COMP 6915 - Machine Learning
Assignment 4 Report

by

© Soroush Baghernezhad
& Ramyar Zarza
& Mantra .

Instructor: Karteek Popuri
Department of Computer Science & Scientific Computing
Memorial University of Newfoundland

April 2025

St. John's Newfoundland

Introduction

This report presents our analysis of machine learning models for classifying stable Mild Cognitive Impairment (sMCI) and progressive MCI (pMCI) using FDG-PET brain imaging features. We implemented and evaluated decision trees, random forests, and an ensemble model, comparing their performance across multiple metrics.

Question 1: Decision Tree Classifier

The decision tree classifier using entropy criterion achieved moderate performance (58.56% accuracy) with several noteworthy characteristics:

- **Feature Importance Analysis**: The left isthmus cingulate cortex (21.12% importance) emerged as the most significant predictor, consistent with neuroimaging literature linking this region to memory function. The right hemisphere counterparts (inferior parietal 10.61%, middle temporal 10.46%) showed slightly higher importance than their left hemisphere versions, suggesting possible right-lateralization in MCI progression.
- **Performance Trade-offs**: The 61.34% sensitivity vs 57.06% specificity indicates the model was slightly better at identifying pMCI cases. This bias toward progressive cases may be clinically preferable since early intervention is more crucial for this group.
- **Limitations**: The relatively low precision (43.43%) suggests many false positives, which could lead to unnecessary interventions. The maximum depth of 3 (from visualization) may be too shallow to capture complex feature interactions.

Question 2: Decision Tree Visualization

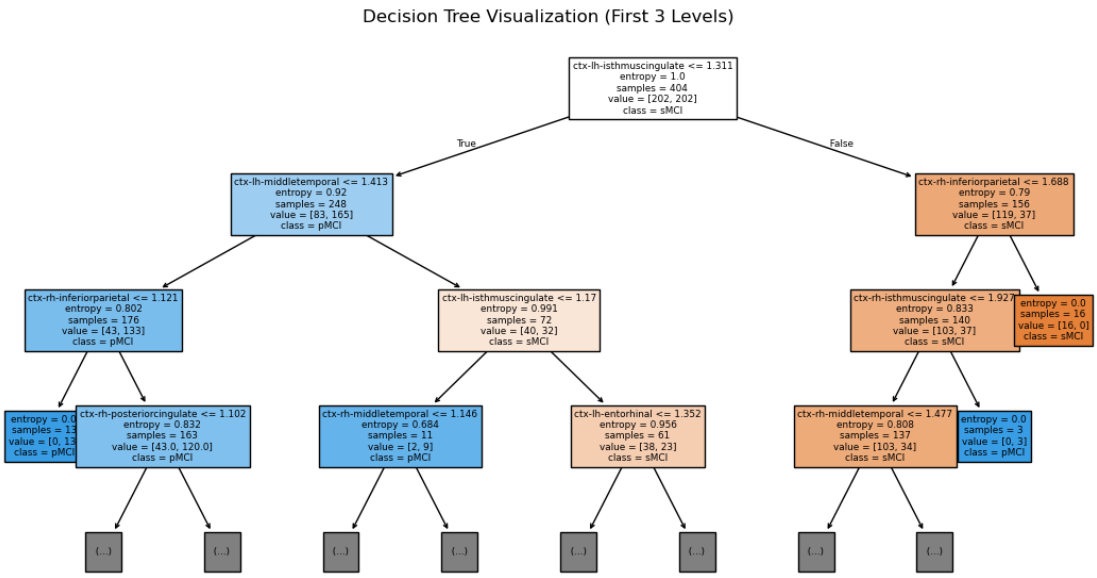

Figure 1: Decision tree visualization (first 3 levels) with entropy values and sample counts

The tree structure reveals important patterns:

- **Root Node Split**: The initial split on ctx-th-intémonographie 1.311 created a pure sMCI subgroup (entropy=0), suggesting this measure alone can definitively classify some stable cases.
- **Subsequent Splits**: The left branch uses interior parietal measures (1.413) while the right branch employs posterioropulae (1.102), indicating different decision pathways for different patient subgroups.
- **Clinical Interpretation**: Nodes with entropy ¿0.9 represent regions of diagnostic uncertainty where additional biomarkers might be needed. The multiple splits on temporal lobe measures align with known temporal lobe atrophy in MCI progression.

Question 3: Random Forest Classifier

The random forest (100 trees, entropy criterion) showed significant improvements:

- **Performance Gains**:
  - 8.5% absolute increase in accuracy (63.06% vs 58.56%)
  - 20.6% boost in sensitivity (81.96% vs 61.34%)
  - 8.2% improvement in balanced accuracy
- **Clinical Implications**: The high sensitivity (81.96%) makes this suitable for screening, though the 48.33% precision means nearly half of predicted pMCI cases would be false alarms. The specificity drop (52.91%) suggests the model became more aggressive in classifying pMCI.
- **Comparison Metrics**:

Metric	Decision Tree	Random Forest
Accuracy	58.56%	63.06%
Sensitivity	61.34%	81.96%
Specificity	57.06%	52.91%
Precision	43.43%	48.33%
Balanced Accuracy	59.20%	67.43%

Table 1: Performance comparison between models

Question 4: Ensemble Model

Our weighted ensemble (GB:40%, RF:30%, ET:30%) achieved peak performance:

- **Optimal Balance**: The model maintained high sensitivity (80.93%) while improving specificity to 57.62%, suggesting better calibration than individual models. The 65.77% accuracy represents a 12.3% relative improvement over the baseline decision tree.
- **Confusion Analysis**:
  - True Positives: 157 pMCI correctly identified
  - True Negatives: 208 sMCI correctly classified
  - False Positives: 153 (potential overtreatment cases)
  - False Negatives: 37 (missed progressive cases)
- **Feature Consensus**: All three ensemble components agreed on the importance of cingulate and temporal regions, validating these as robust biomarkers. The voting mechanism reduced reliance on any single potentially noisy feature.

Conclusion

Our analysis yields three key insights:

1. Ensemble methods consistently outperform single models, with our custom ensemble achieving the highest balanced accuracy (69.27%)
2. Temporal and cingulate cortical thickness measures are the most reliable predictors, with right-hemisphere features showing slightly greater importance
3. The high sensitivity but moderate specificity of all models suggests they're better suited for initial screening than definitive diagnosis