Electronics and Computer Science
Faculty of Physical Sciences and Engineering
University of Southampton

Mirza Ayaz Baig

14 May 2024

Combining imaging and inflammation markers for evaluating the risk of progression to dementia

Project Supervisor: Dr Rahman Attar

Second Examiner: Dr Julian Rathke

A project report submitted for the award of
MEng Computer Science

Abstract

The objective of this project was to evaluate the effectiveness of combining neuroimaging data with inflammatory protetin markers for diagnosing dementia, as opposed to using neuroimaging alone. The ResNet architecture was used, modified to perform classification on 3D volumes and multi-modal inputs. The training data was accessed via the Alzheimer's Disease Neuroimaging Initiative (ADNI) which provided datasets for MRI, FDG-PET and inflammatory protein markers. The Taguchi method was utilised to train several configurations of ResNet model with varying values for learning rate and batch size to obtain an optimal model using the Adaptive Moment Estimation (Adam) optimiser with a learning rate of 0.0001 and a batch size of 6. The results showcase better performance using a standalone MRI dataset with an area under the ROC curve score of 0.89, compared to 0.83 achieved using a combined dataset.

Statement of Originality

- I have read and understood the ECS Academic Integrity information and the University's Academic Integrity Guidance for Students.
- I am aware that failure to act in accordance with the Regulations Governing Academic Integrity may lead to the imposition of penalties which, for the most serious cases, may include termination of programme.
- I consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to verify whether my work contains plagiarised material, and for quality assurance purposes.

You must change the statements in the boxes if you do not agree with them.

We expect you to acknowledge all sources of information (e.g. ideas, algorithms, data) using citations. You must also put quotation marks around any sections of text that you have copied without paraphrasing. If any figures or tables have been taken or modified from another source, you must explain this in the caption and cite the original source.

I have acknowledged all sources, and identified any content taken from elsewhere.

If you have used any code (e.g. open-source code), reference designs, or similar resources that have been produced by anyone else, you must list them in the box below. In the report, you must explain what was used and how it relates to the work you have done.

I have not used any resources produced by anyone else.

You can consult with module teaching staff/demonstrators, but you should not show anyone else your work (this includes uploading your work to publicly-accessible repositories e.g. Github, unless expressly permitted by the module leader), or help them to do theirs. For individual assignments, we expect you to work on your own. For group assignments, we expect that you work only with your allocated group. You must get permission in writing from the module teaching staff before you seek outside assistance, e.g. a proofreading service, and declare it here.

I did all the work myself, or with my allocated group, and have not helped anyone else.

We expect that you have not fabricated, modified or distorted any data, evidence, references, experimental results, or other material used or presented in the report. You must clearly describe your experiments and how the results were obtained, and include all data, source code and/or designs (either in the report, or submitted as a separate file) so that your results could be reproduced.

The material in the report is genuine, and I have included all my data/code/designs.

We expect that you have not previously submitted any part of this work for another assessment. You must get permission in writing from the module teaching staff before re-using any of your previously submitted work for this assessment.

I have not submitted any part of this work for another assessment.

If your work involved research/studies (including surveys) on human participants, their cells or data, or on animals, you must have been granted ethical approval before the work was carried out, and any experiments must have followed these requirements. You must give details of this in the report, and list the ethical approval reference number(s) in the box below.

My work did not involve human participants, their cells or data, or animals.

ECS Statement of Originality Template, updated August 2018, Alex Weddell aiofficer@ecs.soton.ac.uk

Acknowledgements

I would like to thank my supervisor, Dr. Rahman Attar for the continued support throughout this project. I would also like to thank Dr. Sofia Michopoulou for the advice regarding specific medical questions beyond the project scope which provided a clearer insight into dataset selections.

Contents

1 Introduction

Dementia is a progressive neurodegenerative syndrome that causes the loss of cognitive function resulting in impairment of daily activities eventually leading to disability and death. Dementia is not a specific disease but rather a group of symptoms associated with underlying conditions such as Alzheimer's disease, vascular dementia, frontotemporal dementia or Lewy body dementia [16]. In the UK, the number of people with dementia is estimated to be around 920,000 - a figure which is expected to rise to over a million by 2024, mainly due to improved life expectancy [3, 38]. Alzheimer's Disease accounts for around 60-70% of dementia cases in the UK and is most prevalent in patients over the age of 65 [41]. The economic costs are vast and split among healthcare, social care and unpaid care, with the UK spending nearly £35 billion in 2019 [46], costs which are expected to rise to almost £60 billion by 2030 [46].

Aside from the direct economic toll, the social impact of dementia is often profound, deeply affecting families, social relationships and mental health. As the condition progresses, both patients and caregivers may face social isolation, with caregivers finding it increasingly difficult to maintain social relationships or engage in social activities due to caregiving responsibilities and behavioural symptoms of the patient [14]. This isolation often leads to a decrease in social support, resulting in negative effects on mental health [8].

Diagnosis of dementia involves a multifaceted approach that assesses both qualitative and quantitative aspects of cognitive performance. The current clinical approach requires an evaluation of the patient's health and family history, with additional mental cognitive assessments. A common form of such an assessment is the Montreal Cognitive Assessment (MoCA), which is useful in the early detection of cognitive impairment. The outcome of the MoCA is then used to perform further evaluations [6]. Due to the nature of current diagnostic practices being based on assessment of symptoms rather than biomarkers [35], they often contribute to lengthy diagnostic delays, with studies concluding an average of 2 years from initial symptom onset to diagnosis [43]. In young-onset dementia for conditions such as frontotemporal dementia, the process often takes longer, around 4 years on average [43]. This protracted diagnostic timeline underlines an urgent need for faster and precise diagnostic tools, to allow for earlier intervention and aid patients and families in making more informed decisions regarding their life affairs.

Since dementia is a result of various neurological conditions, the variability of symptoms poses a challenge in efficient diagnosis. The reliance on subjective assessments can sometimes lead to inconclusive results [6, 35], which further contributes to diagnostic delays. In such cases, neuroimaging and neuropsychological

testing are used to further determine the presence of dementia. Brain imaging such as Positron Emission Tomography (PET) and Magnetic Resonance Imaging (MRI) are used alongside various radiotracers to detect common characteristics of dementia, such as grey matter deterioration [11]. Whilst neuroimaging techniques for diagnosis are still in the early stages, clinical research has found these methods to be exceedingly helpful in differentiating between multiple subtypes within various etiologies of dementia [9].

Biomarkers and measures of inflammation through fluid samples are also clinically found to be a way to improve diagnostic accuracy. Cerebrospinal fluid is commonly used for this, by measuring levels of amyloid-beta peptides which is seen as one of the main components of amyloid plaques, an indicative feature of Alzheimer's Disease [36].

Neuroimaging techniques and biomarker analysis are often complex and costly processes. Whilst they may produce more accurate diagnoses, interpreting and making deductions from the data requires a high level of expertise and could result in incorrect diagnoses if wrongly analysed. Artificial Intelligence and deep learning models can be used to mitigate this, with research finding AI techniques associated with MRI scans to have a diagnostic accuracy range of 73.9% to 99% [10]. These models can analyse diverse data types, learn from patterns and provide consistent analysis, making it valuable for complex cases with overlapping symptoms. Convolutional neural networks (CNN) and recurrent neural networks (RNN) are commonly used for this task, with past research utilising several different models and techniques such as residual learning and transfer learning to conclude the effectiveness of AI in assisting with Alzheimer's diagnosis and management [18, 33, 48]. The potential of AI to streamline and expedite the diagnostic process showcases its ability to potentially speed up diagnosis with improved accuracy and efficiency [13]. Moreover, AI's ability to handle and analyse complex, multi-modal datasets presents a unique opportunity to explore how different types of diagnostic data can complement each other.

Building on this premise, this project aims to assess the impact of AI in refining diagnostic accuracy and efficiency in Alzheimer's disease. By comparing the effectiveness of neuroimaging as an input and in combination with inflammation biomarkers, this study seeks to analyse whether a combination of these datasets can provide more robust results as opposed to using imaging alone. This comparative analysis can help determine the most effective approach when harnessing AI techniques to better diagnose and manage the progression of dementia.

2 Background Literature Review

The literature review for this project is structured to explore four key areas: an overview of dementia, the challenges and limitations of current diagnostic practices, the role of neuroimaging and inflammation biomarkers in identifying dementia and the application of artificial intelligence in diagnosing dementia. The limitation of current research is also included to identify the gaps in relevant research which could be addressed in this project. This approach was taken to gain a better understanding of the problem at hand and learn from existing research to build an accurate solution which answers the project hypothesis.

2.1 Overview of Dementia

Dementia refers to a clinical syndrome which is characterised by a progressive cognitive decline hindering the ability to function independently [19]. It cannot be attributed to a single cause, but rather a collection of various neurodegenerative disorders which cause an irreversible loss of brain functionality. There are several common types of dementia including Alzheimer's Disease, frontotemporal dementia, Lewy body dementia and vascular dementia, each characterised by distinct causes and symptoms [1]. For example, Alzheimer's Disease, the most prevalent form of dementia which affects around 60% of dementia patients in the UK [41], is primarily characterised by the buildup of amyloid plaques and tau tangles in the brain leading to memory loss and cognitive decline [22]. Vascular dementia, which affects around 15% of dementia patients in the UK [41], is caused by impaired blood flow to the brain, often as a result of stroke and typically causes problems with problem-solving and planning, rather than memory loss [12].

The Alzheimer's Society, in collaboration with the London School of Economics (LSE), conducted research on dementia in the UK, revealing alarming projections on a personal and economic level. In 2019, there were approximately 885,000 older patients in the UK, which is expected to surge to 1.23 million by 2030. Consequently, the economic costs are predicted to rise from £34.7 billion in 2019 to a staggering £59.2 billion by 2030 [46]. These projections highlight the growing significance of dementia and its societal concerns in the UK - with the report concluding how ineffective progress in diagnosis and treatment could lead to damaging economic and societal repercussions. The report notes that the predictive analysis did not account for potential new interventions for dementia, suggesting that advancements in diagnosis and treatment methods have the potential to significantly mitigate the projected numbers.

Aside from the economic costs, the social impact of dementia on patients and their caregivers is profound, affecting the emotional, social and financial aspects

of their lives. Brodaty et. al. [15] in their review of the social impact of dementia on patients and their families state how both patients and caregivers face an increased level of social isolation. The cognitive decline of patients often leads to reduced communication and withdrawal from social activities, which further accelerates symptoms and decline [8, 15]. The review showcases the significant burden caregivers face with high levels of stress, emotional strain and financial pressure, due to the sacrifice of leisure pursuits, hobbies and reduced employment. The report further concludes that with early intervention, the patient and caregiver can more easily manage the progression of dementia with comprehensive plans from doctors and healthcare workers, which highlights the need for early diagnosis.

2.2 Challenges and Limitations of Diagnostic Practices

Diagnosis of dementia is a multifaceted approach which requires several stages of examinations and activity monitoring [6]. These stages include but are not strictly limited to:

- Taking collateral history with collaboration from family members
- Cognitive examinations using screening tools such as the Montreal Cognitive Assessment (MoCA)
- Neurological examinations to evaluate and find evidence for problems such as aphasia, apraxia and agnosia, which are pathognomonic signs of dementia
- Routine blood tests to exclude other causes of symptoms which can be mistaken for dementia
- Neuroimaging analysis, such as positron emission tomography (PET) and magnetic resonance imaging (MRI)

In some cases, additional evaluation is required as patients may show 'normal' function on a screening cognitive test, or in other stages of clinical diagnosis. An example of this is evident in highly educated or functioning individuals in cognitively demanding jobs (such as a doctor or a lawyer). In these cases, the individual may appear to 'pass' the cognitive test despite having a history of cognitive decline in their respective field [6].

Clinical diagnosis is also said to take an average of 2.8 years from initial symptom onset for late-onset dementia, and even more so for young-onset dementia with an average of 4.4 years [43]. This extended timeline for diagnosis demonstrates a clear need for faster and accurate forms of diagnosis which can aid in providing families more time to manage the progression of the disease.

Uncertainty regarding aetiology can also warrant further examination and delays, as dementia covers several symptoms and causes which can go on to affect treatment. Since current diagnosis methods are focused on assessing symptoms, it can be difficult to differentiate between types of dementia and other conditions which present similar symptoms. Neuroimaging analysis can mitigate this by shifting the focus to analysing biomarkers and specific areas of the brain using PET and MRI scans to achieve more concrete diagnoses. These modalities are used to identify patterns of brain atrophy and metabolic dysfunction specific to each subtype, which can help clinicians understand the underlying pathophysiology and make more accurate diagnoses [2].

2.3 Neuroimaging and Inflammation in Dementia

A review on the diagnosis and management of dementia by Arvanitakis et al. [6] suggests that neuroimaging techniques such as PET scans can mitigate the shortcomings of clinical diagnosis by further differentiating the etiologies of different forms of dementia. The review highlights how these techniques can aid in diagnosing less common causes of dementia by measuring brain activity and changes with suitable biomarkers.

Further, Banerjee et al. [9] delve into the efficacy of different neuroimaging techniques in early diagnosis of common etiologies of dementia, namely positron emission tomography (PET), magnetic resonance imaging (MRI) and single-photon emission computer tomography (SPECT). Particularly for Alzheimer's disease (AD), the review summarises how PET scans alongside 11C-labeled Pittsburg Compound B (PiB) radiotracers are effective at demonstrating signs of disease activity at early stages, capable of even differentiating between pre-clinical AD and mild cognitive impairment due to AD. These scans are used to find deposits of amyloid plaques and accumulations of amyloid-beta peptides which are pathognomonic of AD. Amyloid-PET scans for AD were found to be incredibly effective, with a sensitivity of 94% [47], but a lower specificity of 73%.

