Predicting Alzheimer's Disease with Deep Learning

Nina Ebensperger, Karina Martinez, Edison Murairi

December 9, 2023

1 Introduction

Alzheimer's disease is a progressive neurodegenerative disorder characterized by a decline of cognitive functions. It is the most common case of dementia, affecting millions of individuals globally [4]. The disease is generally characterized by an accumulation of β-amyloid (beta-amyloid) plaques and τ-protein (tau-protein) tangles in the brain [13]. The disease has rapidly become a global health issue, and early as well as accurate diagnosis is crucial in implementing proactive interventions. However, most patients experience a sporadic form with a delayed onset [12], making an early diagnosis challenging.

Deep learning has the potential to address this issue. In recent years, deep learning models have shown remarkable capabilities in medical diagnosis, including detecting dementia. For example, Liu and colleagues [1] used a 3D convolutional neural network to distinguish between mild Alzheimer's disease dementia from mild cognitive impairment and cognitively normal individuals using structural MRIs. Moreover, Ref. [5] provides a comparative study of the diagnostic performance of 2D, 3D convolutional neural networks and recurrent neural networks when applied to MRI scan of the brain. This study provides a significant stepping stone into understanding the type of deep learning models required to effectively address this problem.

Our project aims at contributing to the use of deep learning models in stratifying patients with the disease. More specifically, we use the OASIS dataset [10] to compare the performance of three different deep learning models in predicting the stage of dementia in a patient based on the brain's MRI scans and metadata provided in the dataset. The rest of this work is as follows: Sec. 2 will describe the dataset while Sec. 3 will discuss the models that we used. Sec. 4 will detail experimental design, and Sec. 5 will discuss the results. Finally, Sec. 6 will summarize our findings and provide a conclusion.

2 Dataset

The dataset consists of 86,437 MRI images from the Oasis Brains project, an open access resource aimed at making neuroimaging data sets of the brain freely available to the scientific community. Images were retrieved from the Kaggle dataset "OASIS Alzheimer's Detection" and metadata was accessed directly from the study "OASIS-1: Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults [10]."

The dataset includes patients aged 18 to 96, divided into four classes based on Alzheimer's progression. Of the 366 total patients, 285 are non-demented, 58 have very mild dementia, 21 have mild dementia and 2 have moderate dementia. The subjects are all right-handed and include both men and women. The associated metadata also includes information on patient demographics, clinical information and derived anatomic volumes (Table 1).

Image files were converted from original .img and .hdr formats into Nifti format (.nii) using FSL (FMRIB Software Library) and the .nii MRI scans were subsequently converted to .jpg files by the Kaggle author. The brain images were sliced along the z-axis into 256 pieces, and slices ranging from 100 to 160 were selected from each scan.

Variable	Values
Gender (M/F)	M: Male F: Female
Education (Educ)	0-5
Socioeconomic Status (SES)	1: less than high school grad 2: high school grad 3: some college 4: college grad 5: beyond college
Age	number of years old
Mini-Mental State Examination (MMSE)	score in range 0-30
Estimated total intracranial volume (eTIV)	intracranial volume in mm^3
Atlas scaling factor (ASF)	head size normalization factor
Normalized whole brain volume (nWBV)	brain volume normalized for head size
Clinical Dementia Rating (CDR)	0: non-demented 0.5: very mild dementia 1: mild dementia 2: moderate dementia

Table 1: Metadata descriptions

2.1 Pre-processing

The dataset consists of 78% non-demented patients, so it is highly imbalanced. However, all of the subjects diagnosed with very mild to moderate Alzheimer's disease are over the age of 60. Therefore, we restricted the age of the patients to 60 and over and reduced the proportion of non-demented patients to 85 out of 166 total.

Because there were only 2 patients in the moderate dementia group, there was no way to split the dataset into stratified train, test and validation splits. We removed the moderate dementia patients from the dataset and were left with 3 target classes, 164 patients and 39,406 images. The final proportion of subjects was 52% non-demented, 35% very mildly demented, and 13% mildly demented.

