TOWARDS RELIABLE ALZHEIMER'S DIAGNOSIS FROM 3D MRI SCANS: A GENERALIZED APPROACH

by
ZOBIA BATOOL

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of Master of Science

Sabancı University
June 2025

TOWARDS RELIABLE ALZHEIMER'S DIAGNOSIS FROM 3D MRI SCANS: A GENERALIZED APPROACH

Approved by:

Prof. Erchan Aptoula ....
(Thesis Supervisor)

Assoc. Prof. Faik Boray Tek ....

Assoc. Prof. Öznur Taştan ....

Date of Approval: June 23, 2025

ABSTRACT

TOWARDS RELIABLE ALZHEIMER'S DIAGNOSIS FROM 3D MRI SCANS: A GENERALIZED APPROACH

ZOBIA BATOOL

COMPUTER SCIENCE AND ENGINEERING M.Sc. THESIS, June 2025

Thesis Supervisor: Prof. Erchan Aptoula
Thesis Co-Supervisor: Assoc. Prof. Hüseyin Özkan

Keywords: Domain Generalization, Alzheimer's Disease, Contrastive Learning, Morphological Networks, Deep Learning

This thesis aims to address Alzheimer's disease detection from 3D MRI scans under a single-domain generalization setting, where a model is expected to generalize to unseen domains with potentially diverse imaging protocols, patient demographics, and class imbalance levels. Three distinct approaches are investigated. First, a pseudo-morphological augmentation strategy uses learnable modules to produce anatomically coherent, class-specific augmentations, integrated with supervised contrastive learning to extract robust and discriminative features. Second, the MixStyle framework is extended to incorporate higher-order statistical moments including skewness and kurtosis alongside traditional mean and variance, enabling enhanced feature perturbation and focus on disease-specific artifacts. Third, a Mixup-based augmentation method leverages distance transforms to spatially decompose MRI scans into layered components and recompose them from multiple samples, preserving structural integrity while promoting diversity. Extensive experiments across three benchmark datasets, namely NACC, ADNI and AIBL demonstrate that the proposed techniques substantially enhance the generalization capabilities of underlying models, thus providing a strong basis for creating reliable, domain-agnostic tools for early Alzheimer's disease diagnosis.

ÖZET

3B MRG TARAMALARINDAN GÜVENİLİR ALZHEİMER TEŞHİSİNE
DOĞRU: GENELLEŞTİRİLMİŞ BİR YAKLAŞIM

ZOBIA BATOOL

BİLGİSAYAR BİLİMİ VE MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, MAYIS
2025

Tez Danışmanı: Prof. Dr. Erchan Aptoula
Tez Eş-Danışmanı: Doç. Dr. Hüseyin Özkan

Anahtar Kelimeler: Alan Genellemesi, Alzheimer Hastalığı, Karşıtsal Öğrenme, Biçimbilimsel Ağlar, Derin Öğrenme

Bu tez, 3B MRG görüntülerinden Alzheimer hastalığının tespiti üzerine odaklanmakta ve modelin, görüntüleme protokolleri, hasta demografileri ve sınıf dengesizlikleri açısından farklılık gösteren, daha önce görülmemiş alanlara genelleme yapmasının beklendiği tek alanlı genelleme (single-domain generalization) ortamında çalışmaktadır. Üç farklı yaklaşım incelenmiştir. İlk olarak, anatomik olarak tutarlı ve sınıfa özgü veriler üreten, öğrenilebilir modüller kullanan sahte-biçimbilimsel bir veri artırma stratejisi önerilmiştir. Bu strateji, gürbüz ve ayırt edici öznitelikler çıkarmak amacıyla gözetimli karşıtsal öğrenme ile tümleştirilmiştir. İkinci olarak, MixStyle çerçevesi, geleneksel ortalama ve değişintinin yanı sıra çarpıklık ve basıklık gibi daha yüksek dereceli istatistiksel momentleri de içerecek şekilde genişletilmiştir. Bu sayede özellik pertürbasyonu geliştirilmiş ve hastalığa özgü kalıntılara odaklanılmıştır. Üçüncü olarak, Mixup tabanlı bir veri artırma yöntemi, MRG görüntülerini mesafe dönüşümleri ile katmanlara ayırıp, birden fazla örnekten yeniden birleştirerek yapısal bütünlüğü korurken çeşitliliği artırmayı hedeflemiştir. NACC, ADNI ve AIBL olmak üzere üç farklı veri kümesinde yapılan kapsamlı deneyler, önerilen tekniklerin temel modellerin genelleme yeteneğini önemli ölçüde artırdığını ve böylece erken aşamada Alzheimer hastalığı tanısı için güvenilir ve alan bağımsız araçlar geliştirilmesine güçlü bir zemin sağladığını göstermiştir.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to everyone who has supported me throughout this journey.

First and foremost, I want to thank my advisor, Professor Erchan Aptoula. He was always available when I needed help, responded quickly, and gave the kind of guidance that truly made a difference. His support made everything in this thesis and the upcoming publications possible. I have learned so much from him, and will always be grateful for his advice. I am also grateful to Assoc. Prof. Huseyin Ozkan for believing in me and for supporting my work through funding.

I also want to thank the members of my thesis committee for agreeing to be part of the evaluation process and for their time and consideration. Thanks to everyone in the VPA Lab, both those who are still here and those who graduated before me, for making the lab feel like a second home.

I am thankful to Mama and Baba, for whom I have the utmost love and respect. To my sister Maria, who has been like a second mother to me and always offered her guidance. To my brother, for always supporting me and giving thoughtful advice. And to my other sister, Sobia, for always being there for me. A big thank you to my close friends Rabia, Gulsen, Seba, Gokmen, and Ayça, who made life at Sabancı University so much easier and more enjoyable.

Finally, I would like to acknowledge the financial support from The Scientific and Technological Research Council of Türkiye (TÜBİTAK), under grant number 121E452. I also thank Ms. Diala Lteif and Dr. Vijaya Kolachalama from Boston University, for allowing me to access and use their NACC dataset.

Thank you all for being part of this journey.

Dedication
To myself, for not giving up.
To Mama and Baba, for their endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

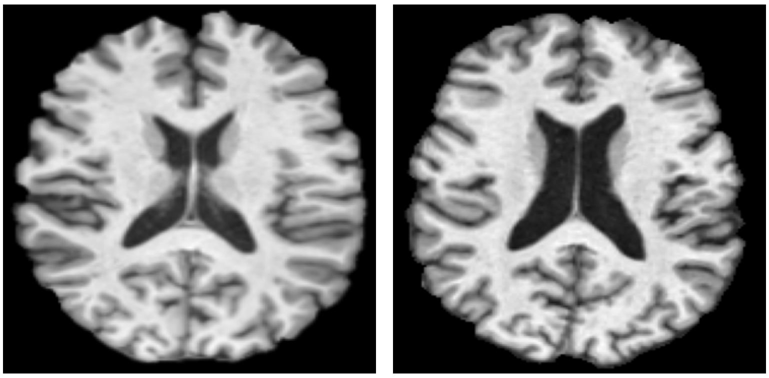
LIST OF FIGURES

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive and irreversible neurodegenerative disorder that primarily affects older adults. It is the most common cause of dementia, with an estimated 55 million people living with dementia globally in 2019. Among the dementia cases, AD accounts for 60-70% of all cases, a number projected to rise significantly due to increasing life expectancy (Selvester & others, 2022). Clinically, AD is characterized by memory loss, cognitive decline, disorientation, and behavioral disturbances that worsen over time. At the pathological level, Alzheimer's is marked by the buildup of harmful substances in the brain. One is called amyloid-β, which forms sticky clumps (plaques) between brain cells and blocks communication. Another is a protein called tau, which becomes abnormal and forms twisted strands (tangles) inside brain cells, stopping them from working properly. These changes cause brain cells to die and lead to shrinkage of the brain, especially in areas like the hippocampus which is important for memory and the outer brain regions involved in thinking and decision-making (Selkoe & Hardy, 2021).

For diagnosis, the clinical assessment of brain structure is commonly supported by neuroimaging techniques such as Magnetic Resonance Imaging (MRI), which provides high-resolution, non-invasive insight into brain morphology. MRI scans can reveal structural abnormalities, including cortical thinning, ventricular enlargement, and hippocampal atrophy, that correspond to disease progression (Teipel & others, 2011; van Oostveen & de Lange, 2024; Weiner & others, 2017). Visually, AD looks like a loss of brain volume and structure when compared to a healthy brain. (Fig. 1.1) illustrates the structural differences observable in MRI scans between a Normal Control (NC) subject and an AD case. The AD brain exhibits clear cortical atrophy, seen as widened grooves on the brain's surface and contracted or enlarged fluid-filled regions such as the ventricles. These changes reflect tissue loss and are especially noticeable around the hippocampus and outer cortical areas. The AD image also contains more dark regions, indicating areas where brain tissue has deteriorated or thinned, which corresponds to loss of neuronal density.

Based on the assessments, early detection of AD is crucial, as it enables timely inter-



(a) NC case (b) AD case

Figure 1.1 MRI samples: (a) NC subject with intact brain structure, and (b) AD case showing pronounced cortical atrophy.

vention, improved quality of life, and better care planning. However, diagnosing AD at an early stage remains challenging. Symptoms often overlap with normal aging or other neurodegenerative conditions, and structural brain changes may be subtle and variable across individuals (Jack, Knopman, Jagust, Shaw, Aisen, Weiner, Petersen & Trojanowski, 2010). In addition, access to advanced diagnostic tools such as PET imaging or cerebrospinal fluid analysis may be limited due to cost, invasiveness, or availability. Furthermore, variability in clinical expertise and subjective interpretation of neuropsychological assessments can lead to inconsistent or delayed diagnosis (Jack, Bennett, Blennow & others, 2018).

On the contrary, machine learning, and in particular deep learning, has become a powerful tool for automated AD diagnosis using MRI. Among various architectures, Convolutional Neural Network (CNN) based models have shown promise in capturing spatial features and structural abnormalities (El-Assy, Amer, Ibrahim & Mohamed, 2024; Folego, Weiler, Casseb, Pires & Rocha, 2020; Han, Li, Fang & Yang, 2023; Li, Zhang, Wu, Zhang, Han & Cui, 2024; Zhang, Chen, Ren, Yang, Yu, Zhang & Zhou, 2022). More recently, attention-based models such as Vision Transformers (ViTs) (Alp, Akan, Bhuiyan, Disbrow, Conrad, Vanchiere, Kevil & Bhuiyan, 2024; Duong, Tran & Gahm, 2025; Joy, Nasrin, Siddiqua & Farid, 2025; Lu, Zhang & Yao, 2025; Shaffi, Viswan & Mahmud, 2024) and Graph Neural Networks (GNNs) (Gamgam, Kabakcioglu, Yüksel Dal & Acar, 2024; Hu, Wang, Zhu, Li & Shi, 2024; Majee, Gupta, Raha & Das, 2024) have been leveraged to capture more complex spatial dependencies and inter-regional connectivity. Despite these advancements, existing approaches often assume that training and test data come from the same distribution. This assumption is rarely satisfied in practice due to domain shifts caused by differences in scanner type, acquisition protocols, preprocessing pipelines, and population demographics. Additionally, class imbalance, particularly the under-representation of AD cases, presents a serious challenge for model learning and generalization.

While many methods have focused on improving model accuracy or incorporating multimodal data, relatively few have considered disease-specific anatomical properties such as brain atrophy. As discussed in Chapter 2, atrophy is a central pathological feature of AD, with different degrees of severity observable across NC, Mild Cognitive Impairment (MCI), and AD classes. Yet, most domain generalization (DG) techniques (Cai, Zhang & Long, 2023; Fiasam, Rao, Sey, Aggrey, Kodjiku, Browne, Danso, Ukwuoma & Gyarteng, 2023; Lteif, Sreerama, Bargal, Plummer, Au & Kolachalama, 2024; Nguyen, Clément, Mansencal & Coupé, 2023; Wang, Chaudhari & Davatzikos, 2022; Zhou, Li, Zhou, Liu & Tu, 2023a) rely on handcrafted priors or statistical perturbations that do not explicitly account for these structural changes. Furthermore, the single-domain generalization (SDG) setting — where only one labeled dataset is available for training and no access to target domain data is assumed — remains relatively underexplored despite being more realistic for clinical deployment.

To address these limitations, this thesis proposes three novel SDG methods designed to improve AD classification under domain shift while taking into account the anatomical properties of neurodegeneration:

- **Learnable pseudo-morphological augmentations with supervised contrastive learning:** As discussed before, atrophy is the main biomarker for AD. Based on that, this method introduces class-specific 3D morphological augmentations using learnable erosion and dilation modules, designed to reflect the variable presence of atrophy across classes. These augmentations are coupled with supervised contrastive loss to enforce representation consistency across domains.
- **Distance transform guided MixUp:** A structure-aware data augmentation strategy is introduced that uses distance transforms to decompose 3D brain volumes into spatial layers. Patch mixing is performed in a way that preserves anatomical boundaries, mitigating the distortions introduced by vanilla MixUp.
- **Extended MixStyle with higher-order statistics:** A generalization of the MixStyle framework is proposed to include higher-order feature statistics (e.g., skewness and kurtosis), providing more expressive feature perturbations that better simulate domain shifts.

All models are trained on the NACC dataset and evaluated on two independent benchmarks, ADNI and AIBL. Extensive experiments conducted on all three afforementioned methods demonstrate improved performance under domain shift. These contributions aim to advance domain-generalizable AD diagnosis by incorporating disease-specific anatomical priors, addressing class imbalance, and improving the robustness of feature learning under SDG constraints.

2. RELATED WORK

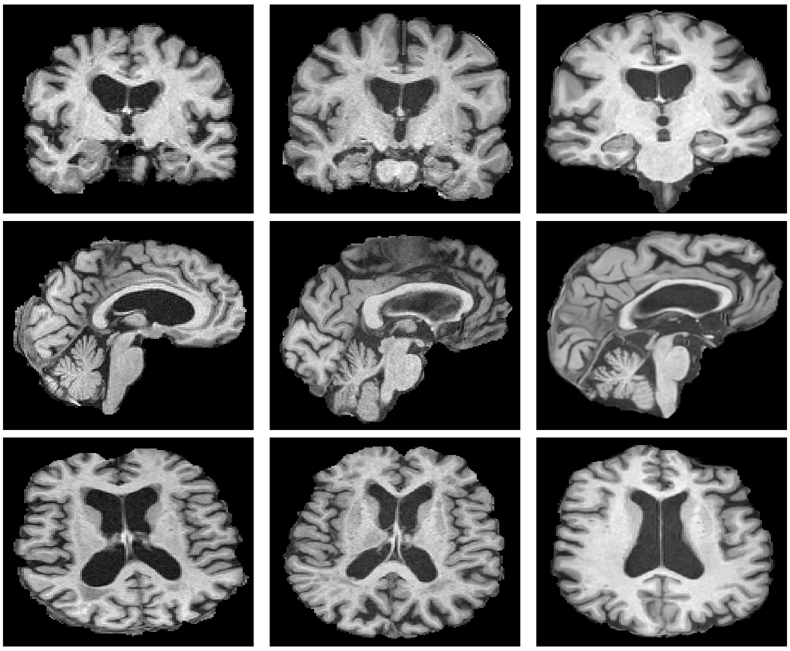
2.1 Background and Challenges in AD Detection

2.1.1 Overview of Alzheimer's Disease

Alzheimer's disease is a progressive neurodegenerative disorder that arises from a complex factors, including genetic susceptibility, environmental exposures, and lifestyle influences, leading to gradual cognitive decline, behavioral impairments, and irreversible structural brain damage (Chen, Pan, Xia & Yuan, 2023; Zhang, Zhang, Wang, Xia, Zhang & Chen, 2024). One of the most significant challenges in managing AD is the difficulty of early diagnosis. This is largely due to the slow and subtle nature of disease progression, which varies considerably across individuals in terms of symptoms, affected brain regions, and rate of cognitive decline. Identifying AD early, especially during prodromal i.e. MCI stages, is crucial for effective clinical intervention and long-term care planning.

Clinically, AD progression is often categorized into three stages: mild (early), moderate (middle), and severe (late). In the mild stage, individuals may function independently but show early signs such as word-finding difficulty, disorganization, or short-term memory loss (McKhann, Knopman, Chertkow, Hyman, Jack, Kawas, Klunk, Koroshetz, Manly, Mayeux & others, 2011). This stage frequently overlaps with MCI, a heterogeneous condition subtyped into early and late MCI depending on severity and likelihood of conversion to AD. As AD advances, memory and behavioral symptoms intensify, eventually requiring full-time care.

At the neuropathological level, AD is defined by the accumulation of two hallmark proteins: amyloid-beta plaques and tau tangles. These protein aggregates disrupt synaptic communication and trigger inflammatory and apoptotic processes, ultimately leading to widespread neuronal death. The hippocampus, critical for memory encoding and consolidation, is among the earliest and most severely affected regions. As the disease progresses, brain atrophy spreads to the temporal and parietal lobes, and cortical thinning becomes more generalized. These anatomical changes are mirrored by worsening memory loss, disorientation, language impairment, emotional dysregulation, and, in later stages, the complete loss of independence (Braak & Braak, 1991; Jack et al., 2018). (Fig. 2.1) shows brain scans from AD, MCI, and NC subjects in the ADNI dataset. The AD scan (a) displays significant cortical atrophy near the hippocampal and parietal regions, while the NC scan (b) shows preserved brain volume. The MCI scan (c) exhibits intermediate features, underscoring the need to detect these subtle structural differences.



(a) AD case (b) NC case (c) MCI case

Figure 2.1 Alzheimer's 2D MRI slices (a) AD case showing pronounced cortical atrophy, (b) NC subject with intact brain structure, and (c) MCI case exhibiting intermediate anatomical features between AD and NC.

2.1.2 Clinical Importance and Imaging-Based Diagnosis

There have been numerous efforts to enable early diagnosis of AD due to its clinical, social, and economic importance. Clinically, early detection allows for timely intervention to slow progression and manage symptoms. It also supports informed care planning. Socially, it reduces uncertainty and engages support networks earlier. Economically, it can lower costs by delaying institutional care and improving treatment efficiency.

Diagnosis typically involves a combination of clinical evaluation, cognitive testing, and biomarker analysis. Although a definitive diagnosis is usually confirmed after death (Jack et al., 2018) tools like brain imaging and spinal fluid tests are important for detecting changes while the person is still alive. Cognitive tests look for problems with memory, language, and thinking skills, while imaging shows structural changes in the brain that are consistent with disease progression.

