Parkinson's Disease Detection from Speech Signals using Explainable Artificial Intelligence

Ghanta Sai Krishna*, Kundrapu Supriya †, Vinay Mishra ‡, Santosh Kumar §
*Department of Computer Science, IIIT Naya Raipur, India
e-mail: ghanta20102@iiitnr.edu.in, kundrapu20100@iiitnr.edu.in, vinay21300@iiitnr.edu.in, santosh@iiitnr.edu.in

***Abstract*—Parkinson's disease (PD) is a neurological condition that is on the rise and disrupts the nervous system. However, there is no specific diagnosis for Parkinson's disease; only a variety of motor signs can be used to identify it. A speech impairment was found in more than 90% of PD patients. This study presents a voice and speech signal data-based model for PD identification. The PD is the speech data set used in this experiment has a great amount of dimension with very few data points. Different data pretreatment techniques, such as data standardization, multicollinearity diagnostic, and dimensionality reduction approach, were used in our suggested model to enhance the quality of the data. Different Machine Learning (ML) classifiers were employed to categorize PD, including k-nearest Neighbor, Support Vector Machine, Random Forest, AdaBoost, and Logistic Regression. In this experiment, grid search, cross-fold validation, and hyperparameter tweaking were used to optimize classifier performance and maintain the class distribution of the unbalanced data set. Our suggested model outperformed the prior tests on the same data set by around 98.00% and reached a maximum accuracy of 98.10**

***Index Terms*—Parkinson Disease, Machine Learning, Explainable AI**

I. Introduction

Millions of people worldwide suffer from progressive neurodegenerative Parkinson's disease (PD) as shown in Fig. **??**. Dopamine-producing brain neurons degenerate, causing motor and non-motor symptoms. Early identification of Parkinson's disease is critical for optimal treatment and care, although symptoms vary and there are no conclusive diagnostics.Early diagnosis of Parkinson's disease is essential for efficient management and therapy, but it may be difficult since symptoms can vary and there are no reliable diagnostic tests [1].

Speech recordings have garnered increasing attention in recent years as a non-invasive, economical way to identify Parkinson's disease early on. Speech is a multifaceted signal that reflects the health of many brain systems, including those responsible for motor control, cognition, and emotion. As a possible biomarker for early identification, changes in speech patterns and features have been connected to the early stages of Parkinson's disease. As effective methods for analysing voice recordings and extracting pertinent aspects that may be utilised for Parkinson's disease diagnosis and monitoring, signal processing and machine learning approaches have evolved. Machine learning algorithms [2] may be used to categorise speech recordings into distinct groups based on these qualities. Signal processing methods employ a variety of digital signal processing techniques to extract speech properties including pitch, intensity, and formant frequencies.

II. Related Works

Speech recordings have drawn more attention recently as a non-invasive, economical method of identifying Parkinson's disease in its early stages. The health of numerous brain systems involved in motor control, cognition, and emotion is reflected in speech, a complicated signal. Parkinson's disease's early phases have been associated with changes in speech patterns and features, making it a potential biomarker for early identification. In order to analyse speech recordings and extract pertinent elements for Parkinson's disease diagnosis and monitoring, machine learning and signal processing methods have developed into useful tools. Based on these variables, machine learning algorithms may categorise voice recordings, whilst signal processing techniques employ various digital approaches to extract speech qualities including pitch, intensity, and formant frequencies. The categorization and severity evaluation approach for PD was developed by the authors of [3]. For the categorization, they combine Support Vector Machines with neural networks. The findings reveal a detection accuracy of 97.64%. The main obstacle to determining accurate classification accuracy is the amount of the database employed in the trials. Less than 60 recordings of voices are often used by the writers, with varying degrees of effectiveness.

