

Paper Type: Original Article

Stacking-Based Machine Learning Approach for Alzheimer's Disease Diagnosis

Nariman A. Khalil¹  and **Mohamed Eassa**^{2,*} ¹ Assistant professor at Egyptian Chinese College for Applied Technology, Suez University, Egypt;
Narimankhalil.eccat@suez.edu.eg.² Department of Computer Science, Faculty of Information Systems and Computer Science, October 6 University, Giza, Egypt;
Mohamed.eassa.cs@o6u.edu.eg.**Received:** 09 Apr 2025**Revised:** 12 Apr 2025**Accepted:** 18 Nov 2025**Published:** 25 Nov 2025**Abstract**

This study investigates the use of machine learning models for diagnosing Alzheimer's disease, based on the Alzheimer's Disease Dataset from Kaggle. The dataset includes a variety of features that are crucial for identifying Alzheimer's in patients. We applied different models such as Random Forest, Support Vector Machine (SVM), XGBoost, and a Stacking Model to predict the disease's presence. After evaluating the performance of each model, the Stacking Model emerged as the most accurate, achieving an impressive accuracy rate of 96%. This result demonstrates the potential of combining multiple models to improve diagnostic accuracy in Alzheimer's disease detection.

Keywords: Alzheimer's Disease; Machine Learning Models; Random Forest; SVM; XGBoost; Stacking Model; Diagnostic Accuracy.

1 | Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder and a leading cause of dementia, with symptoms such as memory loss, language difficulties, and impaired cognitive functions. As the global aging population grows, the prevalence of AD is expected to rise significantly in the coming years, making early diagnosis crucial for effective intervention. However, many patients are diagnosed in the later stages, missing the opportunity for initial treatment [1]. In this study, we aim to improve the early detection of Alzheimer's disease by applying machine learning models to a publicly available dataset. We explore the use of Random Forest, Support Vector Machine (SVM), XGBoost, and Stacking Model to analyze cognitive and brain structural data [2].

**Corresponding Author:** Mohamed.eassa.cs@o6u.edu.eg<https://doi.org/10.61356/j.scin.2025.2619>

Licensee **SciNexuses**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

2 | Data Preprocessing

2.1 | Handling of Data with Numerous Missing Values

Given the minimal amount of missing data in the dataset, we decided not to make significant changes, as the missing values were sparse and unlikely to impact on the overall analysis. Therefore, no further imputation or removal was required for variables with lesser amounts of missing data.

2.2 | Handling of Outliers

During the preprocessing phase, attention was given to potential outliers. Outliers were identified using box plots and the interquartile range (IQR) method. The IQR was calculated by determining the first quartile (Q1) and third quartile (Q3) for each variable, and the outlier range was set as values falling below $Q1 - 1.5 * IQR$ or above $Q3 + 1.5 * IQR$. These identified outliers were then replaced with the median value to minimize their impact on the data distribution.

2.3 | Data Encoding and Standardization

2.3.1 | Data Encoding

After basic preprocessing, categorical variables were encoded to fit machine learning models. Three features were encoded:

- **Gender:** Converted into numerical values (see Table 1).
- **Education Level:** Mapped to ordered numeric values based on educational attainment (see Table 2).
- **Ethnicity:** Label encoded to represent diverse groups numerically (see Table 3).

Table 1. Gender.

Male	Female
0	1

Table 2. Education Level.

None	High School	Bachelor's	Higher
0	1	2	3

Table 3. Ethnicity.

Caucasian	African American	Asian	Other
0	1	2	3

2.3.2 | Standardization

For standardizing quantitative data, we use the following Eq. (1):

$$x_{new} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

3 | Correlation Analysis

To highlight the key features related to Alzheimer's diagnosis, we used a Random Forest classifier. After training the model, we ranked the features based on their importance scores and selected the top ten for further analysis. We then computed a correlation matrix among these top features and visualized it using a

heatmap. As expected, the diagonal values were all one, showing each feature's perfect correlation with itself. However, the off-diagonal values showed very weak correlations, mostly ranging between -0.03 and 0.03. This indicates that while these features are important for the model, they are not strongly correlated with each other [3]. Such low correlation values suggest that the selected features are likely to contribute unique information to the prediction task, which can help reduce redundancy in the model as shown in Figure 1.

