Deep Learning based prediction and monitoring of Parkinson’s Disease using Voice Data

line 1: 1st Given Name Surname   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 4th Given Name Surname  
line 2: *dept. name of organization*  
*(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 2nd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 5th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCIDline 1: 3rd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 6th Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—With recent progress in digitized data acquisition, machine learning and computing infrastructures, the application of AI has stretched its wings far and wide. One such field of application is medical science. Artificial intelligence is gradually changing traditional medical practices and approaches to various problems. Even when medical science is growing with leaps and bounds there are still some problems which are beyond the province of human expertise. The infamous name of Parkinson’s disease haunts the human race even in this era of rapid scientific advancements since it has no cure to date. Thus early detection of the disease is very necessary to be better safe than sorry. PD is a progressive neurological disorder which affects the Central Nervous System (CNS) preventing the proper functioning of the same. This results in tremors, stiffness, slow movements, difficulty in balance and coordination, speech disorders and many more. In this study, we have gathered data on the speech of affected and healthy people. We have analysed the data using acoustic features such as jitter, shimmer, intensity, pitch, etc. The following paper consists of a detailed study of the prediction of Parkinson’s disease using Residual Neural Network which works best in this case. To prove this very point we have experimented with different ML models namely KNN, SVC, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, XGBoost and Gradient Boosting along with modern neural network techniques like ANN and MLP and compared the results. Using ResNet will help us in the early and accurate detection of this deadly disease.

Keywords—Artificial Intelligence, Parkinson’s Disease, Machine Learning models, Neural Networks, Comparative analysis, Dimensionality reduction.

# Introduction

Parkinson’s disease is a chronic and progressive movement disorder that affects the human brain and central nervous system.

It is caused by the degeneration of dopamine-releasing neurons in the brain. Dopamine is a key neurotransmitter responsible for controlling our motor movements. With the decrease in dopamine the person gradually loses motor abilities resulting in tremors, inability to control movements etc.

To this day the cause of Parkinson’s disease is not fully understood and hence no complete cure is also available.

The disease usually progresses slowly and hence the physical symptoms are usually not visible until the patient is already in the advanced stages of the disease.

This is where modern medical science with the help of machine learning can intervene, with early and automatic detection of Parkinson’s disease, patients can undergo different supportive treatments that can reduce the symptoms and slow the progression of the disease.

In this article, we have implemented eight (08) traditional machine learning and three (03) neural network models namely, KNN Classifier, SVM Classifier, Decision Tree Classifier, Random Forest Classifier, Naive Bayes Classifier, Logistic Regression Classifier, XGBoost Classifier, Gradient Boost Classifier, Artificial Neural Network, Multi-Layer Perceptron and ResNet to classify whether a patient is suffering from Parkinson’s disease or not based on their speech patterns and have done a thorough performance analysis of all the models implemented by testing each of them using a dataset.

# Related Work

Agarwal et al. [1] proposed an efficient approach to implement Extreme Machine Learning on a reliable dataset of speech samples of Parkinson's patients sourced from the UCI repository. They were able to achieve an accuracy of 90.76% and 0.81 MCC in distinguishing between Parkinson’s positive and Parkinson’s negative patients. Their work is mainly focused on Neural Networks and Support Vector Machines.

Ouhmida et al. [2] deployed Convolution Neural Networks (CNN) and Artificial Neural Networks (ANN) for the classification of healthy patients from Parkinson’s Disease (PD) positive patients on two datasets from UCI Machine Learning repository databases. The datasets were denoted by database I and database II consisting of 22 and 45 acoustic features respectively. CNN model achieved the highest accuracy of 93.10% on database I.

Ogawa and Yang [3] worked on the early detection of Parkinson’s Disease using 10-layered 1-d Convolution Neural Networks (CNN) and novel Residual Network (ResNet) type 1-d CNN, on a dataset consisting of the vocal features of healthy and PD-positive patients. They were able to achieve an accuracy of 0.888, an F-measure of 0.928 and an MCC of 0.692 in classifying.

Aghzal and Mourhir [4] combined a Histogram of Oriented Gradients with Convolution Neural Networks (CNN) to automate the detection process of Parkinson’s Disease based on the handwriting patterns of both positive and negative patients. Their model was able to achieve an accuracy of 87% and an F1-Score of 83.21%, outperforming the then-present clinical diagnostic techniques.

Anand et al. [5] deployed state-of-the-art machine learning and deep learning models equipped with varying dimensional reduction (DR) techniques to boost the efficiency, precision, recall and F1-Score of the models on the Parkinson’s Speech Dataset gathered from the UCI Machine Learning Repository. A comparative analysis was performed among the implemented models to come up with a conclusion of the best working model.

