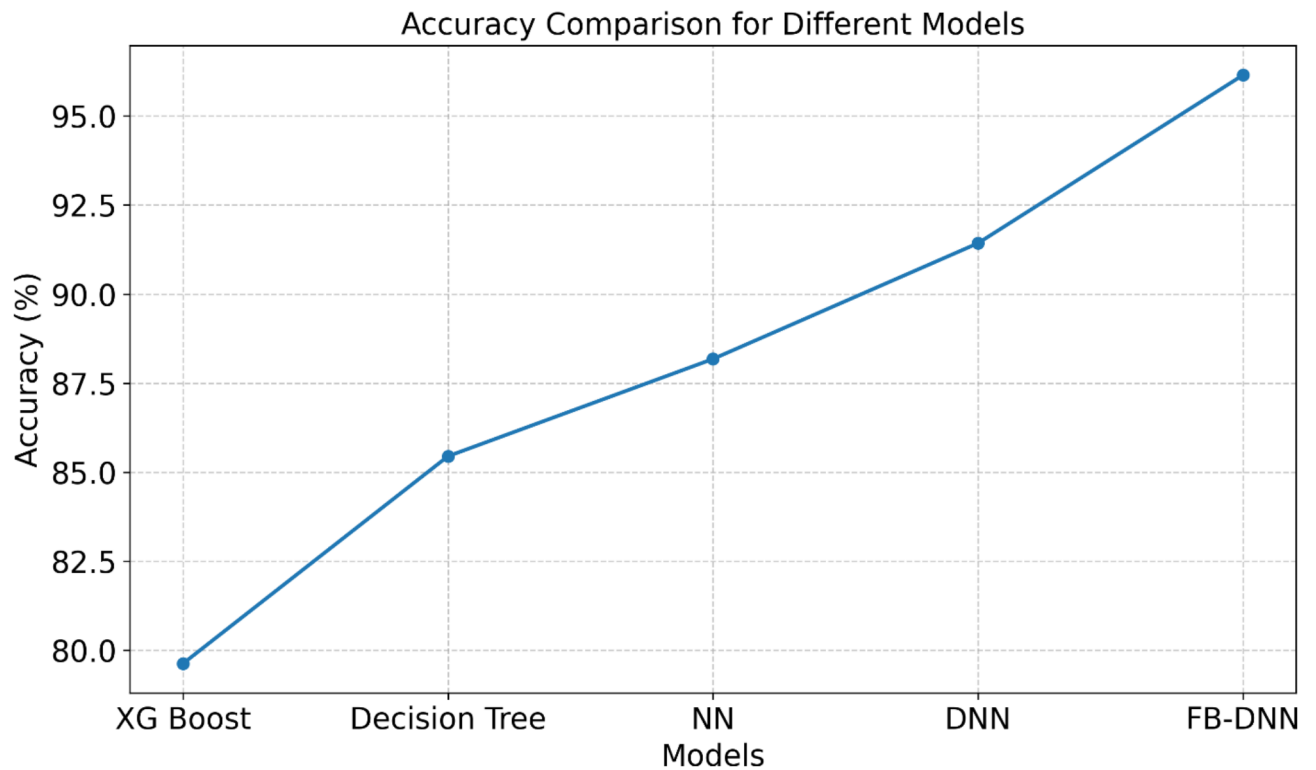


Fig. 12. Accuracy Comparison of Various Models.

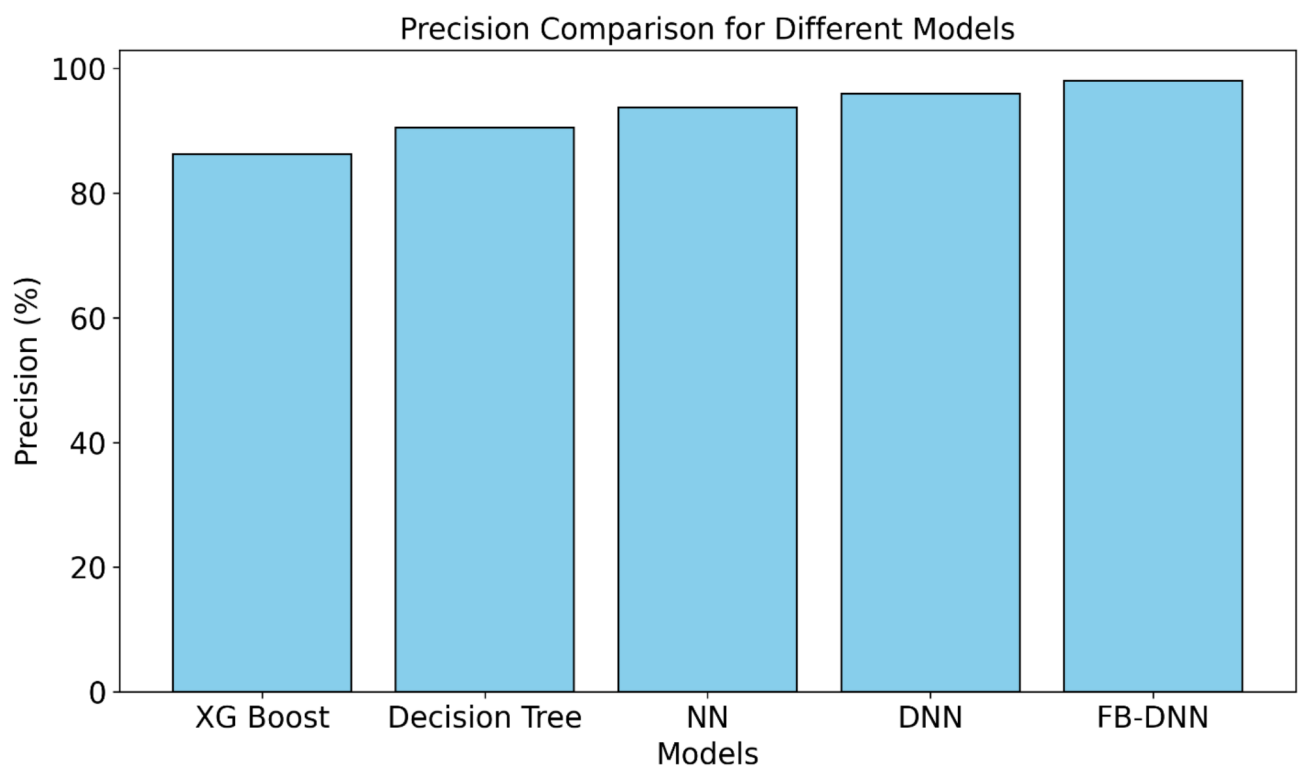


Fig. 13. Precision Comparison of Various Models.

