UNDERSTANDING OF PARKINSON'S DISEASE THROUGH EXPLAINABLE ARTIFICIAL INTELLIGENCE APPLIED TO SPEECH MEASUREMENTS

A DISSERTATION SUBMITTED TO MANCHESTER METROPOLITAN UNIVERSITY
FOR THE DEGREE OF MASTER OF SCIENCE
IN THE FACULTY OF SCIENCE AND ENGINEERING

By
RAMSES MORENO DE LA CRUZ
Department of Computing and Mathematics

WHY…?

- No matter the field or topic, it is very common to see that in a situation of high dimensionality space, some features are dropped or discarded.

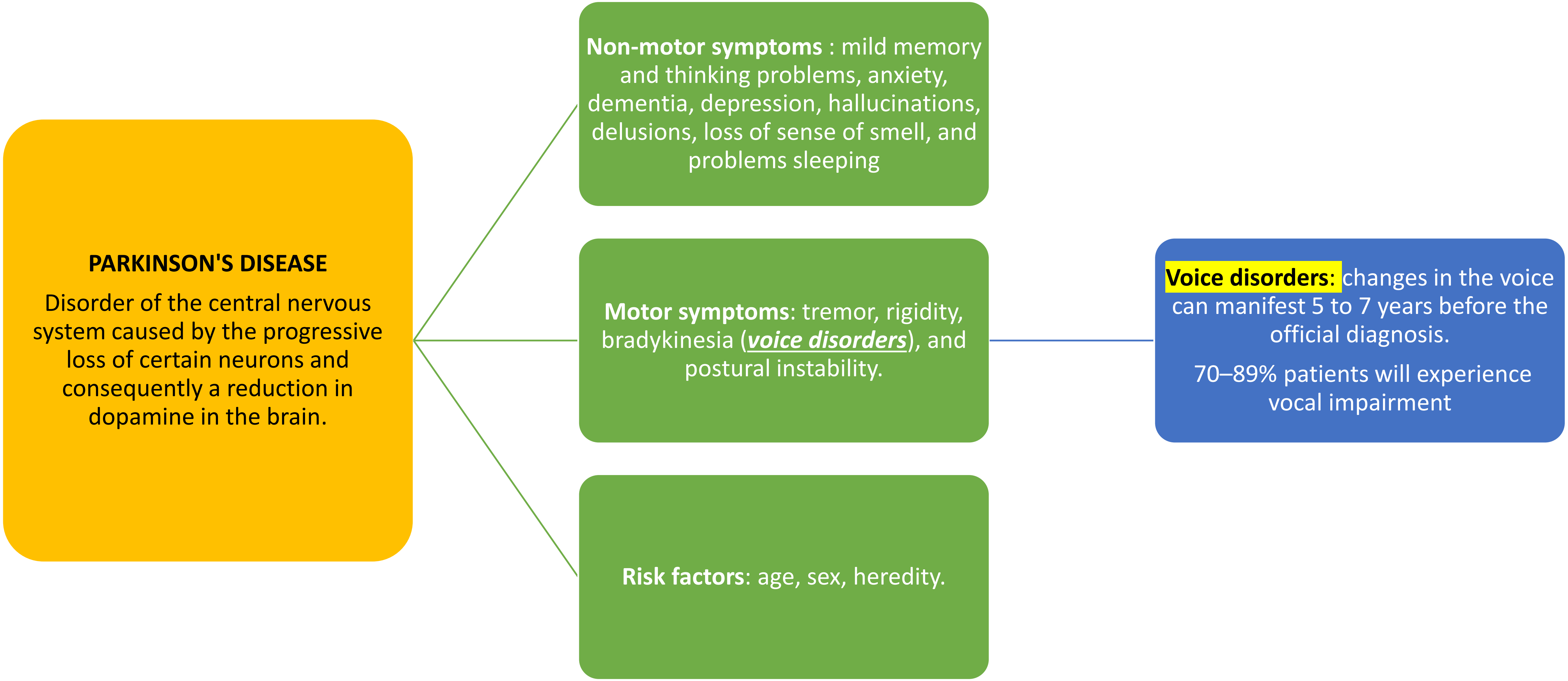
- **BUT.., WHAT IF** in the context of voice biomarkers for the detection and treatment of Parkinson’s disease, there are multidimensional and complex interactions that might be missed or overlooked by traditional machine learning models, but could be harnessed by deep learning models, improving the existing performance metrics, insights, and interpretability of the intricate relationships of the features.

