Intersect, Vol 17, No. 2 (2024)

Evaluation of Deep Learning Models for Early-Stage Alzheimer's Disease Screening

Veer Gupta
Advised by: Dr. Parsa Akbari of the University of Cambridge
Morris Hills High School, Parsippany, New Jersey

Abstract

Alzheimer's disease, a progressive neurodegenerative disorder, poses significant challenges in its early-stage identification and accurate prediction of disease progression. In the early stages of Alzheimer's, patients are usually undetected because symptoms are difficult to identify and previously, no treatments for Alzheimer's were available. In 2022 and 2023 the FDA approved three drugs for treating Alzheimer's disease (Donanemab, Aducanumab, and Lecanemab), a breakthrough in treating the disease. The approved drugs have only been shown to be effective in patients with early cognitive decline. As a result, there is now an urgent need to screen millions of patients for Alzheimer's disease which presents significant financial challenges and substantial time allocation. Machine learning models have shown accuracy in detecting Alzheimer's and other disease outcomes from imaging data including MRI scans. However, previous work on Alzheimer's disease has often focused on identifying patients with late cognitive decline and often using older model architectures and methods. In this study, we train multiple vision-based machine learning models to detect patients in the early stages of Alzheimer's from MRI imaging data. We overcome technical challenges including class imbalance which diminishes model performance. Furthermore, we apply transformer-based models that utilize an attention-based mechanism that has led to breakthroughs in other areas including LLMs (ChatGPT). Our experiments show that ResNet-18 with class weighting, a batch size of 32, and a learning rate of 1E-06 was the optimal model for identifying early cognitive disease. This model produced F-scores as high as 0.984 for the MCI and EMCI stages, and 0.982 for cognitive normal patients.

Introduction

Alzheimer's disease presents a formidable challenge to modern healthcare systems, affecting approximately 6 million individuals in the United States alone. Additionally, the economic burden associated with Alzheimer's care was estimated to be around $355 billion in 2021, a figure that is projected to rise as the population ages (Abubakar et al. 2022).

Early diagnosis and intervention have emerged as crucial factors in managing the disease's debilitating effects and improving the quality of life for those affected. However, the road to effective early-stage identification has been fraught with difficulties, characterized by subtle symptoms that often go unnoticed. The landscape has recently shifted with the introduction of novel therapies, such as Donanemab, Aducanumab, and Lecanemab, which offer newfound hope for individuals in the early stages of cognitive decline ("High-Clearance Anti-Amyloid Immunotherapies in Alzheimer's Disease. Part 1: Meta-Analysis and Review of Efficacy and Safety Data, and Medico-Economical Aspects" 2022). These therapies are only proven effective when administered during the early stages of cognitive decline and have been designed to target the underlying biological processes that contribute to the development of Alzheimer's disease. By intervening at an early stage, when the pathological changes are still relatively mild, these drugs have the greatest potential to slow down or even halt the progression of the disease.

Historically, there were no treatments for patients in the initial stages of Alzheimer's disease. Therefore, there was no clinical need to identify patients with early cognitive impairment. However, these new therapeutics including Donanemab, Aducanumab, and Lecanemab have been shown to be safe and effective, in particular only in patients with early cognitive impairment. Therefore, a suitable way to identify these early stages is of utmost importance.

In this context, ResNet-18, short for Residual Network 18, is a type of deep learning model specifically designed for image recognition tasks. It's a part of a family of neural network architectures called ResNets, which are known for their effectiveness in training very deep networks. What makes ResNet-18 notable is its use of residual connections. These connections allow the model to skip certain layers during training, making it easier for the network to learn and understand complex patterns in images (Ullah et al. 2021). This architecture helps address the "vanishing gradient" problem, which can hinder the training of deep networks. A vanishing gradient arises when information about errors becomes too faint as it flows backward through many layers of a neural network. This can make it difficult for the model to learn effectively from data and can hinder its ability to recognize important patterns. ResNet-18 tackles this issue by introducing special connections and structures that help preserve and transmit this information more effectively, allowing for the training of deeper and more accurate neural networks.

Moreover, vision transformers (VITs) are emerging as a powerful class of models, and have demonstrated remarkable capabilities in various computer vision tasks, including image recognition, object detection, and other visual recognition tasks (Park and Kim 2022). Among the different variants of VITs, one particular model that has gained substantial attention is the DeiT mode. DeiT, short for "Data-efficient Image Transformers," aims to address the challenges of training large-scale transformer models on limited computational resources (Li et al. 2021).

Unlike traditional Convolutional Neural Networks (CNNs), VITs employ a transformer architecture, which leverages self-attention mechanisms to capture long-range dependencies within the image. This grants VITs the ability to process entire images as sequences, allowing for more comprehensive feature interactions. The DeiT model addresses challenges related to training large-scale transformer models on limited computational resources. DeiT utilizes self-attention and transformer-based architectures while integrating techniques for efficient parameter sharing, enabling data-efficient learning (Han et al. n.d.).

CNNs and Vision Transformers are different ways to help the computer learn. ResNet-18 is like a special type of picture learner that's good at recognizing different parts of a picture and putting them together. CNNs are a common type of picture learner that focuses on finding patterns in small parts of the picture and then combining those patterns to understand the whole picture. Vision Transformers, which include DeiT, have different ways of working. They look at the entire picture all at once and try to understand how different parts relate to each other. Each method has its way of processing image data, we use depends on what we want the computer to do and how we want it to learn. In this study, we apply both methods for the identification of patients with Alzheimer's who have early cognitive decline.