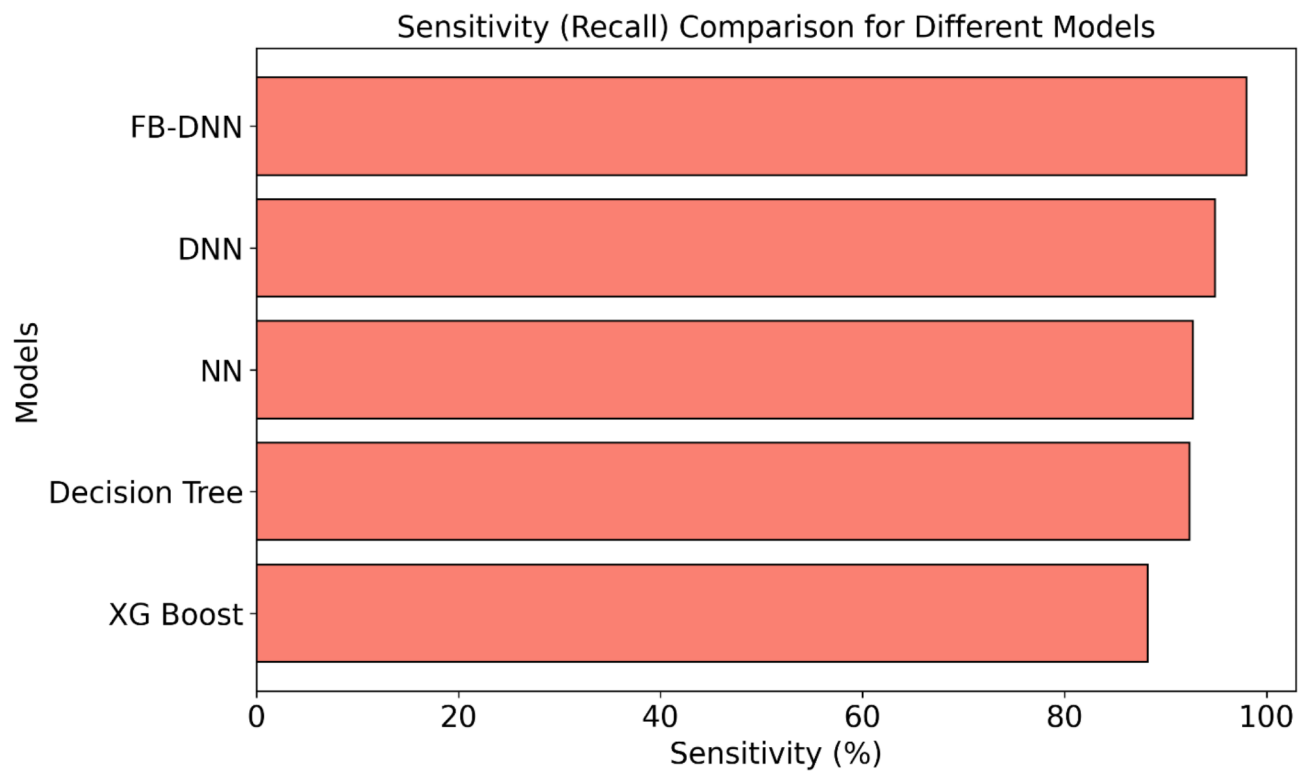


Fig. 14. Sensitivity Comparison of Various Models.

