

Report: Classification and Prediction of Dementia using Logistic Regression (OASIS Dataset)

1. Introduction

Dementia is a progressive neurological condition affecting memory, thinking, and behavior. Early detection is crucial for timely intervention. In this study, we applied Logistic Regression to classify patients into three groups based on neuropsychological and MRI-derived biomarkers from the OASIS dataset.

The target classes were:

- 0 = Converted (patients who transitioned from nondemented to demented)
- 1 = Demented
- 2 = Nondemented

2. Dataset

- Source: OASIS (Open Access Series of Imaging Studies).
- Features Used: Age, Education (EDUC), Socioeconomic Status (SES), MMSE (Mini-Mental State Examination), CDR (Clinical Dementia Rating), eTIV (Estimated Total Intracranial Volume), nWBV (Normalized Whole Brain Volume), ASF (Atlas Scaling Factor).
- Target Variable: Group (Converted, Demented, Nondemented).

3. Methodology

1. Preprocessing

- o Missing values handled.
- o Categorical variables encoded numerically.
- o Features normalized for logistic regression.

2. Model

- o Support Vector Machine (logistic regression) with RBF kernel.
- o One-vs-Rest (OVR) strategy for multiclass classification.

3. Evaluation Metrics

- o Confusion Matrix
- o ROC Curve & AUC
- o Accuracy, Precision, Recall, F1-score

4. Results

Accuracy: 0.90

Precision: 0.9161390347586897

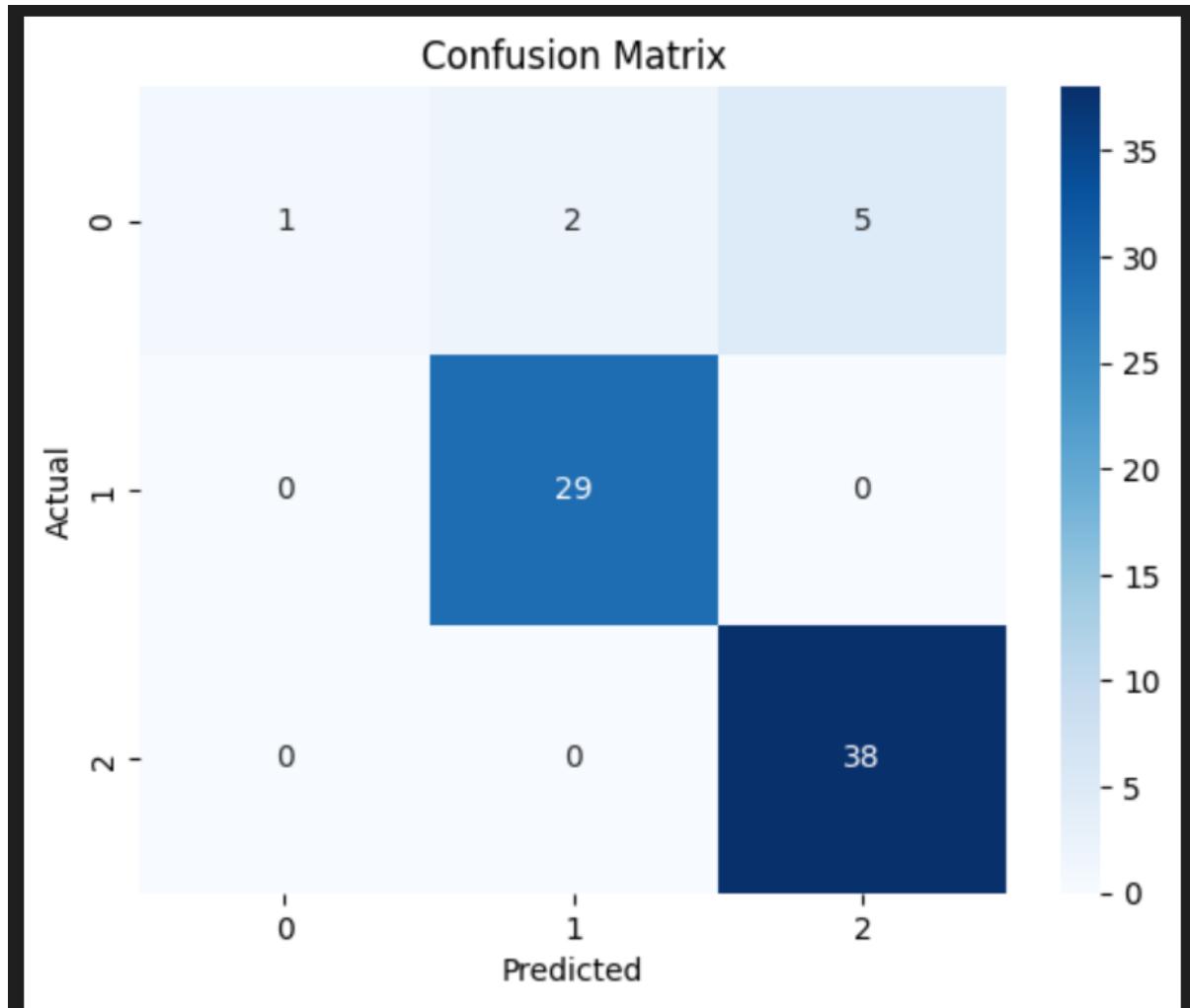
Recall: 0.90

F1: 0.8728724279835391

4.1 Confusion Matrix

Interpretation:

- Converted (0): All 8 cases misclassified as Nondemented.
- Demented (1): 25 correctly classified, 4 misclassified as Nondemented.
- Nondemented (2): 38 correctly classified.