However, these models can run into problems caused by class imbalance. Class imbalance is a common challenge in machine learning, especially in medical applications like Alzheimer's detection. It occurs when one class, such as patients with early-stage Alzheimer's, is significantly underrepresented compared to others, like those with late-stage cognitive decline or without the condition. In such scenarios, machine learning models tend to be biased toward the majority class and may perform poorly on minority classes, which causes the accuracy of the models to decrease

The possibility of promising drug therapies for Alzheimer's disease comes with a significant challenge: accurately identifying patients who are in the initial stages of cognitive decline. The traditional diagnostic landscape has been marked by underdiagnosis and delayed intervention due to the elusive nature of early-stage symptoms. Early-stage symptoms include mild memory loss, difficulty with problem-solving, word-finding problems, decreased judgment, and difficulty completing familiar tasks (Mayeux 2010). Detecting these subtle changes and seeking medical evaluation is essential for timely intervention and effective management of the condition. The introduction of these novel drugs accentuates the urgency of addressing this challenge, emphasizing the need for a robust and reliable mechanism to identify and characterize individuals in the early stages of cognitive impairment.

Furthermore, patients with Early Mild Cognitive Impairment (EMCI) and Mild Cognitive Impairment (MCI), often exhibit subtle or even negligible signs of cognitive decline, making their accurate identification particularly challenging. Given the limitations of conventional diagnostic methods, there is a compelling need to explore innovative approaches that can effectively discern these nuanced cognitive changes.

Regarding the DeiT model, while it offers numerous advantages, it is essential to note that it still demands a significant amount of computational resources compared to traditional Convolutional Neural Networks (CNNs). The complexity of the transformer's structure and the large number of parameters in the DeiT model contribute to increased memory storage requirements and higher computational costs. Consequently, addressing the resource limitations associated with the models becomes crucial for their practical deployment and scalability.

Moreover, it is observed that the traditional CNN model, such as ResNet-18, often outperforms the DeiT and SimpleVit models in certain scenarios. This happens because the choice of the best model depends on various factors like the dataset size, data type, and the problem at hand. CNNs have been around for a while and are good at adapting to smaller datasets and require fewer computing resources. On the other hand, transformer-based models shine when you have a large and diverse dataset. So, it's important to carefully consider these factors when choosing a model, as there is no one-size-fits-all solution in machine learning.

Research Question/Objectives

The primary aim of this paper is to investigate and propose techniques to provide early detection to handle class imbalances, considering the unique characteristics of the ResNet-18 and resource constraints. By addressing this problem, the study seeks to enhance the performance and practicality of VITs in real-world applications, where class imbalance is a common occurrence.

To achieve our research objectives, we will explore the following questions:

1. Which model is the most reliable and can be used to provide earlier detection and why?
2. How does class imbalance impact the performance?
3. What techniques can be developed to mitigate the effects of class imbalance in VITs?
4. How does implementing a SimpleVit model compare to the ResNet-18 and standard DeiT models in terms of performance and resource utilization?

By investigating these research questions, this study aims to provide valuable insights into addressing the class imbalance in VITs and assess the potential of a SimpleVit model for improving the performance of VITs in real-world computer vision applications. The findings of this research endeavor can potentially guide practitioners and researchers in developing more effective and efficient models for computer vision tasks.

Results

Modern monoclonal antibody treatments for Alzheimer's have a preventative mechanism of action that relies on clearing toxic aggregates in the brain before major cognitive decline. To identify target patient populations for these treatments, patients must be screened at an early stage when little to no symptoms are present. Therefore, relevant screening data is likely to include a large proportion of healthy patients to identify a few patients who are in the early stages of cognitive decline. This is reflected in the composition of our training set (Fig. 1).

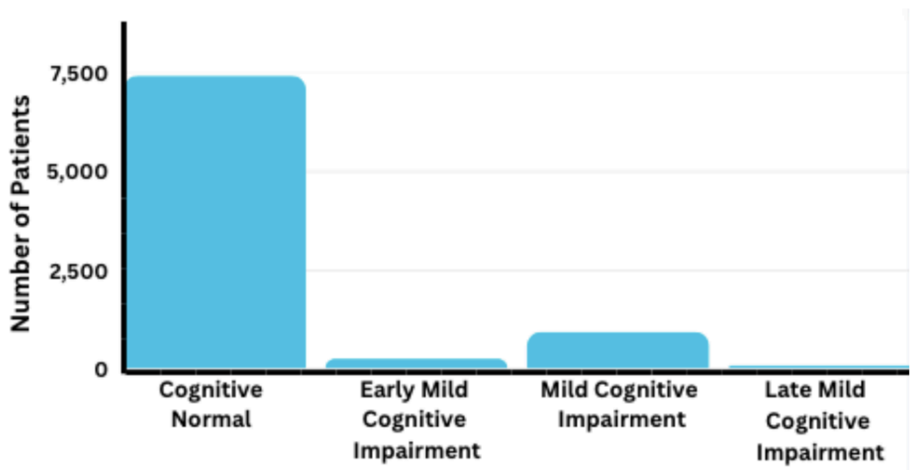


FIGURE 1. Distribution of diagnoses across patients in the dataset.

The majority of screened patients (7,430) show no signs of cognitive decline in the MRI imaging, with smaller proportions in the other stages. Patients are screened with MRI and classified into four categories including cognitively normal, Early Mild cognitive impairment (EMCI), Mild Cognitive impairment (MCI), and Late Mild Cognitive impairment (LMCI). Most patients are classified as cognitively normal, with far fewer patients classified as late MCI.

As shown in Figure 1, there is a significant class imbalance in the number of patients in each category, and previous studies of machine learning models have shown that highly imbalanced datasets lead to very poor predictive performance. When one class significantly outweighs the others in terms of data samples, the model becomes biased toward the majority class leading to poor performance on the minority class. As a result, the model's ability to accurately predict the less-represented class diminishes, affecting its overall generalization and practical utility. In this study, we test approaches that address class imbalance and improve model performance.

We experiment with the batch sizes and learning rates of each model in order for optimal performance, furthermore, we apply weighting techniques to support predictive performance on the minority class by increasing the indicence of the class in the batches. The concept behind batch size is that, instead of the model processing all the MRI imaging data simultaneously, it groups the data into smaller batches and learns from them incrementally. On the other hand, the learning rate serves as a crucial guide for determining how quickly the model adapts its understanding based on the incoming MRI imaging data. If the learning rate is too high, the model might not learn properly; conversely, if it's too low, the learning process may become excessively slow, potentially increasing screening costs.