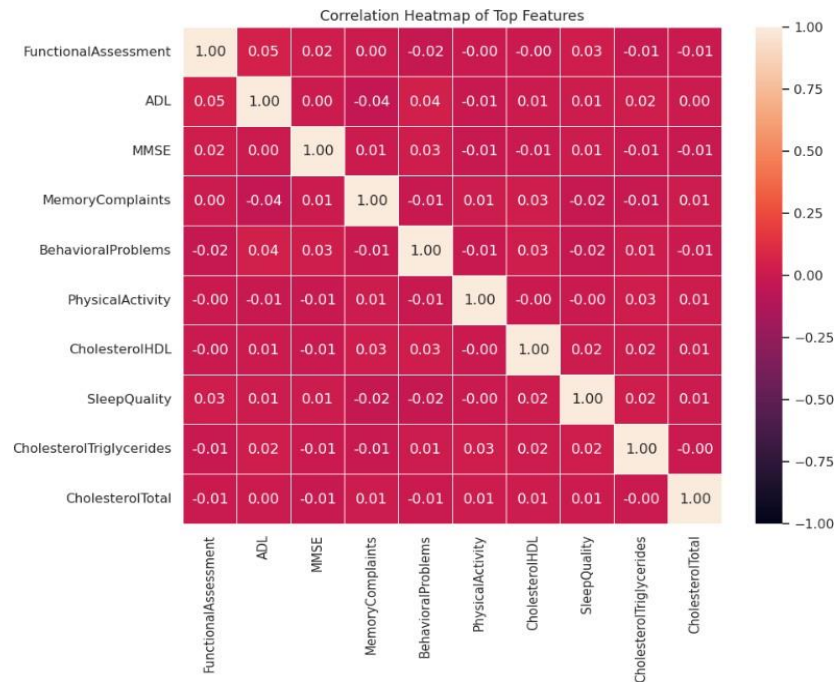


Figure 1. Correlation Heatmap of top features.

4 | Establishment of the Diagnostic Model

4.1 | Random Forest

By training several decision trees on arbitrary data samples, Random Forest develops bagging algorithms. A distinct subset produced by sampling with replacement is used to train each tree. Because just a random subset of characteristics is taken into account at each split, overfitting is decreased and model diversity is increased. A majority vote across all trees determines the final prediction [4]. For Random Forest Equation, we use the following Eq. (2):

$$f_b(x) \sum_{b=1}^B \frac{1}{B} = (x) \hat{f} \quad (2)$$

Figure 2 illustrates the general structure of the Bagging algorithm used in Random Forest, where multiple base learners contribute to the final prediction through averaging.

