Explainable Artificial Intelligence (EXAI) Model for early prediction of Parkinson detection using Deep Learning

**ABSTRACT:**

Parkinson's disease (PD) is a prevalent neurodegenerative disorder impacting millions worldwide, necessitating early detection to enable timely intervention and potential neuroprotective therapies. This research introduces an innovative approach by combining Deep Learning, specifically Long Short-Term Memory (LSTM) networks, with Explainable Artificial Intelligence (XAI) using Local Interpretable Model-agnostic Explanations (LIME) for the early prediction of PD based on speech signals.

The integration of LIME facilitates the generation of visual indicators, providing healthcare professionals with insights into the decision-making process of the LSTM model. This interpretability is crucial for building trust in the model's predictions and aiding clinicians in making informed decisions.

By combining the power of deep learning for accurate classification and explainability through LIME, our model strives to assist healthcare professionals in identifying Parkinson's disease at its prodromal stages. The proactive nature of early detection holds the potential to significantly impact patient outcomes by enabling timely intervention and improving the effectiveness of neuroprotective therapies.

The proposed LSTM model demonstrates remarkable accuracy, achieving 96.5%, with a precision of 0.984, F1-Score of 0.966, and a True Positive Rate (Sensitivity or Recall) of 0.943. The LSTM model serves as the foundation for our Explainable AI (XAI) framework, wherein LIME is applied to enhance transparency and interpretability.

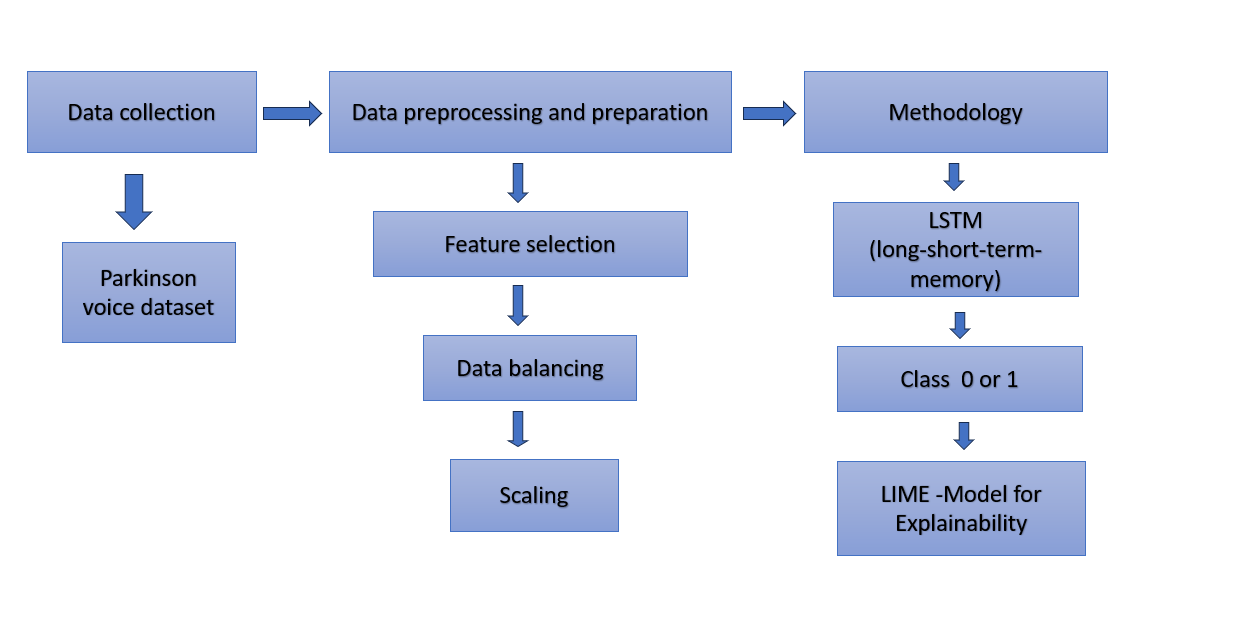
**INTRODUCTION:**

PARKINSON’S disease (PD) is the second most common neurodegenerative disorder after Alzheimer’s disease. PD primarily affects dopaminergic neurons in the substantia nigra of the brain, causing the loss of the neurotransmitter dopamine. PD is diagnosed based on the occurrence of four gross motor dysfunctions, which include bradykinesia, rigidity, resting tremor, and postural instability. However, by the time these dysfunctions are manifested in clinical diagnosis, up to 50% of the dopaminergic neurons are damaged beyond recovery, and the damage of the neurons has been found to increase rapidly during the four-year period after diagnosis. Any neuroprotective therapy performed after clinical diagnosis will be too late to effectively slow down neurodegeneration. Therefore, the early detection of PD in its prodromal stages is essential.

