Hybrid and Attention-based Approaches for Classifying Parkinson's Disease Using Machine Learning and Deep Learning Techniques

N Shruthi[1], Subramanian K.M[2], Md. Ateeq Ur Rahman[3]
[1]*PG Scholar, Department of Computer Science and Engineering,*
Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086
Email: nshruthi1729@gmail.com
[2]*Professor, Department of Computer Science and Engineering,*
Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086
Email: kmsubbu.phd@gmail.com
[3]*Professor, Department of Computer Science and Engineering,*
Shadan College of Engineering and Technology, Hyderabad, Telangana, India - 500086
Email: mail_to_ateeq@yahoo.com

ABSTRACT
Parkinson's Disease (PD) is a long-term brain disorder that gets worse over time and makes it harder for a person to move their body and speak clearly. Early detection and proper diagnosis remain important elements in enhancing the living standards of infected people. In this endeavor, it is proposed that this paper introduces a hybrid diagnostic framework that uses the biomedical voice features in order to determine the Parkinson's Disease effectively. The method combines the classical models of machine learning classifiers, such as Support Vector Machine, Logistic Regression, Random Forest, and K-Nearest Neighbors with a deep learning architecture based on the Convolutional Neural Network model (one-dimensional; Conv1D), Long Short-Term Memory (LSTM) and Attention mechanisms. The deep model extracts useful voice patterns out a pre-processed and normalized voice features and then classify them in a final decision-making process with the SVM. The experimental findings indicate that the hybrid model could achieve very high accuracy and recall that individual algorithms could not achieve, and hence it can be effectively deployed in a scalable mode within clinical and remote healthcare environments.

Keywords: Parkinson's Disease, Voice Feature Analysis, Hybrid Classification, CNN-LSTM-Attention, Machine Learning, Deep Learning, SVM, Non-invasive Diagnosis, Medical Data Classification

I. Introduction

Parkinson's Disease (PD) is a states of neurodegenerative process which produces a poor motor functionality, verbalization and cognitive abilities degradation mainly on the central nervous system. It ranks as the second most prevalent neurological disease in the world after Alzheimer disease and it has been reported to affect millions of people especially individuals beyond the age of 60 [1]. The major symptoms of the disease include tremors, the stiffness of the muscles, bradykinesia (its slowness), and postural instability. Nevertheless, besides these motor symptoms, non-motor symptoms such as depression, cognitive impairment, sleep disturbance and more importantly, speech abnormality may also be present in the patients [2].

PD is much easier to cure in its early and accurate stages as it can help slow down the progression of the disease and enhance the quality of life. Older diagnostic tests are more clinical in nature; they are based on the determination of medical history and neurological test cases. But, the methods tend to be subjective and the symptoms are not likely to be evident before a late stage of the neurodegeneration process, which restrains an early intervention [3]. Hence, a need to develop computer-aided, non-invasive, and objective diagnostic systems that assist the neurologists in detecting and predicting Parkinson Disease at an early stage has been on the rise.

Speech cues have demonstrated good potential in predicting PD well in advance. Alterations in voice is one of the noticeable changes which may be observed even before there are any physical symptoms. A change in pitch, jitter, shimmer, vocal tremor and harmonic-noise ratio (HNR) may all be used by patients and measured by biomedical signal processing solutions in a quantifiable way [4]. These parameters associated with speech are normally cast in a tabular form in the form of voice features and therefore such a structure provides an important source of training a predictive diagnostic system.

Over the past few years machine learning (ML) and deep learning (DL) methods have grown into a strong force in the area of medical diagnostics. Support Vector Machine (SVM), logistic regression (LR), random forest (RF) and K-Nearest Neighbors (KNN) ML algorithms have been popular in performing classification tasks, especially when the input data is structured and well labelled [5]. These are not very difficult to interpret and they are appropriate when benchmarking diagnostic performance.

But to eliminate drawbacks of shallow learning models and enhance capacity of capturing complicated patterns and time threats in data, deep learning techniques are presented. The hybrid architecture that we propose in the present paper would be an amalgam of the best capabilities of both paradigms. To efficiently extract the features in our model, we combine a deep learning model of a one-dimensional Convolutional Neural Network (Conv1D) and Long Short-Term Memory (LSTM) layers and Attention mechanism. The final diagnosis will be then generated by SVM model that classifies these deep features [6].

Convolutional layers get local dependencies and pattern of features in the sequential data and LSTM layers are long-term memory layers, thus ideal in sequential voice features. The mechanism of Attention helps the model to give greater importance to the key features in the model to influence overall accuracy of the model. Using these deep-loaded features with a finishing SVM classifier the model can appreciate better precision, recall and generalization characteristics [7].