Srishti Grover [4] classified the data into the two categories of "severe" and "not severe" using deep learning on the Parkinson's Telemonitoring Voice Dataset from the UCI ML Repository. The neural network has two neurons in the output layer, three hidden layers with 10, 20, and 10 units each, and an input layer with 16 units. The accuracy of the model was 81.6%. Another research [5] used PCA and OFS-based feature sets in an effort to categorise the PD group. Using RF and PCA, nonlinear classifiers including Bagging classification, Regression tree (Bagging CART), Random Forest, and RPART were utilised to classify data. Based on the compromised writing skills, Clayton R. Pereira [6]suggested an alternative strategy. They proposed employing two alternative CNN architectures—ImageNet and LeNet—to learn pen-based characteristics from signals derived from the smart pen's six sensors. With ImageNet for meanders and OPF for spirals, the authors reported the highest accuracy of 83.77%. Additionally, a major feature of PD is a reduction in the quantities of

dopamine generated by brain cells called neuron. It can be discovered using FP-CIT SPECT and dopamine transporter imaging, and the authors of [7] created an automated deep-learning algorithm to decipher a collection of FP-CIT SPECT images taken from the PPMI repository. SPECT pictures are used as inputs for a 3D convolutional layer, which generates 16 3D outputs after going through 777 convolutional filters, max-pooling, and ReLU activation layers in addition to the output layer.

Resul Das [8] conducted a machine learning study comparing four different types of classification algorithms to enable the diagnosis of Parkinson's disease. The study used SAS-based software to model various classifiers that can detect the presence of PD, such as DMNeural, Neural Network, Regression, and Decision Tree. The efficiency of the classifiers was evaluated using various methods, and the accuracy assessment revealed that the Neural Network classifier had a 92.9% correct classification rate. Hariharan et al. [9] distinguish their paper from others by presenting a novel method for detecting and diagnosing Parkinson's disease (PD) using a hybrid intelligent system3. Their proposed system includes model-based clustering pre-processing with a Gaussian Mixture Model, followed by feature reduction with Principal Component Analysis, Linear Discriminant Analysis, Sequential Forward Selection, and Sequential Backward Selection. The authors used three supervised classifiers for classification: least square support vector machine, Probabilistic Neural Network, and General Regression Neural Network. The data used in their study came from the UCI ML5 database. Astrom and Koker [10] found that a parallel feed-forward neural network structure4 was more effective than a single neural network in predicting Parkinson's disease. They used a set of nine parallel neural networks, which resulted in an 8.4% improvement. A rule-based system was used to evaluate the output of each individual neural network to arrive at the final output. Additionally, each network's unlearned data was collected and fed into the training set of the next network in the series.techniques were surpassed by their suggested manner.

III. Methodology

The progression of the classification process as a whole is the topic of discussion in this section. FS was carried out by the 23 voice features of approximately 195 different examples, all of which are multivariate characteristics of Dataset. The conceptual architecture of proposed methodology is illustrated down below in the form of Fig. 1. Data collection, feature selection, the training of a model, and model prediction are the subsequent steps in the suggested method. This research investigates a Machine Learning (ML)-based diagnosis of Parkinson's disease utilising a variety of classifiers, and the accuracy of their predictions is diagnosed based on the established characteristics.

A. Exploratory Data Analysis

During our exploratory data analysis, we observed the presence of outliers in many features. However, these outliers

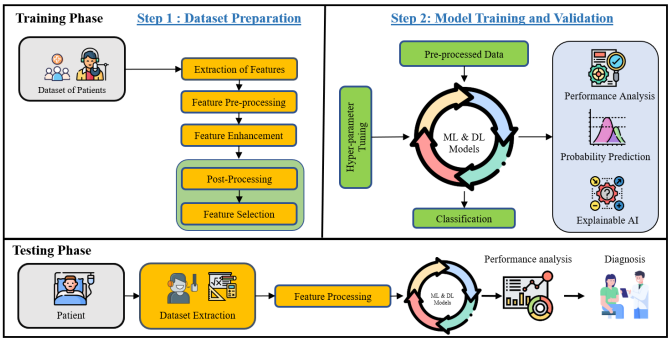

Fig. 1. Conceptual Diagram of Proposed End-to-End Framework

fall within the possible biological ranges for the respective variables, indicating that they could be genuine data points rather than errors. Based on this observation, we decided to let the outliers remain in the dataset and not treat them as anomalies.

