Alzheimer's Detection and Severity Classification by MRI Analysis

Department of Data Science, National University

Kevin Carrillo

Brashon Ford

Misato Nasu

Master's Thesis Approval Form

We certify that we have read the project of Kevin Carrillo, Brashon Ford, and Misato Nasu entitled ALZHEIMER'S DETECTION AND SEVERITY CLASSIFICATION BY MRI ANALYSIS and that in our opinion, it is satisfactory in scope and quality as the thesis for the degree of Master of Science in Data Science at National University.

Approved:

Jedediah Baker, Ph.D, Capstone Project Sponsor
Professor of Data Science, Technology Department
National University

Tabbie Black, Capstone, Ph.D, Capstone Project Advisor
Professor of Data Science, Technology Department
National University

Aeron Zentner, DBA, Capstone Instructor
Part-Time Associate Professor
Department of Engineering, Data and Computer Sciences
School of Technology and Engineering
College of Business, Engineering and Technology
National University

Abstract

The general population, healthcare providers, and healthcare systems stand to benefit significantly from the development and implementation of automated diagnostic tools for Alzheimer's Disease. However, challenges such as data availability, the nature of black box models, and the disconnect between clinicians and data scientists pose significant limitations to the integration of Computer Automated Diagnostic (CAD) tools. The purpose of this project is to develop an automated diagnostic tool using a Convolutional Neural Network (CNN) model to analyze MRI scans for the diagnosis of Alzheimer's Disease. This study utilizes MRI datasets from ADNI, Kaggle, and Washington University in St. Louis, which encompass 14,609 MRI images of patients at various stages of dementia, ranging from non-demented to severely demented. The project aims to enhance diagnostic accuracy and facilitate early detection.

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

In this chapter, the study explores the events that have led up to this research on MRI analyses using Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs). The chapter introduces the reasons why the automation of MRI analyses are important for clinicians and patient outcomes. The models will be compared to assess classification accuracy and performance using accuracy, precision and F1-score.

Background

Alzheimer's Disease (AD) is a significant public health challenge, currently ranking as the seventh leading cause of death in the United States and standing as the most common cause of dementia among older adults (National Institute on Aging, 2023). AD is marked by a progressive decline in cognitive function and memory, affecting millions of older adults worldwide. Despite ongoing research, the precise cause of Alzheimer's remains elusive. Key pathological features include amyloid plaques—abnormal protein clumps in the brain—and neurofibrillary (tau) tangles, which are twisted bundles of fibers. Moreover, Alzheimer's leads to a loss of connections between neurons, disrupting vital communication pathways between the brain and the rest of the body (National Institute on Aging, 2023).

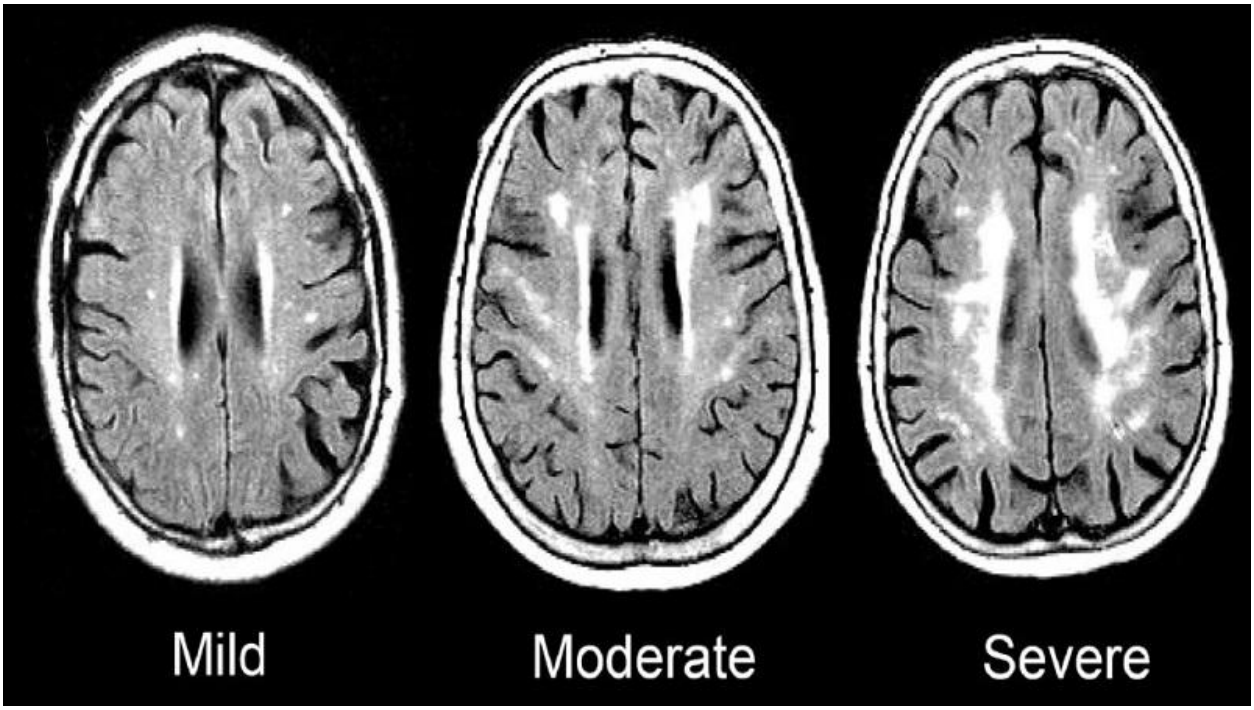
The progression of Alzheimer's is typically categorized into three stages: mild, moderate, and severe (National Institute on Aging, 2023). In the mild stage, individuals may experience difficulties with financial management, exhibit personality and behavioral changes, and develop a tendency to wander. As the disease advances to the moderate stage, it starts to impact brain regions responsible for language, reasoning, and sensory processing. At this point, memory loss and behavioral changes intensify, making it challenging for individuals to learn new tasks or complete multi-step activities. Hallucinations and impulsive behaviors may also arise. By the

severe stage, the damage caused by plaques and tangles is extensive, leaving patients unable to communicate, entirely dependent on others, and often bedridden.

Beyond its devastating health effects, Alzheimer's Disease imposes a staggering economic burden on healthcare systems worldwide. Wimo and Jönsson (2013) estimate that the global costs of dementia, which includes Alzheimer's, reached $604 billion in 2010, with expectations of substantial growth by 2030. These costs, driven by direct medical expenses, long-term care, and lost productivity, highlight the urgent need for innovative solutions. The development of automated diagnostic tools, such as Convolutional Neural Networks (CNNs), could play a crucial role in early detection, potentially reducing the financial strain on healthcare systems. Addressing Alzheimer's Disease with such tools is not just a medical necessity; it is an economic imperative that demands immediate action.

The National Institute on Aging (2023) categorizes Alzheimer's Disease into several stages: mild, moderate, and severe. Mild Alzheimer's Disease is characterized by difficulties in handling money, paying bills, personality and behavioral changes, and a tendency to wander. In the moderate stage, the disease begins to affect areas of the brain that control language, reasoning, conscious thought, and sensory processing. Behavioral changes and memory loss worsen, and the ability to learn new things or complete multi-step tasks may be hindered. Hallucinations and impulsivity may also occur. In the severe stage, extensive brain damage caused by plaques and tangles leaves patients unable to communicate. They may become completely dependent on others and possibly bedridden.

Figure 1: MRI Scans of Mild, Moderate, and Severe Alzheimers



(Johns Hopkins Medicine, 2018)

According to the National Institute on Aging (2023), clinicians use various methods to diagnose Alzheimer's, which include conducting tests of memory, problem-solving, attention, counting, and language. They may also order blood and urine tests to rule out other potential causes, collect cerebrospinal fluid or blood samples to measure levels of proteins associated with Alzheimer's Disease, and perform brain scans, such as MRIs and CTs.

While these assessment methods are effective, they are also time-consuming and subject to variability based on the clinician's expertise. This variability in assessments underscores the need for automated diagnostic tools, which is why many medical researchers are exploring the use of machine learning techniques, particularly Convolutional Neural Networks (CNNs), to automate the diagnosis and classification of Alzheimer's Disease. An automated diagnostic tool

capable of analyzing complex brain scans can reduce diagnostic bias and free up clinicians to focus on other critical tasks.

The journey toward automating diagnoses using machine learning began in 1992 when researchers used Artificial Intelligence (AI) to analyze mammograms in radiology. However, these early models were labor-intensive and did not lead to significant improvements in diagnostic accuracy. With advancements in deep learning, particularly CNNs, AI has transformed, improving the quality of model predictions (Driver et al., 2019).

Milestones in CNN research for Alzheimer's diagnosis began with early studies in 2015, when Adrien Payan and Giovanni Montana developed a 3D CNN model to predict Alzheimer's Disease status using MRI scans. This study was a significant step in applying deep learning to medical imaging. Between 2019 and 2020, advancements focused on optimizing CNN architectures and integrating them with other machine learning techniques, such as Support Vector Machines (SVMs) and Random Forests (RFs), to enhance diagnostic accuracy. Recent developments, from 2021 onward, have refined CNN models to handle complex data and improve interpretability, with researchers like Giridharan et al. (2022) using pre-trained models like ResNet50 to create highly accurate diagnostic tools.

By leveraging these advancements in machine learning, CNNs have become a crucial tool in the early and accurate diagnosis of Alzheimer's Disease, providing a pathway to more efficient and standardized diagnostic processes.

Problem Statement

Alzheimer's Disease impacts around 6.9 million people in the United States, mainly adults aged 65 and older (Alzheimer's Association, 2024). Affected individuals often face severe disorientation, difficulties managing finances, and health challenges due to cognitive impairments. Additionally, memory loss can lead to medication mismanagement, potentially worsening their condition.

Current diagnostic methods for Alzheimer's disease are often slow and vary based on the type of dementia, symptoms, and resource availability. Diagnosis can take months or years, requiring multiple visits and various tests. A shortage of neurologists exacerbates these delays, especially as the aging population increases demand for specialized care. This lengthy process hinders early diagnosis, underscoring the need for more efficient diagnostic approaches to improve patient outcomes (Dementia Care Central, 2023).

An automated diagnostic tool using Convolutional Neural Networks (CNNs) could address these issues by reducing variability and streamlining MRI analysis. This study evaluates whether CNNs can improve diagnostic accuracy and efficiency compared to Support Vector Machines (SVMs), aiming to enhance patient outcomes and healthcare delivery amidst the rising Alzheimer's Disease burden.

Research Hypotheses

The primary purpose of this research is to develop an automated diagnostic tool for Alzheimer's Disease using Convolutional Neural Networks to analyze MRI scans. This study aims to evaluate the accuracy and efficiency of CNNs in diagnosing Alzheimer's Disease compared to traditional diagnostic methods and other machine learning models like Support Vector Machines. The question to be answered is, How does the accuracy and efficiency of CNNs in diagnosing Alzheimer's Disease from MRI scans compare to traditional diagnostic methods and other machine learning models, such as Support Vector Machines? Based on this purpose, and question the research hypotheses are as follows:

Hypothesis 1:

The CNN-based model will demonstrate higher or similar accuracy in identifying the presence and severity of Alzheimer's Disease compared to clinician analysis of MRIs.

Rationale:

Traditional diagnostic methods although effective, are time-consuming and prone to variability based on the clinician's expertise (Diogo, et al., 2022). CNNs have the ability to streamline the diagnostic process, reducing subjectivity and increasing accuracy.

Expected Outcome:

The CNN model will outperform or match traditional diagnostic methods in terms of accuracy, as measured by correlation matrices and loss curves.

Hypothesis 2:

The CNN-based model will demonstrate higher accuracy in identifying the presence and severity of Alzheimer's Disease compared to a Support Vector Machine model.

Rationale:

SVMs are effective for certain classification tasks. However, CNNs are specifically designed for image analysis and have shown superior performance in medical imaging tasks.

Expected Outcome:

The CNN model will achieve higher accuracy metrics compared to the SVM model, demonstrating its superiority in classifying MRI images for Alzheimer's Disease diagnosis.

Hypothesis 3:

The integration of a CNN-based diagnostic tool will significantly reduce the time required for diagnosing Alzheimer's Disease compared to current methods.

Rationale:

Diagnosis of Alzheimer's Disease involves multiple steps including blood and Cerebrospinal fluid tests, cognitive tests, and MRIs. It may take weeks to schedule an MRI and another 2 weeks before receiving information on your MRI scan results. An automated CNN tool could cut weeks off of this process, providing quicker results

Expected Outcome:

The time taken to diagnose Alzheimer's Disease using the CNN-based tool will be reduced to facilitate earlier diagnoses for better patient outcomes.

By testing these hypotheses, this research aims to demonstrate the potential of CNNs to enhance the diagnostic process for Alzheimer's Disease, ultimately contributing to more accurate, efficient, and standardized diagnostic practices in clinical settings.

Objectives

Alzheimer's elusive and progressive nature presents a formidable challenge in the medical field and the individuals plagued by it. With millions of people affected globally and the numbers expected to rise, Alzheimer's Disease has become a focal point of concern not only for its impact on patients and their families but also for the strain it places on healthcare resources (Alzheimer's Association, 2021). The need for timely and accurate diagnostic tools grows greater every year.

Current methods of diagnosis rely heavily on a combination of cognitive assessments and neuroimaging techniques like MRI scans. While these methods have played a crucial role in identifying and managing the disease, recent advancements in machine learning, like CNNs, offer enhancements to the diagnostic process. As previously noted, CNNs are particularly well-suited for image analysis due to their ability to automatically learn and extract features from complex datasets. They are able to identify subtle patterns in MRI scans that may be indicative of Alzheimer's Disease, potentially surpassing the accuracy of current diagnostic methods (LeCun, Bengio, & Hinton, 2015). The application of CNNs in the diagnosis of Alzheimer's

Disease could lead to faster and more accurate detection and classification. This is critical, given that early intervention can significantly impact the progression of Alzheimer's Disease.

Using a trusted diagnostic tool also reduces the variability in current diagnostic methods and increases diagnostic reliability (Wen et al., 2020). This could bring substantial benefits in the clinical setting to decision-making and patient outcomes. Such a tool could optimize resource allocation within healthcare systems and reduce the toll this disease takes on our healthcare system.