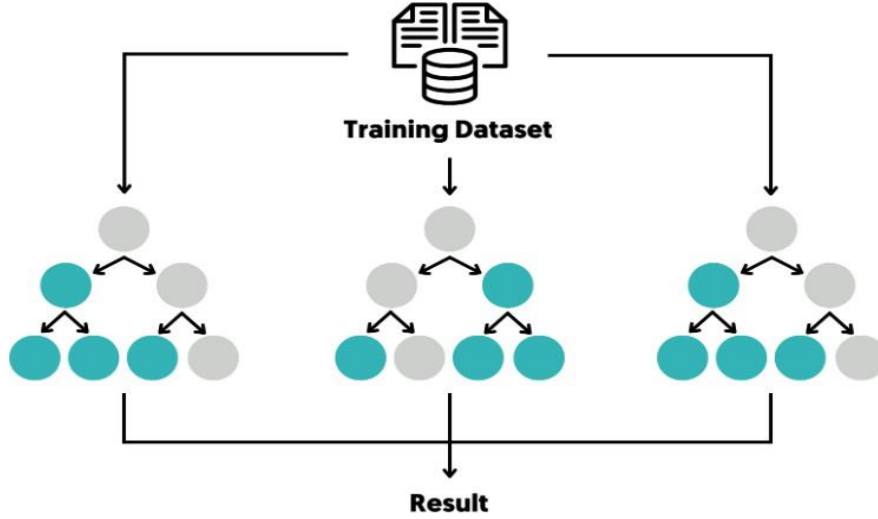


Figure 2. General structure of the Bagging algorithm used in Random Forest.

4.2 | SVM

A popular supervised machine learning approach for classification tasks is Support Vector Machine (SVM). By maximizing the margin between the classes, SVM seeks to identify the hyperplane in the feature space that best divides various classes. When there are more dimensions than data points, it works especially well in high-dimensional spaces [5].

For SVM Equation, we use the following Eq. (3):

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \right) \quad (3)$$

Figure 3 shows the structure of SVM, where a hyperplane is created to separate different classes in the feature space. The model maximizes the margin between the classes, and the final decision boundary is determined by support vectors closest to the hyperplane.

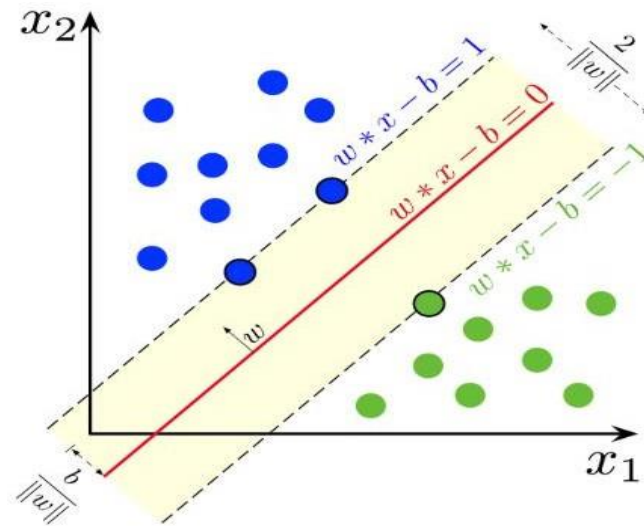


Figure 3. Structure of SVM.

4.3 | XGBoost

Building on the idea of Gradient Boosting, XGBoost (Extreme Gradient Boosting) is a potent machine learning technique. It is well known for its effectiveness and performance in resolving regression and

classification issues. In contrast to conventional boosting algorithms, XGBoost employs a number of sophisticated strategies, including regularisation and tree pruning, to improve model performance and lessen overfitting [6].

For XGBoost, we use the following Eq. (4):

$$f(x) = \sum_{k=1}^K \gamma_k * h_k(x) \quad (4)$$

Figure 4 shows the structure of XGBoost, where multiple decision trees work together to make the final prediction. Each tree is built to correct the mistakes of the previous one, and the results are combined to improve accuracy.

