Performance Evaluation of Early-Stage Dementia Prediction Models

Linglu Li

Evaluation Metrics

The Table 1 shows the performance of five different classifier models: Decision Tree, Random Forest, Support Vector Machine (SVM), XGBoost, and Logistic Regression. The comparison is based on the following metrics:

- **Accuracy**: The proportion of correctly classified result from the total instances.
- **Precision**: The ratio of correctly predicted positive rate to the total predicted positive rates. A good classifier close to value 1.
- **Recall**: The true positive rate. A good classifier close to value 1.
- **F1 Score**: The weighted average of Recall and Precision, which balances both metrics.

The SVM and Logistic Regression models performed the worst in identifying actual dementia patients, as indicated by their lowest Recall of 0.75. This means they only correctly identified 75% of the actual dementia cases in the test set, leading to the highest proportion of misclassified dementia patients. The model best at correctly identifying actual dementia cases is XGBoost with a Recall of 0.88.

XGBoost also ties with the Decision Tree for the highest overall Accuracy (89.36%), with each model reflecting a beneficial aspect of the Precision–Recall trade-off. The Decision Tree offers higher Precision (0.87), effectively reducing False Positives (i.e., fewer healthy patients are misdiagnosed). XGBoost, on the other hand, excels at capturing true dementia cases, reaching a Recall of (0.88) while maintaining solid Precision (0.82).

XGBoost is the recommended model because its high Recall minimizes the risk of missed diagnoses, making it well-suited for clinical decision-making. Although its Precision is slightly lower, its strong ability to avoid false negatives makes it the safest and most reliable choice for detecting disease.

Table 1: Performance comparison of ML models.

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
Decision Tree	89.36%	0.87	0.81	0.84
Random Forest	87.23%	0.81	0.81	0.81
SVM	85.11%	0.80	0.75	0.77
XGBoost	89.36%	0.82	0.88	0.85
Logistic Reg.	82.98%	0.75	0.75	0.75

Note: A positive prediction indicates that the model predicts the patient has Alzheimer's Disease (is Demented), whereas a negative prediction indicates the model predicts the patient does not have Alzheimer's Disease (is Nondemented). All four evaluation metrics—Accuracy, Precision, Recall, and F1-score—are calculated from the fundamental confusion matrix components (TP, TN, FP, and FN). More detailed definitions of these terms are provided in Tables 2 and 3.

Table 2: Correct Outcomes from the Confusion Matrix.

Term	Full Name	Clinical Context
TP	True Positive	The model predicts the patient is Demented, and the patient actually is Demented. (A successful diagnosis).
TN	True Negative	The model predicts the patient is Nondemented, and the patient actually is Nondemented. (A successful exclusion of the disease).

Table 3: Incorrect Outcomes (Errors) from the Confusion Matrix.

Term	Full Name	Clinical Context
FP	False Positive	The model predicts the patient is Demented, but the patient actually is Nondemented. (A misdiagnosis).
FN	False Negative	The model predicts the patient is Nondemented, but the patient actually is Demented. (A missed diagnosis).

Performance Table Visualization

Detailed comparisons of model performance metrics (Accuracy, Precision, Recall, and F1-score) are visually represented in Figure 1.