There are three main objectives of this project. The first is to train and validate a CNN-based diagnostic tool using MRI scans to assess its effectiveness in identifying the presence and severity of Alzheimer's Disease. The second objective is to compare the performance of the CNN model with an SVM model. SVMs have been widely used in classification tasks and serve as a useful baseline for evaluating the advancements offered by deep learning models (Cortes & Vapnik, 1995). Finally, this project seeks to determine the practical application of the CNN-based diagnostic tool in clinical settings. This includes its ability to integrate into existing diagnostic workflows and its impact on clinical decision-making and resource allocation. By completing these objectives, this project aims to provide insight into how much advancements in machine learning can contribute to improving the diagnosis of Alzheimer's, and other elusive diseases.

Limitations of the Study

MRI scans have proven to be the go-to method for imaging Alzheimer's Disease, but other imaging modalities, such as CT and PET scans, are used by clinicians. These alternative imaging techniques provide different information that could complement MRI findings. A comprehensive diagnostic tool would ideally incorporate data from CT and PET scans as well. Future research may need to apply similar machine-learning techniques to these imaging modalities to ensure that the diagnostic tool is robust and applicable across various imaging types.

In the same nature, this project focuses solely on one aspect of the diagnosis process in analyzing brain images. Diagnosis typically involves a combination of cognitive assessments and laboratory tests as well. This project does not encompass the full spectrum of diagnostic criteria, which could potentially limit the tool's applicability in real-world scenarios. Future research should seek to enhance the accuracy and efficiency of these tests as well.

Another limitation is the limited information available within the datasets. While there are more than enough images in the datasets to train a robust CNN model, they are limited in terms of demographic details like age, gender, and comorbidities. The addition of these variables could significantly improve the model's accuracy in diagnosing and classifying Alzheimer's Disease in diverse groups. Without them, the model might miss subtle patterns that could be crucial for improving diagnostic accuracy.

A significant challenge is present in most tools automated by artificial intelligence. That is a lack of model interpretability, or the "Black Box" problem. Models like CNNs alone do not provide clear explanations for their predictions. This lack of interpretability is what hinders

model acceptance in many industries. The end user's ability to understand the rationale behind a diagnosis is as important as the diagnosis itself. Future research should explore Explainable Artificial Intelligence (XAI) methods.

Finally, the resources available for this study, particularly the RAM limitation in Google Colab, impose constraints on the scope and scale of the analysis. The limited memory can restrict the size of the datasets that can be processed simultaneously and the complexity of the models that can be trained. This limitation may affect the overall performance and scalability of the model, especially when dealing with larger datasets.

Definition of Terms

Table 1: Definition of Terms

Term	Definition	Source
Alzheimer's Disease (AD)	A brain disorder that slowly destroys memory and thinking skills, and is the most common type of dementia in older adults. It usually starts in late middle age or old age and worsens over time.	National Institute on Aging, 2023
Amyloid Plaques	Abnormal clumps of a protein called beta-amyloid that form in the brain between nerve cells and other cells.	National Institute on Aging, 2023
Computed Tomography (CT) Scan	A diagnostic imaging procedure that uses a combination of X-rays and computer technology to produce images of the inside of the body.	Johns Hopkins Medicine, n.d.
Convolutional Neural Network (CNN)	A deep learning neural network designed for processing structured arrays of data such as images.	Wood, n.d.

Dementia	A general term for a loss of cognitive functioning that affects a person's ability to think, remember, learn, make decisions, and solve problems.	Alzheimer's Disease and Healthy Aging, 2019
Explainable Artificial Intelligence (XAI)	A set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms.	IBM, n.d.
Magnetic Resonance Imaging (MRI)	A noninvasive medical imaging test that produces detailed images of almost every internal structure of the human body, including the organs, bones, muscles, and blood vessels.	Johns Hopkins Medicine, n.d.
Neurofibrillary (Tau) Tangles	Abnormal accumulations of a protein called tau that collect inside neurons.	National Institute on Aging, 2023
Positron Emission Tomography (PET) Scan	An imaging test that can help reveal the metabolic or biochemical function of your tissues and organs. The PET scan uses a radioactive drug called a tracer to show both	Mayo Clinic, 2023

	typical and atypical metabolic activity.	
Support Vector Machine (SVM)	A supervised machine learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space.	IBM, 2023

Summary

Chapter 1 introduced Alzheimer's Disease as a growing public health concern, affecting millions of individuals, primarily older adults. A detailed description of Alzheimer's Disease was provided, highlighting its stages and the profound impact it has on patients, their families, and the healthcare system. Current diagnostic methods, though effective, are time-consuming and subject to variability, which underscores the need for an automated diagnostic tool.

The purpose and objectives of the research were outlined, emphasizing the development of a convolutional neural network model to diagnose Alzheimer's Disease from MRI images. Research hypotheses were presented, predicting that CNN models will provide high accuracy and consistency in diagnosing Alzheimer's Disease, potentially outperforming current methods and SVMs.

A brief history of the use of CNNs in medical imaging was discussed, tracing their evolution and application in Alzheimer's Disease diagnosis. Self-imposed limits and external constraints were acknowledged, providing a realistic scope for the research.

With a clear understanding of the problem and the research goals described in Chapter One, Chapter Two delves into the literature review. This chapter will explore existing research on the use of CNNs in medical imaging, particularly for diagnosing Alzheimer's Disease. It will also examine studies that have contributed to our understanding of Alzheimer's Disease and the advancements in machine-learning techniques that have paved the way for this research. Through a comprehensive review of relevant literature, Chapter Two will lay the foundation for the methodology and analysis presented in the subsequent chapters.

Chapter 2: Literature Review

Introduction

Early detection of Alzheimer's disease is crucial as it enables individuals to plan for their future, improves the quality of life, supports carers, and can lead to significant societal and economic benefits by delaying the progression of the disease and reducing overall care costs (Rasmussen & Langerman, 2019). This chapter reviews existing literature on convolutional neural networks (CNNs) and magnetic resonance imaging (MRI) for early Alzheimer's disease detection. The review was organized into five main categories: Early Research, Research Progression, Recent Research, Identified Gaps, and Summary.

Early Research:

Much of the knowledge foundational to our modern understanding of Alzheimer's disease can be attributed to the work of the German doctor and researcher Alois Alzheimer (Hippius & Neundörfer, 2003). After receiving his diploma in medicine, a German family hired Alois Alzheimer to accompany them to look over a mentally ill relative for five months. This sparked Alzheimer's interest in neurological research and led him to an assistant position at the Community Psychiatric Hospital in Frankfurt, Germany, where he was able to hone his skills as a psychiatrist. In 1901, at the Frankfurt hospital, Alzheimer met a patient named Auguste Deter, who would become the first patient ever to receive a diagnosis of Alzheimer's Disease.

Auguste Deter came into the Frankfurt hospital at age 51 after displaying a host of unusual symptoms (Yang et al., 2016). Initially, Deter presented with signs of delusions, accusing her husband of adultery. A rapid decline followed this in her memory and function. Her condition continued to deteriorate as she slowly lost the ability to cook or follow directions,

developed insomnia, and exhibited threatening behaviors toward her neighbors. Karl, Auguste's husband, brought her to the hospital after noticing these concerning changes in his wife's condition. This led to Alzheimer's assuming control over Auguste Deter's care. During his evaluation, he checked various functions, such as if Auguste could say her name, name objects, and recall her lunch, among numerous similar assessments. Through this questioning, Alzheimer found that although she was able to name herself and simple objects like pencils when placed in front of her, she struggled with her short-term memory as she was not able to remember what she had for her meal. Following this assessment period, Auguste was diagnosed with 'presenile dementia'.

To further observe Auguste and her health, Alzheimer's recommended that Auguste stay hospitalized in Frankfurt despite concerns raised by Karl Deter related to high costs. In an attempt to address this issue, Alzheimer's allowed the patient to stay in the hospital for free in exchange for unrestricted access to Auguste's medical record as well as possession of her brain after her death for research purposes. During her hospitalization, Auguste was observed to display offensive behaviors toward others and often needed to be secured in a locked room at night to keep her and the hospital safe.

Following the death of his wife, Alzheimer decided to leave Frankfurt to pursue work in Munich. Around the same time, the health of Auguste Deter was also rapidly declining, leading to her death in 1906. As promised, Alzheimer received Auguste's brain and medical records and immediately began investigating. A biopsy was conducted on the brain to investigate if there was any correlation between the medical records, which consisted of the patient's medical history and the patient's symptoms. Through the biopsy, Alzheimer discovered that the cerebral cortex, the region of the brain crucial for memory, language judgment, and thinking, was thinner than usual

and severely impaired. Another significant discovery was the presence of two characteristics that are pathologically characteristic of Alzheimer's Disease: lesions in the brain, which are known to us today as amyloid plaques, and abnormal protein structures known as neurofibrillary tangles. Amyloid plaques are extracellular lesions found in the brain that are circular and usually have a diameter of 10 - 200 micrometers. They form when protein fragments, known as beta-amyloids, clump together and may contain other cellular debris (Kumar et al., 2024). These plaques disrupt the normal function and communication between neurons (Castellani et al., 2010). Tangles, also known as neurofibrillary tangles (NFT), are another critical microscopic feature of Alzheimer's disease. They mainly affect large pyramidal neurons in parts of the brain called Ammon's horn in the hippocampus and the cerebral cortex. These tangles are also found in deeper brain areas, including the midbrain, the brainstem (specifically the pontine tegmentum), the basal nucleus of Meynert, and the hypothalamus (Kumar et al., 2024). Tangles occur when abnormal accumulations of a protein called tau form in neurons. In a healthy neuron, tau helps support structures called microtubules that serve a role in transporting nutrients and molecules within the cell. In neurons with NFTs, tau molecules stick to one another to form threads that tangle together and block the cell's transportation system. This leads to a breakdown of synaptic communication between neurons (Castellani et al., 2010). At the time of the autopsy, tangles were a phenomenon that had never been observed. Additionally, amyloid plaques had previously only been observed in patients over 70, making these extraordinary discoveries.

Following his research breakthroughs, Alzheimer announced his findings at the 37th Psychiatry Conference in southwestern Germany. Addressing a group of his colleagues, he presented the case study of Auguste and the progression of her unusual disease with symptoms such as worsening cognitive impairment, localized neurological issues, hallucinations, delusions,

and psychological and social disabilities (Hippius & Neundörfer, 2003). Alongside this, he revealed images of the findings from the brain biopsy depicting senile plaques, neurofibrillary tangles, and atherosclerotic changes. While at the time of the conference, the scientific community did not receive Alzheimer's findings well, his research would later prove to be instrumental to the discovery of Alzheimer's disease, having researched and analyzed the first Alzheimer's patient known to science .

Despite the significance of Auguste's findings, due to his hostile reception at the conference, they were not taken seriously the first few years following publication. At this time, not much research was conducted on Alzheimer's Disease. It would not be until several years later, in 1964, that there would be a resurgence in interest in the disease due to technological advancements. Another scientist who would prove to be important in advancing Alzheimer's research was Kidd, who studied brains affected by Alzheimer's Disease through the use of electron microscopy.

Kidd collected tissue samples and white matter from five patients to conduct his research. The method of collection involved drilling parietal burr holes using ventriculography to visualize the tissues through which pieces measuring 1 ml in volume were extracted. These one ml-sized pieces were cut into smaller pieces and fixated with two distinct solutions for four hours to stabilize the condition of the tissues and prevent breakdown. The two solutions utilized in this process were osmium tetroxide and potassium permanganate. The tissue was then washed, dehydrated in alcohol, and embedded in Araldite, which is a type of epoxy resin used as an embedding medium to stabilize these fragile tissues. Finally, the tissues were stained with alcoholic phosphotungstic acid or lead citrate for better visualization of detailed structures while

examining the tissues under an electron microscope at a magnification ranging from 2,000 to 80,000 times.

Through this process, Kidd observed some significant differences between healthy and Alzheimer's affected brain tissue. The first change he observed using electron microscopy was the presence of tightly packed filaments bundled in affected neurons around the cell nucleus, also known as neurofibrillary tangles. As the bundles occupy space in the cell's cytoplasm, another abnormality Kidd observed was how other cellular structures were disrupted by the presence of these bundles. There was noted to be lamellated elliptical bodies caught in bundles, which according to some research such as that by Webster, suggests that they appear as the result of a degenerated mitochondria. Vesicular and tubular profiles embedded in a dense mass were found in the bundle as well. These two are believed to originate from the Golgi apparatus and endoplasmic reticulum, suggesting that these two organelles were disrupted by the bundles as well.

Other abnormalities Kidd was able to observe were irregular neuropil and plaques. At low magnification, the neuropil seemed normal. However, with higher magnification, Kidd was able to observe the presence of the helical filaments that comprised the pathological bundles, which is only seen in Alzheimer's-affected brains. The plaques that were observed were defined as regions with multiple irregular cell processes. These plaques often had a central zone with few normal processes and a periphery with a combination of both normal and abnormal structures. While the central region appeared almost homogenous due to the lack of dense particles, the periphery of the plaques showed swollen processes containing numerous highly dense particles, making the plaques easy to distinguish. Throughout the entirety of the plaques, one striking characteristic was the complete absence of synapses and myelin.

Despite the detailed discoveries made by Kidd during his research, there was a high amount of variability in both the appearance and structure of the abnormalities he reported. As such, at this time, Kidd was not able to concretely prove the correlation of these differences between normal brain tissue and AD-affected brain. This serves to highlight how complexity underscores the challenge of fully understanding the pathological changes at the microscopic level in Alzheimer's Disease. Despite the lack of a clear conclusion, Kidd's research still proved to be valuable as it was able to open new possibilities for diagnosing Alzheimer's disease with electron microscopy.

From the time of Kidd's research, technology advanced even more, leading to further innovation. In 1979, there were two methods typically utilized to diagnose Alzheimer's disease, both involving CT scans. The first and most common way was to rate the degree of brain atrophy ranging from average to severe with a 5 point scale. The second method was a quantitative examination of the linear aspects of the ventricular system and the cortical sulci. A pitfall of both these methods is the need for implied normative data and specific linear criteria which can involve a certain level of assumption.

De Leon et al. (1979) suggested a third method using CT scans to diagnose and categorize Alzheimer's disease. His method ranked CT scans continuously based on observed ventricular system changes and cortical sulci changes rather than specific linear measurements. The research consisted of 43 outpatients with a presumptive diagnosis of AD. From the 43, De Leon selected 41 patients for ventricular studies and 31 patients for the cortical studies, where at least two CT scans can be used for each study for each patient. The images were then ordered by two independent observers in a visually continuous sequence of increasing ventricular dilatation or cortical sulcal prominence. This method of continuous ranking aimed to provide a more

refined assessment of brain atrophy without relying on predefined categorical scales or specific linear measurements. The observers were blinded to the behavioral information of the patients to eliminate bias.