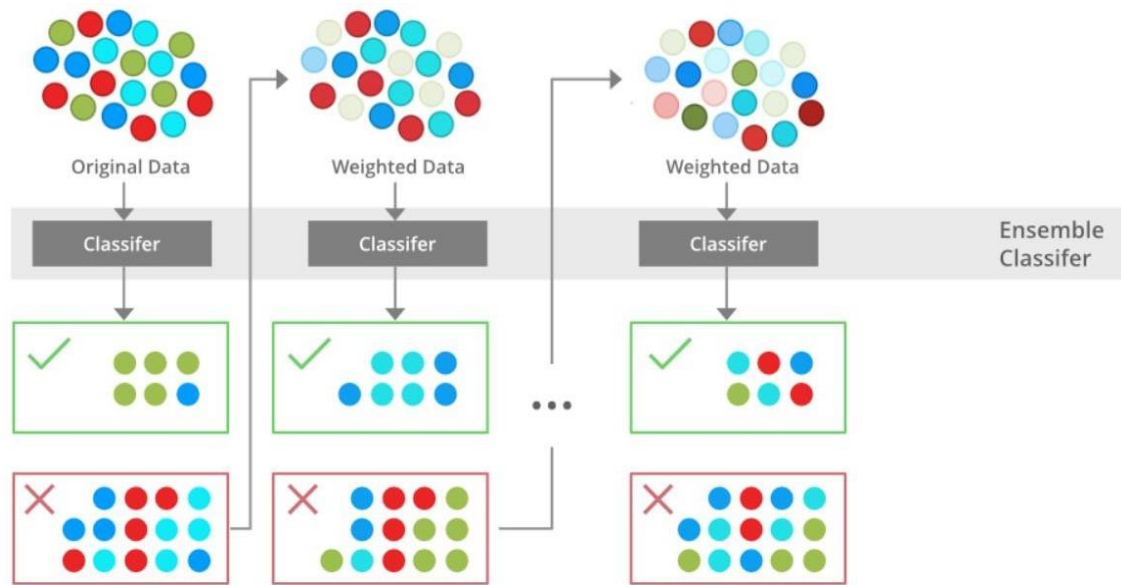


Figure 4. Structure of XGBoost.

4.4 | Stacking Model

The Stacking model combines multiple machine learning algorithms, such as Random Forest, XGBoost, and CatBoost, to improve predictive performance. Each model is trained on the same dataset, and their predictions are then used as input for a meta-model. The meta-model learns how to combine the outputs of the base models to produce a final, more accurate prediction. This ensemble approach leverages the strengths of each individual model, reducing bias and variance, and often leading to better generalization [7].

For Random Forest Equation, we use the following Eq. (5):

$$\alpha_m f_m(x) \sum_{m=1}^M = (x) \hat{f} \quad (5)$$

Figure 5 illustrates the general structure of the Stacking algorithm, where multiple base learners contribute to the final prediction. Each base learner makes an individual prediction, and these predictions are then combined by a meta-model to provide the final outcome."

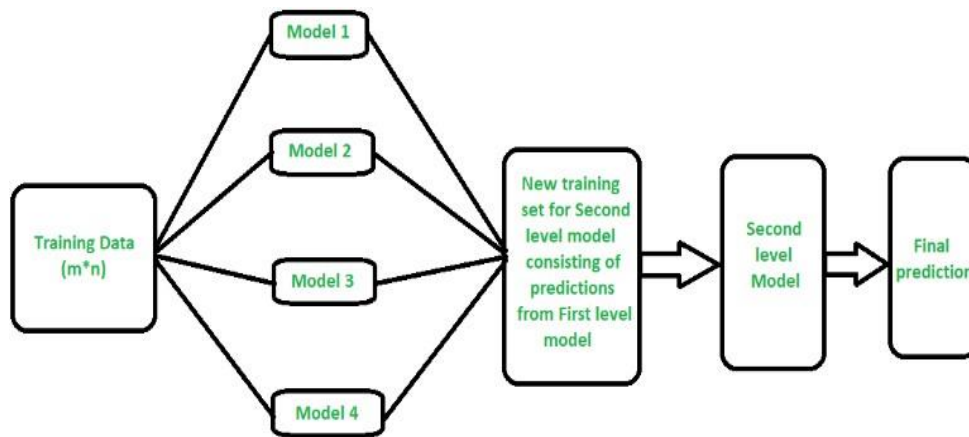


Figure 5. Structure of the Stacking algorithm.