The dataset was divided into 70% train, 15% validation and 15% test splits. To avoid data leakage, patients were stratified so that all images for each patient were included in only one split.

3 Neural Networks

3.1 Gated MLP (GMLP)

GMLP is an archiecture introduced in Ref. [9] as an alternative to transformers [14]. Transformers have enjoyed a great success in computer vision and natural language processing (NLP). One of their main features is their inductive bias, allowing dynamic parametrization of spatial interactions based on the input [3]. GMLPs provide a different construction, where the so-called Spatial Gating Unit accounts for the spatial interactions.

Figure 1 shows the GMLP architecture. First, the input is passed normalized and passed through a channel projection unit. Let the input to the channel projection unit be Z. The projection unit performs the operation $Z\,U$ where U is a projection operator[1]. Then, the result is passed through the activation unit, producing an output $\sigma(ZU)$. For the activation function, the authors recommend using the Gaussian Error Linear Unit (GeLU) [8]. Arguably, the most important element of the network is the Spatial Gating Unit. First, the output of the activation function is split into Z_1 and Z_2. Z_2 is passed through the normalizer, and then through the spatial projection unit. We will denote the output of the normalizer by Z_2 as well. The spatial projection unit is problem-dependent. The simplest choice is to use a linear operation such as $Z_2' = W\,Z_2 + b$ where W is a weight matrix and b is a bias.

[1] An operator U is a projection if and only if $U^2 = U$.

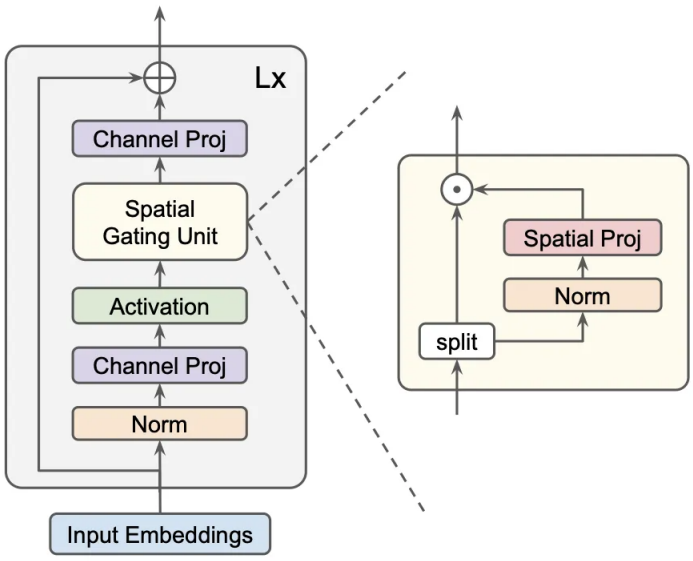

Figure 1: Gated MLP Architecture from Ref. [9]

Finally, Z_1 and Z_2' are combined through a Hadamard product as $s(Z) = Z_1 \odot Z_2'$. Finally, $s(Z)$ is passed through a projection unit, yielding $s(Z)V$ where V is a projector, and the input to the network is then added to the output of the last channel projection. To finish, we note that the whole block has to be repeated L times.

3.2 Residual Network (ResNet) with Metadata

The availability of diverse biomedical data, such as medical imaging, genome sequencing and electronic health records, has paved the way for the emergence of multimodal artificial intelligence in health. Although most applications to date have focused on one data modality, there is a growing trend toward the use of deep learning architectures to integrate multiple data types [2]. Inspired by a paper on feature fusion using neuroimages from brain tumors, we aimed to combine the output of a convolutional neural network with demographic and derived anatomic volumes [11].

Residual Networks were first introduced in 2015 by He and colleagues [7] and have become widely used in computer vision tasks, and particularly image classification. These deep convolutional neural networks effectively address the challenge of training very deep neural networks through the use of residual blocks, which contain skip connections. By skipping over some layers, residual networks mitigated the vanishing gradient problem that results when many convolutional layers are stacked together.