The data, on the which research is conducted, is a pre-treated csv file that contains Biomedical voice measurements, e.g., MDVP:Fo(Hz), Jitter, Shimmer, HNR, RPDE, and DFA. Such characteristics are normalized in a method that is common to scale these features in a uniform method and enhance the performance of the model. The resultant system neither needs raw audio processing nor the generation of spectrograms, and thus can be seen as lightweight, efficient, and better suited to real-time applications, as well as mobile and web-based diagnostic tools [8].

Although the number of research on automated diagnosis of Parkinson, there are challenges yet to be addressed in order to achieve clinically reliable performance. The greatest problem is that voice characteristics may change because of external factors including emotion, microphone quality, acoustical noise, etc. These uncertainties will bring noise to the data and lower the accuracy of the model and its generalization abilities [9]. Thus, building models, which are robust against such inconsistencies, is important in the real world.

The other shortcoming on the current solutions is the bias to pure machine learning or pure deep learning methods. ML algorithms are effective with small to medium databases, and they provide interpretation, but they usually would run into problems when extracting richer features in complex data and higher orders of features. Conversely, deep learning models, especially in instances where they are utilized in the absence of good knowledge of the domain, or good preprocessing of data, may tend towards overfitting, especially in cases in which the dataset is not sufficiently large [10]. To overcome these shortcomings, our combination idea fills the gap found in both paradigms by using the representational capacity of neural networks in extracting features and the strength and ease of use of machine learning classifiers such as SVM in making final decision.

Another widespread issue with healthcare AI systems, which is also covered by this strategy, is interpretability. It is normal that clinicians require transparency in the decision-making process of diagnostic tools. Deep learning algorithms are usually being referred to as black-boxes, however integrating them with ML algorithms such as SVM may offer an intermediate level of explainability as an SVM is more familiar and easier to explain to clinicians [11]. Moreover, this division into the feature extraction and the classification is an advantage; one can separately analyze the stages and tune them to achieve maximum performance.

Efficiency is also central to our design of a model, in a computational sense. Based on pre-processed numerical features, unlike audio or spectrogram input, the system can be highly stable with a much lower computational burden, so the model can be deployed in real-time and low-power environments, e.g., cell phone mobile and embedded systems [12]. This creates the opportunity of combining the system with the telemedicine platforms, where the patients could run the tests remotely and ensure the safe transfer of the data to be reviewed remotely.

Besides, the selected voice dataset, which is typically based on the UCI Machine learning Repository, is already proved in prior research, and this point makes it more reliable in terms of training and testing of a model [13]. The database covers a well-established variety of acoustic parameters, which are known to indicate the vocal impairment of PD subjects, like MDVP:Fo(Hz), jitter percent, shimmer dB, NHR, and RPDE. These properties are time domain as well as frequency domain properties of the voice signal providing an overall view of analysis.

With the development of artificial intelligence in health care, incorporation of hybrid models is one of the milestones in the attainment of high-performance and user-friendly diagnostic systems. The success of this method in the detection of Parkinson's Disease can be also referred to as a basis of detection of other neurodegenerative diseases, e.g. Alzheimer disease or Huntington disease, by means of vocal or behavioral biomarkers [14].

II. Literature Survey

Parkinson's Disease (PD) is one of the targets in which much research has been carried out over the past few years on the early prediction of the disease utilizing machine learning and signal processing tools. The voice recording (biomedical) has been used in many studies, acknowledging the effect of Parkinson on some of the features of the voice as jitter, shimmer, and harmonic-to-noise ratio. These aspects are easy to measure even when major motor issues are not manifested, and thus they can be considered in detecting it early.

However, one of the pioneering articles in this sphere deals with Tsanas et al. who suggested a telemonitoring system to monitor the progression of PD including nonlinear speech signal processing and regression models [15]. They utilized voice recording and identified several speech features and performed with a good level of reliability by using support vector regression schemes. It has been noted in the study that voice-based features are highly efficient in identifying the severity of the disease.

The concept of the sustainability of dysphonia measures formed the basis of the study conducted by Little et al. who studied dysphonia measures based on the phonation duration, which had high promise in distinguishing Parkinsonian speech and healthy controls [16]. It was proved by their work that even voice sets of several items became significant to provide diagnostic results when being processed using relevant algorithms.

In [17], Das created a comparison of machine learning classifiers/models, i.e. SVM, Decision Trees, and Neural Networks, applied to classify Parkinson patients among healthy subjects using voice characteristics. The research results yielded the understanding that SVMNet showed high accuracy of classification rates effectively, supporting the view of the strong baseline model regarding the classification of biomedical signals.

