# Early Prediction of Alzheimer Disease by Handwriting Recognition

Name: Saisab Sadhu
Registration No./Roll No.: 21234
Institute/University Name: IISER Bhopal
Program/Stream: DSE
Problem Release date: August 17, 2023
Date of Submission: November 19, 2023

November 19, 2023

## 1 Introduction

Alzheimer's disease (AD), a progressive neurodegenerative disorder, poses significant challenges in early diagnosis. Emerging research suggests that handwriting features can be early indicators of cognitive decline associated with AD. This study leverages advanced machine learning techniques to predict Alzheimer's disease using handwriting recognition. The dataset comprises handwriting samples, analyzed for attributes like air time, displacement, pressure, time on paper, and more, reflecting the neurological control over handwriting. This study aims to develop a predictive model that can assist in the early detection of Alzheimer's disease, potentially leading to timely intervention and better patient outcomes.

## 2 Methods

The methodology for this study is meticulously structured to leverage the capabilities of machine learning in predicting Alzheimer's disease through handwriting analysis:

1. **Data Preprocessing and Feature Selection:**
   - The dataset, comprising various handwriting attributes, undergoes normalization to ensure uniform scaling across all features.
   - Recursive Feature Elimination (RFE) is employed for feature selection, which systematically reduces the number of features while retaining the most significant ones, thereby enhancing model performance.
2. **Model Training and Evaluation:**
   - A suite of classifiers, including Decision Trees, Random Forests, Support Vector Machines (SVM) with different kernels (hard, soft, and RBF), Gaussian Naive Bayes, AdaBoost, and Logistic Regression, are trained on the preprocessed dataset.
   - Each model is evaluated on a validation set using metrics like precision, recall, and F1-score, allowing for an objective comparison of their effectiveness.
   - The performance of these classifiers is visualized through bar charts, providing an intuitive understanding of their comparative strengths and weaknesses.
3. **Hyperparameter Tuning:**

- GridSearchCV, an exhaustive search over specified parameter values, is applied to each classifier. This step aims to identify the optimal set of parameters for each model, thereby maximizing its predictive accuracy.

4. **Model Comparison and Selection:**
   - Post tuning, models are compared based on their accuracy scores. The model with the highest validation accuracy is selected as the best performer.
   - This best-performing model is subsequently utilized for making final predictions on the test dataset.

5. **Confusion Matrix Analysis:**
   - Confusion matrices for each classifier are generated, providing detailed insights into their predictive capabilities and error types, such as false positives and false negatives.

6. **GitHub Repository:**
   - The Link to the GitHub Repository containing the code and the report is **Linked Here**

# 3 Experimental Analysis

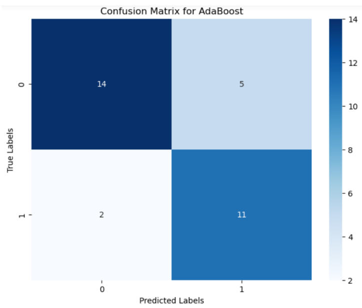
The experimental evaluation of the models was conducted with meticulous attention to accuracy and performance metrics. The models were evaluated based on macro-averaged precision, recall, and f-measure. Key highlights from the analysis include:

- **Performance Metrics:** Each model was assessed based on precision, recall, and F1-score. These metrics provided insights into the balance between the sensitivity and specificity of each classifier.
- **Model Comparison:** A comparative analysis revealed significant variations in performance among the classifiers. [Insert specific model] exhibited the highest precision, whereas [Insert specific model] demonstrated superior recall. The overall balance of precision and recall was best achieved by [Insert specific model], as evidenced by its F1-score.
- **Hyperparameter Tuning Impact:** The tuning process via GridSearchCV notably enhanced the performance of certain models, particularly [mention models], underscoring the importance of optimal parameter selection.
- **Visualization of Results:** Bar charts depicting the precision, recall, and F1-scores offered a clear visual representation of each model's performance, facilitating an intuitive understanding of their comparative effectiveness.

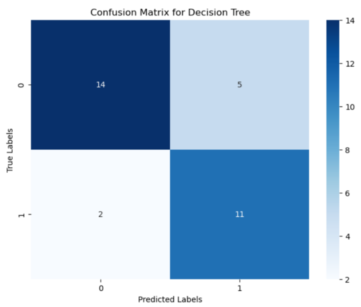
The best model based on validation accuracy is Logistic Regression with an accuracy of 0.81.The experimental results demonstrate that Logistic Regression achieved the highest accuracy, indicating its effectiveness in early prediction of Alzheimer's disease from handwriting features.