18-Fluorodeoxyglucose (FDG) radiotracers can also be used with PET imaging to measure cerebral metabolic rates of glucose [11], which indicates brain synaptic activity and density. In AD patients, these FDG-PET images show widespread metabolic deficits in the neocortical areas (related to structures such as the thalamus, cerebellum and others). The AD metabolic pattern is also characterised by hypometabolism in cognitive-relevant areas such as the medial temporal lobe and extends to the prefrontal cortex as the disease advances. This reduced rate of glucose metabolism in different areas of the brain can render synapses more vulnerable to degeneration which leads to cognitive decline [32]. This modality of

PET scans is reported to show a specificity of 85% to differentiate between AD and other forms of dementia such as frontotemporal dementia [11].

Magnetic resonance imaging (MRI) is another critical tool which has a pivotal role in the early detection and prognosis of dementia. Živanović et al. [49] review the efficacy of MRI scans for detecting dementia, and emphasise its sensitivity to detect structural neurodegenerative changes over time. A major indicator detectable by MRI scans are cerebral atrophy patterns such as medial temporal lobe atrophy which is linked with cognitive decline. In some cases, atrophy can be detectable even in presymptomatic stages, which makes MRI ideal for early diagnosis of dementia. When used in combination with mini-mental state examination (MMSE) scores, structural MRI results in sensitivity and specificity of 93% and 98%, respectively [49].

Whilst these methods are promising for early diagnosis, costs and availability are still limiting factors especially given how neuroimaging in dementia is still in the early stages of research. Previous studies comment that the standalone utility of neuroimaging can be limited. For instance, Lombardi et al. report that standalone structural MRI shows lower specificity and sensitivity, ranging from 64% to 75% [30], which would be improved by using in conjunction with a range of biomarkers to enhance diagnostic accuracy.

The primary pathological hallmarks of AD are the presence of amyloid plaques and neurofibrillary tangles (NFT) [5]. Amyloid plaques are predominantly composed of β-amyloid-40 and β-amyloid-42 peptides [22]. Concurrently, the NFTs are formed by the hyperphosphorylation of tau proteins, which usually serves a necessary role in removing tau from microtubules to allow for transport, normally followed by dephosphorylation to return tau to the microtubules [29]. In AD, hyperphosphorylation at multiple sites causes the excessive removal of tau, resulting in the collapse of microtubule structures and disruption to overall cellular morphology in the brain [5, 44].

Recent research has identified the existence of a third core feature of AD which seemingly provides a link between the two previously mentioned. Evidence of sustained inflammatory responses has shown to progress AD and its severity [26]. It may even suggest a link between the initial β-amyloid pathology and the later development of NFT [29]. This is evident through analysis of pro-inflammatory cytokines such as interleukin (IL)-6 and tumour necrosis factor α (TNF-α), with research showing elevated levels of the cytokines in the cerebrospinal fluid and blood serums of AD patients [26, 29]. Utilising inflammatory markers for diagnosis can be non-invasive and used alongside other diagnostic tools to enhance understanding of the disease.

2.4 Artificial Intelligence in Dementia Diagnosis

Recent advancements in artificial intelligence (AI) with machine learning (ML) and deep learning (DL), have shown promising potential in enhancing the diagnostic processes for dementia and Alzheimer's Disease (AD). One of the pivotal strengths of AI highlighted by Battineni et al. is its ability to engage in causal reasoning and efficiently process extensive datasets from large populations. This can be used in contrast to clinicians' judgement alone to allow for quick responses to large population screenings [10].

2.4.1 Machine Learning

Battineni et al. explore the application of various machine learning techniques for classifying AD patients. One such technique discussed is the support vector machine (SVM), a classification algorithm which maps input vectors into a higher-dimensional feature space to construct a linear decision surface, to allow for both linear and non-linear classifications [17]. The review points to work by Vichianin et al. who demonstrate the practical application of SVMs in AD diagnosis using both structural MRI and clinical parameters [42]. The results achieved show an accuracy range between 83% and 90% when leveraging clinical data such as mental state examination scores. However, the accuracy notably declines to 62.6% when exclusively using structural MRI brain volumetric values, and a combination of both data types gives an accuracy of 88.9%. It is important to note that the sample consisted mainly of patients at low to moderate stages of AD, which might explain the lower accuracies observed with MRI volumetric alone. The study also only uses brain volumetric values as input rather than the full 3D MRI volume which could have provided better results.

To expand on multi-modal diagnostic approaches, Michopoulou et al. evaluate the effectiveness of logistic regression machine learning models by integrating imaging and inflammation markers for AD diagnosis [31]. This was done using perfusion SPECT images, alongside markers of inflammation measured in cerebrospinal fluid, which achieved an area under the receiver characteristic curve (ROC) score of 0.76 for imaging input, 0.76 for inflammation input and 0.85 with a combined dataset. This showcases the potential for combining multiple modalities and datatypes to achieve more accurate diagnosis. This study does not cover the use of different modalities of imaging or detail other classification techniques aside from logistic regression which could potentially provide better results in a classification task for predicting dementia.

In these studies, machine learning models follow a structured pipeline to ensure accurate processing and analysis of data. An example of this is detailed by Ahmed

et al. in their review of neuroimaging and machine learning for dementia diagnosis, which encompasses several fundamental steps that can be seen in Figure 1 [4]. This includes stages like image preprocessing to normalise and reduce noise from images, feature extraction and optimisation algorithms and model selection. To iteratively improve results, most studies often focus on tuning hyperparameters involved in specific stages, which then provides better results for the overall model.

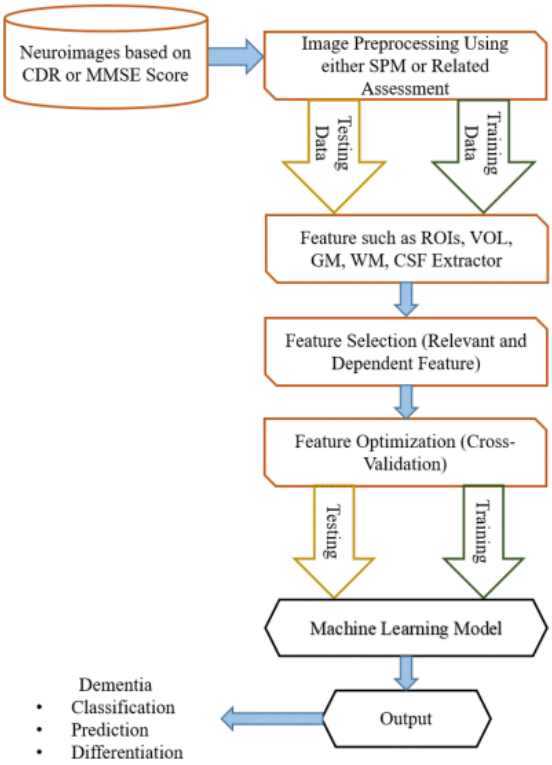

Figure 1: Step by step computer-aided diagnosis of dementia using machine learning approaches [4]

2.4.2 Deep Learning

Arya et al. concluded in their systematic review that traditional ML techniques using SVM classifier have an accuracy of 85.7% which can be improved further by implementing DL techniques with convolutional neural networks achieving an accuracy of 98.6% [7]. Ahmed et al. detail an example of the key steps in a deep learning pipeline approach, which can be seen in Figure 2 [4].

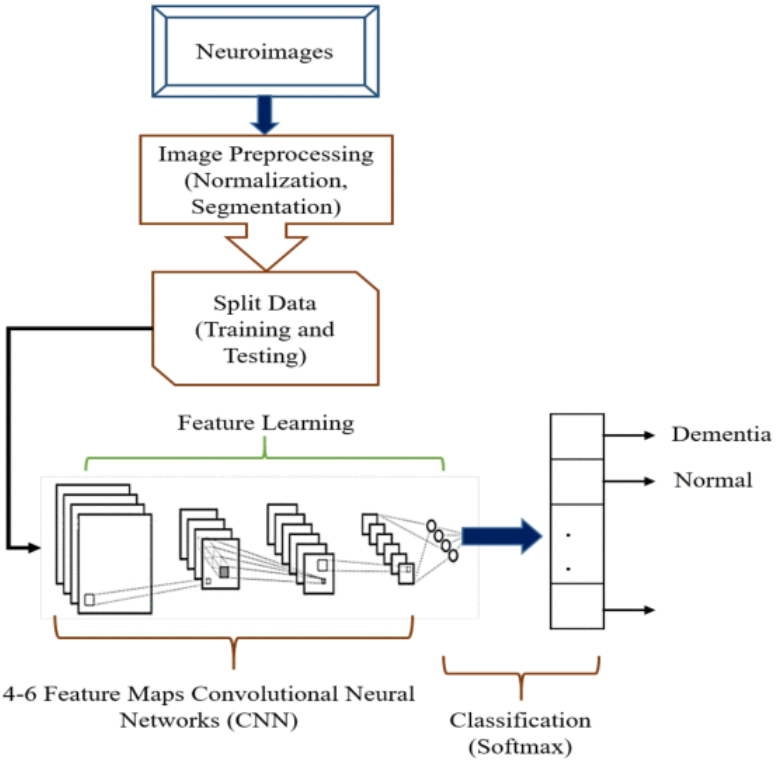

Figure 2: Deep learning CNN based classification pipeline example [4]

When utilising neuroimages for dementia diagnosis, convolutional neural networks provide an advantage due to its ability to learn more complex patterns by focusing on sections of input images. This is done using an initial convolutional layer which can perform large-scale feature extraction [20]. This feature extraction phase is typically followed by several layers that implement techniques such as batch normalisation, pooling to reduce the number of parameters, activation functions, and dropout layers to prevent overfitting. Residual network (ResNet) models have been used for medical imaging classification tasks in past literature [24, 20]. This model is a modification of a traditional CNN which utilises skip connections to jump over some layers to allow for much deeper networks. This modification addresses the vanishing gradient problem, which is an issue seen in deeper CNNs [23, 40]. Ebrahimi et al. [20] reviewed the performance of the ResNet-18 model applied on both 2D and 3D MRI scans to detect dementia. They outline their results of 96.8% accuracy, 100% sensitivity and 93.7% specificity when using an optimised ResNet-18 model with transfer learning on 3D MRI scans, which showcases the potential of implementing these DL models for dementia diagnosis.

2.4.3 Evaluation Techniques

To evaluate the efficacy of AI architectures, most studies focused on three parameters calculated using a confusion matrix [7]:

- Accuracy, which shows the correctness of the prediction by the model

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

- Sensitivity, which is the ratio of correctly positive labelled AD records to all those who suffer from AD in reality

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

- Specificity, which is the ratio of correctly negative labelled AD records to all those who are cognitively normal in reality

$$Specificity = \frac{TN}{FP + TN} \tag{3}$$

In addition to the metrics derived from a confusion matrix, many studies also utilise the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) to evaluate performance. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various thresholds. The AUC shows the likelihood of a model correctly classifying between a positive instance and a negative instance, with a higher AUC indicating better performance [7, 31].

2.5 Limitations of Research

Current research in Alzheimer's diagnosis using AI largely focuses on single-modality inputs and traditional machine learning models. For instance, studies like that of Vichianin et al. [42], which utilised SVMs on MRI volumetric values, often restrict their scope to 2D imaging or brain volumetric values, which could potentially oversimplify the complex nature of dementia. The integration of multi-modal data, especially the combination of 3D neuroimaging and inflammation markers, into more complex models like 3D-modified CNNs also presents a gap in research. Many research projects leverage either 2D imaging or volumetric values, which could potentially fail to capitalise on the spatial information from 3D scans. This is evident in research like that of Michopoulou et al [31],

which utilised statistical parametric mapping of SPECT images with inflammation markers.

Inflammation markers are scarcely covered in research for diagnosing dementia with AI, which is due to the field still being in its early stages. Current studies suggest these markers could be used to link the two primary pathological features of Alzheimer's - amyloid plaque buildup and neurofibrillary tangles [26, 29]. Despite this, there is still a lack of comprehensive studies to fully understand and leverage inflammation markers for early and accurate dementia diagnoses using AI. This presents a gap in research to evaluate the effectiveness of inflammation markers combined with neuroimaging in multi-modal frameworks to classify AD.

It is important to acknowledge that many AI projects and research which have covered dementia diagnosis predominantly focus on technical performance rather than analysis of the medical context. The motivation for such research is to evaluate the accuracy of specific ML and DL architectures in performing classification tasks on medical imaging, which could involve techniques that may not be so easily adaptable in real-world medical settings. For example, the use of overly complex models that require extensive computational resources may achieve high accuracy in controlled environments but could be impractical in clinical environments.

3 Methodology

The main focus of this project is to explore the diagnostic utility of combining neuroimaging with inflammation markers in dementia progression. Specifically, the aim was to assess whether inflammation markers can enhance the reportedly high accuracy of neuroimaging as seen in past research. The evaluation focused on comparing single-modality neuroimaging input metrics against those obtained from integrating neuroimaging and inflammation markers together.

To structure the project, a methodology was adopted similar to the one detailed by Ahmed et al. [4] in Figures 1 and 2, which outline a pipeline for ML and DL projects commonly used in similar past research. The iterative workflow approach from the "Introduction to Practical Use Cases" lecture in the COMP3222 Machine Learning Technologies module was also followed, which helped to refine the project through continuous evaluation and adaptation. The following sections detail the steps taken and how these contributed to the overall effectiveness of the project.

3.1 Problem Characterisation

The primary objective of this project was to develop and evaluate a deep learning model which classifies patients with Alzheimer's disease using both neuroimaging and inflammation markers. The focus was on Alzheimer's specifically since it is the leading form of dementia in the UK, and subsequently has the most available data [41, 46]. The dual-modality approach was chosen to explore whether integrating inflammation markers can enhance diagnostic accuracy achieved with neuroimaging alone.

3.1.1 Success Criteria

To evaluate the success of a model on a specific dataset, the three metrics detailed in the review by Arya et al. [7] and Ebrahimi et al. [20] were used: accuracy, specificity and sensitivity, along with the AUC score taken from a ROC graph [31]. These metrics were selected to ensure model capabilities across both positive and negative cases to build a balanced model which is reflective of real-world needs.

