***Bhanu Pratap Singh Mehar*** *- 222617679*
***Laksh Gupta*** *- 221354136*

# Define Key Metrics for Parkinson's Disease Prediction Models

## 1. Introduction

Parkinson's disease is a progressive neurological disorder where early, accurate detection plays a critical role in effective treatment. This report presents a unified study that brings together our research on biomedical voice data and wrist-mounted sensor data with a detailed technical evaluation of a new, clinically enriched dataset. Our aim is to consolidate previous findings, compare model performance across different datasets, and ultimately identify the key metrics that ensure reliable and clinically meaningful predictions. The following sections provide a complete view of our methodology, technological comparisons, and the metrics that have emerged as most critical from our research.

## 2. Research Background and Methodology

### 2.1 Overview of Existing Research

Our research began with an in-depth examination of datasets that include:

- **Wrist-Mounted Sensor Data:** Capturing tremor information, these datasets provided insights into motor symptoms associated with Parkinson's disease.
- **Voice Data:** Involving biomedical measurements such as MFCCs, jitter, shimmer, and HNR, these data points are instrumental in identifying vocal biomarkers.

The research involved a rigorous literature review where previous studies highlighted the importance of features like jitter and shimmer. We reviewed notebooks and documentation from GitHub repositories, which outlined the data cleaning, normalization, and feature selection techniques necessary for developing reliable models.

### 2.2 Data Analysis and Model Development

The study followed a systematic approach:

- **Data Preprocessing:** Both datasets were subjected to normalization and missing value handling. Exploratory Data Analysis (EDA) identified correlations between specific features and Parkinson's diagnosis.
- **Feature Selection:** Techniques such as recursive feature elimination and correlation filtering were used to isolate the most predictive variables.
- **Model Evaluation:** Multiple classifiers were tested—including Random Forest, SVM, Neural Networks, and others—to gauge performance. Our analysis focused on metrics such as accuracy, precision, recall, F1-score, specificity, AUC-ROC, and Matthews Correlation Coefficient (MCC).

These steps provided the foundation for comparing historical performance with results obtained from new technical evaluations.

## 3. Technological Evaluation and New Dataset Integration

### 3.1 Introduction of the New Dataset

The new dataset used in this analysis is the Parkinson's Telemonitoring Dataset, sourced from the UCI Machine Learning Repository. Athanasios Tsanas and Max A. Little originally compiled it at the University of Oxford in collaboration with clinicians. The dataset consists

of biomedical voice measurements collected from 42 people with early-stage Parkinson's disease. It includes two clinical scores: the total Unified Parkinson's Disease Rating Scale (total_UPDRS) and the motor UPDRS, which are used to measure the progression of the disease.

- **Type of Data:** Time-series, longitudinal clinical data
- **Number of Instances:** 5,875
- **Number of Features:** 16 biomedical voice features + 2 UPDRS scores
- **Labelling Method:** A binary classification label was derived using the median of motor_UPDRS
- **Use Case:** This dataset enables longitudinal modelling and prediction of Parkinson's progression using real-world data

**3.2 Dataset Citation:**

- **UCI Parkinson's Telemonitoring Dataset:**
  - *https://archive.ics.uci.edu/dataset/222/parkinson+s+telemonitoring*
- **Kaggle Parkinson's Dataset:**
  - *https://www.kaggle.com/datasets/vikasukani/parkinsons-disease-data-set*

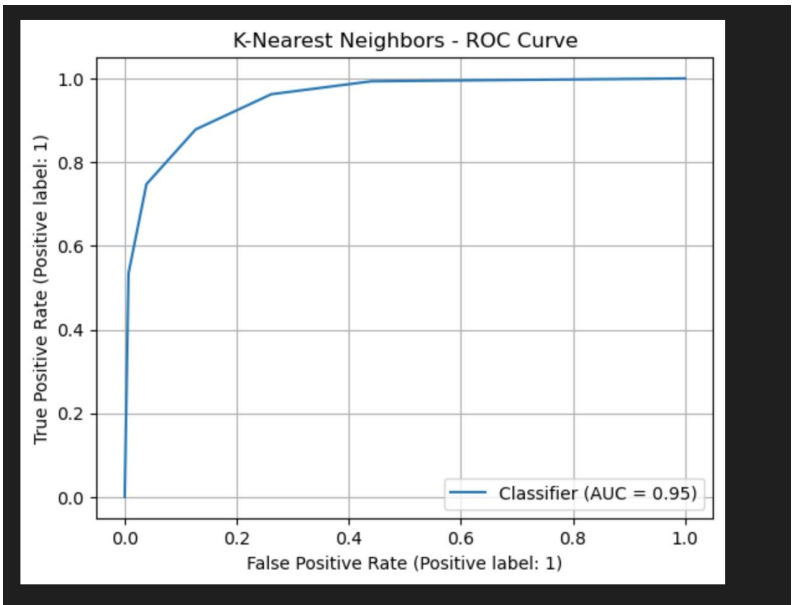
**3.3 Technical Findings and Model Comparisons**