ResNet50 is comprised of 50 layers, with a modular architecture that allows for stacking and repeating blocks to achieve its depth. Due to its success on various image recognition benchmarks, ResNet50 is often used as a pre-trained model for transfer learning in various computer vision applications. For this reason, a ResNet50 architecture was used as the backbone for image feature extraction.

The ResNet50 classifier was replaced with a concatenation step, where the tabular metadata features were fused with the image features. The combined features were then connected to a 5-layer fully connected network (FCN). Each step of the FCN included a dropout layer, batch normalization and ReLU activation function. The ResNet with Metadata architecture is presented in Figure 2.

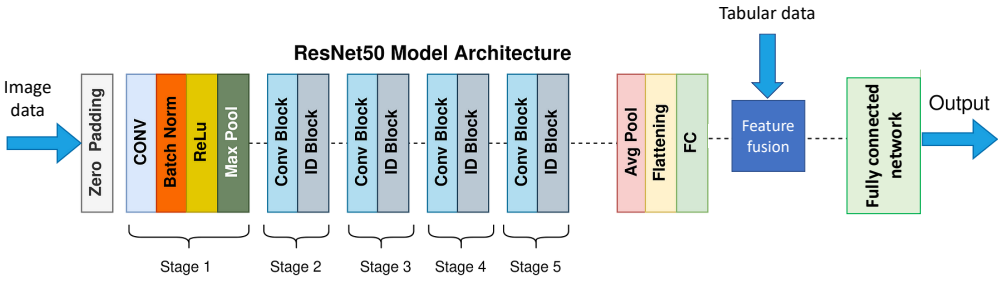

Figure 2: ResNet50 with Metadata architecture. Adapted from

3.3 Convolutional Neural Network (CNN) with Attention Mechanism

3.3.1 Individual Contributions and Algorithm Development

My individual contribution centered on constructing a CNN architecture enriched with CBAM, which introduces channel and spatial attention sequentially. Channel attention prioritizes informative features, and spatial attention concentrates on significant spatial locations within the feature maps. The interaction between these modules is governed by element-wise multiplication of the attention maps with the feature maps.

3.3.2 Attention Mechanisms in CNN

The architecture of the CNN incorporates the Convolutional Block Attention Module (CBAM) at different stages to progressively refine feature representation. The channel attention module directs the model's focus to informative features across channels, while the spatial attention module emphasizes important spatial locations in the feature maps. The equations governing the attention mechanisms are as follows:

$$F' = M_c(F) \otimes F, \tag{1}$$

$$F'' = M_s(F') \otimes F', \tag{2}$$

where F denotes the input feature maps, M_c is the channel attention map, M_s is the spatial attention map, and $\otimes$ represents element-wise multiplication. [15]

3.3.3 Channel Attention

Purpose: To determine which channels (or features) in the feature map are more important.

- It starts by applying both max pooling and average pooling across the spatial dimensions of the feature map. These two pooled features capture different aspects of the feature maps.
- Both pooled features are then passed through a shared multilayer perceptron (MLP) which has one hidden layer. The MLP works as a channel descriptor, learning which channels are more important.
- The outputs from the MLP for both max and average pooled features are then combined using element-wise summation.
- A sigmoid function is applied to the combined features, resulting in a channel attention map where each value is between 0 and 1. This attention map is broadcasted and multiplied with the original feature map to scale the channels accordingly.

