# Gradient Boosting and Minimum Redundancy Maximum Relevance (mRMR) Feature Selection for Diagnosis of Parkinson's Disease through Patient Audio Data

Jagadeepram Maddipatla
Rock Ridge High School
jag.maddipatla@gmail.com

Rishi Athavale
Academy of Engineering and Technology
rishi.athavale1@gmail.com

***Abstract*- Parkinson's Disease (PD) consistently ranks as one of the most common neurodegenerative diseases nationally within the U.S., behind only Alzheimer's. While current research to fight the underlying causes of PD are underway, diagnosis remains a pertinent issue due to an absence of formal lab based tests. Patients with PD often develop abnormal speech patterns, such as the slurring of words. This study aims to utilize and compare various different Gradient Boosting models to diagnose PD based on audio data taken from a patient. Additionally, this study also tests these models with mRMR feature selection, Principal Component Analysis (PCA) dimensionality reduction, Min-Max normalization, and standardization. The dataset used in this study was the UCI Machine Learning Repository's Parkinsons Data Set (available here). The dataset includes 195 voice recordings from 31 subjects with 22 biomedical voice measurements per recording. 48 of the voice recordings in the dataset were from patients without PD and 147 of the voice recordings were from patients with PD. 60% of the voice recordings were used for model training, 20% for cross validation, and 20% for the test set. The HistGradientBoosting model with Min-Max normalization and mRMR feature selection was found to perform the best on the cross validation set, in which it had an accuracy of 94.8718%, a sensitivity of 96.7742%, a specificity of 87.5000%, an AUC score of 0.921371, and an F1 score of 0.967742. This same model also achieved an accuracy of 89.7436%, a sensitivity of 96.6667%, a specificity of 66.6667%, an AUC score of 0.816667, and an F1 score of 0.935484 on the test set. The results in this study show that Gradient Boosting models have the potential to provide quick, efficient, and accurate diagnoses for PD in a clinical setting so patients can receive treatment sooner.**

## 1. Introduction

Parkinson's Disease (PD) is a neurological disease which persists typically amongst the elderly, though not entirely. While the exact causes of the disease vary amongst the affected population, all exhibit injury within the basal ganglia and substantia nigra portions of the brain [8]. These regions are most closely correlated with voluntary movement and dopamine assembly; thus, excessive damage and inhibition of the neurons can lead to noticeable manifestations of PD. For example, common in a majority of PD patients is an involuntary tremor within the hands and sometimes feet [11]. Motor control and movement is inhibited as well, with sudden bouts of muscle rigidity preventing typical bodily actions [11].

Currently, doctors and laboratories tasked with diagnosing PD base their reports upon symptom descriptions and brain scans, particularly dopamine mapping. In particular, tremors and muscle stiffness symptoms typically reported by PD patients result in the official diagnosis of the disease due to its connection to the substantia nigra [3]. Rigorous PD diagnosis, however, is unable to be conducted properly simply due to the unavailability of clinical PD tools and unique symptoms of PD. While the disease results in the manifestation of numerous symptoms, these are oftentimes not related solely to PD, and can also be attributed to a variety of other diseases and disorders. Utilizing dopamine mapping as the basis for PD diagnosis also leads to dramatically limited accessibility to patients worldwide due to the lack of proper imaging tools.

In conjunction with the aforementioned symptoms, PD patients also experience a change in speech patterns, often with slight variations in enunciation [3]. While such changes are typically too

insignificant to be noticed by the human observer on individual cases, tell-tale patterns are clear within frequency metrics that are derived from vocal recordings. The prevalence of such a symptom within the PD population as a whole allows for an opportunity to differentiate active PD patients through a machine learning approach. Unlike other widely considered symptoms, PD voice discrepancies are specific to the PD disease, and a diagnosis tool based around voice fluctuations could prove to be comparatively rigorous in the diagnostic process.