From the literature review, a range of 60% to 80% for accuracy, specificity and sensitivity and 0.6 to 0.8 AUC score were chosen as acceptable comparisons [7, 13, 20, 31]. While higher values for these metrics would be advantageous, the project's focus was to validate the potential benefit of using inflammation markers in a multi-modal approach. The literature review of similar projects showed

that most research achieved even higher values than the above, but this usually required more complex and computationally demanding models or evaluation of a wide range of models which may not fit within the constraints of the project scope and timeline.

3.1.2 Data Requirements

For this project, the primary datasets required are neuroimaging brain scans and inflammation biomarker values from both Alzheimer's patients and normal controls. Similar past classification projects commonly use Amyloid-PET, MRI and FDG-PET scans, which can be processed as either 2D images or 3D volumes. Additionally, a range of inflammation biomarkers such as TNF-α, IL-16 and IP-10 are potentially effective for early diagnosis [26, 29], which are measured in blood plasma and serum. The final dataset for this project would need to include either of the three neuroimaging modalities and quantified inflammation biomarkers for each patient.

3.1.3 Data Collection

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

ADNI is the most common dataset used in similar deep learning projects and medical research into dementia [21], and includes data for subjects aged 55 to 90 across the US and Canada.

Whilst there is an abundance of neuroimaging scans, datasets for inflammation biomarkers are limited, likely due to the relatively early stage of research into their role in dementia diagnosis. During this project, the only available data on inflammation biomarkers available on ADNI was the Hu Lab CSF Inflammatory proteins dataset [27]. This dataset measured values of non-β-amyloid and non-tau biomarkers of AD, focusing on inflammatory biomarkers that could potentially indicate disease progression. It encompasses data from 90 subjects diagnosed with AD, 113 normal controls (NC) and 185 subjects with mild cognitive impairment (MCI). The dataset includes measures of 13 inflammatory proteins in cerebrospinal fluid (CSF), including pro-inflammatory markers like IL-7 and ILp40,

anti-inflammatory markers such as IL-4 and IL-10, and T-helper cell-associated markers like IL-6 and TGF-β, measured in picograms per millilitre (pg/mL) [27].

After refining the inflammatory markers dataset by removing duplicates and entries with missing values (detailed cleaning methods are described in section **3.4, Data Preprocessing**), the patient count was reduced to 385 patients. For these select patients, ADNI did not include Amyloid-PET scans, so this modality was not considered further. After consultations were conducted with Dr Sofia Michopoulou, the specific search criteria for FDG-PET from ADNI was defined:

- Radioisotope: F-18
- Radiopharmaceutical: Fluorodeoxyglucose F-18 and 18F-FDG

This search yielded around 4000 scans, with each patient having between 3 to 10 scans. This included scans taken on different dates and processed through various pre-processing steps as part of the ADNI acquisition procedure (discussed further in section **3.4, Data Preprocessing**). To refine the dataset further, only the most recent scans that had undergone the same pre-processing steps were selected, which resulted in a final dataset of FDG-PET scans for 251 patients, split into 45 Alzheimer's patients, 120 patients with mild cognitive impairment and 86 normal controls.

A similar approach was used to retrieve MRI images, which resulted in MRI scans for 358 patients, split into 90 Alzheimer's patients, 158 patients with mild cognitive impairment and 110 normal controls. T1 MRI was selected as the choice of weighting due to its capability of providing contrast between white and grey matter and assessing cortical atrophy, a key landmark of Alzheimer's disease [9, 49].

3.2 Data Characterisation

3.2.1 Neuroimaging Dataset

The FDG-PET scans from ADNI were available as 2D image slices in DICOM format, whereas the MRI scans were available as 3D volumes in NIFTI format. Additional patient metadata was also available as a comma-separated values (CSV) file, containing information for each patient such as patient ID, scan acquisition date and group ID. The group ID categorised the patients into three groups: 'AD' for Alzheimer's patients, 'MCI' for those with mild cognitive impairment, and 'NC' for normal controls. This categorisation was used as the labels for each patient scan, which was utilised when training and validating model predictions.

Due to the difference in format of the two modalities, both 2D and 3D approaches were taken when implementing the classification model. For the FDG-PET scans, the Python library dicom2nifti was utilised to compile the slices into 3D NIFTI volumes. On the other hand, the MRI scans were split into individual 2D slices using the nii2dcm Python library. This reformation allowed for consistent handling and analysis of the neuroimages for both modalities.

After obtaining the dataset from ADNI, the original folder format was reorganised to separate the scans of patients with AD from those without. Figure 3 illustrates an example of the folder structure before and after this reorganisation, which is similarly applied for the FDG-PET scans. This reorganisation was done by using the metadata csv which showed the group ID for each patient, and the scan acquisition date, so that only the latest scan was considered for each patient.

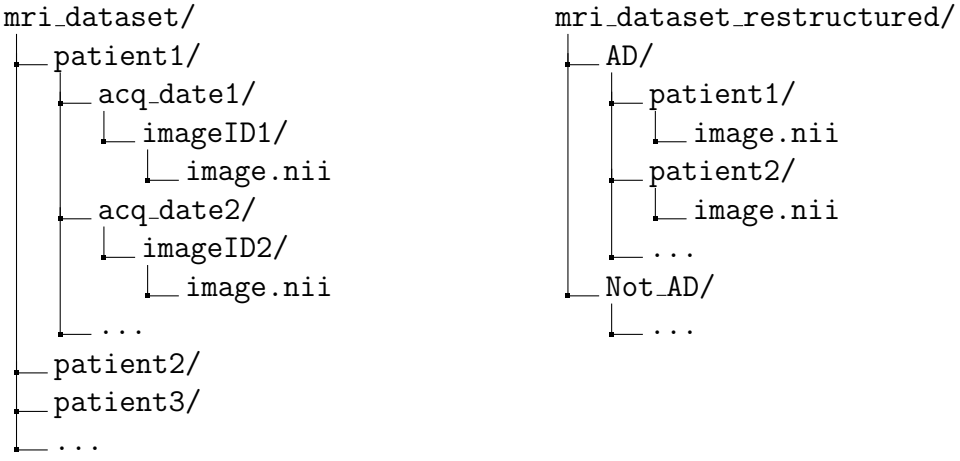

Figure 3: Comparison of original and restructured MRI dataset folder structures

3.2.2 Inflammation Markers Dataset

The inflammation markers dataset from ADNI, provided as a CSV file, included 398 entries, with each entry consisting of 13 columns representing 13 different inflammatory protein markers, alongside metadata such as patient ID number and examination date. The use of a patient ID number presented a mismatch challenge since the neuroimaging dataset used alphanumeric strings for patient IDs which differed in format.

To resolve this mismatch, ADNI provides a data linkage file which includes both types of patient IDs and their associated testing group to align different datasets. Using the pandas Python library, which is commonly used for data frame manipulation, the inflammation markers dataset was grouped by patient

ID taken from the neuroimaging dataset metadata, and then cleaned to remove duplicates. To further filter the dataset, the entries were sorted by acquisition date in descending order, and only the most recent inflammation marker values for each patient were retained, as similarly done for the neuroimaging scans. Whilst this method mostly removed any duplicate entries and redundant values, there were still some patients who had been tested multiple times on the same date. To properly account for this, and after consultations with Dr Sofia Michopoulou, an average of the inflammatory protein marker values was taken for each value present to ensure consistency and accuracy in the dataset.

3.2.3 Ethical Approval

Access to the ADNI data was obtained following approval through the ADNI website (https://adni.loni.usc.edu/data-samples/access-data/). Ethical approval was also required from the University, as the project involves data from human participants, although obtained from a secondary source. This approval was facilitated through the University's Ethics and Research Governance Online (ERGO) management system (https://www.ergo2.soton.ac.uk). The ERGO2 submission ID for this project is 89875.

3.3 Data Analysis

Due to the complex nature of Alzheimer's disease, it is vital to develop an understanding of the input data and how specific features may relate to key characteristics. An initial analysis of the datasets provided an insight into these features which an AI model would need to learn when performing classification. This was then used to form a hypothesis which relates to the overall goal of this project.

3.3.1 MRI Images

A hallmark feature of Alzheimer's observable in MRI scans is brain atrophy, particularly in the region surrounding the medial temporal lobe. T1-weighted MRI scans effectively reveal these patterns in subregions such as the hippocampus and other surrounding structures [49]. An example of this is evident when comparing Figure 4, which shows the middle slices of the three axes (axial, coronal and sagittal) from an MRI scan of a normal control patient, contrasted with Figure 5 which shows a scan of an Alzheimer's patient. The slices in Figure 5 illustrate a generally symmetric pattern of atrophy, particularly in the lower regions of the brain corresponding to the medial temporal lobe.

Medial temporal lobe atrophy, particularly in the cognitive-relevant subregions

such as the hippocampus, is closely related to cognitive decline and is commonly symmetric in AD patients [49].

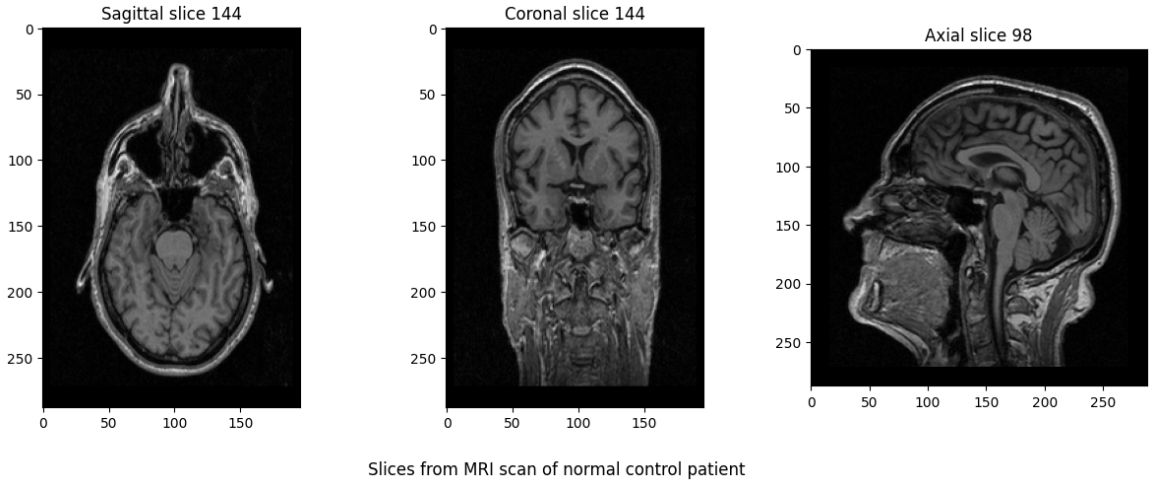

Figure 4: Middle sagittal, coronal and axial slices of normal control patient from MRI scan

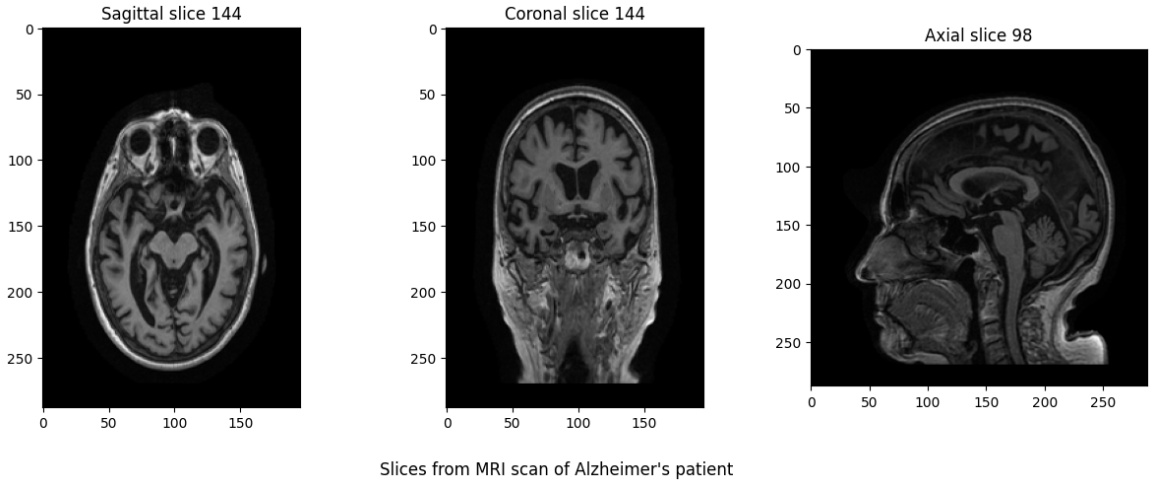

Figure 5: Middle sagittal, coronal and axial slices of Alzheimer's patient from MRI scan

3.3.2 FDG-PET Images

Reduced metabolic rates of glucose and the consequent metabolic deficits is another hallmark feature of AD. FDG-PET scans can be used to visualise this buildup of glucose in areas of the brain such as the aforementioned medial temporal lobe. This is evident by comparing Figure 6, which shows slices from an FDG-PET scan of a normal patient, contrasted with Figure 7 which shows the slices from a scan of an Alzheimer's patient. However, the quality of the FDG-PET scans from ADNI was an issue during this project, which resulted in less favourable outcome metrics further discussed in section **4, Results & Evaluation**. This resulted in FDG-PET scans not being used during the final stages of training since MRI input was producing better results.