Addressing class imbalance by weighting training batches increases model performance

The class weighting technique has been applied in various other domains such as fraud and anomaly detection, where fraudulent transactions are a minority compared to legitimate ones, and in natural language processing tasks like sentiment analysis, where extremely positive or negative sentiment classes might be imbalanced (Byrd and Lipton 2019). In the context of Alzheimer's disease and the urgent need for early-stage detection, addressing class imbalance in machine learning plays a critical role in ensuring fair representation of different patient groups within our dataset. This becomes particularly important when dealing with imbalanced data, where one category, such as patients with late-mild Alzheimer's, is underrepresented compared to others. We adopted a class weighting approach to rectify this imbalance and improve the performance of our machine learning models. Weighting in machine learning strives to achieve an equitable distribution of images or samples across each category or class within a classification. This approach is pivotal for correcting datasets when one class is overrepresented compared to others. By using class weighting, machine learning algorithms can better capture the nuances of the underrepresented class, leading to more balanced predictions and enhanced overall performance, especially in scenarios where accurate forecasts for the minority class are crucial.

The evaluation of the ResNet model aimed at assessing the impact of class imbalance on its performance. The investigation unveiled that class imbalance does indeed pose a challenge. The F-scores significantly improved when class weighting was applied compared to the models without weighting (Fig 2). Notably, the improvement was most pronounced when comparing the best models within each category.

F-scores are a measure of a machine learning model's accuracy that combines precision and recall into a single metric. Precision measures how many of the predicted positive instances were correct, while recall measures how many of the actual positive instances were predicted correctly. The F-score provides a way to balance these two metrics by taking their harmonic mean. Higher F-scores indicate that the model has both high precision in avoiding false positives and high recall in detecting true positives.

The following key observations were made regarding the impact of class imbalance and the effectiveness of class weighting in the ResNet model: The ResNet models with class weighting turned on consistently exhibited substantially improved F-scores across various categories in comparison to models with weighting turned off. In the context of the best-performing models for each category, the effect of class weighting on the LMCI category was particularly striking. The F-score for LMCI in the training set increased significantly from an almost negligible value of 3.50E-47 to a remarkable 0.999. Similarly, the LMCI F-score on the testing set also showed a substantial rise from 1.78E-11 to 0.995 (Fig 2 and Fig 3).

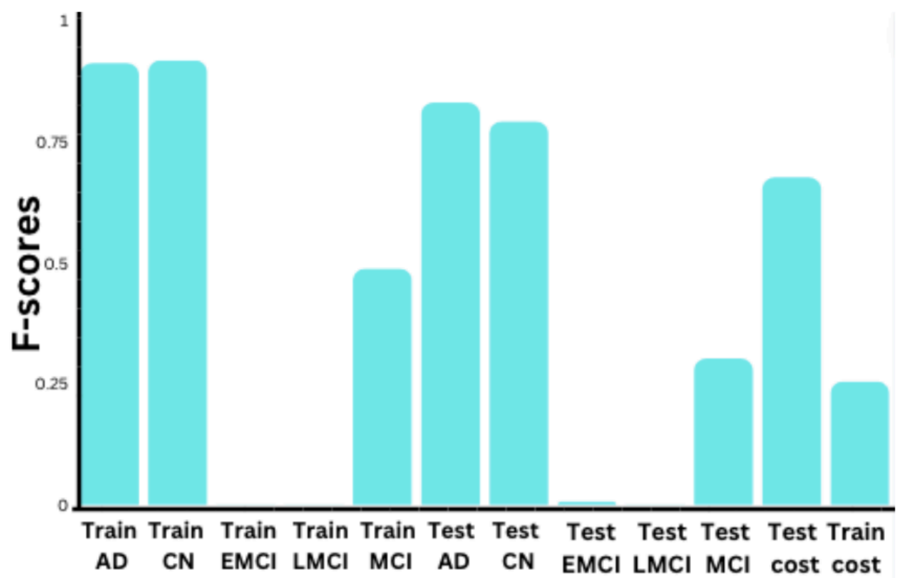


FIGURE 2. ResNet-18 without class weighting, batch size 32, learning rate 1E-06.

It can be seen that the F-scores of train AD and train CN are high (0.913 and 0.918 respectively). The model has low predictive performance for cognitive disease categories such as EMCI and MCI.

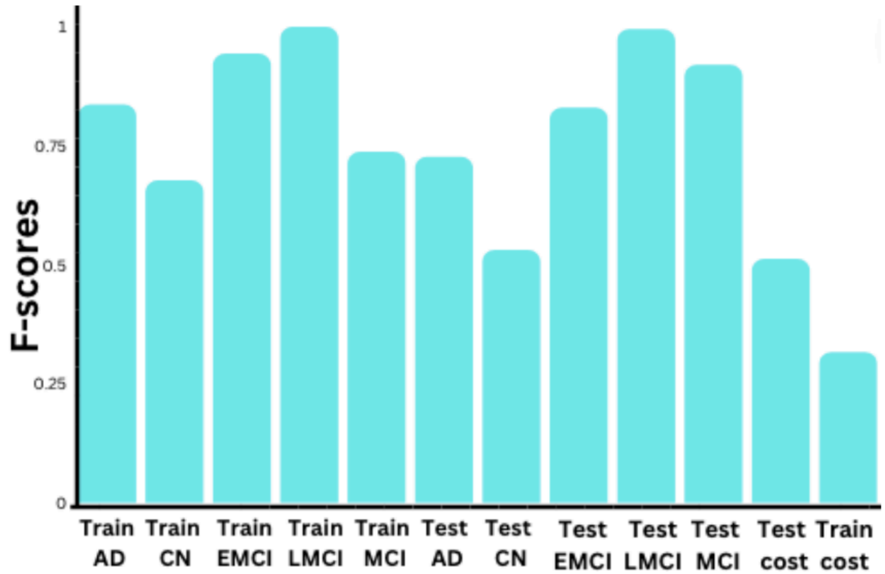


FIGURE 3. ResNet-18 with class weighting, batch size 32, learning rate 1E-06.