Among various imaging techniques, magnetic resonance imaging (MRI) is widely used in the diagnosis of AD due to its accessibility, safety, and ability to detect structural brain changes. Unlike Positron Emission Tomography (PET), which requires radioactive tracers, MRI is non-invasive and does not involve radiation. It provides high-resolution images that reveal early signs of brain atrophy, particularly in the hippocampus and medial temporal lobe—areas affected early in the disease. These changes often appear before noticeable symptoms and help distinguish AD from other types of dementia. MRI also supports disease staging by monitoring cortical thinning over time, making it valuable for both diagnosis and ongoing assessment (Frisoni, Fox, Jack Jr, Scheltens & Thompson, 2010).

2.1.3 Limitations of Current Diagnostic Methods

Despite recent advances, current methods for detecting AD, both clinical and AI-driven, still face significant limitations. Clinically, early symptoms often overlap with those of normal aging or other neurodegenerative and psychiatric conditions, such as depression or vascular dementia. This overlap contributes to frequent misdiagnoses or delayed detection (Fotuhi, Hachinski & Whitehouse, 2009). Cognitive screening tools are commonly used but have limited sensitivity for early-stage or preclinical AD and can be influenced by factors such as education level, language, and cultural background (Manly, Bell-McGinty, Tang, Schupf, Stern & Mayeux, 2005). Biomarker-based methods, such as specialized brain imaging and fluid tests, can improve diagnostic accuracy but tend to be expensive, invasive, and not widely accessible in standard healthcare settings.

AI-based approaches offer promise in overcoming some of these clinical limitations, yet they introduce new challenges. Many models are trained on datasets that lack demographic diversity, reducing their effectiveness across broader populations. Imaging data often originate from specific research cohorts, which may not reflect the variability encountered in typical clinical environments. Furthermore, differences in MRI scanners, acquisition protocols, and data preprocessing can introduce inconsistencies that degrade model performance. These models also tend to rely on high-quality or multi-modal inputs that are not routinely available in many clinics. Another major concern is the limited interpretability of AI systems, which makes it difficult for clinicians to understand or trust the decisions being made. Additionally, technical barriers such as high computational costs and the absence of standardized evaluation practices hinder the deployment of these tools at scale (Lundervold & Lundervold, 2019).

2.1.4 Importance of Generalization in AD Detection

Generalization is a cornerstone of any reliable AI system intended for clinical deployment, particularly in the context of AD detection using 3D MRI data. In real-world settings, patients vary widely in terms of age, ethnicity, and cognitive baselines, and MRI data can differ significantly depending on scanner manufacturer, magnetic field strength, and acquisition parameters. A model that performs well on a single dataset or site may struggle when applied to new, unseen populations or imaging conditions. Such variability can lead to poor clinical performance and unreliable diagnostic outcomes, undermining the potential benefits of AI in early AD detection.

To address this, recent research has focused on methods that explicitly enhance generalization. These include multi-site training, domain adaptation, harmonization of imaging data, and validation across diverse cohorts. Federated learning is also gaining attention as a privacy-preserving strategy to train models on decentralized data from multiple sources (Li, Sahu, Talwalkar & Smith, 2020). Importantly, models must not only generalize across imaging sites but also maintain robustness across varying cognitive trajectories and disease presentations. Ensuring generalization is essential to move beyond controlled experimental conditions and toward building clinically trustworthy systems. As such, improving generalization is not merely a technical refinement but it is fundamental to safe, and effective application of AI in AD diagnostics.

2.2 Data and Evaluation Landscape

2.2.1 Public Datasets for AD Research

Public datasets have been fundamental to the advancement of AD research, particularly for training and evaluating machine learning models on neuroimaging and clinical data. Among the most widely used is the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen, Aisen, Beckett, Donohue, Gamst, Harvey, Jr, Jagust, Shaw, Toga & Trojanowski, 2010), which provides a rich repository of longitudinal MRI, PET, genetic, and cognitive assessments collected from multiple clinical sites. The National Alzheimer's Coordinating Center (NACC) (Ellis, Bush, Darby, Fazio, Foster, Hudson, Lautenschlager, Lenzo, Martins, Maruff, Masters & the AIBL Research Group, 2009) dataset complements this with standardized clinical data from a large, diverse patient population across Alzheimer's Disease Centers in the United States. Similarly, the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) (Ellis et al., 2009) offers multi-modal data with an emphasis on lifestyle and biomarker interactions, while the Framingham Heart Study (FHS) (Seshadri, Beiser, Kelly-Hayes, Kase, Au, Au, Hoffmann, Benjamin, Vasan & Wolf, 2006) provides valuable insights through its long-term, community-based cohort tracking both cardiovascular and cognitive health. Additionally, the Open Access Series of Imaging Studies (OASIS) (Marcus, Wang, Parker, Csernansky, Morris & Buckner, 2007) offers freely available cross-sectional and longitudinal MRI datasets, spanning cognitively normal individuals to those with early-stage dementia, making it a critical resource for studying brain aging and Alzheimer's progression.

These datasets enable robust model development and validation by offering varied demographic, clinical, and imaging characteristics. Their availability supports reproducibility and cross-study comparisons, which are essential for scientific rigor. Additionally, the diversity across datasets allows for the development of models that are more generalizable to real-world clinical settings. However, challenges such as label inconsistencies, missing modalities, and variations in imaging protocols remain under-investigated and need attention.

2.2.2 Evaluation Metrics and Validation Protocols in AD Research

Robust evaluation metrics and validation protocols are essential for assessing the reliability and clinical applicability of AD detection models. Commonly used metrics in classification tasks include accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) (Bravo-Ortiz, Holguin-Garcia, Quiñones-Arredondo, Mora-Rubio, Guevara-Navarro, Arteaga-Arteaga, Ruz & Tabares-Soto, 2024). These metrics help quantify a model's ability to distinguish between diagnostic categories such as cognitively normal, mild cognitive impairment, and AD. However, over-reliance on accuracy can be misleading, especially in datasets with class imbalances which is a common issue in medical datasets. In such cases, metrics like AUC, balanced accuracy, and sensitivity-specificity trade-offs offer more informative insights into model performance. Moreover, in imbalanced datasets, macro metrics are needed because they equally weight each class, ensuring that minority classes are not overshadowed by majority class performance.

Beyond metrics, the choice of validation strategy plays a critical role in determining how well a model generalizes. Cross-validation, particularly k-fold or stratified variants, is widely used to assess performance within a dataset. However, to truly evaluate generalization, external validation which means testing on an independent dataset from a different cohort or imaging site is increasingly recognized as the preferred standard. Additionally, consistent preprocessing pipelines and train-test splits are crucial for reproducibility and fair comparisons across studies. In practice, validation results can be unreliable due to inconsistencies in preprocessing steps, varying dataset sizes, and the lack of standardized sampling strategies. These factors may lead to biased evaluations and inflated performance estimates, thereby limiting the reliability and comparability of reported results (Zech, Badgeley, Liu, Costa, Titano & Oermann, 2018).

2.3 Model Architectures

Recent advances in deep learning have led to the development of specialized architectures tailored for medical imaging tasks, including AD detection from MRI scans. Among these, encoder-decoder-based architectures, attention mechanisms, and 3D CNNs have shown strong performance in capturing both spatial patterns and anatomical features relevant to neurodegeneration. This section outlines representative architectures commonly used for AD-related tasks, with a focus on their design principles and suitability for structural brain imaging.

2.3.1 Convolutional Neural Networks

CNNs form the backbone of many deep learning approaches in medical imaging, including AD diagnosis from MRI scans. Their ability to learn hierarchical spatial features makes them well-suited for identifying structural brain abnormalities associated with neurodegeneration. Traditional CNNs operate on 2D slices, often processing sagittal, coronal, or axial planes independently and later aggregating results across views (Sarraf & Tofighi, 2016). While effective in many contexts, such 2D approaches may miss inter-slice dependencies that are crucial in volumetric data.

To address the limitations of 2D approaches, 3D CNNs apply convolutions across all three spatial dimensions, enabling direct modeling of volumetric context. Many CNN variants have been explored for AD classification and feature extraction, but this thesis focuses on UNet and ResNet architectures due to their relevance and widespread use. Despite higher computational costs, 3D CNNs remain well-suited for structural neuroimaging tasks because of their ability to capture subtle anatomical patterns.

2.3.1.1 ResNet Architecture

The ResNet (Residual Network) architecture is a deep CNN model designed to address the vanishing gradient problem in very deep networks through the use of residual connections. These connections allow the network to learn identity mappings by directly passing input features across layers, facilitating more effective gradient flow during backpropagation. As shown in (Fig. 2.2), a ResNet is composed of multiple stacked residual blocks, each containing convolutional layers and shortcut connections that bypass one or more layers.

In context of AD detection using 3D MRI, ResNet is commonly adapted into 3D variants by replacing 2D operations with 3D convolutions and pooling. This allows the model to learn volumetric features and detect subtle anatomical changes across the brain. Its residual blocks and modular design support deep architectures while preserving efficient training and representational depth.

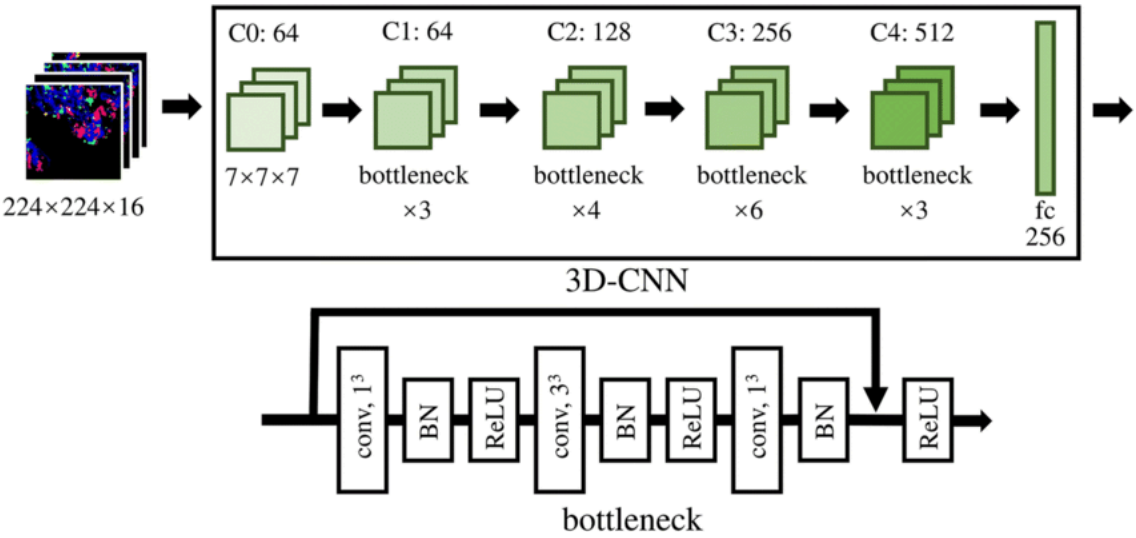


Figure 2.2 3D ResNet Architecture (adapted from (Guan et al., 2021)).

2.3.1.2 U-Net Architecture

The UNet architecture is a widely adopted CNN architecture originally developed for biomedical image segmentation, and it has been effectively adapted for tasks such as brain tissue segmentation, lesion detection, classification, and regional volumetric analysis from MRI scans. As shown in (Fig. 2.3), the symmetric encoder-decoder structure is designed to capture both high-level semantic information and fine-grained spatial details. The encoder path consists of repeated convolutional and downsampling layers that learn hierarchical features from input images. These features are then passed through a decoder path, which performs upsampling and integrates information from earlier encoder layers through skip connections. This design helps retain spatial context and sharp boundaries, which is particularly useful when identifying subtle anatomical changes associated with early-stage AD.

For 3D MRI data, 3D UNet variants are often used to improve model volumetric context across the sagittal, coronal, and axial planes. These models replace traditional 2D operations with 3D convolutions and pooling layers, enabling more comprehensive analysis of structural brain changes.

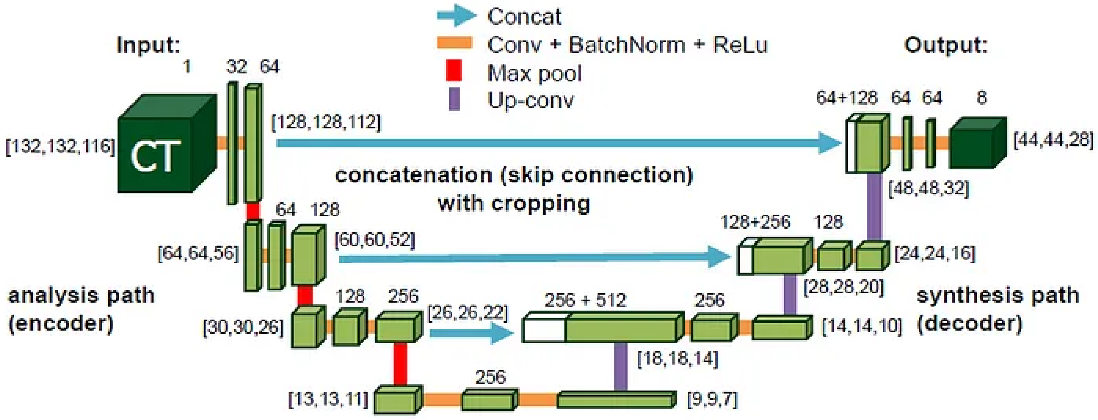


Figure 2.3 U-Net3D Architecture (adapted from (Roth et al., 2018)).

2.3.2 Morphology-Inspired Neural Architectures

2.3.2.1 Mathematical Morphology

Mathematical morphology is a nonlinear image processing framework focused on the analysis and transformation of geometrical structures within images. Unlike linear techniques that manipulate intensity values through operations like convolution and frequency filtering, morphological methods emphasize the shape preservation and modification. Central to these methods is the structuring element, a predefined shape such as a disk, square, or line which slides across the image to interact with local pixel neighborhoods. Fundamental morphological operations such as dilation and erosion apply the structuring element to expand or shrink image regions based on their spatial configuration (Soille, 2003). These operations are particularly effective in highlighting structural boundaries, suppressing noise, and extracting shape-based features critical for medical and biomedical image analysis.

2.3.2.2 Morphological Networks

Morphological networks implement the operations of mathematical morphology such as dilation and erosion within neural architectures by reformulating them in a differentiable, learnable manner (Davidson & Hummer, 1993). Based on the utility of morphological operations in this thesis, only dilation and erosion operations will be discussed. Traditional morphological operations are non-linear and non-differentiable due to the use of max and min functions. To integrate these into gradient-based training pipelines, dilation and erosion can be defined as follows:

$$y(i) = \max_{j \in W} \left(x(i+j) + w(j) \right), \tag{2.1}$$

$$y_{\text{ero}}(i) = \min_{j \in W} \left(x(i+j) - w(j) \right), \tag{2.2}$$

where x is the input feature map, w is a learnable structuring element, and W is the neighborhood window. Similarly, erosion replaces the max operation with a min. Since max and min are non-differentiable, smooth approximations such as the log-sum-exp function are often used to enable gradient flow during training. The maximum can be approximated as:

$$\max(a_1, a_2, \ldots, a_n) \approx \frac{1}{\beta} \log \left(\sum_{i=1}^{n} e^{\beta a_i} \right), \tag{2.3}$$

where $\beta > 0$ controls the sharpness of the approximation. A higher β yields a result closer to the true max, while a lower β produces a smoother output. Similarly, the minimum can be approximated using the identity $\min(a) = -\max(-a)$, which gives:

$$\min(a_1, a_2, \ldots, a_n) \approx -\frac{1}{\beta} \log \left(\sum_{i=1}^{n} e^{-\beta a_i} \right). \tag{2.4}$$

These smooth functions allow morphological layers to remain differentiable and trainable within standard backpropagation-based optimization frameworks.

An alternative implementation strategy, introduced by (Hu, Belkhir, Angulo, Yao & Franchi, 2022), avoids explicit approximations by simulating morphological behavior through pseudo morphological modules. These approximate dilation and erosion using a combination of convolution and pooling. A convolutional layer first expands the input channels, and a pixel shuffle operation reshapes the output to form local receptive fields analogous to a morphological window. To simulate dilation, a max pooling layer is applied over this rearranged feature map, extracting the local maximum equivalent to traditional dilation. Erosion is approximated by first negating the feature map, applying max pooling, and then re-negating the result—equivalent to computing a minimum via the identity. this approach is further expanded on in Section 4.2. Since all operations are differentiable, the module enables efficient, end-to-end learning of morphological transformations within deep networks.

2.4 Deep Learning Approaches for AD Detection and Generalization

2.4.1 AD Detection Approaches

Numerous studies have investigated AD detection using MRI data. A Multi-Attention-based Global 3D ResNet architecture is introduced in (Li et al., 2024), enhancing feature representations using channel and spatial attention mechanisms, along with a non-local block to capture long-range dependencies. Similarly, the 3D Global Fourier Network (Zhang et al., 2022) uses global frequency filtering instead of spatial convolutions for long-range dependency modeling, achieving strong performance on ADNI and AIBL datasets. AMSNet (Wu, Zhou, Zeng, Qian & Song, 2022) employs a multi-scale integration block and soft attention to extract hierarchical spatial features. Lightweight CNN models such as LHAttNet (Jabason, Ahmad & Swamy, 2025) apply dual attention modules to extract contextual features from 2D MRI slices.

Other studies propose architectures like MMRN (Han et al., 2023), which treat multiple brain template registrations as self-supervised augmentations and introduce meta-information regularization. The Dense Attention Network proposed in (Gan, Lan, Huang, Su & Huang, 2025) leverages linear attention within dense connections to reduce parameters while maintaining accuracy. A dual-branch CNN in (El-Assy et al., 2024) combines different receptive fields and pooling strategies for diverse feature capture. Similarly, a 3D densely connected CNN with connection-wise attention (Zhang, Zheng, Gao, Feng, Liang & Long, 2021) and the ADNet model (Folego et al., 2020) further demonstrate full-volume 3D CNN pipelines for efficient biomarker learning. Ensemble methods, such as in (Rahim, Ahmad, Ullah, Bedi & Jung, 2025), integrate Bayesian-optimized classifiers and CNN attention modules for early-stage AD detection. The Biceph-Net architecture (Rashid, Gupta, Gupta & Tanveer, 2022) also offers a compact dual-branch solution for extracting both structural and similarity-based features from 2D MRIs.