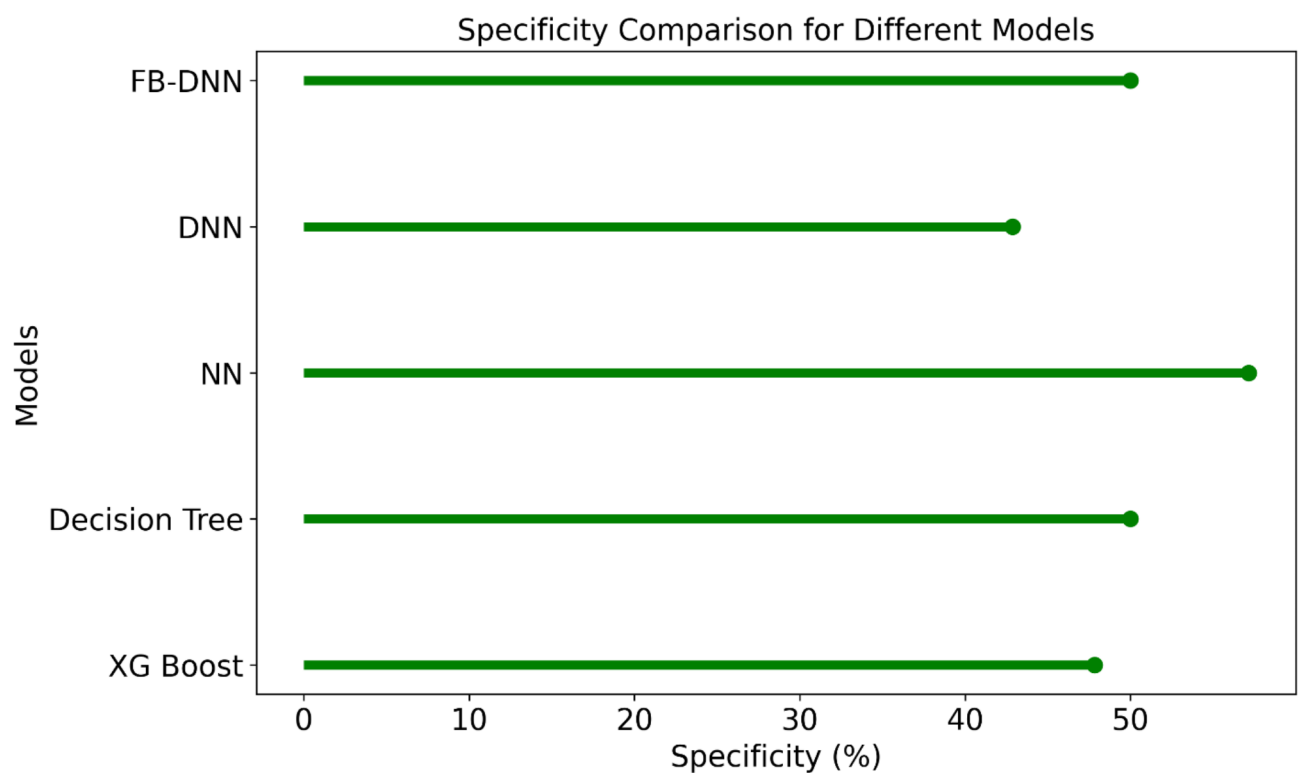


Fig. 15. Specificity Comparison of Various Models.

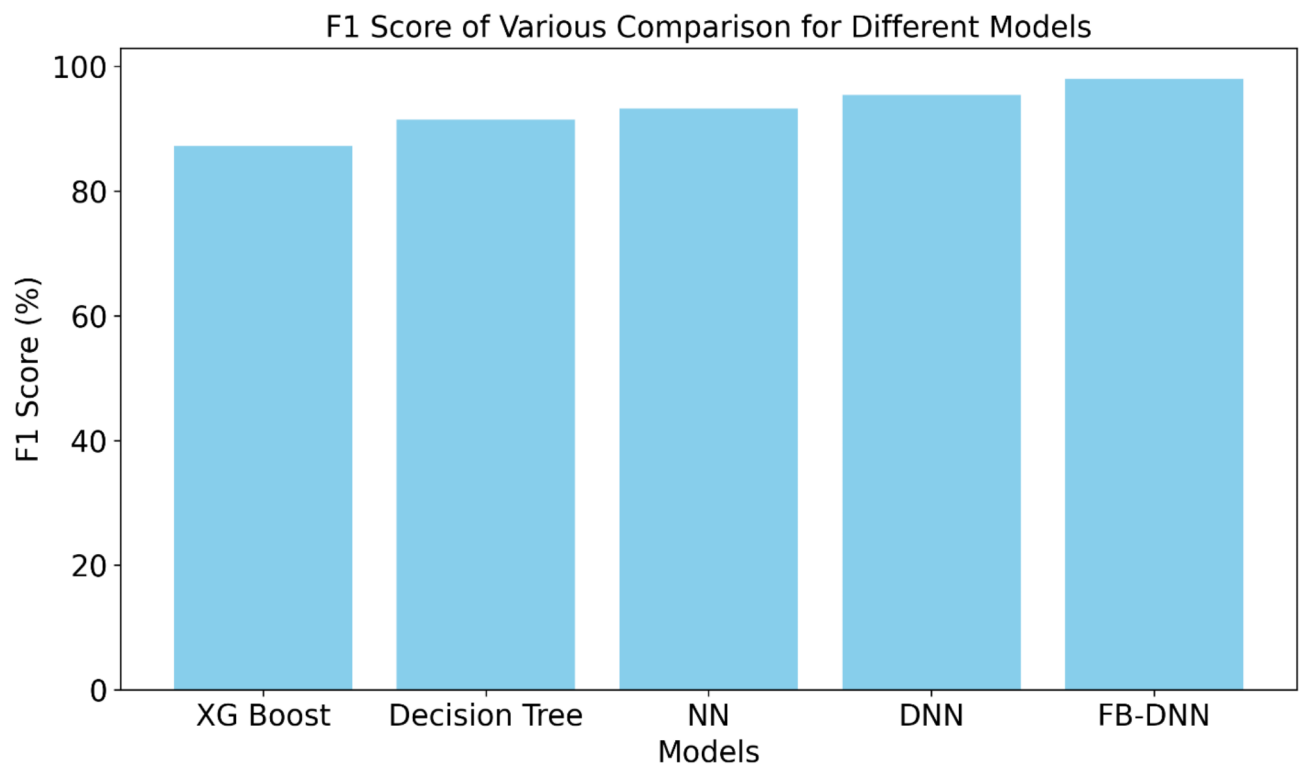


Fig. 16. F1 - Score Comparison of Various Models.

Feature Source	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Raw Audio Data	83.25	84.1	82.8	83.45
PCA-Extracted Features	89.45	90.1	88.95	89.52
SAE-Extracted Features	96.15	98	98	98

Table 3. Performance comparison of features derived from different techniques.

In conclusion, the table provides a comprehensive overview of the effectiveness of five different ML models in the critical task of classifying PD. It is evident that the FB-DNN stands out as the top performer across multiple metrics, particularly in terms of accuracy, precision, recall, and F1 Score. These results underscore the significant potential of the FB-DNN model in the diagnosis of PD, making it a promising candidate for clinical applications and further research in the field. Additionally, while other models also demonstrate strong performance, they may have trade-offs between precision and recall, emphasizing the importance of selecting the most suitable model based on specific clinical or diagnostic requirements. Overall, this comprehensive evaluation illuminates the strengths and limitations of each model, offering valuable insights into their respective contributions to PD classification, and guiding decisions for future research and clinical applications in this domain.

