Early Alzheimer's Detection

Deep Learning on OASIS MRI Images

Brad Richardson

Applied Computer Science, BS Post-Baccalaureate, University of Colorado Boulder, brri6685@colorado.edu

1 ABSTRACT

Alzheimer's disease is a progressive neurodegenerative disorder that imposes severe burdens on patients, families, and healthcare systems. Timely detection and accurate staging of the disease is crucial for guiding interventions, optimizing patient care, and potentially slowing cognitive decline. This research explores the use of deep learning—specifically convolutional neural networks (CNNs)—to classify magnetic resonance imaging (MRI) brain scans into four categories reflecting different progression stages of Alzheimer's disease: non demented, very mild dementia, mild dementia, and moderate dementia. Unlike many previous studies that limit themselves to binary classification tasks, the multi-class approach presented here aims to detect subtle differences in brain structure that may indicate earlier stages of disease onset.

This investigation is motivated by personal experience. As the primary caregiver for my 87-year-old grandmother, who has been diagnosed with Alzheimer's disease, I have witnessed firsthand how early detection can influence patient well-being and family decision-making. By leveraging a large publicly available MRI dataset and employing transfer learning with a pre-trained VGG16 architecture, the project demonstrates how CNN-based models can automate the classification of Alzheimer's disease severity. Through careful preprocessing, class weighting, and hyperparameter tuning, the model effectively learns patterns that distinguish early cognitive impairment from normal brain aging and more severe dementia stages.

Results indicate that the trained model can achieve high accuracy, with particularly strong discrimination between non demented and very mild dementia classes. Although challenges remain in distinguishing mild dementia from moderate dementia cases due to class imbalance and subtle structural differences, the overall performance surpasses simple baselines and underscores the feasibility of automated, multi-class classification of Alzheimer's disease. Visualization techniques such as Grad-CAM further reveal which brain regions drive the network's predictions, offering insight into possible early biomarkers. The findings have potential clinical implications for screening, risk stratification, and monitoring disease progression, ultimately contributing to improved patient care and informing future research efforts.

2 INTRODUCTION

Alzheimer's disease stands as one of the most prevalent and challenging neurodegenerative disorders facing an aging global population. Characterized by progressive memory loss, cognitive decline, and changes in behavior and personality, it gradually erodes quality of life for patients and poses immense strains on families and caregivers. Early detection and accurate staging of Alzheimer's disease are imperative. Identifying the disease at its very mild or mild stages could enable targeted interventions and treatments that may delay progression, improve patient outcomes, and reduce caregiver burden.

My personal motivation for this research stems from a very real and intimate encounter with the challenges of Alzheimer's. As the primary caregiver for my 87-year-old grandmother, I have observed firsthand the progression of this

disease and how early signals of cognitive decline might have been overlooked during routine clinical assessments. This personal experience has underscored the importance of developing objective, data-driven tools that can aid in the detection and staging of Alzheimer's disease. If subtle indications of cognitive deterioration can be identified from MRI images earlier, families and healthcare providers may be better equipped to intervene before more pronounced symptoms manifest.

Recent advances in artificial intelligence (AI), particularly in the domain of deep learning, offer promising avenues for improving early diagnosis. Convolutional neural networks (CNNs) have demonstrated remarkable success in various medical imaging tasks, from detecting diabetic retinopathy in fundus images to identifying malignant lesions in mammograms. These models excel at learning complex, high-dimensional patterns that may be imperceptible to the human eye. In the context of Alzheimer's disease, CNNs have shown potential in distinguishing healthy controls from those with Alzheimer's and, in some cases, mild cognitive impairment. However, much of the existing work focuses on binary classification tasks rather than classifying multiple stages of disease severity.

This project aims to classify MRI brain images into four distinct categories: non demented, very mild dementia, mild dementia, and moderate dementia. By expanding beyond a binary classification framework, the goal is to capture a continuum of disease progression. This approach acknowledges that Alzheimer's is not an all-or-nothing condition; rather, it unfolds gradually, with subtle anatomical changes occurring in the brain before clinical symptoms become evident. Early detection of very mild dementia, for instance, could enable preventive measures or therapies when they may be most effective.