Transformer-based models have also garnered increasing attention due to their ability to capture global context and long-range dependencies. For instance, a modified ViT pipeline for MRI enhancement is proposed in (Joy et al., 2025), aiming to better preserve structural details during image reconstruction. Similarly, RanCom-ViT (Lu et al., 2025) introduces token compression techniques to reduce computational overhead while maintaining classification performance. In a related approach, Joint Transformer models (Alp et al., 2024) are designed to simultaneously model spatial features across brain regions and the temporal progression of AD. Moreover, multimodal fusion strategies, such as the one in (Duong et al., 2025), leverage complementary information from MRI and PET using both self-attention and cross-attention mechanisms to improve diagnostic accuracy. In addition, cross-modal ViTs (Jabason et al., 2025) integrate features across imaging modalities to enhance robustness against variations in the input data. Ensemble ViTs (Shaffi et al., 2024) further improve performance in imbalanced and limited data settings by aggregating predictions from multiple transformer branches.

Graph-based methods also offer promising alternatives. DAGNN (Gamgam et al., 2024) uses disentangled attention on brain connectivity graphs. While, AD-GNN (Wu et al., 2022) and CSWCL-GCN (Hu et al., 2024) capture spatial and demographic relationships with graph pooling and contrastive loss. Disentanglement-based architectures (Chen et al., 2023) fuse modality-specific and shared representations for imputation and classification. Hybrid frameworks like 3D HCCT (Majee et al., 2024) combine convolution and transformer elements with minimal preprocessing pipelines. Similarly, semi-supervised models like (Aviles-Rivero, Runkel, Papadakis, Kourtzi & Schönlieb, 2022) apply hypergraph learning to capture high-order intermodal relationships. While these models achieve strong performance when trained and evaluated on data from the same distribution, their ability to generalize across domains remains a challenge, motivating recent research in domain generalization for AD detection.

2.4.2 Domain Generalization in AD Detection

Domain Generalization in AD detection aims to build models that perform robustly across domains with different distributions. This is especially challenging in the SDG setting, where models are trained on a single source domain. A disease-informed framework using a two-stage 3D U-Net with saliency-guided attention was proposed in (Lteif et al., 2024) to enhance generalization. Another method (Zhou et al., 2023a) employed a patch-free ResNet with domain-specific encoding to disentangle invariant and variant features. Prototype-guided multi-scale domain adaptation (Cai et al., 2023) addressed both marginal and conditional distribution shifts using a combination of adversarial and metric-based alignment. In contrast, (Wang et al., 2022) embraced domain-specific variations by incorporating auxiliary demographic tasks and intra-study fine-tuning. Similarly, attention-guided deep domain adaptation (Guan, Liu, Yang, Yap, Shen & Liu, 2021) aligned features at both the representation and region levels without requiring labels from the target domain. Collective AI strategies (Nguyen et al., 2023) used ensembles of region-specific 3D U-Nets and GNNs to manage domain variability. Domain contrastive learning (Zhou, Li, Zhou, Liu & Tu, 2023b) applied a self-supervised contrastive loss in the latent space to extract domain-invariant features. A related method (Fiasam et al., 2023) combined a ResNet-based 3D CNN with a domain-aware module and multiple classifiers to learn both shared and domain-specific representations. Despite these advances, most DG approaches remain generic and do not incorporate the disease-specific anatomical patterns associated with AD. This thesis aims to address this gap by explicitly integrating disease-aware information into the generalization process.

2.4.3 General DG Approaches for Medical Image Analysis

Several DG techniques from general computer vision have been adapted for AD detection. MixUp (Zhang, Cissé, Dauphin & Lopez-Paz, 2018) and MixStyle (Zhou, Yang, Qiao & Xiang, 2021) offer simple yet effective augmentation strategies by interpolating samples or by erturbing feature statistics, respectively. However, their inherent randomness may unintentionally distort subtle pathological patterns critical for AD diagnosis. More targeted techniques include adversarial Bayesian augmentation (Cheng, Gokhale & Yang, 2023), which generates domain-shifted variants through Bayesian inference, and frequency-mixed methods (Li, Li, Zhao, Fu, Su, Hu & Liu, 2023), which operate in the spectral domain to simulate realistic distribution shifts. Style-transfer-based perturbations (Li, Zimmer, Ding, Wu, Huang, Schnabel & Zhuang, 2021) aim to replicate domain-specific appearance changes while preserving semantic content.

Beyond augmentation, Representation Self-Challenging (RSC) approach (Huang, Wang, Xing & Huang, 2020) improves generalization by identifying and suppressing overly dominant features via gradient analysis, prompting the model to learn more diverse and informative representations without architectural changes. Contrastive DG strategies like CCSDG (Hu, Liao & Xia, 2023) introduce style-augmented image pairs to disentangle style and structure at the channel level, allowing segmentation or classification to rely primarily on structure-relevant cues. Exact Feature Distribution Matching (EFDM) (Zhang, Li, Li, Jia & Zhang, 2022) enhances feature alignment by directly matching empirical cumulative distributions in latent space through a fast Sort-Matching algorithm, offering a more precise alternative to conventional Gaussian-based methods.

While these approaches have advanced DG in medical imaging, they remain largely generic and are not specifically designed for the structural complexity and progressive nature of Alzheimer's disease. This highlights the need for DG methods that explicitly incorporate anatomical and morphological cues associated with neurodegeneration.

2.4.4 Morphological Modules for AD Detection

Mathematical morphology has been extensively explored in the context of medical imaging (Gurcan, Boucheron, Can, Madabhushi, Rajpoot & Yener, 2009; James & Dasarathy, 2014; Sotiras, Davatzikos & Paragios, 2013) before the advent of deep learning due to its inherent capacity for shape analysis (Aptoula & Lefèvre, 2009). Consequently, multiple attempts have been made in order to implement its data-driven counterpart, even though its non-linearity complicates differentiability and the back-propagation of errors via gradient descent. For instance, one approach (Hu et al., 2022) introduced differentiable dilation and erosion using neural architecture search to integrate them into end-to-end models. Another method (Ghosh & Das, 2024) proposed multi-scale modules with learnable structuring elements, and a separate effort (Guzzi, Zuluaga, Lareyre, Di Lorenzo, Goffart, Chierici, Raffort & Delingette, 2024) developed soft morphological filters to optimize operations like dilation. A hybrid strategy (Canales-Fiscal & Tamez-Peña, 2023) combined fixed morphological layers with shallow convnets, though it lacked flexibility and full end-to-end training.

More specifically in the context of morphology-oriented AD detection, graph-theoretical metrics have been applied to morphological networks for early detection (Ding, Wang, Tang, Wang, Qi, Dou, Qian, Gao, Zhong, Yang & others, 2023), while mesh-based graph convolutional networks have been used in (Azcona, Besson, Wu, Punjabi, Martersteck, Dravid, Parrish, Bandt & Katsaggelos, 2020) to model cortical structure. Finally, a transformer-based model (Wang, Chen & Wang, 2024) incorporated morphology-aware augmentations but relied on external, non-learnable priors. In contrast, one of the proposed approaches in this thesis utilizes fully differentiable, class-aware pseudo-morphological modules for AD diagnosis. It learns transformations specific to each class and generates meaningful augmentations to improve neurodegeneration modeling.

2.4.5 Contrastive Learning based Approaches for AD Detection

Contrastive learning has recently gained popularity in AD detection as a powerful self-supervised and semi-supervised representation learning technique, particularly valuable in medical imaging domains. By training models to distinguish between similar and dissimilar data pairs, contrastive frameworks help uncover subtle neurodegenerative patterns characteristic of AD. One approach, called SMoCo, introduced a self-supervised model utilizing 3D amyloid-PET to predict MCI-to-AD conversion (Kwak, Park, Lee & others, 2023). Another method, known as Brain-Aware Replacements, leveraged region-specific mixup and supervised contrastive loss to boost classification performance (Seyfioğlu, Liu, Kamath, Gangolli, Wang, Grabowski & Shapiro, 2022). In a separate study, a 3D CNN-based multichannel contrastive learning framework was developed to diagnose AD and MCI using T1-weighted MRI. This method combines supervised classification loss with unsupervised contrastive loss by applying multiple data transformations (e.g., noise injection), enhancing both accuracy and generalization on the ADNI dataset (Wang, Hu & others, 2023). Furthermore, a technique called Domain Contrastive Learning was introduced to address variability in multi-site MRI datasets by extracting domain-invariant features within an autoencoder's latent space. Evaluated across three sites, the approach significantly improved classification robustness and transferability, outperforming baselines in AD vs. NC tasks (Li, Zhang & others, 2023). These diverse approaches collectively underscore the growing utility of contrastive learning in achieving more reliable, generalizable, and interpretable AD detection across modalities and data environments.

3. DATASET AND IMPLEMENTAITON DETAILS

3.1 Dataset

This thesis utilizes three publicly accessible datasets: NACC (Beekly, Ramos, van Belle, Deitrich, Clark, Jacka & Kukull, 2004), ADNI (Petersen et al., 2010), and AIBL (Ellis et al., 2009). Each dataset contains 3D MRI scans grouped into three diagnostic categories: NC, MCI, and AD. Only subjects aged 55 and older were considered to ensure age consistency. The 3D MRI is presented as 2D slices for visualization in (Fig. 3.1) prior to any preprocessing steps.

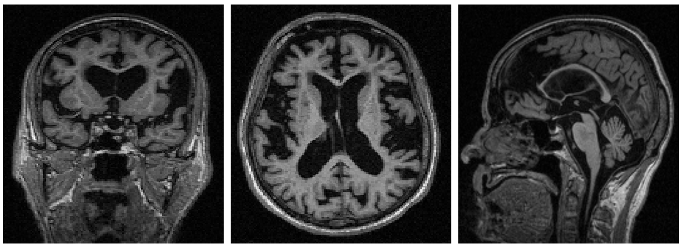
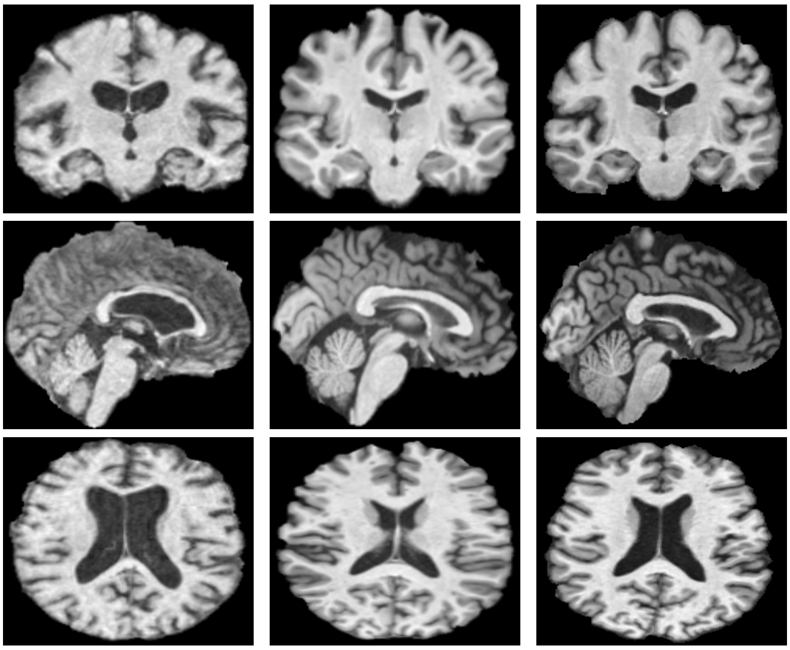


Figure 3.1 MRI samples without preprocessing.

To maintain uniformity across datasets, all MRI volumes underwent preprocessing using a standardized pipeline adapted from (Qiu, Miller, Joshi, Lee, Xue, Ni, Wang, De Anda-Duran, Hwang, Cramer & others, 2022), which includes registration to the MNI152 template, skull stripping to eliminate extra brain tissues, and bias field correction to minimize intensity inhomogeneity. Due to the differing acquisition protocols and scanner configurations across datasets, domain shift is inherently present. Labels were used as provided, and no subtypes were distinguished within the MCI class. For model development, 80% of the NACC data was used for training and 20% for validation. Generalization performance was evaluated using 80% subsets of ADNI and AIBL.

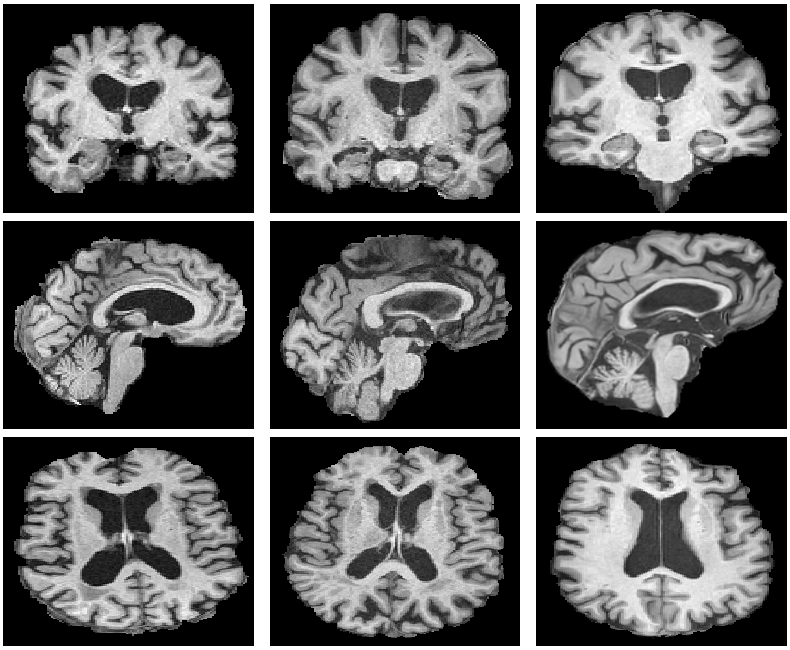
Figures 3.2, 3.3 and 3.4 present representative MRI slices from the NACC, ADNI and AIBL datasets, respectively. These 2D slices are extracted from 3D MRI volumes only for the purpose of visualization. For each dataset, subfigure (a) shows a subject with AD, (b) shows a subject with NC, while (c) shows a subject with MCI. The slices highlight both anatomical variability across diagnostic groups and structural differences between datasets. The NACC volumes have a resolution of $182 \times 218 \times 182$ pixels with isotropic voxel spacing of 1 mm. ADNI scans exhibit a larger volume size of $166 \times 256 \times 256$ pixels, also with 1 mm isotropic spacing. In contrast, AIBL volumes are sized $160 \times 240 \times 256$ pixels with anisotropic spacing of approximately $[1.20, 1.00, 1.00]$ mm/pixel, yielding physical dimensions of roughly $192 \times 240 \times 256$ mm. Moreover, while intra-dataset variations such as cortical thickness, ventricle size, and localized atrophy pose inherent classification challenges, inter-dataset differences also complicate model generalization. Together, these examples underscore the need for robust, domain-aware learning strategies capable of adapting to both clinical and technical variability.



(a) AD case (b) NC case (c) MCI case

Figure 3.2 NACC dataset samples: (a) AD case showing pronounced cortical atrophy, (b) NC subject with intact brain structure, and (c) MCI case exhibiting intermediate anatomical features between AD and NC.

Moreover, demographic and class distribution details of the datasets are presented in Table 3.1. Class imbalance remains a key challenge, particularly due to the relatively low number of AD samples, as evidenced in Table 3.1. This imbalance not only affects model learning but also significantly hinders generalization, especially when distinguishing early-stage AD from MCI or NC. Cross-validation was not conducted,



(a) AD case (b) NC case (c) MCI case

Figure 3.3 ADNI dataset samples: (a) AD case showing pronounced cortical atrophy, (b) NC subject with intact brain structure, and (c) MCI case exhibiting intermediate anatomical features between AD and NC.

primarily due to computational and time constraints since, 3D nature of the input volumes significantly increases training time per epoch.

Table 3.1 Demographic Characteristics of Participants in NACC, ADNI, and AIBL Datasets.

Dataset	**Group**	**Age, years** mean ± std	**Gender** (male count)
NACC (Beekly et al., 2004)	NC (n=2524)	69.8± 9.9	871 (34.5%)
	MCI (n=1175)	74.0 ± 8.7	555 (47.2%)
	AD (n=948)	75.0 ± 9.1	431 (45.5%)
ADNI (Petersen et al., 2010)	NC (n=684)	72.3 ± 6.9	294 (43.0%)
	MCI (n=572)	73.8 ± 7.5	337 (58.9%)
	AD (n=317)	75.1 ± 7.7	168 (53.0%)
AIBL (Ellis et al., 2009)	NC (n=465)	72.3 ± 6.2	197 (42.4%)
	MCI (n=101)	74.5 ± 7.2	53 (52.5%)
	AD (n=68)	73.0 ± 8.2	27 (39.7%)

Another critical limitation lies in the skull stripping step of the preprocessing pipeline. Despite using widely accepted tools, errors in skull stripping occasionally result in unintended removal of brain tissue, particularly at the boundaries of the frontal and temporal lobes. (Fig. 3.5) illustrates an example of such a failure. These artifacts not only distort the anatomical structure but can also mislead the model during training, causing it to focus on erroneous regions unrelated to disease

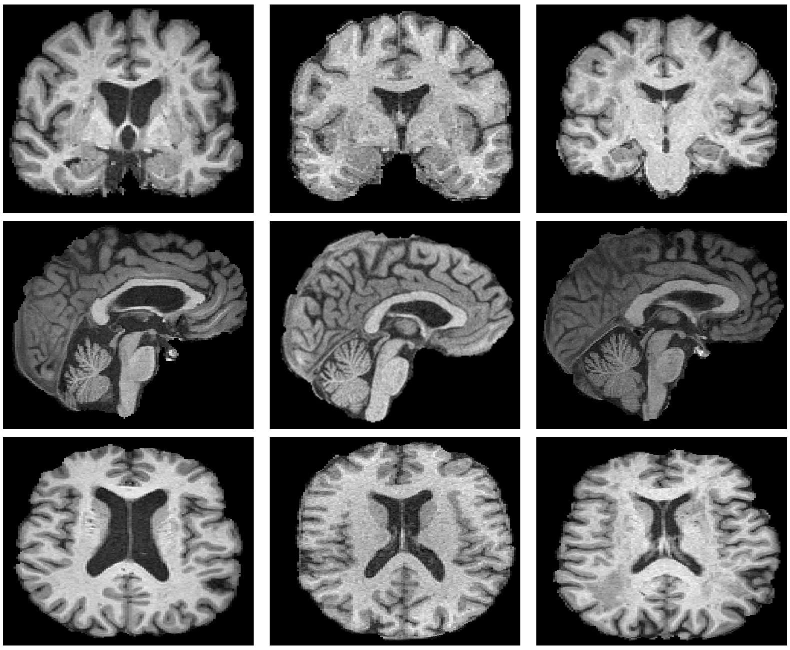

(a) AD case (b) NC case (c) MCI case

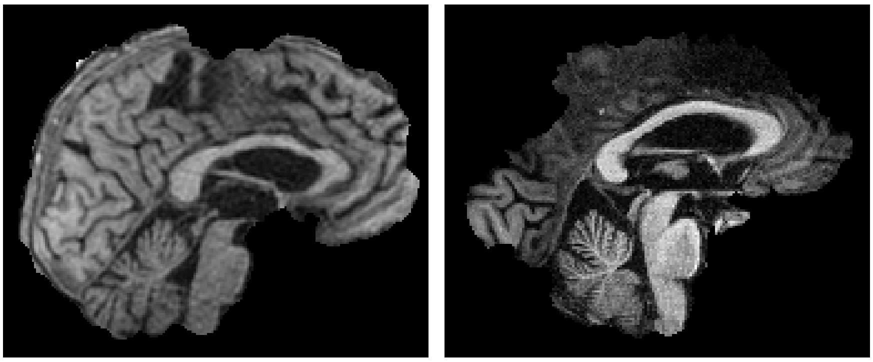
Figure 3.4 AIBL dataset samples: (a) AD case showing pronounced cortical atrophy, (b) NC subject with preserved brain volume, and (c) MCI case with intermediate structural features.

pathology. This can ultimately degrade the model's ability to generalize and correctly detect disease patterns. Although such artifacts can affect learning, removing all problematic samples is impractical due to the already limited size of the datasets and the class imbalance.