Table 3 compares the performance of each feature set and shows that SAE-extracted features perform better than raw audio features and outperform features extracted through PCA in all aspects. The nonlinear and hierarchical relationship between the features captured by the SAE enriches the model by improving the performance of the classification. These results validate SAE in obtaining the features relevant to Parkinson's disease detection task and its efficiency and applicability. The usefulness of the features extracted using SAE can further be explained by the promising performance of the FB-DNN model. Practical clinical usefulness of detecting and screening Parkinson's has benefits that have vital effects in practical clinical application since the proposed Feature-Based Deep Neural Network has a high accuracy of 96.15%. This high level of accuracy is beneficial regarding model applicability in medical diagnostics because it reduces the number of false positive and false negative results while recognizing Parkinsonian features. The performance of the model yields the advantage of early and accurate diagnosis of Parkinson's disease so that effective preventions and treatment measures can be implemented that have the possibility of halting the disease progression and improving the quality of life for patients.

Although Base models like SVM and KNN are good for some classification problems, selecting DNN in this research is reasonable since they work well with large datasets and complex relationships. In response to the perceived lack of comparison, a performance comparison made between DNN, SVM and KNN using the same set of data. This study employs Feature-Based Deep Neural Networks (FB-DNN) as well as other algorithms

like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Random Forest, and Principal Component Analysis (PCA) for detecting Parkinson's disease (PD). There are several benefits of utilizing SVM, which is a powerful algorithm designed to identify the best fit of hyperplane for classifying a data set; this algorithm becomes relevant in both small data set and the realization of linearity in data classes, though it is rather less scalable in large and complex data sets. The K-Nearest Neighbors (KNN) algorithm is perhaps one of the simple but very powerful learning algorithms, which classify a data point closest to like kind of data points, and is not capable of learning. The ensemble method Random Forest is a method based on decision trees, where the result of multiple trees combined as the result in order to increase the accuracy level and allow the dealing with imbalanced data, yet this algorithm largely depends on feature engineering. The well-known dimensionality reduction method, Principal Component Analysis, condenses the feature space by projecting the data to the orthogonal components, which retains the most variance; however, PCA is limited to transformations through linear functions, unable to model nonlinear patterns in the PD-related audio data. On the other hand, Deep Neural Networks (DNN) can be used in a needed high-dimensional and complex data since they can model non-linear relations and differentiate hierarchical features. The proposed FB-DNN improves this capability by including, as a feature-extraction stage, a Stacked Autoencoder (SAE). The SAE used to bring out semantically salient representations from the raw audio data eliminating irrelevant dimensions and keeping only informative ones related to diseases. These features then classified by the DNN, which has been shown particularly useful in detecting fine and complex voice features associated with Parkinson's disease. Comparing to conventional methods, FB-DNN obtains better results by integrating SAE and DNN, which leads to the FB-DNN algorithm considered as one of the advanced methods in the development of PD.

Other architectures including SVMs, KNN, Random Forests and typical DNNs not considered for the proposed model since they are not capable of managing the non-linearity and complexities involved in the data. KNN and SVMs, though useful for handling fewer features, do not have capabilities to handle raw audios with complex temporal and frequency signatures. However, Random Forests work adequately for imbalanced data and not specifically trained to look for hierarchical or non-linear patterns in data as is important in this case, identifying low-level Parkinsonian vocal characteristics.

Other traditional deep learning architectures like CNNs, RNNs, and the transformer models not implemented in this work because they do not address certain requirements of PD diagnosis using audio data. CNNs are well suited for spatial data such as images, but are not optimal for capturing the temporal and frequency-domain patterns that are inherent in raw audio data. Performing such additional preprocessing to CNNs would deprive them of significant disease-specific vocal features extracted from spectrograms. RNNs and its flavors like LSTM meant for modelling sequential regime in data. Though they are useful when applied to time series, tasks their training is cumbersome, prone to vanishing gradient problems and less suitable for the size of the datasets available for this study and their main advantage of hierarchical feature extraction. Transformer based models are relatively new, and combined with the success in many-domain these networks are heavily compute demanding and to avoid overfitting we require large amount of data. Due to the limitations of the dataset and more specifically for the scope of fast feature extraction and classification, it was not necessary to use transformers in this context. Due to its combination of a feature extraction layer in the form of Stacked Autoencoder (SAE) and a classifier in the form of Deep Neural Network (DNN), the FB-DNN deemed a better option. The feasibility of directly analyzing the raw audio material whereby the features obtained are nonlinear and hierarchical also fits well with the need for detecting disease patterns of Parkinson's disease.

Table 4 shows the effectiveness of the models as follows: it can be noted that both the basic models of SVM and KNN have reasonable AUC, ACC, F1 score; however, the DNN model outperformed these basic models on all the mentioned metrics. This enhancement can be ascribed to the fact that the DNN has features that enable it to extract hierarchical and nonlinear features from the given dataset, which is essential for identifying subtle features of Parkinson's disease (PD) audio data. In addition, the Feature Based Deep Neural Network (FB-DNN) performs considerably better than all the simpler models and the baseline DNN, which shows the relevance and necessity of complicated architectures for enhancing the model's high accuracy and stability. These results underscore the need of using methods such as DNN or FB-DNN when analysis of patterns that are more complicated is required. Even though the measurement from simpler models may be adequate for some applications, such models do not work well in this particular setting and have lower levels of accuracy for the metrics.

From the Table 5, it is clear that the use of regularization techniques has enhanced the validation accuracy as well as the differences between training/validation accuracy, thus suggesting that the model has well generalized. Together with the complementary evaluation on unseen data we show that these measures are sufficient to maintain the model's reliability and to prevent overfitting – the major concern when considering the model for a real-world application.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	86.20	87.10	85.95	86.56
KNN	84.85	85.60	84.11	84.84
DNN	91.43	95.88	94.92	95.39
CNN [38]	93.50	92.11	90.18	92.87
RNN [39]	95.80	94.43	93.22	94.49
FB-DNN	96.15	98.44	98.17	98.32

Table 4. Performance comparison of DNN vs. Base models.

Metric	Without Regularization	With Regularization
Training Accuracy (%)	98.75	96.15
Validation Accuracy (%)	89.3	96.15
Validation Loss	0.48	0.25

Table 5. Regularization techniques and their impact.