- **Correlation Analysis :** We conducted a correlation analysis to identify any patterns or relationships among the variables in the dataset. Several interesting observations were made during this analysis
- **Clusters of Strong Correlations :** We found clusters of strong correlations between specific groups of variables. For example, measurements of Jitter are strongly correlated with each other, suggesting that they may share some underlying commonality. Additionally, there is a decent correlation observed between Shimmer and Jitter variables, which might imply a relationship between these two groups of features.
- **High Negative Correlation :** We identified a strong negative correlation between the Harmonics-to-Noise Ratio (HNR) and the rest of the variables in the dataset. This finding indicates that as HNR increases, the other variables tend to decrease, and vice versa.
- **Novel Measurements :** The novel measurements such as Recurrence Period Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Entropy (PPE) are not strongly correlated with other variables in the dataset. This observation can be attributed to the fact that these measurements are nonlinear and therefore may not exhibit a linear relationship with the other variables.

In conclusion, our exploratory data analysis revealed some intriguing patterns and relationships among the variables in the dataset. These insights will be valuable when we proceed to the next stages of our analysis, such as feature selection and model building.

B. Principle Component Analysis

A statistical method called principal component analysis (PCA) is used to make datasets smaller while preserving as much variance as feasible. To do this, the original data is transformed into a new coordinate system, where the first axis (principal component) denotes the direction of highest variation and each succeeding axis denotes the direction of

maximum variation orthogonal to the preceding axis. In data analysis and machine learning, PCA is often used to find patterns in data, minimise the number of variables in a dataset, and enhance the accuracy of prediction models. When dealing with high-dimensional datasets, where it is difficult to visualise the data or find significant associations between variables, it is very helpful.

The following are the steps for doing PCA:

- Data standardisation is important because PCA works best with variables that have comparable scales. As a result, it's crucial to standardise the data by dividing by the normal deviation and removing the mean.
- Calculate the covariance matrix to reveal the relationships between the variables in the dataset. It is calculated by adding the transpose to the conventional data matrix.
- Determine the eigenvectors and eigenvalues: Eigenvectors are the directions in which the data vary most, and eigenvalues are the degree to which each eigenvector varies the data. They are computed by solving the eigenvalue problem, which entails locating the covariance matrix's eigenvectors and eigenvalues.
- Choose the principle components: The eigenvectors with the highest eigenvalues are the principal components. They stand for the directions in which the data vary most widely.
- Transform the data: To convert the original data into the new coordinate system, multiply it by the eigenvector matrix.

C. Predictive Modelling

1) Random Forest: A popular machine learning approach for classification and regression applications is called Random Forest. Multiple decision trees are combined using an ensemble learning approach to create a model that is more accurate and stable. The step by step process of random forest is described below:

2) Decision Tree: A graphical depiction of a decision-making process that has a tree-like form is called a decision tree. Decision-making processes are modelled and visualised using it in a variety of industries, including business, medical, engineering, and finance. Both classification and regression issues may be solved using decision trees. The step by step process of decision tree is described below:

3) SVM: A supervised machine learning approach called Support Vector Machine (SVM) is used for classification and regression analysis. It operates by identifying the hyperplane in a high-dimensional space that best classifies the data.

4) KNN: A simple but effective machine learning approach called K-Nearest Neighbors (KNN) is utilised for classification and regression applications. The instance-based or slow learning algorithms group includes KNN. In a KNN, the input consists of the k training instances that are closest to each other in the feature space, and the output, for classification or regression problems, is a class membership.

5) CNN + LSTM: The LSTM + CNN model is a deep learning architecture that combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks. This architecture is commonly used in tasks such as image captioning, where both visual and textual information need to be processed. Working of LSTM + CNN model:

- Data preprocessing: The CSV data is first preprocessed into a format suitable for deep learning. This typically involves scaling the data and splitting it into training and testing sets.
- CNN layer: The input data is then passed through a 1D CNN layer to extract features:

$$conv_{i,j} = \sigma(b_i + \sum_{k=1}^{n} w_{k,i} x_{j+k-1}) \quad (1)$$

  where i indexes the filter, j indexes the time step, n is the filter size, x_{j+k-1} is the input data at time step j+k-1, $w_{k,i}$ is the weight of the filter, b_i is the bias term of the ith feature map, and σ is the activation function.
- LSTM layer: The output of the CNN layer is then fed into an LSTM layer, which processes the temporal sequence of features:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o) \quad (5)$$