The results of this approach demonstrated that 65% of the cognitive measures correlated significantly with ventricular dilatation, while 27% correlated with sulcal prominence. The difference in these percentages was statistically significant, indicating a stronger and more consistent relationship between ventricular changes and cognitive function compared to cortical changes. Specifically, 17 out of 28 memory measures and 5 out of 7 psychomotor performance tasks were significantly associated with ventricular dilatation. In contrast, only 7 out of 28 memory measures and 1 out of 7 performance tasks significantly correlated with cortical sulcal prominence. Additionally, both global measures of cognitive deterioration were significantly correlated with both rankings.

These findings suggested that the continuous ranking method provided a more detailed and reliable assessment of the relationship between brain atrophy and cognitive decline in Alzheimer's Disease. By avoiding the assumptions required by traditional methods and providing a more nuanced view of brain changes, this method offered potential improvements in diagnosing and categorizing Alzheimer's Disease. The researchers noted that their approach could be compared to existing CT evaluation techniques to validate its effectiveness and applicability further. De Leon's research contributed to a deeper understanding of how CT imaging can be utilized to better align brain changes with cognitive symptoms. It highlighted the potential for more accurate and meaningful diagnostic practices in the field of Alzheimer's Disease, which might be able to transition into MRI scans with further investigations.

De Leon et al. (1988) investigated the link between Alzheimer's disease and hippocampal atrophy, building on earlier research by Sapolsky and his team. Sapolsky's studies found that aging rats showed prolonged glucocorticoid secretion following stress, an abnormal adrenocortical stress response that impaired the hippocampus's role in regulating stress hormones. This impairment led to a cycle of prolonged stress and further hippocampal damage. De Leon hypothesized that this abnormal increase in cortisol secretion and resulting hippocampal damage could be related to Alzheimer's disease. If these abnormalities are observed, they serve as early indicators of AD.

To test this hypothesis, De Leon studied nine patients with mild to moderately severe Alzheimer's disease and eight age-matched controls, all of whom were free from other diseases that could affect the brain. All participants underwent intravenous glucose-tolerance tests (GTT) to measure cortisol levels.

The GTT results showed that AD patients had significantly higher cortisol levels 90 minutes to 2 hours after glucose administration than the control group. At the 2-hour mark, elevated cortisol levels in AD patients were strongly correlated with hippocampal atrophy, as observed on CT scans. These elevated cortisol levels were also closely associated with cognitive impairment, as measured by the global deterioration scale, and with reduced glucose utilization in the brain, as measured by positron emission tomography (PET) in the temporal lobe. This suggests a strong link between physical damage to the hippocampus and abnormal stress hormone responses in Alzheimer's disease.

Among the control group, some individuals with minimal hippocampal atrophy also showed higher cortisol levels, indicating that even minor hippocampal damage might be linked to abnormal cortisol responses. When comparing these findings with those in patients with multi-

infarct dementia (MID), another form of dementia, it was found that most MID patients did not exhibit the same abnormal cortisol response as AD patients. This finding suggests that the abnormal cortisol response might be specific to Alzheimer's disease, potentially aiding in its early diagnosis. However, further research is needed to fully understand the relationship between these factors and Alzheimer's disease.

In 1996, Gómez-Isla et al. (1996) tried to tackle the problem of distinguishing the early onset of Alzheimer's disease from aging. Even though neurofibrillary tangles and senile plaques are a great hallmark of Alzheimer's disease-affected brains, these two can also be found in the brains of aging seniors, which made some researchers question if Alzheimer's disease is just an accelerated form of aging. However, others argue that aging and Alzheimer's disease are two completely different processes that both affect memory and other brain functions.

To investigate this, Gomez-Isla investigated 20 entorhinal cortex of dead subjects whose medical information before death was known. The entorhinal cortex is a crucial component in the brain that acts as a gateway connecting the neocortex and the hippocampal formation. These 20 subjects were divided into two groups: those with Alzheimer's disease or not. All subjects' brain tissues were fixed using buffered 10% formalin and cut into 1 cm thick slices from the left cerebral hemisphere for neuropathological examination, consisting of a density check of neurofibrillary tangles and senile plaques. Tissue blocks from the right cerebral hemisphere were taken for neuronal counts. The neuronal counts were explicitly taken from the entorhinal cortex, using the same techniques for consistent data collection.

Finally, the data collected were processed through computer software using linear regression and ANOVA for statistical investigation. This investigation found that neuronal numbers remain stable in cognitively normal aging, with no significant differences observed

between individuals from the sixth to the ninth decades. However, in Alzheimer's Disease, there is a substantial reduction in EC neurons, with a 48% decrease in total neuron count and a 40% reduction in EC volume compared to controls. The most pronounced loss occurs in layer II, which shows a dramatic 72% reduction in severe AD cases. Neuronal loss correlates significantly with neurofibrillary tangles (NFTs) and neuritic plaques but not with amyloid plaques alone. The study suggests that neuronal loss in the EC is a significant early feature of AD, occurring even in very mild stages of the disease, highlighting the need for early diagnostic tools to detect AD before extensive neuronal damage occurs.

The early research into Alzheimer's disease laid the groundwork for our understanding of its pathology and diagnostic challenges. From Alois Alzheimer's pioneering observations of senile plaques and neurofibrillary tangles to subsequent advancements using electron microscopy and CT scans, each phase of research has progressively deepened our insight into the disease. These foundational studies have highlighted the intricate changes in brain structures and underscored the need for more refined diagnostic methods. As we transition into the discussion of Research Progression, it becomes evident that these early findings have significantly influenced the evolution of diagnostic techniques and research methodologies. This next section will synthesize how the research landscape has evolved, illustrating the advancements made and how these have shaped current approaches to early detection of Alzheimer's disease, mainly through convolutional neural networks (CNNs) and magnetic resonance imaging (MRI).

Research Progression

Early researchers helped society understand Alzheimer's disease and everything surrounding it. As technology advanced, machine learning started to be used in healthcare,

including Alzheimer's diagnosis. One of the earliest Alzheimer's disease diagnosis research done with machine learning techniques and images was from 2008 by Stefan Kloppel and his team. Before Kloppel's research, the primary way to diagnose Alzheimer's disease was based on clinical history and examination supported by neuropsychological evidence of the pattern of cognitive impairment. There were ways to increase the accuracy of diagnosing Alzheimer's disease, such as NINCDS-ADRDA and the Diagnostic and Statistical Manual, but these were not time-effective. Around the same time, MRI images were usually used to exclude the possibility of other dementia causes and not used to diagnose Alzheimer's. A new possibility grew that it is possible to diagnose Alzheimer's disease positively. Kloppel et al. (2008) considered automating the observation of MRI scans using support vector machines (SVMs) to incorporate MRI scans in diagnosis.

The research brought in patients and their MRI scans from different research centers; these patients were split into four groups. Group one consisted of 20 patients and 20 control patients from a community and referral based in Rochester, Minnesota. The second group comprised 14 patients and 14 control patients from the Dementia Research Centre, University College London. Group three had 33 patients with probable mild Alzheimer's and 57 control patients, and all came from the same source as Group one. The only difference is that the diagnosis of these patients was 'probable' according to DSM-III-R and NINCDS-ADRDA criteria for Alzheimer's disease. Group four consisted of 19 subjects with different forms of dementia other than Alzheimer's to help differentiate Alzheimer's and other dementia types.

The MRI scanning method for each group was different, so the images were ensured to have good quality control and were normalized for machine learning algorithm usage. The specific machine learning algorithm used in this study is the SVM, a supervised, multivariate

classification method. This machine-learning technique was used for other neuroimaging research, but Kloppel was one of the first to use it for Alzheimer's. SVM classifies images fishing a hyperplane that maximizes the margin between different classes. This research will be conducted using MRI scans of Alzheimer patients and control patients. The SVMs are trained using a leave-one-out method to ensure generalizability, and the importance of voxels in classification is assessed. It will also be programmed to analyze the whole brain and region-specific for better analysis.

The machine learning algorithm showed promising results in categorizing the MRI scans into their class and diagnosing Alzheimer's disease. When using the whole brain for image analysis, all groups had a classification accuracy of at least 81%. In the SVM classification, when only anteromedial lobe volume of interest was used for analysis, two categories had a lower accuracy of around 70%, but the rest had at least 85% accuracy. This research became one of the pioneers of machine learning and brain scans to categorize Alzheimer's.

Following Kloppel's research, other researchers started to look into using machine learning techniques to categorize and diagnose Alzheimer's disease. 2009 Hinrichs et al. (2009) published a different machine learning compared to the SVMs to categorize and diagnose Alzheimer's disease, Linear UProgram (LP) boosting. Hinrichsa observed a few problems using SVMs from Kloppel's research and others using the same machine algorithm. The first problem was that using all image voxels made the SVM algorithm less accurate in diagnosing and categorizing the MRI scans, leading to extensive feature selection and dimensionality reduction. Secondly, SVM cannot account for spatial relationships between voxels. This meant that spatial information needed to be incorporated in preprocessing and post-processing. This, with SVM

limitations, led to extensive processing being required, which could have been more cost-effective.

The suggested model in this paper will address the limitation of spatial relationships for SVMs. LP boosting allowed the research team to bypass the lengthy processing phase so that the algorithm could incorporate spatial relationships. The research team decided to use large datasets to address the dataset being negligible when using the boosting method. The algorithm used images from ADNI and was preprocessed by segmenting the images into gray matter, white matter, and cerebrospinal fluid(CSF), a process similar to what Kloppel had with his research. The photos were also normalized, corrected for bias, modulated, smoothed out, and employed an absolute threshold masking to minimize white matter inclusion to focus on the gray matter.

Collecting MR images from Alzheimer's patients and regular control patients from ADNI datasets, the images were split into two groups—group one for the Alzheimer's patients and group two for the control. The images were put through the algorithm and the classification, accuracy, and ROC curve. The algorithm had an 82% accuracy with a sensitivity of 85% and specificity of 80%. The ROC curve had an AUC of 0.88, showing good predictive accuracy. The research team also looked into the impact of spatial constraints and learned that spatial constraints are necessary for accurate prediction when diagnosing Alzheimer's. This research also showed a possible future of early detection of Alzheimer's using machine learning. There are a few limitations, such as overfitting, so it was determined that more research is needed to investigate the best option for the detection of Alzheimer's using machine learning.

To dive further into the research done by Kloppel and Hinrichsa, Kohannim et al. (2010) created an SVM algorithm and a boosting algorithm to incorporate multiple biomarkers to classify

subjects from ADNI into three categories. These three categories are subjects with Alzheimer's disease, mild cognitive impairment, and normal subjects.

Other than MRI scans, other biomarkers can be used to diagnose and distinguish mild cognitive impairment, an intermediary stage between normal aging and Alzheimer's disease. One biomarker is the PET scan, which can detect the accumulation of amyloid plaques and tau protein tangles. The second biomarker is PET-FDG, which measures metabolic activity in the brain. In this research, the area with reduced glucose metabolism will be checked, which is associated with neurodegeneration in Alzheimer's. Thirdly, the Cerebrospinal Fluid Biomarkers will check for low or high levels to check for amyloid plaque deposition and tau tangles, respectively. Lastly, the APoE genotype can be a biomarker, as the presence of this genotype is associated with an increased risk of developing Alzheimer's disease.

After testing various biomarker combinations with boosting algorithms, Kohannima found that using MRI measures of hippocampal volume and ventricular size, along with ApoE genotype and age, achieved 82.21% accuracy in diagnosing Alzheimer's Disease (AD) and an ROC AUC of 0.945. This demonstrates the value of these MRI metrics in understanding AD. The classification of Mild Cognitive Impairment (MCI) was less accurate, but hippocampal volume, ApoE genotype, and age were still crucial predictors. Adding CSF and PET-FDG biomarkers improved the accuracy further, with CSF tau levels being critical for AD and CSF Aβ42 for MCI. Predicting cognitive decline in MCI subjects was difficult, but baseline PET-FDG, MRI temporal and ventricular data, and ApoE genotype effectively forecast a decrease over a year. This approach suggests that using multiple biomarkers can help reduce clinical trial sample sizes by identifying participants more likely to show significant atrophy, improving trial

efficiency, and detecting treatment effects. Combining these biomarkers with machine learning shows promise for better diagnosing and predicting outcomes in AD and MCI.

Besides SVM and boosting a third machine-learning technique was used in 2012 to detect Alzheimer's disease. Westman et al. (2012) investigated if the combination of baseline MRI and CSF could improve the classification of Alzheimer's disease compared to using each method individually and whether it is possible to predict the progression of MCI using orthogonal partial least squares to latent structures (OPLS). OLPS is a multivariate data analysis technique that handles complex datasets like the MRI and CSF biomarkers. OPLS is a data analysis method that separates functional patterns from noise and focuses on the relationship between predictors and outcomes. It works well with complex datasets with more variables than samples and offers insights into which variables are most important, similar to linear SVM.

Like Kohannim et al. (2010) research, Westman tested out different combinations of biomarkers, including MRI (hippocampus, cerebral cortex, etc.) and CSF (Amyloid-beta 42, total tau, etc.) measures to see the best combination of biomarkers to diagnose Alzheimer's disease. Combining MRI and CSF measures resulted in the most accurate classification of 91.8%, distinguishing Alzheimer's from the control. MRI alone had 87%, and CSF had 81.6% accuracy, which proved that the combination yielded a better prediction algorithm. This was the same for MCI vs Control, where the combined algorithm had a 77.6% accuracy while individual biomarkers had less than 72% accuracy. However, only 54.3% of stable MCI subjects were correctly categorized, while 90.1% of the MCI converters to Alzheimer's disease were categorized correctly. The algorithm also tried to predict MCI conversion, but the accuracy fell between 58% - 86.5%, depending on the number of months it takes for the MCI case to work. This research showed the potential to use machine learning to diagnose and detect Alzheimer's.

Still, there will need to be more research on the algorithms and cost-effectiveness, especially if the machine learning algorithm will be used to predict MCI conversions.