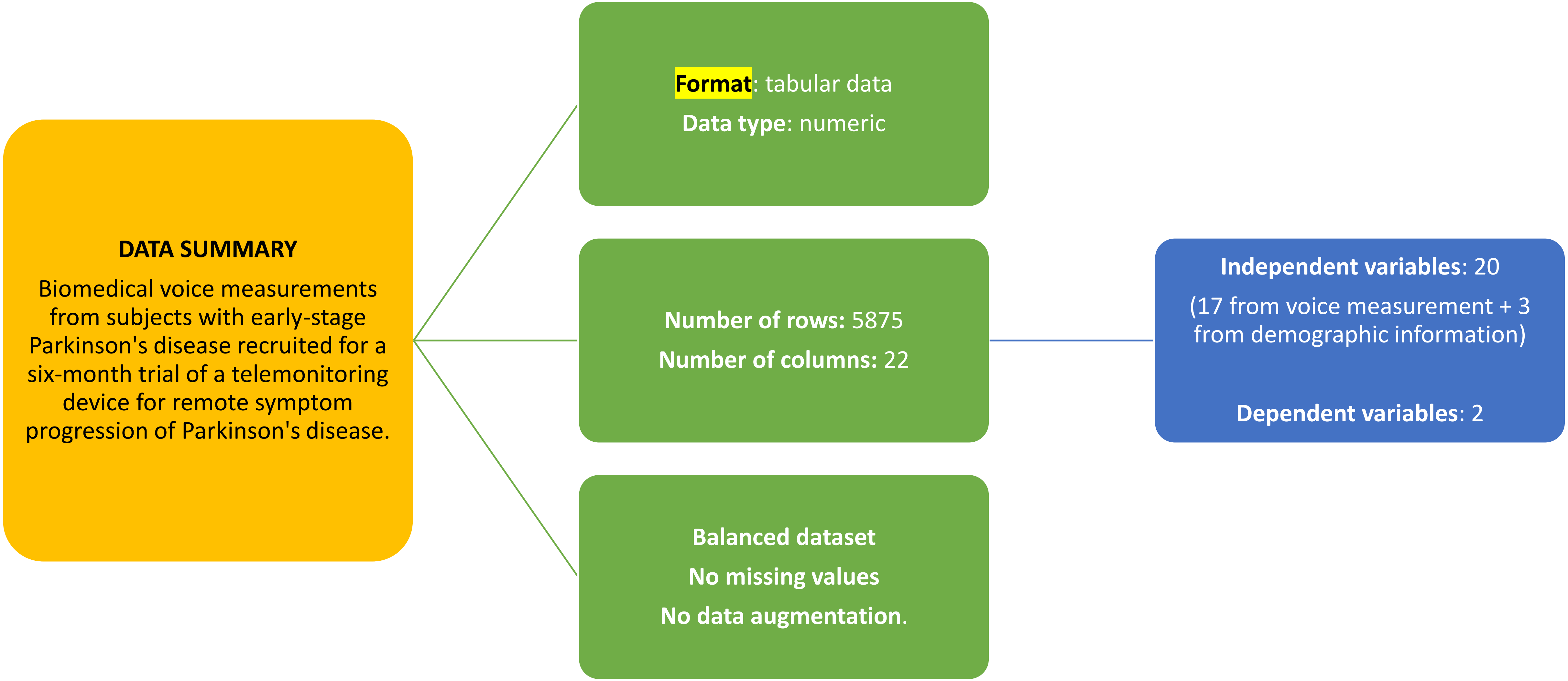
AIM

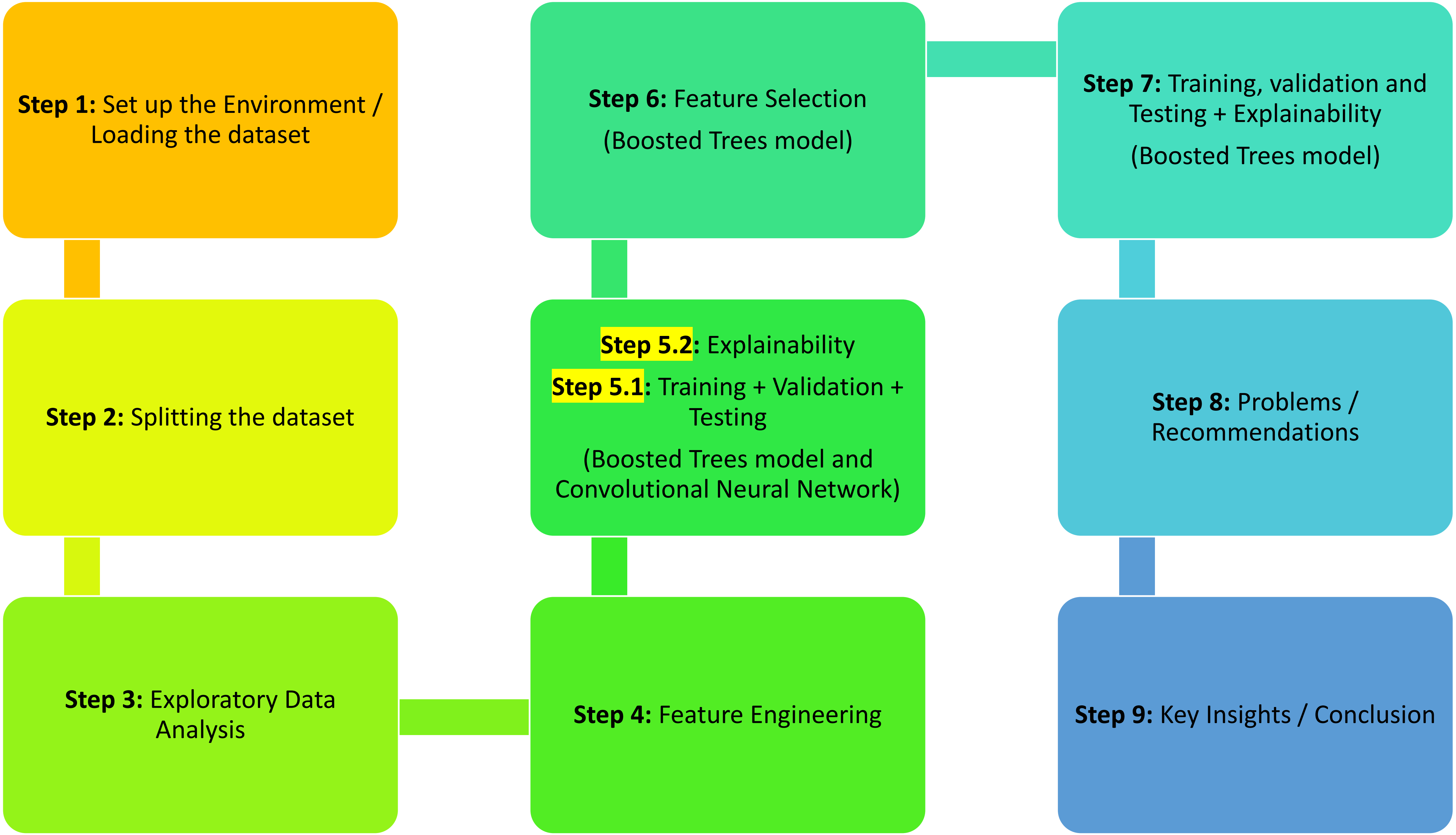
- ***Provide to healthcare professionals with an effective and easy-to-use technological tool based on machine learning, deep learning, and explainable artificial intelligence that sheds light on the way the models make decisions and describes the most important features related to speech measurements for understanding Parkinson's disease.***

WHAT TO DO? → OBJECTIVES

- Literature survey.
- Exploratory Data Analysis.
- Design, build, and run a Machine Learning model (Boosted Trees) and a Deep Learning model (Convolutional Neural Network) for a regression task.
- Apply Explainable Artificial Intelligence technique (Shapley Values).


