DS4420 MLDM2 Project

Shreyas Shukla, Suneet Panthangay

Alzheimer’s Prediction from MRI

Introduction

Alzheimer’s disease is often framed as a binary problem —“dementia or not”—but in reality it progresses through gradual stages that are hard to distinguish even for experts. In this project, we investigate whether machine learning can reliably stage cognitive impairment from structural brain MRI into four levels: No Impairment, Very Mild, Mild, and Moderate Impairment. Using a curated dataset of T1-weighted slices, we train a convolutional neural network (CNN) to classify images directly and then fit Bayesian logistic and ordinal models on top of the CNN’s predicted probabilities to study uncertainty and feature importance. Our goal is not just to chase high accuracy, but to see how well these models separate neighboring stages, how trustworthy their predictions are, and what their limitations reveal about using MRI-based ML systems for early Alzheimer’s detection.

Methodolgy & Data

We use a balanced Alzheimer’s MRI dataset of axial T1 brain slices with four classes (No, Very Mild, Mild, Moderate Impairment), where minority classes are augmented with high-fidelity WGAN-GP–generated scans. Each slice is resized (150×150 for the main model, 64×64 for the manual model) and normalized to [0,1][0,1][0,1]. We train three models: (1) an Extensive CNN in Keras with three conv–BN–max-pool blocks plus a 256-unit dense layer and dropout, trained with data augmentation to directly predict the 4 classes; (2) a Manual CNN implemented entirely in NumPy with a single conv layer, ReLU, max-pool, and a fully connected softmax layer trained via hand-coded backprop; and (3) Bayesian models in R (brms) that take the CNN’s class probabilities as inputs to a binary logistic and an ordinal cumulative-logit model, allowing us to study uncertainty and calibration rather than just accuracy.

Citations

Hechkel, Wided, and Abdelhamid Helali. 2025. “Early Detection and Classification of Alzheimer’s Disease through Data Fusion of MRI and DTI Images Using the YOLOv11 Neural Network.” *Frontiers in Neuroscience* 19: 1554015. https://doi.org/10.3389/fnins.2025.1554015.

El-Latif, Ahmed A. Abd, Samia Allaoua Chelloug,Maali Alabdulhafith, and Mohamed Hammad. 2023. “Accurate Detection of Alzheimer’s Disease Using Lightweight Deep Learning Model on MRI Data.” *Diagnostics* 13 (7): 1216. https://doi.org/10.3390/diagnostics13071216.

Results

Overall performance

- The extensive CNN achieves X% test accuracy on the 4-class task.
- However, the row-normalized confusion matrix shows that Very Mild, Mild, and Moderate cases are often predicted as No Impairment, meaning the model is conservative but misses exactly the early-stage patients we care about.

CNN vs. other models

- The manual NumPy CNN performs worse than the Keras CNN, confirming that deeper architectures and data augmentation are important for this dataset.
- The Bayesian binary model appears to get high accuracy for AD vs. non-AD, but its confusion matrix shows it mostly predicts the majority class; the ordinal Bayesian model is close to chance on the 4-class task.

Uncertainty / interpretability

- Posterior coefficient plots for the Bayesian models show that only a few features have reasonably tight credible intervals away from zero.
- Most coefficients are highly uncertain and near zero, indicating strong collinearity and that the models are not reliably identifying which features truly drive the predictions.