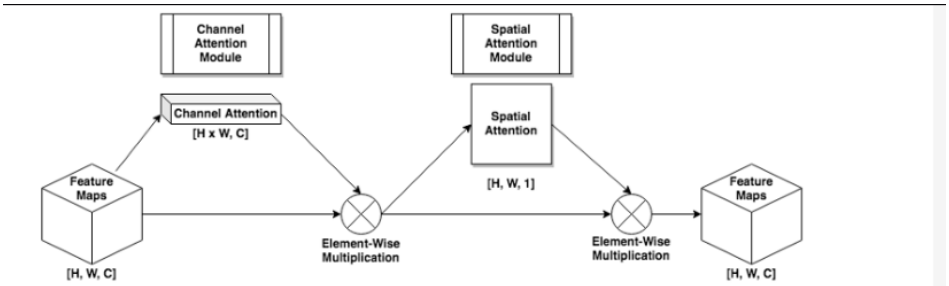

Figure 3: Block diagram of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

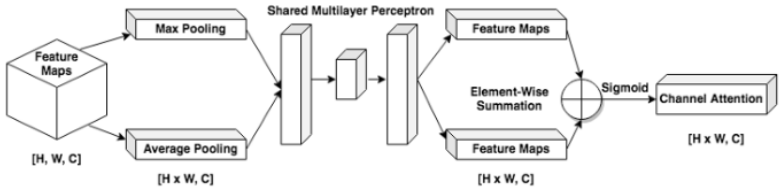

Figure 4: Block diagram of the *Channel Attention Module* of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

Figure 3: Graph illustrating the model's training progress over epochs. Adapted from [6].

Purpose: To determine which channels (or features) in the feature map are more important.

- It starts by applying both max pooling and average pooling across the spatial dimensions of the feature map. These two pooled features capture different aspects of the feature maps.
- Both pooled features are then passed through a shared multilayer perceptron (MLP) which has one hidden layer. The MLP works as a channel descriptor, learning which channels are more important.
- The outputs from the MLP for both max and average pooled features are then combined using element-wise summation.
- A sigmoid function is applied to the combined features, resulting in a channel attention map where each value is between 0 and 1. This attention map is broadcasted and multiplied with the original feature map to scale the channels accordingly.

3.3.4 Spatial Attention

Purpose: To determine which spatial regions of the feature map should be emphasized.

- After channel attention has been applied, the model computes the average and max features across the channel dimension, which are then concatenated together.
- This concatenated feature map is then passed through a convolutional layer followed by a sigmoid function to produce a spatial attention map. This map is a single-channel 2D map where higher values indicate regions of interest.
- The spatial attention map is then used to scale the feature map after channel attention has been applied.

3.3.5 Combining Channel and Spatial Attention

- The CBAM module applies channel attention first to the input feature map, producing a feature map that has been scaled channel-wise.

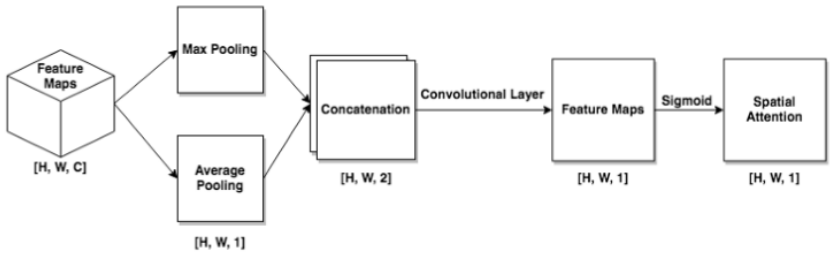

Figure 5: Block diagram of the *Spatial Attention Module* of the *Convolutional Block Attention Module* (CBAM), according to Woo et al. [2018].

Figure 4: Graph illustrating the model's training progress over epochs. Adapted from [6].

- It then applies spatial attention to this channel-refined feature map, producing the final output feature map that has been refined both channel-wise and spatially.
- These attention mechanisms allow the neural network to adaptively focus on important features and suppress less relevant ones, potentially improving the performance of the network for tasks such as classification, detection, or segmentation.

4 Experimental Setup

5 Results

5.1 GMLP

Figure 5 shows the loss and accuracy on training set and the test set for ten epochs. The plot clearly shows that although the loss on the training set is decreasing, it is increasing on the validation set. This result is concerning because it is suggesting that the model does not generalize well. Moreover, the accuracy is also increasing on the training set while it decreases (or almost remains constant on the validation set). These two results strongly indicates that this model overfits the training set, and does not generalize. Table 2 shows the performance metrics, confirming that the model does not perform well.