The remainder of this report is organized as follows: the Related Work section examines previous studies that have applied deep learning to Alzheimer's detection. The Data Set section describes the MRI dataset used, detailing its composition and preprocessing steps. The Techniques Applied section discusses the neural network architecture, training strategies, and evaluation methodology. The Results section presents the model's performance metrics, confusion matrices, and visualizations of learned features. Applications consider how these findings may translate into real-world clinical and research settings. Finally, the Conclusion summarizes the key insights gained and suggests directions for future work.

3 RELATED WORKS

Deep learning approaches have increasingly been applied to Alzheimer's detection and classification from MRI data, often achieving promising results. Oh et al. (2019) developed deep convolutional neural networks to classify Alzheimer's disease stages and offered interpretable visualizations of their models, highlighting the importance of explainability in medical AI. Islam and Zhang (2018) adopted ensemble systems of CNNs to improve Alzheimer's detection, demonstrating how multiple models can complement each other to enhance accuracy and robustness.

Korolev et al. (2017) compared plain and residual convolutional neural networks for the classification of 3D brain MRI volumes. Their findings suggested that deeper and more advanced architectures, such as residual networks, can effectively capture the complex, volumetric features associated with neurodegenerative changes. Basaia et al. (2019) designed automated classification systems leveraging deep neural networks to distinguish Alzheimer's disease and mild cognitive impairment from a single MRI, streamlining the

diagnostic process without relying heavily on manual feature extraction. Wen et al. (2020) conducted a reproducible evaluation of various CNN architectures, emphasizing rigorous validation and transparency in reporting results to ensure reproducibility and comparability across studies.

While these studies represent significant strides, many emphasize binary classification tasks, primarily differentiating between healthy controls and Alzheimer's patients. Fewer works have tackled the more nuanced challenge of multi-class classification. Yet, understanding intermediate stages—very mild or mild dementia—could be key to slowing disease progression. The present work builds upon these foundational efforts by developing a multi-class CNN classifier that attempts to map the subtle gradations of Alzheimer's disease and identify early indicators that might guide timely interventions.

4 DATA SET

This research utilizes the OASIS MRI dataset, a publicly available resource widely used in Alzheimer's disease studies. The dataset comprises over 86,000 MRI images sourced from 461 patients. Each patient's MRI volume has been sliced to generate multiple 2D images along a given axis, ensuring data in a format amenable to CNN-based image classification techniques. The dataset is organized into four categories corresponding to the disease stage:

1. **Non demented**: Representing most of the dataset (~77.77%), this class includes MRI scans from individuals without clinical signs of dementia. Such images often serve as a baseline, reflecting normal variations in brain structure and aging.
2. **Very mild dementia**: Accounting for roughly 15.88% of the images, this category encompasses patients who exhibit extremely subtle cognitive impairments that may precede clinically diagnosed Alzheimer's. Detecting shifts at this stage could greatly influence early intervention strategies.
3. **Mild dementia**: Comprising about 5.79% of the images, this class corresponds to more apparent but still relatively early cognitive decline. Patients may display mild memory issues or difficulty with complex tasks. Distinguishing mild dementia from very mild dementia is challenging due to subtle morphological changes.
4. **Moderate dementia**: Representing only about 0.56% of the dataset, these images capture more pronounced structural brain changes consistent with moderate-stage Alzheimer's disease. The extreme minority representation of this class presents significant challenges for training, as CNNs may be biased toward classes with greater representation.

All images in the dataset share consistent dimensional properties (originally around 496x248 pixels), facilitating standardized preprocessing. Before training, images are resized to 128x128 pixels to reduce computational overhead. The pixel intensity values are normalized by dividing by 255.0, ensuring that all image data lie within [0,1] and preventing large gradients during training.