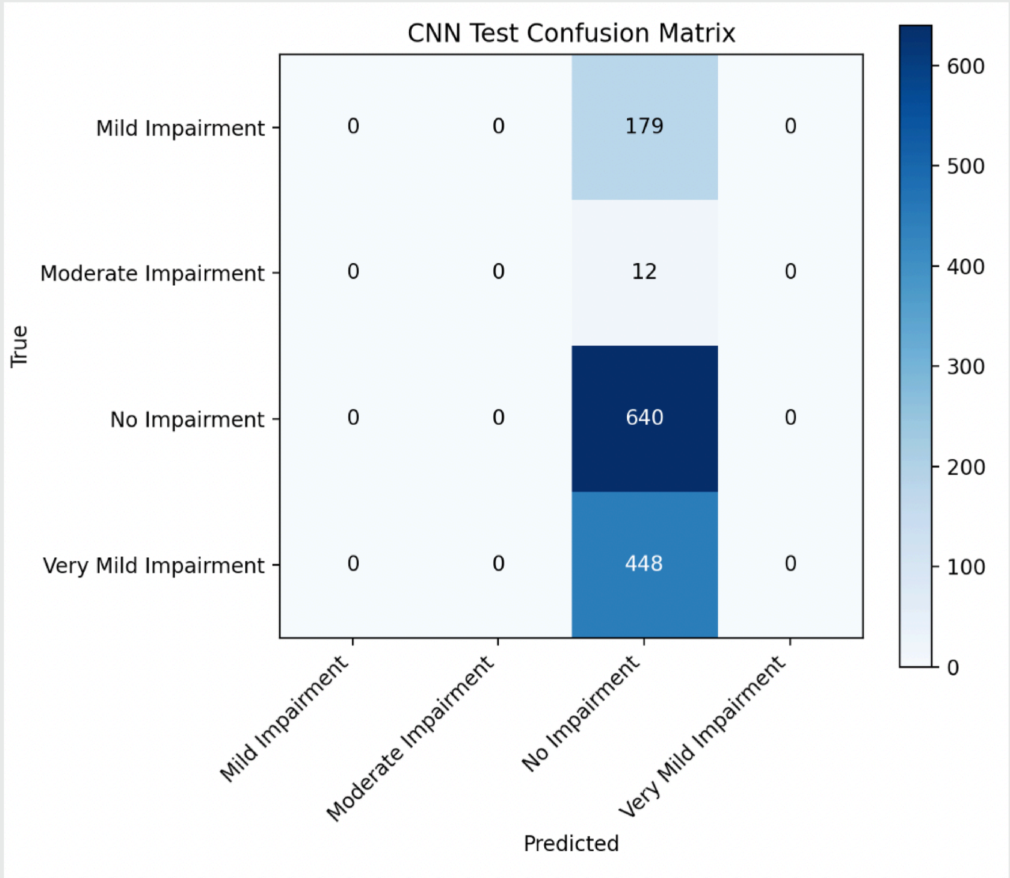

Fig 1 – CNN Confusion Matrix
“Row-normalized confusion matrix. The CNN detects No Impairment reasonably well but often mislabels Very Mild, Mild, and Moderate cases as healthy.”

model	accuracy	precision	recall
0.5004	0.2504	0.5004	0.3338
0.8516	0.0	0.0	0.0
0.8102	0.8388	0.9096	0.8728

Fig 2 – Model Comparison
“Test performance across models. The extensive CNN beats the manual CNN and both Bayesian models; the binary Bayesian model mainly exploits the majority class and the ordinal model struggles with 4-way staging.”

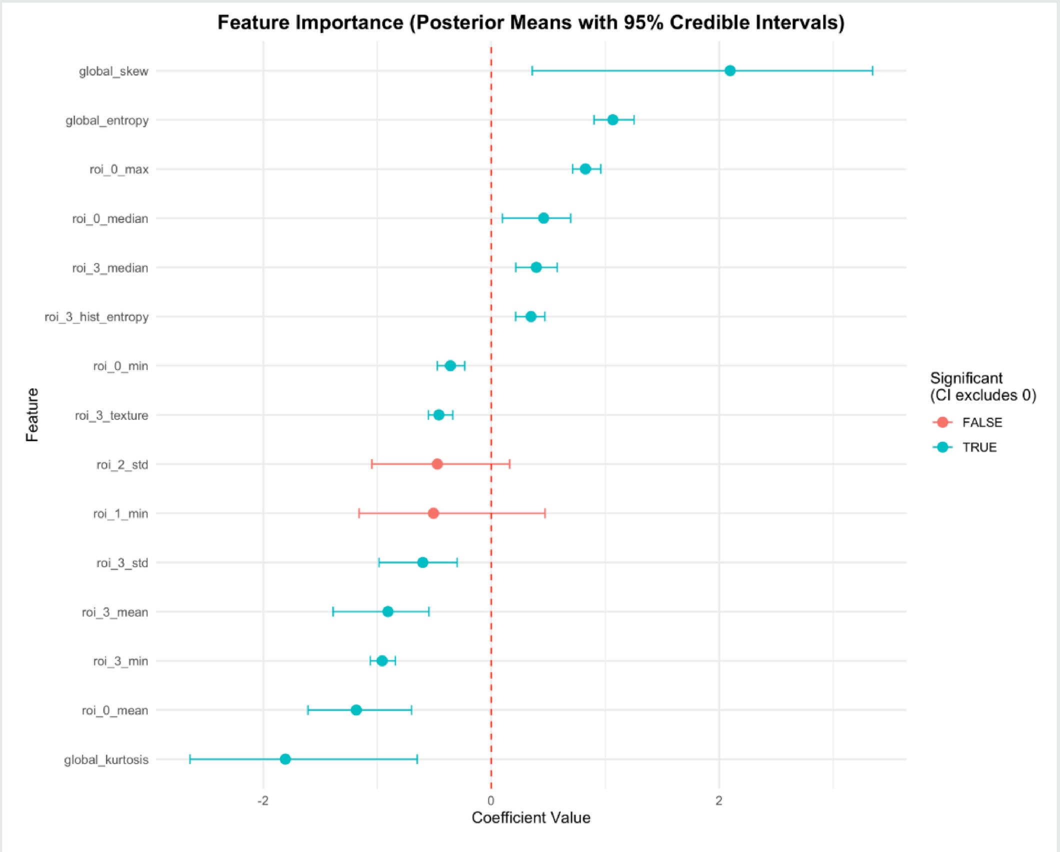

Fig 3 – Bayesian Coefficients
“Posterior means and 95% credible intervals for Bayesian coefficients. Wide intervals around zero show high uncertainty and collinearity, so feature ‘importance’ is unstable.”