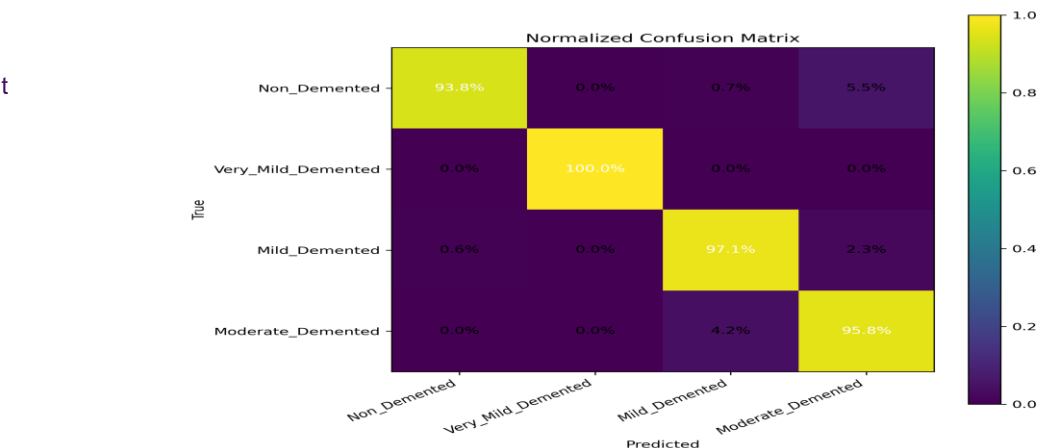
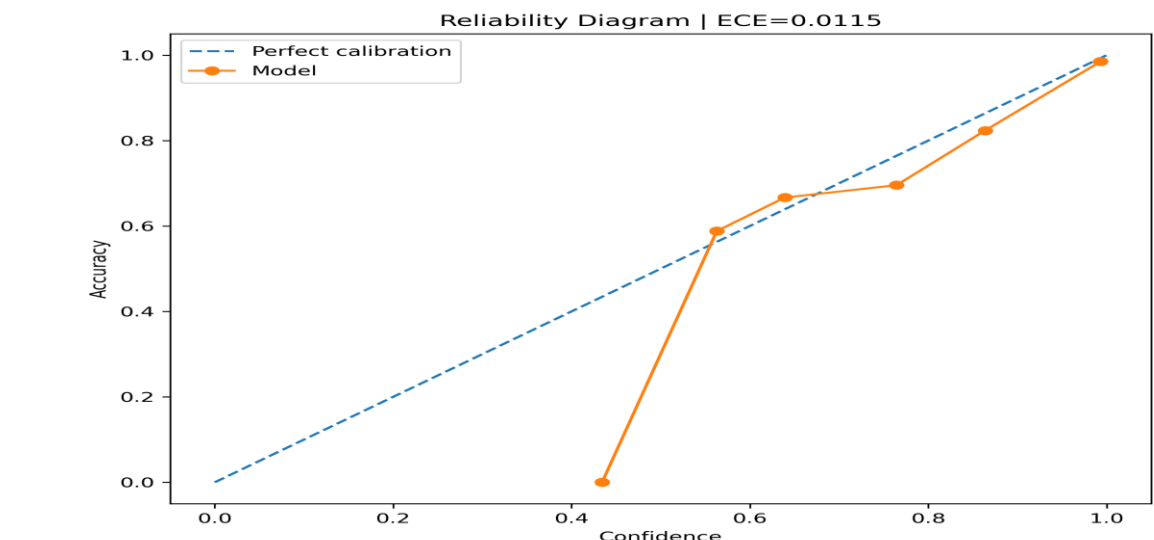
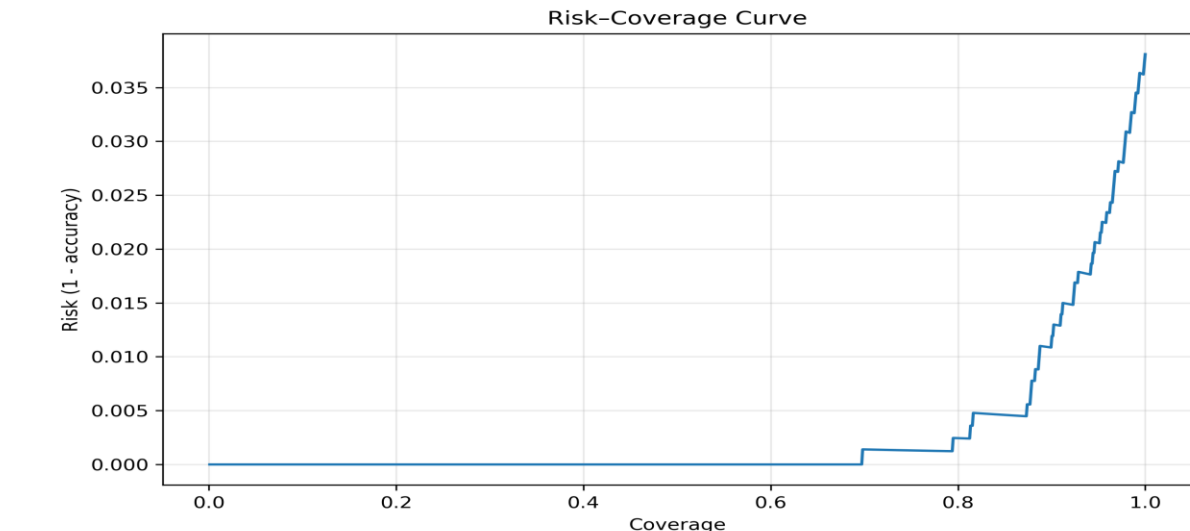