$$h_t = o_t \cdot tanh(c_t) \quad (6)$$

  where i_t, f_t, c_t, and o_t are the input, forget, cell, and output gates, respectively, x_t is the input feature vector at time t, h_{t-1} is the previous hidden state, c_{t-1} is the previous cell state, W and b are the weight and bias matrices, and tanh and σ are the hyperbolic tangent and sigmoid activation functions, respectively.
- Dense layer: The output of the LSTM layer is passed through a time-distributed dense layer, which applies a fully connected layer to each time step:

$$y_t = W_y h_t + b_y \quad (7)$$

  where y_t is the output at time t, W_y and b_y are the weight and bias matrices of the dense layer, respectively.
- Softmax layer: The output of the time-distributed dense layer is passed through a softmax layer to generate the final predictions:

$$P_t = softmax(y_t) \quad (8)$$

  where P_t is the predicted probability distribution over the classes at time t.

6) GRU: The Gated Recurrent Unit (GRU) is a type of recurrent neural network (RNN) that is widely used for sequential data processing, such as speech recognition, language translation, and image captioning. The GRU was introduced by Cho et al. in 2014 as a simpler alternative to the long short-term memory (LSTM) architecture. The GRU has fewer parameters than the LSTM and is easier to train, while still achieving state-of-the-art performance on many tasks. The GRU architecture consists of a hidden state vector h and two gating mechanisms: the reset gate r and the update gate z. The reset gate determines how much of the previous hidden state to forget, while the update gate determines how much of the new information to incorporate into the new hidden state.

The equations for the GRU are as follows:

$$Resetgate : r_t = sigmoid(W_r[x_t, h_{t-1}] + b_r) \quad (9)$$

$$Updategate : z_t = sigmoid(W_z[x_t, h_{t-1}] + b_z) \quad (10)$$

$$Candidateactivation : h'_t = tanh(W[x_t, r_t * h_{t-1}] + b) \quad (11)$$

$$Hiddenstate : h_t = (1 - z_t) * h_{t-1} + z_t * h'_t \quad (12)$$

where, x_t is the input vector at time t, h_{t-1} is the previous hidden state, r_t is the reset gate vector, z_t is the update gate vector, $h_{t'}$ is the candidate activation vector, W_r, W_z, and W are weight matrices for the reset gate, update gate, and candidate activation, respectively b_r, b_z, and b are bias vectors for the reset gate, update gate, and candidate activation, respectively.

The reset gate r_t determines how much of the previous hidden state h_{t-1} to forget. If r_t is close to 1, then the previous hidden state is fully retained, while if r_t is close to 0, then the previous hidden state is mostly ignored.

Finally, the new hidden state h_t is computed as a weighted average of the previous hidden state h_{t-1} and the candidate activation $h_{t'}$, where the weights are determined by the update gate z_t. If z_t is close to 1, then most of the new hidden state is based on the candidate activation, while if z_t is close to 0, then most of the new hidden state is based on the previous hidden state.

D. Evaluation Metrics

1) Accuracy: Accuracy is a term commonly used in statistics and machine learning to measure the degree of correctness of a prediction or classification model. In classification tasks, accuracy is defined as the ratio of correctly predicted instances to the total number of instances in the dataset. It is often expressed as a percentage, with higher values indicating a more accurate model [11].

2) Precision: Precision is a performance metric commonly used in machine learning and statistics to evaluate the quality of a classification model. It is defined as the ratio of true positive instances to the total number of instances predicted as positive. In other words, precision measures how many of the positive predictions made by the model are actually correct.

$$Precision = \frac{T_P}{T_P + F_P} \quad (13)$$

3) Recall: Another evaluation metric that can be used for evaluating the performance of the model is recall. It is defined as the ratio of true positive instances to the total number of instances that are actually positive. In other words, recall measures how many of the positive instances in the dataset the model correctly identified.

$$Recall = \frac{T_P}{T_P + F_N} \quad (14)$$

4) F1-Score: F1-Score is the harmonic mean of precision and recall, which provides a balanced measure of both metrics.