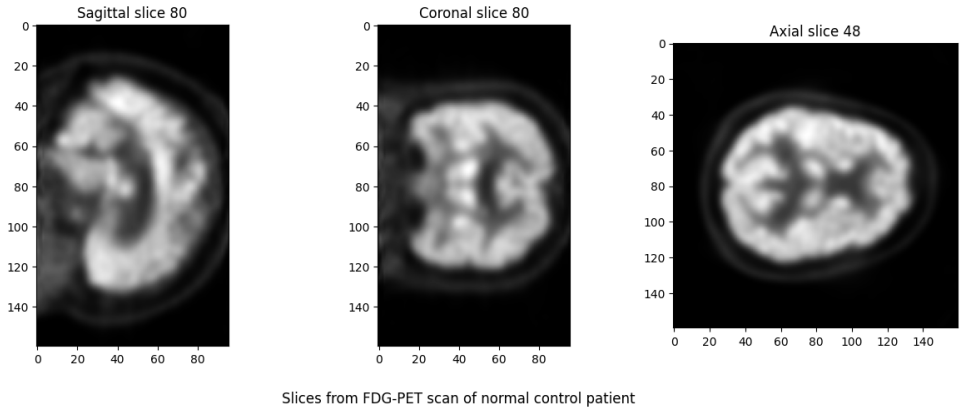

Figure 6: Middle sagittal, coronal and axial slices of normal control patient from FDG-PET scan

3.3.3 Inflammation Markers

The ADNI dataset for inflammation markers presents values for 13 different inflammatory protein markers, each measured within differing ranges. While the majority of patients had values recorded for all 13 markers, there were a subset of patients who had missing data, particularly for the markers IL-9, IP-10 and IL-12 P40. Figure 8 showcases the number of missing values per biomarker, which highlights that these three markers were missing for around 50 patients. Throughout this project, only the patients from the inflammation markers dataset who also had either MRI or FDG-PET scans were considered so that they could be used for both classification models.

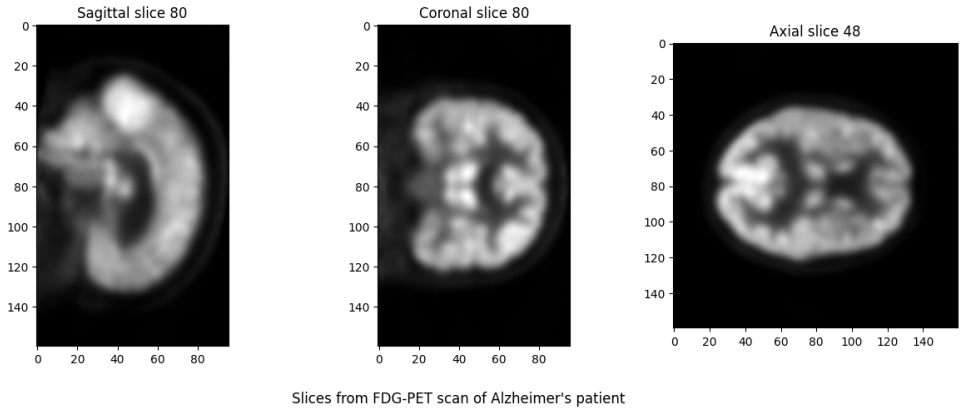

Figure 7: Middle sagittal, coronal and axial slices of Alzheimer's patient from FDG-PET scan

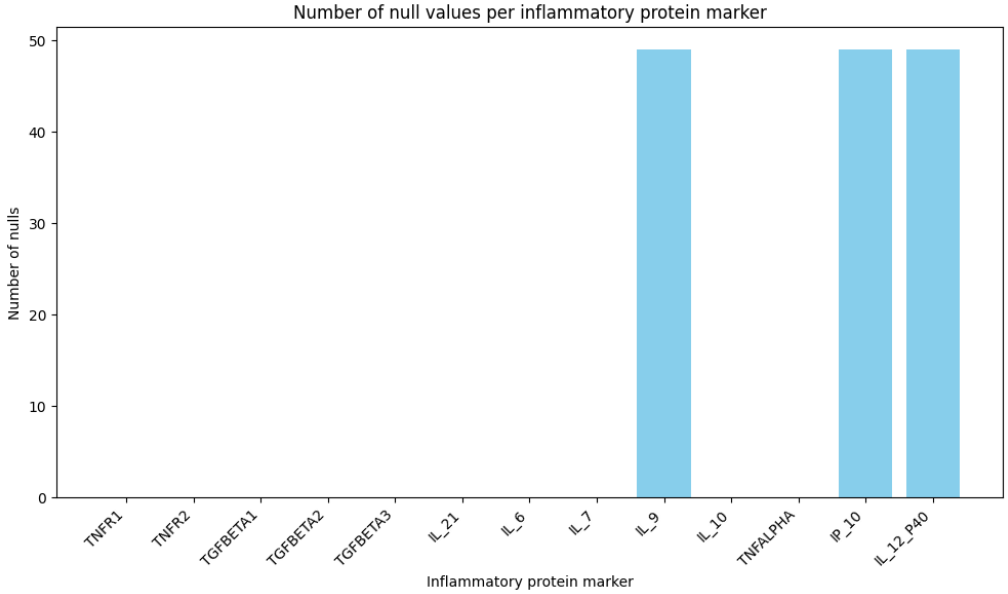

Figure 8: Number of null values per inflammatory protein marker for patients who also have MRI/FDG-PET scans

After excluding the three markers with significant missing data, the distributions of the remaining protein markers were analysed and can be seen present as box plots in Figure 9. This analysis showed varying ranges of values among the protein markers. For example, IL-6 values ranged from approximately 0.8 pg/mL to around 9 pg/mL, while TNFR-1 exhibited a broader range from about 400 pg/mL to over 1400 pg/mL. These varying distributions showed a necessity of normalising the data to mitigate bias towards certain protein markers and ensure

comparability across all markers, which is further discussed in section **3.4, Data Preprocessing**.

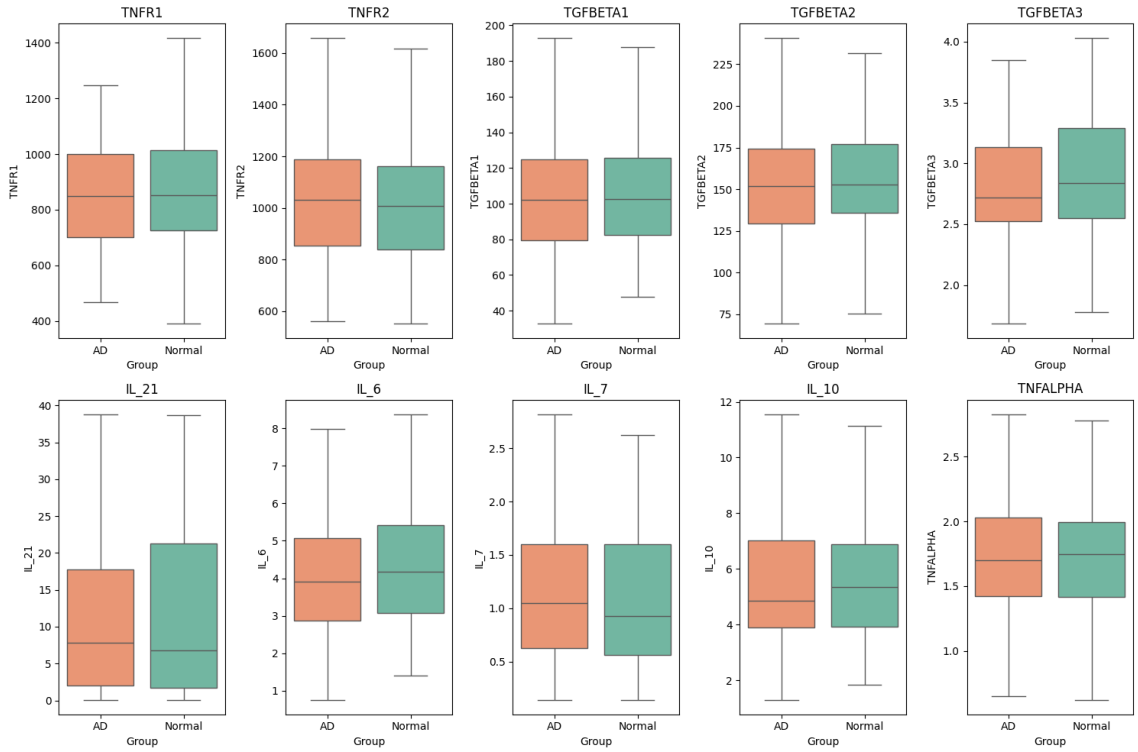

Figure 9: Boxplots to show the inflammatory marker values for AD and normal control patients

3.3.4 Hypothesis

From the initial analysis of the datasets, our formulated hypothesis suggested that combining imaging and inflammation markers as inputs in a deep learning model would improve the classification accuracy of AD. This can be explained by the premise that combining multimodal data can capture a broader range of AD characteristics, such as atrophy and the presence of inflammation markers, which could lead to more precise diagnoses.

For the inflammation markers, from observing Figure 9, it is evident that not all markers demonstrate significant differences in value distribution when comparing AD patients and normal controls. This shows that a selective approach would be necessary to use different subsets of protein markers in order to optimise the model's performance.

3.4 Data Preprocessing

The significant concern when applying preprocessing techniques to medical imaging is the potential loss of crucial information, such as spatial relationships of layers, which could lead to models not being effective for real-world application. To mitigate this risk, ADNI allows the selection of several preprocessing protocols when searching for imaging datasets for the various neuroimaging modalities. These preprocessing steps were chosen to ensure uniformity and accuracy among the data, in order to reduce the chances of bias and inaccurate predictions. The following preprocessing steps discussed were selected in accordance with past research projects, such as those covered in the review by Ebrahimi et al. [20, 21].

3.4.1 MRI Images

For MRI scans, the preprocessing protocols are as follows:

- **Gradwarp**: A correction of image geometric distortion caused by gradient non-linearity. This ensures that spatial information is accurate and maintains the integrity of spatial relationships.
- **B1 non-uniformity**: A correction of non-uniform intensity of the images, which is caused by differing transmission and reception properties of radiofrequency coils in MRI equipment. This ensures the variations of intensity in the images represent actual differences in tissue properties and are not affected by the hardware used.
- **N3 intensity normalisation**: A histogram peak sharpening algorithm applied after grad warp and B1 correction. This reduces non-uniform intensity caused by dielectric effects, to ensure a more homogenous representation of the same tissue types across a scanned volume.

Full details of the ADNI preprocessing protocols for MRI analysis can be seen at https://adni.loni.usc.edu/methods/mri-tool/mri-pre-processing/.

3.4.2 FDG-PET Images

For FDG-PET scans, similar preprocessing protocols are applied by ADNI to enhance the uniformity and comparability of PET scans across different systems. The preprocessing protocols are detailed below:

- **Co-registered dynamic images**: This initial step is applied to raw PET images to convert them into a standard file format and uploaded as DICOM files onto ADNI. The separate frames of the raw images are co-registered to the first frame to mitigate the effects of patient motion.

- **Frame averaging**: This is applied to the co-registered dynamic image set from the previous step by averaging 6 five-minute frames to create a single 30-minute PET image.

- **Image and voxel size standardisation**: This method applies spatial re-orientation and intensity normalisation on the original PET scan to resize the images into a standard 160×160×96 image grid with 1.5 mm cubic voxels. This allows for easier comparisons between PET scans from different scanner models.

- **Uniform resolution**: The final step to smooth the scans to produce a uniform resolution, which is achieved by applying a scanner-specific filter function to adjust the resolution. This ensures PET scans from different scanner models are comparable and minimises resolution disparities.

Full details of the ADNI preprocessing protocols for PET analysis can be seen at https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/.

3.4.3 Additional Imaging Preprocessing

In addition to the ADNI preprocessing protocols, further generic preprocessing steps were implemented to ensure consistency across all images to allow for easy comparability. This was particularly necessary for MRI images, as the dataset had a range of sizes and orientations that required adjustment to ensure uniformity among all scans.

- **Spatial re-orientation**: This method was applied to MRI images due to the variations in axes across the ADNI MRI dataset. The re-orientation adjusted the images to a common shape of (Width, Height, Depth). This adjustment was limited to spatial re-orientation to minimise the risk of losing crucial spatial information between slices.

- **Resizing**: This was also done on the MRI dataset due to the variability in the sizes of the MRI images. To ensure uniformity, all images were zero-padded to match the size of the maximum observed image which had dimensions of 288×288×196.

- **Normalisation**: This was done locally to normalise the image intensities of both the MRI and FDG-PET datasets. Each voxel value was scaled to a range between 0 and 1, which helps to reduce variance in the intensity values and prevents higher values from dominating the learning process.

3.4.4 Inflammation Markers

For the inflammation markers dataset, several cleaning and preprocessing methods were implemented to ensure uniformity and standardised values and to handle any missing data.

- **Handling missing values**: Some entries in the dataset had missing values for specific protein markers. To address this, the dataset was first sorted by acquisition date in descending order, to prioritise the most recent measurements which typically had fewer missing values. After sorting, only three protein markers - IL-9, IP-10 and IL-12-P40 - were missing for around 50 patients. Given the potential inaccuracies of data imputation in a medical context, these three markers were not considered in further analysis during this project.

- **Removing duplicates**: Even after sorting by acquisition date, some patients had multiple test results on the same date. After consulting with Dr Sofia Michopoulou, the decision was made to average the values of these entries to result in single values for each patient.

- **Normalisation**: After applying the cleaning steps, a min-max scaler was used to normalise the values of each protein marker to a range of 0 to 1. This normalisation ensured that each marker contributed equally to the analysis.

It was evident from analysis into the inflammation markers dataset that not all markers had much variance when comparing values for AD patients and normal controls (see Figure 9). In past research studies regarding inflammation markers for Alzheimer's [34, 26, 45], the most prevalent protein markers were IL-6, IL-10, TNF-α, TNFR-1 and TNFR-2. This subset of protein markers were therefore selected to ensure variance between both classes.

3.5 Model Implementation

The main objective of this project is to assess the efficacy of combining neuroimaging and inflammation markers in dementia diagnosis. To achieve this, a range of preprocessing techniques and deep learning classification models were researched and implemented, following the practices in similar research [20, 28, 31]. Convolutional neural networks (CNNs) are a type of deep-layered neural networks commonly used for image classification tasks, as discussed by Ebrahimi et al. [20, 21]. These deep neural networks present the advantage of being able to combine feature extraction and classification which leads to better learning performance compared to training a classifier independently from the feature extraction stage [21].

3.5.1 Residual Neural Networks

In recent research on image classification, evidence shows that deeper CNNs have demonstrated superior classification results [37]. Studies on the ImageNet dataset have shown that networks with larger depths of layers are particularly effective due to their larger parameter space which theoretically increases model adaptability, as shown by Simoyan et al. with their evaluation of 16 to 19-layer VGG models [37]. A challenge presented, however, is a loss in generalisation capability demonstrated by extensively deeper networks, which can be due to the vanishing gradient problem [23, 40], and performance degradation due to saturation of accuracy [24].