## 2. METHODS

### *2.1 Dataset*

This study utilized biomedical voice measurements of voice recordings to both train and evaluate Gradient Boosting models to detect PD. This data was taken from the UCI Machine Learning Repository's Parkinsons Data Set. This dataset includes a total of 195 voice recordings taken from 31 subjects, 23 of whom have PD and 8 of whom are healthy. Subjects who have PD are labeled with a 1 and healthy subjects are labeled with a 0. For each voice recording, 22 biomedical voice measurements are included. These biomedical voice measurements include the average vocal fundamental frequency (MDVP:Fo(Hz)), the maximum vocal fundamental frequency (MDVP:Fhi(Hz)), the minimum vocal fundamental frequency (MDVP:Flo(Hz)), five measures of variation in fundamental frequency (MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP), six measures of variation in amplitude (MDVP:Shimmer, MDVP:Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA), two measures of ratio of noise to tonal components in the voice (NHR, HNR), two nonlinear dynamical complexity measures (RPDE, D2), signal fractal scaling exponent (DFA), and three nonlinear measures of fundamental frequency variation (spread1, spread2, PPE). The dataset is then split, with 60% of the voice recordings (117 voice recordings) going to the training set, 20% of the voice recordings (39 voice recordings) going to the cross validation set, and 20% of the voice recordings (39 voice recordings) going to the test set. The training set is used to train the Gradient Boosting models and the cross validation set is used to compare them. Once the best model is determined based on cross validation set performance, it is then evaluated on the test set to measure its performance on new data outside of the training or cross validation sets.

### *2.2 Gradient Boosting*

Gradient Boosting is an ensemble machine learning model that utilizes numerous decision trees [1]. Gradient Boosting starts with an initial prediction $F_0(x)$ is shown below:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, \gamma) \quad (1)$$

For Equation (1), L is the loss function and $y^i$ is the $i^{th}$ label.

Once an initial prediction is made, regression trees are then constructed based on the pseudo residuals of the previous prediction [1]. The equation for calculating the pseudo residuals is shown below:

$$r_{im} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)} \text{ for } i = 1...n \quad (2)$$

For Equation (2), $r_{im}$ is the pseudo residual for sample $i$. This pseudo residual will be used to create regression tree $m$.

To create regression tree $m$, a regression tree is fitted to the pseudo residuals [1]. The terminal regions of the regression tree are denoted by $R_{jm}$, where $j$ is the number of the terminal region in the regression tree and $m$ is the number of the regression tree [1]. The output for each leaf node the tree is then computed with the following equation:

For $j$ = i...$J_m$:

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (3)$$

For Equation (3), $J_m$ is the number of terminal regions for regression tree $m$.

Using the outputs from the tree, the predictions (denoted by $F_m(x)$) are now updated.

$$F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}) \quad (4)$$

For Equation (4), $v$ is the learning rate.

This process is then repeated $M$ times, with each iteration generating a new tree [1]. The final output is denoted by $F_M(x)$.

There are multiple other variations of Gradient Boosting which introduce improvements to the base algorithm. XGBoost (eXtreme Gradient Boosting) is a version of Gradient Boosting that improves on scalability [12]. LightGBM (Light Gradient Boosting) decreases memory usage and training time [4]. CatBoost allows for automatic handling of categorical features and reduces overfitting [7].

### *2.3 Model Implementation*

This study utilized five different Gradient Boosting models: XGBoost, HistGradientBoosting, GradientBoosting, LightGBM, and CatBoost.

XGBoost was implemented using the XGBClassifier class from the xgboost Python. HistGradientBoosting and GradientBoosting were implemented using sklearn's HistGradientBoostingClassifier and GradientBoostingClassifier classes. LightGBM was implemented using the LGBMClassifier class from the lightgbm Python library. CatBoost was implemented using the CatBoostClassifier class from the catboost Python library.

*2.4 Data Preprocessing*

The biomedical voice measurements in the data were also normalized using Min-Max scaling. For a given feature, MinMax scaling subtracts the minimum value for that feature and then divides the result by the maximum value for that feature. The equation for Min-Max scaling is shown below:

$$x_i' = \frac{x_i - min(x_i)}{max(x_i) - min(x_i)} \quad (5)$$

For Equation (5), $x^i$ is the $i^{th}$ feature. To prevent data leakage, the maximum and minimum feature values used were taken from the training set.