Among many other symptoms, speech disorders are manifested in PD patients at the prodromal stages as early as five years before the occurrence of gross motor dysfunctions. Speech disorders caused by PD can be characterized by symptoms such as reduced vocal tract volume and tongue flexibility, inappropriate pauses, impairments in voice quality, and reduction in pitch range and voice intensity. Speech-based assessment of PD has attracted increasing interest among researchers as an automatic, low-cost, and easy-to-administer method for detecting early PD. This study focuses the detection of PD by classifying speech signals into those produced by PD patients and those produced by healthy speakers.

This research endeavors to harness the potential of Deep Learning while prioritizing transparency and interpretability through an Explainable Artificial Intelligence (EXAI) model, aiming to assist healthcare professionals in identifying Parkinson’s disease at an early stage by displaying visual indicators produced by the model during its prediction.

**ARCHITECTURE DIAGRAM:**



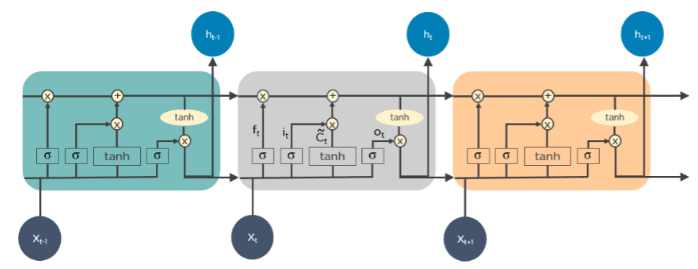
**METHODOLOGY**

**Dataset Description:**

The dataset comprises biomedical voice measurements from 31 individuals, including 23 with Parkinson's disease (PD). Each row represents one of 195 voice recordings, identified by the "Name" column. The dataset features various voice measures, such as fundamental frequency, frequency variation, amplitude variation, noise-to-tonal components ratio, nonlinear dynamical complexity measures, and nonlinear measures of fundamental frequency variation. The primary objective is to discriminate between healthy individuals (status 0) and those with PD (status 1). Notable measures include MDVP (Mean, Maximum, Minimum) for frequency, Jitter, Shimmer, noise-to-harmonics ratio (NHR), harmonics-to-noise ratio (HNR), and nonlinear complexity measures like RPDE, D2, and DFA. The dataset is in ASCII CSV format, facilitating analysis for early detection and understanding of Parkinson's disease based on voice characteristics.

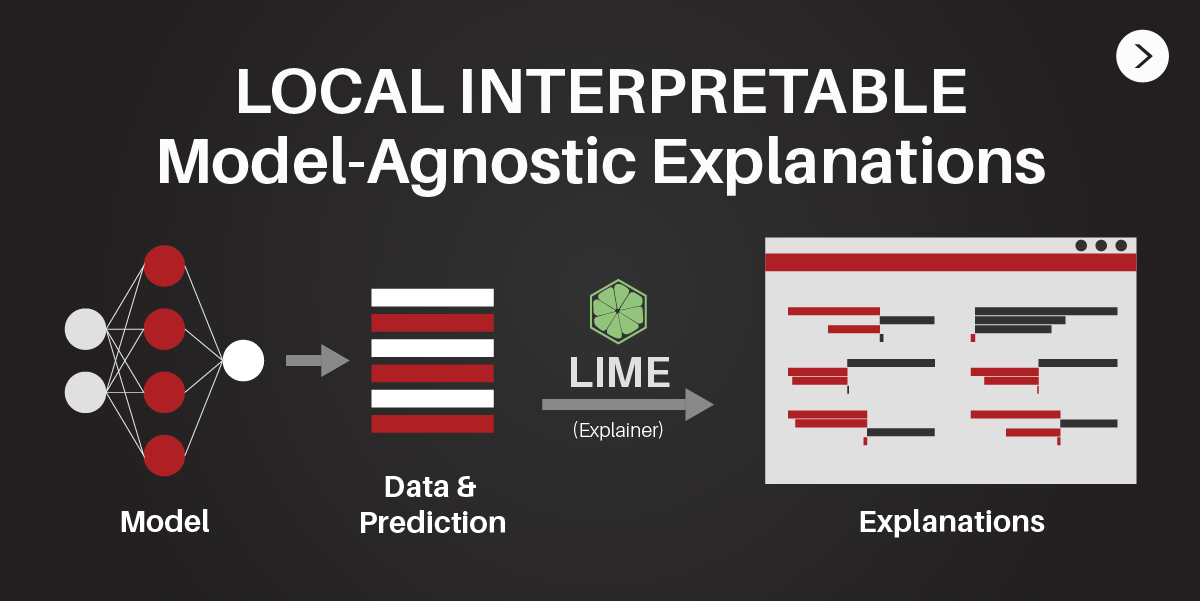
**Long Short-Term Memory Networks (LSTMs)**

LSTM work in a sequence of events. First, they don't tend to remember irrelevant details attained in the previous state. Next, they update certain cell-state values selectively and finally generate certain parts of the cell-state as output. Below is the diagram of their operation.