3.1.1 Limitations in Dataset Expansion and Preprocessing Considerations

Although incorporating additional datasets could improve model generalizability, we chose to limit our work to three, well-curated dataset due to several practical and methodological constraints. As shown in (Fig. 3.1), raw images often include non-brain structures such as the skull and neck, necessitating skull stripping. However, this step is error-prone and can lead to loss of critical brain tissue, as illustrated in (Fig. 3.5). These distortions not only compromise anatomical accuracy but also introduce noise into model training, reducing reliability.

In addition, the preprocessing pipeline is highly time-intensive, requiring multiple stages such as reorientation, intensity normalization, and quality control—each demanding both computational resources and manual oversight. Every single sam-



(a) Skull-stripping error in ADNI sample

(b) Skull-stripping error in AIBL sample

Figure 3.5 Skull-stripping quality visualization for (a) ADNI and (b) AIBL datasets.

ple must be manually inspected to ensure preprocessing integrity. Expanding to more datasets would further amplify these demands. Furthermore, inconsistencies in imaging protocols, label definitions, and data access across sources pose significant integration challenges. To maintain consistency, reduce complexity, and ensure data quality, we limited our analysis to a single dataset with controlled preprocessing.

3.2 Implementation Details

All experiments were carried out using an A6000 GPU (48 GB). Due to GPU memory constraints, training began with a batch size of 2, which was effectively scaled to 16 via gradient accumulation. The reason for lower batch size is due to high memory demands of processing full-resolution 3D MRI volumes. The model was optimized using stochastic gradient descent with a learning rate of 0.01, momentum set to 0.9, and a weight decay of 0.0005. An exponential learning rate scheduler was applied, decaying the learning rate by 5% after each training epoch. To mimic domain shift scenarios encountered in real-world applications, the standard SDG protocol described in (Qiao, Zhao & Peng, 2020) was followed. Specifically, **the model was trained and validated exclusively on the NACC dataset, then tested on two distinct target domains: ADNI and AIBL which differ in scanner hardware, imaging protocols, and demographic characteristics.** This setup allows us to evaluate the model's ability to generalize under out-of-distribution conditions.

3.2.1 Baselines

The proposed methods were evaluated against a baseline encoder without any SDG components, along with several established SDG techniques. MixUp (Zhang et al., 2018) was applied with an interpolation factor of $\alpha = 0.3$ to encourage smoother decision boundaries through sample interpolation. RSC (Huang et al., 2020) was configured with 20% feature dropout, 5% background dropout, and a 0.3 mixing probability to suppress dominant features and promote learning from alternative cues. EFDM (Zhang et al., 2022) used a patch replacement probability of $p = 0.5$ and $\alpha = 0.1$, aiming to maintain consistent feature distributions across domains. Finally, MixStyle (Zhou et al., 2021) was applied with $\alpha = 0.1$ and $p = 0.5$, facilitating AD detection by mixing instance-level feature statistics.

All baselines used a 3D pretrained U-Net architecture (Zhou, Sodha, Siddiquee, Feng, Tajbakhsh, Gotway & Liang, 2019) with four convolutional blocks and filter sizes increasing from 32 to 512. Performance was assessed using accuracy, F1 score, sensitivity, and specificity to provide a comprehensive evaluation of model behavior across key clinical metrics.

3.2.2 Pretraining with Chest CT Data: Rationale and Limitations

The pretrained checkpoint used for the 3D U-Net architecture in this work was derived from (Zhou et al., 2019), which was trained on unlabeled chest CT scans. Model Genesis is a self-supervised learning approach that trains models to learn transferable 3D representations by solving tasks like image restoration, inpainting, and geometric transformations, enabling structural understanding without labels.

Although chest CT and brain MRI belong to different imaging domains, both share essential 3D structural characteristics, making such pretraining beneficial. To the best of our knowledge, there are no publicly available 3D U-Net checkpoints pre-trained specifically on brain MRI using self-supervised or large-scale annotated datasets. Given this gap, the Model Genesis checkpoint trained on chest CT offers a strong initialization, improving convergence and generalization for AD detection with limited or imbalanced brain imaging data.

3.2.3 Weighted Cross-Entropy Loss for Class Imbalance

Class imbalance is a common challenge in medical imaging datasets, particularly in disease classification tasks where certain diagnostic categories may be significantly underrepresented. This imbalance can bias the model toward majority classes, resulting in degraded performance on minority categories such as AD.

To mitigate this issue, a weighted cross-entropy (WCE) loss function is employed. The standard cross-entropy (CE) loss is defined as:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{C} y_i \log(p_i) \tag{3.1}$$

where C denotes the number of classes, y_i is the one-hot encoded ground truth label, and p_i is the predicted probability for class i.

In the weighted variant, each class is assigned a weight w_i, resulting in:

$$\mathcal{L}_{WCE} = -\sum_{i=1}^{C} w_i y_i \log(p_i) \tag{3.2}$$

The weights w_i are typically computed as the inverse of the class frequency. By increasing the penalty for misclassifying underrepresented classes, this approach encourages the model to treat all classes more equally and reduces bias toward dominant categories such as NC.

4. SINGLE DOMAIN GENERALIZATION FOR ALZHEIMER'S DETECTION FROM 3D MRI SCANS WITH PSEUDO-MORPHOLOGICAL AUGMENTATIONS AND CONTRASTIVE LEARNING

4.1 Overview

While numerous deep learning approaches have been proposed for AD diagnosis from 3D MRI data, most fail to incorporate disease-specific anatomical priors, particularly the progressive nature of brain atrophy that characterizes AD. As discussed in Section 2, existing methods primarily focus on architectural modifications or DG techniques, with limited attention to the underlying structural changes that differentiate stages such as AD, MCI, and NC. This oversight limits the model's ability to capture biologically meaningful variations that are crucial for fine-grained classification. This is a critical gap, as atrophy is not only a defining hallmark of AD but also varies in severity across classes: it is more pronounced in AD, minimal in NC, and intermediate in MCI. Furthermore, the challenge is compounded by domain shifts across sites, and severe class imbalance. These factors call for a method that can both generalize across domains and meaningfully encode the anatomical progression of the disease.

To address this gap in literature, the proposed method combines a 3D U-Net encoder with class-specific 3D learnable morphological augmentations and supervised contrastive learning in order to improve AD detection under a realistic SDG setting. In this setting, it is assumed that only a single labeled training dataset is available, without access to any target domain data or labels. This combination of components encourages the model to generate more diverse, anatomically relevant variations while simultaneously enforcing representation consistency across domains. During training, batches with only MCI samples undergo CutMix (Yun, Han, Oh, Chun, Choe & Yoo, 2019), a technique that swaps patches between images to increase intra-class diversity in addition to standard augmentations. For the remaining two classes, AD and NC, 3D learnable morphological augmentations are computed via erosion and dilation respectively, motivated by the fact that these operators can visually intensify or alleviate the effect of brain atrophy. Erosion is applied to simulate more pronounced atrophy for AD, while dilation reduces the atrophy effect for NC. Both original and augmented samples are processed through the encoder and a projection head, to compute contrastive embeddings. The total loss combines weighted cross-entropy loss with inverse class frequency weights and supervised contrastive loss as illustrated in (Fig. 4.1) (Algorithm 1).

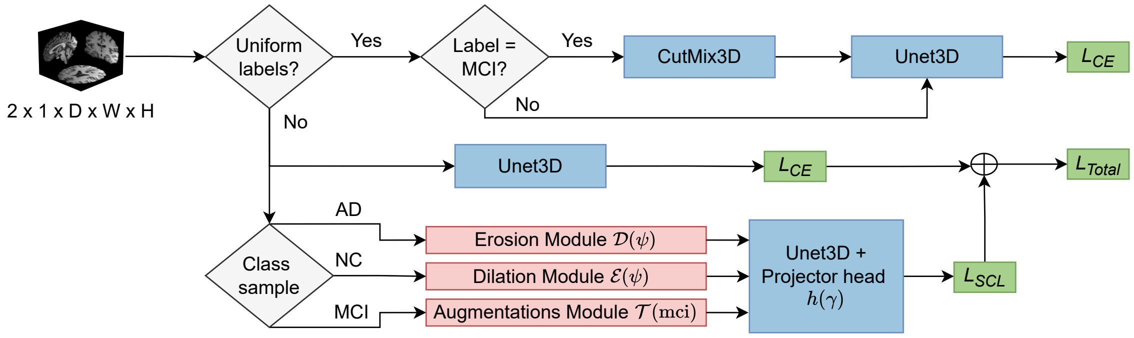


Figure 4.1 Overview of the proposed training pipeline. Class-specific augmentations (erosion, dilation, or diverse augmentations) and CutMix3D (for uniform MCI batches) enhance feature diversity. A shared 3D U-Net encoder processes original and augmented views, optimized using a combined cross-entropy and supervised contrastive loss.

4.2 Model Architecture

The classification framework is built upon a pretrained 3D U-Net architecture (Zhou et al., 2019), modified for feature extraction by removing the decoder and using only the encoder (Fig. 4.2). The encoder processes input 3D MRI volumes through four convolutional blocks, each consisting of two 3D convolutional layers, batch normalization, and a non-linear activation function. As the input passes through these blocks, the spatial resolution decreases while feature dimensionality increases: the number of channels progresses from 32 to 64, 128, 256, and finally 512 across the four blocks. This hierarchical structure enables the network to capture increasingly abstract anatomical and structural patterns associated with Alzheimer's disease. The deepest layer produces a 512-channel volumetric feature map embedded with contextual and pathological information, which is then global average-pooled into a compact feature vector for final classification.

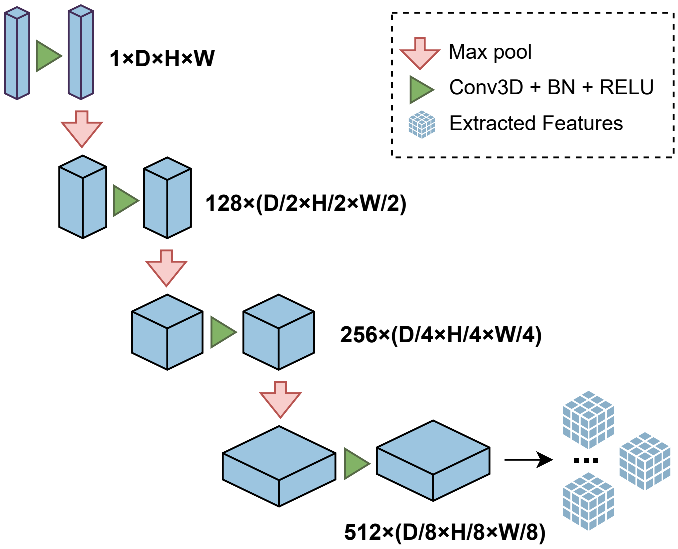


Figure 4.2 Encoder architecture of 3D U-Net.

4.3 Class-Specific Augmentations

To address distributional shifts between training and testing domains, class-specific learnable morphological 3D augmentation is incorporated. While standard augmentation techniques that apply identical transforms may improve robustness, they often miss disease-specific changes, potentially blurring class differences. To prevent this, the proposed method uses two learnable morphological modules (Hu et al., 2022) that approximate erosion and dilation with differentiable operations. Erosion and dilation are the two fundamental operations of mathematical morphology, known for expanding image regions that are respectively darker and brighter w.r.t. their surroundings (Serra, 2023). As such, when applied to a brain's MRI sample, erosion's effect becomes visually similar to intensifying brain atrophy (i.e. less brain matter). These modules learn transformations based on local image structure, which is an essential property in 3D brain MRI, whereas fixed kernels may fail across different patients or regions. Since these modules mimic morphological behavior without an underlying complete lattice foundation, they are referred to as "pseudo-morphological". Both modules are trained jointly with the main model and optimized to generate augmentations that enhance intra-class variability.

4.3.1 Pseudo-Dilation Module ($\mathcal{D}_\psi$)

This module's task is to generate synthetic NC samples, corresponding to healthy non-AD brain images. It achieves this by expanding regions of greater pixel intensity w.r.t. their surrounding via a learnable 3D dilation. More formally, given a 3D grayscale input $I : \mathbb{Z}^3 \to [0, 255] \cap \mathbb{Z}$ and a cubic structuring element $S_k \subseteq \mathbb{Z}^3$ of size $k \times k \times k$ pixels, 3D grayscale (flat) dilation is defined as:

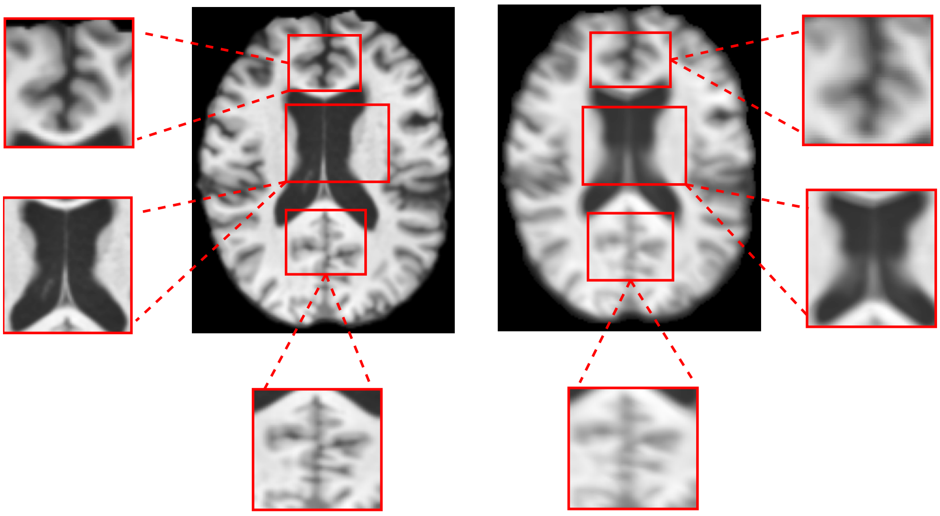
$$(I \oplus S_k)(x, y, z) = \max_{(s,t,u) \in S_k} I\Big(x - s,\ y - t,\ z - u\Big) \tag{4.1}$$

In the context of the present task, this operation simulates the increase of brain matter and thus helps diversify NC samples to reduce the risk of overfitting to the NC class and strengthens inter-class separability.

We adopt the learnable pseudo-dilation module design presented in (Hu et al., 2022), where the process is parameterized and learned via $\mathcal{D}_\psi$, and ψ denotes the learnable parameters. Let $x^{(l-1)} \in \mathbb{R}^{B \times C_{\text{in}} \times D \times H \times W}$ be the input to layer l, where B is the batch size, C_{in} the number of input channels, and D, H, W are the spatial dimensions. At each layer l, the input sample is passed through a 3D convolution with a randomly selected kernel size $k \in \{3, 5\}$ to simulate anatomical variation at different scales. The convolutional output has shape $B \times (C_{\text{out}} \cdot k^3) \times D \times H \times W$, where C_{out} is the number of output channels. The learnable dilation is then performed by taking the maximum over the kernel positions. The following operation mimics the effect of dilation on the input:

$$x^{(l)} = \max_{\text{kernel channels}} (\text{Conv3D}_k(x^{(l-1)})) \tag{4.2}$$

where Conv3D_k denotes a 3D convolution with kernel size k, and $x^{(l)}$ is the resulting feature map. To preserve anatomical boundaries, and restrict the transformation to foreground regions, a binary foreground mask $M = \mathbb{1}_{x \neq 0} \in \{0, 1\}^{B \times 1 \times D \times H \times W}$ is applied at each layer. As shown in (Fig. 4.3), the dilation-based transformation produces outputs that mimic the expanded appearance typical of healthy brain anatomy in NC cases.



(a) Input NC image (b) Output after dilation

Figure 4.3 Visualization of a pseudo-dilation result. (a) Original NC image from the dataset. (b) Resulting image after applying the pseudo-dilation module.