Following methods such as SVM, boosting, and OPLS, Payan and Montana (2015) introduced a fourth machine learning method—convolutional neural networks (CNNs)—for categorizing and diagnosing Alzheimer's Disease (AD) using MRI data. They utilized MRI scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI), normalizing the scans into a 68x95x79 voxel format, equating to 510,340 voxels per image. The images underwent a two-stage process involving both 2D and 3D convolutional networks. The training process started by training a sparse autoencoder to learn convolutional filters. The study used 2,265 historical scans, with 1,731 for training, 306 for validation, and 228 for the final evaluation of the algorithms.

The study tested the images across four classification categories: Alzheimer's vs. healthy control, Alzheimer's vs. MCI, healthy control vs. MCI, and a three-way classifier for Alzheimer's, MCI, and healthy control. The 3D algorithm showed higher accuracy in all but one classification category, Alzheimer's vs. healthy control, where both 2D and 3D methods achieved an accuracy of 95.39%. This category also exhibited the highest accuracy overall, while the lowest was 82.24% for the 2D Alzheimer's vs. MCI classification. Notably, all 3D classifiers achieved over 86% accuracy. The enhanced accuracy of the 3D algorithm is attributed to its ability to capture volumetric and spatial relationships between adjacent slices, allowing it to detect the subtle and widespread structural changes in the brain associated with MCI or Alzheimer's. This research became a pioneer and foundation in using CNN algorithms for Alzheimer's diagnosis.

CNNs have been used on other datasets as well. Islam and Zhang (2017) used 100 patient data from the Open Access Series of Imaging Studies (OASIS), which consists of those over 60 with diagnoses ranging from very mild to moderate Alzheimer's Disease. Because the dataset was small for this algorithm, the algorithm needed to be run with 5-fold cross-validation where each fold consisted of 70% as training data, 10% as validation, and 20% as test. The size was set to 299X299X1, and the convolutional filter sizes fit the MRI data. Using this algorithm, Islam and Zhang achieved an accuracy of 73.75%. This process only took a little less than an hour, which is much more efficient than manually diagnosing with humans, which will take weeks. One of the main issues for this research was that the data pool was small compared to other research. This means that the CNN algorithm did not have enough data to train and diversify the result, which may lead to data overfitting. Compared to the other researchers using CNNs to categorize Alzheimer's, the accuracy was lower, which suggests that it is crucial to have an extensive dataset for a more accurate and better categorizing algorithm.

In 2018, a more complicated CNN was used to classify MRI scans into three categories using data collected from the ADNI database. Jain et al. (2019) implemented transfer learning with VGG16 CNN architecture and pre-trained on ImageNet. VGG16, a deep convolutional network with 16 weight layers, achieves high classification accuracy through its deep architecture, which includes 13 convolutional layers with short 3×33 \times 33×3 filters and three fully connected layers. Its performance on the ImageNet dataset is notable, with a top-1 validation error rate of 24.8% and a top-5 validation error rate of 7.5% when using multi-scale and multi-crop evaluations. VGG16 secured 2nd place in the ILSVRC-2014 competition, outperforming previous models and demonstrating that increased depth and practical scale jittering significantly enhance classification accuracy.

After the 150 subjects and 3,840 MRI slices from the ADNI dataset, the research went through three-way classification and three binary classifiers like Westman et al. (2012). The accuracy for the three-way classification, Alzheimer vs. control, Alzheimer vs. MCI, and control vs. MCI were 95.73%, 99.14%, 99.30%, and 99.22%. This algorithm generally had higher accuracy throughout all classifiers compared to past research, which showed the positive progression of CNN and Alzheimer's disease detection research over the years.

Research into machine learning for diagnosing Alzheimer's disease (AD) has evolved significantly from traditional methods to advanced computational techniques. Early studies, such as Kloppel et al. (2008), utilized support vector machines (SVMs) for automating MRI scan analysis, achieving an accuracy of 81% across various patient groups. This was refined by Hinrichs et al. (2009), who introduced LP boosting to enhance spatial analysis of MRI scans, achieving 82% accuracy with an ROC AUC of 0.88. Kohannim et al. (2010) further expanded the field by integrating multiple biomarkers, including PET scans and cerebrospinal fluid (CSF) measures, reaching an accuracy of 82.21% for AD classification. Westman and colleagues later employed orthogonal partial least squares (OPLS) to combine MRI and CSF data, achieving 91.8% accuracy in distinguishing AD from controls. Payan and Montana (2015) advanced this work by using convolutional neural networks (CNNs) to capture subtle brain changes, with accuracies up to 95.39%. More recent studies by Islam and Zhang (2017) have employed CNNs and transfer learning to improve categorization and prediction accuracy, with some algorithms achieving up to 99.30%. This progression highlights the increasing sophistication and effectiveness of machine learning techniques in early AD detection and diagnosis, paving the way for ongoing advancements in the field.

Recent Research

The emergence of researchers using machine learning to diagnose and distinguish Alzheimer's disease from normal and other brain disease symptoms shed new light on the possibility of a new early diagnosis system. The problem now lies in improving, perfecting, and making a more complete algorithm to diagnose Alzheimer's using MRI scans. My researchers investigated the process of improvement, especially in recent years.

Ganesh and Vanamu (2022) sought to improve the CNN method to detect Alzheimer's in 2022. Their research utilized data from ADNI and OASIS, which have been used multiple times in the past throughout research on Alzheimer's detection and machine learning. All the images were resized to a consistent resolution of 224X224 pixels. OASIS images were T1-weighted images, while ADNI were T2-weighted images. These images were put into three types of CNN, which were Resnet-50, VGG-16, and DenseNet-169, where DenseNet-169 has the most layers, and VGG-6 has the least.

The set of images collected from OASIS had four types of images to categorize. They were "very mild demented," "mild demented," "moderate demented," and "non-demented." The dataset was split into three sets, where 64% of the dataset(4098 images) was for training, 16% for validation, and 20% for testing. The images were then tested in the three different algorithms, and their AUC values were checked. VGG-16 had an AUC of 0.9046, Resnet-50 had an AUC of 0.8334, and DenseNet-169 had an AUC of 0.8898. The images collected from ADNI had five categories, one more than the OASIS set. The five categories were early mild cognitive impairment (EMCI), mild cognitive impairment (MCI), late mild cognitive impairment (LMCI), Alzheimer's disease, and control. The dataset was split into 80% training (1035 images), 10% validation, and 10% for training. Again, the data was put through the three algorithms where the

AUC values were 0.7383, 0.7332, and 0.7113 for VGG-15, Resent-50, and DenseNet-169, respectively. For both datasets, VGG-15 outperformed the other two algorithms. This might suggest that there was overfitting for Resent-50 and DenseNet-169. The ADNI had a worse AUC overall than OASIS data as OASIS had more data to train the system, pointing out the importance of data amount. The researchers also observed that the VGG-15 and DenseNet-169 algorithms improved as more data and epoch increased, while Resnet-50 did not show this improvement, which might suggest that there are better algorithms than the Resnet-50 for this instance. This research will aid the improvement of understanding machine learning research on Alzheimer's detection and will also help our research as we use the same datasets.

Another research tested the accuracy of CNN compared to other machine learning techniques when diagnosing Alzheimer's disease. Ahmadi et al. (2024) went back to the basics to see which machine learning technique best categorizes MRI scans. Ahmadi utilized a dataset from Kaggle that consisted of 5,000 images divided into four severity classes: moderate demented, mild demented, very mild demented, and non-demented, which are the same categories as the OASIS dataset. This dataset went through 12 feature extractions, which included GLCM, LBP, RLBP, LTP, SGM, BIBIS, PCA Filter, ICA Filter, Gabor Filter, Log-Energy, Model-based Feature, and conventional shape signature. These analyzed various aspects of image texture, structure, and patterns, including spatial relationships, local intensity variations, and frequency components, to capture and represent essential features for diagnosing Alzheimer's severity. Feature reduction was used to simplify the dataset by reducing the number of features or dimensions while retaining the most critical information.

After all the processing, the images were put into six traditional machine algorithm methods and CNN. These algorithms were compared with accuracy and AUC statistics. The

accuracy was 67.5% for Naïve Bayes, 72.3% for SVM, 74.5% for KNN, 65.6% for LDA, 62.4% for DT, 73.8% for Ensemble Learning, and 95.3% for the CNN algorithm. The AUC was 87% for Naïve Bayes, 89% for SVM, 92% for KNN, 87% for LDA, 82% for DT, 93% for Ensemble Learning, and 99% for CNN. For both statistics, we can see that CNN was superior, which backs up our decision to use CNN for our research and test the paper theory on different datasets.

One of the most recent studies by Ali et al. (2024) and his team used the Kaggle dataset used by Ahmadi and ADNI to test CNN and its capabilities to diagnose Alzheimer's disease again. In this study, Ahmadi used a multistage CNN framework for Alzheimer's disease and also subclassified the images. After preprocessing, Ahmadi used a 26-layer CNN model to classify MRI scans into two categories in Stage 1: healthy and dementia. These layers included convolutional layering, batch normalization, rectified linear unit activation, max pooling, fully connected layers, dropout layer, softmax function, and cross-entropy loss. The second stage was transfer learning from the previous stage. Stage 1 was fine-tuned, which involved the final three layers of the original CNN model to adapt for subclassifying dementia into three categories: mild, moderate, and very mild dementia.

After running the Developed 26-layer CNN, the results were compared to pre-trained CNN models using training accuracy, training loss, validation accuracy, validation loss, and training time. All of the CNN models had 100% training accuracy. The training losses were all very similar as well. On the other hand, the developed CNN had the highest validation accuracy of 97.45% and had a sufficient difference from the second-highest algorithm(DenseNet-201) with 93.93% accuracy. The training time was also the shortest at 8 minutes 53 seconds, which was significantly lower than GoogleNet's 40 min 30 seconds. These values showed that the efficiency and accuracy of the newly suggested CNN model were far superior to those of the pre-

trained CNN models. The CNN model for subclassification also showed promising numbers, with the validation accuracy at 99.70%, only missing one classification out of more than 500 images. This research showed that CNN is steadily becoming a possible detection and classification method for Alzheimer's following more research.

Contrary to the previous research, another research by Suchitra et al. (2024) used a pre-trained CNN architecture to push the research of early Alzheimer's disease detection to a more efficient and helpful direction. The pre-trained CNN algorithm is named EfficientNetB7 and offers high accuracy, low computational complexity, and efficient training. Like a lot of the other research, the ADNI dataset was used. The MRI images were processed, formatted, and divided into training and testing sets. The training sets used 80% of the images, while the testing set used 20%. Next, the EfficientNetB7 was run through EfficientNetB7, which consisted of convolutional layers, ReLU and pooling layers, and a fully connected layer. The convolutional layer was utilized to extract features from the init images. The ReLU layer was used to introduce non-linearity, and the pooling layers were used to reduce dimensionality.

The outcome was measured with four metrics: Sensitivity, Specificity, Accuracy, and F-score. The accuracy of the research came out to be 98.68%, which is high compared to past research. The sensitivity was at 98.08%, which shows that the model effectively detects true positive cases. The specificity was 98.20%, which showed that the model was proficient at avoiding false positives. Finally, the F-score was at 98.95%, which showed that both false positives and false negatives were minimal. These statistics illustrate that it is possible to create a pre-trained CNN to create an algorithm that helps early detection systems. They showed endless possibilities for machine learning and Alzheimer's detection.

Although CNN seems to be the best option to pursue from this past and recent research, some researchers are still pursuing new algorithms that might surpass CNN's efficiency and accuracy in diagnosing Alzheimer's disease. One research study by Irfan et al. (2023) sought to see if a method other than CNN would work better. To answer this curiosity, Irfan suggested a novel ensemble machine-learning technique. First, the vital cognitive features are selected from the ADNI dataset using the Neighborhood Component Analysis and Correlation-based Filtration (NCA-F). This suggested NCA-F is a novel feature selection technique that incorporates all the classifier's advantages for various types of disease detection. Then, the selected features NCA-F will be trained in seven machine learning techniques, which include AdaBoost, artificial neural network (ANN), SVM, and NB (Naõve Bayes). Out of the seven, five top-performing classifiers will be selected and will be given a weight that is based on their performance. The final prediction will then use an adaptive voting mechanism, where the predictions of these weighted classifiers are aggregated to enhance the overall detection accuracy and robustness of the model. This approach leverages the strengths of each classifier while mitigating individual weaknesses, leading to improved performance in early Alzheimer's disease detection.

To compare all the possible algorithms, accuracy, precision, recall, F1, and error were reported in percent. Seven traditional machine learning and adaptive voting were compared. In all statistics, adaptive voting had the best results when all features were used to detect Alzheimer's. It was significantly better as adaptive voting was the only algorithm with 90% or more accuracy, precision, recall, and F1, which showed that it was superior to traditional methods. When done with selected features from NCA-F, the adaptive voting method was still better in every category compared to the traditional method, even though a general improvement was seen with all algorithms.

Research by Khosroazad et al. (2023) has explored the use of sleep disorders as an early indicator of Mild Cognitive Impairment (MCI). Early detection of MCI might be feasible using data from an advanced sleep monitoring device called SleepMove. This device recorded body movements and respiratory signals at a frequency of 16 Hz over two nights. This data developed a new diagnostic parameter, Time-Lag (TL), to quantify the temporal relationship between high-frequency movements and corresponding respiratory changes. To evaluate the utility of TL in distinguishing between MCI and normal control subjects, the team applied two machine learning techniques: Kernel Density Estimation (KDE) and Neural Networks (NN). The KDE analysis demonstrated that TL is a strong indicator of MCI, while the NN results showed TL's effectiveness with high sensitivity (86.75%), specificity (89.25%), and accuracy (88%). This approach may offer a promising alternative to traditional diagnostic methods like MRI scans and CNNs for early MCI detection and Alzheimer's prevention.

Another research looked into diagnosing MCI using ADNI datasets and Dual-Fusion Graph Convolutional Networks (DFCGCN) (Meng & Zhang, 2023). This research used functional MRI and diffusion tensor imaging to capture complex brain connectivity patterns. DFCGCN is a machine learning technique that is well-suited to handling graph-structured data. By employing the graph convolutional networks and advanced graph pooling techniques, the model achieves an impressive accuracy of 90.7%, demonstrating superior performance compared to traditional machine learning methods such as Multilayer Perceptron, Random Forest, and SVM. Integrating multi-modal data and innovative feature extraction strategies allows the DFCGCN to effectively preserve and utilize spatial features, resulting in enhanced classification of MCI. The study carefully tests the DFCGCN model using 10-fold cross-validation and

compares it with other methods. The results show that DFCGCN performs better, with higher sensitivity and specificity, in diagnosing Mild Cognitive Impairment (MCI). The model better captures complex brain activity patterns by using graph kernels and non-imaging data. Overall, the research demonstrates that DFCGCN is a powerful tool for improving brain imaging diagnostics and accurately detecting MCI, which may open up new opportunities for the early detection of Alzheimer's disease using machine learning techniques other than CNN.