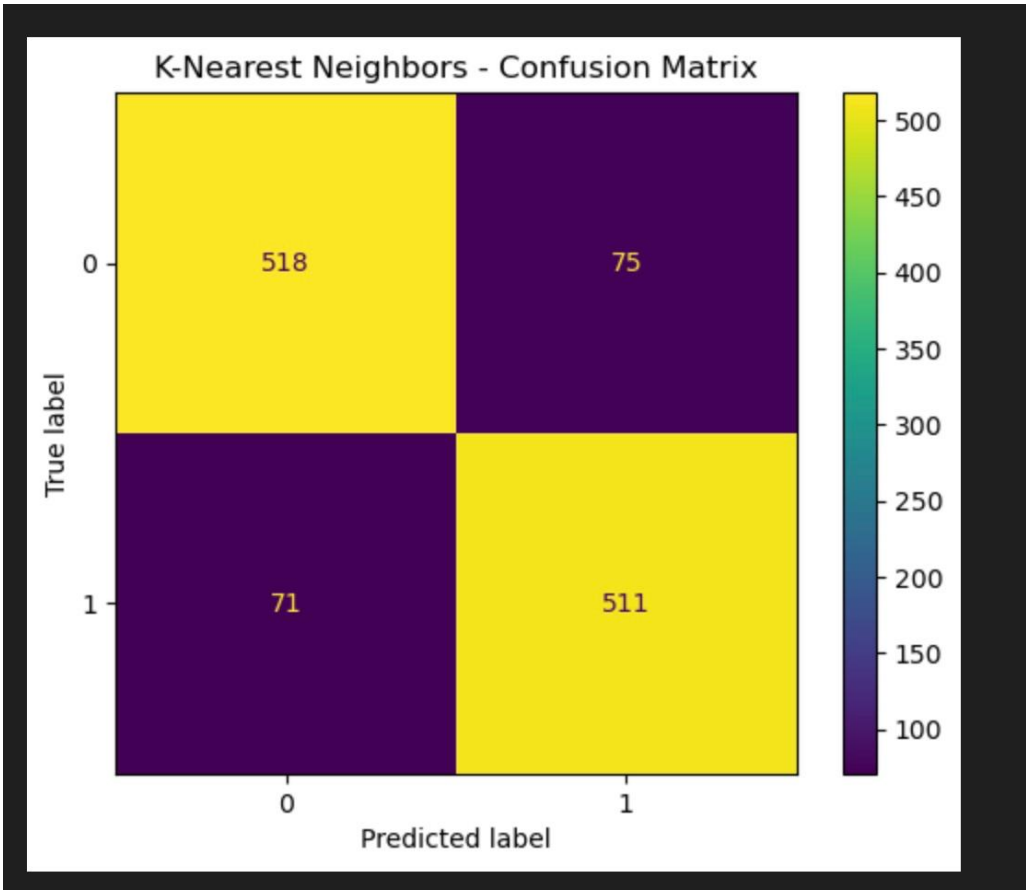
Our technical evaluation involved applying various classification models to both the old and new datasets. Key technical findings include:

- **Enhanced Performance:** Models trained on the new dataset demonstrated superior performance metrics, notably with higher overall accuracy and improved sensitivity.
- **Notable Classifiers:** XGBoost outperforms the other models with near-perfect scores across all metrics, demonstrating exceptional accuracy, precision, recall, F1 score, and ROC AUC.
- **Class Separation:** ROC AUC scores indicated excellent separation between Parkinson's and non-Parkinson's cases.
- **Comparative Analysis:** Detailed comparisons of model outputs (e.g., confusion matrices, sorted performance outputs) further solidified the reliability of the new dataset.