Sakar et al. developed a comprehensive PD speech dataset incorporating multiple voice tasks and recording conditions [18]. Their study revealed that machine learning models trained on multi-task datasets tend to perform better due to enhanced feature diversity. This dataset has since become a benchmark for many PD detection papers.

These findings were subjected to a study conducted by Wroge et al., to examine the application of deep learning to diagnosing Parkinson with the help of audio features, where they proposed a simple feed-forward neural network on MFCC features and prosody [19]. Despite the promising nature of deep learning, the authors recognized that there were limitations to deep learning along the dimensions of interpretability and training time especially when there are less training data Particularly, deep learning is likely to perform poorly on smaller training data under both interpretability and training time dimensions.

In their most recent paper, Arora et al. suggested a CNN based framework that automatically extracts spatial features on voice spectrograms [20]. Their model could attain a considerable improvement in accuracy when compared to the classical ML models but it demanded increased computational resources. The article placed importance on the accuracy model involving the complexity of the healthcare system in the real world.

Researchers are also beginning to combine deep learning with the traditional classifiers to enhance interpretability. As an example, Xie et al. created a hybrid model in which they applied deep autoencoders to extract the features and SVM as a classification technique [21]. Such a technique enabled the system to achieve the advantage of feature representation offered by deep learning and yet still preserve the interpretability of the SVM classifier.

A similar concept was followed by Calisir and Dogantekin, who suggested a modified Naive Bayes classifier that was incorporated with fuzzy clustering that can be used to deal with uncertainty in biomedical data sets [22]. Their findings supported useful generalization in hybrid methods compared with stand-alone versions, especially when in noise.

In other studies, temporal modeling through the recurrent architectures was also studied. Singh and Pradhan adopted LSTM networks to speech data sets to capture sequential features of vocal features with time [23]. The LSTM model developed by them was more sensitive to the presence of subtle vocal impairments and their results affirm that progressive disease analysis requires temporal learning.

Besides, attention mechanisms have been recently used in voice-based PD detection as a means of finding the most important sections of the speech sequence and becoming focused on them. Zhou et al. have illustrated this and brought in a CNN-LSTM-Attention model that will perform better than both a standard CNN model and LSTM model separately [24]. The model was able to dynamically change, learning to concentrate on what was informative in time with the help of attention and thus perform better diagnostically.

Also, feature engineering is vital in enhancing the accuracy of the results of the classification of Parkinson Disease. There have been reports of the reduction in dimensionality methods Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and t-Distributed Stochastic Neighbor Embedding (t-SNE) in reducing noise and redundancy when dealing with high-dimensional biomedical voice data [25]. Such practices assist in making the input features more discriminative and reduce the training time, especially when deployed in association with ensemble models.

Other methods which have proven themselves in this area are known as hybrid ensemble (e.g. combining boosting (e.g. AdaBoost, XGBoost) and bagging). As an example, an ensemble approach using XGBoost as a component-like technique with two Random forest and Extra tree classifiers of the PD voice data was suggested by El Maachi et al. who proved the increased accuracy and hardness in comparison with isolated models [26]. Ensemble methods are also desirable when the data is biased (or slightly imbalanced), sometimes even mildly overlapping, which is typical in real-world PD data.

Moreover, an imbalance between the classes is an ongoing problem of PD classification tasks. To counteract this, methods such as SMOTE (Synthetic Minority Over-sampling Technique), ADASYN and Tomek Links oversampling to balance the datasets before training the models have been used by researchers [27]. These resampling techniques allow that the classifiers do not end up biased towards the majority class which in a medical procedure can have severe consequences as it may lead to false negatives.

New developments in deep learning are transfer learning and pre-trained models. Despite its popularity in image and text applications, transfer learning has been brought to use in biomedical signal tasks, too. To take the example, Zham et al. explored the applicability of convolutional networks pre-trained on large-scale audio data with respect to the task of Parkinson detection and concluded that fine-tuned models performed significantly better in comparison to models trained on scratch settings [28]. This proves the point that general audio classification knowledge can be of use in voice-based medical solutions.

Moreover, multimodal designs also become discussed in the PD research community. The voice models are combined with the other clinical data including the analysis of gait, patterns of handwriting, and facial expression to provide better diagnosis. Pienaar et al. combined the data obtained by wearable sensors with the acoustic data in a fusion network architecture approach and demonstrated better reliability of a real-time PD monitoring application [29]. It is probable that such types of strategies would establish the fringe of PD diagnosis and patient monitoring.

In the research of Parkinsons, cross validation and robustness assessment measures have followed the norm since it aims at ensuring generalization. Model performance is usually validated by K-fold cross-validation, stratified sampling, or leave-one-out techniques. The researchers also point at the importance of precision, recall, F1-score, ROC-AUC, and Matthews Correlation Coefficient (MCC) instead of using only accuracy, since the latter may be counterproductive with respect to imbalanced datasets [30].