Standardization was also tested as an additional data preprocessing method. Standardization works similarly to Min-Max scaling, except that in Standardization the mean of the values for a feature are subtracted from each feature value and the result is then divided by the standard deviation for the feature values.

$$x_i' = \frac{x_i - \mu}{\sigma} \quad (6)$$

For Equation (6), μ represents the mean of xi and σ represents the standard deviation of xi . For standardization, the mean and standardization values used were taken from the training set.

Min-Max scaling, and Standardization were all tested for each model and with both mRMR feature selection and PCA.

*2.5 mRMR Feature Selection*

Minimum redundancy maximum relevancy (mRMR) feature selection is an algorithm that identifies the best group of K features in a dataset [13]. Unlike other feature selection algorithms like Boruta that seek to identify features that have any predictive capability, mRMR identifies a small subset of features that will be the most useful [13]. For this dataset in particular, which includes numerous redundant features (such as the six different measures of variation in amplitude), decreasing the number of features to the most essential will help to eliminate redundant features and potentially improve model performance. For this study, mRMR feature selection was used to reduce the number of features from 22 to 20. Additionally, to prevent data leakage, the features will be selected based on training data. The models will be tested both with and without mRMR feature selection.

*2.6 PCA Dimensionality Reduction*

Similar to mRMR, Principal Component Analysis (PCA) reduces the number of features inputted to the model [2]. However, unlike mRMR which selects features to utilize, PCA condenses features that correlate with one another into a new feature [2]. This allows PCA to both reduce the number of dimensions and minimize the amount of information lost in the process [2]. For this study, PCA was used to reduce the number of dimensions from 22 to 20. Additionally, to prevent data leakage, PCA was performed based on data from the training set. The models will be tested both with and without PCA.

*2.7 Model Training and Evaluation*

The five different Gradient Boosting models were trained on the data with different combinations of Min-Max, Standardization, mRMR, and PCA. One group on data with MinMax scaling, one group on data with Min-Max scaling and mRMR, one group on data with Min-Max scaling and PCA, one group on data with Standardization, one group on data with Standardization and mRMR, and one group on data with Standardization and PCA. In total, 30 models were trained on the training set and then evaluated and compared on the cross validation set. The models were evaluated on the cross validation set using the metrics of accuracy, sensitivity, specificity, AUC score, and F1 score.

*1) Sensitivity and Specificity*: Sensitivity is the model's accuracy on positive examples (examples where the patient has PD). Specificity is the model's accuracy on negative examples (examples where the patient is healthy). These metrics are useful because they can specifically identify how a model does on a specific class of data. This is especially useful for this study because the dataset that is being utilized has a high degree of class imbalance in favor of positive examples.

*2) AUC Score*: The AUC score is the area under the Receiver Operating Characteristic (ROC) Curve. An ROC Curve is generated by varying the model's threshold and plotting the different false positive and true positive rates. The area under this curve works as a measure of how likely a model is to output a higher probability for a positive example than a negative example. For example, an AUC score of 0.7 would represent that if given a positive example and a negative example, the model will

output a higher probability for the positive example than the negative one 70% of the time.

*3) F1 Score*: A model's F1 score is the harmonic mean of the model's precision and recall. Precision is the likelihood that if the model predicts that a given example is positive that the example is actually positive. This metric is also known as the Positive Probability Value (PPV). Recall is the same as sensitivity (the model's accuracy on positive examples). The term recall is used in this context because recall is most commonly used when concerning F1 score. The equation for the F1 score is shown below:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{7}$$

For Equation (7), it is common to also add a value $\epsilon$ to the denominator. $\epsilon$ is often a very small value (such as $1e-100$) and serves to prevent dividing by zero. Equation (7) with the $\epsilon$ term included is shown below:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall + \epsilon} \tag{8}$$

## 3. Results

The five models were trained on the training set with different combinations of Min-Max scaling, Standardization, mRMR feature selection, and PCA. The models were then evaluated on the cross validation set based on their accuracy, sensitivity, specificity, AUC score, and F1 score. The results for each model on the cross validation set are shown in Tables 1-6.