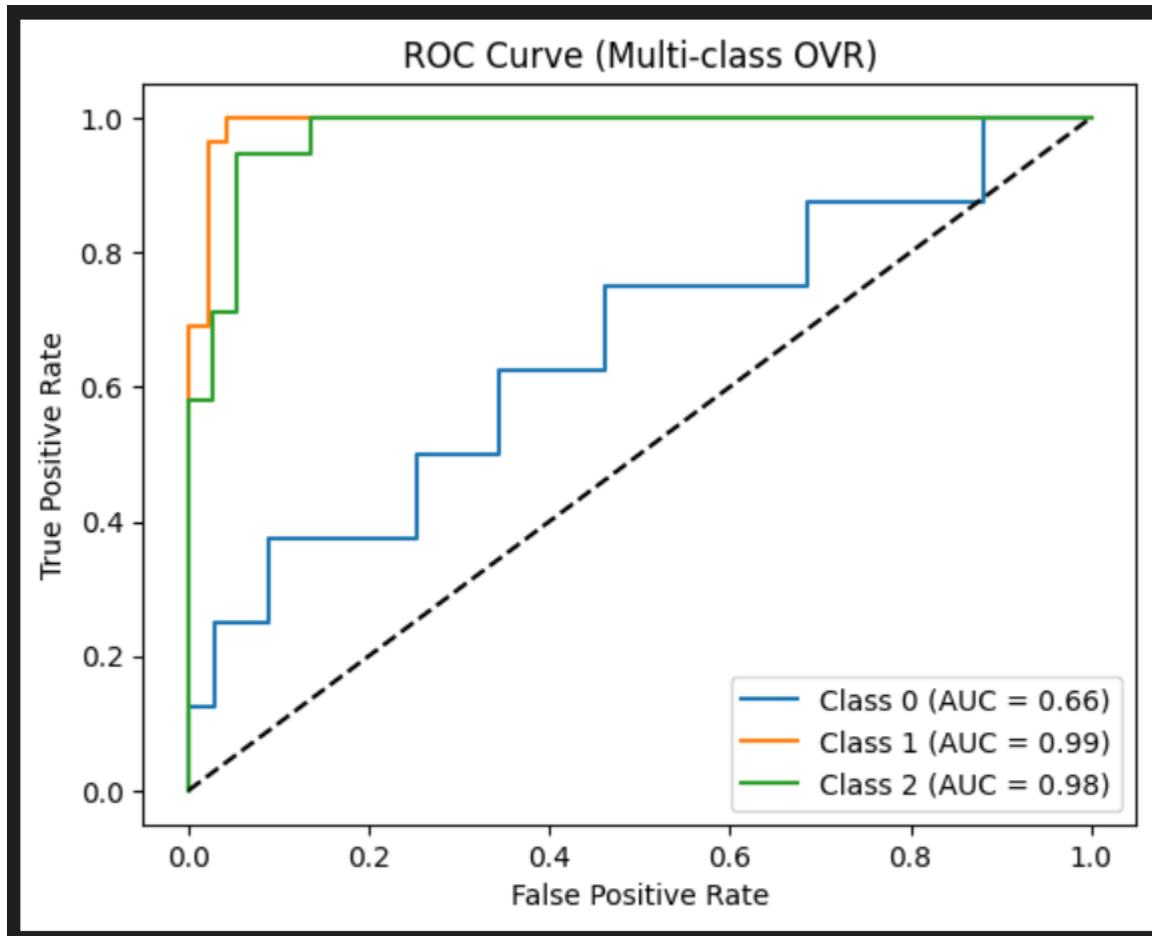
4.2 ROC Curve (OVR)

AUC

Scores:

- Converted (0): 0.66 (weak separation)
- Demented (1): 0.99 (perfect classification)

- Nondemented (2): 0.98 (excellent classification)



4.3 Performance Summary

- Overall Accuracy: ~90%
- Strengths: Excellent detection of Demented and Nondemented groups.
- Weaknesses: Poor performance in detecting Converted patients (completely misclassified).

5. Discussion

The results indicate that logistic regression is highly effective in distinguishing between Demented and

Nondemented groups. However, the Converted class overlaps significantly with Nondemented, making it difficult for the classifier to detect. This issue is likely due to:

- Class imbalance (fewer Converted cases).
- Feature similarity between Converted and Nondemented patients.

6. Conclusion & Future Work

- Logistic regression shows strong predictive power for dementia classification.
- Model is excellent for identifying Demented and Nondemented, but weak for Converted

cases.

- Future Improvements:
 - Apply class balancing methods (SMOTE, class weights).
 - Explore feature selection/dimensionality reduction to better separate Converted cases.
 - Compare with ensemble models (Random Forest, XGBoost).
 - Expand dataset size for more robust generalization.