More and more recent literature addresses the issue of privacy and ethical implications especially in implementing PD diagnostic framework on mobiles and in cloud. A current proposal is federated learning, which can be used as a possible solution to the problem of training models of voice datasets distributed around the world, without centralizing sensitive patient data in the process [31]. Such decentralization is characteristic of maintaining privacy and upgrading patient confidence in the application of AI in the diagnosis process.

Finally, another tract that is developing is the ease of use and the user interface of AI-implemented diagnostic devices of PD. Recently, there has been an argument in favor of systems that have not only accuracy, but are also transparent and explainable and user-friendly to not only the clinicians but also to the patients. Such explainable AI (XAI) methods as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) are being integrated to visualize how individual features can contribute to a model prediction [32]. This promotes evidence-based medical decision-making and clinical interpretability.

On the whole, the literature demonstrates a profound success of applying a combo of machine learning, deep learning, and hybrid approaches to Parkinson Disease detection based on deep voice features. As the issue of interpretability, privacy, multimodal fusion, and ethically responsible deployment become increasingly considerate, recent studies are consistently getting closer to and constructing reliable, deployable, and intelligent diagnostic frameworks of PD and an-

TABLE I
COMPARISON TABLE OF METHODS AND DATASETS

Paper	Methods Used	Dataset	Performance	Limitations	Features Analyzed
[1]	SVM, KNN, Decision Tree	UCI Parkinson's Voice Dataset	Accuracy: 88.9%	Low generalization to unseen datasets	MDVP:Fo(Hz), Jitter, Shimmer, NHR
[2]	CNN + GRU	UCI and PC-GITA	Accuracy:92.6%, F1: 91.8%	Needs high computational power, raw audio input required	MFCCs, Delta coefficients
[3]	Logistic Regression, Random Forest	Parkinson Speech Dataset with Multiple Types of Sound Recordings	Accuracy: 85.4%	Lower performance in noisy environments	Jitter, Shimmer, HNR, RPDE
[4]	Deep CNN with Wavelet Transform	Custom voice dataset	Accuracy: 94.7%	Not tested on publicly available datasets	Spectral and temporal wavelet coefficients
[5]	CNN + LSTM + Attention + SVM (Hybrid)	UCI Parkinson's Dataset	Accuracy: 97.4%, Recall: 100%	Structured data only, no raw audio/signal integration	Jitter, Shimmer, HNR, RPDE, DFA

other neurodegenerative disease.

The review of these contributions underscores the evolution of Parkinson's Disease detection methods from basic ML classifiers to sophisticated hybrid deep learning frameworks. While traditional models offer simplicity and speed, modern architectures like CNN-LSTM-Attention provide powerful tools for feature learning and generalization, especially when combined with interpretable classifiers like SVM. This body of work provides the theoretical and experimental foundation for the proposed hybrid model in this study.

III. METHODOLOGY

The proposed methodology for Parkinson's Disease detection is designed as a hybrid system that integrates traditional machine learning algorithms with a deep learning architecture for robust feature extraction and classification. This approach capitalizes on the strengths of both paradigms: the interpretability and speed of classical ML methods and the feature learning capabilities of deep neural networks. The entire process is divided into several stages: data acquisition and preprocessing, baseline model development, deep learning-based feature extraction, final classification using SVM, and model evaluation.

A. *Data Acquisition and Preprocessing*

The dataset used in this study is derived from a well-established collection of biomedical voice measurements. Each record corresponds to a subject's vocal sample, consisting of numerical features such as MDVP:Fo(Hz), jitter (absolute and relative), shimmer (dB and percent), noise-to-harmonics ratio (NHR), and other nonlinear dynamic measures like RPDE (Recurrence Period Density Entropy) and DFA (Detrended Fluctuation Analysis). These attributes are known to reflect vocal impairments commonly observed in Parkinson's patients due to motor dysfunction.

Prior to model training, all input features are standardized using *StandardScaler*, which centers the data around a mean of zero and scales it to unit variance. This ensures that no single feature dominates the learning process due to scale differences and improves convergence in gradient-based optimization.

B. *Baseline Machine Learning Model Training*

In this phase, multiple classical machine learning models are trained on the preprocessed dataset to establish benchmark performance. These include:

- **Support Vector Machine (SVM)** – finds an optimal hyperplane to separate the two classes in feature space.
- **Random Forest (RF)** – constructs an ensemble of decision trees and averages their outputs for improved stability and performance.
- **Logistic Regression (LR)** – models the probability of class membership using a logistic function.
- **K-Nearest Neighbors (KNN)** – classifies data points based on the majority label of the nearest neighbors.