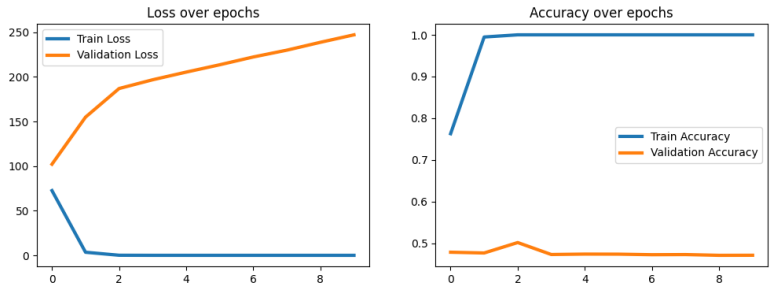

Figure 5: Loss and Accuracy on training set and validation set over 10 epochs

Metric	Value
Accuracy	0.51
F1 Score	0.37
COH Score	0.06

Table 2: Performance Metrics of the GMLP model

This overfitting result is probably due to the class in-balance. This diagnostic is made even clearer when examining the confusion matrix, Fig. 6. In this confusion matrix, 0 refers to 'No dementia', 1 refers to 'Very mild dementia' and 2 refers to 'Mild dementia'. The confusion matrix shows that the model overpredicts 'No dementia'. More specifically, the model is clearly not capable of distinguishing between 'No dementia' and 'Very mild dementia'.

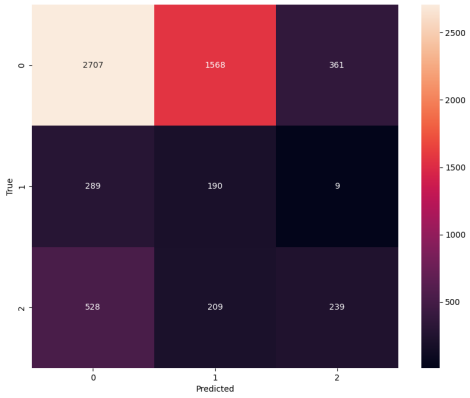

Figure 6: GMLP Confusion Matrix. 0 refers to 'No dementia', 1 refers to 'Very mild dementia' and 2 refers to 'Mild dementia'.

5.2 ResNet with Metadata

The ResNet50 model was trained with the three derived metadata features: eTIV, ASF and nWBV. The results presented in Figure 7 show that the validation loss was actually lower than the training loss, while the training accuracy was higher than that of the validation. Possible reasons for this are that the validation set is easier than the training set, the validation set is too small, or it wasn't properly sampled. Because each of the MRI images was treated as a separate observation, the same patient metadata was presented multiple times as slices from the same MRI scan were presented to the model. The pleateau of the train loss, and the low accuracy of both sets could also be due to sub-optimal concatenation of the features from the image and tabular data.

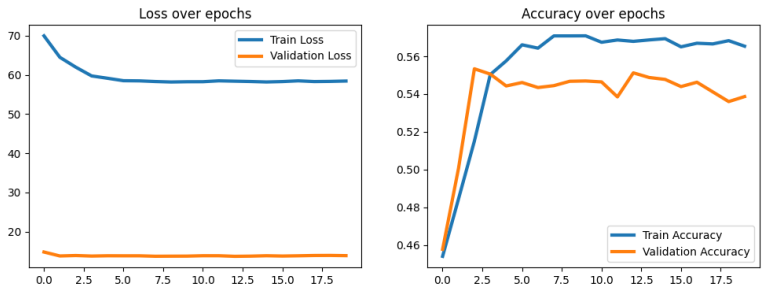

Figure 7: Loss and Accuracy on training set and validation set over 20 epochs

Metric	Value
Accuracy	0.58
F1 Score	0.35
COH Score	0.09

Table 3: Performance Metrics of the ResNet with Metadata model

The accuracy score of the ResNet plus metadata model was the highest of the three models, while the F1 and Cohen's Kappa scores fell in between the other models (Fig 3). The confusion matrix indicates that this model performed best on the majority class, non-demented, but underperformed on the minority classes compared to the other models (Fig 8). The high number of false negatives in the minority groups indicates that this model was biased toward the majority group, and would not be helpful as an early prediction tool for dementia.