**LIME (Local Interpretable Model-agnostic Explanations Model Explanation):**

It is a method for explaining predictions of Machine Learning and deep learning models. The explainable model is usually exploited to understand which variables are the most important for the prediction for the specific individual.



**IMPLEMENTATION:**

**Data Collection:**

The dataset is sourced from the UCI Machine Learning Repository, consisting of a comprehensive set of features related to Parkinson's disease. Features include demographic information, clinical voice measurements, and other relevant parameters.

**Data Preprocessing:**

The data undergoes thorough preprocessing steps to ensure its quality and suitability for model training. Missing values and outliers are handled, and numerical features are standardized or normalized. Categorical variables are encoded, and any specific issues related to the dataset are addressed.

**Feature Selection:**

The SelectKBest method is applied for feature selection to identify the most informative features for the prediction task. This method is chosen for its simplicity and ability to handle both continuous and categorical features.

**Data Balancing:**

To address imbalanced classes, the Synthetic Minority Over-sampling Technique (SMOTE) is implemented. This technique generates synthetic samples to balance the distribution of the target variable, enhancing the model's ability to generalize across classes.

**Data Splitting:**

Trials are conducted with varying test sizes (0.2, 0.3, 0.4, 0.5) during data splitting. Different test sizes are explored to evaluate the model's sensitivity to the training-test partition.

**Model Selection:**

The choice of the LSTM model is motivated by its suitability for sequential data and its potential to capture temporal patterns in Parkinson's disease progression.

**Model Configuration**:

The LSTM model architecture is defined with an embedding layer, multiple LSTM layers, and dense layers. Hyperparameters such as the number of units, activation functions, and learning rates are specified based on experimentation and literature review.

**Model Training:**

The dataset is split into training and testing sets, and the LSTM model is trained using the training set. Model performance is monitored using relevant metrics during training.

**Model Evaluation:**

The trained LSTM model is evaluated on the testing set using metrics such as accuracy, precision, recall, F1 score, True positivity rate, True negative rate, False positive rate, False negative rate. Potential challenges, limitations, and areas for improvement are discussed.

**Comparative Analysis:**

A comparative analysis is conducted to compare the performance of the LSTM model with baseline models or other relevant algorithms. This analysis provides insights into the strengths and weaknesses of the proposed model.

**Fine-Tuning:**

Based on the evaluation results, hyperparameters are fine-tuned to optimize the model's performance. The rationale behind adjustments is explained.

**XAI Model for Explainability (LIME):**

LIME is introduced as the chosen Explainable AI (XAI) model for interpreting the predictions of the LSTM model. LIME generates perturbed samples and fits a local interpretable model, providing insights into the features influencing individual predictions.

**Results Interpretation:**

The results obtained from the LSTM model and LIME are interpreted. The importance of selected features in predicting Parkinson's disease is discussed. Any unexpected findings are addressed and contextualized.

**SYSTEM TESTING**

|  |  |
| --- | --- |
| 0.2 test size |  |
| accuracy | 0.9651 |
| Precision | 0.9856 |
| F1-Score: | 0.9641 |
| Recall | 0.9435 |
| Specificity | 0.9863 |
| False positive Rate | 0.0136 |
| False Negative Rate | 0.0564 |
| 0.3 test size |  |
| accuracy | 0.9625 |
| Precision | 0.9873 |
| F1-Score: | 0.9619 |
| Recall | 0.9377 |
| Specificity | 0.9877 |
| False positive Rate | 0.0122 |
| False Negative Rate | 0.0622 |

|  |  |
| --- | --- |
| 0.4 test size |  |
| accuracy | 0.9553 |
| Precision | 0.9765 |
| F1-Score: | 0.9543 |
| Recall | 0.9331 |
| Specificity | 0.9776 |
| False positive Rate | 0.0223 |
| False Negative Rate | 0.0668 |

|  |  |
| --- | --- |
| 0.5 test size |  |
| accuracy | 0.9524 |
| Precision | 0.9870 |
| F1-Score: | 0.9511 |
| Recall | 0.9177 |
| Specificity | 0.9877 |
| False positive Rate | 0.0122 |
| False Negative Rate | 0.0822 |

Based on the analysis, the 0.3 test size appears to be the best choice. It offers a good balance of:

High accuracy: Similar to the 0.2 test size, it correctly predicts most PD and non-PD cases.