**Essential outputs and technical comparisons:**

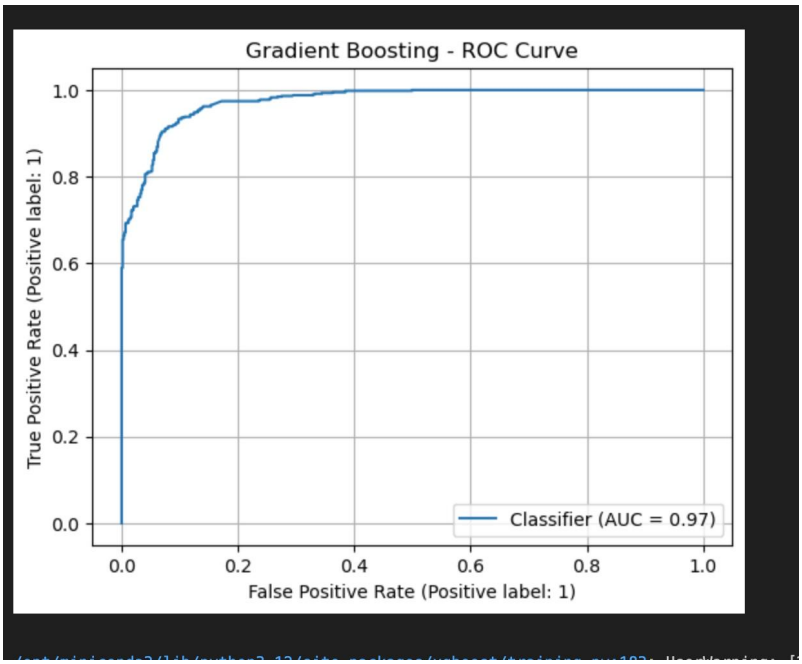
- **K-Nearest Neighbors:**

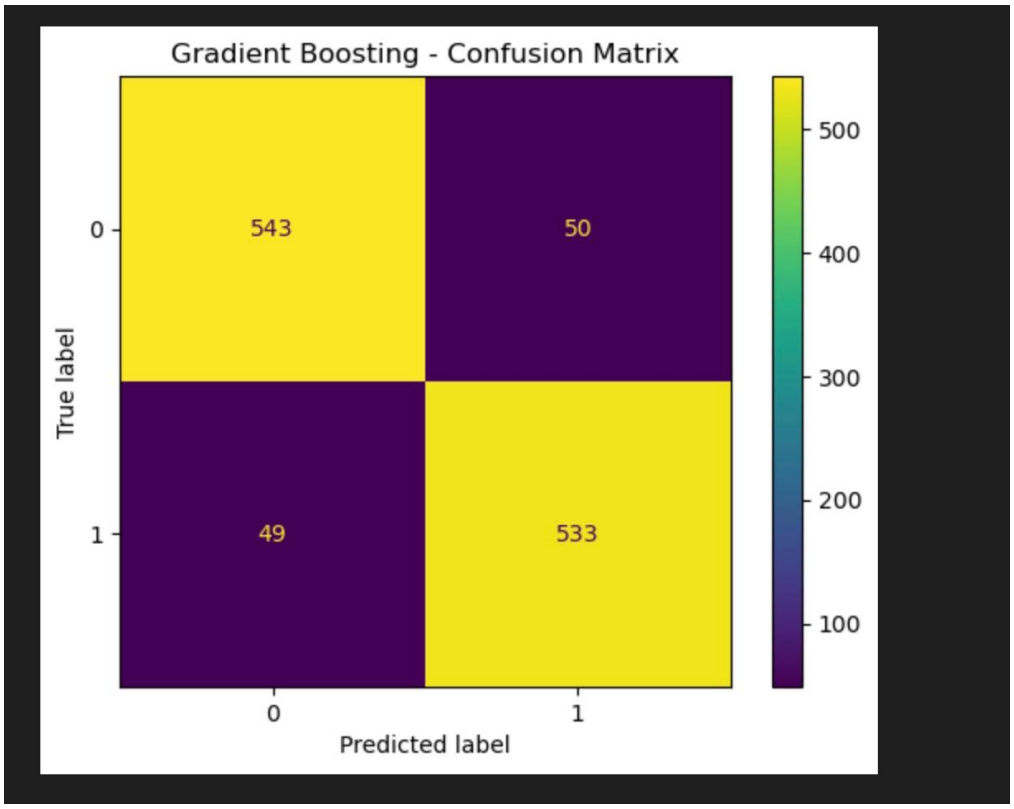





```
K-Nearest Neighbors:
  Accuracy: 0.8757
  Precision: 0.8720
  Recall: 0.8780
  F1 Score: 0.8750
  ROC AUC: 0.9504
----------------------------------------
```