Kaiming He et al. [24] introduce a solution to these issues through deep residual learning. This architecture modifies traditional CNNs by utilising skip connections between layers. These connections enable gradients to flow more efficiently through the network, reducing performance degradation due to increased depth. This led to the development of residual neural networks (ResNets), which can be structured with significantly more layers than previous models without increasing in computational complexity.

For example, Kaiming He et al. demonstrated the efficiency of a 34-layer ResNet compared to a 19-layer VGG model by Simonyan et al. [24, 37]. The ResNet model required only 3.6 billion floating point operations (FLOPs), significantly fewer than the 19.6 billion FLOPs of the VGG-19 model. Improved performance was also observed, with a 7.4% error rate on the ImageNet validation set compared to a 9.33% error rate for VGG-16.

Due to the efficient performance properties of ResNet with medical imaging and its usage prevalence in similar past studies, this was the model of choice in this project for the neuroimaging classification.

3.5.2 Proposed Residual Unit Architecture

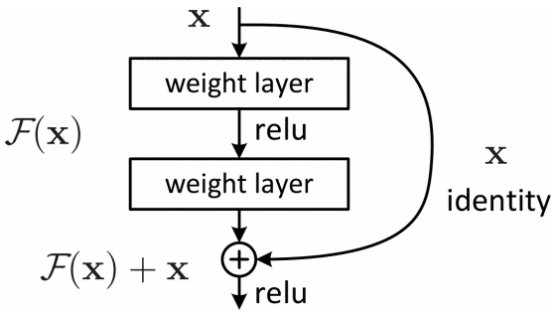

Figure 10: Basic residual unit block with RELU activation [24]

Residual networks are built using a fundamental residual unit architecture, which can be seen in Figure 10 [24]. These units use a skip connection which adds the original input directly to the output of the neural network layer transformations. This modification allows the network to learn an identity function which improves the performance of deeper layers to match shallower layers, which mitigates the performance degradation problem [25]. Formally this can be represented as [24]:

$$\mathbf{y} = \mathcal{F}(\mathbf{x},\ \{W_i\}) + \mathbf{x}. \tag{4}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the input and output vectors of the layers in the residual unit. The function $\mathcal{F}(\mathbf{x}, \{W_i\})$ represents the residual mapping which the model aims to learn which is hypothesised by He et al. to be simpler than learning the original underlying mapping between layers [25]. Since no extra parameters are introduced, the resulting computational complexity is maintained at a level similar to shallower networks.

The residual mapping functions which ResNet models aim to learn are a form of identity mapping which helps reduce the effect of the vanishing gradient problem caused during backpropagation in deeper networks. This simplifies the layers by providing clearer paths for gradient descent which ensures the identity of the input is preserved in deeper layers [25]. This property of residual units make ResNets capable of classifying complex structures, such as medical imaging [21].

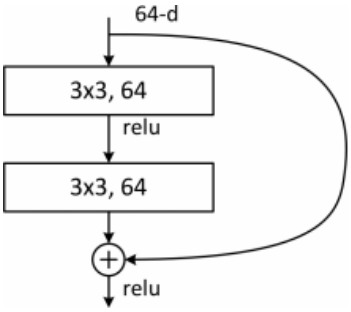

Figure 11: Residual unit structure for a 34-layer ResNet [24]

A standard residual unit discussed by He et al. can be seen in Figure 11 which consists of a convolutional layer followed by a rectified linear unit (ReLU) activation function. This architecture is used in the Keras Python library, which includes an inbuilt ResNet model that currently supports only 2D imaging inputs. Modifications were therefore necessary to handle the 3D neuroimaging datasets used for this project. Following the practices outlined by Karasawa et al. [28] and Ebrahimi et al. [20], the proposed 3D adaptions of the residual unit can be seen in Figure 12.

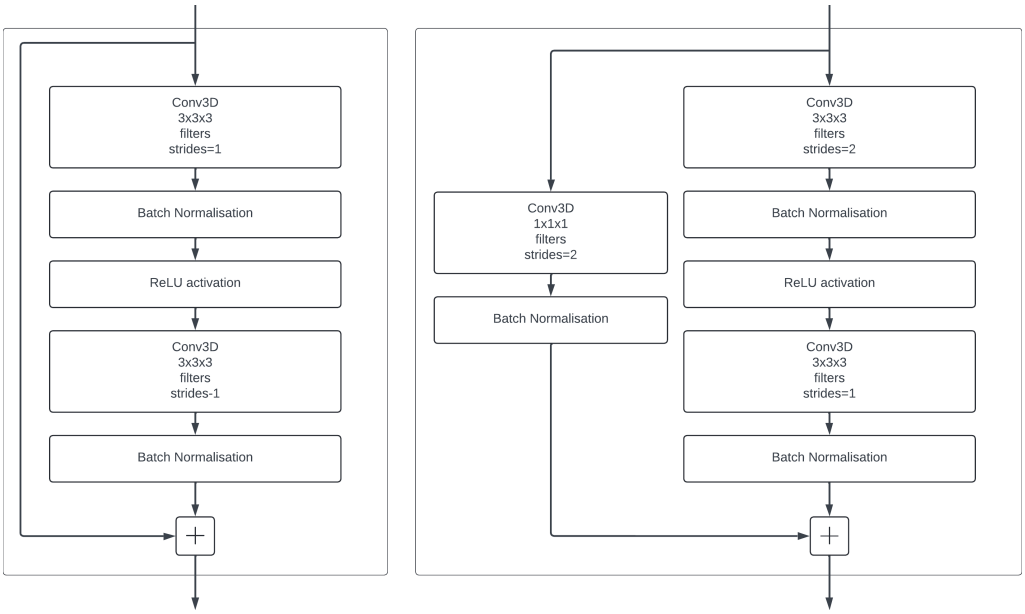

Figure 12: Proposed residual units with 3D convolutional layers

This configuration includes:

- **3D convolutional layers**: Each convolutional layer uses the Keras inbuilt Conv3D layer with a kernel size of 3×3×3 for the convolutional window, as

used by Karasawa et al. [28]. This 3D convolution uses a cubic filter to compute features from an input volume whilst maintaining spatial relationships.

- **Batch normalisation**: As discussed by He et al. [24] and Ebrahimi et al. [20], each convolutional layer is followed by a batch normalisation operation to accelerate the learning process by normalising the output of the previous convolutional layer based on the mini-batch mean and standard deviation.

- **ReLU activation**: This is used after batch normalisation to introduce non-linearity so that the network can learn complex patterns in the data.

Following the structure of the ResNet by He et al. [24], the initial residual units have convolution filters of size 64 and a stride value of 1. The sequence transformation outputs are combined directly with the original input by element-wise addition, as seen in Figure 12 (left). This reflects the skip connection to align with the principles of residual learning.

As the network layers deepen, the convolution filter size is increased by a factor of 2 until 512 filters. Aside from the initial 64-filter residual units, for each later set, the first unit uses a stride value of 2 in order to perform spatial dimensionality reduction to focus on select features of the image and to reduce computational load as depth increases. Since there is a change in dimension shape, to execute the addition at the end requires the original input to also be dimensionally reduced. The skip connection in this case uses an additional convolution layer with kernel volume $1\times1\times1$ and a stride value of 2, followed by a standard batch normalisation, as seen in Figure 12 (right). The remaining residual units in each set use a stride value of 1 which maintains the dimensionality.

The Keras Python library was used to implement this, utilising the inbuilt Conv3D layer.

3.5.3 Proposed Model Architecture

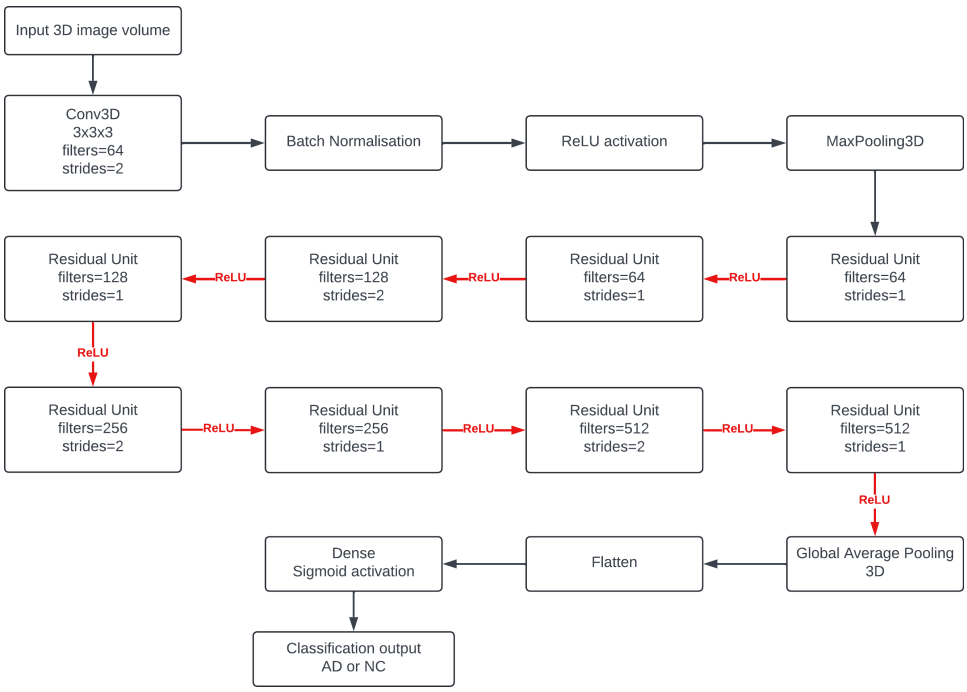

Figure 13: Proposed ResNet-18 model for 3D imaging classification

The proposed model in Figure 13 is a modification based on the ResNet-18 architecture described by Ebrahimi et al. [20]. The model utilises the aforementioned residual units to perform classification on the volumetric input data. The breakdown of the architecture is as follows:

- **Initial layers**: The model applies the 3D input image to a 3D convolutional layer with a filter size of 64 and a kernel window volume of 7×7×7, as practised by He et al. [24]. This reduces the volume size, which is then followed by batch normalisation to stabilise the learning and a ReLU activation function to introduce non-linearity.

- **Max Pooling**: A 3D max pooling layer is applied to further reduce the dimensionality, which allows the model to focus on complex features [20, 24].

- **Residual Units**: The residual units form the core of the model. The ResNet-18 architecture uses a first set of 64-filter size residual units with a stride of 1 to maintain dimensionality. The subsequent sets of units have a filter size of 128, 256 and 512 across 2 units each. The first units in each of

these sets use a stride of 2 to perform spatial dimensionality reduction as the network progresses deeper.

- **Global Average Pooling**: After the final residual unit, a global average pooling layer condenses the 3D feature map to a single scalar by averaging its elements.
- **Output classification**: A final dense layer is applied with the sigmoid activation function to perform binary classification.

A ResNet-34 model was also used in this project to evaluate whether a deeper network would perform improved classification of a dual-modality dataset. This model follows a similar structure as the ResNet-18 model detailed in Figure 13, following the ResNet-34 architecture outlined by He et al. [24]. A differing amount of core residual units were used with 3 sets of 64-filter units, 4 sets of 128-filter units, 6 sets of 256 filter units and 3 sets of 512-filter units.

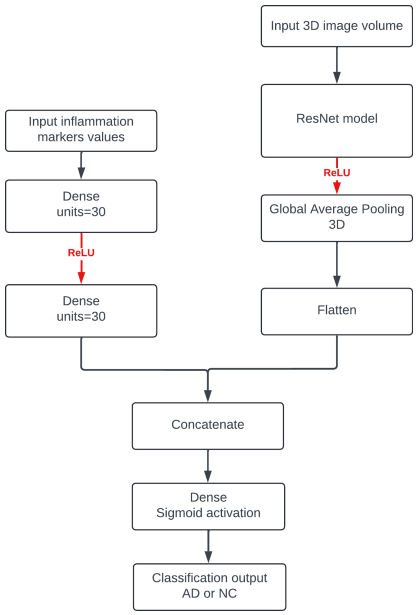

Figure 14: Proposed classification architecture for combined neuroimaging and inflammation marker input

To perform classification with multi-modal datasets, the inflammation markers are input into 2 dense layers to perform dimensionality reduction. A concatenation layer is then used to combine the flattened neuroimaging output and the inflammation markers, which is then input into a final dense layer to perform classification, as exemplified in Figure 14.

3.6 Iterative Improvement

3.6.1 Parameter Optimisation

Backpropagation is a fundamental step when training neural networks to adjust node weights based on the error rate from the previous epochs. Optimisation functions can be used to improve the efficiency of convergence by backpropagation. Inspired by Ebrahimi et al. in their architecture of a 3D ResNet [20], the Stochastic Gradient Descent (SGD) and Adaptive Moment estimation (Adam) optimisers were used to minimise the loss function.

3.6.2 Hyperparameter Tuning

To optimise the performance of the models and achieve the best classification results, the Taguchi method was applied to systematically test combinations of the hyperparameters, as followed by Ebrahimi et al. [20]. The Taguchi method is an experimental methodology to find the minimum number of experiments to be performed with a range of different factors [39]. This allows for conducting robust design experiments using orthogonal arrays to study large combinations of variables with a reduced number of trials.