5 | Model Comparison

5.1 | Evaluation of the Three Models

- Random Forest achieved an accuracy of 90%, with a good balance between precision (0.89) and recall (0.96) for class 0, and solid precision (0.91) and recall (0.78) for class 1. Its F1-score for class 0 (0.92) and class 1 (0.84) are also commendable.
- Support Vector Machine (SVM) achieved an accuracy of 83%, with precision and recall values of 0.84 and 0.92 for class 0, and 0.82 and 0.69 for class 1. The SVM's performance is a bit lower in comparison to Random Forest and XGBoost, especially for class 1, where the recall is lower.
- XGBoost performed the best with an accuracy of 92%, along with extremely high precision (0.96) and recall (0.80) for class 1, and high precision (0.90) and recall (0.98) for class 0. The F1-scores for class 0 and 1 are 0.94 and 0.87, respectively.

Based on the evaluation of all three models as shown in Figure 6, XGBoost performed the best, achieving the highest accuracy (92%) and demonstrating robust performance in both precision and recall for both classes. It consistently outperforms Random Forest and Support Vector Machine (SVM), particularly for class 1. Therefore, XGBoost is the most effective model among the three in this analysis.

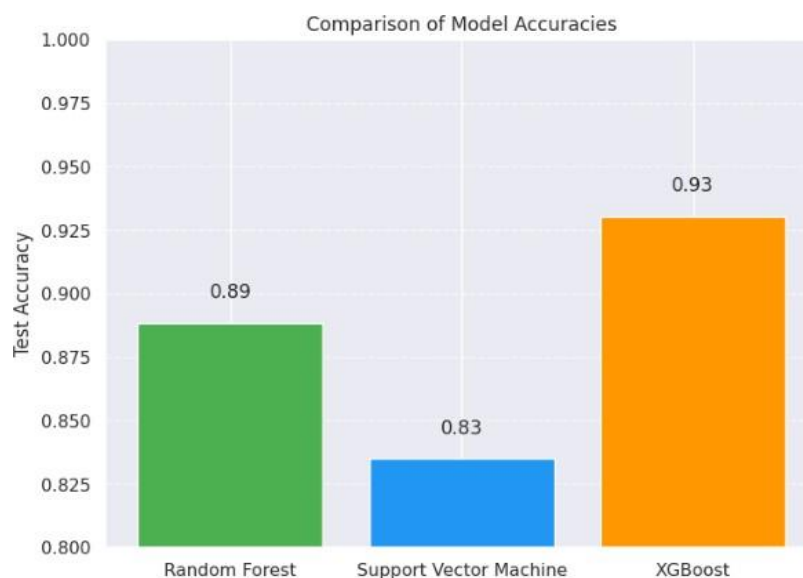


Figure 6. Comparisons of model accuracies.

5.2 | Best Performing Model Comparison with Stacking Model

The Stacking Model outperforms XGBoost in all key metrics, achieving a higher accuracy of 96% compared to 92%. The Stacking Model also shows improved precision, recall, and F1-scores, particularly for class 0. Overall, the Stacking Model provides a more balanced and effective prediction, making it the better model for this task as shown in Table 4.

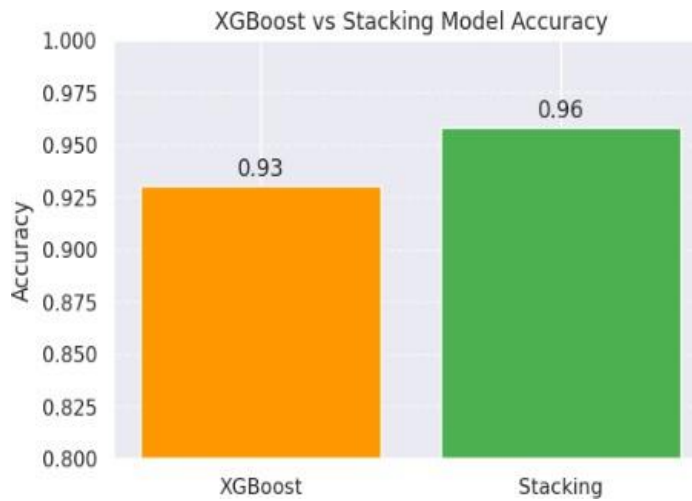


Figure 7. XGBoost vs. Stacking Model Accuracy.