PARKINSON'S DISEASE
Disorder of the central nervous system caused by the progressive loss of certain neurons and consequently a reduction in dopamine in the brain.
Non-motor symptoms : mild memory and thinking problems, anxiety, dementia, depression, hallucinations, delusions, loss of sense of smell, and problems sleeping
Motor symptoms: tremor, rigidity, bradykinesia (voice disorders), and postural instability.
Risk factors: age, sex, heredity.
Voice disorders: changes in the voice can manifest 5 to 7 years before the official diagnosis.
70–89% patients will experience vocal impairment


DATA SUMMARY
Biomedical voice measurements from subjects with early-stage Parkinson's disease recruited for a six-month trial of a telemonitoring device for remote symptom progression of Parkinson's disease.
Format: tabular data
Data type: numeric
Number of rows: 5875
Number of columns: 22
Balanced dataset
No missing values
No data augmentation.
Independent variables: 20
(17 from voice measurement + 3 from demographic information)
Dependent variables: 2


Step 1: Set up the Environment / Loading the dataset
Step 2: Splitting the dataset
Step 3: Exploratory Data Analysis
Step 4: Feature Engineering
Step 5.1: Training + Validation + Testing
Step 5.2: Explainability
(Boosted Trees model and Convolutional Neural Network)
Step 6: Feature Selection
(Boosted Trees model)
Step 7: Training, validation and Testing + Explainability
(Boosted Trees model)
Step 8: Problems / Recommendations
Step 9: Key Insights / Conclusion

Step 2: Splitting the dataset

Table 5. Data Splitting

	Percentage of instances from the whole dataset	Number of instances from the whole dataset
Data Subset		
Training set	~ 60%	3584
Validation set	~ 20%	1145
Test set	~ 20%	1146
Total	100%	5875

Step 3: Exploratory Data Analysis

Comparison of Pearson and Spearman Correlation Coefficient

	Metric Variable	Pearson Correlation	Metric Variable	Spearman Correlation
0	motor_UPDRS	1.000000	motor_UPDRS	1.000000
1	total_UPDRS	0.947476	total_UPDRS	0.957649
2	age	0.282680	age	0.312043
3	subject	0.209668	subject	0.197384
4	PPE	0.151383	PPE	0.149625
5	Shimmer:APQ11	0.116734	Shimmer:APQ11	0.142993
6	RPDE	0.103239	Shimmer(dB)	0.122358
7	Shimmer(dB)	0.091120	NHR	0.118600
8	Shimmer	0.084116	Shimmer	0.117644
9	Jitter(%)	0.074269	Jitter(%)	0.114440
10	Shimmer:APQ5	0.072014	Jitter:PPQ5	0.106485
11	test_time	0.069940	Shimmer:APQ5	0.100105
12	Jitter:PPQ5	0.069145	Shimmer:APQ3	0.095644
13	Shimmer:APQ3	0.066514	Shimmer:DDA	0.095633
14	Shimmer:DDA	0.066514	Jitter:DDP	0.094104
15	NHR	0.066099	Jitter:RAP	0.094024
16	Jitter:DDP	0.063351	RPDE	0.087109
17	Jitter:RAP	0.063341	test_time	0.062601
18	Jitter(Abs)	0.039854	Jitter(Abs)	0.058422
19	sex	-0.050313	sex	-0.056787
20	HNR	-0.133199	HNR	-0.132891
21	DFA	-0.133771	DFA	-0.146948

Figure 3. Comparison of the Pearson and Spearman Correlation Coefficients of the features related to the target variable (motor_UPDRS)