TABLE I
GRADIENT BOOSTING MODELS ON VALIDATION SET (MIN-MAX)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 87.1795 | 90.3226 | 75.0000 | 0.826613 | 0.918033 |
| HistGradientBoosting | 92.3077 | 93.5484 | 87.5000 | 0.905242 | 0.950820 |
| GradientBoosting | 82.0513 | 87.0968 | 62.5000 | 0.747984 | 0.885246 |
| LightGBM | 92.3077 | 93.5484 | 87.5000 | 0.905242 | 0.950820 |
| CatBoost | 84.6154 | 87.0968 | 75.0000 | 0.810484 | 0.900000 |

TABLE II
GRADIENT BOOSTING MODELS ON VALIDATION SET (MIN-MAX + PCA)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 76.9231 | 83.8710 | 50.0000 | 0.669355 | 0.852459 |
| HistGradientBoosting | 89.7436 | 96.7742 | 62.5000 | 0.796371 | 0.937500 |
| GradientBoosting | 79.4872 | 83.8710 | 62.5000 | 0.731855 | 0.866667 |
| LightGBM | 87.1795 | 93.5484 | 62.5000 | 0.780242 | 0.920635 |
| CatBoost | 89.7436 | 100.000 | 50.0000 | 0.750000 | 0.939394 |

TABLE III
GRADIENT BOOSTING MODELS ON VALIDATION SET (MIN-MAX + MRMR)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 89.7436 | 90.3226 | 87.5000 | 0.889113 | 0.933333 |
| HistGradientBoosting | 94.8718 | 96.7742 | 87.5000 | 0.921371 | 0.967742 |
| GradientBoosting | 76.9231 | 80.6452 | 62.5000 | 0.715726 | 0.847458 |
| LightGBM | 92.3077 | 93.5484 | 87.5000 | 0.905242 | 0.950820 |
| CatBoost | 87.1795 | 90.3226 | 75.0000 | 0.826613 | 0.918033 |

TABLE IV
GRADIENT BOOSTING MODELS ON VALIDATION SET (STANDARDIZATION)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 87.1795 | 93.5484 | 62.5000 | 0.780242 | 0.920635 |
| HistGradientBoosting | 87.1795 | 93.5484 | 62.5000 | 0.780242 | 0.920635 |
| GradientBoosting | 84.6154 | 90.3226 | 62.5000 | 0.764113 | 0.903226 |
| LightGBM | 84.6154 | 93.5484 | 50.0000 | 0.717742 | 0.906250 |
| CatBoost | 82.0513 | 87.0968 | 62.5000 | 0.747984 | 0.885246 |

TABLE V
GRADIENT BOOSTING MODELS ON VALIDATION SET (STANDARDIZATION + PCA)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 84.6154 | 96.7742 | 37.5000 | 0.671371 | 0.909091 |
| HistGradientBoosting | 89.7436 | 100.0000 | 50.0000 | 0.750000 | 0.939394 |
| GradientBoosting | 87.1795 | 93.5484 | 62.5000 | 0.780242 | 0.920635 |
| LightGBM | 87.1795 | 100.0000 | 37.5000 | 0.687500 | 0.925373 |
| CatBoost | 89.7436 | 100.0000 | 50.0000 | 0.750000 | 0.939394 |

TABLE VI
GRADIENT BOOSTING MODELS ON VALIDATION SET (STANDARDIZATION + MRMR)

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| XGBoost | 84.6154 | 93.5484 | 50.0000 | 0.717742 | 0.906250 |
| HistGradientBoosting | 89.7436 | 96.7742 | 62.5000 | 0.796371 | 0.937500 |
| GradientBoosting | 84.6154 | 90.3226 | 62.5000 | 0.764113 | 0.903226 |
| LightGBM | 84.6154 | 90.3226 | 62.5000 | 0.764113 | 0.903226 |
| CatBoost | 84.6154 | 90.3226 | 62.5000 | 0.764113 | 0.903226 |

The HistGradientBoosting model with Min-Max and mRMR performed the best on the cross validation set because, as shown in Table 3, it obtained the highest accuracy, specificity, AUC score, and F1 score and obtained the second highest sensitivity. This model was then evaluated on the test set. Its performance on the test set is shown in Table 7 and Fig. 1.