Due to the severe class imbalance, the dataset's utility hinges on employing strategies that help the model treat minority classes with sufficient importance. Class weighting provides a straightforward mechanism to penalize the model more heavily for misclassifying underrepresented classes. Additionally, stratified splits are used to maintain class proportions across training, validation, and test sets. The

training set was formed by taking 85% of the data, with the remainder split into validation (15% of the training set) and test subsets (15% of the total data) to ensure robust performance estimates.

Leveraging the code and scripts stored in a dedicated GitHub repository, the dataset was loaded, preprocessed, and partitioned. Automated integrity checks confirmed consistent image dimensions, resolved any corrupted files, and ensured proper labeling. Exploratory analysis revealed interesting intensity patterns—non demented brain scans appeared slightly different in average pixel intensity distributions compared to mild and moderate dementia classes. Although subtle, these differences may be key indicators that the model can exploit to distinguish cognitive states.

5 MAIN TECHNIQUES APPLIED

The core methodology involves leveraging transfer learning with a pre-trained CNN architecture. This choice was influenced by the fact that training a deep model from scratch on a relatively limited medical dataset can be challenging. By initializing the model's convolutional layers from a network pre-trained on ImageNet—a large and diverse dataset of natural images—we can capitalize on robust, low-level feature representations, such as edges and textures, which can transfer effectively to the MRI domain.

A VGG16-based architecture served as the backbone of the model. The initial convolutional layers of VGG16 were frozen during training, preserving learned feature detectors. On top of these layers, custom dense layers were added, culminating in a softmax output layer for four-class classification. The code, maintained in my public GitHub repository, executed the following steps:

1. **Model Construction**:
   The pre-trained VGG16 network (without its top classification layers) was loaded and integrated into a Sequential model. New layers—Flatten, Dense, Dropout, and a final Dense output layer—were appended. Only these newly added layers and the top convolutional blocks were trainable, enabling the network to adapt generic features to Alzheimer's-specific patterns while maintaining computational efficiency.

2. **Optimization and Loss**:
   The Adam optimizer was selected for its efficiency in handling sparse gradients and adaptive learning rates. A low learning rate (e.g., 0.0001) was used to fine-tune the model gently. The categorical cross-entropy loss function ensured that the network's probability outputs aligned with the one-hot encoded labels.

3. **Class Weighting**:
   Given the severe imbalance favoring the non demented class, class weights were applied. For example, the moderate dementia class was assigned a higher weight to emphasize its underrepresented status. These weights ensure that misclassifying a moderate dementia image incurs a greater penalty than misclassifying a non demented image. Over successive iterations, this approach encourages the model to learn distinguishing features of minority classes.

4. **Early Stopping and Regularization**:
   Early stopping monitored the validation loss and halted training when no improvement was observed after several epochs. This practice prevented overfitting and reduced the risk of

memorizing training examples. Dropout layers were incorporated into the fully connected part of the architecture to further mitigate overfitting by randomly deactivating neurons during each training iteration, thus encouraging robust feature representations.

5. **Evaluation and Visualization**: After training, the model's performance was evaluated on the test set. Various metrics—accuracy, precision, recall, and F1-score—were computed for each class. A confusion matrix offered insights into common misclassifications and class overlap. Additionally, Grad-CAM (Gradient-weighted Class Activation Mapping) techniques were employed to visualize which regions of the MRI the model relied upon most heavily when making its predictions. This step improves model interpretability and can help identify clinically relevant brain regions related to each disease stage.

The final trained model was saved, along with its training history, for subsequent analysis. The code facilitated a reproducible pipeline, ensuring that experiments could be retraced, modified, and improved iteratively. This systematic methodology allowed careful control of hyperparameters and architectural decisions, ultimately yielding a model capable of learning subtle structural features indicative of cognitive decline.

6 KEY RESULTS

The trained CNN model achieved strong overall performance, surpassing baseline expectations and demonstrating its capacity to detect subtle shifts in brain structure correlating with Alzheimer's disease stages. All results were obtained from the final model after tuning hyperparameters and applying class weighting. The test set, held out from training and validation, provided an unbiased estimate of generalization.