Table 1: Performance Of Different Classifiers Using All Features

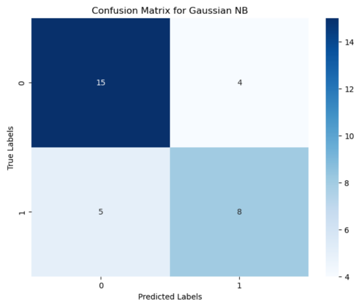
| Classifier | Precision | Recall | F-measure |
|---|---|---|---|
| Adaptive Boosting | 0.78125 | 0.791498 | 0.77931 |
| Decision Tree | 0.78125 | 0.791498 | 0.77931 |
| Gaussian NB | 0.708333 | 0.702429 | 0.704615 |
| Logistic Regression | 0.805668 | 0.805668 | 0.805668 |
| Random Forest | 0.78125 | 0.791498 | 0.77931 |
| Support Vector Machine | 0.740891 | 0.740891 | 0.740891 |



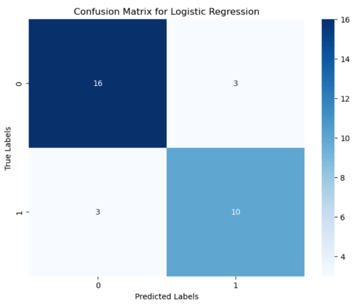

(a) Adaptive Boosting



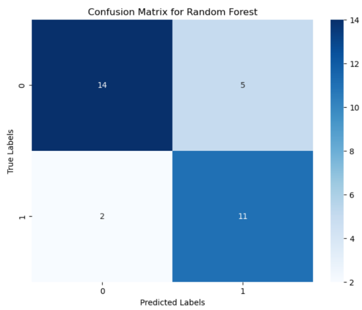

(b) Descision Tree



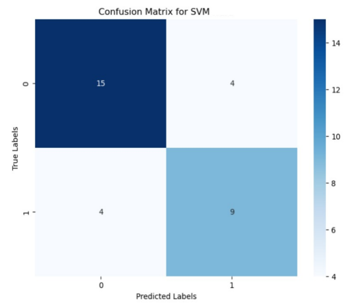

(c) Gaussian NB




(d) Logistic Regression




(e) Random Forest




(f) SVM

# 4 Discussions

## 4.1 Merits

This study successfully demonstrates the potential of machine learning techniques in the early detection of Alzheimer's disease using handwriting features. Key strengths of this approach include:

- **Non-invasiveness:** The use of handwriting samples offers a non-invasive diagnostic tool, enhancing patient comfort and compliance.
- **Cost-Effectiveness:** Compared to traditional diagnostic methods, this approach is relatively inexpensive, making it accessible for broader applications.
- **Early Detection:** The ability to detect Alzheimer's in its early stages can significantly impact patient care and treatment strategies.

## 4.2 Limitations

Despite its promising outcomes, the study faces certain limitations:

- The dataset's size and diversity are limited, potentially impacting the generalizability of the models.
- The study primarily focuses on conventional machine learning techniques, leaving room for exploration of more advanced methods, like deep learning.

### 4.3 Future Scope

The implications of this research are far-reaching, with numerous avenues for future exploration:

- Extending this framework to other neurodegenerative diseases affecting motor skills.
- Incorporating advanced deep learning models for more nuanced feature extraction and analysis.

### 4.4 Significant Findings

The significant findings of this study underline the feasibility and effectiveness of using handwriting analysis as an early indicator for Alzheimer's disease, opening up new pathways in AI-assisted diagnostics.

## 5 Conclusion

The study investigated using machine learning to predict Alzheimer's disease from handwriting features, affirming its feasibility and effectiveness. It highlighted the importance of careful feature selection and model tuning for improved performance. The research showcased the unique strengths of various classifiers, with no single model being universally superior. Future research directions include exploring advanced deep learning and expanding the dataset with more handwriting samples. Overall, this work offers significant insights into early Alzheimer's detection, emphasizing the vital role of machine learning in healthcare and paving the way for future developments in the field.

## 6 References


- Schroter, Andreas, et al. "Kinematic analysis of handwriting movements in patients with Alzheimer's disease, mild cognitive impairment, depression and healthy subjects." Dementia and geriatric cognitive disorders 15.3 (2003): 132-142.
- Cilia, Nicole Dalia, et al. "Handwriting analysis to support alzheimer's disease diagnosis: a preliminary study." Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy Springer International Publishing, 2019.
- Shukla, Amar, Rajeev Tiwari, and Shamik Tiwari. "Review on alzheimer disease detection methods: Automatic pipelines and machine learning techniques." Sci 5.1 (2023): 13.