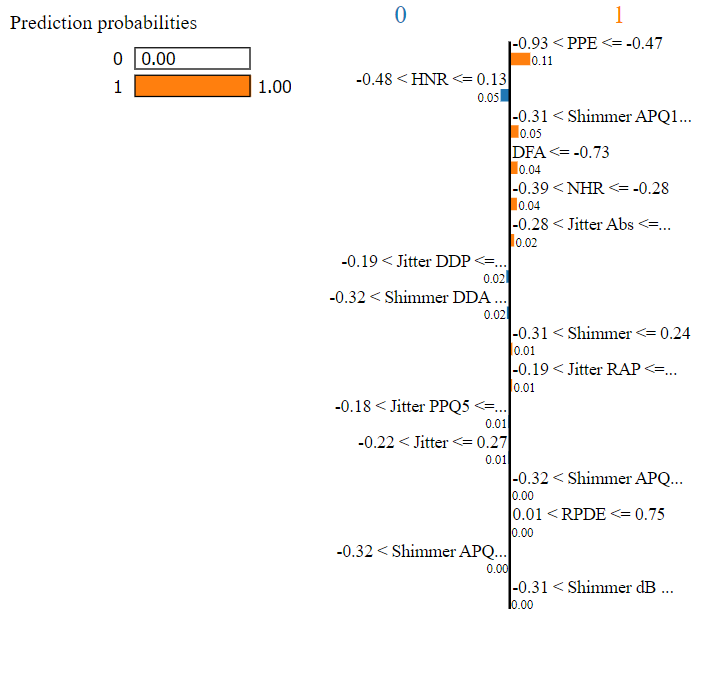
High precision: It minimizes false positives, accurately identifying non-PD cases.

Moderate recall: It captures a good majority of PD cases, although slightly less than the 0.2 test size.

Good F1-score: It combines precision and recall effectively, indicating overall good performance.

While the 0.2 test size has slightly higher recall, the trade-off is a lower precision. The larger test sizes (0.4 and 0.5) offer higher precision but have significantly lower recall, potentially missing more PD cases, which is crucial for early diagnosis. Therefore, the 0.3 test size provides a good balance of accuracy, precision, recall, and F1-score, making it the most suitable choice for this model.

Class 1 Prediction:

A screenshot of a cell phone

Description automatically generated

A blue and orange sign with white text

Description automatically generatedThe image shows two sets of probabilities: one for predicting class 0 (no Parkinson's disease) and one for predicting class 1 (Parkinson's disease). The left bar for class 0 is much shorter than the right bar for class 1, indicating that the model is much more confident in predicting Parkinson's disease in this case.

The features listed below the probabilities are the ones that the LIME model has identified as being most important for the model's prediction. Each feature is represented by a bar, with the length of the bar indicating the contribution of that feature to the model's prediction. A positive value means that the feature is associated with an increased probability of Parkinson's disease, while a negative value means that the feature is associated with a decreased probability of Parkinson's disease.

Here are the top features in this case:

PPE > 1.05: This feature refers to the sound pressure level of the voice signal, which is often elevated in people with Parkinson's disease.

-0.73 < DFA <= -0.35: This feature refers to the range of values that the DFA (detrended fluctuation analysis) feature can take on. DFA is a measure of the long-term variability in a signal, and it has been found to be associated with Parkinson's disease.

0.05 < NHR <= 0.28: This feature refers to the range of values that the NHR (noise-to-harmony ratio) feature can take on. NHR is a measure of the amount of background noise relative to the vocal signal, and a higher value can be indicative of Parkinson's disease.

Let's take a look at some of the lowest features in current LIME interpretation:

APQ5 < 0.19: This feature represents a low shimmer (variation in amplitude) in the APQ5 frequency band (around 5 kHz). This feature suggests a relatively steady voice, which is less likely to be associated with Parkinson's disease.

NHR > 0.07: This feature indicates a higher noise-to-harmony ratio, meaning there's more background noise relative to the vocal signal. While an excessively high NHR can be a symptom of Parkinson's, a moderately high value can also be due to external factors like a noisy environment, and in this case, it seems to be pushing the prediction away from Parkinson's.

-0.31 < Jitter RAP <= -0.14: This feature represents a range of low jitter (variation in frequency) values in the RAP frequency band (around 150 Hz). Low jitter suggests a stable voice, again pointing towards a lower likelihood of Parkinson's disease.