Overall Metrics:
On the test set, the model attained an overall accuracy of approximately 89%. This high-level metric, while useful, does not fully capture class-specific nuances. Nonetheless, achieving close to 90% accuracy in a multi-class classification problem with extreme class imbalance is encouraging. The network not only distinguished clearly between non demented and more advanced stages but also managed to identify a substantial number of very mild dementia cases, a finding that supports the feasibility of early detection.

Class-Specific Performance:
Precision, recall, and F1-scores were computed for each class:

- **Non demented:** This majority class was predicted with a precision of around 92% and a recall of about 94%. The strong performance here was expected, given the abundance of training examples. High recall indicates that the model rarely missed non demented samples. Some misclassifications occurred when very mild dementia cases were mistaken as non demented, reflecting subtle differences that can be challenging to detect.
- **Very mild dementia:** Distinguishing this class from non demented was a key objective. The model achieved a precision of roughly 85% and a recall of about 82% for very mild dementia. These metrics indicate that the model identified early changes consistently. The success in detecting very mild dementia is particularly noteworthy, given that clinical

differentiation at this stage is often difficult.

- **Mild dementia:**
  For mild dementia, the model attained around 78% precision and 75% recall. While these figures are slightly lower than those for the very mild dementia class, they still represent a substantial improvement over random guessing. Confusions commonly arose between very mild dementia and mild dementia, suggesting that the anatomical differences between these stages are subtle. Nonetheless, capturing such fine-grained differences remains a valuable step forward.
- **Moderate dementia:**
  Despite comprising the smallest fraction of the dataset, the moderate dementia class achieved approximately 65% precision and 60% recall. While this performance is modest compared to the other classes, it is remarkable given that less than 1% of the dataset fell into this category. Class weighting proved beneficial here, guiding the model to pay increased attention to moderate cases. Identifying moderate dementia reliably can guide clinicians in recognizing patients who may require more intensive management or support.

Confusion Matrix Insights:

A confusion matrix provided a more granular view of performance. It confirmed that non demented and moderate dementia were the easiest to distinguish. Very mild and mild dementia classes overlapped considerably, sometimes confusing one stage for the next. This pattern aligns with the clinical reality that these transitional states may not have pronounced imaging biomarkers distinguishable with current deep learning models and MRI data alone.

ROC Curves and AUC:

Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) metrics supported these findings. The model achieved AUCs above 0.90 for the non demented and very mild dementia classes, indicating high discriminative capability. Mild dementia and moderate dementia displayed AUCs in the range of 0.85 to 0.88. Although lower than those of the other classes, these AUC values still reflect a model that is learning meaningful patterns. In a clinical setting, such discrimination may justify more targeted follow-up screenings or assessments.

Grad-CAM Visualization:

Grad-CAM was employed to highlight the regions in MRI scans that most influenced the model's predictions. Visualizations commonly showed heightened attention to regions associated with memory and cognition—structures such as the medial temporal lobe or hippocampal areas. For very mild dementia scans, the model often focused on subtle tissue density variations in these critical regions. Such insights not only enhance interpretability but could also inspire further neuroimaging research into early anatomical markers of Alzheimer's.

These results collectively highlight the promise and limitations of CNN-based multi-class classification in Alzheimer's detection. While the model excels in identifying non demented and very mild dementia classes, distinguishing mild and moderate stages remains more challenging. Future refinements might include supplementing MRI with additional biomarkers, using more advanced architectures (e.g., 3D CNNs or transformers), or integrating clinical metadata to improve specificity. Nevertheless, these findings affirm that deep learning can assist in earlier and more nuanced detection of Alzheimer's progression stages, aligning well with the original motivation of supporting

patients, clinicians, and caregivers in timely decision-making.

7 APPLICATIONS

The potential impact of these findings on clinical practice and research is significant. By pushing beyond binary classification to a four-class model that includes very mild and mild dementia stages, this research offers a tool that can better reflect the complex continuum of Alzheimer's progression.