Exp	Model	Optimizer	Learning Rate	Batch Size	Dataset
1	ResNet18	SGD	0.0005	2	Imaging
2	ResNet18	SGD	0.0001	6	Combined
3	ResNet18	Adam	0.0005	6	Imaging
4	ResNet18	Adam	0.0001	2	Combined
5	ResNet34	SGD	0.0005	6	Combined
6	ResNet34	SGD	0.0001	2	Imaging
7	ResNet34	Adam	0.0005	2	Combined
8	ResNet34	Adam	0.0001	6	Imaging

Table 1: L8 Orthogonal Array for Hyperparameter Tuning

For this project, the Taguchi method was used to find the optimal combination of hyperparameters without having to exhaustively test every combination, which would have been computationally expensive. Table 1 shows the L8 orthogonal array used for this project, which tests combinations of optimiser choice, learning rate, batch size and input dataset. This is then followed by an analysis of the performance metrics by calculating the signal-to-noise (S/N) ratio for each metric which measures the robustness of each trial. The signal-to-noise ratio (S/N ratio) for the "higher-the-better" scenario is given by [39]:

$$\text{S/N Ratio} = -10 \log_{10} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{y_i^2} \right) \tag{5}$$

where y_i represents the performance metric value for each trial, and n is the number of trials. The S/N ratios for each factor level were then averaged across all trials where that level was used, with the highest average S/N ratios being selected as the optimal values. A final validation run was performed with the optimal values obtained for both neuroimaging and combined dataset inputs to evaluate the effectiveness of both.

In an effective implementation of the Taguchi method, multiple trails of the different configurations are recommended. However, due to the time and computational constraints of this project, only 2 trials were run per configuration, with the results then being averaged before calculating the S/N ratios.

4 Results & Evaluation

The 8 trials outlined in Table 1 followed a consistent data split of 70% training, 15% validation and 15% testing. Training was done on the Google Collab Jupyter environment using the NVIDIA L4 GPU runtime, which provided 22.5 GB of GPU RAM and 62.8 GB of system RAM. The training process followed the methodology of Ebrahimi et al. [20] and Michopoulou et al. [31], by training until 100% accuracy was achieved on both training and validation sets, or when the validation loss did not improve for 15 consecutive epochs. The training dataset was reshuffled after each epoch to mitigate overfitting. In general, each epoch took around 45 to 90 seconds with the model usually running for all 50 epochs.

Initially, the plan was to train with both MRI and FDG-PET as separate input options for neuroimaging. However, likely due to the poor quality of the FDG-PET dataset, the metrics were usually between a random range of 0% to 50%, which did not improve with different configurations.

4.1 Evaluation Metrics

A consistent set of metrics as those used by Ebrahimi et al. [20] and Michopoulou et al. [31] were used to perform evaluation. This was chosen since the architecture of the model for this project and the overall goal was similar to those studies, so a similar evaluation approach was taken.

- **Accuracy**: This metric shows the correctness of the predictions by the model, which is calculated using Equation 1. Accuracy provides a quick snapshot into a model's efficacy across all cases.
- **Sensitivity**: The proportion of AD cases that are correctly predicted by the model as seen in Equation 2. In a medical context, high sensitivity is vital to minimise false negatives.
- **Specificity**: The proportion of correctly identified cognitively normal patients as seen in Equation 3. High specificity shows a lesser chance of false positive predictions.
- **Area under ROC curve**: This metric shows the likelihood of a model correctly classifying between positive and negative instances. A higher AUC shows a balance between sensitivity and specificity across various thresholds.

These metrics were analysed using the Taguchi method to find the optimal hyperparameter configurations from the L8 orthogonal array in Table 1. The results

are summarised in Table 2 which shows the scores for each metric across the different configurations. The S/N ratio for each configuration was then calculated for each metric, which is detailed in Table 3. Based on these results, the optimal configuration was found to be the ResNet-18 model with the Adam optimiser, a learning rate of 0.0001, and a batch size of 6.

Exp	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
1	0.629	0.714	0.600	0.67
2	0.519	0.900	0.294	0.64
3	0.741	0.700	0.765	0.88
4	0.667	0.600	0.706	0.85
5	0.667	0.500	0.765	0.78
6	0.556	0.438	0.727	0.65
7	0.407	0.417	0.400	0.47
8	0.593	0.857	0.500	0.83

Table 2: Performance Metrics for Each Configuration

Exp	**Accuracy S/N**	**Sensitivity S/N**	**Specificity S/N**	**AUC S/N**
1	-4.027	-2.926	-4.437	-3.479
2	-5.697	-0.915	-10.633	-3.876
3	-2.604	-3.098	-2.327	-1.110
4	-3.517	-4.437	-3.024	-1.412
5	-3.517	-6.021	-2.327	-2.158
6	-5.099	-7.171	-2.769	-3.742
7	-7.808	-7.597	-7.959	-6.558
8	-4.539	-1.340	-6.021	-1.618

Table 3: S/N Ratios for Performance Metrics

Factor	Level	Average S/N Ratio (dB)
Model	ResNet18	-3.595
Model	ResNet34	-4.765
Optimizer	SGD	-4.300
Optimizer	Adam	-4.061
Learning Rate	0.0001	-4.113
Learning Rate	0.0005	-4.247
Batch Size	2	-4.748
Batch Size	6	-3.613
Dataset	MRI	-3.519
Dataset	Combined	-4.841

Table 4: Average Composite S/N Ratios for Each Factor Level

4.2 Performance on Neuroimaging Input

The performance of the previously mentioned optimal model was evaluated by tracking the model binary accuracy and loss per epoch for both training and validation sets, as seen in Figure 15. A relatively steady training accuracy and loss was observed, with the accuracy surpassing 80% before the 15th epoch, and 100% by the end of training, and low loss throughout. The validation metrics were more variable with common dips of low accuracy and high loss observed, but during the later epochs the validation set also displayed high accuracy and low loss.

The fluctuations in validation accuracy suggest that the model was overfitting during the training process. This can be explained due to the small dataset being used.

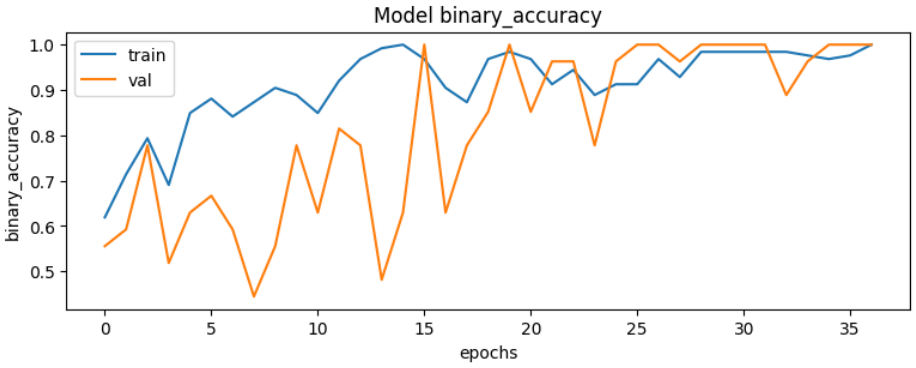

Figure 15: Model Binary Accuracy of Training and Validation sets per Epoch with standalone MRI input

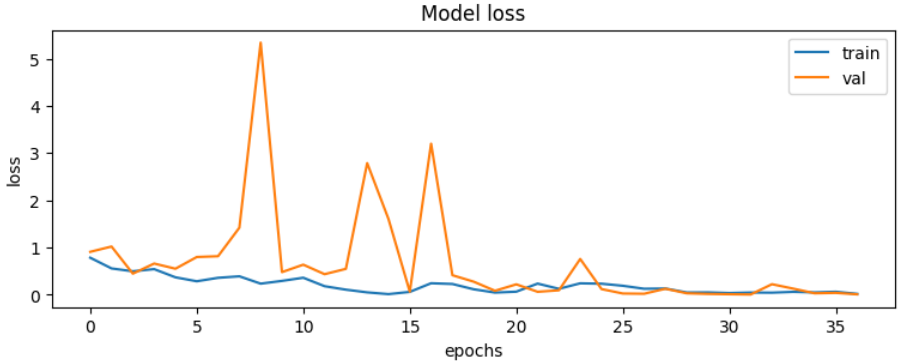

Figure 16: Model Loss Value in Training and Validation sets per Epoch with standalone MRI input

The optimal model was then run on a testing set to obtain the values shown in the confusion matrix in Figure 17. The metrics derived from the confusion matrix are as follows:

- **Accuracy**: 78.6%, which shows the model's overall correctness.
- **Sensitivity**: 85.7%, indicates strong likelihood of identifying true AD cases and minimising false negatives.
- **Specificity**: 71.4%, shows slightly weaker performance for identifying true positives.

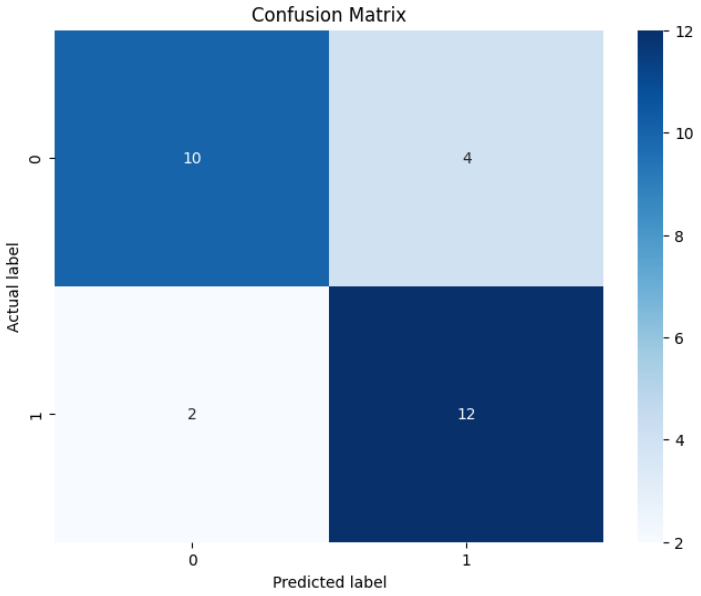

Figure 17: Confusion Matrix of Optimal ResNet Model with Standalone MRI Input

4.3 Performance on Combined Input

The performance of the combined MRI and inflammation markers dataset was slightly worse, as shown in Figure 18. Whilst the training set showed steady improvement with generally high accuracy and low loss, the model seemed to perform worse on the validation set. This indicates the model had problems with generalisation when using the combined dataset, which could be due to the added complexity of combining and classifying multi-modal datasets. The small size of the dataset was presented the issue of overfitting which can be seen by the high training accuracy throughout, but relatively low for the validation set.

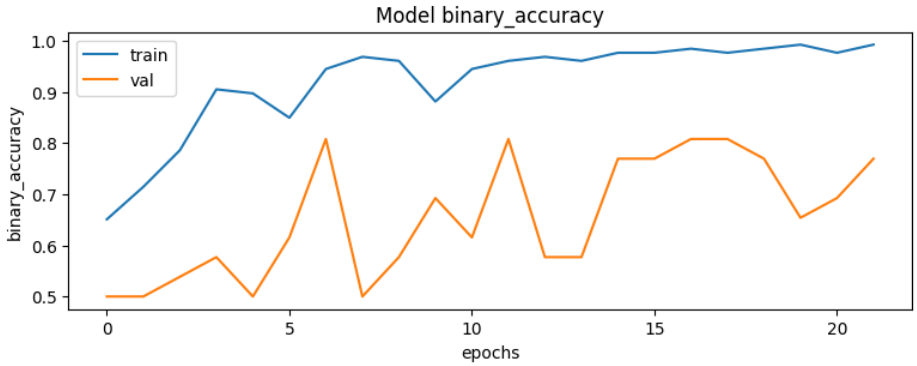

Figure 18: Model Binary Accuracy of Training and Validation sets per Epoch with Combined MRI and Inflammation Markers Dataset

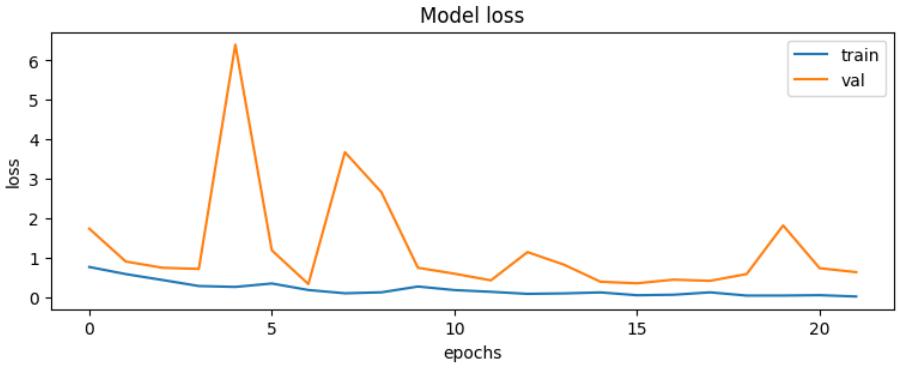

Figure 19: Model Loss Value in Training and Validation sets per Epoch Combined MRI and Inflammation Markers Dataset

The optimal model on the combined dataset performed slightly worse than the standalone MRI, albeit still producing acceptable results. The confusion matrix in Figure 20 details these results, with the metrics achieved as follows:

- **Accuracy**: 75%, slightly lower than the MRI dataset but still generally makes correct predictions
- **Sensitivity**: 78.6%, although lower than with standalone MRI input, sensitivity is still high which shows less likelihood of false negatives.
- **Specificity**: 71.4%, identical to the MRI-only input which shows generally strong performance for predicting negative cases.