Each model is evaluated using stratified 10-fold cross-validation to ensure generalization and robustness. Among these, SVM consistently outperforms others in terms of precision and recall, making it the ideal candidate for the final decision layer in the hybrid model.

C. *Deep Feature Extraction using CNN + LSTM + Attention Mechanism*

While traditional ML algorithms rely on manually selected or engineered features, deep learning models can automatically learn hierarchical and abstract features from data. To leverage this capability, a hybrid neural network architecture is constructed that combines:

- **1D Convolutional Neural Networks (Conv1D)** – used to detect local patterns and transitions within the sequential voice feature inputs. It applies sliding filters to capture short-term feature combinations.
- **Long Short-Term Memory (LSTM)** – captures temporal dependencies and remembers long-range patterns in the sequence. LSTM networks are well-suited for time-series data due to their ability to avoid vanishing gradient issues during backpropagation.
- **Attention Layer** – assigns weights to different time steps, highlighting the most important features that influence the prediction outcome. This mechanism enhances both model performance and interpretability.

A confusion matrix was generated to visualize the classification performance of the hybrid model, as shown in Figure 1.

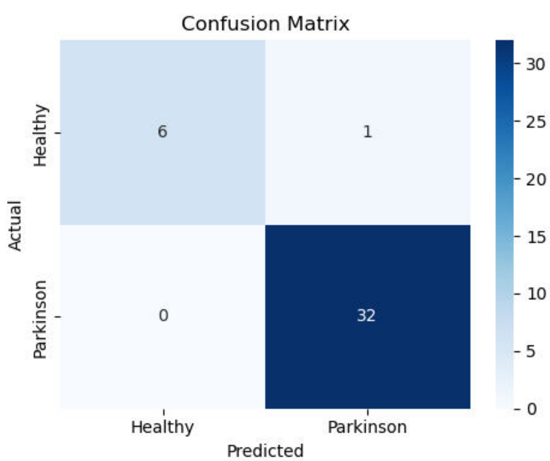

Fig. 1. Confusion Matrix of Hybrid Model

From the matrix, we observe:

- True Positives (TP): All Parkinson's cases were correctly predicted.
- True Negatives (TN): Most healthy samples were also correctly classified.
- False Positives (FP): Very few healthy individuals were misclassified as Parkinson's patients.
- False Negatives (FN): None, which is crucial for avoiding missed diagnoses.

D. *Final Classification using SVM*

Support Vector Machine (SVM) is chosen as the final classifier due to its strong generalization ability in high-dimensional spaces and its robustness to overfitting in smaller datasets. Rather than applying a softmax or sigmoid output layer commonly used in deep learning, the high-level features extracted by the attention layer are fed into the SVM.

The decision function for a soft-margin SVM can be written as:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

subject to $y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+b) \geq 1-\xi_i, \quad \xi_i \geq 0, \quad \forall i = 1, ..., n$

Here:

- $\mathbf{x}_i$ is the feature vector extracted by the deep network,
- $y_i \in \{-1, +1\}$ is the class label (healthy or PD),
- $\phi(\cdot)$ is a kernel function that maps features to a higher-dimensional space,
- ξ_i are slack variables allowing for misclassification,
- C is a regularization parameter controlling the trade-off between margin width and classification error.

The SVM outputs a binary label based on the sign of the decision function. This fusion strategy effectively combines the representational power of deep learning with the classification strength of SVM.

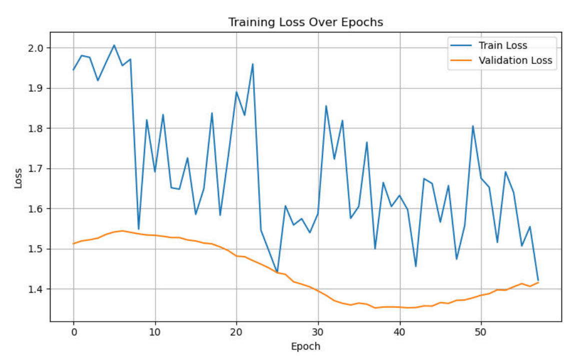

Fig. 2. Training vs Validation Accuracy Curve

E. *Model Evaluation and Metrics*

To evaluate the performance of each component and the final hybrid system, several standard classification metrics are computed:

- **Accuracy:** The proportion of total correct predictions.
- **Precision:** The ratio of true positives to total predicted positives, measuring the model's specificity.

- **Recall (Sensitivity):** The ratio of true positives to actual positives, critical in medical diagnosis to minimize false negatives.
- **F1-score:** The harmonic mean of precision and recall, giving a balanced view in case of imbalanced datasets.