TABLE VII
HISTGRADIENTBOOSTING (MIN-MAX + MRMR) MODEL ON TEST SET

| Model | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC Score | F1 Score |
|---|---|---|---|---|---|
| HistGradientBoosting (Min-Max + mRMR) | 89.7436 | 96.6667 | 66.6667 | 0.816667 | 0.935484 |

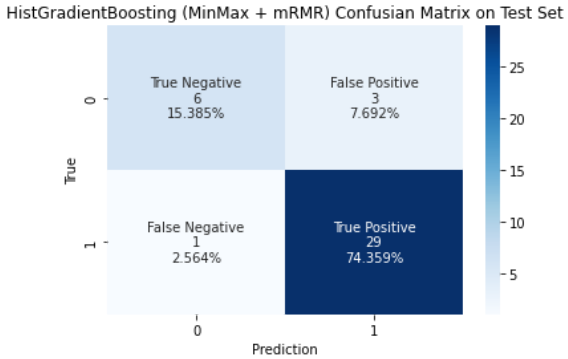


Fig 1. HistGradientBoosting (Min-Max + mRMR) Confusion Matrix on Test Set

## 4. DISCUSSION & CONCLUSION

PD currently affects roughly 1 million people in the U.S. alone, with 60,000 U.S. citizens being positively diagnosed for the disease annually [9]. With this statistic only expected to rise in the future, it is becoming increasingly important to diagnose PD in its early stages. Prolonged diagnosis delays have proven to be catastrophic in the livelihood of families and patients due to a lack of proper medication and attention. While alternatives for diagnosis such as dopamine screening and symptom checklists exist, they require medical professionals and extensive equipment to allow for proper execution; not only is this not accessible to many populations, but it can also be extremely expensive. Furthermore, many of the tested symptoms of PD also overlap with the known symptoms for other diseases. Creating a viable and accurate solution for the rapid diagnosis of PD is an essential asset in the race to stem disease progression. Voice analysis, being a PD specific symptom, can easily be scaled to meet the needs of analysis due to its prevalence in positively diagnosed patients. A machine learning algorithm to detect discrepancies in patient voices for diagnosis tackles the issues of both accessibility and cost by creating a readily available software solution. Voice is also unique with regards to PD, and so can be used as a relatively accurate metric for diagnosis.

To improve the accuracy and efficiency of speech-based PD diagnosis, this study aims to apply Gradient Boosting to classify a patient as either having PD or being healthy based on various biomedical voice measurement features. After training on the biomedical voice measurements from 117 voice recordings, the best performing Gradient Boosting method was found to be HistGradientBoosting with Min-Max feature scaling and mRMR feature selection. On the cross validation set, this model achieved an accuracy of 94.8718%, a sensitivity of 96.7742%, a specificity of 87.5000%, an AUC score of 0.921371, and an F1 score of 0.967742. When tested on the test set, this model was found to have an accuracy of 89.7436%, a sensitivity of 96.6667%, a specificity of 66.6667%, an AUC score of 0.816667, and an F1 score of 0.935484. This relatively high performance on a limited amount of data highlights Gradient Boosting's high applicability to speech-based PD diagnosis and usefulness in clinical practice.

Gradient Boosting often produces models that take little memory and are able to both run and train very quickly. This would further increase their accessibility, as they wouldn't require intensive hardware to run. Additionally, since they only require an audio sample from a user, they may also be applicable to smart phone applications so that users may obtain diagnoses from their home. The performance of Gradient Boosting on classifying PD also indicates that it may also have applications in diagnosing other neurological diseases, such as Alzheimer's Disease, from audio samples. Based on the findings of this study, Gradient Boosting has the potential to provide accessible and efficient diagnosis for PD and possibly many other neurological diseases.