Another research conducted by Mitra and Rehman (2024) took a novel approach to early Alzheimer's disease (AD) detection by integrating machine learning with handwriting analysis. Utilizing the Diagnosis AlzheimeR WIth haNdwriting (DARWIN) dataset, which comprises data from 174 individuals, the study applied a sophisticated stacking ensemble method to enhance prediction accuracy. This method structured the machine learning models into a two-tiered system: base-level classifiers and meta-classifiers. The base-level classifiers included Extra Trees, Random Forest, XGBoost, Gaussian Naive Bayes, Multilayer Perceptron, Logistic Regression, and Support Vector Machine. These models were trained on the dataset to identify patterns indicative of AD. The meta-classifier then aggregated the predictions from the base-level classifiers to form the final diagnosis.

The methodology also incorporated advanced feature selection techniques such as Recursive Feature Elimination (RFE) and Analysis of Variance (ANOVA) to identify the most relevant features for each classifier. This rigorous approach led to the ensemble model achieving impressive performance metrics, including 97.14% accuracy, 95% sensitivity, and 100% specificity. The study's findings highlight the efficacy of combining multiple machine learning models and employing feature selection techniques to improve predictive performance.

A paper by Rajasree and Brintha Rajakumari (2023) introduced a novel approach to Alzheimer's disease detection by integrating advanced optimization and deep learning techniques. The methodology encompasses four phases: preprocessing, feature extraction, feature selection, and disease classification. The data consisted of MRI scans, which were cleaned and normalized for preprocessing. Feature extraction involved obtaining higher-order statistical features, weighted modified correlation-based features, and statistical features using WGM-PCA. Feature selection was performed using the I-AR method to identify the most relevant features. In the classification phase, an ensemble model combining Bi-GRU, MLP, and QDNN was used, where Bi-GRU, optimized with inputs from MLP and QDNN, made the final classification decision.

The proposed model, EMOAOA+EC, demonstrated significant improvements over traditional methods. It achieved notably lower false negative rates (FNR) and false favorable rates (FPR) compared to AOA+EC, PRO+EC, and PSO+SVM models. At the 90th learning rate, EMOAOA+EC recorded the lowest FNR and superior overall accuracy, highlighting its effectiveness. The study emphasizes the value of early detection of Alzheimer's disease. It suggests future development of a web-based diagnostic interface utilizing technologies like ONNX and TensorFlow.js for real-time, browser-based predictions. This advancement aims to make the predictive model more accessible and practical for early Alzheimer's diagnosis.

Other than MRI scans, EEGs are a great way to monitor brain activity. Electroencephalography, or EEG, is a non-invasive way to collect signals corresponding to electrical activity from the brain with electrodes placed on the scalp. There are differences between the signals of an average person and an Alzheimer's patient. Using data from two distinct datasets, Jia et al. (2023) proposed a new data augmentation method by decomposing

EEG signals into intrinsic mode functions(IMFs), recombining these IMFs across trials, and generating artificial trials to enhance the training dataset. By doing so, they hoped to improve the performance of machine learning techniques on smaller datasets to detect Alzheimer's.

Jia tested to see if the data augmentation improved the classification accuracy using the neural network classifiers ResNet, BrainNet CNN, and EEGNet.The results demonstrated significant improvements in classification accuracy, particularly for ResNet and BrainNet CNN. ResNet's accuracy increased by 5.24% for mild AD, and BrainNet CNN's accuracy improved by 2.38%. On the MCI dataset, ResNet achieved a 4.50% gain, while BrainNet CNN saw a 0.75% increase. These improvements underscore the system's effectiveness in enhancing model performance. Despite these gains, the paper highlights the need for further research to fine-tune the system, including determining the optimal number of artificial trials and training epochs to avoid overfitting and achieve the best results in early Alzheimer's diagnosis. This research done by Hao may hint that data augmentation possibly improves CNN and Alzheimer's detection using MRI scans.

Recent research in Alzheimer's disease (AD) detection highlights significant advancements using machine learning techniques. Studies by Ganesh and Vanamu (2022) and Ahmadi et al. (2024) demonstrated that Convolutional Neural Networks (CNNs), particularly VGG-16, outperform other algorithms in MRI scan classification with superior accuracy and AUC metrics. Ali et al. (2024) further improved CNN models, achieving high validation accuracy and efficiency, while Suchitra et al. (2024) used the EfficientNetB7 CNN model, which yielded exceptional early AD detection results with impressive accuracy and low computational complexity. A novel approach by integrating advanced optimization and deep learning techniques, exemplified by the EMOAOA+EC model, utilized MRI data with preprocessing,

feature extraction, selection, and classification phases, showing significant improvements in false negative and false favorable rates. Additionally, Irfan et al. (2023)'s ensemble method and Somayeh Khosroazad's use of sleep disorder analysis for detecting Mild Cognitive Impairment (MCI) demonstrate promising alternatives. Jia et al. (2023)'s research on data augmentation for EEG signals significantly improved classification accuracy for ResNet and BrainNet CNN. These studies reflect the progress in machine learning for AD diagnosis and underscore the need for continued research into dataset diversity, multimodal data integration, and innovative diagnostic tools.

Identified Gaps

Though significant advancements in Alzheimer's disease detection through machine learning and novel methodologies have been made, several gaps and opportunities for further research remain. Addressing these areas could enhance current methods' diagnostic accuracy, efficiency, and applicability. One gap is with the diversity and size of the data.

In most research done by researchers using MRI scans, the dataset comes from ADNI, Kaggle, or OASIS. These researchers only use one dataset and can only collect analysis from the one dataset they use. For example, Ganesh and Vanamu (2022) utilized ADNI and OASIS datasets. Even though it is more diverse to use two different datasets, they did not combine the datasets and instead trained them separately. If it is possible to combine datasets and train them, it is possible to diversify the algorithm and observe Alzheimer's disease detection in a more vast area. Another example from Ahmadi et al. (2024) utilized a dataset from Kaggle, which is a minimal pool of samples and will only reflect some of the population. In our thesis, we can close that gap by combining datasets to create a more diverse scope of subjects.

Another area for improvement is the different types of data being used. Usually, in past research, researchers used a single or a limited combination of data. For example, Khosroazad et al. (2023) study explored sleep disorders and their relationship with MCI but only used EGG. On the other hand, Meng and Zhang (2023) used both MRI and diffusion tensor imaging, demonstrating the advantages of combining different imaging modalities for enhanced diagnostic performances. Research could explore the integration of multiple modalities. Combining imaging data with physiological signals may enhance diagnostic results and provide a more comprehensive understanding of Alzheimer's disease.

A third issue that can be looked into is the transparency of the "black box." When using methods like CNNs, there are criticisms that the decision-making done to predict by the algorithm is not transparent. Additionally, feature selection techniques vary widely, and their effectiveness in different contexts is only sometimes apparent. Ali et al. (2024) emphasized the development of a 26-layer CNN model, which reported a high validation accuracy. However, the transparency was limited in how these models made their decisions, which can be a significant barrier to their adoption in clinical settings as it is crucial to understand the rationale behind a diagnosis. Addressing the gap requires a focus on developing more interpretable models where Mitra and Rehman (2024) are a good example. Their stacking ensemble method incorporated ANOVA and Recursive Feature Elimination, which made it possible to create more interpretability of machine learning models while maintaining high performance. Further research in this area could focus on improving selection methods and exploring ways to make the decision-making process transparent for clinicians by CNNs.

While CNNs have shown significant promise in Alzheimer's detection, other algorithms and ensemble methods also offer potential benefits. However, comprehensive studies comparing the efficacy of these different approaches are limited. Ahmadi et al. (2024) compared CNNs with traditional machine learning algorithms like Naïve Bayes, SVM, and KNN, highlighting the need for more thorough comparisons to identify the most effective techniques for AD detection. Irfan et al.'s (2023) study, which explored a novel ensemble machine-learning technique, further underscores this point. His research demonstrated the effectiveness of this approach compared to traditional methods, suggesting that there is still much to learn about optimizing algorithms for Alzheimer's detection. Future research could focus on conducting more in-depth comparative studies, exploring newer algorithms, and developing hybrid models that combine the strengths of different approaches. This could lead to more efficient and accurate diagnostic tools.

Many studies achieve impressive results in controlled environments, but they often need to address the practical challenges of implementing these models in real-world clinical settings. For instance, Rajasree and Brintha Rajakumari's (2023) study proposed an ensemble model that significantly improved diagnostic accuracy. Still, their research needed to delve into the challenges of scaling this model for widespread clinical use. Similarly, Jia et al. (2023) research on data augmentation for EEG signals focused on improving model performance but needed to fully consider the computational demands of applying such techniques in real-time clinical scenarios. To bridge this gap, future research should prioritize developing models that are accurate, computationally efficient, and easy to integrate into existing clinical workflows. This includes addressing challenges related to computational resource limitations, scalability, and the creation of user-friendly interfaces. Research could also explore developing cloud-based or edge-

computing solutions that allow for real-time analysis and diagnosis in diverse healthcare settings, ensuring that these advanced models can be practically applied in routine clinical practice.

The complexity of Alzheimer's disease detection necessitates an interdisciplinary approach, yet many studies are conducted within single fields of expertise. For instance, developing advanced fCNN models, such as those by Ali et al. (2024), could greatly benefit from collaboration between data scientists and clinicians. Such interdisciplinary efforts could help tailor these models to meet the specific needs of clinical practice. Encouraging collaboration between computer scientists, clinicians, neuroscientists, and other experts could lead to more comprehensive and effective solutions for Alzheimer's diagnosis and treatment. Rajasree and Brintha Rajakumari's (2023) study, which focused on optimization techniques in deep learning models, also illustrates the potential benefits of interdisciplinary collaboration. By combining insights from data science, clinical neurology, and other fields, researchers can address the multifaceted challenges of Alzheimer's disease detection more effectively. This approach could lead to scientifically rigorous and practically applicable innovations in diverse healthcare settings.

Summary

Chapter 2 of this thesis has provided a comprehensive review of the existing literature on using convolutional neural networks (CNNs) and magnetic resonance imaging (MRI) for early detection of Alzheimer's Disease. The review began with an exploration of early research, highlighting the foundational discoveries in Alzheimer's diagnosis and the initial applications of diagnostic imaging techniques. The progression of research was traced to recent advancements, with a particular focus on integrating machine learning methods, including CNNs, to enhance diagnostic accuracy.

Significant developments have been made, especially in applying CNNs for analyzing MRI scans. These advancements offer promising avenues for improving diagnostic accuracy and efficiency. However, the review also identified critical gaps in the current research landscape. Issues such as datasets' limited diversity and size, the challenges of integrating multiple data modalities, and the opacity of machine learning models were discussed. Additionally, the practical challenges of deploying these models in real-world clinical settings and the need for interdisciplinary collaboration were highlighted as areas requiring further exploration.

The insights gathered in this chapter underscore the importance of addressing these gaps to advance Alzheimer's disease detection. Future research can develop more robust, accurate, and clinically applicable diagnostic tools by combining diverse datasets, enhancing model transparency, and fostering interdisciplinary collaboration. This chapter sets the stage for the subsequent thesis sections, which will delve into the methodology and experimental approach designed to tackle these identified gaps.

Chapter 3: Methodology

Gathering significant data comes with using appropriate collection to be able to uphold or reject our research hypotheses. In Ch.2, the progression of technology has been revolutionary with early detection of Alzheimer's. In the beginning, data that was used for detection mainly relied on MRI scans and Diagnostic and Statistical Manual. The objective seeks to determine the application of CNN-based diagnostic tools in medical settings. This will provide insight into how machine learning has advanced in healthcare and the diagnosis of Alzheimer's. Let's investigate how Team CNN collected their data for the analysis question.

Description of Subjects

The Oasis datasets are from Washington University Medical School where they gathered data from a cross-sectional collection of 416 subjects aged 18 to 96 that had participated in MRI studies at the university (Marcus et al., 2007). Out of the 416 subjects, one hundred of the included subjects older than 60 years have been clinically diagnosed with very mild to moderate Alzheimer's. For older adults, subjects had to fit at least three acquired T1-weighted images, right hand dominance, and a recent clinical evaluation. The young and middle-aged subjects were recruited from the Washington University community.

Kaggle is a popular database that hosts competitions, tutorials, and datasets that are publicly available for people to use. The dataset we are using is an Alzheimer's images that consist of two files: training and testing. The data includes 14,609 MRI images segregated into: Mild Demented, Very Mild Demented, Moderate Demented, and Non-Demented.

Data Models

For our research questions, we will be utilizing convolutional neural networks and support vector machines which are algorithms used in deep learning. Recent research studies for Alzheimer's detection have involved CNN and SVM models which had promising outcomes; CNNs are effective for learning complex patterns in image data but it does require large datasets. SVMs are robust for classification in high-dimensional spaces but it does struggle with analyzing very large datasets. We will be primarily using CNN to classify the stages of Alzheimer's and proceed to compare it against a SVM model.

For model training and evaluation, we will be optimizing CNNs for feature extraction from the MRI images. From there the models will be evaluated using metrics such as: precision, recall, accuracy and confusion matrices to assess the reliability of the models. Accuracy will provide us with a metric to determine which ML model is working and usable. Confusion matrix is a detailed view into accuracy

Description of Research Instrument

Oasis-1 used The Clinical Dementia Rating is a 5-point scale used to characterize six domains of cognitive and functional performance pertaining to Alzheimer disease. The six domains include: memory, orientation, judgment and problem solving, function in community affairs, home and hobbies, and personal care. A global CDR of 0 indicates no dementia, and CDRs 0.5, 1, 2, and 3 represent very mild, mild, moderate, and severe dementia, respectively. ADNI also used the CDR 5-point scale to categorize each subject into three groups: Normal, MCI, and Mild AD. The subjects underwent a clinical/cognitive assessment which asked questions on

obtaining consent, medical history, activities of daily living, and other questions to gain more information on each subject.