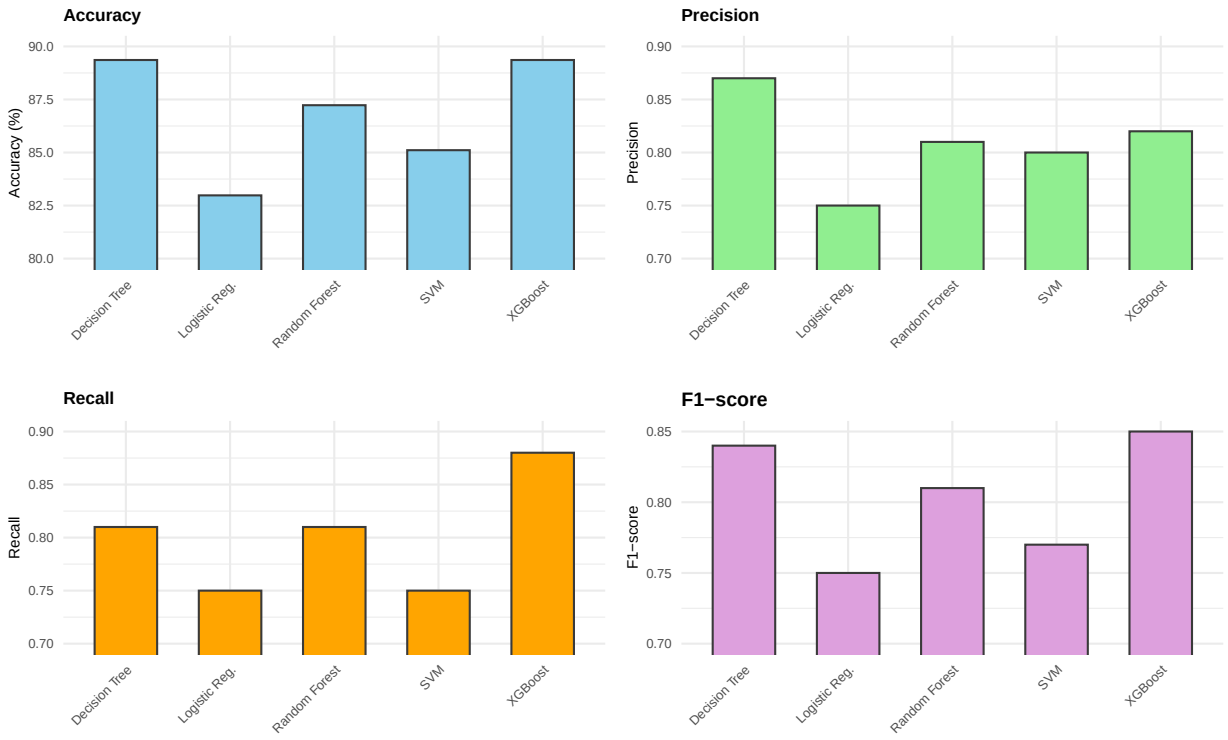

Figure 1: Model comparison across Accuracy, Precision, Recall, and F1-score.

Confusion Matrix Analysis

The individual confusion matrices (Figure 2) provide the underlying count data that justifies the calculated performance metrics (Table 1). The confusion matrix data visually and numerically validates the conclusion that XGBoost is the optimal model. Its ability to minimize False Negatives (FN = 2) is the defining factor, ensuring the highest rate of true dementia case detection necessary for early intervention.

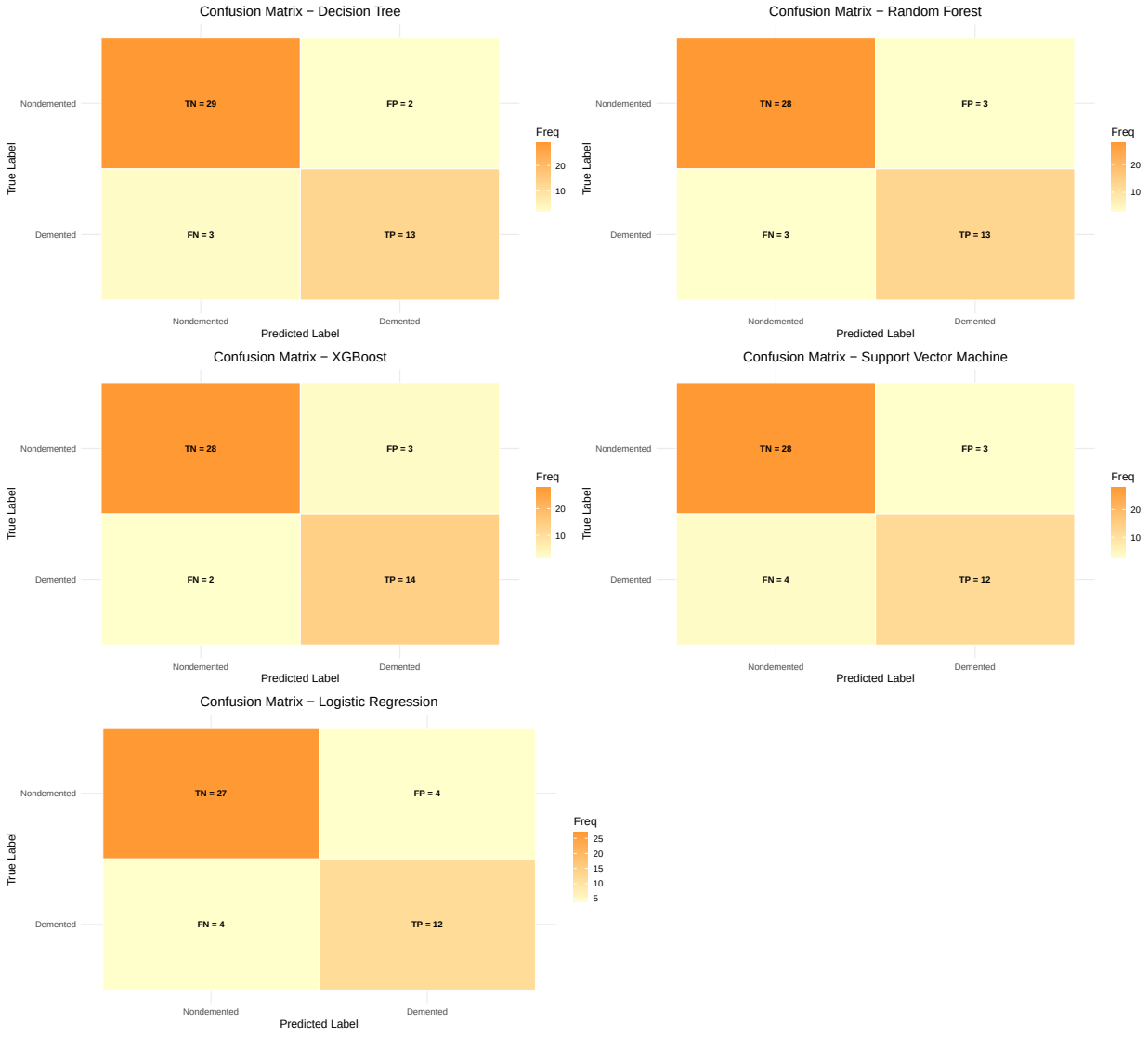

Figure 2: Individual Confusion Matrices for All Five Models.