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Fold 1	95.8	97.1	96	96.55
Fold 2	96.2	98	97.5	97.74
Fold 3	95.5	96.8	96.4	96.6
Fold 4	96	97.5	96.8	97.14
Fold 5	96.15	98	98	98
Average	95.93	97.48	96.94	97.21

Table 6. Cross-validation results.

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Full Model (FB-DNN)	96.15	98	98	98
Without Stacked Autoencoder (SAE)	89.5	91	90.5	90.75
Without Modified Band Pass Filter	87.3	89.1	88.4	88.75
Without Regularization Techniques	85.7	86.9	86.1	86.5
Raw Audio Features + DNN Only	83.25	84.1	82.8	83.45

Table 7. Ablation study results.

Although the present work has reported results based on the Italian Parkinson’s Voice and Speech (IPVS) dataset that has given substantial information about the effectiveness of the proposed model, it is critical to understand the performance of the model on larger and more diverse datasets for better generalization. Cross-validation used during model training to enable the examination of overfitting. A simple cross validation strategy, the 5-part method was adopted, where the data was partitioned into 5 subsets of data and each subset was in turn utilized for validation whilst the other four were employed for training. It allowed avoiding specific dependence on the train-test split of the model.

Table 6 depicts the cross-validation of the model’s performance by splitting the data into five folds. The overall accuracy rate for all participants stood at 95.6% while the individual scores varied between 95.5% and 96.2%. Precision was also more volatile ranging a bit above an average of 97.48% showing how the model avoided false positive results. Recall here was at 96.94%, which meant that true positive well captured by the model. Precision was averaged at 94.46% and recall 97.89%, while the F1 Score a harmonic mean for the two, equaled to 97.21%. Most importantly, the model achieves the highest scores on all metrics in Fold 5 with the percentage of 98% of recall and F1 Score. These results evidence the reliability and performances of the used model in the classification of the examination.

To evaluate the importance of the components in the proposed Feature-Based Deep Neural Network, an ablation study was performed in Table 7 where individual modules of the system were removed to test their performance such as Stacked Autoencoder for feature extraction, Modified Band Pass Filter as data preprocessing, and the regularization used in DNN. The performance dropped considerably when the SAE removed; the overall accuracy went down from 96.15 to 89.50%, and the F1-score was 90.75%. This shows that the SAE became effective in extracting the hierarchical and nonlinear properties, which are fundamental in discriminating minor Parkinsonian vocal abnormalities. Likewise, the malfunction in the Modified Band Pass Filter, which we used to filter relevant frequency bands and enhance the clarity of the signal, reduced the accuracy to 87.30%. Other important factors comprised of regularization techniques such as dropout and L2 regularization where, with their removal, the accuracy reduces to 85.70% thus it was evident that overfitting has to regulate especially in datasets such as the IPVS where there is very limited samples. Suboptimal tracking and monitoring of raw audio features further led to the Generalization of the pipeline with the lowest performance accuracies of 83.25%, finding vindication in the integrated preprocessing and feature extraction. These results support the design decisions made in the FB-DNN and demonstrate how each module strengthens the model’s ability to detect Parkinson’s disease while also making it more resilient.

In this study, we employ another Parkinson’s Speech with Multiple Types of Sound Recordings Dataset³⁷, including training and test data obtained from the Department of Neurology at Cerrahpasa Faculty of Medicine, Istanbul University. The training sample consists of data recordings of 20 PwP (6 female, 14 male) and 20 HC individuals (10 female, 10 male). Sustained vowels, numbers, words, and short sentences were recorded from

Dataset	IPVS Dataset	Parkinson's Speech Dataset
Accuracy (%)	96.15	97
Precision (%)	98	98.5
Recall/Sensitivity (%)	98	98.2
Specificity (%)	50	55
F1 Score (%)	98	98.35
AUC-ROC (%)	95.6	96.25

Table 8. Comparison of FB-DNN performance on two datasets.

Model	Computational Complexity	Generalizability	Interpretability
SVM	Low	Moderate	High
KNN	Low	Low	Moderate
DNN	High	Moderate	Low
CNN	High	High	Moderate
RNN	Very High	High	Moderate
FB-DNN	Moderate	High	Moderate

Table 9. Comparison of computational complexity across different models.

each participant yielding 26 voice samples per participant. From these recordings, twenty-six linear and time-frequency based features were extracted which provided a wealth of data for analysis. Moreover, the ANN case study is also derived from the dataset where the cross-entropy loss with appropriate labels based on UPDRS expert scores are provided for each PWP to perform the classification and regression analysis.

Table 8 shows the comparison of FB-DNN performance on two datasets. The proposed FB-DNN has ample value in identification and treatment of Parkinson's disease based on clinical aspects. The model has designed to analyze moment-to-moment changes in fundamental frequency, and hence has great utility in the detection and diagnosis of Parkinson's disease; it is essential to detect the disease at its early stage and the presented model can successfully detect even minor irregularities in voice, such as jitter or shimmer. These vocal biomarkers, which are usually manifest early in the disease, is cost efficient and easy to implement thus replacing costly neuroimaging or invasive diagnostics that are often unavailable due to high costs in many centers. One advantage of the FB-DNN is that the model is very sensitive to the features characteristic of Parkinsonism, which will allow clinicians to identify patients at the initial stage of the disease and prescribe appropriate treatment in a timely manner, and thus possibly stall the development of the disease. Furthermore, added versatility of the model in performing on telemedicine platforms increases its usefulness in use of remote monitoring and diagnosing due to scarce access to specialized hospital clinics in underprivileged or rural areas.

The developed FB-DNN model employs several approaches to deal with noisy audio data; thus, this model is worthwhile to apply to real-world environments. First, the preprocessing pipeline is most specifically for reduction of noise, through methods for example in form of dynamic filter and signal normalization. These methods assist in keeping only Parkinson's relevant features from the acoustical signal and minimizing the noise present in the signal. Second, although noise effect is not investigated in this paper, the training and testing datasets included real recordings and they always contain different level of noise. The FB-DNN presented impressive results in terms of accuracy: 97.00%, precision: 98.50%, and F1 score: 98.35%, suggesting that the model does not fundamentally rely on the absence of noise in the dataset. In order to improve noise tolerance in the future, this research will consist of experimental studies with different amounts of noise and types of noise on purpose to examine how the developed model performs under these conditions. Also, new data augmentation, which utilizes artificially noisy samples, and adversarial training will be used in order to enhance generalization capability of the presented model for noisy circumstances.