Procedures

Data was gathered from the Alzheimer's Disease Neuroimaging Initiative, Kaggle, and Washington University in St. Louis. For each dataset, I will be reviewing how the data was collected and other statistical analysis that was conducted.

The Open Access Series of Imaging Studies (OASIS-1) is a project that is conducted at Washington University Alzheimer Disease Research Center that provides publicly available data sets for study and analysis. For older adults, subjects had to fit at least three acquired T1-weighted images, right hand dominance, and a recent clinical evaluation. The young and middle-aged subjects were recruited from the Washington University community. Each subject was screened for inclusion, with young and middle-aged adults being questioned before the image acquisition about their medical histories and use of psychoactive drugs. The older subjects went through the full ADRC clinical assessment, which is based off the CDR scale. To analyze the dataset, Oasis-1 researchers used mean absolute percent differences, effect size, and ANOVA.

Kaggle is a data science competition platform and online community that encourages data enthusiasts to connect and enhance their skills using their public datasets. The Alzheimer's Dataset was hand collected from various websites with every label being verified. The data consists of MRI images which are split into four classes both in training and testing which was mentioned earlier in our subject descriptions; there are a total of 14,609 images in this dataset.

In terms of how Team CNN will be analyzing our data, I will go over our pre-processing steps for data cleaning, reduction, and transformation. First, we will be removing any unnecessary artifacts that will discard irrelevant information from the scans. Normalizing the data will make

sure that all our images have consistent intensity levels which will help differentiate the technology over time. Our data must be augmented to increase diversity, robustness, and standardizing the images will ensure consistency in the dimensions for the neural networks. For model training and evaluation, we will be optimizing CNNs for feature extraction from the MRI images. From there the models will be evaluated using metrics such as accuracy and confusion matrices to assess the reliability of our models. With our target variable pertaining to classification, we will use the MRI scans and place them into various stages of Alzheimer's from: very mild, mild, moderate, to no Alzheimer's. Our datasets come from different sources which compiled over 14,609 images, these images are 2D and 3D axial slices in various file formats that we'll be converting in order to work collectively in Google Colab throughout the process.

Summary

Our team is dedicated to collecting data that is robust and thorough so that we can do a reproducible analysis and find prediction models that can assist with early detection of Alzheimer's. By using SVM and CNN methods from deep learning, we'll be able to determine the accuracy and confusion matrix for each model.

Chapter 4: Results and Analysis

This chapter presents and analyzes the results of the research. The primary objective is to evaluate the ability of Convolutional Neural Networks (CNNs) in diagnosing Alzheimer's Disease from MRI scans and compare these findings with those obtained from Support Vector Machines (SVMs). This chapter is structured to align with the research hypotheses, providing an objective analysis supported by visualizations and statistical data.

Descriptive Statistics

This chapter presents and analyzes the data and results obtained from the study. The dataset used for the analysis includes a total of 14,609 images, each with a resolution of 128x128 pixels and three color channels. These images are categorized into four distinct groups: Very Mild Demented, Moderate Demented, Mild Demented, and Non Demented. The distribution of images across these categories is relatively balanced, which is important for ensuring that the models are trained and evaluated on a fair representation of each class. Specifically, the dataset contains 4,792 images labeled as Very Mild Demented, 540 images labeled as Moderate Demented, 3,717 images labeled as Mild Demented, and 5,560 images labeled as Non Demented. Understanding the distribution of images across these labels is crucial for interpreting the results and assessing the models' performances.

Figure 2: Distribution of Images

Label	Label Name	Count of Images
0	Very Mild Demented	4792
1	Moderate Demented	540
2	Mild Demented	3717
3	Non Demented	5560

Hypothesis Testing

CNN Model Performance

The Convolutional Neural Network (CNN) was trained and evaluated using the combined dataset. The CNN architecture consisted of several convolutional layers followed by pooling layers, with dense layers and dropout added for regularization. The training process was run for a maximum of 50 epochs, with early stopping based on validation loss to prevent overfitting. The results of the CNN model showed that it achieved a peak training accuracy of 99.02%, with a validation accuracy of 99.28%. The training and validation loss curves in Figures 2 and 3 demonstrate that the model converged well, with minimal signs of overfitting.

Figure 3: Loss Curve

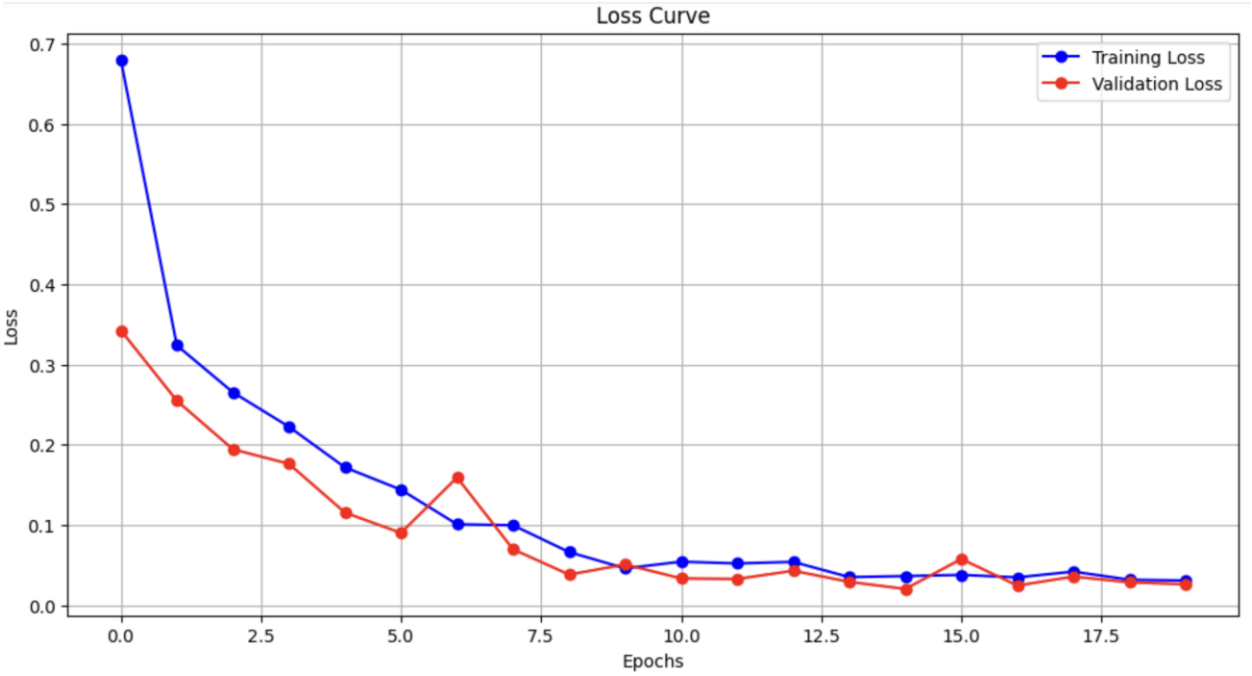
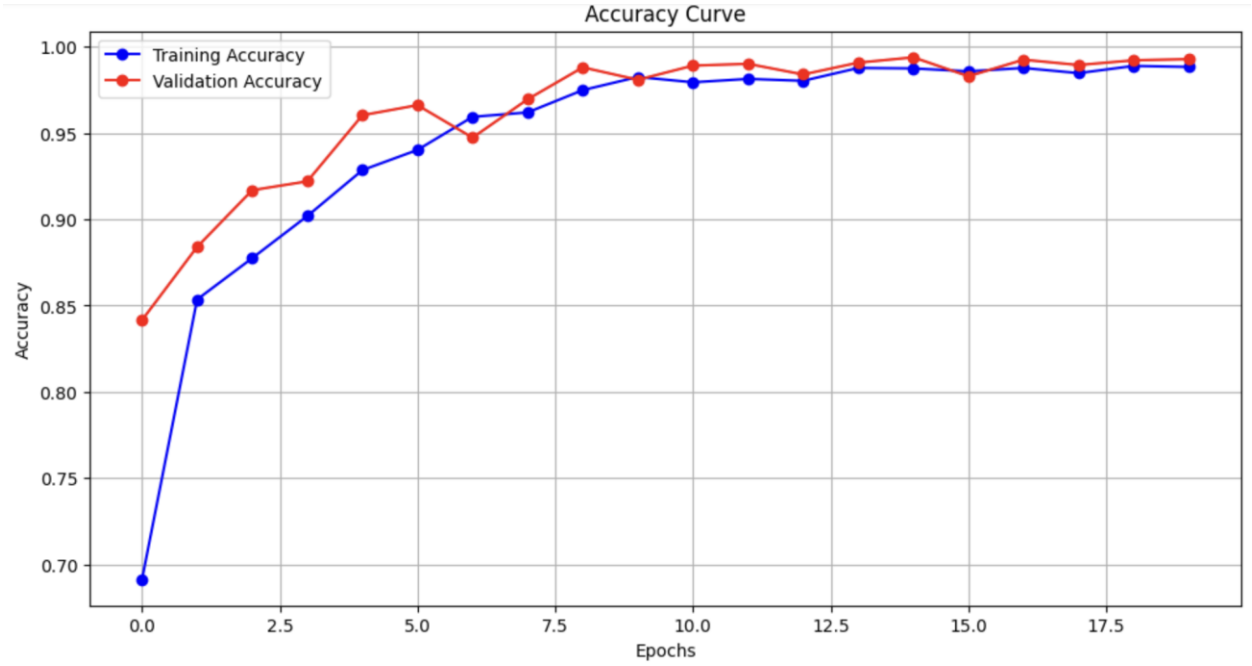


Figure 4: Accuracy Curve



The CNN model's classification report in Figure 4 further highlights its effectiveness in classifying the images into their respective categories. For the Very Mild Demented class, the model achieved a precision of 0.99, a recall of 0.99, and an F1-score of 0.99. The Moderate Demented class had perfect scores across all metrics, with a precision, recall, and F1-score of 1.00. Similarly, the Mild Demented class showed strong performance, with a precision of 1.00, a recall of 0.99, and an F1-score of 1.00. Finally, for the Non Demented class, the model attained a precision, recall, and F1-score of 0.99. Overall, the CNN model performed exceptionally well across all categories, as evidenced by the high precision, recall, and F1-scores for each class. These metrics suggest that CNNs could offer a reliable alternative to more traditional diagnostic tools.

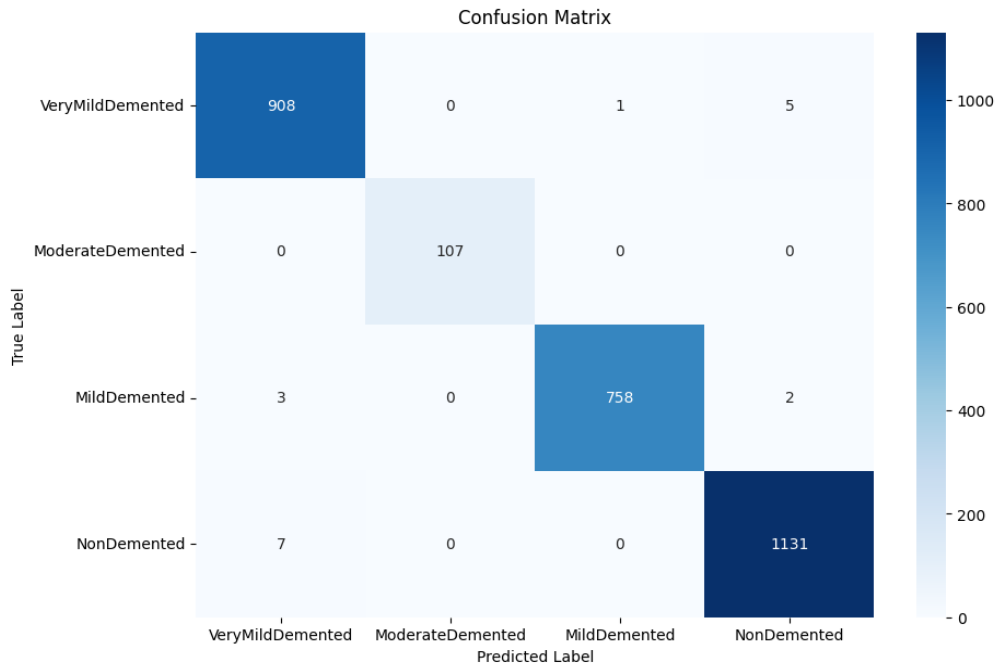
Figure 5: CNN Model Classification Report On Test Dataset

Classification Report:

	precision	recall	f1-score	support
VeryMildDemented	0.99	0.99	0.99	914
ModerateDemented	1.00	1.00	1.00	107
MildDemented	1.00	0.99	1.00	763
NonDemented	0.99	0.99	0.99	1138
accuracy			0.99	2922
macro avg	1.00	1.00	1.00	2922
weighted avg	0.99	0.99	0.99	2922

Additionally, the confusion matrix for the CNN model, shown in Figure 5, visualizes the true and false classifications across the four classes, offering a clearer understanding of the model's performance.

Figure 6: CNN Model Confusion Matrix



The confusion matrix in Figure 6 offers a visual representation of the model's true and false classifications across the four diagnostic categories. This is a critical tool for understanding where the model might be making errors and how well it distinguishes between similar stages of dementia. A well-performing model should have a high number of true positives (on the diagonal), which our CNN model achieves.

SVM Model Performance

In addition to the CNN, a Support Vector Machine (SVM) with a radial basis function (RBF) kernel was evaluated using the same dataset. The SVM model also performed well, though slightly less so than the CNN. The SVM classification report revealed that for the Very Mild Demented class, the model achieved a precision of 0.97, a recall of 0.96, and an F1-score of 0.97. For the Moderate Demented class, the SVM performed very well, with a precision of 1.00,

a recall of 0.99, and an F1-score of 0.99. The Mild Demented class saw equally high results, with a precision of 1.00, a recall of 0.96, and an F1-score of 0.98. Lastly, for the Non Demented class, the model attained a precision of 0.96, a recall of 1.00, and an F1-score of 0.98. The overall accuracy of the SVM model was 97%, and its classification report and confusion matrix provide a more detailed view of how the model performed on each class.