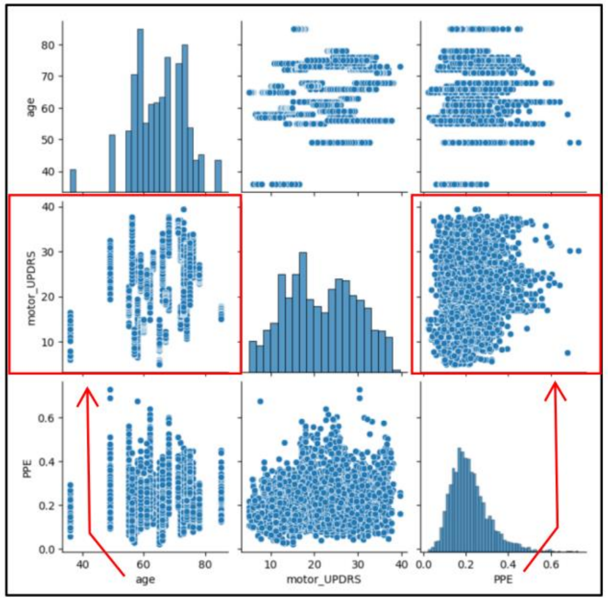


Figure 4. Pair plot of some features related to the target variable (motor_UPDRS)

Step 5.1: Training, Validation and Testing

(Boosted Trees model and Convolutional Neural Network)

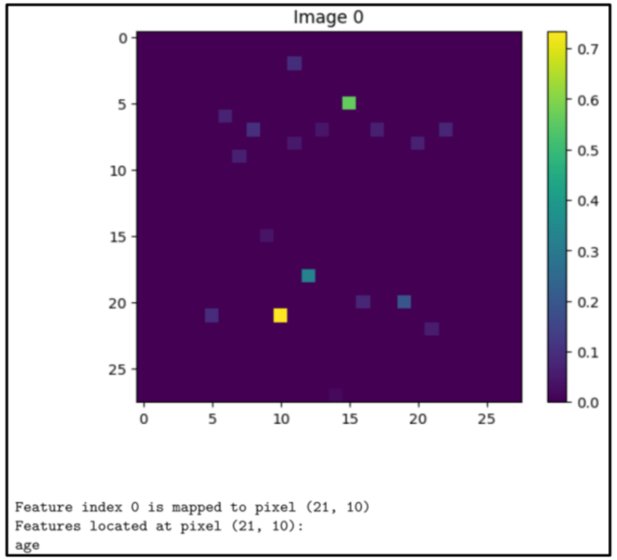


Figure 5. Instance of the training set transformed into an image

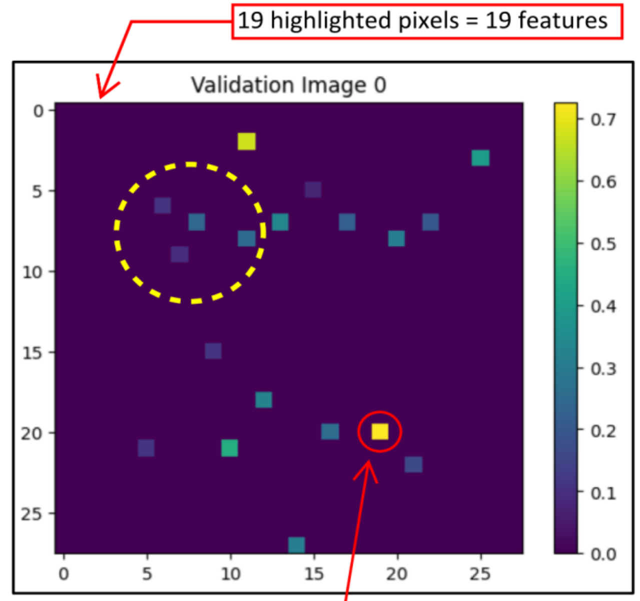


Figure 6. Instance of the validation set transformed into an image

Phase	Model	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Mean Square Error (MSE)	Coefficient of Determination (R^2)
Training	Boosted Trees	1.21	1.66	2.77	0.96
	Convolutional Neural Network	6.43	7.66	58.66	-
Validation	Boosted Trees	7.46	-	-	-
	Convolutional Neural Network	7.17	8.70	75.66	-
Test	Boosted Trees	1.23	2.59	6.70	0.90
	Convolutional Neural Network	6.52	7.64	58.43	-