When the use of weighting is employed, the F-scores of early cognitive disease notably increase from $1x10^{-4}$ to 0.943 for EMCI and from $3.50x10^{-47}$ to 0.999 for LMCI. This corresponds with our main goal of making early detection more reliable.

These outcomes signify that class imbalance has a detrimental effect on the model's performance, and applying class weighting serves as a remedy to enhance its predictive capabilities. The data-driven insights derived from these experiments emphasize the critical importance of addressing class imbalance when training machine learning models for early Alzheimer's disease detection. When all experiments were conducted with class weighting, the dot plot reinforces the clear advantage gained by the ResNet model when tackling class imbalance challenges.

Merging of MCI and EMCI to enhance performance

In our pursuit of achieving excellence in early Alzheimer's detection while optimizing screening costs, a strategic decision was made to merge the categories of Mild Cognitive Impairment (MCI) and Early Mild Cognitive Impairment (EMCI) into a single group (MCI + EMCI), effectively eliminating the late Mild Cognitive Impairment (LMCI) stage from our classification. This decision was rooted in a thorough analysis of our dataset composition, the critical importance of early detection, and the need for resource-efficient screening.

Our dataset exhibited a significant class imbalance, with the majority of patients classified as cognitively normal, while the numbers of patients with EMCI, MCI, and LMCI were considerably lower. This class imbalance posed challenges in training machine learning models effectively, especially for the underrepresented EMCI and MCI classes. Given the urgency of early Alzheimer's detection and the introduction of preventative treatments, we aimed to optimize our models for identifying patients in the early stages of cognitive decline. Focusing on late-stage LMCI, which is closer to severe cognitive decline, could divert resources away from the critical early detection task. Additionally, the cost-effectiveness of screening plays a pivotal role in making early detection accessible to a larger population. By merging MCI and EMCI, we reduced the complexity of our classification task, which could lead to more efficient and cost-effective screening processes.

This decision to merge MCI and EMCI categories yielded notable improvements in our machine learning models' performance, particularly for early Alzheimer's detection. This consolidation allowed the models to allocate resources more effectively, resulting in better predictive capabilities for early cognitive impairment. The F-scores for EMCI and MCI, now MCI + EMCI, of both the training and testing sets, showed significant increases, indicating the enhanced ability to identify patients in the early stages of Alzheimer's disease. Moreover, the models demonstrated improved resource utilization, as reflected in lower screening costs, making early detection more reliable and accessible.

ResNet-18 model optimal for identifying MCI and EMCI (Early cognitive disease) compared to vision transformer models

While the ResNet-18 model demonstrated excellence in early detection, we sought to investigate the potential advantages of using models like DeIT (Data-efficient Image Transformer) and SimpleViT in comparison. Our decision to include DeIT and SimpleViT in the analysis was based on several key factors and logical considerations. Firstly, DeIT and SimpleViT are both transformer-based models, that leverage attention mechanisms. Transformers have shown remarkable success in various machine learning tasks, including natural language processing (e.g., GPT-3) and computer vision. These models are known for their ability to capture long-range dependencies and patterns in data, making them promising candidates for complex image analysis tasks. Beyond performance, we also considered the resource efficiency of these models. Given the importance of cost-effective early detection, we wanted to assess whether DeIT could offer a balance between performance and computational resources, making it a suitable choice for practical application

In order to achieve reliable and early detection of Mild Cognitive Impairment and Early Mild Cognitive Impairment (MCI + EMCI), the assessment of model performance plays a pivotal role. We recognize that achieving excellence in early detection must harmonize with the imperative of minimizing costs. Therefore, the ResNet-18 model, with its optimal configuration, not only excels in early detection but also aligns seamlessly with our intention to lower screening costs effectively. Based on the evaluation of the provided models, the ResNet-18 model emerges as the most reliable candidate for early detection (Fig 5).

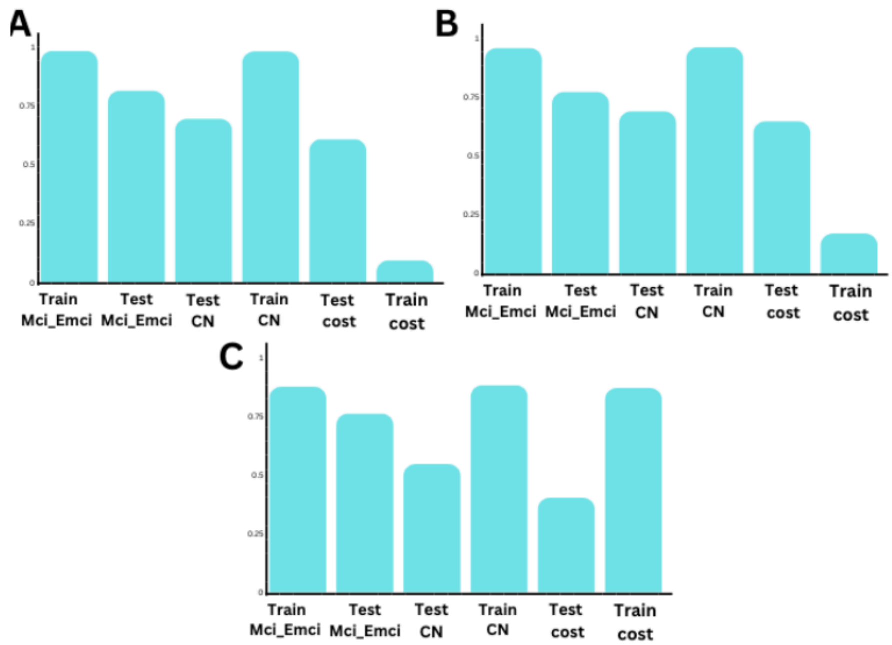


FIGURE 4. A) ResNet-18 with class weighting, batch size 32, learning rate 1E-06 B) DeiT with class weighting, batch size 8, learning rate 1E-06 C) SimpleVit with class weighting, batch size 32, learning rate 1E-06.