Confidence Calibration And Uncertainty Analysis In Deep Learning-based Alzheimer’s Disease Neuroimaging Classification

Ben George Ephrem, PhD¹, Abraham Varghese, PhD¹

(1) University of Technology and Applied Sciences, Muscat, Muscat, Sultanate of Oman,

A confidence-aware ResNet-18 model achieved 96.2% accuracy and excellent calibration (ECE = 0.012) for four-stage Alzheimer’s disease MRI classification, demonstrating that uncertainty-guided prediction can reduce errors from 3.8% to 0.5% by referring low-confidence cases for expert review

INTRODUCTION	METHODS	RESULTS	CONCLUSIONS	REFERENCES
- Deep learning models are increasingly applied to neuroimaging in Alzheimer's Disease (AD) research. - However, classification accuracy alone does not guarantee clinical usefulness, as predictions must be reliable. - Poorly calibrated models can erode trust, especially when distinguishing between closely related disease stages. - This study evaluates both predictive performance and confidence calibration in multi-class AD classification using structural MRI.	- A ResNet-18 CNN was trained on 6400 structural MRI slices across four categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. - Images were resized to 224×224 pixels and z-score normalized. - An 80-20 train-validation split used, and training employed cross-entropy loss with early stopping. - Performance was assessed using accuracy, precision, recall, and macro-F1 score. - Calibration analysis was done using reliability diagrams with Expected Calibration Error (ECE) and risk-coverage curves to evaluate selective prediction strategies.	 The model achieved 96.19% validation accuracy with a macro-F1 score of 96.63%  Calibration analysis show a low ECE of 0.012, indicates model's confidence estimates closely matched actual prediction accuracy  Risk-coverage analysis demonstrates that the model's confidence scores are highly informative for selective prediction strategies.	- This work shows that properly trained deep learning models can achieve both high accuracy and excellent calibration for AD neuroimaging. - Confidence and risk-coverage analysis help determine which predictions can be automated and which require expert review. - These findings support the deployment of such models as decision-support tools in research and clinical settings. - Validation on established clinical datasets and comparative benchmarking remain essential next steps to assess generalizability beyond this single dataset.	1. Falah G. Salieh. <i>Alzheimer MRI Dataset</i>. Hugging Face Datasets, 2023. 2. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. <i>ICML</i>. 2017. CONTACT  Ben George Faculty Department of IT, University of Technology and Ap...  www.linkedin.com/in/bengeorge28  ALZHEIMER'S ASSOCIATION AAIC 26