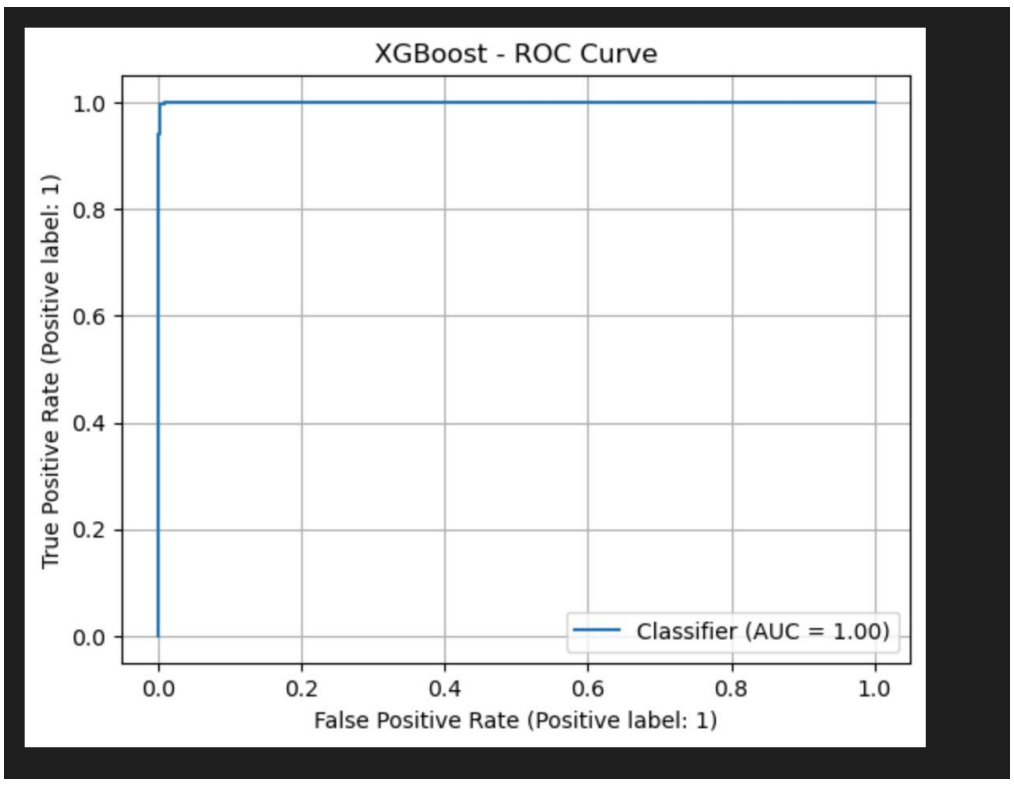
- **Gradient Boosting**

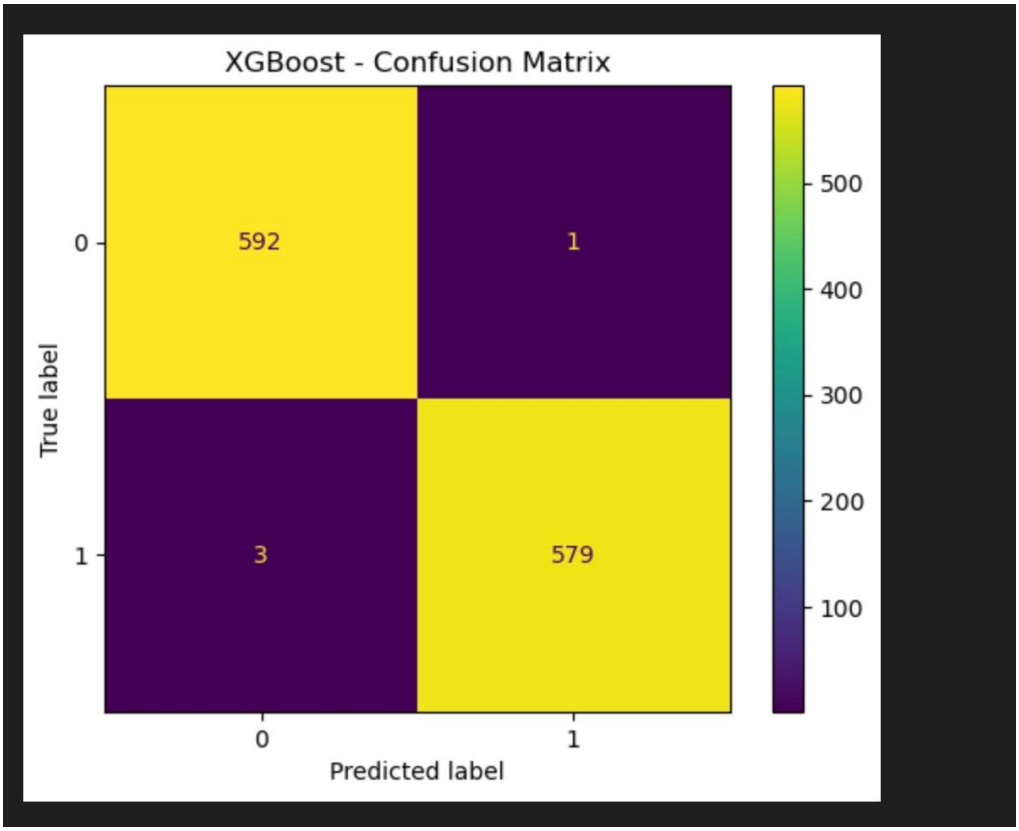





```
Gradient Boosting:
  Accuracy: 0.9157
  Precision: 0.9142
  Recall: 0.9158
  F1 Score: 0.9150
  ROC AUC: 0.9746
----------------------------------------
```