# Data Set

We have used a Parkinson’s Speech Pattern Dataset to showcase our work. The dataset we have used has 756 data points, with each data point having 754 attributes. The dataset has 564 Parkinson’s positive data points and 192 Parkinson’s negative data points.

Although the dataset is relatively small, the relatively high amount of attributes although good for training our models made our algorithm susceptible to overfitting, we have taken care of that by using Principal Component analysis to reduce the number of attributes without losing the training ability. With this, we have been able to train the traditional machine learning algorithms and especially the neural network algorithms to a very high level of precision and accuracy. However, we are certain that with the inclusion of more data points, we would be able to improve our models further.

# Methodology

Data pre-processing [6] is a very important step in machine learning. The goal of data preprocessing is to prepare and clean the raw dataset so that the efficiency and accuracy of the machine learning algorithms can be maximized.

Similarly, before processing our dataset through the different machine learning algorithms we had to prepare our dataset for optimal use.

The Processes we had to go through to prepare our dataset were:

1. *Null Value Replacement: [7] Real World datasets can have missing data which needs to be replaced by the mean value of the missing attribute. We checked our dataset for missing values and as the dataset did not have any missing values we did not have to do anything.*
2. *Skewness Reduction: [8] Skewness is the measure of the asymmetry of the probability distribution of an attribute. Excessive skewness can lead to bias in the final model. For our dataset, we first check each attribute for its skewness and for any attribute which has an absolute skewness value of greater than 1 we append that attribute column to a list named skewedCols. Below is the list of 5 attributes each which has the highest and lowest skewness before skewness reduction:*

TABLE I.

|  |  |  |
| --- | --- | --- |
| **Attribute Name** | **Skew Value** | **Skew Value** |
| tqwt\_TKEO\_mean\_dec\_32 | 0.19846778387365632 | 26.48258509147365 |
| tqwt\_TKEO\_std\_dec\_32 | 0.01987384966738003 | 26.0620838755717 |
| tqwt\_TKEO\_mean\_dec\_33 | 0.3104628329189008 | 24.9442663610479 |
| tqwt\_TKEO\_std\_dec\_33 | 0.07507677689667193 | 24.283816068288253 |
| det\_TKEO\_mean\_3\_coef | 1.4434955464066024 | 20.874872200923843 |
| det\_LT\_entropy\_shannon\_7\_coef | -4.780452134423335 | -21.4150979727166 |
| tqwt\_medianValue\_dec\_29 | -0.31596805061181005 | -21.623643949849406 |
| tqwt\_skewnessValue\_dec\_24 | -1.690083153162617 | -22.684339987002808 |
| tqwt\_entropy\_shannon\_dec\_33 | -2.023628408215805 | -25.06135227711703 |
| tqwt\_entropy\_shannon\_dec\_32 | -2.120342043665047 | -25.672811274750888 |

Fig. 1. Table showing skewness of 5 most skewed attributes

*Then we further classify the skewedCols based on whether that attribute contains any positive, zero or negative values into three separate lists named skewedCols\_PositiveVals, skewedCols\_ZeroVals, and SkewedCols\_NegativeVals respectively.*

*Then for the attributes present in the list skewedCols\_PositiveVals we first use Box-Cox Transformation to reduce the skewness of the attributes.*

*Box-Cox transformation [12] works by applying a power function to the dependent variable which allows it to be transformed into a normal distribution and reduce its skewness.*

*We also use cube root transformation to reduce the skewness of the attributes which are present in the lists skewedCols\_ZeroVals, SkewedCols\_NegativeVals. Cube Root transformation works by taking the cube root of each value of the attribute and making the attribute more closely resemble a normal distribution and reduce its skewness. Below is the final skewness of those attributes:*

1. *Kurtosis Reduction: [9] Kurtosis of an attribute is the measure of the peakedness/flatness of the probability distribution when measured relative to the normal distribution. The kurtosis of the normal distribution is considered zero. High kurtosis can lead to biases in the final model and hence we need to reduce kurtosis before processing our data. For our dataset, we check for attributes which have a kurtosis value greater than 3. For our dataset, we did not have any such attribute and hence we did not need to do any kurtosis reduction.*
2. *Outlier Detection: [10] Outliers are data points in an attribute which are significantly different from the rest of the data. Outliers are bad for machine learning as they can bias the results of the final model. For our model, We check each attribute and look for data points which are lower than the 25 percentile or higher than the 75 percentile of that attribute any data point which falls outside of this criteria is replaced by the mean of that attribute.*
3. *PCA: [11] For our dataset we loop through different number of principal components and finally got 148 principal components which was giving us the optimal results.*

Supervised Machine Learning models used in the process are listed as follows:

### K-Nearest Neighbors: KNN predicts the values of new entries using values of the k nearest data points.

### Support vector Classifier: SVC creates a separating hyperplane with the help of support vectors.

### Decision tree Classifier: DTC classifies the data by dividing the dataset into various segments based on certain decisions.