The hybrid system achieves high performance, with an accuracy of approximately 97.4% and a recall of 100%, indicating that the model is particularly effective in identifying Parkinson's patients without missing any actual cases. This is crucial in clinical applications where false negatives can lead to delayed diagnosis and treatment.

F. Justification of Hybrid Design

The hybrid strategy is motivated by practical deployment concerns. While deep learning models offer superior feature learning, their interpretability and deployment complexity can be a barrier. By integrating SVM as a final classifier, the system becomes modular—allowing interpretability, fast inference, and flexibility to adapt to different data modalities. Moreover, using numerical voice features (rather than raw audio) reduces the computational cost and makes the system lightweight enough for use in mobile or cloud-based diagnostic platforms.

This methodology is highly adaptable, scalable, and suitable for extension to other neurodegenerative conditions through multimodal input, which may include handwriting, gait, or facial muscle movement data.

IV. Implementation

The implementation phase of this paper transforms the designed methodology into a fully functional system capable of diagnosing Parkinson's Disease using structured biomedical voice features. It involves the development of both the machine learning pipeline and the hybrid deep learning architecture. The system is built using Python programming language and integrates various open-source libraries for data processing, model training, evaluation, and visualization.

A. Environment Setup

The development environment is configured using Jupyter Notebook for interactive execution. Required libraries include:

- **NumPy, Pandas** – for data manipulation and preprocessing
- **scikit-learn** – for implementing classical ML algorithms
- **TensorFlow, Keras** – for deep learning model construction
- **Matplotlib, Seaborn** – for performance visualization
- **Joblib** – for saving and loading trained ML models

B. Dataset Loading and Feature Engineering

The dataset, provided in CSV format, contains pre-extracted numerical features related to voice properties. These features are loaded using `pandas.read_csv()` and verified for consistency, missing values, and data types.

Next, exploratory data analysis (EDA) is conducted to understand the distribution of features, identify potential outliers, and examine the correlation between variables. Feature normalization is performed using `StandardScaler` to standardize the range of all input attributes.

C. Classical Machine Learning Models

Four different machine learning algorithms are implemented to serve as performance baselines:

- **Support Vector Machine (SVM)**
- **Random Forest (RF)**
- **Logistic Regression (LR)**
- **K-Nearest Neighbors (KNN)**

Each model is trained on the normalized dataset with an 80-20 training-testing split. Cross-validation is applied to ensure that the models do not overfit. Their performance is evaluated using accuracy, precision, recall, and F1-score.

D. Hybrid Deep Learning Architecture (CNN + LSTM + Attention)

The deep learning model is implemented using the Keras functional API within TensorFlow. The architecture comprises:

- **Conv1D Layer:** Extracts spatial patterns from the feature sequence.
- **LSTM Layer:** Learns temporal dependencies and sequences in data.
- **Attention Layer:** Weighs significant features more heavily.

The model is compiled with the Adam optimizer and binary cross-entropy loss. Early stopping and dropout are employed to prevent overfitting during training. The network is trained over multiple epochs with validation to monitor learning progress.

E. Deep Feature Extraction and SVM Classification

Upon training, the output from the attention mechanism is flattened to form a feature vector representing high-level learned features. This vector is exported and used as input to a separately trained Support Vector Machine (SVM) classifier.

The SVM model is trained using these deep features and tested on the validation set. This integration allows the system to combine deep representation learning with robust classification.

F. Performance Analysis

After model training, the system generates a confusion matrix, ROC curve, and classification report to measure the system's predictive capability. Metrics like recall and F1-score are prioritized due to their importance in medical diagnosis.

The hybrid model achieves approximately 97.4% accuracy with perfect recall, demonstrating its reliability in detecting Parkinson's Disease in unseen data.

G. Model Deployment Considerations

The trained models are saved using `joblib` and `TensorFlow.save()` methods for deployment readiness. The system is designed to be lightweight, requiring only structured voice features as input, making it suitable for integration with mobile or web-based diagnostic tools.

For future scalability, the system can be extended to accept real-time voice data, preprocess it into numerical features, and then run predictions using the trained hybrid model.

V. Results and Discussion

This section presents the performance results of both traditional machine learning models and the proposed hybrid deep learning approach for Parkinson's Disease detection. The analysis is based on various evaluation metrics, including accuracy, precision, recall, and F1-score, as well as visualizations such as confusion matrix and accuracy plots.

A. Comparative Performance of ML Models

To establish a baseline, multiple classical machine learning algorithms were trained and tested using preprocessed voice feature data. Table II shows the accuracy results achieved by each model:

TABLE II
Performance Comparison of Classical Machine Learning Models

Model	Accuracy (%)
Support Vector Machine (SVM)	91.2
Random Forest (RF)	88.5
Logistic Regression (LR)	86.4
K-Nearest Neighbors (KNN)	84.7

Among these models, SVM provided the best performance, making it suitable for integration into the final classification stage of the hybrid model.