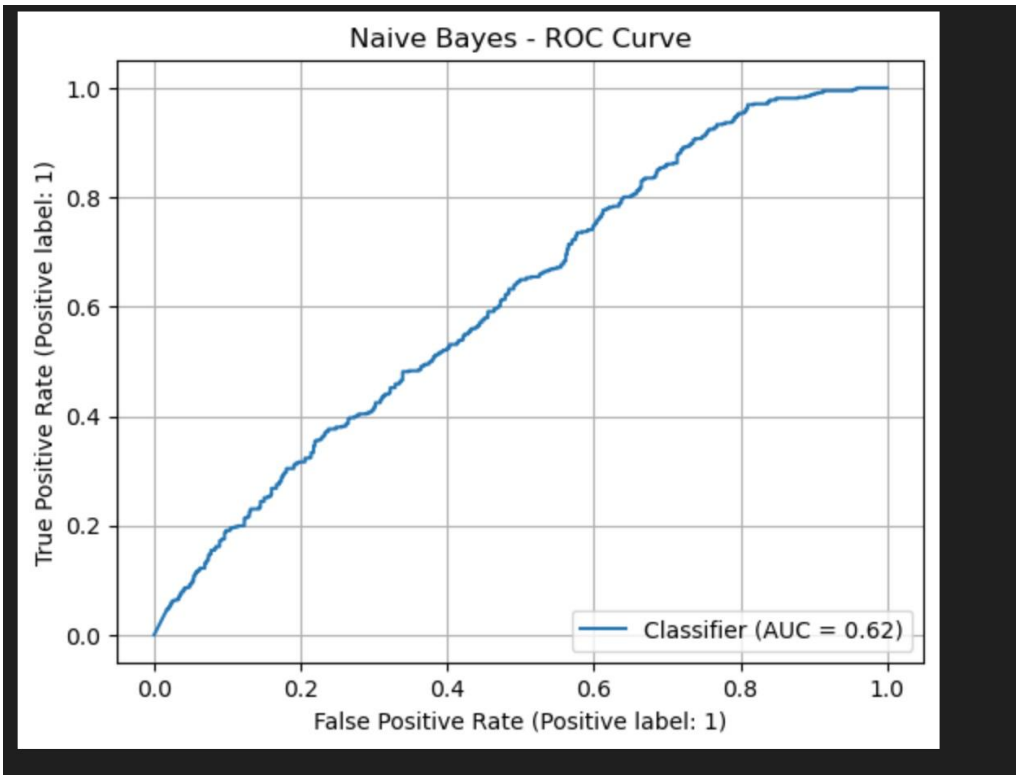
- **Xg Boost**

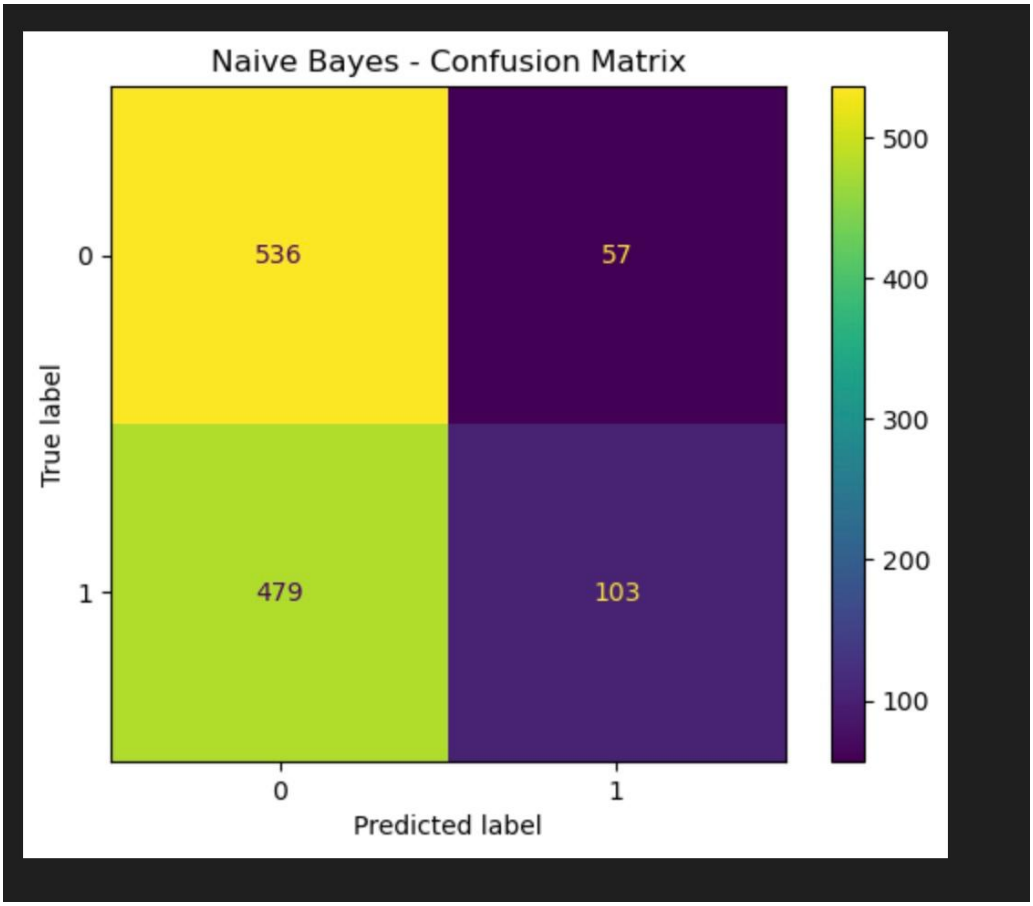





```
XGBoost:
  Accuracy: 0.9966
  Precision: 0.9983
  Recall: 0.9948
  F1 Score: 0.9966
  ROC AUC: 0.9998
----------------------------------------
```