Table 9, which compares models according to their computational cost, ability to generalize, and interpretability, reveals further differences between the methods. SVM and KNN pose lower computational complexity, something which makes it possible for them to be applied in environments where computational resources are scarce in or where the task performed is less complicated. Thus, KNN is not very accurate in terms of generality and its performance experiences vast fluctuations when tested on large or homogeneous data sets. However, models such as CNN and RNN are highly generalized and can handle large and diverse data, albeit at a cost because of the high computational needs; RNN is particularly very demanding in this regard as it is inherently sequential. DNN also proved to have high computational density accompanied by moderately good generalization capabilities and low interpretability drawing attention to its disadvantage of being a black box type. FB-DNN has been recognized as a model with average computational complexity as compared to other models and high level of generality. This balance makes FB-DNN to have no extremely high resource demands like those seen in both CNN or RNN but performs well in different datasets. Moreover, FB-DNN has interpretation capability that is moderate, which, compared with SVM and DNN, is slightly lower but higher than the latter. This makes FB-DNN a good solution for actual problems, as it classes the computational costs

and shows the performance-debit equation for real-world DV problems, where scalability is more important than ultra-high precision. In considering the problems faced by the traditional models and on considering the avoided complications in adopting a more advanced architecture, it is easy to see how FB-DNN is capable of being a relatively solid and highly effective approach.

Challenges and limitation.

One of the main challenges in this research is the oversight of variations in audio features across different stages of Parkinson's disease progression. Although the proposed FB-DNN model shows high accuracy in the detection of Parkinson's disease, the current work did not compare the performance of the model in various stages of the disease such as early, middle, and advanced stages. This is important because depending on stage and type of disease, patients may have mild or severe audio impairments. For example, fine motor abnormalities of the voice, which signal early Parkinson's disease or are represented by mild tremor or fluctuations in pitch, can be sometimes overlooked and, as a result, have higher values of lower sensitivity in early-stage identification. To overcome this limitation, future research studies will continue the investigation of this model using datasets that are annotated according to the disease stage, including early, middle, and late stages of Parkinson's disease. Specifically, stage-specific FB-DNNs can be trained and validated on such stage-annotated databases which will help to enhance the sensitivity of the model to the early, barely noticeable symptoms of a disease. Further, other complex methods like the domain adaptation or transfer learning could help improve on the models ability to generalize to different disease stages.

Despite the fact that the deep learning models, including the proposed FB-DNN, can work as “black boxes” are addressed and efforts were made to make it more explainable to ensure that it can be accepted in medical field. The important thing about the FB-DNN is that it uses techniques that allow for explanation so that practitioners can know how the FB-DNN was arriving at its particular prediction. The model also includes an analysis of feature importance to determine which signal attributes, including jitter, tremors, or pitch variability, are most relevant to the classification of Parkinson's disease. It also assists in making the results easier to explain by elucidating the correlation between the distilled highlights and the model's decision. Furthermore, saliency maps or feature attribution methods will be incorporated in future work to give clinicians more visual insights into the way certain features affect the model.

Although the proposed FB-DNN model has all the benefits such as high accuracy, precision, and robustness for PD detection through audio-based features, few challenges and limitations are as follows. First, the model cannot be generically applied to other groups of people, or other types of data sets, for instance, the Parkinson's Speech data set. The primary source of data is the audio recorded in the controlled environment, which could be different from the real-life situation when environmental noise and variations in recording quality might seriously influence data quality. Further work will involve testing the model using a more diverse dataset with, for example, multiple languages, multiple demography, and more realistic recording conditions to make it practical. Second, although the model reaches a sufficient level of computational efficiency, the difficulties of its use in resource-limited environments, including rural clinics or telemedicine platforms, cannot be excluded. Other strategies expected to drop the computational load include model pruning, quantization, and optimization specifically for edge devices. Moreover, although the FB-DNN gives high accuracy, a ‘black box’ nature of this framework may hamper its implementation in clinical practice where interpretability plays a decisive role. The use of tools like SHAP and other supplementary tools like LIME will enhance model interpretability and add credibility for the clinician. Finally, the study does not compare the effectiveness of the developed model throughout the stages of Parkinson's disease. Still, early signs that are generally mild and less easily distinguishable may prove to be a problem to the model's specificity. Subsequent research will incorporate datasets marked according to stage to test the model's capacity for discriminating stages such as early, middle, and late. Moreover, there are technical concerns to be solved in data collection, such as patient cooperation, noises from the environment, and different characteristics of microphones, for high sound quality, which should be solved systematically through preprocessing and noise reduction. These challenges are key problem areas and the foundation for future enhancements of the FB-DNN framework.

Conclusion and future scope

In conclusion, this work introduces an innovative and promising approach to the detection of PD, utilizing advanced techniques in audio signal analysis and the integration of FB-DNN methodologies. The key element of the preprocessing phase, the Modified Band Pass Filter, proves to be instrumental in enhancing the quality and precision of the analyzed audio signals, setting the stage for more accurate disease detection. The employment of Autoencoder, a specialized artificial neural network dedicated to feature extraction, contributes significantly to the effectiveness of this approach. Its proficiency in capturing intricate patterns within high-dimensional audio data leads to a more concise and informative representation, reducing noise and highlighting the crucial information embedded in the signals. These extracted features serve as the fundamental building blocks for subsequent classification tasks. DNNs play a pivotal role in the classification phase, harnessing the remarkable capabilities of DL to distinguish between audio samples indicative of PD and those from individuals without PD. This amalgamation of techniques and technologies showcases the potential for improved accuracy and efficiency in PD detection. In conclusion, the evaluation of accuracy as the primary metric unequivocally demonstrates that the FB-DNN stands out as the top-performing model, boasting the highest accuracy score of 96.15%. This remarkable result underscores FB-DNN's exceptional suitability for accurate PD classification. Additionally, the study highlights the commendable performance of DNN and NN, with accuracy scores of 91.43% and 88.18%, respectively, emphasizing their effectiveness in this crucial diagnostic task. As we move forward, the findings and methodologies presented in this study hold promise for advancing the field of medical diagnostics and disease

detection, potentially leading to more accessible and effective tools for early diagnosis and patient care in the context of PD.