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

E. Explainable AI

XAI approaches like SHAP and LIME explain machine learning model predictions. Explaining feature relevance in a forecast is popular with SHAP. SHAP values each characteristic in a forecast based on its effect on the outcome. This helps users understand which elements are most relevant for prediction and how they interact. LIME explains forecasts. It develops a local model around a forecast and illustrates which characteristics most affected its result. This helps users understand how the model made a prediction. SHAP and LIME work with any machine learning model, independent of its training method. This makes them handy tools for many uses.

IV. Experimental Results

A. Dataset Description

This dataset contains biomedical voice measurements of 31 individuals, including 23 individuals diagnosed with Parkinson's disease (PD) and 8 healthy individuals. The dataset aims to identify and discriminate healthy individuals from those with PD based on voice-related characteristics derived from voice recordings. Each row of the dataset represents one voice recording, and each column represents a particular voice measure. There are 195 voice recordings in total, with approximately six recordings per patient. The first column of the dataset identifies the name of the patient. The "status" column is the target variable that indicates the health status of each individual. The value of 0 represents a healthy individual, and the value of 1 represents an individual with PD. The dataset is provided in ASCII CSV format, which contains 196 rows (including header) and 23 columns. The header row specifies the name of each voice measure, and the remaining rows contain the values of each measure for each voice recording.

B. Dimensionality Reduction :

In our analysis, we addressed the issue of multicollinearity and high-dimensional data by applying dimensionality reduction techniques. Rather than simply dropping highly correlated variables, we chose to utilize Principal Component Analysis (PCA) on different subsets of variables. This approach allowed us to condense the information contained in multiple related

TABLE I
THE DESCRIPTION OF DATASET

Voice measure	MEANING
MDVP-Fo (Hz)	Average vocal frequency
MDVP:Fhi (Hz)	Maximum vocal frequency
MDVP:Flo (Hz)	Minimum vocal frequency
MDVP- Iitter (%)	Measures of variation in
MDVP:Jitter (Abs)	Frequency
MDVP:RAP	MDVP relative amplitude perturbation
MDVP:PPQ	MDVP five-point quotient
Jitter:DDP	Average difference between jitter cycles
MDVP-Shimmer	Measures of variation in amplitude
MDVP-Shimmer (dB)	Measures of variation in amplitude in (dB)
Shimmer:APQ3	Three-point perturbation quotient
Shimmer:APQ5	Five-point perturbation quotient
MDVP:APQ11	MDVP 11-point perturbation quotient
Shimmer:DDA	Average differences between the amplitudes
NHR	Two measures of ratio of noise to tonal
HNR	components in the voice
RPDE	Two nonlinear dynamical complexity
D2	measures
DFA	Signal fractal scaling exponent
spread 1	Three nonlinear measures of fundamental
spread 2	frequency variation
PPE	Pitch period entropy
status	Healthy or Not

variables into a smaller number of uncorrelated components. For instance, we applied PCA to the group of Jitter variables, which consisted of five different measures.

C. The Evaluation of Proposed Architecture

1) Evaluation of Machine Learning Models and Deep Learning Models: Different machine learning and deep learning models have been implemented on the PD dataset and the results of all the machine learning models have been shown in Table II. From the table, we can observe that out of all the implemented machine learning models KNN has obtained a higher accuracy. A total of 4 deep learning models has been deployed and the results of the deep learning models. CNN + LSTM has obtained higher accuracy when trained and tested on data without PCA whereas GRU has better accuracy when trained and tested on data PCA.

2) Evaluation based on Metrics: For the purpose of evaluating our system, the standard standards for assessing classification models have been used. Fig. 2 is a representation of the learning curves of accuracy and loss for the training and validation of the models. These learning plots are evidence of an effective learning algorithm since the validation curve and the training curve both retain a point of stability with a minimum difference between them. In order to get the best possible results, the training of the effective model was designed to integrate three separate but interrelated tasks at the same time: 1) the calculation of output, 2) the correction of mistakes, and 3) the fine-tuning of the hyper-parameters. Following a number of rounds during which the hyper-parameters were fine-tuned, the highest training and validation accuracies, respectively, were found to be 99.5% and 98.1% when using a particular combination of hyper-parameters. The precision, recall, and F1 scores, along with other accuracy measures, were computed, as given in Table II.