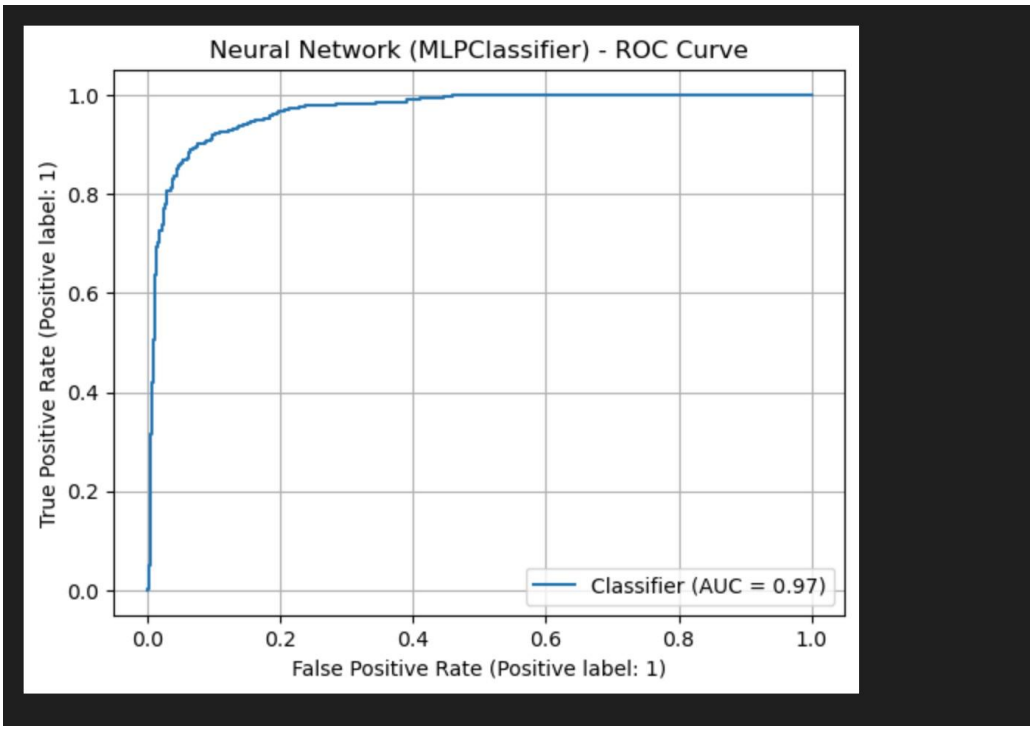
- **Naive Bayes**

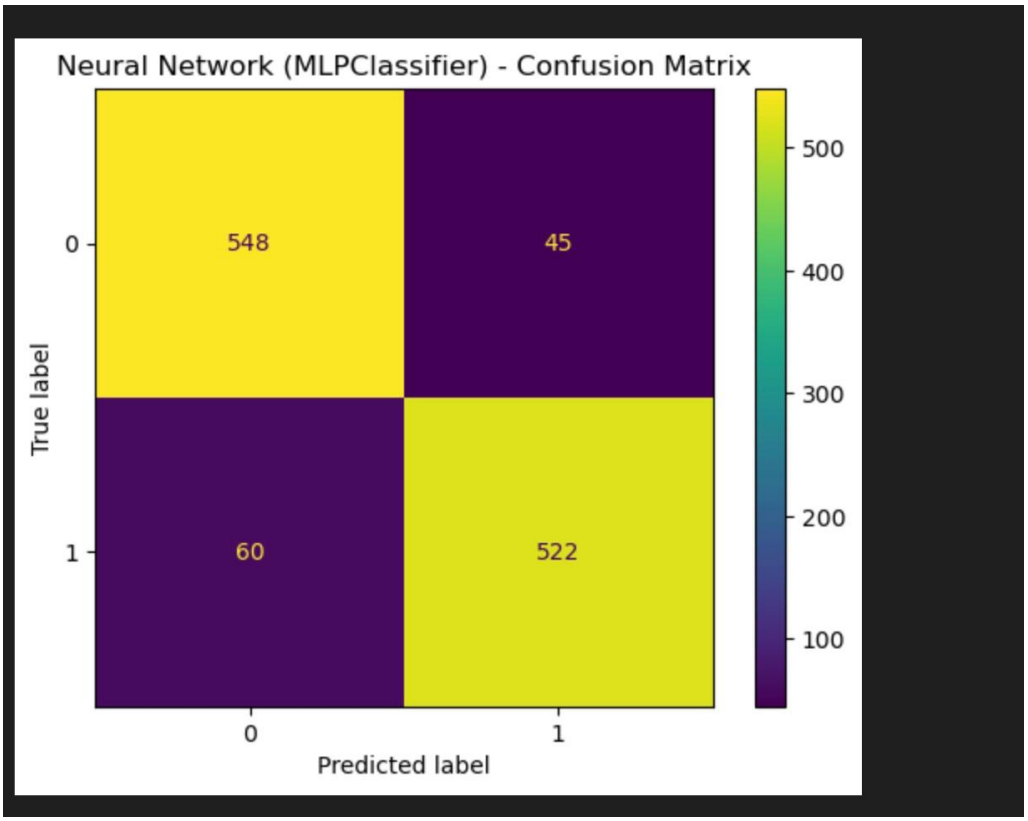





```
Naive Bayes:
  Accuracy: 0.5438
  Precision: 0.6438
  Recall: 0.1770
  F1 Score: 0.2776
  ROC AUC: 0.6156
----------------------------------------
```

