Probable Title:-

**Graph-Based Machine Learning for Alzheimer’s Disease Detection: A Comparative Study of GNNs and Traditional Models**

**Abstract**

This study presents a novel graph-based and ensemble learning approach for the early detection and classification of Alzheimer’s Disease (AD) using structured clinical, cognitive, and biomarker data from the DARWIN dataset. Traditional diagnostic methods often rely on subjective assessments or expensive imaging techniques, which can delay timely intervention. To overcome these limitations, we construct patient similarity graphs based on biomarker profiles and clinical records and apply Graph Neural Networks (GNNs), including GraphSAGE and Graph Attention Networks (GAT), for patient classification into different stages of AD.

As a novel contribution, we also introduce an ensemble learning framework consisting of XGBoost, Random Forest, LightGBM, and CatBoost classifiers. We evaluate and compare the performance of graph-based models, the ensemble, and traditional machine learning classifiers to identify the most effective strategy for AD diagnosis. The results highlight the strength of combining relational modeling through GNNs with the robustness of ensemble classifiers to improve diagnostic accuracy and reliability.

**Introduction**

Alzheimer’s Disease (AD) is a neurodegenerative disorder that is progressive in nature and has a tremendous effect on aspects of cognition and memory as well as an individual’s daily life. Early and accurate diagnosis is key to effective intervention but traditional diagnostic methods can be based on subjective assessments or expensive imaging techniques. Machine learning (ML) and deep learning have emerged as new tools of choice for automated and data-driven AD detection.

Here, we propose a graph based approach to Alzheimer's detection based on biomarker information, cognitive test scores, and patient clinical records contained in the Darwin dataset. This dataset based on structured clinical and biomarker data enables us to model complex relationships among patients based on disease characteristics. In summary, we use biomarker similarities to build patient similarity graphs and then use Graph Neural Networks (GNNs) to classify the patients into different stages of AD.

In order to evaluate the performance of GNNs, we contrast the performance of GraphSAGE and Graph Attention Networks (GAT) with traditional machine learning algorithms like Random Forest, SVM and Logistic Regression. The findings further offer a glimpse into the benefits of the graph-based learnings compared to traditional ML approaches in identifying complex disease patterns.

This study aims to highlight the potential of graph-based models in Alzheimer’s disease diagnosis, demonstrating how patient similarity graphs can improve classification accuracy and aid in early detection efforts.

**Methodology**

We use the DARWIN dataset containing handwriting-based cognitive task features for 174 participants. Each participant's data comprises 452 features derived from 25 tasks. The preprocessing includes missing value imputation, standardization, and optional PCA for dimensionality reduction. To capture inter-patient relationships, we construct a patient similarity graph using cosine similarity or k-nearest neighbors. Each node represents a patient, and edges encode similarity scores.

We implement two Graph Neural Network (GNN) models: GraphSAGE, which aggregates neighborhood information for node embeddings, Graph Attention Network (GAT), which uses attention mechanisms to weigh neighbor importance. In parallel, we develop an ensemble model comprising XGBoost, Random Forest, LightGBM, and CatBoost classifiers.  
Predictions are combined through soft voting to enhance robustness.

We evaluate models using accuracy, precision, recall, F1-score, and ROC-AUC. Comparative analysis is done between GNNs, ensemble methods, and traditional classifiers like SVM and Logistic Regression. We also explore data augmentation and ablation studies to assess model generalizability.

**Results and Future Work**

* Our experiments demonstrate that graph-based models, particularly GraphSAGE and GAT, outperform traditional classifiers like SVM and Logistic Regression in capturing complex inter-patient relationships. The ensemble model combining XGBoost, Random Forest, LightGBM, and CatBoost further boosts classification accuracy and stability across different folds.
* GNNs show strong performance on structured patient similarity graphs.
* Ensemble models offer robust, competitive results even on high-dimensional tabular data.
* Data augmentation has mixed impact—boosts traditional models but may degrade GNN performance.

Future work may include:

* Investigate multi-modal fusion by integrating MRI features with handwriting-based data.
* Apply advanced graph construction techniques (e.g., dynamic or attention-based graphs).
* Explore self-supervised learning to reduce dependence on labelled data.
* Optimize model scalability for real-time clinical deployment.

**Reference**

Vimaladevi, M., Thangamani, R., Priyadharshini, S., Varshini, B., & Tarun, A. (2024, August). Prediction of Alzheimer's Disease by Analyzing Handwriting Dynamics Using Machine Learning Algorithms. In *2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1298-1304). IEEE.