### Random Forest Classifier: RFC uses an ensemble of DTC and uses the mode of the predictions to

### K-Nearest Neighbors: KNN predicts the values of new entries using values of the k nearest data points.

### Support vector Classifier: SVC creates a separating hyperplane with the help of support vectors.

### Decision tree Classifier: DTC classifies the data by dividing the dataset into various segments based on certain decisions.

### Random Forest Classifier: RFC uses an ensemble of DTC and uses the mode of the predictions to

### K-Nearest Neighbors: KNN predicts the values of new entries using values of the k nearest data points.

### Support vector Classifier: SVC creates a separating hyperplane with the help of support vectors.

### Decision tree Classifier: DTC classifies the data by dividing the dataset into various segments based on certain decisions.

# Comparative Study

After preprocessing our dataset, we trained each of our machine learning and neural network models using the dataset. We split the dataset into training and testing parts and after developing the models using the training split, we carefully tested each model using the test split. We have accumulated the following parameters to evaluate the performance of the different algorithms against our dataset.

* **Confusion Matrix**

TABLE II.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name Of Model** | **True Positive** | **False Positive** | **False Negative** | **True Negative** |
| KNN | 14 | 3 | 1 | 58 |
| SVC | 11 | 6 | 1 | 58 |
| DTC | 12 | 5 | 5 | 54 |
| RFC | 10 | 7 | 1 | 58 |
| NBC | 11 | 6 | 3 | 56 |
| LR | 12 | 5 | 5 | 54 |
| XGBC | 15 | 2 | 5 | 54 |
| GBC | 12 | 5 | 4 | 55 |
| ANN | 165 | 18 | 10 | 525 |
| MLP | 14 | 3 | 1 | 54 |
| ResNet | 1810 | 110 | 110 | 5530 |

Fig. 5. Table for confusion matrix of ML models

* **Accuracy -** We have plotted the accuracies of the different models for the training split in the table above. From the above table, it is evident that the Residual Neural Network and the ANN Classifier have the highest accuracy in classifying whether a patient has Parkinson’s disease or not based on their attributes.
* Error - Error is the difference between the true result and the result predicted by the model. We have plotted the different error graphs for the different models.

From the above graphs, it is evident the models are able to learn to a certain accuracy where the loss or error is minimum.

* **Recall -** We have accrued the recall values of the different models for the training split in the table above. Recall measures the proportion of actual positive cases that are correctly identified by the model as positive. It is a very important metric for our use case as a high recall signifies that the model does not miss many positive cases as negative. From the above table, it is evident that Residual Neural Network and the ANN Classifier are the best-performing models in terms of recall.
* **Precision -** We have accrued the precision values of the different models for the training split in the table above. Precision is the measure the fraction of correctly predicted positive instances out of all the instances the model has predicted as positive. It is also a very important metric for our use case as a high precision value implies that the model does not misclassify too many positive patients as negative and vice versa. From the above table, it is also evident that the Residual Neural Network is the optimal model in terms of precision.
* **F1-Score -** We have accrued the F1-scores of the different models for the training split in the table above.

F1 - Score is calculated by combining both the Precision and recall values of the model and hence is a important metric to consider for our use case as a high F1 - Score implies that the model is making accurate predictions and also minimizes false negatives reducing the risk of miss-diagnosis. From the above table it is evident that Residual Neural Network is the optimal model as it has the highest F1 - Score for both the classes.

* **AUC-ROC Curve -** The AUC-ROC curve is made by plotting the model’s true positive rate against its false positive rate. A high AUC value in the ROC curve (Close to 1) implies that the model can precisely distinguish between Parkinson’s Positive and Parkinson’s Negative patients.

From the above graphs, it is evident that the Neural Networks are the optimal model as they have the highest AUC-ROC.

* AUC-PR Curve - The AUC-PR curve is made by plotting the model’s precision against recall of the model. In medical use cases it is a particularly important metric to consider as it can provide a more accurate evaluation of the performance model when the class distribution is imbalanced. The same is true for our dataset as well.

From the above graphs it is evident that the Neural Networks are the optimal model as they have the highest AUC-PR.

# Conclusion ans scope for future work

The research article focuses on a comparative study between Traditional Machine Learning Algorithms and Neural Networks. From the study, it is evident that both Models have their own positive and negative aspects. The Traditional Model focuses on a simplistic approach to automate the learning process of machines and find hidden and unseen patterns. The Neural Networks, on the other hand, try to mimic the working of the human brain, by setting up processing nodes called neurons in a systematic plan and each neuron is interconnected with each others giving rise to a fully connected network.