Table 4. Evaluation of the applied models.

Model	Class 0 Precision	Class 0 Recall	Class 0 F1-score	Class 1 Precision	Class 1 Recall	Class 1 F1-score	Accuracy
Random Forest	0.89	0.96	0.92	0.91	0.78	0.84	0.9
Support Vector Machine	0.84	0.92	0.88	0.82	0.69	0.75	0.83
XGBoost	0.9	0.98	0.94	0.96	0.8	0.87	0.92
Stacking Model	0.96	0.98	0.97	0.96	0.92	0.94	0.96

6 | Conclusion

From the model evaluation, it is clear that the Stacking Model outperforms the individual models in terms of overall performance. With an accuracy of 96%, it demonstrates excellent precision and recall across both classes, surpassing XGBoost, which is the second-best with an accuracy of 92%. The Random Forest model and Support Vector Machine (SVM) performed well, but with lower accuracy and F1-scores compared to the Stacking model. The Stacking Model, combining multiple classifiers such as Random Forest, XGBoost, and SVM, capitalizes on the strengths of each, leading to improved predictive accuracy. Therefore, for this task, Stacking proves to be the best model choice for high accuracy and reliable performance.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data, but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] Li, Mingyang; Liu, Hongyu; Li, Yixuan; Wang, Zejun; Yuan, Yuan; Dai, Honglin. (2024). Intelligent diagnosis of Alzheimer's disease based on machine learning. arXiv preprint arXiv:2402.08539.
- [2] Abd El Hamid, Marwa M.; Mabrouk, Mai S.; Omar, Yasser M. K. (2019). Developing an early predictive system for identifying genetic biomarkers associated to Alzheimer's disease using machine learning techniques. *Biomedical Engineering: Applications, Basis and Communications*, 31(05), Article 1950040. <https://doi.org/10.4015/S1016237219500406>
- [3] Abd El Hamid, Marwa M.; Omar, Yasser M. K.; Mabrouk, Mai S. (2016). Identifying genetic biomarkers associated to Alzheimer's disease using Support Vector Machine. In *Proceedings of the 2016 8th Cairo International Biomedical Engineering Conference (CIBEC)* (pp. 5–9). IEEE. doi:10.1109/CIBEC.2016.7836087.
- [4] Abd El Hamid, Marwa M.; Shaheen, Mohamed; Mabrouk, Mai S.; Omar, Yasser M. K. (2021). Machine learning for detecting epistasis interactions and its relevance to personalized medicine in Alzheimer's disease: Systematic review. *Biomedical Engineering: Applications, Basis and Communications*, 33(06), Article 2150047. <https://doi.org/10.4015/S1016237221500473>
- [5] Abd El Hamid, Marwa M.; Omar, Yasser M. K.; Shaheen, Mohamed; Mabrouk, Mai S. (2022). Discovering epistasis interactions in Alzheimer's disease using deep learning model. *Gene Reports*, 29, 101673. <https://doi.org/10.1016/j.genrep.2022.101673>
- [6] Abd El Hamid, Marwa M.; Ali, Nehal M.; Saad, Mohamed N.; Mabrouk, Mai S.; Shaker, Olfat G. (2020). Multiple sclerosis: An associated single-nucleotide polymorphism study on Egyptian population. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1), 48. <https://doi.org/10.1007/s13721-020-00255-6>
- [7] Mirzaei, Golrokh; Adeli, Hojjat. (2022). Machine learning techniques for diagnosis of Alzheimer disease, mild cognitive disorder, and other types of dementia. *Biomedical Signal Processing and Control*, 72, 103293. <https://doi.org/10.1016/j.bspc.2021.103293>

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.