4.3.2 Pseudo-Erosion Module ($\mathcal{E}_\psi$)

This module is designed to simulate brain tissue loss, similar to that observed in AD (Hu et al., 2022). It works by expanding image regions darker than their surroundings, producing more severe-looking AD samples from existing ones. More formally, given a 3D grayscale input $I : \mathbb{Z}^3 \to [0,255] \cap \mathbb{Z}$ and a cubic structuring element $S_k \subseteq \mathbb{Z}^3$ of size $k \times k \times k$ pixels, 3D grayscale (flat) erosion is defined as:

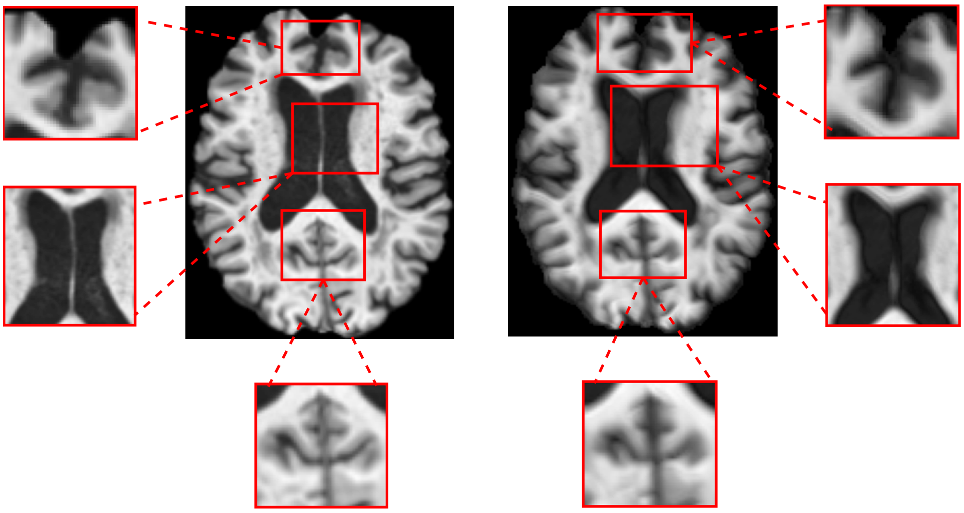
$$(I \ominus S_k)(x,y,z) = \min_{(s,t,u)\in S_k} I\Big(x-s,\ y-t,\ z-u\Big) \tag{4.3}$$

$\mathcal{E}_\psi$ approximates erosion using learnable 3D convolutional layers. At each layer, the input is passed through a 3D convolution with a randomly selected kernel size $k \in \{3,5\}$. Contrary to the pseudo-dilation module, a binary mask M is applied before the convolution so that background zeros do not influence the minimum.

To simulate erosion, the convolutional output is reshaped and the local minimum is taken over the kernel dimension. The erosion operation is then computed as:

$$x^{(l)} = -\min_{\text{kernel channels}} (-\text{Conv3D}_k(x^{(l-1)})) \tag{4.4}$$

Here, the negation of the output followed by a minimum operation is mathematically equivalent to erosion (Hu et al., 2022). Finally, the foreground mask M is reapplied to restrict updates to valid tissue regions only. This method enables the network to learn how to apply erosion-like transformations in a differentiable way. The resulting output is expected to resemble more advanced cases of AD (Fig. 4.4).

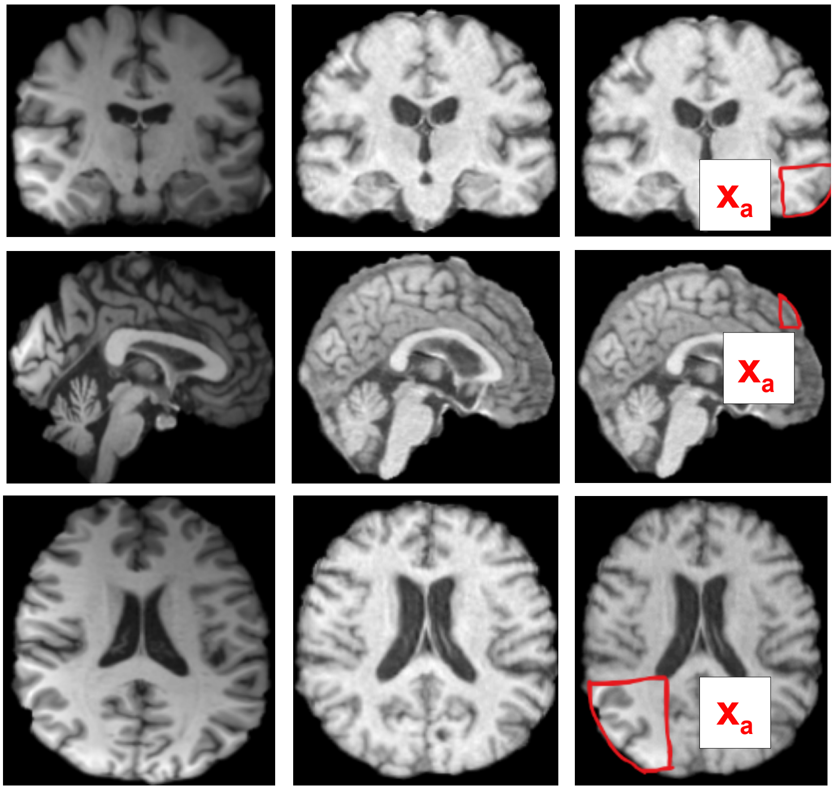


(a) Input AD image (b) Output after erosion

Figure 4.4 Visualization of a pseudo-erosion result. (a) Original AD image from the dataset. (b) Resulting image after applying the pseudo-erosion module, showing features akin to more severe AD.

4.3.3 Augmentations module for MCI ($\mathcal{T}_{\mathbf{mci}}$)

As a transitional state between NC and AD, MCI shows anatomical changes that may overlap with features of either class. This makes it particularly challenging for models to differentiate MCI from neighboring classes. Given an MCI sample, random affine transformations are applied in order to introduce slight translation, scaling and contrast adjustments so as to produce an augmented sample. These augmentations do not involve morphological changes, as such operations could risk rendering MCI samples more similar to either of its neighboring classes. Instead, they introduce moderate variations that increase diversity while keeping the anatomy realistic and helping the model generalize better to borderline cases. To further diversify MCI representations, if a batch consists of only MCI samples, sub-volume mixing is applied, where a random 3D volume is extracted from one sample and inserted at the same position in the other sample. This encourages the model to capture variations within each class while still keeping the classes separated (Fig. 4.5).



(a) x_a (b) x_b (c) $\tilde{x}_a$

Figure 4.5 CutMix pipeline: (a) Source image x_a, (b) Source image x_b, (c) Region-wise patches are extracted and swapped to generate mixed image $\tilde{x}_a$ in 3D space.

4.3.4 Weighted Supervised Contrastive Learning

To encourage domain-invariant and class-discriminative representations, the features are provided as input to a projection head.

$$q_i = \frac{h_\gamma(f_\theta(x_i))}{\|h_\gamma(f_\theta(x_i))\|_2} \tag{4.5}$$

where x_i denotes the i-th input sample, which may be an original or augmented image; $f_\theta(\cdot)$ is the encoder network parameterized by θ, used to extract feature representations; and $h_\gamma(\cdot)$ is the projection head parameterized by γ, consisting of 3D convolutions, that maps the features to a lower-dimensional embedding space. Lastly, the output $q_i \in \mathbb{R}^{1024}$ is L2-normalized using the Euclidean norm $\|\cdot\|_2$.

These projected features are then used to compute supervised contrastive loss (Khosla, Teterwak, Wang, Sarna, Tian, Isola, Maschinot, Liu & Krishnan, 2020). In more detail, given $2N$ samples (i.e. originals and their augmentations), the loss is defined as:

$$\mathcal{L}_{\text{SCL}} = \sum_{i=1}^{2N} \frac{-1}{|P(i)|} \sum_{j \in P(i)} w_{y_i} \cdot \log \frac{\exp(q_i^\top q_j/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(q_i^\top q_k/\tau)} \tag{4.6}$$

where $\mathcal{L}_{\text{SCL}}$ is the supervised contrastive loss, $q_j \in \mathbb{R}^{1024}$ denotes the L2-normalized embedding of a positive sample from the same class as i, computed in the same way as q_i. The set $P(i)$ contains the indices of all positive samples that share the same class label as i. The temperature parameter τ controls the sharpness of the distribution, where $\tau > 0$. Additionally, $w_{y_i} \in \mathbb{R}$ represents the weight assigned to class y_i, allowing for class-aware reweighting to address class imbalance. Hence, supervised contrastive learning pulls together representations of samples from the same class while pushing apart those from different classes, leading to more structured and discriminative feature embeddings.

The overall training algorithm is shown at (Algorithm 1).

4.4 Results and Discussion

The proposed method consistently achieved superior generalization performance compared to baseline approaches on both the ADNI and AIBL datasets. Tables 4.1 and 4.2 summarize the results across multiple evaluation metrics.

Table 4.1 shows that compared to the strongest baseline (RSC), the proposed approach improved accuracy by 4.77 percentage points, F1 score by 1.7 percentage points, sensitivity by 2.7 percentage points, and specificity by 1.6 percentage points. These improvements indicate enhanced domain-invariant feature learning and better separation between classes. It also validates the effectiveness of morphological priors in handling domain shifts in AD detection.

Table 4.2 presents the generalization results on the AIBL dataset. While the EFDM method attained the highest accuracy of 69.94%, its F1 score and sensitivity were notably lower, suggesting that higher accuracy is due to bias towards the majority class in the imbalanced dataset. In contrast, the proposed method demonstrated a strong trade-off between these metrics, with improvements of 15.5 percentage points in F1 score and 12.3 percentage points in sensitivity over EFDM.

Across methods, gains on AIBL exceed those on ADNI, highlighting the challenge

Algorithm 1: Training with Class-Specific Augmentation and Supervised Contrastive Learning

```
Input: Model with projection head $h_\gamma$; augmenters $\mathcal{D}_\psi$, $\mathcal{E}_\psi$, $\mathcal{T}_{\text{mci}}$;
       CutMix3D module $\mathcal{C}$; training set $\mathcal{X}$; temperature $\tau$; contrastive
       weight $\lambda$
foreach minibatch $\{(x_i, y_i)\}_{i=1}^{N}$ do
    if all labels in minibatch have the same label then
        if labels are MCI then
            Apply CutMix3D: $x_i^{\text{mix}} \leftarrow \mathcal{C}(x_i)$
            Compute classification loss $\mathcal{L}_{\text{CE}}$
            return $\mathcal{L}_{\text{CE}}$
        else
            Compute classification loss $\mathcal{L}_{\text{CE}}$
            return $\mathcal{L}_{\text{CE}}$
    Compute classification loss $\mathcal{L}_{\text{CE}}$
    Initialize augmented set $\mathcal{A} \leftarrow \emptyset$
    foreach $(x_i, y_i)$ do
        if $y_i = 0$ (NC) then
            $x_i^{\text{aug}} \leftarrow \mathcal{D}_\psi(x_i)$
        else if $y_i = 1$ (MCI) then
            $x_i^{\text{aug}} \leftarrow \mathcal{T}_{\text{mci}}(x_i)$
        else
            $x_i^{\text{aug}} \leftarrow \mathcal{E}_\psi(x_i)$
        Add $(x_i^{\text{aug}}, y_i)$ to $\mathcal{A}$
    foreach $(x_i, y_i) \in A$ do
        Compute normalized projection $q_i$
    Compute supervised contrastive loss $\mathcal{L}_{\text{SCL}}$
    $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda \cdot \mathcal{L}_{\text{SCL}}$
```

Table 4.1 Results with the ADNI dataset.

Method	**ADNI**			
	ACC(%)	F1	SEN	SPE
Baseline (Zhou et al., 2019)	38.04	0.359	0.359	0.679
Mixup (Zhang et al., 2018)	48.29	0.339	0.392	0.703
RSC (Huang et al., 2020)	46.14	0.407	0.410	0.713
CCSDG (Hu et al., 2023)	39.55	0.396	0.419	0.700
EFDM (Zhang et al., 2022)	45.35	0.249	0.353	0.679
Proposed	**50.91**	**0.424**	**0.437**	**0.729**

Table 4.2 Results with the AIBL dataset.

Method	**AIBL**			
	ACC(%)	F1	SEN	SPE
Baseline (Zhou et al., 2019)	38.50	0.338	0.392	0.699
Mixup (Zhang et al., 2018)	65.42	0.382	0.382	0.721
RSC (Huang et al., 2020)	51.27	0.414	0.449	0.737
CCSDG (Hu et al., 2023)	40.82	0.396	0.401	0.699
EFDM (Zhang et al., 2022)	**69.94**	0.301	0.329	0.678
Proposed	62.27	**0.456**	**0.452**	**0.742**

of domain shift in medical imaging. In contrast, the proposed method consistently achieves the highest macro F1 on both datasets, with a 4.2 percentage points gain over the next best method (RSC) on AIBL and a 1.7 percentage points gain on ADNI. Notably, since the training is conducted on NACC and validated on ADNI and AIBL, the result in Tables 4.1 and 4.2 suggest that the NACC distribution is more closely aligned with ADNI, enabling better adaptation and performance, while the greater divergence from AIBL poses a more significant generalization challenge. Despite the distribution shifts, the improvement in results through the proposed approach validate that integrating class-specific augmentations and contrastive learning improve domain generalization and preserve class boundaries.

4.5 Ablation study

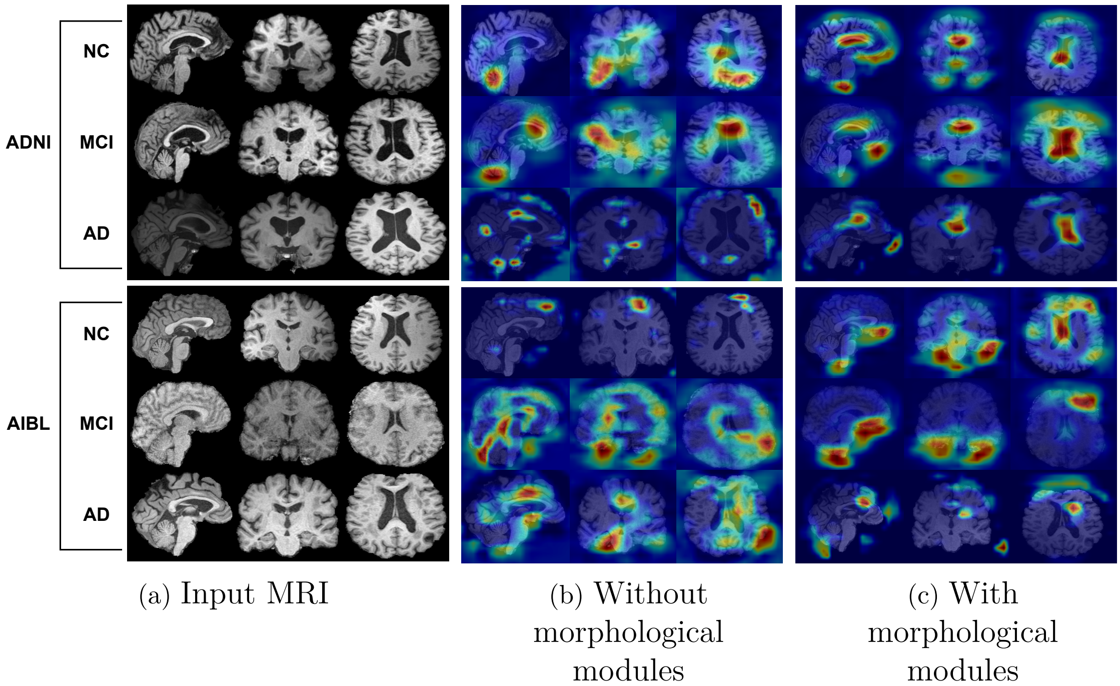
Table 4.3 presents the results of an ablation study conducted to assess the individual contributions of key components in the proposed approach. Removing the morphological modules results in a drop in macro F1 score by 6.5 percentage points on ADNI and 11.8 percentage points on AIBL. This shows the importance of structural information in tackling domain shift. Without the supervised contrastive loss, the F1 score drops by 10.2 percentage points and sensitivity by 6.6 percentage points on ADNI. This means the loss helps the model separate the classes better. Removing CutMix has the biggest impact on AIBL, with accuracy dropping by 23.2 percentage points, indicating that AIBL has more borderline MCI cases that look similar to other classes. ADNI is less affected, so its class boundaries may be clearer. These findings collectively demonstrate that each component contributes meaningfully to the model's overall effectiveness, and their integration is essential for achieving balanced performance across datasets.

Table 4.3 Ablation study results on ADNI and AIBL datasets.

	ACC(%)	F1	SEN	SPE
	ADNI			
With morphology	35.58	0.345	0.385	0.679
No pseudo-morphology	38.04	0.359	0.359	0.679
No CutMix	45.51	**0.436**	0.434	0.717
No supervised contrastive loss	42.09	0.322	0.371	0.686
Proposed	**50.91**	0.424	**0.437**	**0.729**
	AIBL			
With morphology	29.67	0.294	0.438	0.712
No pseudo-morphology	38.50	0.338	0.392	0.699
No CutMix	39.09	0.355	0.449	0.720
No supervised contrastive loss	**73.28**	0.305	0.343	0.671
Proposed	62.27	**0.456**	**0.452**	**0.742**

From a qualitative standpoint, Fig. 4.6 further presents Grad-CAM visualizations comparing the baseline 3D U-Net model and the proposed pseudo-morphological modules for AD detection. The leftmost column shows original MRI slices from both ADNI and AIBL datasets across three classes: NC, MCI, AD. The middle column illustrates attention maps generated by the baseline U-Net3D model, which are less focused, suggesting limited class-specific feature localization. In contrast, the rightmost column shows Grad-CAM results from the model enhanced with pseudo-morphological modules. These attention maps are more focused on areas in the brain which are known to be affected by AD, such as the hippocampus and surrounding medial temporal lobe. This indicates that incorporating class-specific morphological augmentations helps the model learn more discriminative and generalizable features for AD detection.

As far as computational costs are concerned, Table 4.4 shows that the proposed method uses 21.2 M parameters, requires 2065.34 GFLOPs per forward pass, and consumes 494.56 MB of memory—1.6 M more parameters, 398.23 additional GFLOPs, and over 6 times more memory than the strongest baseline (RSC: 19.6 M parameters, 1667.11 GFLOPs, 78.40 MB memory). This relatively modest overhead is justified by the significant accuracy and macro-F1 improvements shown in Tables



(a) Input MRI (b) Without morphological modules (c) With morphological modules

Figure 4.6 Grad-CAM visualizations on MRI scans from ADNI and AIBL datasets. (a) Input images for NC, MCI, and AD groups. (b) Attention maps from the model without morphological modules. (c) Attention maps with morphological modules, showing improved focus on disease-relevant regions.

Table 4.4 Computational comparison of the proposed approach w.r.t. its counterparts.