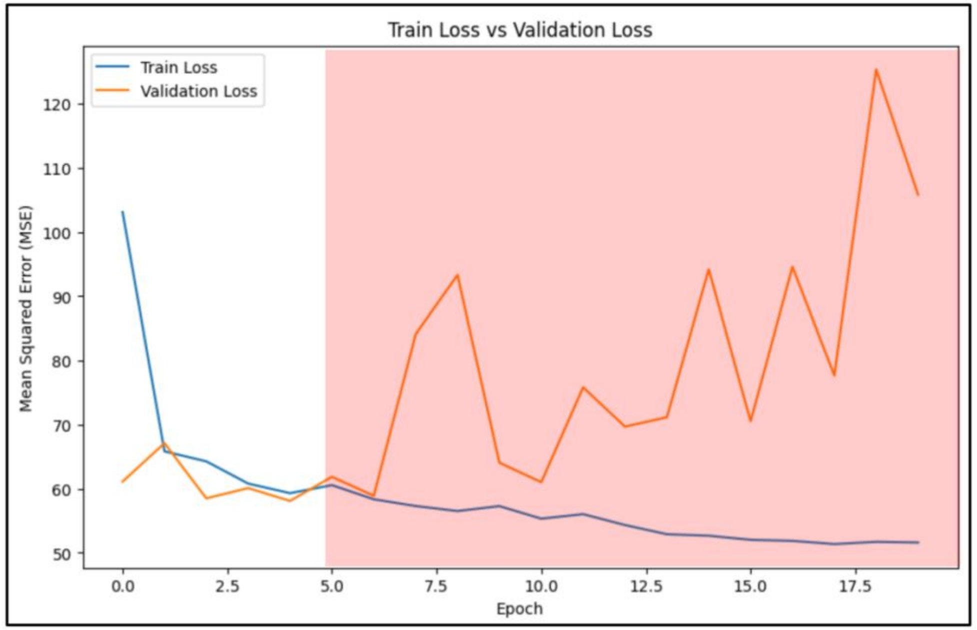


Figure 7. Train Loss vs Validation Loss, CNN

Step 5.2: Explainability

(Boosted Trees model and Convolutional Neural Network)