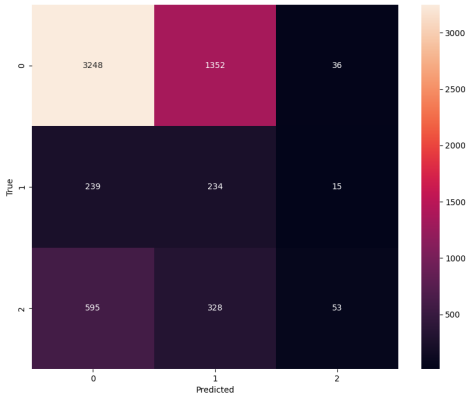

Figure 8: ResNet with Metadata Confusion Matrix. 0 refers to 'No dementia', 1 refers to 'Very mild dementia' and 2 refers to 'Mild dementia'.

5.3 CNN with Attention

Upon reevaluation of the model's training dynamics, signs of overfitting were identified. As shown in Figure 9, the model's training loss decreased significantly over epochs, suggesting improved learning on the training dataset. However, the validation loss plateaued and began to slightly increase, indicative of the model's diminishing ability to generalize to unseen data. Similarly, the training accuracy approached near-perfection, while the validation accuracy stagnated at a lower level, reinforcing the presence of overfitting.

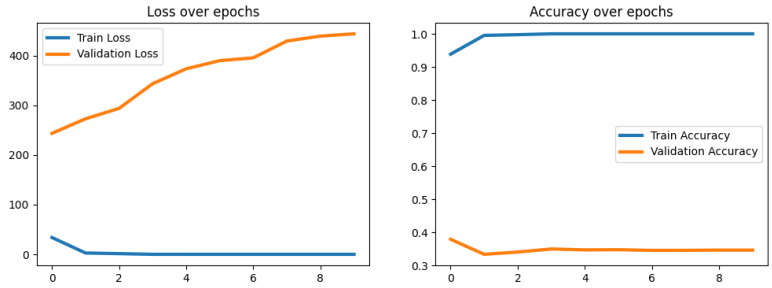

Figure 9: Loss and accuracy over epochs indicating overfitting: validation metrics do not improve in tandem with training metrics.

The recalibrated models underwent retraining with the revised dataset, yielding new performance metrics that painted a more authentic picture of the models' capabilities. The confusion matrix, depicted in Figure 10, provided a detailed breakdown of the model's predictive accuracy across Alzheimer's disease stages, showcasing an improved balance between sensitivity and specificity.

Furthermore, Grad-CAM visualizations were employed to understand the model's focus areas during prediction. As depicted in Figure 5, the model appears to be concentrating on the skull and the space between the skull and the brain. This unexpected focus raises concerns about the interpretability of the model's learning and the potential need for further refinement to ensure that medically relevant features within the brain are being adequately considered.

This refined approach led to a more reliable assessment of the model's performance and underscored the critical role of rigorous validation in the development of machine learning models for medical applications.

Metric	Value
Accuracy	0.50
F1 Score	0.39
COH Score	0.10

Table 4: Performance Metrics of the CNN + Attention model

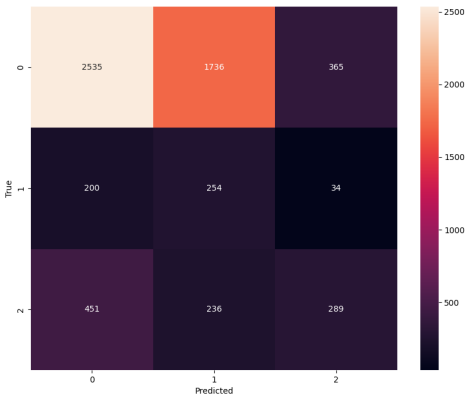

Figure 10: Confusion matrix showcasing the model's predictive accuracy across different stages of Alzheimer's disease after addressing overfitting.