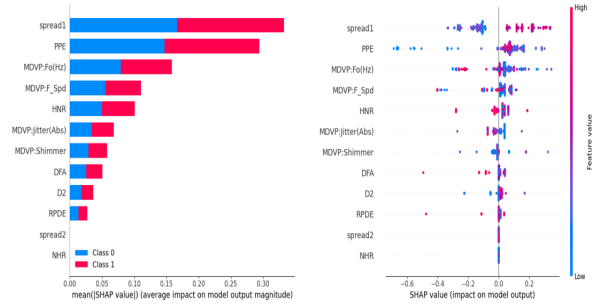

Fig. 2. Explainable AI (SHAP) Results

3) Impact of Training and Validation Splits: After being educated on our source data, it is imperative that we do an analysis of our model. When a model keeps the parameters of a periodic function and is then experimented on using the same data, this might lead to overfitting problem.

As a result, we verified our model by taking into account five distinct instances of data splits. Following a number of trials, the model's best training and validation accuracy are both 100%, with an 80:20 split of the results, as depicted in Fig. 3. Because of the highly generalized nature of the trained model, we were able to achieve a test accuracy of close to one hundred percent when we developed an independent test set that consisted of ten samples and was not a part of the training or validation process.

4) Explainable AI: SHAP bar plot shows the overall impact of each feature on the model's predictions. The bars represent the average SHAP value for each feature, and their length indicates the magnitude and direction of the impact. If a feature has a positive SHAP value, it means that increasing its value tends to increase the model's output, and vice versa for negative values. The SHAP bar plot show that spread1 has the highest positive impact on the model's predictions, followed by PPE and then MDVP:HFo(hz). This suggests that if the data consists larger spread1 value then it tends predict the given data point as disease.

Similarly SHAP dot plot shows the impact of each feature on individual predictions. Each dot represents a single observation in the dataset, and its position on the x-axis corresponds to the SHAP value for the corresponding feature. The y-axis can represent the actual value of the prediction or some other variable of interest. From the fig we can see that the spread1 had a high positive impact on its predicted price, while the PPE had a smaller positive impact, and the MDVP:HFo(hz) had a small negative impact. This suggests that the if the spread1 value is larger than average and has more PPE than average, which tends to increase its value.

V. CONCLUSION

In this research paper, we aimed to perform a comparative analysis of machine learning and deep learning-based approaches for Parkinson disease (PD) identification. Our

TABLE II
THE DESCRIPTION OF EVALUATION RESULTS OF MACHINE LEARNING MODELS

Model	Train Accuracy	Test Accuracy	F1 Score	Recalis]	Area Under Curve(AUC
KNN	94,85	94.92	96.47	95.35	94.55
SVM	70.47	3458	85.15	100	5312
Naive Baves	57.55	85.44	90.24	86,05	85.77
Logistic Regression	84.56	8644	91.49	100	90
Meta Cassfier	94.85	9492	96.35	97.67	9259
Decision Tree	8.76	86.44	91.49	100	75
Bagging	87.5	88.14	92.47	100	7812
Adaboost	96.32	88.14	91.95	93.02	84.01
Gradient Boost	96.37	91.53	9451	100	84.38
Random Forest	95.3	92.22	95.35	95.35	91.42