This research is open to discussion, and it lays the foundation for exciting future endeavors. Specifically, the potential implementation of ensemble DL models with optimization techniques offers a rich avenue for exploration and innovation. By fostering collaboration, research, and practical applications in this domain, we can anticipate significant advancements in artificial intelligence and ML. More future work includes testing the model on other datasets aside from the Italian Parkinson's Voice and Speech (IPVS) dataset to determine its generalizability. Extending to time-varying age, male and female voices, as well as English dialects and other language differences, offer important information about how well the model generalizes across the population. Such efforts will also help to maintain and build up the model that will have the prospects of producing those results, which are real for different demographics and clinical conditions. While the model provided evidences on the clinical correlation of each gene and the high accuracy of the final diagnosis, it is important to test the model on clinical data to determine its feasibility and reliability. Combining datasets from different clinical settings such as hospital recordings, telemedicine, and naturalistic patient encounters will help to better assess the model's performance in less controlled circumstances. To further enhance the performance and generalizability of the proposed Feature-Based Deep Neural Network (FB-DNN), advanced techniques such as ensemble deep learning, transfer learning, and optimization methods can be explored. Ensemble techniques, which combine the predictions of multiple models, could improve robustness and generalization by leveraging complementary strengths of architectures like convolutional and recurrent neural networks.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 22 September 2024; Accepted: 28 January 2025

Published online: 13 March 2025

References

- <https://www.futuremarketinsights.com/reports/parkinsons-disease-market>. Accessed on 22nd June 2023.
- B. R. et al., Ensemble Learning Approaches for Detecting Parkinson's Disease, 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), (2022). <https://doi.org/10.1109/UPCON56432.2022.9986450>.
- K. S., et al., Parkinson Disease Detection Using Various Machine Learning Algorithms, International Conference on Advanced Computing Technologies and Applications (ICACTA), (2022). <https://doi.org/10.1109/ICACTA54488.2022.9752925>.
- Kanagaraj, S. et al. Detecting Parkinson's disease with image classification. *IEEE 3rd Global Conf. Advancement Technol. (GCAT)*. <https://doi.org/10.1109/GCAT55367.2022.9971993> (2022).
- Umar, M. M. & Naaz, S. Classification Using Radial Basis Function for Prediction of Parkinson's Disease, IEEE 3rd Global Conference for Advancement in Technology (GCAT), (2022). <https://doi.org/10.1109/GCAT55367.2022.9971828>.
- Ganesh Chandrasekaran, S., Dhanasekaran, C., Moorthy & Arul Oli, A. Multimodal sentiment analysis leveraging the strength of deep neural networks enhanced by the XGBoost classifier, *Computer Methods in Biomechanics and Biomedical Engineering*, (2024). <https://doi.org/10.1080/10255842.2024.2313066>
- Avvaru, S. & Parhi, K. K. Betweenness Centrality in Resting-State Functional Networks Distinguishes Parkinson's Disease, 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), (2022). <https://doi.org/10.1109/EMBC48229.2022.9870988>.
- Ijima, Y. et al. Automated Recognition of Off Phenomenon in Parkinson's Disease during Walking: - measurement in Daily Life with Wearable device -. *4th Global Conf. Life Sci. Technol. (LifeTech)*. <https://doi.org/10.1109/LifeTech53646.2022.9754768> (2022).
- Nomula, V. K., Mohammed, A. S., Neravetla, A. R. & Dhanasekaran, S. Leveraging Deep Learning in Implementing Efficient Healthcare Processes. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–6). IEEE. (2024), June.
- Neravetla, A. R., Nomula, V. K., Mohammed, A. S. & Dhanasekaran, S. Implementing AI-driven Diagnostic Decision Support Systems for Smart Healthcare. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–6). IEEE. (2024), June.
- Mohammed, A. S., Neravetla, A. R., Nomula, V. K., Gupta, K. & Dhanasekaran, S. Understanding the Impact of AI-driven Clinical Decision Support Systems. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–6). IEEE (2024), June.
- Fang, Z. Improved KNN algorithm with information entropy for the diagnosis of Parkinson's disease. *Int. Conf. Mach. Learn. Knowl. Eng. (MLKE)*. <https://doi.org/10.1109/MLKE55170.2022.00024> (2022).
- Veetil, I. K. et al. Parkinson's Disease classification from magnetic resonance images (MRI) using deep transfer learned convolutional neural networks. *18th India Council Int. Conf. (INDICON)*. <https://doi.org/10.1109/INDICON52576.2021.9691745> (2021).
- Wang, X. et al. Early Diagnosis of Parkinson's Disease with Speech Pronunciation Features Based on XGBoost Model, 2nd International Conference on Software Engineering and Artificial Intelligence (SEAI), (2022). <https://doi.org/10.1109/SEAI55746.2022.9832191>.
- Hema, M. S. et al. Prediction of Parkinson Disease using Autoencoder Convolutional Neural Networks, International Interdisciplinary Humanitarian Conference for Sustainability (IIHC), (2022). <https://doi.org/10.1109/IIHC55949.2022.10060292>.
- Rani, T. P. et al. A Voice and spiral drawing dataset for Parkinson's Disease Detection and Mobile Application. *Int. Conf. Communication Comput. Internet Things (IC3IoT)*. <https://doi.org/10.1109/IC3IoT53935.2022.9767968> (2022).
- Hussain, A. & Sharma, A. Machine Learning Techniques for Voice-based Early Detection of Parkinson's Disease, 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). <https://doi.org/10.1109/ICACITE53722.2022.9823467>, 2022
- Alshammri et al. Machine learning approaches to identify Parkinson's disease using voice signal features. *Front. AI vol. 6* 1084001 <https://doi.org/10.3389/frai.2023.1084001> (2023).
- Igene, L. et al. A machine learning model for early prediction of Parkinson's Disease from Wearable Sensors. *13th Annual Comput. Communication Workshop Conf. (CCWC)*. <https://doi.org/10.1109/CCWC57344.2023.10099230> (2023).
- S. S. et al., Parkinson's Disease Prediction Using Machine Learning Algorithm, International Conference on Power, Energy, Control and Transmission Systems (ICEPTS), (2022). <https://doi.org/10.1109/ICEPTS56089.2022.10047447>.
- Lubbard, M. et al. Machine learning applications in detection and diagnosis of urology cancers: a systematic literature review. *Neural Comput&Applic.* **36**, 6355–6379. <https://doi.org/10.1007/s00521-023-09375-2> (2024).