Resnet18 32 1E-06 achieves a training MCI + EMCI F-score of 0.9843, the highest of all three models. It also achieves the best F-scores in the other 5 categories, including the lowest test and train costs.

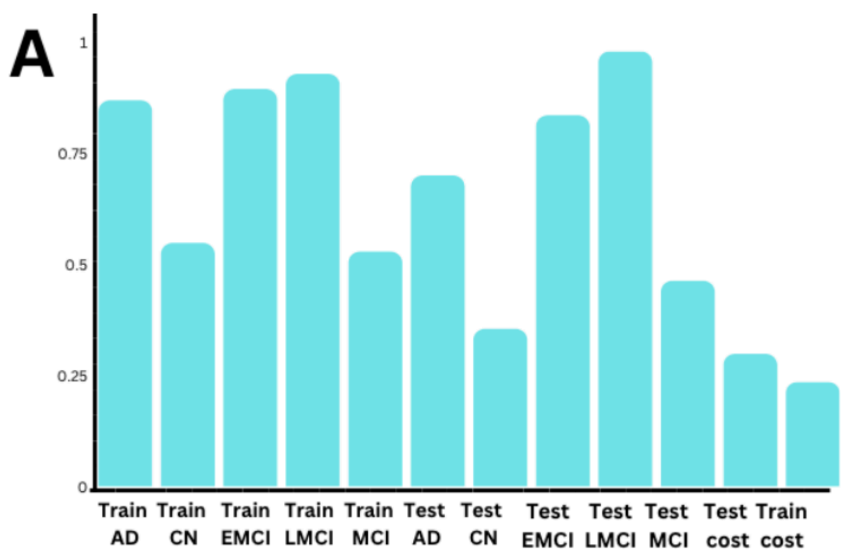
The F-scores are important measures of how well the model is at identifying MCI and EMCI, which are stages linked to cognitive diseases. The ResNet-18 model consistently achieves higher F-scores during both the training and testing phases.

Notably, the ResNet-18 model demonstrates exceptionally high F-scores for Train MCI + EMCI F-score and Train CN F-score categories, indicating its capacity to accurately detect these conditions. The model also manages to use resources efficiently. This is reflected in the lower cost values during both training and testing. In simpler terms, the model does a great job using its computing power wisely, making it practical for real-world use.

Given these insights, the ResNet-18 model emerges as the most reliable candidate for early detection of MCI and EMCI due to its consistently high F-scores across training and testing phases, coupled with its resource-efficient performance.

DeiT model performance reaches 92% F1 score with hyperparameter tuning

Optimizing the DeiT model's performance for the critical task of detecting patients in the early stages of Alzheimer's from MRI imaging data while also considering the cost of screening each patient, makes the selection of optimal hyperparameters, such as batch size and learning rate, even more crucial. As the need for earlier detection increases, a significant increase in the expected costs of screening patients can occur, as more individuals require assessments to detect these initial stages accurately. These hyperparameters influence how effectively the machine learning model learns from the data for early Alzheimer's detection, and they can impact the overall cost of the screening process. Our aim with the DeiT model was not only to achieve accurate early Alzheimer's detection but also to do so in a cost-effective manner. After extensive experimentation with various combinations, we arrived at a particularly successful configuration: a batch size of 8 and a learning rate of 1E-06, which yielded highly favorable outcomes (Fig 5) despite being inferior to ResNet models in a head-to-head comparison (Fig 4).



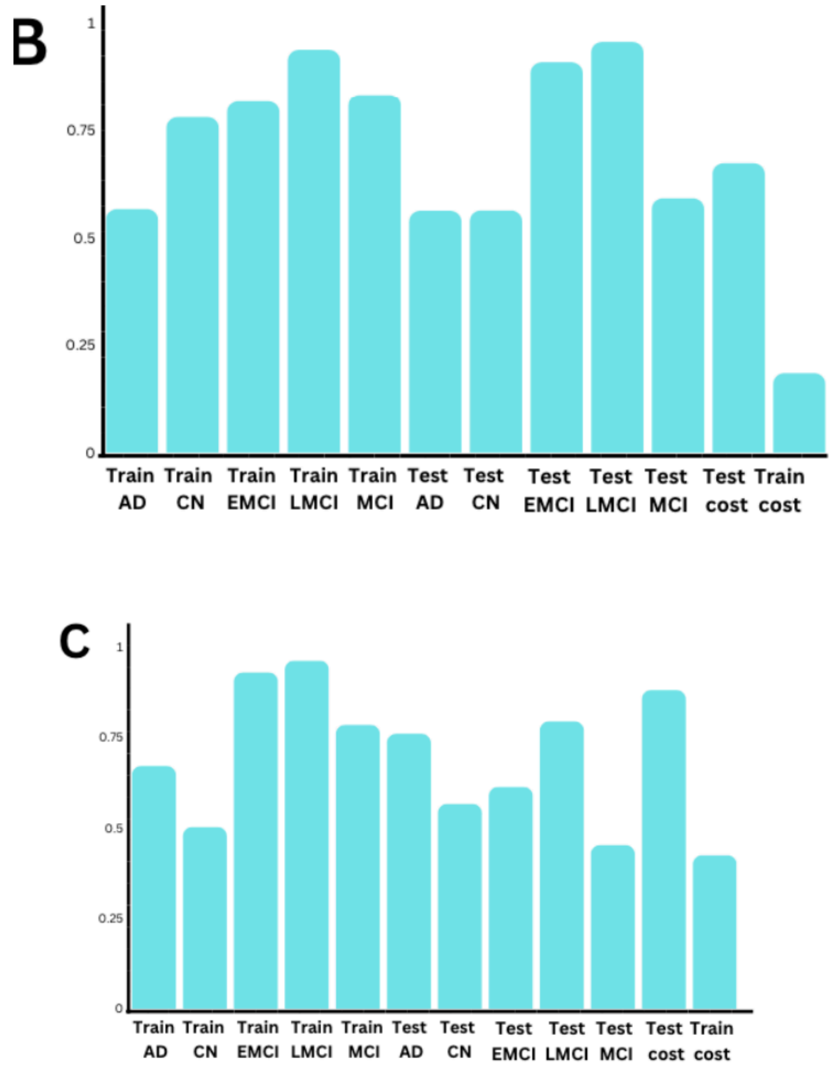


FIGURE 5. DeiT with class weighting, learning rate 1E-06, and batch sizes A) 8 B) 16 C) 32.