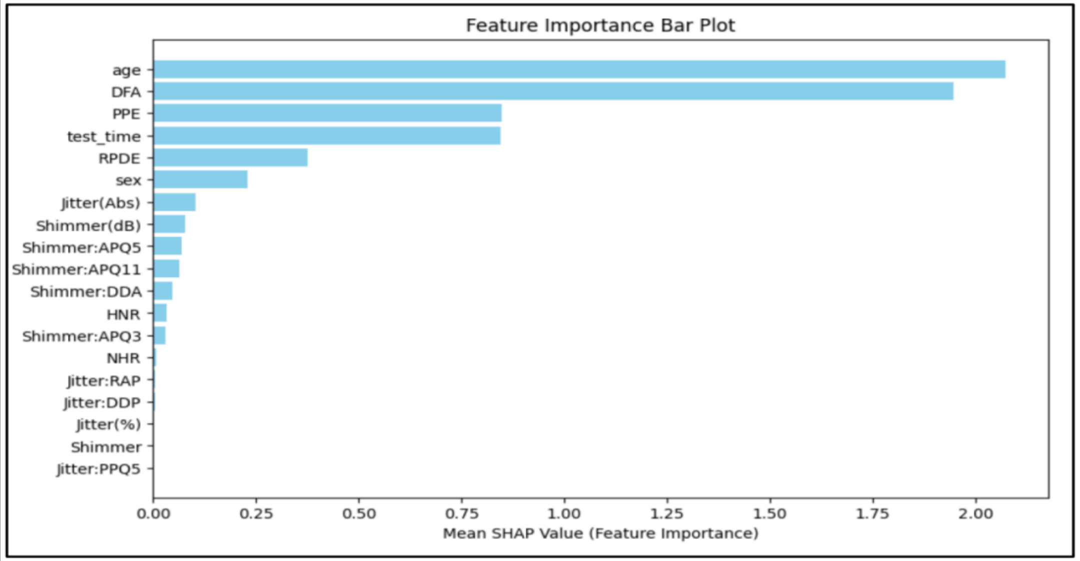


Figure 10. Mean SHAP Value (Feature Importance) – Global Impact, CNN

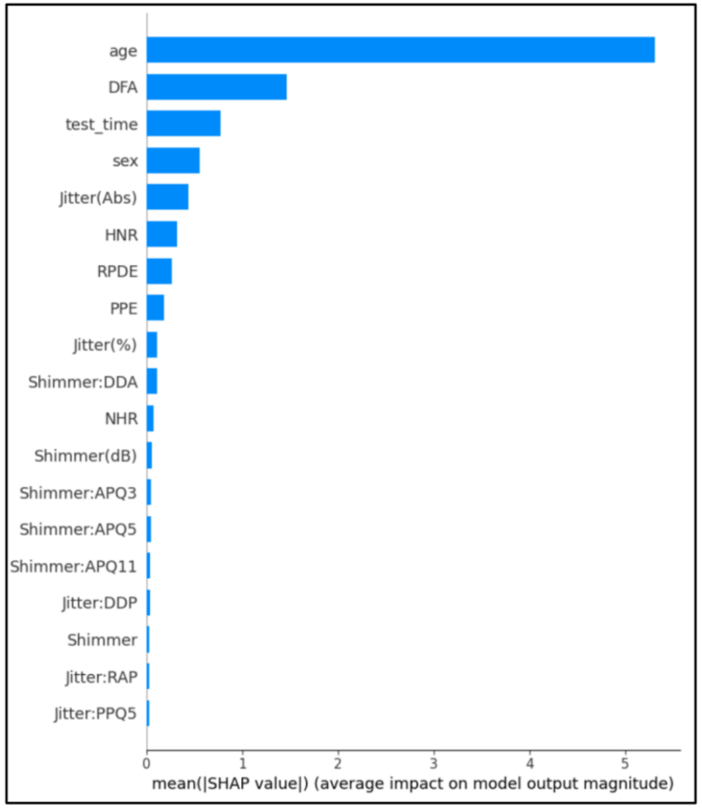


Figure 8. Mean SHAP Value (Average impact on model output), Boosted Trees

Table 10. Top variables according to SHAP Value for Boosted Trees and CNN

N°	Boosted Trees Model Mean Shapley Value	Convolutional Neural Network Mean Shapley Value
1	age	age
2	DFA	DFA
3	test_time	PPE
4	sex	test_time
5	Jitter (Abs)	RPDE
6	HNR	sex
7	RPDE	Jitter (Abs)
8	PPE	Shimmer(dB)
9	Jitter (%)	Shimmer: APQ5
10	Shimmer: DDA	Shimmer: APQ11
11	NHR	Shimmer: DDA
12	Shimmer(dB)	HNR
13	Shimmer: APQ3	Shimmer: APQ3
14	Shimmer: APQ5	NHR
15	Shimmer: APQ11	Jitter: RAP
16	Jitter: DDP	Jitter: DDP
17	Shimmer	Jitter (%)
18	Jitter: RAP	Shimmer
19	Jitter: PPQ5	Jitter: PPQ5

Step 6: Feature Selection

(Boosted Trees model)

Step 7: Training, Validation and Testing + Explainability

(Boosted Trees model)