Method	Params (M)	GFLOPs	Model size (MB)
Baseline (Zhou et al., 2019)	19.6	1667.11	78.40
Mixup (Zhang et al., 2018)	19.6	1667.11	78.40
RSC (Huang et al., 2020)	19.6	1667.11	78.40
CCSDG (Hu et al., 2023)	20.5	1689.59	81.94
EFDM (Zhang et al., 2022)	19.6	1667.11	78.40
Proposed	**21.2**	**2065.34**	**494.56**

4.1 and 4.2: on ADNI, we achieve a 4.77 percentage points accuracy boost and a 1.7 percentage points F1 gain over RSC; on AIBL, we record a 4.2 percentage points macro-F1 increase compared to the next best method. These results demonstrate a favorable trade-off between computational cost and generalization performance.

5. DISTANCE TRANSFORM GUIDED MIXUP FOR ALZHEIMER'S DETECTION

5.1 Overview

This chapter introduces an SDG approach for AD detection, leveraging a 3D U-Net feature extractor with distance transform-based mixup augmentation. The objective is to improve model robustness and generalization while mitigating class imbalance in AD detection from 3D MRI scans.

The training process begins with T1-weighted 3D MRI scans, which are preprocessed for quality enhancement and standardization. The proposed Distance Transform Guided Mixup augmentation technique is then applied, where selected regions from different MRI scans are mixed using distance transform to generate diverse training samples. The augmented images are processed through a U-Net 3D architecture (Zhou et al., 2019), which extracts hierarchical features. The U-Net 3D model architecture is described in Section 4.2. These features are fed into a classifier, predicting one of three classes: NC, AD, or MCI. The model is trained using weighted soft cross-entropy loss to improve diagnostic accuracy and generalization (Fig. 5.1).

5.2 Distance Transform-Based Mixup Augmentation

To further improve the model's generalization, for each input MRI scan x, the corresponding distance transform $D(x)$ is computed offline and stored. Each voxel's value represents its distance to the nearest anatomical boundary. The transform is

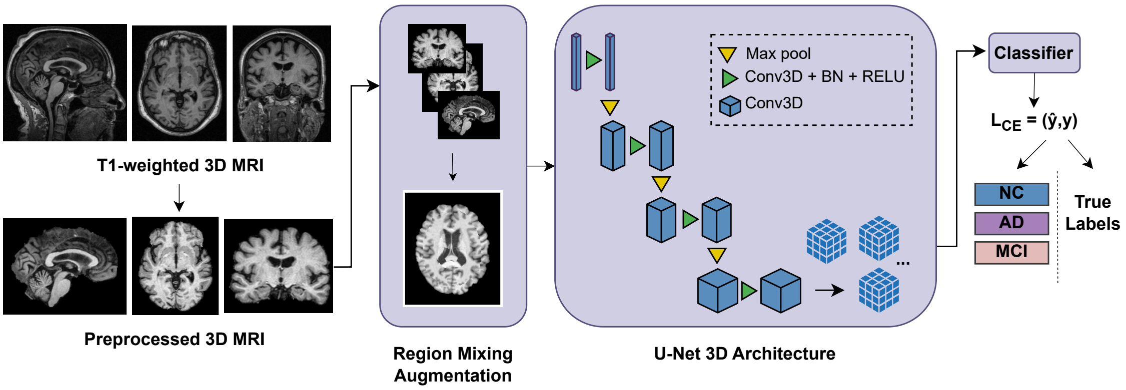


Figure 5.1 Overview of the proposed training pipeline.

mathematically defined as:

$$D(p) = \min_{q \in B} \|p - q\| \tag{5.1}$$

where $D(p)$ is the distance transform value at pixel p, B is the set of all background pixels and $\|\cdot\|$ is the Euclidean distance. After computing the distance transform,

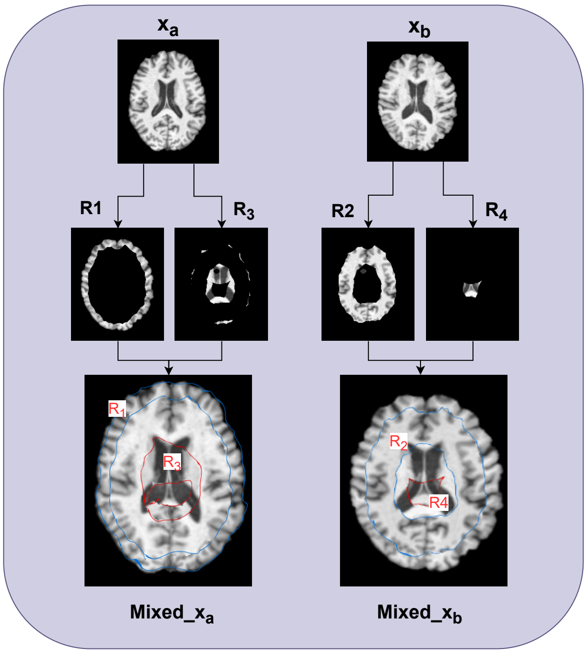


Figure 5.2 Overview of the mixing strategy. Given two input MRI scans (x_a and x_b), region-wise masks (R_1, R_2, R_3, R_4) are extracted to generate mixed samples (Mixed_x$_a$ and Mixed_x$_b$).

two thresholds t_1 and t_2 are set as the minimum and maximum values of the distance transform for each input, ensuring that at least 10% of the brain structure is preserved in each region. This is done to avoid creating excessively small regions.

Given a pair of 3D MRIs (x_a, x_b), the regions are defined as:

$$R_1 = \mathbb{1}(D_a \leq \mathrm{t}_1) \tag{5.2}$$

$$R_2 = \mathbb{1}(\mathrm{t}_1 < D_b \leq \mathrm{t}_2) \cdot (1 - R_1) \tag{5.3}$$

$$R_3 = \mathbb{1}(\mathrm{t}_1 < D_a \leq \mathrm{t}_2) \cdot (1 - R_1) \cdot (1 - R_2) \tag{5.4}$$

$$R_4 = \mathbb{1}(D_b > \mathrm{t}_2) \cdot (1 - R_1) \cdot (1 - R_2) \cdot (1 - R_3) \tag{5.5}$$

where $\mathbb{1}(\cdot)$ represents an indicator function that returns 1 if the condition inside is true and 0 otherwise. The variables R_1, R_2, R_3, and R_4 are mutually exclusive binary masks, each defining different spatial regions within the image. The terms D_a and D_b correspond to the distance transforms of the images x_a and x_b, respectively. These thresholds divide the MRI into four non-overlapping regions (Fig. 5.2). The mixed image is constructed as:

$$\tilde{x} = (R_1 \cdot x_a) + (R_2 \cdot x_b) + (R_3 \cdot x_a) + (R_4 \cdot x_b) \tag{5.6}$$

where R_1, R_2, R_3, R_4 represent binary masks corresponding to the different thresholded regions.

5.3 Label Mixing

To balance label contributions from the spatially mixed images, the probabilities are calculated based on the pixel count from each region. Given the masks defining the segmented regions, regions R_1 and R_3 correspond to the first image x_a while regions R_2 and R_4 correspond to the second image x_b. The number of pixels assigned from each image is computed as:

$$P_a = \sum R_1 + \sum R_3 \tag{5.7}$$

$$P_b = \sum R_2 + \sum R_4 \tag{5.8}$$

where P_a and P_b represent the number of pixels originating from x_a and x_b, respectively.

The relative contribution of each image to the final mixed sample is determined as:

$$\alpha_a = \frac{P_a}{P_a + P_b}, \quad \alpha_b = \frac{P_b}{P_a + P_b} \tag{5.9}$$

where α_a and α_b represent the proportion of pixels coming from each source image. Using these proportions, the final mixed label $\tilde{y}$ is computed as a weighted sum of the labels from both images:

$$\tilde{y} = \alpha_a \cdot y_a + \alpha_b \cdot y_b \tag{5.10}$$

where y_a and y_b are the original class labels of the input images x_a and x_b, respectively.

A soft cross-entropy loss function is used to handle the soft labels. Additionally, to address the class imbalance in AD detection, class weights w_i are introduced, ensuring that underrepresented classes contribute more significantly to the loss computation. The weighted soft cross-entropy loss is defined as:

$$L = -\frac{1}{\sum w_i} \sum_{i=1}^{C} w_i \cdot y_i \cdot \log(\hat{y}_i) \tag{5.11}$$

where C represents the number of classes, y_i represents the true class label, $\hat{y}_i$ denotes the predicted probability for class i, w_i is the class weight, computed based on the inverse frequency of each class to mitigate the effects of data imbalance.

5.4 Results and Discussion

This section presents the results of the proposed Distance Transform method and baseline approaches on the ADNI and AIBL datasets. Generalization Results for ADNI and AIBL are summarized in Tables 5.1 and 5.2.

As shown in Table 5.1, on the ADNI dataset, the proposed Distance Transform method outperformed all baselines, achieving the highest accuracy (48.37%) and F1 score (0.460), along with superior sensitivity (0.461) and, specificity (0.733). This demonstrates its effectiveness in distinguishing between classes, outperforming

Table 5.1 Results with the ADNI dataset

Method	**ACC (%)**	**SEN**	**SPE**	**F1**
Baseline (Zhou et al., 2019)	38.04	0.359	0.679	0.359
Mixup (Zhang et al., 2018)	48.29	0.392	0.703	0.339
RSC (Huang et al., 2020)	46.14	0.410	0.713	0.407
CCSDG (Hu et al., 2023)	39.55	0.419	0.700	0.396
Proposed	**48.37**	**0.461**	**0.733**	**0.460**

techniques such as Mixup (Zhang et al., 2018), RSC (Huang et al., 2020), and CCSDG (Hu et al., 2023).

Table 5.2 Results with the AIBL dataset

Method	**ACC (%)**	**SEN**	**SPE**	**F1**
Baseline (Zhou et al., 2019)	38.50	0.392	0.699	0.338
Mixup (Zhang et al., 2018)	**65.42**	0.382	0.721	0.382
RSC (Huang et al., 2020)	51.27	0.449	**0.737**	0.414
CCSDG (Hu et al., 2023)	40.82	0.401	0.699	0.396
Proposed	52.25	**0.454**	0.726	**0.430**

As shown in Table 5.2, on the AIBL dataset, the distance transform-based approach yielded promising results, achieving an accuracy of 52.25%, an F1 score of 0.430, and a sensitivity of 0.454. This demonstrates its robustness in handling domain shifts across datasets. Comparatively, the baseline method achieved only 38.50% accuracy, while Mixup and RSC provided modest improvements.

The proposed distance transform method showed superior performance across the ADNI and AIBL datasets compared to existing baseline methods. On the ADNI dataset, it outperformed all baselines in terms of accuracy, F1 score, and sensitivity, demonstrating its ability to classify AD, MCI, and NC groups effectively. Similarly, the method achieved competitive results on the AIBL dataset. This reflects the importance of leveraging domain adaptation techniques like Distance Transform to enhance model performance in AD detection.

6. TOWARDS SINGLE-DOMAIN GENERALIZATION IN ALZHEIMER'S DETECTION VIA EXTENDED MIXSTYLE

6.1 Overview

While domain generalization methods like MixStyle improve robustness by simulating inter-domain variability, they are primarily designed for scenarios involving multiple training domains. Therefore, standard MixStyle may be less effective when only a single domain is available (Zhang et al., 2022; Zhou et al., 2021). To address these challenges, this thesis proposes a mixstyle variant, named Extended MixStyle. This extension enhances the original method by incorporating higher-order moment statistics into feature perturbations. It allows for more subtle variations within the same domain, improving model generalization to unseen data. During training, each MRI volume is passed through a 3D U-Net backbone integrated with Extended MixStyle, enabling the model to learn robust representations under diverse internal variations. The network is optimized using a weighted cross-entropy loss, where class imbalance is addressed by assigning inverse class frequency weights. Integrated into this framework, Extended MixStyle supports improved AD detection performance without relying on multi-domain training data.

6.2 Model Architecture

The classification framework is based on a 3D U-Net architecture described in section 4.2 which serves as the backbone for feature extraction. In more detail, the architecture comprises four stacked convolutional blocks, with MixStyle regularization applied to the second intermediate layer. This specific placement introduces style variation at a mid-level semantic representation, which is empirically found to have an effective balance between low-level noise and high-level abstraction.

6.3 Extended MixStyle with Higher-Order Moments

Given a 3D MRI volume $x \in \mathbb{R}^{B\times C\times D\times H\times W}$, where B is the batch size, C is the number of channels, and (D, H, W) represent the spatial dimensions, the objective is to improve SDG by perturbing feature distributions during training.

MixStyle, originally proposed in (Zhou et al., 2021), performs feature-level domain mixing by interpolating the statistical moments of feature maps, specifically the mean and standard deviation, across spatial dimensions. Given a feature map $x \in \mathbb{R}^{B\times C\times H\times W}$, MixStyle first computes the per-channel spatial mean and standard deviation as defined in (6.1) and (6.2):

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{6.1}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 + \epsilon} \tag{6.2}$$

where $x_1, x_2, \ldots, x_N$ represent the spatial values within a given feature map, and $N = D\times H\times W$. The mean and standard deviation are computed independently for each feature map, and ϵ is a small constant added for numerical stability.

During training, MixStyle is applied with a probability p. For each sample, a mixing coefficient $\alpha \sim \mathrm{Beta}(\alpha, \alpha)$ is drawn, and a feature-level permutation is applied within the batch to obtain shuffled statistics (μ', σ'). The mixed statistics are then computed as:

$$\mu_{\text{mix}} = \alpha\mu + (1-\alpha)\mu' \tag{6.3}$$

$$\sigma_{\text{mix}} = \alpha\sigma + (1-\alpha)\sigma' \tag{6.4}$$

The normalized feature is then reparameterized using the mixed statistics:

$$x' = \left(\frac{x-\mu}{\sigma}\right) \cdot \sigma_{\text{mix}} + \mu_{\text{mix}} \tag{6.5}$$

This process as defined in (Zhou et al., 2021) perturbs the feature distributions dynamically during training, simulating domain shifts and encouraging the model to learn domain-invariant representations.

While perturbing first- and second-order statistics (mean and variance) can improve generalization, this may be insufficient in context of 3D MRI-based AD detection, especially in SDG settings. MRIs contain complex anatomical structures and non-Gaussian intensity patterns that cannot be fully captured by just mean and variance. Subtle changes related to AD, such as hippocampal atrophy or enlarged ventricles, may alter the shape of feature distributions in ways that affect higher-order moments like skewness and kurtosis. Moreover, in single-domain scenarios where no external datasets are available, the model must rely solely on intra-domain variability, which includes scanner noise, demographic differences, and disease heterogeneity. These subtle variations often go beyond simple shifts in average intensity or spread, making it essential to account for the asymmetry and tail behavior of feature distributions. Therefore, incorporating higher-order statistics enables better feature perturbations, helping the model learn domain-invariant representations.

To address this limitation, an extension to the original MixStyle is proposed that incorporates higher-order moments: **skewness** (third-order) and **kurtosis** (fourth-order). These moments capture the asymmetry and tail behavior of feature distributions, which are particularly relevant in medical imaging where subtle structural variations can affect diagnostic accuracy.

Specifically, given the same input feature map $x \in \mathbb{R}^{B\times C\times D\times H\times W}$, we compute the skewness S and excess kurtosis K for each feature channel as:

$$S = \frac{\mathbb{E}[(x-\mu)^3]}{\sigma^3+\epsilon} \tag{6.6}$$

$$K = \frac{\mathbb{E}[(x-\mu)^4]}{\sigma^4+\epsilon} - 3 \tag{6.7}$$

Here, $\mathbb{E}[\cdot]$ denotes the spatial average over the $D\times H\times W$ dimensions, μ and σ are the spatial mean and standard deviation as previously defined, and ϵ is a small constant for numerical stability.

During training, as in MixStyle, a sample-wise mixing coefficient $\alpha \sim \text{Beta}(\alpha,\alpha)$ is drawn, and feature statistics are shuffled across samples within the batch. Let S' and K' denote the skewness and kurtosis of a shuffled sample in the batch. Then the mixed higher-order moments are computed by interpolating between the original and shuffled statistics:

$$S_{\text{mix}} = \alpha S + (1-\alpha) S' \tag{6.8}$$

$$K_{\text{mix}} = \alpha K + (1-\alpha) K' \tag{6.9}$$

Finally, the perturbed feature map x' is reconstructed by extending the standard MixStyle reparameterization:

$$x' = \frac{x-\mu}{\sigma} \cdot \sigma_{\text{mix}} + \mu_{\text{mix}} + S_{\text{mix}} \cdot (x-\mu_{\text{mix}})^2 + 0.1 \cdot K_{\text{mix}} \cdot (x-\mu_{\text{mix}})^3 \tag{6.10}$$

where the skewness term introduces asymmetry correction, and the kurtosis term adjusts the tail behavior of the feature distribution.

The enhanced MixStyle is applied to intermediate layers of the 3D U-Net, specifically the second layer, with mixing probability of 0.7 and interpolation parameter α of 0.7. To promote domain generalization, MixStyle is applied only during training. By introducing richer statistical perturbations, Extended MixStyle enables the model to focus on disease-relevant features while ignoring dataset-specific biases. This leads to improved generalization and detection performance on unseen MRI data.

6.4 Results and Discussion

The proposed approach consistently outperforms baseline approaches. Generalization results for ADNI and AIBL are summarized in Tables 6.1 and 6.2.

Table 6.1 shows that the proposed method achieved the highest performance across all evaluation metrics on the ADNI dataset. Specifically, it reached an accuracy of 50.25%, surpassing the next best method, MixStyle (Zhou et al., 2021), by 4.95 percentage points. In terms of sensitivity and specificity, the proposed approach improved upon MixStyle by 2.1 and 2.0 percentage points, respectively. Additionally, the F1 score increased by 3.3 percentage points (0.486 vs. 0.453), indicating better overall class balance. Compared to MixUp (Zhang et al., 2018), a widely used

Table 6.1 Results with the ADNI dataset

Method	**ACC (%)**	**SEN**	**SPE**	**F1**
Baseline (Zhou et al., 2019)	38.04	0.359	0.679	0.359
Mixup (Zhang et al., 2018)	48.29	0.359	0.703	0.339
RSC (Huang et al., 2020)	46.14	0.410	0.713	0.407
CCSDG (Hu et al., 2023)	39.55	0.419	0.700	0.396
MixStyle (Zhou et al., 2021)	45.30	0.461	0.729	0.453
EFDM (Zhang et al., 2022)	45.35	0.353	0.679	0.249
Proposed	**50.25**	**0.482**	**0.749**	**0.486**

baseline, the accuracy gain was nearly 2 percentage points. Other methods, such as RSC (Huang et al., 2020), EFDM (Zhang et al., 2022), and CCSDG (Hu et al., 2023), also fell short across all metrics, highlighting the advantage of our proposed framework in generalizing to unseen domain variations in AD detection.