DeiT with a batch size of 8 and a learning rate of 1E-06 achieved the highest F-scores in the Early Mild Cognitive Impairment (EMCI) and Mild Cognitive Impairment (MCI) categories among the three configurations. It scored an F-score of 0.895 for EMCI and 0.9795 for MCI on the testing set, indicating its strong ability to identify early cognitive impairment.

In comparison to the other configurations, DeiT with a batch size of 8 and a learning rate of 1E-06 Weighted, consistently demonstrates superior performance, particularly in the critical task of identifying early cognitive impairment. Its balanced performance, efficient training, and strong ability to detect early cognitive decline make it a compelling choice for Alzheimer's disease detection, especially when early detection is of utmost importance.

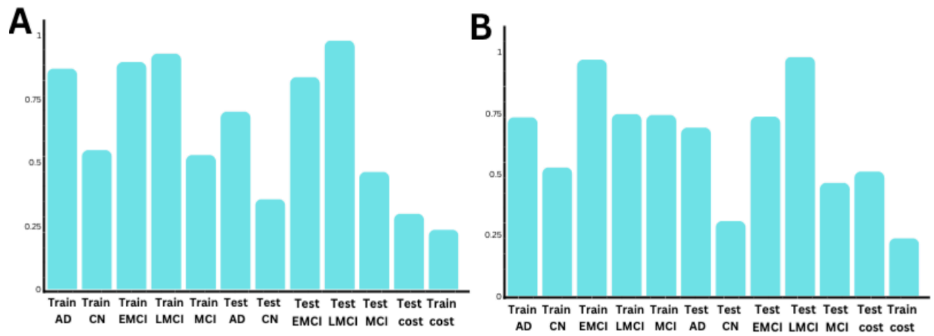


FIGURE 6. DeiT with class weighting, batch size 8, learning rates A) 1E-06 B) 1E-07

While the model with 1E-07 as the learning rate does have a better F-score in the training set of EMCI (0.968), the 1E-06 model has higher F-scores in both testing sets of EMCI and LMCI (0.835 vs. 0.734 and 0.980 vs. 0.975).

Our experimentation revealed that using a batch size of 8 for DeIT proved to be the most effective approach for early Alzheimer's detection while keeping computational costs manageable. This means that the model processes 8 portions of MRI imaging data at a time, incrementally improving its understanding of the early stages of Alzheimer's before moving on to the next batch, thus controlling computational expenses. We also determined that a learning rate of 1E-06, which represents a relatively small value, facilitated effective learning for the DeiT model while maintaining computational efficiency, which translates to cost-effectiveness.

The combination of a batch size of 8 and a learning rate of 1E-06 yielded highly promising results for both early Alzheimer's detection and cost control in the screening process. This configuration allowed the DeiT model to learn from the MRI data in an intelligent and efficient manner, instilling greater confidence in the reliability of its predictions for early Alzheimer's detection while keeping the cost of screening each patient at a manageable level.

SimpleVIT performance does not surpass the DeiT model

SimpleVIT is a type of machine learning model used for various computer vision tasks, involving leveraging self-attention to discern objects, patterns, and intricate details within images. This mechanism contributes significantly to its capacity for accurate image classification and object detection. It enables the model to focus on different parts of the input data, giving it the ability to understand contextual relationships. The key to SimpleVIT's efficiency is its simplicity. It reduces the redundancy found in other Vision Transformers through a simplified tokenization process and the removal of complex components like distillation tokens. At its core, SimpleViT still utilizes the power of self-attention to understand relationships between image components (Xu et al. 2022).

In order to achieve better results in the early detection of Alzheimer's disease, it was important to determine the most effective configuration for achieving high performance in various metrics of the SimpleVit model. Based on the evaluation of our experiments, the SimpleVit model with a batch size of 32 and a learning rate of 1E-06 is identified as the optimal configuration for achieving favorable results. The following experimental results showcase the performance of the SimpleVit model under different batch sizes and learning rate combinations (Fig 4).

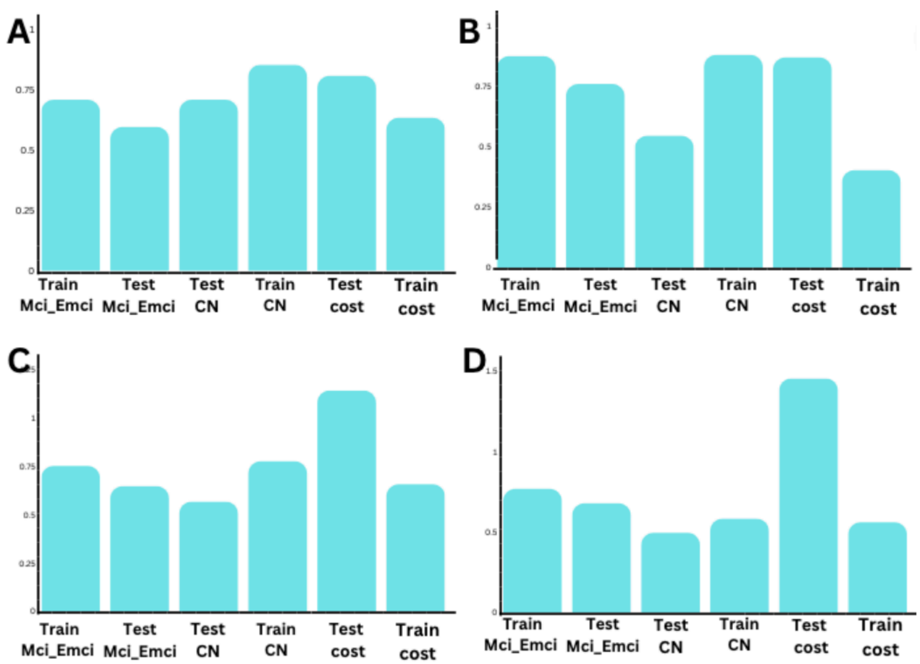


FIGURE 7. SimpleViT with class weighting, learning rate 1e-6, and batch size A) 16, B) 32, C) 64, and D) 128.