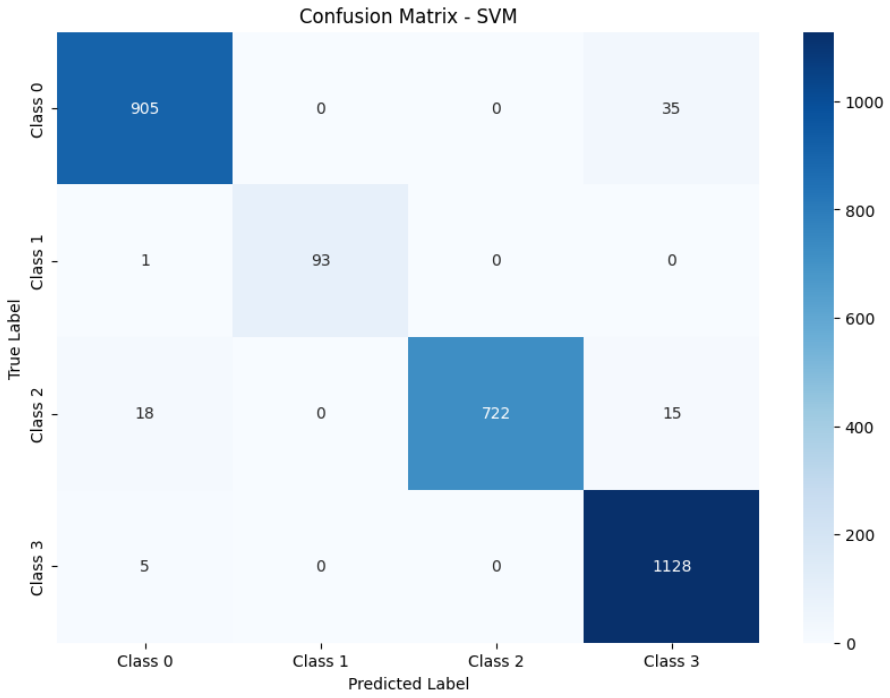
Figure 7: SVM Model Classification Report

Classification Report:

	precision	recall	f1-score	support
Class 0	0.97	0.96	0.97	940
Class 1	1.00	0.99	0.99	94
Class 2	1.00	0.96	0.98	755
Class 3	0.96	1.00	0.98	1133
accuracy			0.97	2922
macro avg	0.98	0.98	0.98	2922
weighted avg	0.98	0.97	0.97	2922

The confusion matrix for the SVM model is depicted in Figure 7. This visualization helps to compare the SVM's performance with that of the CNN by illustrating the distribution of true and false classifications.

Figure 8: SVM Model Confusion Matrix



Comparative Analysis

When comparing the CNN and SVM models, it is clear that the CNN outperformed the SVM in most performance metrics. The CNN achieved an overall accuracy of 99%, whereas the SVM's accuracy was slightly lower at 97%. The CNN also showed higher precision and recall across all categories, making it a more effective model for classifying the different stages of dementia. While both models demonstrated strong results, the CNN consistently provided superior performance, particularly in its ability to generalize across all four categories. This can

likely be attributed to CNN's superior ability to capture complex image patterns due to its deep architecture, whereas SVMs, though robust, may struggle with large image datasets that require high-dimensional analysis. Furthermore, the CNN was significantly more efficient in computational tasks compared to the SVM. This makes the CNN model the better choice for this specific task of dementia classification based on MRI images.

Summary

In conclusion, both the CNN and SVM models performed well in classifying dementia stages from MRI images, but the CNN showed consistently better results. The CNN model achieved higher accuracy and F1-scores across all categories compared to the SVM, making it a more effective tool for this classification task. The accuracy and loss curves for the CNN, as well as the classification reports and confusion matrices for both models, provide a comprehensive view of the models' performances. These results support the hypothesis that deep learning models, such as the CNN, are better suited for image classification tasks in medical imaging than traditional machine learning models like the SVM. The visualizations, which include the CNN's accuracy curve, loss curve, and confusion matrix, as well as the SVM's confusion matrix, will be presented in the subsequent figures to further illustrate the findings discussed in this chapter.

Chapter 5: Conclusions and Recommendations

In this chapter, the research study will be summarized for researchers and readers for future research. The objectives that were placed include: training and validating the CNN-based tool using MRI scans, comparing the performance of the CNN model to the SVMs, and determining how to add the CNN diagnostic tool into a clinical setting. Our goal was to determine how machine learning can contribute to improving the diagnosis of Alzheimer's using Deep Learning techniques to identify subjects that are at risk of the disease.

Summary of Study

Alzheimer's Disease is the most common type of Dementia, and symptoms of this disease can first appear after the age of 60 (*What Is Alzheimer's Disease? | CDC*, n.d.). This study tested the hypothesis that CNNs would outperform traditional machine learning methods such as SVM in diagnosing Alzheimer's from MRI scans, which was supported by the results. The conducted research underscores the potential of CNNs to enhance early detection of Alzheimer's disease, with the long-term goal of developing automated diagnostic tools that improve patient outcomes and reduce the time clinicians spend on imaging analysis. The data utilized in the study included MRI scans from Oasis-1 from Washington University Medical School and Kaggle, which were combined to create one dataset. Oasis dataset consists of a longitudinal collection of 150 subjects aging from 60-96 years old where each subject was scanned on two or more visits for a total of 373 imaging sessions.

After completing the CNN and SVM models, comparisons between the two were analyzed to discover which one achieved a better accuracy. Both the CNN and SVM models performed well in classification, but CNN showed better results overall. CNNs outperform SVMs because they can learn more complex features of images while SVMs require complex feature engineering and can struggle with larger complex datasets. The Convolutional Neural

Network was trained using the combined data resulting in a training accuracy of 99.02% and a validation accuracy of 99.28%. Accuracy is used to classify the severity of Alzheimer's from the MRI images. For each category of Very Mild Demented, Moderate Demented, and Non-Demented classes each one performed well for high precision, recall, and F1 scores. For example, Very Mild Demented had a precision score of 0.99, a recall of 0.99, and a F1-score of 0.99. Support vector machine classifier works on developing a linear decision to classify multiple categories. The SVM model performed well with an accuracy of 97%. Overall, we conclude that the CNN model is a better choice especially with each class of demented categories.

Limitations

Alzheimer's Disease diagnosis is normally determined by neurological exams, functional assessments, brain images, and cerebrospinal fluid. Brain images include MRI and PET scans, the study that we are focusing on only includes MRI scans. A limitation that was discovered early in the study was the number of images that were acquired and the amount of space that was available in Google Colab. Imaging data has become an advanced tool for early detection, but there have been discussions on the lack of available data. Our data consists of only MRI scans without any type of demographic attributes which makes it difficult to pinpoint the characteristics of the subjects. Due to the size of space needed to create both CNN and SVM models, our data set was reduced to 14,609 images which affected how much we could train and test. With more memory and processing power, we could have trained the models with larger datasets. After combining the datasets and reducing the images to 128x128 pixels there could have been potential loss of features.

Implications for Practitioners

The study's findings offer a great insight for researchers to keep pushing for new opportunities to use Artificial Intelligence for Alzheimer's research. Recently Boston University has been working on a model that can predict whether someone with mild cognitive impairment can remain stable over six years or fall into the early stages of Alzheimer's (Thurston, 2024). The Department of Psychology at the University of Cambridge recently developed a model to predict whether and how fast a person with mild cognitive impairments will progress to develop Alzheimer's. The researchers were able to find that the model they developed correctly identified those who progressed to having Alzheimer's, only using data from cognitive tests and MRI scans that they gathered from a USA research cohort (Lee et al., 2024). Practitioners and researchers should expand their studies into different areas of dementia such as frontotemporal dementia, and use data that includes biomarkers from blood tests. The use of Deep Learning and AI in the medical community continues to make strides each day with new discoveries in drug discovery and diagnostic treatments. The integration of AI-based techniques into clinical practice can encourage future work and greatly enhance diagnostic capabilities and patient care.

Recommendations

There have been many studies and questions on how to combat and cure Alzheimer's Disease, but currently there is still no defined cure. Researchers are now trying to incorporate different AI tools to better predict early onset Alzheimer's, and it has been shown that gender and age play a prominent factor in diagnosing the disease (Basheer et al., 2021). Further studies on Alzheimer's and early detection continue to make strides, but there can be a lot more done with the use of Machine Learning and Artificial Intelligence. With researchers and academic hospitals,

they have a significant amount of tools that they can use to implement AI into medical research and combating Alzheimer's Disease.

Conclusion

The primary goal for this study was to develop an automated diagnostic tool for Alzheimer's using CNN to analyze MRI scans. The initial hypothesis assumed that the CNN-based model would perform higher in identifying the presence of AD compared to clinical analysis of MRIs. We found the Convolutional Neural Networks performed the best with an accuracy of 99.28% compared to our SVM model. We discovered different limitations that can be used to generate better and larger models using more computational space.

References

Leming, M. J., Bron, E. E., Bruffaerts, R., Ou, Y., Iglesias, J. E., Gollub, R. L., & Im, H. (2023). Challenges of implementing computer-aided diagnostic models for neuroimages in a clinical setting. *NPJ Digital Medicine, 6*(129). https://doi.org/10.1038/s41746-023-00868-x

National Institute on Aging. (n.d.). *Alzheimer's Disease Fact Sheet*. U.S. Department of Health and Human Services. Retrieved August 5, 2024, from https://www.nia.nih.gov/health/alzheimers-and-dementia/alzheimers-disease-fact-sheet

Driver, C. & Bowles, Bradley & Bartholmai, Brian & Greenberg-Worisek, Alexandra. (2019). Artificial Intelligence in Radiology: A Call for Thoughtful Application. Clinical and Translational Science. 13. 10.1111/cts.12704.

Geras KJ, Mann RM, Moy L. Artificial Intelligence for Mammography and Digital Breast Tomosynthesis: Current Concepts and Future Perspectives. Radiology. 2019 Nov;293(2):246-259. doi: 10.1148/radiol.2019182627. Epub 2019 Sep 24. PMID: 31549948; PMCID: PMC6822772.

Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks. arXiv preprint arXiv:1502.02506. Retrieved from https://arxiv.org/abs/1502.02506

Alzheimer's Association. (2024). *2024 Alzheimer's Disease Facts and Figures*. Alzheimers Dement, 20(5). Retrieved from https://www.alz.org/media/Documents/alzheimers-facts-and-figures.pdf

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. Retrieved from https://link.springer.com/article/10.1007/BF00994018

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. Retrieved from https://www.nature.com/articles/nature14539

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-Gonzalez, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., & Colliot, O. (2020). Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Medical image analysis*, 63, 101694. Retrieved from https://pubmed.ncbi.nlm.nih.gov/32417716/

Johns Hopkins Medicine. (2018, July 2). *Mid- to late-life increases in marker of chronic inflammation tied to dementia*. https://www.hopkinsmedicine.org/news/newsroom/news-releases/2018/07/mid--to-late-life-increases-in-marker-of-chronic-inflammation-tied-to-dementia

Mayo Clinic Staff. (2023, April 18). Positron emission tomography scan care at Mayo Clinic. Mayo Clinic. https://www.mayoclinic.org/tests-procedures/pet-scan/about/pac-20385078
Centers for Disease Control and Prevention. (2019, April 5). About dementia. Alzheimer's Disease and Healthy Aging. https://www.cdc.gov/aging/dementia/index.html

Wood, T. (n.d.). Convolutional neural network. DeepAI. https://deepai.org/machine-learning-glossary-and-terms/convolutional-neural-network

IBM. (2023, December 27). What are support vector machines (SVMs)? IBM. https://www.ibm.com/topics/support-vector-machine

IBM. (n.d.). What is explainable AI? IBM. https://www.ibm.com/topics/explainable-ai

Nasr Mohammed, A., Albagul, A., & Musbah Ahmad, M. (2023). Automated Alzheimer's Disease diagnosis using Convolutional Neural Networks and Magnetic Resonance Imaging. *WSEAS Transactions on Signal Processing*, 19, 1-9. https://doi.org/10.37394/232014.2023.19.13

Ahmadi, M., Javaheri, D., Khajavi, M., Danesh, K., & Hur, J. (2024). A deeply supervised adaptable neural network for diagnosis and classification of Alzheimer's severity using multitask feature extraction. *PLOS ONE, 19*(3). https://doi.org/10.1371/journal.pone.0297996

Ali, M. U., Kim, K. S., Khalid, M., Farrash, M., Zafar, A., & Lee, S. W. (2024). Enhancing Alzheimer's disease diagnosis and staging: A multistage CNN framework using MRI. *Frontiers in Psychiatry, 15*. https://doi.org/10.3389/fpsyt.2024.1395563

Castellani, R. J., Rolston, R. K., & Smith, M. A. (2010). Alzheimer disease. *Disease-a-Month, 56*(9), 484–546. https://doi.org/10.1016/j.disamonth.2010.06.001

De Leon, M. J., Ferris, S. H., Blau, I., George, A. E., Reisberg, B., Kricheff, I. I., & Gershon, S. (1979). Correlations between computerized tomographic changes and behavioural deficits in senile dementia. *The Lancet, 314*(8147), 859–860. https://doi.org/10.1016/s0140-6736(79)92217-7

De Leon, M. J., McRae, T., Tsai, J. R., George, A. E., Marcus, D. L., Freedman, M., Wolf, A. P., & McEwen, B. (1988). Abnormal cortisol response in Alzheimer's disease linked to hippocampal atrophy. *The Lancet, 332*(8607), 391–392. https://doi.org/10.1016/s0140-6736(88)92855-3

Ganesh, A., & Vanamu, G. (2022). A novel approach for early detection of Alzheimer's disease using deep neural networks with magnetic resonance imaging. *Journal of Emerging Investigators*. https://doi.org/10.59720/21-053

Gómez-Isla, T., Price, J. L., McKeel Jr., D. W., Morris, J. C., Growdon, J. H., & Hyman, B. T. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *The Journal of Neuroscience, 16*(14), 4491–4500. https://doi.org/10.1523/jneurosci.16-14-04491.1996

Hinrichs, C., Singh, V., Mukherjee, L., Xu, G., Chung, M. K., & Johnson, S. C. (2009). Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. *NeuroImage, 48*(1), 138–149. https://doi.org/10.1016/j.neuroimage.2009.05.056

Hippius, H., & Neundörfer, G. (2003). The discovery of alzheimer's disease. *Dialogues in Clinical Neuroscience*, *5*(1), 101–108. https://doi.org/10.31887/dcns.2003.5.1/hhippius