Table 6.2 Results with the AIBL dataset

Method	**ACC (%)**	**SEN**	**SPE**	**F1**
Baseline (Zhou et al., 2019)	38.50	0.392	0.699	0.338
Mixup (Zhang et al., 2018)	65.42	0.382	0.721	0.382
RSC (Huang et al., 2020)	51.27	0.449	0.737	0.414
CCSDG (Hu et al., 2023)	40.82	0.401	0.699	0.396
MixStyle (Zhou et al., 2021)	69.94	0.480	0.742	0.412
EFDM (Zhang et al., 2022)	61.49	0.329	0.678	0.301
Proposed	**73.28**	**0.483**	**0.759**	**0.455**

As shown in Table 6.2, proposed method also achieved the best generalization performance on the AIBL dataset. It led with an accuracy of 73.28%, which is 3.34 percentage points higher than the closest alternative, Mixstyle (Zhou et al., 2021) (69.94%). Improvements were also consistent across other metrics: a 0.3 percentage points gain in sensitivity (0.483 vs. 0.480) and 1.7 points in specificity (0.759 vs. 0.742). When compared to earlier approaches like RSC (Huang et al., 2020) and EFDM (Zhang et al., 2022), our model achieved notably stronger results, outperforming EFDM by 11.79 percentage points margin in accuracy and over 15 points in specificity, highlighting the limitations of those methods under AIBL domain conditions.

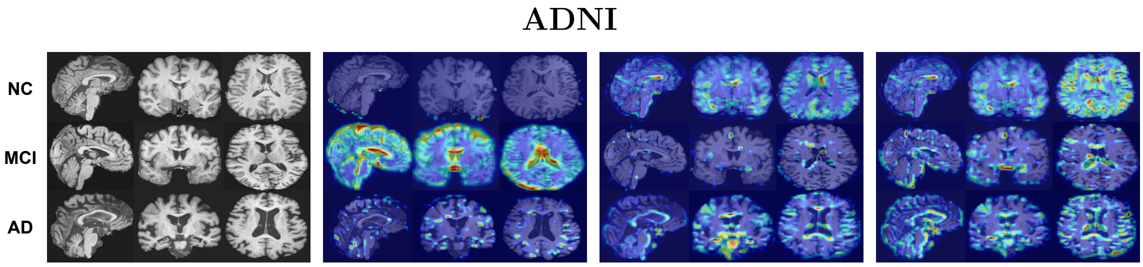
Across both ADNI and AIBL datasets, the proposed method consistently demonstrated superior generalization performance in the presence of domain shift. Although accuracy was substantially higher on AIBL (73.28%) than on ADNI (50.25%), both reflect meaningful improvements over state-of-the-art alternatives. This performance gap is likely due to inherent differences in domain characteristics. AIBL's imaging protocol and cohort demographics may align more closely with NACC, resulting in reduced domain discrepancy. Nonetheless, our approach outperformed all baselines on both datasets, indicating its robustness to variations in scanner hardware, imaging settings, and subject populations. These results support the effectiveness of our design choices in handling domain generalization and highlight the model's capacity to generalize well to unseen clinical environments.

6.5 Ablation Study

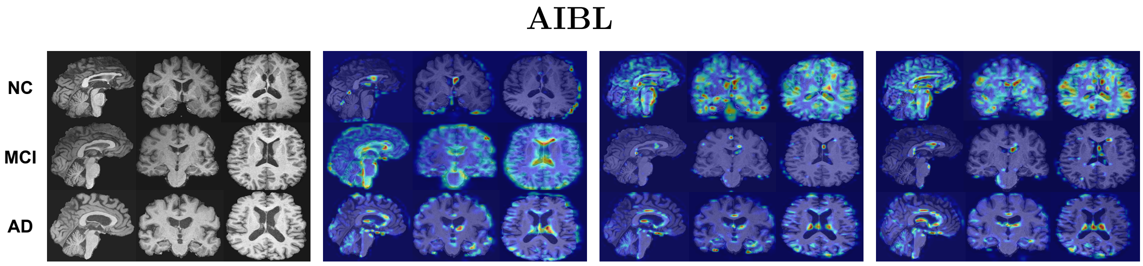
To investigate how different levels of style mixing influence model attention, Grad-CAM visualizations for both the ADNI and AIBL datasets are examined in (Fig. 6.1). Each dataset presents MRI scans across three diagnostic categories: NC, MCI, and AD. In the ADNI row, the model without MixStyle (column b) exhibits broad and somewhat noisy attention, with limited focus on disease-relevant brain regions. Introducing MixStyle (mean + variance, column c) results in a clearer and more localized attention pattern, particularly in the medial temporal regions, which are commonly affected in AD. The extended MixStyle (mean, variance, skewness, and kurtosis, column d) further sharpens this focus, showing tighter and more anatomically aligned activation. This progression suggests that the inclusion of higher-order moments helps the model attend more precisely to relevant features in the ADNI cohort.

In contrast, for the AIBL dataset (bottom row) in (Fig. 6.1), the distinction between variants is more subtle. The without MixStyle model already identifies some informative regions, though the attention remains relatively scattered. The standard MixStyle improves spatial consistency, yet in some instances (e.g., the MCI row), the attention still focuses on some non-informative areas. With extended MixStyle, the activations appear slightly more confined and centered compared to the previous variants, but the visual improvement is less pronounced than in ADNI. This suggests that while extended MixStyle contributes to refined attention in both datasets, the degree of improvement may vary with dataset characteristics, such as sample distribution or imaging protocol differences. Overall, the qualitative results support the notion that extended MixStyle helps regularize attention maps and potentially improves domain robustness, though its impact is more visually evident in ADNI than in AIBL.

The computational comparison in Table 6.3 shows that the proposed Extended MixStyle method introduces no additional overhead compared to the baseline. It



(a) Original MRI (b) Without MixStyle (c) MixStyle (d) Extended MixStyle



(e) Original MRI (f) Without MixStyle (g) MixStyle (h) Extended MixStyle

Figure 6.1 Grad-CAM visualizations on 3D MRI scans from the ADNI (top row) and AIBL (bottom row) datasets. Each row shows: (a) original input, (b) attention maps without MixStyle, (c) with standard MixStyle using mean and variance, and (d) with extended MixStyle incorporating all four moments (mean, variance, skewness, kurtosis).

retains the same number of parameters (19.6M), GFLOPs (1667.12), and model size (78.40 MB) as methods like Mixup, RSC, EFDM and MixStyle. On the contrary, CCSDG increases complexity slightly with 20.5M parameters and 1689.59 GFLOPs. These results confirm that our approach achieves improved generalization performance without increasing computational burden, making it highly suitable for practical and resource-constrained deployment.

To assess the impact of applying the Extended MixStyle module at different network layers, an ablation study was conducted across individual and combined insertion points. In Table 6.4, among the single-layer configurations, inserting Extended MixStyle at Layer 2 consistently yields the best generalization performance on the ADNI dataset, achieving 55.00% accuracy, outperforming all other variants. This is an improvement of 4.75 percentage points over the baseline MixStyle (45.30%) and 6.63 points over RSC (46.14%). The F1 score (0.486) and sensitivity (0.482) also show notable gains compared to earlier methods. On the AIBL dataset, however, Layer 3 performs slightly better in some metrics. It achieves an F1 score of 0.482 and the highest sensitivity (0.495), compared to 0.455 F1 and 0.483 sensitivity in Layer 2. However, the accuracy in Layer 3 (57.95%) is 15.13 points lower than the peak AIBL performance (73.08%) observed in multi-layer configurations.

Table 6.3 Computational comparison of the proposed approach w.r.t. its counterparts.

Method	**Params (M)**	**GFLOPs**	**Model size (MB)**
Baseline (Zhou et al., 2019)	19.6	1667.11	78.40
Mixup (Zhang et al., 2018)	19.6	1667.11	78.40
RSC (Huang et al., 2020)	19.6	1667.11	78.40
CCSDG (Hu et al., 2023)	20.5	1689.59	81.94
EFDM (Zhang et al., 2022)	19.6	1667.11	78.40
MixStyle (Zhou et al., 2021)	19.6	1667.12	78.40
Proposed	**19.6**	**1667.12**	**78.40**

Table 6.4 Performance comparison of Extended MixStyle applied at different U-Net layers on the ADNI and AIBL datasets.

	ACC(%)	F1	SEN	SPE
	ADNI			
Layer	46.30	0.419	0.417	0.710
Layer 2	**55.00**	**0.486**	**0.482**	**0.749**
Layer 3	48.45	0.425	0.427	0.715
Layer (1 + 2)	48.13	0.455	0.451	0.727
Layer (2 + 3)	47.25	0.314	0.377	0.696
Layer (1 + 3)	48.53	0.433	0.434	0.724
	AIBL			
Layer 1	64.04	0.481	0.471	0.758
Layer 2	55.00	0.455	0.483	0.759
Layer 3	57.95	**0.482**	**0.495**	**0.766**
Layer (1 + 2)	51.66	0.400	0.429	0.743
Layer (2 + 3)	**73.08**	0.347	0.367	0.702
Layer (1 + 3)	20.82	0.200	0.415	0.686

While Layer 3 slightly outperforms Layer 2 on AIBL in terms of F1 and sensitivity, it results in a noticeable drop on ADNI — 6.55 percentage points lower in accuracy (from 55.00% to 48.45%) and 0.061 lower in F1 (from 0.486 to 0.425). This trade-off indicates that although Layer 3 is more effective on AIBL, it generalizes less robustly on ADNI. Considering that the ADNI dataset often serves as a foundational source, this degradation is significant.

In more detail, combinations of Extended MixStyle across two layers were also analyzed. Interestingly, combining layers does not always lead to better performance. For instance, Layer 1+2 achieves 48.13% accuracy on ADNI, which is 6.87 points lower than using Layer 2 alone. Similarly, the (1+3) and (2+3) layer combinations yield either comparable or degraded results on AIBL, despite achieving a higher raw accuracy of 73.08%. These combinations underperform in terms of F1 score and sensitivity, suggesting that increased complexity may lead to less consistent feature representations.

In terms of balancing ADNI and AIBL Performance, although Layer 3 performs marginally better on AIBL, adopting it as the primary layer would lead to a substantial performance drop on ADNI, with a 6.55 point decrease in accuracy and lower F1 and sensitivity. In contrast, Layer 2 maintains strong generalization across both datasets and demonstrates the most balanced trade-off. This consistency is valuable for real-world deployment scenarios, where models are expected to perform reasonably well across varying domains.

Table 6.5 ADNI — MixStyle with skewness

α	p	ACC(%)	F1	SEN	SPE
0.1	0.5	45.35	0.339	0.373	0.690
0.1	0.7	48.13	0.433	0.431	0.718
0.1	0.9	44.71	0.421	0.423	0.714
0.3	0.5	47.41	0.420	0.422	0.717
0.3	0.7	45.19	0.413	0.412	0.705
0.3	0.9	46.22	**0.455**	0.463	**0.727**
0.5	0.5	46.94	0.433	0.431	0.722
0.5	0.7	37.72	0.369	0.385	0.686
0.5	0.9	40.42	0.399	**0.464**	0.719
0.7	0.5	**48.21**	0.430	0.430	0.717
0.7	0.7	33.99	0.337	0.381	0.684
0.7	0.9	43.52	0.427	0.434	0.713

Table 6.6 ADNI — Extended MixStyle

α	p	ACC(%)	F1	SEN	SPE
0.1	0.5	47.89	0.373	0.400	0.717
0.1	0.7	48.13	0.451	0.452	0.728
0.1	0.9	35.55	0.367	0.430	0.702
0.3	0.5	45.59	0.456	0.471	0.727
0.3	0.7	50.91	0.449	0.450	0.732
0.3	0.9	49.24	0.426	0.432	0.724
0.5	0.5	51.15	0.406	0.429	0.726
0.5	0.7	53.13	0.424	0.423	0.717
0.5	0.9	45.11	0.436	0.436	0.718
0.7	0.5	32.32	0.304	0.414	0.689
0.7	0.7	**55.00**	**0.486**	**0.482**	**0.749**
0.7	0.9	41.54	0.409	0.422	0.707

Tables 6.5, 6.6, 6.7 and 6.8 explore the relationship between mixing strength (α) and application probability (p) for both MixStyle with skewness and the proposed Extended MixStyle approach, evaluated on ADNI and AIBL datasets. The trends reveal that the optimal configuration differs across datasets, reflecting distinct generalization dynamics.

On the ADNI dataset, Extended MixStyle with $\alpha = 0.7$ and $p = 0.7$ achieves the best overall performance, reaching 55% accuracy, 0.486 F1, and high sensitivity

Table 6.7 AIBL — MixStyle with skewness

α	**p**	**ACC(%)**	**F1**	**SEN**	**SPE**
0.1	0.5	**66.79**	0.366	0.382	0.718
0.1	0.7	53.24	0.414	0.427	0.742
0.1	0.9	37.72	0.340	0.439	0.712
0.3	0.5	46.95	0.383	0.410	0.722
0.3	0.7	59.33	0.410	0.411	0.741
0.3	0.9	43.22	0.398	**0.514**	0.752
0.5	0.5	48.72	0.420	0.396	0.716
0.5	0.7	29.86	0.314	0.462	0.711
0.5	0.9	32.61	0.300	0.469	0.723
0.7	0.5	63.26	**0.439**	0.456	**0.775**
0.7	0.7	28.29	0.287	0.469	0.717
0.7	0.9	33.33	0.306	0.406	0.706

Table 6.8 AIBL — Extended MixStyle

α	**p**	**ACC(%)**	**F1**	**SEN**	**SPE**
0.1	0.5	61.49	0.377	0.405	0.739
0.1	0.7	57.95	**0.482**	**0.495**	0.766
0.1	0.9	31.82	0.300	0.482	0.721
0.3	0.5	35.55	0.351	0.488	0.723
0.3	0.7	50.88	0.376	0.398	0.740
0.3	0.9	54.42	0.399	0.428	0.753
0.5	0.5	**72.69**	0.428	0.443	**0.776**
0.5	0.7	40.27	0.334	0.370	0.698
0.5	0.9	41.65	0.373	0.435	0.732
0.7	0.5	20.82	0.200	0.415	0.686
0.7	0.7	55.00	0.455	0.483	0.759
0.7	0.9	26.91	0.266	0.415	0.693

and specificity (0.482 and 0.749, respectively). This configuration outperforms all MixStyle variants, whose best accuracy peaks at 48.13% under $\alpha = 0.1, p = 0.7$. Notably, for both methods, moderate values of α and p tend to be more effective than extremes. As α or p increase beyond optimal points, performance drops substantially, likely due to excessive feature distortion. This pattern is especially visible when $\alpha = 0.7, p = 0.9$, where ADNI accuracy falls below 42% for both methods. Therefore, Extended MixStyle demonstrates stronger stability across configurations, maintaining high performance across a broader range of values. This suggests that higher-order moment mixing improves generalization without compromising feature integrity when the perturbation is appropriately used.

Unlike ADNI, AIBL exhibits a different optimal pattern. The best accuracy, 72.69%, is achieved by Extended MixStyle at $\alpha = 0.5, p = 0.5$, which also yields strong specificity (0.776) and F1 score (0.428). While $\alpha = 0.7, p = 0.7$, which was chosen as the best setting for ADNI still performs reasonably on AIBL (55.00% accuracy, 0.455

F1), although it does not reach the same effectiveness. Interestingly, MixStyle performs relatively better on AIBL than on ADNI. For instance, with $\alpha = 0.1, p = 0.5$, it achieves 66.79% accuracy, outperforming several Extended MixStyle configurations. However, this comes with much lower sensitivity (0.382) and F1 (0.366), highlighting a possible imbalance between true positive and negative predictions.

These observations indicate that AIBL responds better to slightly lower mixing intensities (α) and more conservative application probabilities. The results emphasize the importance of dataset-specific calibration of mixing hyperparameters, especially when the source and target domain characteristics differ significantly.

7. CONCLUSION AND FUTURE WORK

7.1 Conclusion

This thesis explored methods to enhance the generalization capacity of models for AD detection from 3D MRI data under the SDG setting. Recognizing the performance limitations of conventional models when faced with domain shifts, the study proposed three approaches that aim to improve robustness across unseen imaging domains while maintaining sensitivity to disease-specific patterns. The first proposed method employed class-specific, learnable pseudo-morphological augmentations with supervised contrastive learning to simulate disease specific anatomical variations linked to neurodegeneration. The second approach introduced distance transform-guided Mixup to generate structurally coherent recompositions of MRI scans, addressing limitations of conventional Mixup. The third method enhanced MixStyle by incorporating higher-order statistical moments, enabling richer feature perturbations to better simulate domain shifts.

All methods were evaluated using the NACC dataset for training and the ADNI and AIBL datasets for out-of-distribution validation. Across these benchmarks, the proposed models consistently outperformed established baselines, with noticeable improvements in accuracy, sensitivity, and F1 scores. These gains show the importance of task-aware augmentation strategies and distribution-aware feature manipulation in mitigating the effects of domain shift. Moreover, qualitative analysis revealed that the proposed models were better aligned with disease-relevant regions, providing further evidence of their enhanced interpretability.

7.2 Future Work

Despite the performance gains achieved in this study, certain limitations highlight directions for future work. The primary concern lies in the higher model complexity and memory consumption introduced by the learnable augmentation modules. Future research could explore more efficient implementations through model compression, knowledge distillation, or lightweight architectures that maintain performance with reduced overhead. Furthermore, the proposed methods were validated primarily on benchmark datasets such as ADNI and AIBL, which, though well-curated, may not reflect the full heterogeneity of real-world clinical settings. Expanding validation to larger, multi-institutional datasets with varied imaging protocols and patient demographics will be essential for assessing generalizability in practical deployment scenarios.