"SimpleVit 32 (batch size) 1E-06 (learning rate) with class weighting " achieves the highest training MCI + EMCI F-score of 0.875, indicating high accuracy in classifying cases related to MCI and EMCI during training. It also performs very well on the test data with a test MCI + EMCI F-score of 0.760. Even though its test and train costs (0.870 and 0.403) are relatively high, they are the lowest of the 4 models.

Based on the trends observed in the provided experiments, the SimpleVit model consistently performs well with a batch size of 32 and a learning rate of 1E-06. This configuration yields higher F-scores across various categories while maintaining relatively lower costs, both in the training and testing phases. Therefore, the SimpleVit model with a batch size of 32 and a learning rate of 1E-06 can be considered the optimal configuration for achieving reliable and effective performance in detecting

Mild Cognitive Impairment (MCI) and Early Mild Cognitive Impairment (EMCI).

Discussion/Conclusion

This study delved into the critical area of early detection of Alzheimer's disease, addressing the challenges posed by class imbalance and computational resource constraints. Our research unveiled several key findings, including class imbalance, effective configurations of learning models, and ResNet-18's superior predictive performance. The ResNet-18 model produced F-scores as high as 0.9843 for the MCI and EMCI stages, and 0.9826 for the cognitive normal stage.

We demonstrated that addressing class imbalance through class weighting significantly enhances the performance of machine learning models, particularly the ResNet-18 model, in detecting early cognitive decline. The F-scores improved notably across various categories when class weighting was applied, underscoring the importance of this technique in improving predictive accuracy. We also identified optimal hyperparameters for the DeiT and SimpleVit models. A batch size of 8 and a learning rate of 1E-06 were found to be highly effective for the DeiT model, resulting in accurate early Alzheimer's detection while controlling screening costs. Similarly, the SimpleVit model performed optimally with a batch size of 32 and a learning rate of 1E-06, achieving favorable F-scores and cost-effectiveness. However, In the pursuit of early detection of Mild Cognitive Impairment (MCI) and Early Mild Cognitive Impairment (EMCI), the ResNet-18 model emerged as the most reliable candidate. It consistently achieved higher F-scores for MCI and EMCI detection while efficiently utilizing computational resources, making it a practical choice for real-world applications.

The original challenge posed by Alzheimer's disease was the accurate identification of patients in the early stages of cognitive decline, particularly with the introduction of novel therapies that are effective only at this stage. Our study directly addresses this challenge by providing methodologies to improve the performance of machine learning models in early Alzheimer's detection. By addressing class imbalance and optimizing hyperparameters, we demonstrate a reliable and automated method that is cost-effective for screening Alzheimer-related early cognitive decline.

Our study aligns with and builds upon existing works that have aimed to address similar challenges. Several notable research studies have contributed to the field of Alzheimer's disease detection and early cognitive decline prediction. Notably, studies have addressed class imbalance, model configurations, and the choice of neural architectures.

Numerous studies have focused on using machine learning models, including convolutional neural networks (CNNs) and more recently, vision transformers (VITs), for the early detection of Alzheimer's disease. These works share our overarching goal of improving diagnostic accuracy. While many earlier studies concentrated on late cognitive decline (TanveerM et al. 2020), our research distinguishes itself by targeting early-stage cognitive impairment, aligning with the emerging importance of early intervention and the introduction of novel therapies.

Additionally, class imbalance is a pervasive issue in medical diagnostics and has been extensively studied. Our approach to addressing class imbalance through class weighting draws from this body of work (Hu et al. n.d.). Class imbalance is a well-known challenge in machine learning, impacting the model's ability to accurately predict minority classes. Our research reinforces the importance of class weighting, particularly in the context of VITs, and demonstrates its significant impact on model performance.

The scope of this research paper encompasses an in-depth investigation to determine the best model for the early detection of Alzheimer's and addressing class imbalance in Vision Transformers (VITs). The study leverages the publicly available ADNI dataset, which provides a diverse set of neuroimaging and clinical data for training and evaluation.

However, there are certain limitations to this study. Firstly, while the ADNI dataset is a widely used and well-established resource for Alzheimer's disease research, it may still suffer from inherent biases and limitations due to factors such as data collection protocols and participant demographics. Mitigating these biases is a complex challenge that may not be fully addressed by the proposed techniques. To further strengthen our findings, we would replicate our analyses in an independent dataset collected from a distinct patient population.

Additionally, the generalizability of the findings may be limited to the specific context of Alzheimer's disease prediction, and the performance improvements observed in this study may not translate directly to other computer vision tasks or datasets with different characteristics.

Despite these limitations, our study shows valuable insights and methodologies to tackle the class imbalance problem in VITs. The findings are expected to guide researchers and practitioners in developing more efficient and effective models for addressing complex medical diagnostic tasks like Alzheimer's disease early-stage prediction.

Our research contributes to the goal of improving the lives of individuals at risk of Alzheimer's disease by facilitating early detection and intervention. By making the screening process more reliable and cost-effective, we enable healthcare providers to identify patients in the early stages of cognitive decline, when interventions like novel drug therapies are most effective. This has the potential to significantly enhance the quality of life for those affected by Alzheimer's disease and reduce the economic burden on healthcare systems.

Going forward, it would be valuable to validate these findings on additional independent datasets from diverse patient populations to further assess the generalizability of the ResNet-18 model's performance. Another important next step is to deploy and prospectively evaluate the model in clinical settings to understand its real-world impact on early Alzheimer's screening and patient outcomes. Continued research into innovative deep-learning architectures and techniques for handling class imbalance could also yield even higher-performing models for this critical application. Finally, integrating these AI-based screening tools into clinical decision support systems and exploring ways to explain the model predictions could increase clinician trust and adoption for facilitating early intervention.