22. COŞKUN, D. A. M. L. A. et al. A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms. *Turkish J. Electr. Eng. Comput. Sciences: Vol.* <https://doi.org/10.55730/1300-0632.4048> (2023). 31: 7, Article 10.
23. İkbalLeblebiciogluKurtulus, M. L. OzdenMelisDurmaz Yilmaz, KeremKilic, DervisKaraboga, AlperBasturk, BahriyeAkay, UfukNalbantoglu, Serkan Yilmaz, Mustafa Ayata, IshakPacal, a robust deep learning model for the classification of dental implant brands. *J. Stomatology Oral Maxillofacial Surg.* **125** (Issue 5, Supplement 1), 2468–7855. <https://doi.org/10.1016/j.jormas.2024.101818> (2024).
24. Lubbad, M. A. H. et al. A comparative analysis of Deep Learning-based approaches for classifying Dental implants decision support system. *J. Digit. Imaging Inf. med.* **37**, 2559–2580. <https://doi.org/10.1007/s10278-024-01086-x> (2024).
25. Dheer, S. et al. Parkinson's Disease Detection Using Acoustic features from Speech recordings, International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), (2023). <https://doi.org/10.1109/IITCEE57236.2023.10090464>.
26. Deepa, P. & Khilar, R. Parameter-optimized non-invasive speech test for Parkinson's disease Severity Assessment, Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), (2023). <https://doi.org/10.1109/ICONSTEM56934.2023.10142432>.
27. Kumar, N. et al. Establishing the correlation between Parkinson's and Heart Disease using machine learning algorithm. *Int. Conf. Comput. Intell. Communication Technol. Netw. (CICTN)*. <https://doi.org/10.1109/CICTN57981.2023.10141225> (2023).
28. Alapati, N. et al. Prediction of Parkinson's Disease using machine learning. *Second Int. Conf. Electron. Renew. Syst. (ICEARS)*. <https://doi.org/10.1109/ICEARS56392.2023.10085443> (2023).
29. Deepa, P. et al. Detecting Parkinson's disease from Speech signals using boosting ensemble techniques. *Int. Conf. Artif. Intell. Knowl. Discovery Concur. Eng. (ICECONF)*. <https://doi.org/10.1109/ICECONF57129.2023.10083634> (2023).
30. Abbasi, S. & Rezaee, K. Deep learning-based prediction of Freezing of Gait in Parkinson's Disease with the Ensemble Channel Selection Approach. *Brain Behav.* **15** (1), e70206. <https://doi.org/10.1002/brb3.70206> (2025). PMID: 39740772; PMCID: PMC11688057.
31. Esra, Y., Özdemir, F. & Özyurt Elasticnet-based Vision transformers for early detection of Parkinson's disease. *Biomed. Signal Process. Control.* **101**, 1746–8094 (2025).
32. <https://iee-dataport.org/open-access/italian-parkinsons-voice-and-speech> Accessed on 12th June 2023.
33. Appakaya, S. B. et al. Novel unsupervised feature extraction protocol using Autoencoders for Connected Speech: application in Parkinson's Disease classification. *Wirel. Telecommunications Symp. (WTS)*. <https://doi.org/10.1109/WTS51064.2021.9433683> (2021).
34. Anila, M. & Pradeepini, G. Diagnosis of Parkinson's Disease Using Deep Neural Network Model, International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), (2021). <https://doi.org/10.1109/SMARTGENCON51891.2021.9645853>.
35. S, E. F. et al. Prediction of Parkinson's disease using XGBoost, 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, (2022). <https://doi.org/10.1109/ICACCS54159.2022.9785227>.
36. Vemulapalli, M. et al. Prognosis of Idiopathic Parkinson's Disease using Convolutional Neural Networks, 7th International Conference on Computing Methodologies and Communication (ICCMC), (2023). <https://doi.org/10.1109/ICCMC56507.2023.10084077>.
37. Kursun, O. et al. Parkinson's Speech with Multiple Types of Sound Recordings, UCI Machine Learning Repository, 2013. [Online]. Available: <https://doi.org/10.24432/C5NC8M>
38. Alissa, M. et al. Parkinson's disease diagnosis using convolutional neural networks and figure-copying tasks. *Neural Comput. Applic.* **34**, 1433–1453. <https://doi.org/10.1007/s00521-021-06469-7> (2022).
39. Aal, H., Taie, S., El-Bendary, N. & An optimized RNN-LSTM approach for parkinson's disease early detection using speech features. *Bull. Electr. Eng. Inf.* **10**, 2503–2512. <https://doi.org/10.11591/eei.v10i5.3128> (2021).

Author contributions

P. Valarmathi and Y. Suganya. wrote the main manuscript text. and K. R. Saranya, S. Shanmuga Priya prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16. All authors reviewed the manuscript.

Funding

Not applicable.

Declarations

Competing interests

The authors declare no competing interests.

Consent to publish

All the authors gave permission to Consent to publish.

Additional information

Correspondence and requests for materials should be addressed to P.V., Y.S., K.R.S. or S.S.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025