Recommendation:

While high PPE, DFA, and NHR suggest a higher risk of Parkinson's, low APQ5, high NHR (within a specific range), and low Jitter RAP indicate a lower risk.

Class 0 PredictionA screenshot of a graph

Description automatically generated

A blue and orange sign with white text

Description automatically generated

A blue and orange rectangular box with white text

Description automatically generated

High Features (Contributing to no Parkinson's disease prediction):

APQ3 < 0.27: This indicates a low shimmer in the APQ3 frequency band (around 3 kHz). Shimmer measures variations in amplitude, and a low value suggests a steady voice, less likely associated with Parkinson's.

0.20 < DFA <= 0.84: This range of DFA (detrended fluctuation analysis) values suggests relatively less long-term variability in the signal. High DFA can be linked to Parkinson's, so this range favors no disease.

-0.60 < Jitter DDA <= -0.27: This range of Jitter DDA (variation in frequency in the DDA band) falls within a zone potentially less indicative of Parkinson's disease. High Jitter DDA can be a symptom, but this specific range might not raise concerns here.

NHR > 0.16: This feature shows a higher noise-to-harmony ratio, meaning more background noise relative to the vocal signal. While an excessively high NHR can be a concern, a moderate value like this can also be due to external factors and doesn't necessarily suggest Parkinson's in this case.

Low Features (Further supporting no Parkinson's disease prediction):

-0.53 < Jitter PPQ5 <= -0.02: This range of Jitter PPQ5 (variation in frequency in the PPQ5 band) falls within a zone potentially less indicative of Parkinson's disease. High Jitter PPQ5 can be a symptom, but this specific range might not raise concerns here.

Shimmer dB < 0.26: This indicates a low Shimmer dB, another measure of amplitude variation. Similar to APQ3, a low value suggests a steady voice, less likely associated with Parkinson's.

-0.60 < Jitter Abs <= -0.33: This range of Jitter Abs (absolute jitter) falls within a zone potentially less indicative of Parkinson's disease. High Jitter Abs can be a symptom, but this specific range might not raise concerns here.

Recommendation :

While some features like low shimmer and DFA suggest a lower risk of Parkinson's, others like moderate NHR remain within a range that wouldn't definitively rule out Parkinson.

**SCOPE FOR FEATURE ENHANCEMENTS**

Future enhancements for the Parkinson's disease detection project could involve the integration of multi-modal data, including imaging and genetic information, to provide a more comprehensive understanding of the disease. Longitudinal data analysis may capture dynamic disease progression over time, enhancing the model's sensitivity to subtle changes. Implementing online learning would enable continuous monitoring and adaptive predictions based on evolving patient data. External validation on diverse datasets and real-world testing in clinical settings would assess the model's generalizability and practical applicability. Moving towards patient-specific predictions by considering individual factors and developing a mobile application for self-assessment could encourage proactive healthcare seeking. Develop a user-friendly mobile application for easy and widespread deployment of the model. The application could enable individuals to self-assess their speech patterns, providing an early indication of potential Parkinson's disease symptoms and encouraging proactive healthcare seeking.Improving explainability through additional XAI techniques, addressing ethical considerations, and integrating the model into clinical trials would further enhance the project's impact on early Parkinson's disease detection and intervention.

**CONCLUSION**

In conclusion, this paper presents a groundbreaking approach to Parkinson's disease (PD) detection through the integration of Deep Learning, specifically Long Short-Term Memory (LSTM) networks, and Explainable Artificial Intelligence (XAI) using Local Interpretable Model-agnostic Explanations (LIME). The fusion of these technologies not only achieves an impressive accuracy of 96.5% but also addresses the critical need for interpretability in medical AI models. By integrating LIME, we unlock the model's decision-making process, shedding light on the features that guide its predictions. This transparency is critical in building trust and empowering healthcare professionals with actionable insights; they gain a nuanced understanding of the model's reasoning, enabling them to make informed decisions with greater confidence. This synergy between Deep Learning's predictive power and Explainable AI's transparency opens doors to a paradigm shift in the early detection of Parkinson's disease. Proactive intervention, spurred by timely diagnosis, becomes a tangible possibility. The effectiveness of neuroprotective therapies can be significantly enhanced, offering hope for improved patient outcomes and potentially even disease modification.

In essence, the synergy between deep learning for accurate classification and explainability through LIME positions our model as a valuable tool for healthcare professionals. The robust performance metrics, coupled with the interpretability provided by XAI, underscore the potential of our approach to revolutionize Parkinson's disease detection, setting the stage for improved patient care and proactive intervention strategies.

Signature of the guide with date