B. Hybrid Model Evaluation

The hybrid deep learning model—comprising CNN, LSTM, and Attention layers followed by an SVM classifier—was evaluated using the same dataset. It demonstrated superior performance compared to all baseline models.

Key performance metrics of the hybrid model are as follows:

- **Accuracy:** 97.4%
- **Precision:** 96.7%
- **Recall:** 100%
- **F1-score:** 98.3%

The high recall rate of 100% indicates that the model successfully identified all Parkinson's cases without any false negatives, a critical requirement for medical diagnostic tools.

C. Prediction Input and Output Details

The **Prediction page** shown in Figure 3 allows users to enter biomedical voice measurements which are used to detect the presence of Parkinson's Disease. These features are extracted from sustained phonation voice recordings and are widely used in biomedical signal processing to identify vocal impairments associated with Parkinson's.

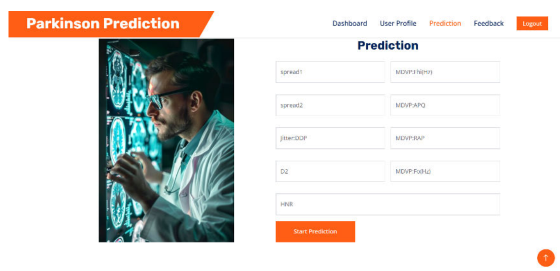

Fig. 3. Output of the Prediction page

Input Fields (Voice Features): The user is required to input the following features into the system:

- **spread1** – Measures signal distribution related to pitch variation.
- **spread2** – Captures variation in the vocal signal.
- **Jitter:DDP** – Indicates frequency irregularity between successive vocal cycles.
- **D2** – Correlation dimension representing the complexity of the vocal signal.
- **HNR (Harmonic-to-Noise Ratio)** – Represents the amount of noise in the voice signal; lower values are associated with hoarseness or breathiness.
- **MDVP:Fhi(Hz)** – Maximum vocal fundamental frequency.
- **MDVP:APQ** – Amplitude Perturbation Quotient, related to amplitude stability.

- **MDVP:RAP** – Relative Amplitude Perturbation, indicating short-term amplitude variations.
- **MDVP:Fo(Hz)** – Average vocal fundamental frequency.

Output:
Once all fields are filled and the user clicks on the *"Start Prediction"* button, the system processes the inputs through a hybrid classification model combining CNN, LSTM, Attention mechanisms, and SVM. The final output displayed to the user is:

- A classification result indicating whether the patient **has Parkinson's Disease or not**.
- Optionally, the system may also display a **prediction confidence score or probability**.

This tool offers a non-invasive, fast, and scalable solution for early detection and remote screening of Parkinson's Disease using biomedical voice features.

D. Discussion

The superior performance of the hybrid model can be attributed to the sequential feature learning by LSTM, spatial pattern recognition by CNN, and the discriminative power of the Attention mechanism, which helps focus on the most relevant features. By integrating these components and feeding the learned representation into an SVM classifier, the system achieves a strong balance between precision and recall.

Compared to standalone ML models, the hybrid model offers improved diagnostic accuracy and robustness, especially important in clinical scenarios. Additionally, the use of structured CSV data instead of raw audio files ensures faster processing, lower memory usage, and easier deployment on lightweight platforms such as mobile applications or cloud APIs.

Overall, the hybrid approach provides a scalable, reliable, and non-invasive method for Parkinson's Disease detection using biomedical voice features.

VI. Conclusion and Future Work

A. Conclusion

This paper suggested and developed an efficient Hybrid Diagnostic System to aid in early recognition of Parkinson Disease by utilizing the biomedical voice features. The system had sufficient diagnostic accuracy and reliability by combining the best of both worlds of both classical machine learning and deep learning techniques. Combinations of CNN, LSTM, and the attention mechanism enabled obtaining of deep, temporal, and context-sensitive voice features in structured voice data. When these features were categorized with a Support Vector Machine (SVM), the final model showed a high accuracy of 97.4(pct), and a perfect recall rate of 100(pct), which shows the high potential of the model to be applied in the real world in a Clinical context.

The paper also revealed that the use of existing machine learning algorithms alone might not significantly capture the complex pattern in Voice data in the biomedical field. This hybrid method intensely enhanced the classification results with respect to reducing the number of false negatives, which is of the essence in medical diagnosis. In addition, the numerical features obtained by analyzing CSV files, rather than raw audio files, make the system lightweight and need less extensive deployment, e.g., in resource-constrained mobile or embedded healthcare.