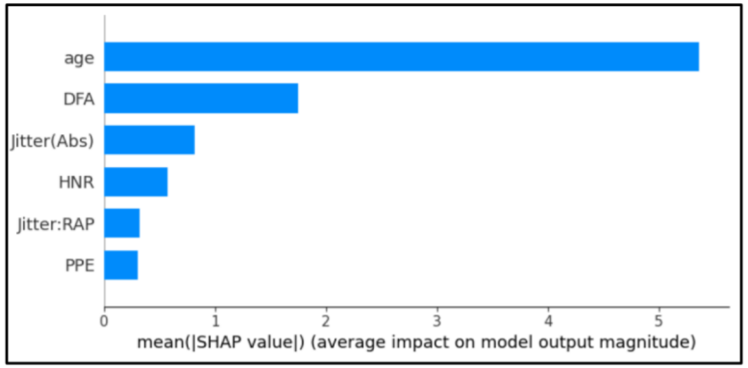


Figure 14. Mean SHAP Value (Average impact on model output) – Global Impact, Boosted Trees

Table 12. Summary of the Feature Selection Process

Feature selection Technique	Number of Features (Before "feature selection" → After "feature selection")	Feature name
K-best	19 → 4	age
		HNR
		DFA
		PPE
Feature Elimination with Cross Validation	19 → 1	Jitter (RAP)
Sequential Feature Selector	19 → 3	age
		Jitter (Abs)
		DFA

Phase	Model	Mean Absolute Error (MAE)	Root Mean Square Error (RMSE)	Mean Square Error (MSE)	Coefficient of Determination (R^2)
Training	Boosted Trees	1.21	1.66	2.77	0.96
	Boosted Trees (Improved Version)	2.00	2.72	7.42	0.88
Validation	Boosted Trees	7.46	-	-	-
	Boosted Trees (Improved Version)	7.10	-	-	-
Test	Boosted Trees	1.23	2.59	6.70	0.90
	Boosted Trees (Improved Version)	2.31	3.52	12.41	0.81

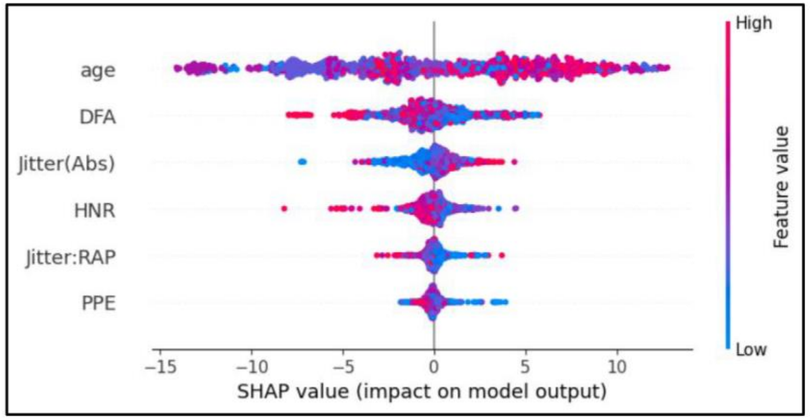


Figure 15. SHAP Value (Impact on model output) – Global Impact, Boosted Trees

Step 8: Problems / Recommendations

Main problems:

1. Lack of publicly available data.
2. Error / issues in the cloud-based platform Google Colaboratory.

Recommendations:

1. Data augmentation.
2. Pre-trained convolutional neural networks.
3. Other approaches for data splitting.
4. Different strategies at the moment to transform the tabular data into image.
5. Fine-tuning process.
6. GPUs.

Step 9: Key Insights / Conclusion

Key Insights:

1. The input mismatch (i.e., not the same feature space) in the models may have made the interpretation and comparison of the results difficult.
2. The transformation into images, which in this case was through the novel algorithm called "Image Generator for Tabular Data," could potentially introduce some complexity, loss of information, or noise degrading the performance of the convolutional neural network.

Conclusion:

1. The experimental evidence suggests that the Boosted Trees model is more convenient than a convolutional neural network considering several performance metrics, computational resources used, complexity during the fine-tuning process, and interpretability. The context of this conclusion takes into account the conditions of the dataset of biomedical speech measurements of subjects with early-stage Parkinson's disease in tabular format and as images after its respective transformation.