There are a variety of other methods that may improve on these results and that weren't implemented in this study. This study had to work with a limited number of voice recordings (195) that were taken from a small range of patients (31). Due to the differences in speech based on language and accent, voice samples from a large and diverse number of subjects would help make a more universal model. Other than collecting more data from more participants, which may be time consuming and costly, it may also be possible to increase the diversity in the dataset by utilizing machine learning algorithms that convert speech samples to different accents. Additionally, since this study found success by using mRMR feature selection, other feature selection methods such as Boruta and Fisher's Score may be worth testing to see how they may perform differently than mRMR. Finally, other boosting algorithms such as AdaBoost may be worth testing on this problem based on the performance of Gradient Boosting.

Based on Gradient Boosting's relatively impressive results on a small dataset of 195 voice recordings, Gradient Boosting is a promising method for providing accessible and efficient PD diagnosis throughout the world, especially following further research.

## 5. References


[1] *(pdf) Gradient Boosting Machines, a tutorial - researchgate*. (n.d.). Retrieved February 16, 2022, from https://www.researchgate.net/publication/259653472_Gradient_Boosting_Machines_A_Tutorial

[2] Gewers, F. L., Ferreira, G. R., de Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. da F. (2018, June 19). *Principal component analysis: A natural approach to data exploration*. arXiv.org. Retrieved February 16, 2022, from https://arxiv.org/abs/1804.02502

[3] *How parkinson's disease is diagnosed*. Johns Hopkins Medicine. (n.d.). Retrieved February 16, 2022, from https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/how-parkinson-disease-is-diagnosed

[4] Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W.,

Ma, W., Ye, Q., & Liu, T.-Y. (2019, August 6). *LightGBM: A highly efficient gradient boosting decision tree*. Microsoft Research. Retrieved February 16, 2022, from https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/

[5] Niccolini F, Su P, Politis M. (2014, December). *Dopamine receptor mapping with PET imaging in parkinson's disease*. Journal of neurology. Retrieved February 16, 2022, from https://pubmed.ncbi.nlm.nih.gov/24627109/

[6] Mayo Foundation for Medical Education and Research. (2022, January 14). *Parkinson's disease*. Mayo Clinic. Retrieved February 16, 2022, from https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/diagnosis-treatment/drc-20376062

[7] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019, January 20). *CatBoost: Unbiased boosting with categorical features*. arXiv.org. Retrieved February 16, 2022, from https://arxiv.org/abs/1706.09516

[8] Spine, M. B. &. (n.d.). *Parkinson's disease*. Parkinson's Disease (PD) Mayfield Brain & Spine Cincinnati, Ohio. Retrieved February 16, 2022, from https://mayfieldclinic.com/pe-pd.htm

[9] *Statistics*. Parkinson's Foundation. (n.d.). Retrieved February 16, 2022, from https://www.parkinson.org/Understanding-Parkinsons/Statistics#:~:text=Nearly%20one%20million%20people%20in,to%201.2%20million%20by%202030.

[10] UCI Machine Learning Repository: Parkinsons data set. (n.d.). Retrieved February 16, 2022, from https://archive.ics.uci.edu/ml/datasets/parkinsons

[11] U.S. Department of Health and Human Services. (n.d.). *Parkinson's disease*. National Institute on Aging. Retrieved February 16, 2022, from https://www.nia.nih.gov/health/parkinsons-disease

[12] *XGBoost: A scalable tree boosting system - researchgate*. (n.d.). Retrieved February 16, 2022, from https://www.researchgate.net/publication/31

0824798_XGBoost_A_Scalable_Tree_Boosting_System

[13] Zhao, Z., Anand, R., & Wang, M. (2019, August 15). *Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform*. arXiv.org. Retrieved February 16, 2022, from https://arxiv.org/abs/1908.05376