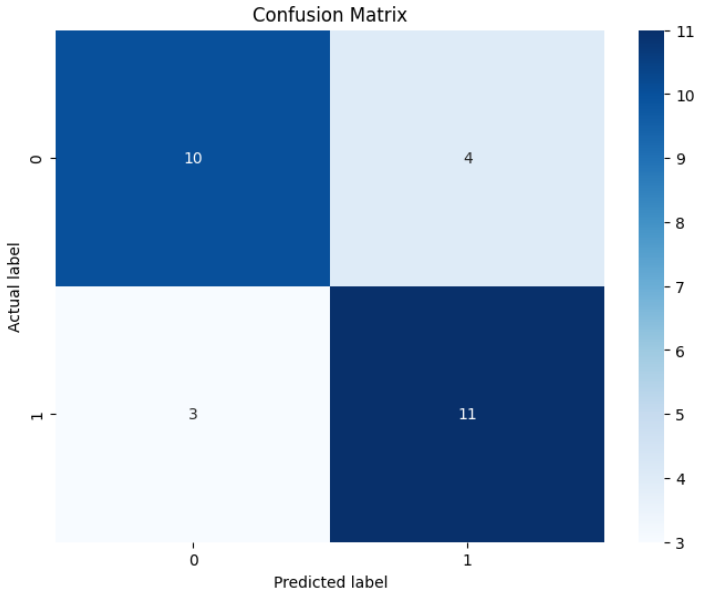

Figure 20: Confusion Matrix of Optimal ResNet Model with Combined MRI and Inflammation Markers Dataset

4.4 Comparative Analysis

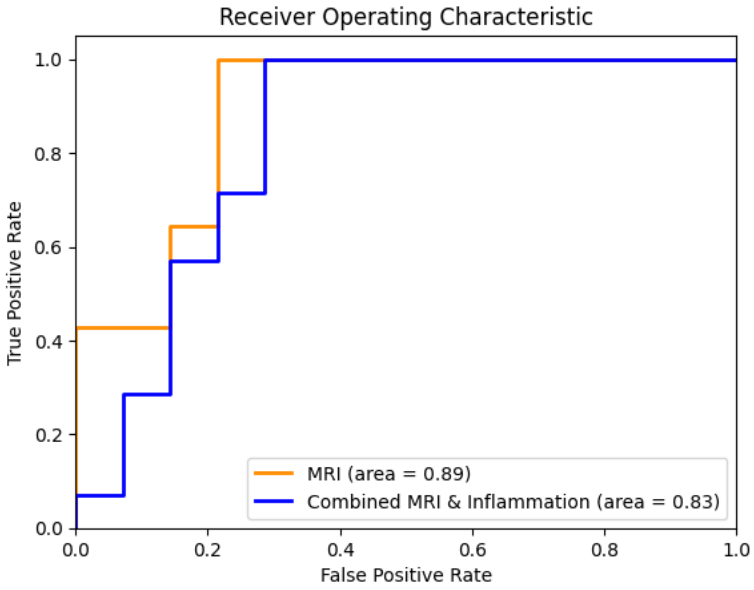

Figure 21: ROC Curve for both Models

The ROC curve in Figure 21 compares the performance of both the standalone MRI and combined datasets. The MRI-only dataset achieved a higher AUC score of 0.89 showing better classification ability across various thresholds compared to the combined dataset's score of 0.83. This suggests that MRI scans are robust in classifying AD, as was well-documented in past research such as Ebrahimi et al. [20].

The initial hypothesis predicted that combining with inflammation markers would improve diagnostic accuracy by capturing a broader range of AD features, but the results show otherwise. Although the combined dataset still performed with relatively high metrics, the difference in performance could be due to the added complexity of processing multi-modal data, especially in a context of 3D volumes with additional values.

Furthermore, overfitting was observed in both setups, shown by the high training accuracy compared to lower validation and testing performances. This is more emphasised with the limitation of a smaller dataset, which reduces the model's capability to generalise to new data.

Model	Accuracy	Sensitivity	Specificity	AUC
MRI Input	78.6%	85.7%	71.4%	0.89
Combined Input	75%	78.6%	71.4%	0.83

Table 5: Final Model Results for MRI Input and Combined Input

Despite the difference, both inputs showed relatively high metric scores as seen in Table 5, which exemplifies the potential of AI in this field. To determine if the observed differences were statistically significant, paired t-tests were performed on the 4 metrics obtained across all configuration runs. The results show:

- **Accuracy:** t-statistic = 1.027, p-value = 0.380.
- **Sensitivity:** t-statistic = 0.538, p-value = 0.628.
- **Specificity:** t-statistic = 1.474, p-value = 0.237.
- **AUC:** t-statistic = 0.704, p-value = 0.532.

Using a significance level of 0.05, the analysis shows that the difference of scores were not statistically significant. This shows that the addition of inflammation markers may not lead to significant improvement in diagnostic performance.

5 Project Management

Due to the scope of this project, there had to be a strong focus on careful project management to ensure the goals of the project were achieved in a suitable time frame. Several project management techniques such as agile methodology and Gantt charts were used to track progress.

5.1 Agile Methodology

Sprint meetings were conducted every week with the project supervisor who acted as the owner stakeholder. These meetings discussed the progress of work completed during the week and plans for the next week. The meetings also allowed for any issues to be solved by consulting with the supervisor who gave technical and background advice for the project. Google Docs were used to take notes and monitor progress throughout the project.

5.2 Risk Analysis

Before starting key areas of the project, a risk assessment was undertaken to ensure solutions were available on the off-chance a delay occurred. The biggest delay for this project was the failure of the FDG-PET dataset for classification, which was originally considered a low probability risk due to its efficacy in past projects. Another delay was due to personal reasons arising from a surgery during the year which slowed potential model training and analysis during that period. Computational constraints were considered a high probability risk since this project would require high computational power. Originally, an attempt was made to use the University's IRIDIS cluster for training. However, due to issues with transferring the datasets, Google Colab was used instead.

5.3 Gantt Chart

The Gantt chart for this project can be seen in Figure 22, which was used to track progress throughtout the project timeline. There were some differences to the previous Gantt chart seen in Appendix A, Figure 24, due to changes in project focus as more research was done in regards to model implementation. The final report was written alongside model training as a continuous process so that the progress could be recorded as the model was developed resulting in more detail.

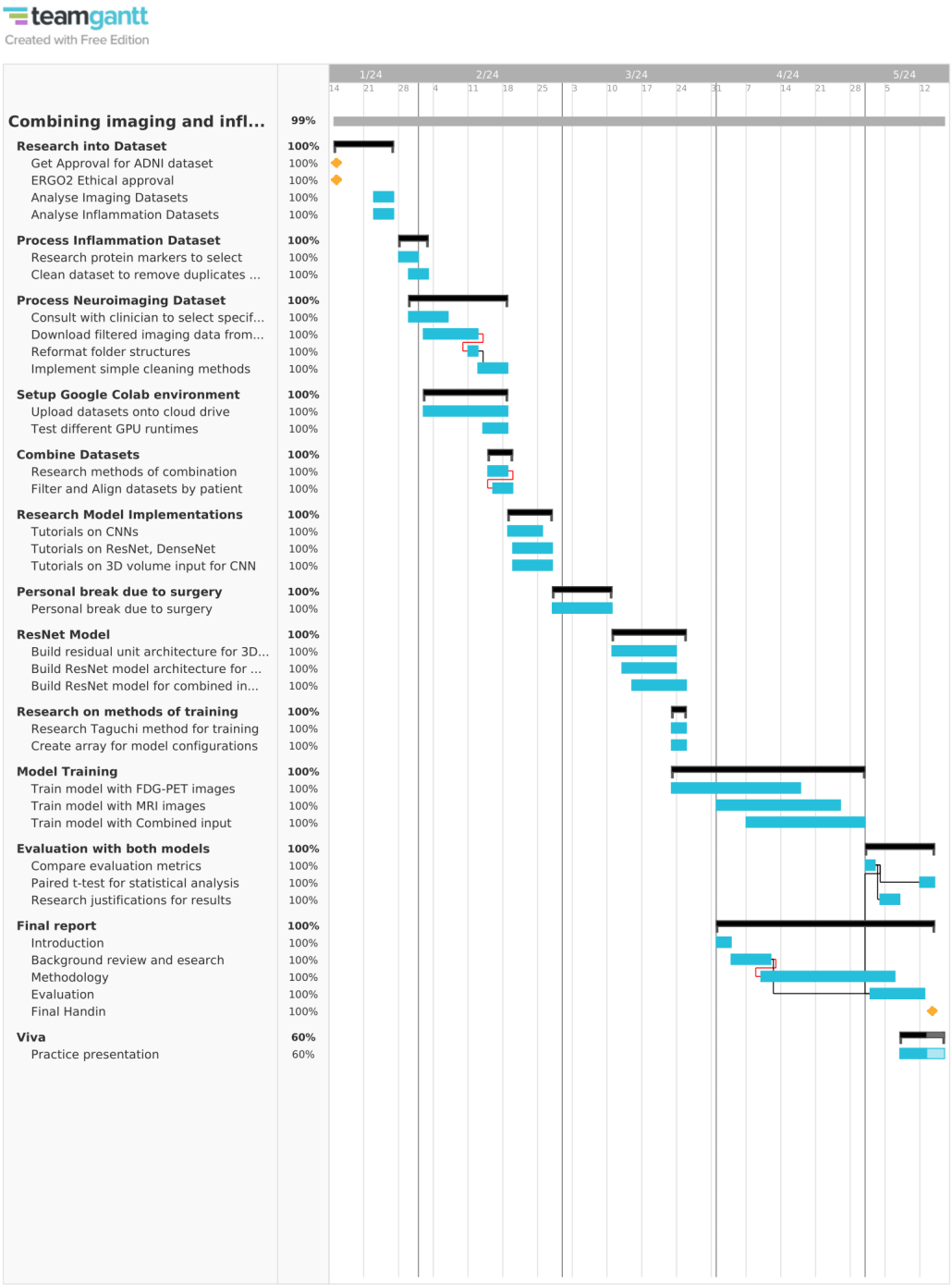

Figure 22: Final Gantt Chart for Project Plan and Estimated Progress in Semester 2

6 Conclusion & Future Work

The project detailed a study into the effectiveness of neuroimaging and inflammatory protein markers in diagnosing dementia. With a focus on Alzheimer's disease, an array of neuroimaging modalities, including MRI, FDG-PET and Amyloid-PET, were utilised alongside neural network techniques. The modalities were processed as 3D volumes provided by ADNI, alongside inflammation markers known to be associated with Alzheimer's.

Throughout this project, substantial research was conducted to develop an understanding of dementia and its impact on society. The commitment was to implement medically relevant techniques which would form the basis of future research. Background research included past reviews of inflammation markers and different modalities to identify a range of characteristics pathognomonic to Alzheimer's. Despite the medical aspect being beyond the initial scope, foundational research helped to make informed decisions during the development phase, particularly for choosing specific modalities and methods for implementing the classification.

The final dataset comprised of 90 AD patients and 110 normal controls, alongside 10 protein markers which were subset to 5 specific markers most relevant in research. Various configurations were tested using a modified ResNet architecture designed for 3D volumes and multi-modal inputs. The Taguchi method was utilised during the testing to reduce computational costs. Standalone MRI exhibited higher performance, achieving accuracy, sensitivity, and specificity of 78.6%, 85.7%, and 71.4%, respectively, with an AUC score of 0.89. The combined inputs displayed slightly lower results, with accuracy of 75%, sensitivity of 78.6%, and specificity of 71.4%, and an AUC score of 0.83, the discrepancy likely due to overfitting the small sample size. Statistical analysis using paired t-tests showed that the difference in performance was not significant to make any final conclusions regarding the effect of adding inflammation markers.

Despite the initial challenges, the results show the potential of AI in medical diagnostics. For future work, expanding hyperparameter tuning and exploring additional preprocessing strategies could possibly improve results. In similar past projects, techniques such as data augmentation, regularisation with Dropout layers and variable learning rates have shown promise. Computational and time constraints limited additional testing during this project, but these techniques show valuable pathways for future research to yield better results.

7 References

[1] National Institute on Aging. *What is Dementia? Symptoms, Types, and Diagnosis*. https://www.nia.nih.gov/health/alzheimers-and-dementia/what-dementia-symptoms-types-and-diagnosis. Dec. 2022.

[2] F. Agosta, F. Caso, and M. Filippi. "Dementia and neuroimaging". In: *Journal of Neurology* 260 (2013), pp. 685–691. DOI: 10.1007/s00415-012-6778-x.

[3] S. Ahmadi-Abhari et al. "Temporal trend in dementia incidence since 2002 and projections for prevalence in England and Wales to 2040: modelling study". In: *The BMJ* 358 (2017). DOI: 10.1136/bmj.j2856.

[4] Md Rishad Ahmed et al. "Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects". In: *IEEE Reviews in Biomedical Engineering* 12 (2019), pp. 19–33. DOI: 10.1109/RBME.2018.2886237.

[5] A Anoop et al. "CSF Biomarkers for Alzheimer's Disease Diagnosis". en. In: *Int J Alzheimers Dis* 2010 (June 2010).

[6] Zoe Arvanitakis, Raj C. Shah, and David A. Bennett. "Diagnosis and Management of Dementia: Review". In: *JAMA* 322.16 (Oct. 2019). Accessed: 2023-12-19. ISSN: 0098-7484. DOI: 10.1001/jama.2019.4782. URL: https://doi.org/10.1001/jama.2019.4782.

[7] Akhilesh Deep Arya et al. "A systematic review on machine learning and deep learning techniques in the effective diagnosis of Alzheimer's disease". en. In: *Brain Inform* 10.1 (July 2023).

[8] Lílian Viana Dos Santos Azevedo et al. "Impact of Social Isolation on People with Dementia and Their Family Caregivers". In: *Journal of Alzheimer's disease : JAD* (2021). DOI: 10.3233/JAD-201580.

[9] Dipanjan Banerjee et al. "Neuroimaging in Dementia: A Brief Review". en. In: *Cureus* 12.6 (June 2020). Accessed: 2023-12-19.

[10] Gopi Battineni et al. "Artificial Intelligence Models in the Diagnosis of Adult-Onset Dementia Disorders: A Review". en. In: *Bioengineering (Basel)* 9.8 (Aug. 2022). Accessed: 2023-12-20.

[11] Valentina Berti, Alberto Pupi, and Lisa Mosconi. "PET/CT in diagnosis of dementia". en. In: *Ann N Y Acad Sci* 1228 (June 2011). Accessed: 2023-12-19.

[12] Shyamal C Bir et al. "Emerging Concepts in Vascular Dementia: A Review". en. In: *J Stroke Cerebrovasc Dis* 30.8 (May 2021), p. 105864.

[13] R. Borchert et al. “Artificial intelligence for diagnosis and prognosis in neuroimaging for dementia; a systematic review”. In: (2021). DOI: 10.1101/2021.12.12.21267677.