Traditional ML models are very useful for small to medium-sized datasets, whereas, neural networks are well known for their ability to handle complicated and high-dimensional datasets. Due to the simplicity of traditional ML models, their computing time is often faster than neural networks. Thus, Neural Networks are much more resource hungry and complex to implement.

The factors responsible for choosing between implementing the traditional ML models or the Neural Networks depend upon the type and quality of the dataset, the computing resources and the budget.

As the field of Machine Learning and Artificial Intelligence continues to evolve at an exponential rate, both the Traditional Machine Learning Models and the Neural Networks will continue to upgrade themselves in terms of accuracy, computation time and resource needs. It is important to continue to explore and develop new approaches that combine the strengths of these models and address their limitations, to improve accuracy, interpretability, and scalability.

From our comparative analysis of the selected dataset, we conclude, Residual Neural Network (ResNet) has worked exceptionally better than all the traditional Machine Learning Models with an accuracy of 99%, a precision of 0.985, a recall of 0.99 and an F1-Score of 0.985. Standing next to the ResNet is ANN with an accuracy of 98%, a precision of 0.98, a recall of 0.97 and an F1-Score of 0.975. Thus, we can conclude, for our chosen dataset, the Neural Networks have performed exceptionally better in terms of the evaluation metrics. With the advent of more precise attributes in the future, the accuracy of the ML Models and Neural Networks can be boosted to a great extent using Feature Selection, Dimension Reduction and all the other effective and essential Data Pre-processing Techniques.

# Reference Work

1. A. Agarwal, S. Chandrayan and S. S. Sahu, "Prediction of Parkinson's disease using speech signal with Extreme Learning Machine," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, India, 2016, pp. 3776-3779, doi: 10.1109/ICEEOT.2016.7755419.
2. A. Ouhmida, O. Terrada, A. Raihani, B. Cherradi and S. Hamida, "Voice-Based Deep Learning Medical Diagnosis System for Parkinson's Disease Prediction," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, 2021, pp. 1-5, doi: 10.1109/ICOTEN52080.2021.9493456.
3. M. Ogawa and Y. Yang, "Residual-Network -Based Deep Learning for Parkinson's Disease Classification using Vocal Datasets," 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech), Nara, Japan, 2021, pp. 275-277, doi: 10.1109/LifeTech52111.2021.9391925.
4. M. Aghzal and A. Mourhir, "Early Diagnosis of Parkinson’s Disease based on Handwritten Patterns using Deep Learning," 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 2020, pp. 1-6, doi: 10.1109/ICDS50568.2020.9268738
5. A. Anand, M. A. Haque, J. S. R. Alex and N. Venkatesan, "Evaluation of Machine learning and Deep learning algorithms combined with dimentionality reduction techniques for classification of Parkinson’s Disease," 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Louisville, KY, USA, 2018, pp. 342-347, doi: 10.1109/ISSPIT.2018.8642776.
6. Sen, Sohom; GHOSH, ANKIT (2022): Analysis and Prediction of Parkinson's Disease using Machine Learning Algorithms. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.20005703.v1>
7. G. A. Mirza, "Null Value Conflict: Formal Definition and Resolution," 2015 13th International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 2015, pp. 132-137, doi: 10.1109/FIT.2015.32.
8. P. L. Kevin Ding, S. Martin and B. Li, "Improving Batch Normalization with Skewness Reduction for Deep Neural Networks," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 7165-7172, doi: 10.1109/ICPR48806.2021.9412949.
9. T. Thaiupathump and R. Chompu-inwai, "Impact of kurtosis on performance of mixture control chart patterns recognition using Independent Component Analysis and neural networks," 2015 4th International Conference on Advanced Logistics and Transport (ICALT), Valenciennes, France, 2015, pp. 94-99, doi: 10.1109/ICAdLT.2015.7136600.
10. H. C. Mandhare and S. R. Idate, "A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2017, pp. 931-935, doi: 10.1109/ICCONS.2017.8250601.
11. K. Li and S. Warren, "Principle component analysis on photoplethysmograms: Blood oxygen saturation estimation and signal segmentation," 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 2011, pp. 7171-7174, doi: 10.1109/IEMBS.2011.6091812.
12. Yue-Shiang Liu, S. F. Su, C. -C. Chuang and J. -T. Jeng, "Box-Cox transformation-based annealing robust radial basis function networks for skewness noises," 2010 International Conference on System Science and Engineering, Taipei, Taiwan, 2010, pp. 537-541, doi: 10.1109/ICSSE.2010.5551736.