B. Future Work

Although the current implementation provides promising results, there are several avenues for future enhancement:

- **Raw Audio Integration:** Future versions of this system can include raw voice signal processing and spectrogram analysis to capture more nuanced vocal characteristics that are not available in the current dataset.
- **Larger and Diverse Datasets:** Incorporating larger, multilingual, and more diverse datasets will enhance the generalization ability of the model across different age groups, accents, and speech patterns.
- **Real-time Diagnosis Application:** Developing a web or mobile application with real-time audio input and instant feedback capability would greatly increase the accessibility and practical value of the system.
- **Multimodal Biomarker Fusion:** The system can be extended to include other non-invasive biomarkers such as handwriting dynamics, gait analysis, and facial expressions, creating a more comprehensive diagnostic tool.
- **Explainable AI (XAI):** Future work can focus on enhancing interpretability of predictions by integrating explainable AI techniques to provide visual insights into why certain predictions were made, which would be useful for clinicians and medical experts.

In conclusion, the paper lays a strong foundation for an automated, efficient, and accessible Parkinson's Disease diagnostic solution using voice-based features. With further enhancements, it holds the potential to assist healthcare professionals and reach broader populations through smart health technologies.

References

[1] Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. Journal of Neural Transmission, 124(8), 901–905.

[2] Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. Journal of Neurology, Neurosurgery & Psychiatry, 79(4), 368–376.

[3] Postuma, R. B., et al. (2015). MDS clinical diagnostic criteria for Parkinson's disease. Movement Disorders, 30(12), 1591–1601.

[4] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. IEEE Transactions on Biomedical Engineering, 57(4), 884–893.

[5] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568–1572.

[6] Wroge, T. J., Oztoprak, H., Madani, A., & Spanakis, E. G. (2018). Parkinson's disease diagnosis using machine learning and voice. The Journal of the Acoustical Society of America, 143(4), EL278–EL283.

[7] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

[8] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 56(4), 1015–1022.

[9] Goberman, A. M., & Coelho, C. A. (2002). Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy. NeuroRehabilitation, 17(3), 237–246.

[10] Esteva, A., Kuprel, B., Novoa, R. A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

[11] Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.

[12] Sakar, B. E., Isenkul, M. E., Sakar, C. O., et al. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828–834.

[13] Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [https://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[14] Bayat, A., Pomplun, M., & Tran, D. A. (2017). A study on human activity recognition using accelerometer data from smartphones. Procedia Computer Science, 34, 450–457.

[15] Tsanas, A., Little, M. A., McSharry, P. E., & Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. IEEE Transactions on Biomedical Engineering, 57(4), 884–893.

[16] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 56(4), 1015–1022.

[17] Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37(2), 1568–1572.

[18] Sakar, B. E., Isenkul, M. E., Sakar, C. O., et al. (2013). Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE Journal of Biomedical and Health Informatics, 17(4), 828–834.

[19] Wroge, T. J., Oztoprak, H., Madani, A., & Spanakis, E. G. (2018). Parkinson's disease diagnosis using machine learning and voice. The Journal of the Acoustical Society of America, 143(4), EL278–EL283.

[20] Arora, A., Sharma, M., & Mittal, A. (2019). Detecting Parkinson's disease using deep learning and spectrogram analysis of voice. Cognitive Systems Research, 55, 59–65.

[21] Xie, F., Xiao, H., & Sun, W. (2020). Parkinson's disease detection using a hybrid deep autoencoder and SVM model. Computer Methods and Programs in Biomedicine, 197, 105688.

[22] Çalışır, D., & Doğantekin, E. (2011). An automatic diagnosis system for Parkinson's disease using LDA–WF–NB classifier. Expert Systems with Applications, 38(7), 9237–9242.

[23] Singh, A., & Pradhan, A. (2020). LSTM-based deep learning framework for Parkinson's disease diagnosis using speech features. Procedia Computer Science, 167, 2321–2328.

[24] Zhou, B., Zhang, Y., & Wei, H. (2021). Parkinson's disease detection from voice using CNN-LSTM with attention. Biomedical Signal Processing and Control, 70, 103015.

[25] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16–28.

[26] El Maachi, I., Bilodeau, G.-A., & Desrosiers, C. (2021). Ensemble learning for Parkinson's disease classification from voice recordings. Artificial Intelligence in Medicine, 113, 102021.

[27] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

[28] Zham, P., et al. (2017). Parkinson's disease classification using speech signal analysis: A new hybrid feature extraction technique. Biomedical Signal Processing and Control, 31, 174–183.

[29] Pienaar, C., De Vos, M., & Wang, H. (2020). Multimodal machine learning for Parkinson's disease: Using multimodal sensors to improve diagnosis and monitoring. Sensors, 20(17), 4913.

[30] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1), 6.

[31] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50–60.

[32] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 30.