[14] Henry Brodaty and Marika Donkin. “Family caregivers of people with dementia”. In: *Dialogues in Clinical Neuroscience* 11 (2009), pp. 217–228. DOI: 10.31887/DCNS.2009.11.2/hbrodaty.

[15] Henry Brodaty and Marika Donkin. “Family caregivers of people with dementia”. In: *Dialogues in Clinical Neuroscience* 11.2 (2009). PMID: 19585957, pp. 217–228. DOI: 10.31887/DCNS.2009.11.2/hbrodaty. eprint: https://doi.org/10.31887/DCNS.2009.11.2/hbrodaty. URL: https://doi.org/10.31887/DCNS.2009.11.2/hbrodaty.

[16] Keystone Clinic. *Dementia*. https://keystonemedical.com.sg/dementia/. Accessed: 2023-11-22. 2023.

[17] Corinna Cortes and V. Vapnik. “Support-Vector Networks”. In: *Machine Learning* 20 (1995), pp. 273–297. DOI: 10.1023/A:1022627411411.

[18] Mark Dashwood et al. “Artificial intelligence as an aid to diagnosing dementia: an overview”. In: *Progress in Neurology and Psychiatry* 25 (2021). DOI: 10.1002/pnp.721.

[19] Silvia Duong, Tejal Patel, and Feng Chang. “Dementia: What pharmacists need to know”. en. In: *Can Pharm J (Ott)* 150.2 (Feb. 2017). Accessed: 2023-12-20.

[20] Amir Ebrahimi, S. Luo, and R. Chiong. “Introducing Transfer Leaming to 3D ResNet-18 for Alzheimer’s Disease Detection on MRI Images”. In: *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)* (2020), pp. 1–6. DOI: 10.1109/IVCNZ51579.2020.9290616.

[21] Mr Amir Ebrahimighahnavieh, Suhuai Luo, and Raymond Chiong. “Deep learning to detect Alzheimer’s disease from neuroimaging: A systematic literature review”. In: *Computer Methods and Programs in Biomedicine* 187 (2020), p. 105242. DOI: https://doi.org/10.1016/j.cmpb.2019.105242. URL: https://www.sciencedirect.com/science/article/pii/S0169260719310946.

[22] Mark A Findeis. “The role of amyloid beta peptide 42 in Alzheimer’s disease”. en. In: *Pharmacol Ther* 116.2 (July 2007).

[23] Xavier Glorot and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks". In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Vol. 9. Proceedings of Machine Learning Research. PMLR, 2010, pp. 249–256. URL: https://proceedings.mlr.press/v9/glorot10a.html.

[24] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015), pp. 770–778. DOI: 10.1109/cvpr.2016.90.

[25] Kaiming He et al. *Identity Mappings in Deep Residual Networks*. 2016. arXiv: 1603.05027 [cs.CV].

[26] Michael T Heneka et al. "Neuroinflammation in Alzheimer's disease". en. In: *Lancet Neurol* 14.4 (Apr. 2015).

[27] William T. Hu and J. Christina Howell. *CSF Inflammatory Proteins Methods - Hu Lab*. Available from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI, 2011. URL: https://adni.loni.usc.edu/data-samples/access-data/.

[28] Hiroki Karasawa, Chien-Liang Liu, and Hayato Ohwada. "Deep 3D Convolutional Neural Network Architectures for Alzheimer's Disease Diagnosis". In: *Intelligent Information and Database Systems*. Ed. by Ngoc Thanh Nguyen et al. Cham: Springer International Publishing, 2018, pp. 287–296. ISBN: 978-3-319-75417-8.

[29] Jefferson W Kinney et al. "Inflammation as a central mechanism in Alzheimer's disease". en. In: *Alzheimers Dement (N Y)* 4 (Sept. 2018).

[30] Gemma Lombardi et al. "Structural magnetic resonance imaging for the early diagnosis of dementia due to Alzheimer's disease in people with mild cognitive impairment". en. In: *Cochrane Database Syst. Rev.* 3 (Mar. 2020), p. CD009628.

[31] Sofia Michopoulou et al. "Perfusion Imaging and Inflammation Biomarkers Provide Complementary Information in Alzheimer's Disease". en. In: *J Alzheimers Dis* 96.3 (2023), pp. 1317–1327.

[32] L Mosconi et al. "Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease". en. In: *Neurology* 72.6 (Feb. 2009), pp. 513–520.

[33] S. Muhammed et al. "Improved Classification of Alzheimer's Disease With Convolutional Neural Networks". In: *2023 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)* (2023), pp. 1–7. DOI: 10.1109/SPMB59478.2023.10372725.

[34] Jong-Chan Park, Sun-Ho Han, and Inhee Mook-Jung. "Peripheral inflammatory biomarkers in Alzheimer's disease: a brief review". en. In: *BMB Rep* 53.1 (Jan. 2020).

[35] Melinda C Power et al. "Risks and Benefits of Clinical Diagnosis Around the Time of Dementia Onset". en. In: *Gerontol Geriatr Med* 9 (Nov. 2023), p. 23337214231213185.

[36] Pedro Rosa-Neto et al. "Fluid biomarkers for diagnosing dementia: rationale and the Canadian Consensus on Diagnosis and Treatment of Dementia recommendations for Canadian physicians". en. In: *Alzheimers Res Ther* 5.Suppl 1 (Nov. 2013). Accessed: 2023-12-19.

[37] K. Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *CoRR* abs/1409.1556 (2014).

[38] Social Care Institute for Excellence. *Dementia: At a glance*. `https://www.scie.org.uk/dementia/about/`. Accessed: 2023-11-23. 2020.

[39] Genichi Taguchi. *Introduction to Quality Engineering: Designing Quality into Products and Processes*. Asian Productivity Organization, 1986.

[40] Kshitij Tripathi, A. Gupta, and R. Vyas. "Deep Residual Learning for Image Classification using Cross Validation". In: *International Journal of Innovative Technology and Exploring Engineering* (2020). DOI: `10.35940/ijitee.f4131.049620`.

[41] Azlheimer's Research UK. *Subtypes of Dementia Statistics*. `https://dementiastatistics.org/about-dementia/subtypes/`. Accessed: 2024-04-22. Nov. 2019.

[42] Yudthaphon Vichianin et al. "Accuracy of Support-Vector Machines for Diagnosis of Alzheimer's Disease, Using Volume of Brain Obtained by Structural MRI at Siriraj Hospital". en. In: *Front Neurol* 12 (May 2021), p. 640696.

[43] D. van Vliet et al. "Time to diagnosis in young-onset dementia as compared with late-onset dementia". In: *Psychological Medicine* 43.2 (2013). Accessed: 2023-12-19. DOI: `10.1017/S0033291712001122`.

[44] Susanne Wegmann, Jacek Biernat, and Eckhard Mandelkow. "A current view on Tau protein phosphorylation in Alzheimer's disease". en. In: *Curr Opin Neurobiol* 69 (Apr. 2021).

[45] A. Wennberg et al. "The Cross-Sectional and Longitudinal Associations Between IL-6, IL-10, and TNF-alpha and Cognitive Outcomes in the Mayo Clinic Study of Aging". In: *The journals of gerontology. Series A, Biological sciences and medical sciences* (2018). DOI: `10.1093/gerona/gly217`.

[46] Raphael Wittenberg et al. *Projections of older people with dementia and costs of dementia care in the United Kingdom, 2019–2040*. Accessed: 2023-11-23. Nov. 2019. URL: https://www.alzheimers.org.uk/sites/default/files/201911/cpec_report_november_2019.pdf.

[47] Daniel E Wollman and Isak Prohovnik. "Sensitivity and specificity of neuroimaging for the diagnosis of Alzheimer's disease". en. In: *Dialogues Clin Neurosci* 5.1 (Mar. 2003).

[48] Kuo Yang and E. Mohammed. "A Review of Artificial Intelligence Technologies for Early Prediction of Alzheimer's Disease". In: *ArXiv* abs/2101.01781 (2020).

[49] Milica Živanovic et al. "The role of magnetic resonance imaging in the diagnosis and prognosis of dementia". en. In: *Biomol Biomed* 23.2 (Mar. 2023), pp. 209–224.

Appendix A

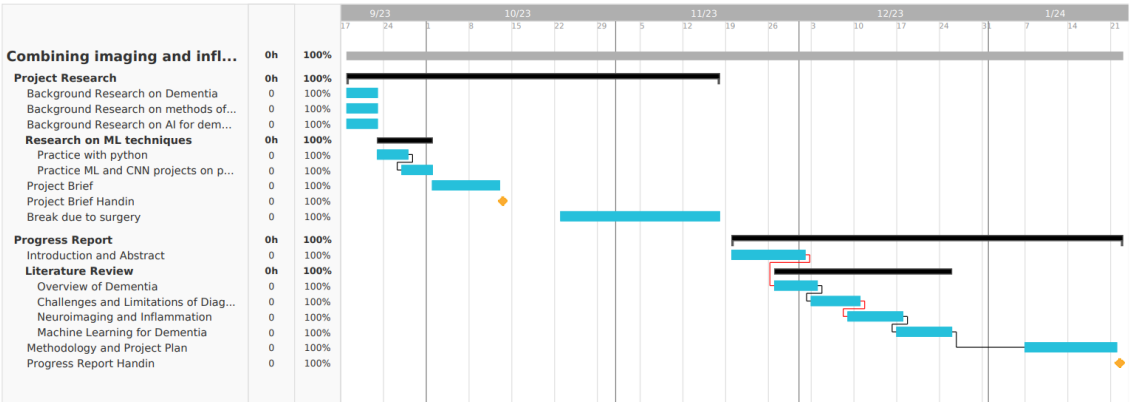

Figure 23: Gantt Chart for Project Schedule from September to Progress Report Handin

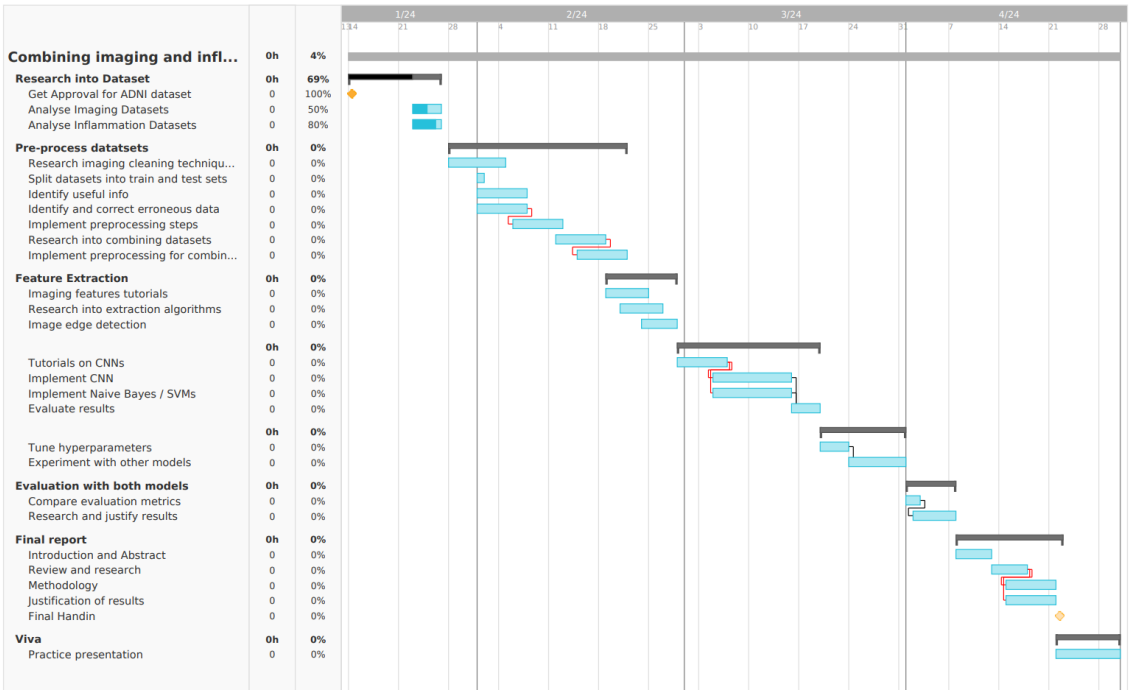

Figure 24: Old Planned Gantt Chart for Project Schedule for Semester 2

Appendix B

B.1 Problem

Dementia is a progressive neurodegenerative disease that causes loss of cognitive function eventually leading to disability and death. Dementia diagnosis is a complex, subjective and slow process. It is estimated that it takes upwards of 2 years on average from symptom onset to diagnosis. There is a clear need for objective methods for early diagnosis of dementia and prediction of the likelihood of progression.

Positron Emission Tomography (PET) and Single Photon Emission Computed Tomography (SPECT) scans can support diagnosis of dementia by identifying changes in brain metabolism and perfusion. Similarly, inflammation markers in the blood and cerebrospinal fluid may help identify patients at risk of progression of dementia. The problem at hand is trying to ensure a faster way of interpreting imaging data and providing diagnosis for patients. Artificial Intelligence (AI) can be used to provide rapid and accurate image interpretation to improve diagnosis and workflow for healthcare systems, and may be more accurate and efficient than using imaging markers alone.

B.2 Goals

The main functionality of the model will be to classify patients according to their risk of progression, using PET & SPECT imaging scans and inflammation markers, and then to evaluate the ability of this model. The data will be retrieved from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public dataset and local data from the Biomarker Assessment of Inflammation in Neurodegeneration (BRAIN) study collected at University Hospital Southampton (UHS).
The breakdown of goals is as follows:

- Retrieve and perform feature extraction on relevant data from ADNI public dataset and UHS studies
- Develop a machine learning model using algorithms such as convolutional neural networks (CNNs) to perform classification on the dataset
- Evaluate the model by splitting the dataset into training, validation and test sets
- Potentially present the findings to imaging physics and neuroimmunology experts at UHS

B.3 Scope

The scope of this project includes but is not limited to:

- Research into existing machine learning models for dementia diagnosis
- Data retrieval from ADNI public database and UHS studies
- Development and evaluation of a machine learning model to classify patient data

The project will **not** include:

- New data collection
- Clinical diagnosis or real-time processing use on current patients