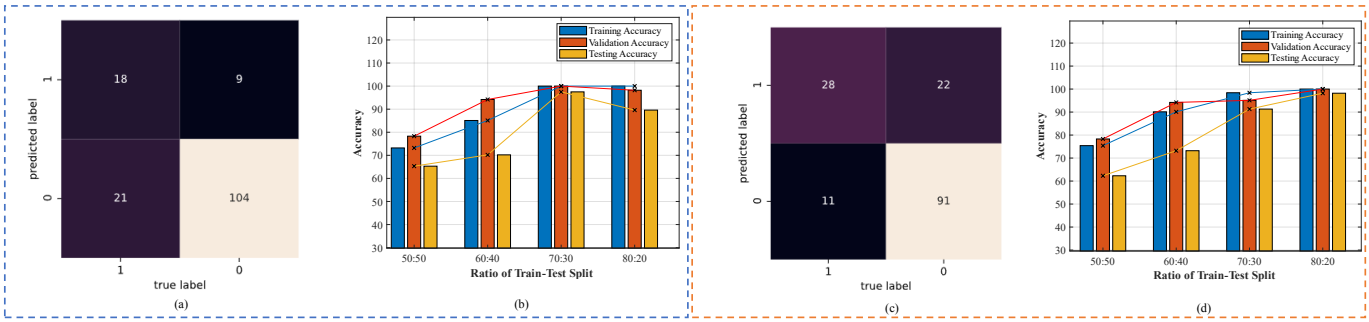

Fig. 3. Confusion Matrix and Impact of Train-Test-Validation Split of on (a)-(b) CNN+LSTM and (c)-(d) GRU

primary objective was to identify key predictors, especially speech-related characteristics of PD, using explainable AI. Furthermore, we aimed to build an end-to-end framework for PD identification, starting from patient data collection to the final diagnosis. Our results showed that both machine learning and deep learning-based approaches are effective in identifying PD from voice signals. However, deep learning models outperformed machine learning models in terms of accuracy and robustness. We also found that various speech-related characteristics, such as jitter, shimmer, and harmonic-to-noise ratio (HNR), were important predictors for PD identification. To make the diagnosis process more interpretable and explainable, we developed an explainable AI-based approach that not only provides accurate predictions but also highlights the key features that contribute to the prediction. Our approach can aid medical professionals in making informed decisions and improving the quality of care for PD patients. Finally, we built an end-to-end framework for PD identification, which involves data collection, preprocessing, feature extraction, model training, and prediction. Our framework can be used in clinical settings for early detection and diagnosis of PD, which is crucial for improving patient outcomes. Overall, our research provides insights into the use of machine learning and deep learning-based approaches for PD identification and highlights the importance of speech-related characteristics as key predictors. We hope that our work will contribute to the development of more accurate and interpretable models for PD diagnosis and ultimately improve the quality of life for PD patients.

REFERENCES

[1] P. S. Kumar and G. Akshay, "Deep learning-based spatio temporal facial feature visual speech recognition," *arXiv preprint arXiv:2305.00552*, 2023.

[2] J. Vijava, P. S. Kumar, M. Sorgile *et al.*, "Optimization techniques for deep learning based house price prediction," in *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*. IEEE, 2023, pp. 202–207.

[3] I. Mandal and N. Sairam, "Accurate telemonitoring of parkinson's disease diagnosis using robust inference system," *International journal of medical informatics*, vol. 82, no. 5, pp. 359–377, 2013.

[4] S. Grover, S. Bhartia, A. Yadav, K. Seeja *et al.*, "Predicting severity of parkinson's disease using deep learning," *Procedia computer science*, vol. 132, pp. 1788–1794, 2018.

[5] S. Aich, K. Younga, K. L. Hui, A. A. Al-Absi, and M. Sain, "A nonlinear decision tree based classification approach to predict the parkinson's disease using different feature sets of voice data," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 2018, pp. 1–2.

[6] C. R. Pereira, S. A. Weber, C. Hook, G. H. Rosa, and J. P. Papa, "Deep learning-aided parkinson's disease diagnosis from handwritten dynamics," in *2016 29th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*. Ieee, 2016, pp. 340–346.

[7] H. Choi, S. Ha, H. J. Im, S. H. Paek, and D. S. Lee, "Refining diagnosis of parkinson's disease with deep learning-based interpretation of dopamine transporter imaging," *NeuroImage: Clinical*, vol. 16, pp. 586–594, 2017.

[8] R. Das, "A comparison of multiple classification methods for diagnosis of parkinson disease," *Expert Systems with Applications*, vol. 37, no. 2, pp. 1568–1572, 2010.

[9] M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent system for accurate detection of parkinson's disease," *Computer methods and programs in biomedicine*, vol. 113, no. 3, pp. 904–913, 2014.

[10] F. Åström and R. Koker, "A parallel neural network approach to prediction of parkinson's disease," *Expert systems with applications*, vol. 38, no. 10, pp. 12 470–12 474, 2011.

[11] P. Santhosh Kumar, K. Supriya, and M. Rao K, "Covid-19 detection leveraging vision transformers and explainable ai," *arXiv e-prints*, pp. arXiv–2307, 2023.