1. **Clinical Decision Support:** Automated classification systems could serve as valuable assistants to radiologists and neurologists. Given a patient's MRI scan, such a tool can quickly and objectively gauge whether subtle changes indicative of very mild dementia are present. This early warning could prompt additional assessments, lifestyle interventions, or therapies during a window when treatment may be most beneficial.
2. **Personalized Patient Monitoring:** Patients already diagnosed with mild dementia could benefit from periodic MRI scans analyzed by this system. Tracking changes over time may help clinicians understand how quickly a patient's disease is progressing and adjust care plans accordingly. Early detection of a transition from very mild to mild dementia might inspire interventions to maintain cognitive function and independence for as long as possible.
3. **Enhancing Clinical Trials and Research:** Stratification of patient populations in clinical trials often relies on accurate staging of the disease. By providing a more nuanced classification, researchers can better select participants at specific stages of cognitive decline. This can refine the evaluation of potential treatments and improve the sensitivity of detecting drug effects. For example, recruiting participants at very mild dementia stages might increase the likelihood of observing a treatment's efficacy before the disease has advanced too far.
4. **Healthcare Resource Allocation:** As Alzheimer's disease incidence grows, healthcare systems must allocate resources efficiently. Automated screening tools can help identify patients who need more specialized follow-up, while those with less severe findings may remain in general monitoring programs. This stratification could ultimately reduce costs and ensure that specialized resources—geriatricians, advanced imaging modalities, and interventions—reach those who need them most.
5. **Integration with Multimodal Data:** In the future, combining CNN-based MRI classification with other diagnostic indicators, such as cerebrospinal fluid biomarkers, PET imaging, or cognitive test scores, may produce even more robust predictive models. Such multimodal systems could overcome the inherent limitations of single-modality data and improve predictive power, offering a more holistic view of a patient's neurological health.

8 CONCLUSION

This research demonstrates that deep learning techniques, specifically CNN-based models with transfer learning, can classify MRI brain images into multiple Alzheimer's disease stages with a high degree of accuracy. By moving beyond a

binary classification framework, the study provides a more nuanced understanding of disease progression, aligning with clinical realities where cognitive decline unfolds gradually rather than suddenly.

Through extensive preprocessing, careful handling of class imbalance, and the application of class weighting, the model successfully differentiated non demented and very mild dementia with notable accuracy and sensitivity. Although distinguishing mild and moderate dementia remains a challenge, the results indicate that even subtle structural changes may be detected. The integration of Grad-CAM interpretation highlighted that the model's focus aligns with neuroanatomical regions known to be associated with memory and cognition, increasing the plausibility and clinical relevance of the model's decision-making process.

The personal motivation behind this research—my experience as a caregiver for my grandmother—underscores the importance of developing tools that can support early detection and intervention. While no model currently can replace clinical judgment or serve as a standalone diagnostic tool, the methods and insights presented here pave the way for improved screening programs, more precise participant selection in clinical trials, and more tailored patient management strategies.

Moving forward, multiple avenues remain for future improvement. Incorporating 3D convolutional layers could capture volumetric information lost in 2D slices. Using more advanced architectures like residual networks or transformers might enhance discriminative power. Integrating multimodal data sources, including genetic factors, laboratory results, or cognitive test scores, could refine predictions and better explain individual patient trajectories. Ultimately, the goal remains to bridge the gap between early radiological manifestations of Alzheimer's disease and timely clinical interventions that preserve patient function and quality of life.

9 REFERENCES

[1] Oh, K., Chung, Y. C., Kim, K. W., & Kim, W. S. (2019). Classification and visualization of Alzheimer's disease using deep convolutional neural networks. Scientific Reports, 9, 18150.

[2] Islam, J., & Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Informatics, 5(2), 2.

[3] Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (pp. 835-838). IEEE.

[4] Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., & Filippi, M. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage: Clinical, 21, 101645.

[5] Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-Gonzalez, J., Routier, A., Bottani, S., D., Dormont, S., Durrleman, N., Burgos, & Colliot, O. (2020). Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Medical Image Analysis, 63, 101694.