From a healthcare system perspective, substantial cost savings are possible by reducing the need for more expensive interventions later in the disease progression. Moreover, the improved accuracy of early diagnosis allows patients to plan their next steps and access support services earlier. This provides patients and families dealing with a devastating diagnosis more time to adjust lifestyle factors, enroll in clinical trials, and proactively manage healthcare decisions, which can enhance quality of life and reduce uncertainties. In summary, this work could help healthcare institutions deliver better patient-centered care while optimizing resource utilization for managing Alzheimer's disease. The positive impact on both healthcare infrastructure and human lives makes this a valuable application of AI in healthcare.

Methods

For this research, we utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The ADNI dataset is a well-established and widely used resource in the field of Alzheimer's disease research. It comprises a comprehensive collection of neuroimaging and clinical data from individuals with different stages of Alzheimer's disease, as well as from cognitively normal controls. The dataset provides a valuable resource for training and evaluating machine learning models for Alzheimer's disease prediction, allowing us to assess the performance of VITs with class imbalance in mind.

The ADNI dataset includes a diverse range of imaging modalities, such as structural MRI scans, along with corresponding clinical information, such as age, gender, and visit information. Each sample in the dataset is labeled with its respective stage of Alzheimer's disease, including Cognitive Normal (CN), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), Early Mild Cognitive Impairment (EMCI), and Alzheimer's Disease (AD) stages. The discrimination between different stages of AD are considerably important issue for future pre-dementia treatment. The categorical class labels (CN, MCI, LMCI, EMCI, and AD) were encoded into numerical values to facilitate model training. Additionally, the dataset contains information on cognitively normal individuals who serve as controls. It is important to highlight that the dataset is prone to class imbalance, as the number of samples in different Alzheimer's disease stages and the control group are not evenly distributed. This class imbalance can potentially impact the performance of machine learning models, especially VITs, which was addressed in this research.

The dataset was organized into separate training and test folders containing the MRI scans. These were loaded using PyTorch ImageFolder datasets, which automatically assigns class labels based on the folder structure. The MRI images were resized to 224x224 pixels and normalized to the tensor data type using torchvision transforms. Different versions of the dataset were used by passing root file paths, including removing the LMCI class or restricting it to only EMCI and MCI.

Models like ResNet-18, DeiT, and SimpleVit were implemented in PyTorch and initialized with pre-trained weights from ImageNet. The models were trained for 3 epochs using the Adam optimizer, which adopts adaptive learning rates based on parameter-specific momentum. A learning rate of 1e-6 was selected through hyperparameter tuning and supplied to the Adam optimizer. The objective function for optimization was cross-entropy loss, commonly used for multi-class classification, weighted to account for class imbalance.

Model performance was evaluated every 4 iterations on a held-out test set. Metrics like cost, precision, recall, and F1-score were computed for each class. The sklearn functions were used to calculate precision-recall metrics. Weights & Biases were used to track experiments, log metrics, and visualize performance. The models were compared to identify optimal configurations based on the highest test F1-scores for early cognitive decline classes like EMCI and MCI.

References

Abubakar, Murtala Bello, Kamaldeen Olalekan Sanusi, Azizah Ugusman, Wael Mohamed, Haziq Kamal, Nurul Husna Ibrahim, Ching Soong Khoo, and Jaya Kumar. 2022. "Alzheimer's Disease: An Update and Insights Into Pathophysiology." Frontiers in Aging Neuroscience 14 (March): 742408.

Byrd, Jonathon, and Zachary Lipton. 2019. "What Is the Effect of Importance Weighting in Deep Learning?" In International Conference on Machine Learning, 872–81. PMLR.

Han, Kai, Yunhe Wang, Hanting Chen, Xinghao Chen, Jianyuan Guo, Zhenhua Liu, Yehui Tang, et al. n.d. "A Survey on Vision Transformer." Accessed September 26, 2023. https://ieeexplore.ieee.org/abstract/document/9716741.

"High-Clearance Anti-Amyloid Immunotherapies in Alzheimer's Disease. Part 1: Meta-Analysis and Review of Efficacy and Safety Data, and Medico-Economical Aspects." 2022. Revue Neurologique 178 (10): 1011–30.

Hu, Shengye, Wen Yu, Zhuo Chen, and Shuqiang Wang. n.d. "Medical Image Reconstruction Using Generative Adversarial Network for Alzheimer Disease Assessment with Class-Imbalance Problem." Accessed September 26, 2023. https://ieeexplore.ieee.org/abstract/document/9344912.

Li, Yawei, Kai Zhang, Jiezhang Cao, Radu Timofte, and Luc Van Gool. 2021. "LocalViT: Bringing Locality to Vision Transformers." http://arxiv.org/abs/2104.05707.

Mayeux, Richard. 2010. "Early Alzheimer's Disease," June. https://doi.org/10.1056/NEJMcp0910236.

Park, Namuk, and Songkuk Kim. 2022. "How Do Vision Transformers Work?" http://arxiv.org/abs/2202.06709.

TanveerM, RichhariyaB, U. KhanR, H. RashidA, KhannaP, PrasadM, and T. LinC. 2020. "Machine Learning Techniques for the Diagnosis of Alzheimer's Disease." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), April. https://doi.org/10.1145/3344998.

Ullah, Asad, Hassan Elahi, Zhaoyun Sun, Amna Khatoon, and Ishfaq Ahmad. 2021. "Comparative Analysis of AlexNet, ResNet18 and SqueezeNet with Diverse Modification and Arduous Implementation." Arabian Journal for Science and Engineering 47 (2): 2397–2417.

Xu, Yufei, Jing Zhang, Qiming Zhang, and Dacheng Tao. 2022. "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation." Advances in Neural Information Processing Systems 35 (December): 38571–84.