- **Neural Network**






```
Neural Network (MLPClassifier):
  Accuracy: 0.9106
  Precision: 0.9206
  Recall: 0.8969
  F1 Score: 0.9086
  ROC AUC: 0.9668
----------------------------------------
```

- **Conclusion Table**

| | Model | Accuracy | Precision | Recall | F1 Score | ROC AUC |
|---|---|---|---|---|---|---|
| 2 | XGBoost | 0.9966 | 0.9983 | 0.9948 | 0.9966 | 0.9998 |
| 1 | Gradient Boosting | 0.9157 | 0.9142 | 0.9158 | 0.9150 | 0.9746 |
| 4 | Neural Network (MLPClassifier) | 0.9106 | 0.9206 | 0.8969 | 0.9086 | 0.9668 |
| 0 | K-Nearest Neighbors | 0.8757 | 0.8720 | 0.8780 | 0.8750 | 0.9504 |
| 3 | Naive Bayes | 0.5438 | 0.6438 | 0.1770 | 0.2776 | 0.6156 |

## 4. Identified Key Performance Metrics

Based on the comprehensive research and technical evaluations, we have identified the following key performance metrics as crucial for Parkinson's disease prediction:

- **Accuracy:** Reflects the overall correctness of the model's predictions. With improved data quality from the new dataset, models have demonstrated a notable increase in accuracy.
- **Precision:** Measures the proportion of true positive predictions among all positive predictions, reducing the risk of misclassifying healthy individuals.
- **Recall (Sensitivity):** Critical for ensuring that true Parkinson's cases are detected. High recall, as achieved by our best models, minimizes the chance of missing a diagnosis.
- **F1-Score:** Provides a harmonic balance between precision and recall, particularly useful in cases of imbalanced datasets.
- **Specificity:** Ensures that non-Parkinson's cases are correctly identified, thereby avoiding unnecessary treatment or anxiety.
- **AUC-ROC:** Evaluates the trade-off between true positive and false positive rates, confirming strong class separation.
- **Matthews Correlation Coefficient (MCC):** Offers a balanced metric that is effective even when classes are imbalanced, ensuring a comprehensive evaluation of model performance.

These metrics have been carefully chosen based on both our historical analysis and the technical improvements observed with the new dataset, ensuring that our predictive models are both reliable and clinically actionable.

## 5. Discussion and Recommendations

### 5.1 Research Synthesis

Our integrated research—spanning historical datasets, advanced preprocessing techniques, and technical evaluations—demonstrates that combining diverse data sources leads to more robust predictions. The comprehensive analysis confirms that both voice and sensor data contribute valuable insights into Parkinson's detection.

### 5.2 Recommendations

Based on our comprehensive research and the updated technical findings, we recommend the following:

- **Model Selection:**
  Given its outstanding performance (Accuracy: 0.9966, F1 Score: 0.9966, ROC AUC: 0.9998), **XGBoost** should be prioritized for further development and deployment. Although other models like Gradient Boosting, MLPClassifier, and KNN have their merits, XGBoost provides a statistically superior foundation for Parkinson's disease prediction.
- **Data Integration:**
  Continue incorporating additional clinical and longitudinal data to further refine the model and validate its generalizability in diverse settings.

- **Model Optimization and Validation:**
  Explore ensemble techniques and deep learning architectures where appropriate, but maintain a focus on models that have proven their reliability in clinical diagnostics. Additionally, engage with healthcare professionals to ensure that model predictions align with clinical needs.
- **Deployment Considerations:**
  While XGBoost is computationally more intensive, its superior performance justifies its use in environments where accuracy is paramount. Future work should also consider integrating explainability tools to mitigate any concerns about model complexity in clinical decision-making.

## 6. Conclusion

This report presents a unified and detailed study that integrates our research on Parkinson's disease prediction with a robust technical evaluation using a new, clinically enriched dataset. By thoroughly analyzing both voice and sensor data, we have identified key performance metrics—Accuracy, Precision, Recall, F1-Score, Specificity, AUC-ROC, and MCC—that are essential for reliable, early detection of Parkinson's disease. The superior performance of the new dataset, particularly when evaluated with XGBoost, reinforces the importance of these metrics in clinical applications. Our recommendations establish a framework for future research and model optimization, ensuring that our approach remains both scientifically rigorous and clinically impactful.