Another promising direction involves developing adaptive augmentation strategies that respond to data-specific characteristics such as localized atrophy patterns. Incorporating patient demographic information such as age and gender into pretext tasks may further enhance the model's ability to improve generalization. Moreover, integration of multimodal data, including PET imaging, genetic profiles, and cognitive assessments, could offer a more comprehensive diagnostic framework, particularly for borderline cases like early-stage MCI. Finally, future research may also benefit from embedding causal reasoning frameworks to move beyond purely correlative learning and foster more robust, interpretable decision-making.

BIBLIOGRAPHY

Alp, S., Akan, T., Bhuiyan, M. S., Disbrow, E. A., Conrad, S. A., Vanchiere, J. A., Kevil, C. G., & Bhuiyan, M. A. (2024). Joint transformer architecture in brain 3d mri classification: its application in Alzheimer's disease classification. *Scientific Reports*, *14*(1), 8996.

Aptoula, E. & Lefèvre, S. (2009). Multivariate mathematical morphology applied to colour image analysis. In T. Géraud & L. Najman (Eds.), *Multivariate Image Processing: Methods and Applications* (pp. 303–337). ISTE Ltd and John Wiley & Sons Inc.

Aviles-Rivero, A. I., Runkel, C., Papadakis, N., Kourtzi, Z., & Schönlieb, C.-B. (2022). Multi-modal hypergraph diffusion network with dual prior for Alzheimer classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 717–727). Springer.

Azcona, E. A., Besson, P., Wu, Y., Punjabi, A., Martersteck, A., Dravid, A., Parrish, T. B., Bandt, S. K., & Katsaggelos, A. K. (2020). Interpretation of brain morphology in association to Alzheimer's disease dementia classification using graph convolutional networks on triangulated meshes. In *International Workshop on Shape in Medical Imaging*, (pp. 95–107). Springer.

Beekly, D. L., Ramos, E. M., van Belle, G., Deitrich, W., Clark, A. D., Jacka, M. E., & Kukull, W. A. (2004). The national Alzheimer's coordinating center (nacc) database: an Alzheimer disease database. *Alzheimer Disease & Associated Disorders*, *18*(4), 270–277.

Braak, H. & Braak, E. (1991). Neuropathological stageing of alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259.

Bravo-Ortiz, M. A., Holguin-Garcia, S. A., Quiñones-Arredondo, S., Mora-Rubio, A., Guevara-Navarro, E., Arteaga-Arteaga, H. B., Ruz, G. A., & Tabares-Soto, R. (2024). A systematic review of vision transformers and convolutional neural networks for Alzheimer's disease classification using 3d mri images. *Neural Computing and Applications*, *36*(35), 21985–22012.

Cai, H., Zhang, Q., & Long, Y. (2023). Prototype-guided multi-scale domain adaptation for Alzheimer's disease detection. *Computers in Biology and Medicine*, *154*, 106570.

Canales-Fiscal, M. R. & Tamez-Peña, J. G. (2023). Hybrid morphological-convolutional neural networks for computer-aided diagnosis. *Frontiers in Artificial Intelligence*, *6*, 1253183.

Chen, Y., Pan, Y., Xia, Y., & Yuan, Y. (2023). Disentangle first, then distill: A unified framework for missing modality imputation and Alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging*, *42*(12), 3566–3578.

Cheng, S., Gokhale, T., & Yang, Y. (2023). Adversarial bayesian augmentation for single-source domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, (pp. 11400–11410)., Paris, France. IEEE.

Davidson, J. L. & Hummer, F. (1993). Morphology neural networks: An introduction with applications. *Circuits, Systems and Signal Processing*, *12*(2), 177–210.

Ding, H., Wang, Z., Tang, Y., Wang, T., Qi, M., Dou, W., Qian, L., Gao, Y., Zhong, Q., Yang, X., et al. (2023). Topological properties of individual gray matter morphological networks in identifying the preclinical stages of Alzheimer's disease: a preliminary study. *Quantitative Imaging in Medicine and Surgery*, *13*(8), 5258.

Duong, Q. A., Tran, S. D., & Gahm, J. K. (2025). Multimodal surface-based transformer model for early diagnosis of Alzheimer's disease. *Scientific Reports*, *15*(1), 5787.

El-Assy, A., Amer, H. M., Ibrahim, H., & Mohamed, M. (2024). A novel cnn architecture for accurate early detection and classification of Alzheimer's disease using mri data. *Scientific Reports*, *14*(1), 3463.

Ellis, K. A., Bush, A. I., Darby, D., Fazio, D. D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, R., Masters, P., & the AIBL Research Group (2009). The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*, *21*(4), 672–687.

Fiasam, L. D., Rao, Y., Sey, C., Aggrey, E. S. E., Kodjiku, S. L., Browne, J. A., Danso, J. M., Ukwuoma, C. C., & Gyarteng, E. S. (2023). Domain contrastive learning for multi-site Alzheimer's disease classification. In *2023 20th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, (pp. 1–6). IEEE.

Folego, G., Weiler, M., Casseb, R. F., Pires, R., & Rocha, A. (2020). Alzheimer's disease detection through whole-brain 3d-cnn mri. *Frontiers in bioengineering and biotechnology*, *8*, 534592.

Fotuhi, M., Hachinski, V., & Whitehouse, P. J. (2009). Milder stages of Alzheimer's disease: an overview. *Nature Reviews Neurology*, *5*(11), 635–643.

Frisoni, G. B., Fox, N. C., Jack Jr, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural mri in Alzheimer disease. *Nature Reviews Neurology*, *6*(2), 67–77.

Gamgam, G., Kabakcioglu, A., Yüksel Dal, D., & Acar, B. (2024). Disentangled attention graph neural network for Alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 219–228). Springer.

Gan, Y., Lan, Q., Huang, C., Su, W., & Huang, Z. (2025). Dense convolution-based attention network for Alzheimer's disease classification. *Scientific Reports*, *15*(1), 5693.

Ghosh, S. & Das, S. (2024). Multi-scale morphology-aided deep medical image segmentation. *Engineering Applications of Artificial Intelligence*, *137*, 109047.

Guan, H., Liu, Y., Yang, E., Yap, P.-T., Shen, D., & Liu, M. (2021). Multi-site MRI harmonization via attention-guided deep domain adaptation for brain disorder identification. *Medical image analysis*, *71*, 102076.

Guan, Y., Hu, W., & Hu, X. (2021). Abnormal behavior recognition using 3d-cnn combined with lstm. *Multimedia Tools and Applications*, *80*(12), 18787–18801.

Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, *2*, 147–171.

Guzzi, L., Zuluaga, M. A., Lareyre, F., Di Lorenzo, G., Goffart, S., Chierici, A., Raffort, J., & Delingette, H. (2024). Differentiable soft morphological filters for medical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 177–187)., Marrakesh, Marocco.

Han, K., Li, G., Fang, Z., & Yang, F. (2023). Multi-template meta-information regularized network for Alzheimer's disease diagnosis using structural MRI. *IEEE Transactions on Medical Imaging*, *43*(5), 1664–1676.

Hu, S., Liao, Z., & Xia, Y. (2023). Devil is in channels: Contrastive single domain generalization for medical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 14–23)., Vancouver, Canada.

Hu, Y., Belkhir, N., Angulo, J., Yao, A., & Franchi, G. (2022). Learning deep morphological networks with neural architecture search. *Pattern Recognition*, *131*, 108893.

Hu, Y., Wang, J., Zhu, H., Li, J., & Shi, J. (2024). Cost-sensitive weighted contrastive learning based on graph convolutional networks for imbalanced Alzheimer's disease staging. *IEEE Transactions on Medical Imaging*, *43*(9), 3126–3136.

Huang, Z., Wang, H., Xing, E. P., & Huang, D. (2020). Self-challenging improves cross-domain generalization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, Virtual Conference. Springer.

Jabason, E., Ahmad, M. O., & Swamy, M. N. S. (2025). A lightweight deep convolutional neural network extracting local and global contextual features for the classification of Alzheimer's disease using structural mri. *IEEE Journal of Biomedical and Health Informatics*, *29*(3), 2061–2073.

Jack, C. R., Bennett, D. A., Blennow, K., et al. (2018). Nia-aa research framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's and Dementia*, *14*(4), 535–562.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128.

James, A. P. & Dasarathy, B. V. (2014). Medical image fusion: A survey of the state of the art. *Information fusion*, *19*, 4–19.

Joy, M. A. M., Nasrin, S., Siddiqua, A., & Farid, D. M. (2025). Vitad: Leveraging modified vision transformer for Alzheimer's disease multi-stage classification from brain mri scans. *Brain Research*, *1847*, 149302.

Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., & Krishnan, D. (2020). Supervised contrastive learning. *Advances in neural information processing systems*, *33*, 18661–18673.

Kwak, M. G., Park, J., Lee, S., et al. (2023). Self-supervised contrastive learning to predict the progression of Alzheimer's disease with 3d amyloid-pet. *Bioengineering*, *10*(10), 1141.

Li, H., Li, H., Zhao, W., Fu, H., Su, X., Hu, Y., & Liu, J. (2023). Frequency-mixed single-source domain generalization for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 127–136). Springer.

Li, L., Zimmer, V. A., Ding, W., Wu, F., Huang, L., Schnabel, J. A., & Zhuang, X.

(2021). Random style transfer based domain generalization networks integrating shape and spatial information. In *Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges: 11th International Workshop, STACOM 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers 11*, (pp. 208–218). Springer.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, *37*(3), 50–60.

Li, X., Zhang, Y., et al. (2023). Domain contrastive learning for multi-site Alzheimer's disease classification. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE. Multi-site MRI generalization.

Li, Y., Zhang, Y., Wu, J., Zhang, X., Han, L., & Cui, X. (2024). Multi-attention-based global 3d resnet for Alzheimer's disease diagnosis. In *2024 International Joint Conference on Neural Networks*, (pp. 1–8)., Yokohama, Japan.

Lteif, D., Sreerama, S., Bargal, S. A., Plummer, B. A., Au, R., & Kolachalama, V. B. (2024). Disease-driven domain generalization for neuroimaging-based assessment of Alzheimer's disease. *Human Brain Mapping*, *45*(8), e26707.

Lu, S.-Y., Zhang, Y.-D., & Yao, Y.-D. (2025). An efficient vision transformer for Alzheimer's disease classification using magnetic resonance images. *Biomedical Signal Processing and Control*, *101*, 107263.

Lundervold, A. S. & Lundervold, A. (2019). An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, *29*(2), 102–127.

Majee, A., Gupta, A., Raha, S., & Das, S. (2024). Enhancing mri-based classification of Alzheimer's disease with explainable 3d hybrid compact convolutional transformers. In *2024 International Joint Conference on Neural Networks (IJCNN)*, (pp. 1–8). IEEE.

Manly, J. J., Bell-McGinty, S., Tang, M.-X., Schupf, N., Stern, Y., & Mayeux, R. (2005). Cultural bias in cognitive assessment: A comparison of african american and white elderly on the mini-mental state examination. *Journal of the International Neuropsychological Society*, *11*(2), 175–183.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, *19*(9), 1498–1507.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia*, *7*(3), 263–269.

Nguyen, H.-D., Clément, M., Mansencal, B., & Coupé, P. (2023). Towards better interpretable and generalizable AD detection using collective artificial intelligence. *Computerized Medical Imaging and Graphics*, *104*, 102171.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jr, C. R. J., Jagust, W. J., Shaw, L. M., Toga, A. W., & Trojanowski, J. Q. (2010). Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, *74*(3), 201–209.

Qiao, F., Zhao, L., & Peng, X. (2020). Learning to learn single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 12556–12565)., Seattle, WA, USA. IEEE.

Qiu, S., Miller, M. I., Joshi, P. S., Lee, J. C., Xue, C., Ni, Y., Wang, Y., De Anda-Duran, I., Hwang, P. H., Cramer, J. A., et al. (2022). Multimodal deep learning for Alzheimer's disease dementia assessment. *Nature communications*, *13*(1), 3404.

Rahim, N., Ahmad, N., Ullah, W., Bedi, J., & Jung, Y. (2025). Early progression detection from MCI to AD using multi-view mri for enhanced assisted living. *Image and Vision Computing*, *157*, 105491.

Rashid, A. H., Gupta, A., Gupta, J., & Tanveer, M. (2022). Biceph-net: A robust and lightweight framework for the diagnosis of Alzheimer's disease using 2d-mri scans and deep similarity learning. *IEEE Journal of Biomedical and Health Informatics*, *27*(3), 1205–1213.

Roth, H. R., Oda, H., Zhou, X., Shimizu, N., Yang, Y., Hayashi, Y., Oda, M., Fujiwara, M., Misawa, K., & Mori, K. (2018). An application of cascaded 3d fully convolutional networks for medical image segmentation. *Computerized Medical Imaging and Graphics*, *66*, 90–99.

Sarraf, S. & Tofighi, G. (2016). Deepad: Alzheimer's disease classification via deep convolutional neural networks using mri and fmri. In *Proceedings of the IEEE international conference on bioinformatics and biomedicine (BIBM)*, (pp. 1015–1020). IEEE.

Selkoe, D. J. & Hardy, J. (2021). The amyloid-beta pathway in Alzheimer's disease. *Molecular Psychiatry*, *26*, 5485–5499.

Selvester, N. J. et al. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. *The Lancet Public Health*, *6*(10), e18–e25.

Serra, J. (2023). Mathematical morphology. In *Encyclopedia of mathematical geosciences* (pp. 820–835). Springer.

Seshadri, S., Beiser, A. S., Kelly-Hayes, M., Kase, C. S., Au, P. A., Au, R., Hoffmann, U., Benjamin, E. J., Vasan, R. S., & Wolf, P. A. (2006). The lifetime risk of stroke: estimates from the framingham study. *Stroke*, *37*(2), 345–350.

Seyfioğlu, M. S., Liu, Z., Kamath, P., Gangolli, S., Wang, S., Grabowski, T., & Shapiro, L. (2022). Brain-aware replacements for supervised contrastive learning in detection of Alzheimer's disease. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 461–470)., Singapore.

Shaffi, N., Viswan, V., & Mahmud, M. (2024). Ensemble of vision transformer architectures for efficient Alzheimer's disease classification. *Brain Informatics*, *11*(1), 25.

Soille, P. (2003). *Morphological Image Analysis: Principles and Applications.* Springer Science & Business Media.

Sotiras, A., Davatzikos, C., & Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, *32*(7), 1153–1190.

Teipel, S. J. et al. (2011). Quantitative structural mri for early detection of Alzheimer's disease. *Frontiers in Neuroscience*, *5*, 128.

van Oostveen, W. M. & de Lange, E. C. M. (2024). Early detection of Alzheimer's disease in structural and functional mri. *Frontiers in Medicine*, *11*, 223.

Wang, C., Hu, Q., et al. (2023). 3-d cnn-based multichannel contrastive learning for Alzheimer's disease automatic diagnosis. In *IEEE International Conference on Biomedical and Health Informatics (BHI)*. IEEE.

Wang, R., Chaudhari, P., & Davatzikos, C. (2022). Embracing the disharmony in medical imaging: A simple and effective framework for domain adaptation. *Medical image analysis*, *76*, 102309.

Wang, Y., Chen, K., & Wang, H. (2024). Adapt: Alzheimer diagnosis through adaptive profiling transformers. https://arxiv.org/abs/2401.06349. arXiv preprint arXiv:2401.06349.

Weiner, M. W. et al. (2017). The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia*, *13*(1), 1–18.

Wu, Y., Zhou, Y., Zeng, W., Qian, Q., & Song, M. (2022). An attention-based 3d cnn with multi-scale integration block for Alzheimer's disease classification. *IEEE Journal of Biomedical and Health Informatics*, *26*(11), 5665–5673.

Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (pp. 6023–6032)., Seoul, South Korea.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLoS Medicine*, *15*(11), e1002683.

Zhang, H., Cissé, M., Dauphin, Y. N., & Lopez-Paz, D. (2018). mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*, Vancouver, Canada. ICLR / OpenReview.

Zhang, J., Zhang, Y., Wang, J., Xia, J., Zhang, J., & Chen, L. (2024). Recent advances in Alzheimer's disease: Mechanisms, clinical trials and new drug development strategies. *Signal Transduction and Targeted Therapy*, *9*(1), 211.

Zhang, J., Zheng, B., Gao, A., Feng, X., Liang, D., & Long, X. (2021). A 3d densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification. *Magnetic Resonance Imaging*, *78*, 119–126.

Zhang, S., Chen, X., Ren, B., Yang, H., Yu, Z., Zhang, X.-Y., & Zhou, Y. (2022). 3d global fourier network for Alzheimer's disease diagnosis using structural mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 34–43). Springer.

Zhang, Y., Li, M., Li, R., Jia, K., & Zhang, L. (2022). Exact feature distribution matching for arbitrary style transfer and domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (pp. 8035–8045)., New Orleans, LA, USA. IEEE.

Zhou, K., Yang, Y., Qiao, Y., & Xiang, T. (2021). Domain generalization with mixstyle. In *International Conference on Learning Representations (ICLR)*, Virtual Conference. ICLR / OpenReview.

Zhou, Y., Li, Y., Zhou, F., Liu, Y., & Tu, L. (2023a). Learning with domain-knowledge for generalizable prediction of Alzheimer's disease from multi-site structural MRI. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 452–461)., Vancouver, Canada.

Zhou, Y., Li, Y., Zhou, F., Liu, Y., & Tu, L. (2023b). Learning with domain-knowledge for generalizable prediction of Alzheimer's disease from multi-site structural mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 452–461). Springer.

Zhou, Z., Sodha, V., Siddiquee, M. M. R., Feng, R., Tajbakhsh, N., Gotway, M. B., & Liang, J. (2019). Models Genesis: Generic autodidactic models for 3D medical image analysis. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (pp. 384—-393)., Shenzhen, China.

APPENDIX A

Publications

The following publications and manuscripts are based on the work conducted during the course of this thesis:

1.1 Z. Batool, H. Özkan, and E. Aptoula, “Distance Transform Guided Mixup for Alzheimer’s Detection,” *Accepted at Signal Processing and Communications Applications Conference (SIU)*, 2025.

1.2 Z. Batool, H. Özkan, and E. Aptoula, “Single Domain Generalization for Alzheimer’s Detection from 3D MRIs with Pseudo-Morphological Augmentations and Contrastive Learning,” *Submitted to Computer Vision and Image Understanding (CVIU)*. Available at: `https://arxiv.org/abs/2505.22465`

1.3 Z. Batool, H. Özkan, and E. Aptoula, “Towards Single Domain Generalization in Alzheimer’s Detection via Extended MixStyle,” *In preparation for IEEE Transactions on Medical Imaging (TMI).*