Irfan, M., Shahrestani, S., & Elkhodr, M. (2023). Early detection of Alzheimer's disease using cognitive features: A voting-based ensemble machine learning approach. *IEEE Engineering Management Review, 51*(1), 16–25. https://doi.org/10.1109/emr.2022.3230820

Islam, J., & Zhang, Y. (2017). A novel deep learning-based multi-class classification method for Alzheimer's disease detection using brain MRI data. *Lecture Notes in Computer Science*, 213–222. https://doi.org/10.1007/978-3-319-70772-3_20

Jain, R., Jain, N., Aggarwal, A., & Hemanth, D. J. (2019). Convolutional neural network-based Alzheimer's disease classification from magnetic resonance brain images. *Cognitive Systems Research, 57*, 147–159. https://doi.org/10.1016/j.cogsys.2018.12.015

Jia, H., Huang, Z., Caiafa, C. F., Duan, F., Zhang, Y., Sun, Z., & Solé-Casals, J. (2023). Assessing the potential of data augmentation in EEG functional connectivity for early detection of Alzheimer's disease. *Cognitive Computation, 16*(1), 229–242. https://doi.org/10.1007/s12559-023-10188-7

Khosroazad, S., Abedi, A., & Hayes, M. J. (2023). Sleep signal analysis for early detection of Alzheimer's disease and related dementia (ADRD). *IEEE Journal of Biomedical and Health Informatics, 27*(5), 2264–2275. https://doi.org/10.1109/jbhi.2023.3235391

KIDD, M. (1964). Alzheimer's disease —an electron microscopical study. *Brain*, *87*(2), 307–320. https://doi.org/10.1093/brain/87.2.307

Kloppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., Fox, N. C., Jack, C. R., Ashburner, J., & Frackowiak, R. S. (2008). Automatic classification of MR scans in Alzheimer's disease. *Brain, 131*(3), 681–689. https://doi.org/10.1093/brain/awm319

Kohannim, O., Hua, X., Hibar, D. P., Lee, S., Chou, Y.-Y., Toga, A. W., Jack, C. R., Weiner, M. W., & Thompson, P. M. (2010). Boosting power for clinical trials using classifiers based on multiple biomarkers. *Neurobiology of Aging, 31*(8), 1429–1442. https://doi.org/10.1016/j.neurobiolaging.2010.04.022

Kumar, A., Sidhu, J., Lui, F., & others. (2024). Alzheimer disease. In *StatPearls* (Updated February 12, 2024). StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK499922/

Meng, L., & Zhang, Q. (2023). Research on early diagnosis of Alzheimer's disease based on dual fusion cluster graph convolutional network. *Biomedical Signal Processing and Control, 86*, 105212. https://doi.org/10.1016/j.bspc.2023.105212

Mitra, U., & Rehman, S. U. (2024). ML-powered handwriting analysis for early detection of Alzheimer's disease. *IEEE Access, 12*, 69031–69050. https://doi.org/10.1109/access.2024.3401104

Payan, A., & Montana, G. (2015). Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks. *arXiv*. https://arxiv.org/abs/1502.02506

Rajasree, R., & Brintha Rajakumari, S. (2023). Ensemble-of-classifiers-based approach for early Alzheimer's disease detection. *Multimedia Tools and Applications, 83*(6), 16067–16095. https://doi.org/10.1007/s11042-023-16023-3

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *arXiv*. https://arxiv.org/abs/1409.1556

Suchitra, S., Krishnasamy, L., & Poovaraghan, R. J. (2024). A deep learning-based early Alzheimer's disease detection using magnetic resonance images. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-19677-9

Westman, E., Muehlboeck, J.-S., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage, 62*(1), 229–238. https://doi.org/10.1016/j.neuroimage.2012.04.056

Yang, H. D., Kim, D. H., Lee, S. B., & Young, L. D. (2016). History of Alzheimer's disease. *Dementia and Neurocognitive Disorders, 15*(4), 115–121. https://doi.org/10.12779/dnd.2016.15.4.115

Diogo, V.S., Ferreira, H.A., Prata, D. *et al.* Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alz Res Therapy* 14, 107 (2022). https://doi.org/10.1186/s13195-022-01047-y

Wimo, A., & Jönsson, L. (2013). "The worldwide societal costs of dementia: Estimates for 2010 and 2030." *Alzheimer's & Dementia, 9*(1), 1-11. https://doi.org/10.1016/j.jalz.2012.11.006

Johns Hopkins Medicine. (2018). Mid- to late-life increases in marker of chronic inflammation tied to dementia. *Clinical Connection*. Retrieved from https://clinicalconnection.hopkinsmedicine.org/news/mid-to-late-life-increases-in-marker-of-chronic-inflammation-tied-to-dementia

Dementia Care Central. (2023, April 10). *Process for diagnosing Alzheimer's & other related dementias*. https://www.dementiacarecentral.com/aboutdementia/diagnosing/

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. https://doi.org/10.1162/jocn.2007.19.9.1498

CDR® Dementia Staging Instrument | Knight Alzheimer Disease Research Center | Washington University in St. Louis. (n.d.). https://knightadrc.wustl.edu/professionals-clinicians/cdr-dementia-staging-instrument/

Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., DeCarli, C., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Landau, S. M., Morris, J. C., Okonkwo, O., Perrin, R. J., Petersen, R. C., Rivera, M. M., Saykin, A. J., Shaw, L. M., Toga, A. W., Tosun, D., … Initiative, A. D. N. (2022). Using the Alzheimer's Disease Neuroimaging Initiative to improve early detection, diagnosis, and treatment of Alzheimer's disease. *Alzheimer's & Dementia*, *4*.

What is Alzheimer's Disease? | CDC. (n.d.). https://www.cdc.gov/aging/aginginfo/alzheimers.htm

Thurston, A. (2024, July 9). New AI Program from BU Researchers Could Predict Likelihood of Alzheimer's Disease. *Boston University*. https://www.bu.edu/articles/2024/new-ai-program-could-predict-alzheimers-disease/#:~:text=Now%2C%20Boston%20University%20researchers%20say,shows%20the%20power%20of%20AI.%E2%80%9D

Basheer, S., Bhatia, S., & Sakri, S. B. (2021). Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset. *IEEE Access*, *9*, 42449–42462. https://doi.org/10.1109/access.2021.3066213

Lee, L. Y., Vaghari, D., Burkhart, M. C., Tino, P., Montagnese, M., Li, Z., Zühlsdorff, K., Giorgio, J., Williams, G., Chong, E., Chen, C., Underwood, B. R., Rittman, T., & Kourtzi, Z. (2024). Robust and interpretable AI-guided marker for early dementia prediction in real-world clinical settings. *EClinicalMedicine*, 102725. https://doi.org/10.1016/j.eclinm.2024.102725

Appendix A

Code

Import Necessary Libraries

```
from google.colab import drive
```

```
drive.mount('/content/drive')

import os
import numpy as np
import pandas as pd
from tensorflow.keras.preprocessing.image import img_to_array,
load_img,ImageDataGenerator
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, Conv3D, MaxPooling2D, Flatten,
Dense, Dropout
from tensorflow.keras.optimizers import Adam
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns
```

Mount Google Drive

```
from google.colab import drive
drive.mount('/content/drive')
```

Set File Paths and Load Datasets

```
base_dir = '/content/drive/My Drive/ANA699/2.Data_Literature/Datasets'

print("Base directory contents:", os.listdir(base_dir))

alzheimers_train_dir = os.path.join(base_dir,
'alzheimers_imaging_dataset/train')
print("Alzheimer's Imaging Train directory contents:",
os.listdir(alzheimers_train_dir))

alzheimers_test_dir = os.path.join(base_dir,
'alzheimers_imaging_dataset/test')
print("Alzheimer's Imaging Test directory contents:",
os.listdir(alzheimers_test_dir))

oasis_dir = os.path.join(base_dir, 'oasis_subset_9488')
print("OASIS Subset directory contents:", os.listdir(oasis_dir))
```

Load Images and Labels

```
def load_images_from_folder(folder, label, img_size=(128, 128)):
    images = []
    labels = []
```

```
    for filename in os.listdir(folder):
        img_path = os.path.join(folder, filename)
        if filename.lower().endswith(('.png', '.jpg', '.jpeg')):
            img = load_img(img_path, target_size=img_size)
            img_array = img_to_array(img)
            images.append(img_array)
            labels.append(label)
    return np.array(images), np.array(labels)

def load_dataset(base_path, labels_dict, img_size=(128, 128)):
    images = []
    labels = []
    for folder_name, label in labels_dict.items():
        folder_path = os.path.join(base_path, folder_name)
        imgs, lbls = load_images_from_folder(folder_path, label, img_size)
        images.extend(imgs)
        labels.extend(lbls)
    return np.array(images), np.array(labels)

alzheimers_imaging_labels = {
    'VeryMildDemented': 0,
    'ModerateDemented': 1,
    'MildDemented': 2,
    'NonDemented': 3
}

oasis_subset_labels = {
    'Moderate Dementia': 1,
    'Mild Dementia': 2,
    'Non Demented': 3,
    'Very mild Dementia': 0
}

alzheimers_train_images, alzheimers_train_labels =
load_dataset(alzheimers_train_dir, alzheimers_imaging_labels)

oasis_images, oasis_labels = load_dataset(oasis_dir, oasis_subset_labels)
```

Checking the Datasets

```
def check_dataset(images, labels, dataset_name):
    print(f"{dataset_name} - Images shape: {images.shape}, Labels shape:
{labels.shape}")
```

```
    print(f"{dataset_name} - Unique labels: {np.unique(labels)}")

check_dataset(alzheimers_train_images, alzheimers_train_labels, "Alzheimer's
Imaging Train")
check_dataset(alzheimers_test_images, alzheimers_test_labels, "Alzheimer's
Imaging Test")
check_dataset(oasis_images, oasis_labels, "OASIS Subset")
```

Combining the Datasets

```
import numpy as np

combined_images = np.concatenate((alzheimers_train_images, oasis_images),
axis=0)


combined_labels = np.concatenate((alzheimers_train_labels, oasis_labels),
axis=0)

print(f"Combined Images shape: {combined_images.shape}")
print(f"Combined Labels shape: {combined_labels.shape}")
print(f"Unique labels in combined dataset: {np.unique(combined_labels)}")

import numpy as np
import pandas as pd

label_mapping = {
    0: 'Very Mild Demented',
    1: 'Moderate Demented',
    2: 'Mild Demented',
    3: 'Non Demented'
}

label_counts = np.bincount(combined_labels)

df_label_counts = pd.DataFrame({
    'Label': np.arange(len(label_counts)),
    'Label Name': [label_mapping[i] for i in range(len(label_counts))],
    'Count of Images': label_counts
})

print(df_label_counts)
```

Build, Train, and Test the CNN Model

```
from sklearn.metrics import classification_report
import numpy as np
from tensorflow.keras.utils import to_categorical
from sklearn.model_selection import train_test_split
from sklearn.utils import shuffle
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

combined_images = combined_images.astype('float32') / 255.0

combined_labels = to_categorical(combined_labels, num_classes=4)

combined_images, combined_labels = shuffle(combined_images, combined_labels,
random_state=42)

X_train, X_test, y_train, y_test = train_test_split(combined_images,
combined_labels, test_size=0.2, random_state=42)

from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense,
Dropout

model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(128, 128, 3)),
    MaxPooling2D((2, 2)),

    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D((2, 2)),

    Conv2D(128, (3, 3), activation='relu'),
    MaxPooling2D((2, 2)),

    Flatten(),
    Dense(128, activation='relu'),
    Dropout(0.5),

    Dense(4, activation='softmax')
])


model.compile(optimizer='adam',
              loss='categorical_crossentropy',
              metrics=['accuracy'])

model.summary()
```

Add Early Stop, Train & Test the Model

```
from tensorflow.keras.callbacks import EarlyStopping

early_stopping = EarlyStopping(
    monitor='val_loss',
    patience=5,
    restore_best_weights=True
)

history = model.fit(X_train, y_train,
                    epochs=50,
                    batch_size=32,
                    validation_data=(X_test, y_test),
                    callbacks=[early_stopping])

y_pred_probs = model.predict(X_test)
y_pred = np.argmax(y_pred_probs, axis=1)
y_true = np.argmax(y_test, axis=1)


class_report = classification_report(y_true, y_pred,
target_names=['VeryMildDemented', 'ModerateDemented', 'MildDemented',
'NonDemented'])
print("Classification Report:\n", class_report)

plt.figure(figsize=(12, 6))
plt.plot(history.history['loss'], 'bo-', label='Training Loss')
plt.plot(history.history['val_loss'], 'ro-', label='Validation Loss')
plt.title('Loss Curve')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.grid(True)
plt.show()

plt.figure(figsize=(12, 6))
plt.plot(history.history['accuracy'], 'bo-', label='Training Accuracy')
plt.plot(history.history['val_accuracy'], 'ro-', label='Validation
Accuracy')
plt.title('Accuracy Curve')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.grid(True)
plt.show()
```

Print Test Results of Confusion Matrix

```
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

conf_matrix = confusion_matrix(y_true, y_pred)

plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
            xticklabels=['VeryMildDemented', 'ModerateDemented',
'MildDemented', 'NonDemented'],
            yticklabels=['VeryMildDemented', 'ModerateDemented',
'MildDemented', 'NonDemented'])
plt.title('Confusion Matrix')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()
```

Build & Train the SVM Model

```
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import classification_report, confusion_matrix

import numpy as np

X_flat = combined_images.reshape(14609, -1)

X_flat = X_flat / 255.0

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X_flat, combined_labels,
test_size=0.2, random_state=42)

scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

svm_model = SVC(kernel='rbf', random_state=42)

svm_model.fit(X_train, y_train)
```

Test the SVM Model and Visualize Results

```
from sklearn.metrics import classification_report, confusion_matrix
```

```
import matplotlib.pyplot as plt
import seaborn as sns

y_pred_svm = svm_model.predict(X_test)

print("Classification Report:")
print(classification_report(y_test, y_pred_svm, target_names=['Class 0',
'Class 1', 'Class 2', 'Class 3']))

conf_matrix_svm = confusion_matrix(y_test, y_pred_svm)

plt.figure(figsize=(10, 7))
sns.heatmap(conf_matrix_svm, annot=True, fmt='d', cmap='Blues',
            xticklabels=['Class 0', 'Class 1', 'Class 2', 'Class 3'],
            yticklabels=['Class 0', 'Class 1', 'Class 2', 'Class 3'])
plt.title('Confusion Matrix - SVM')
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.show()
```