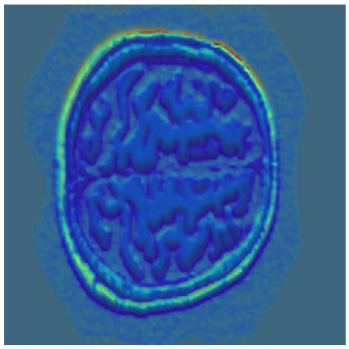

Figure 11: Grad-CAM visualization highlighting the regions of interest within a brain MRI scan that the CNN model focused on when predicting the presence of Alzheimer's disease. Areas in warmer colors indicate higher influence on the model's decision.

6 Summary and Conclusion

The study aimed to predict Alzheimer's disease stages using deep learning models with a focus on Gated MLP (GMLP), Residual Network (ResNet) with Metadata, and Convolutional Neural Network (CNN) with Attention Mechanism. The models were evaluated on the OASIS dataset, consisting of MRI images and metadata.

The GMLP model demonstrated overfitting, as evidenced by increasing validation loss and poor generalization. This overfitting may be attributed to class imbalance, particularly in distinguishing 'No dementia' and 'Very mild dementia.'

The ResNet with Metadata showed promise, but further refinement is needed to improve its predictive accuracy. The CNN with Attention demonstrated overfitting initially, which was addressed through model recalibration. The revised model exhibited improved performance, emphasizing the importance of rigorous validation in medical applications.

However, the study underscores the need for interpretability in deep learning models for medical diagnosis. Attention mechanisms and Grad-CAM visualizations revealed unexpected model focus areas, prompting consideration for refining features and ensuring medical relevance.

The issues we have faced may be addressed by artificially generating examples of the under-represented classes so that the models may have a better chance at learning to classify them. In addition, it is possible to aggregate the model predictions for all the MRI slices of a unique patient. One way to perform this aggregation is by taking the majority prediction as the prediction for a given patient. Finally, it is possible to use models such as 3D CNNs to directly account for the spatial relationship of the brain slices for each patient.

In conclusion, while deep learning models show potential for predicting Alzheimer's disease stages, careful model selection, addressing overfitting, and interpretability are crucial aspects for their successful application in medical contexts. Future work should focus on refining model architectures, incorporating domain knowledge, and ensuring robust interpretability for reliable clinical use.

References

[1] Generalizable deep learning model for early alzheimer's disease detection from structural mris. *Scientific Reports*, 12(1):17106, 2022.

[2] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol. Multimodal biomedical ai. *Nature Medicine*, 28(9):1773–1784, Sept. 2022.

[3] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. 2014.

[4] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones. Alzheimer's disease. *Lancet*, 377(9770):1019–1031, Mar. 2011.

[5] A. Ebrahimi, S. Luo, and Alzheimer's Disease Neuroimaging Initiative. Convolutional neural networks for alzheimer's disease detection on MRI images. *J Med Imaging (Bellingham)*, 8(2):024503, Apr. 2021.

[6] T. Gonçalves, I. Rio-Torto, L. F. Teixeira, and J. S. Cardoso. A survey on attention mechanisms for medical applications: are we moving towards better algorithms?, 2022.

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

[8] D. Hendrycks and K. Gimpel. Gaussian error linear units (GELUs). 2016.

[9] H. Liu, Z. Dai, D. R. So, and Q. V. Le. Pay attention to MLPs. 2021.

[10] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 19(9):1498–1507, Sept. 2007.

[11] D. Nie, J. Lu, H. Zhang, E. Adeli, J. Wang, Z. Yu, L. Liu, Q. Wang, J. Wu, and D. Shen. Multi-channel 3d deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Scientific Reports*, 9(1), Jan. 2019.

[12] C. Pinel. Alzheimer's disease. *Nurs. Times*, 71(3):105–106, Jan. 1975.

[13] P. Scheltens, K. Blennow, M. M. B. Breteler, B. de Strooper, G. B. Frisoni, S. Salloway, and W. M. Van der Flier. Alzheimer's disease. *Lancet*, 388(10043):505–517, July 2016.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. 2017.

[15] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module, 2018.