

Pravir Kumar

Thesis Plag (2)

 Nishita Singh

Document Details

Submission ID**trn:oid:::27535:138632229****Submission Date****May 12, 2026, 12:00 PM GMT+5:30****Download Date****May 12, 2026, 12:11 PM GMT+5:30****File Name****Thesis Plag (2).pdf****File Size****4.9 MB****33 Pages****10,477 Words****64,032 Characters**

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

-  **44 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 4%  Internet sources
- 2%  Publications
- 3%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 44 Not Cited or Quoted 5%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 0 Missing Citation 0%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 4% Internet sources
- 2% Publications
- 3% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	www.frontiersin.org	<1%
2	Student papers	University of Sunderland on 2025-01-09	<1%
3	Internet	www.mdpi.com	<1%
4	Internet	www.techscience.com	<1%
5	Student papers	Istanbul Bilgi University on 2025-01-03	<1%
6	Student papers	Kaplan College on 2024-12-16	<1%
7	Student papers	Guru Ghasidas Vishwavidyalaya, Bilaspur on 2026-05-08	<1%
8	Student papers	Pondicherry University on 2023-05-10	<1%
9	Internet	admin.lidsen.com	<1%
10	Student papers	bsmrstu on 2025-06-06	<1%

11	Publication	Jha, Anshul. "Neural Networks for Intelligent Healthcare: Transfer Learning and ..."	<1%
12	Internet	d-nb.info	<1%
13	Internet	safeandsecure1.blogspot.com	<1%
14	Internet	spandidos-publications.com	<1%
15	Internet	www.biorxiv.org	<1%
16	Publication	Rifat Burak Ergül, Poyraz Doğan, Muhammed Edip Görür, Mert Emre Erden et al. "..."	<1%
17	Student papers	University of Hull on 2025-12-14	<1%
18	Internet	content.iospress.com	<1%
19	Internet	pubmed.ncbi.nlm.nih.gov	<1%
20	Internet	researchoutput.csu.edu.au	<1%
21	Internet	www.biomedcentral.com	<1%
22	Publication	Li, Qian. "Unraveling the Contribution and Molecular Phenotypes of Bone Marro..."	<1%
23	Internet	drepo.sdl.edu.sa	<1%
24	Internet	ir.vanderbilt.edu	<1%

25	Student papers	BB9.1 PROD on 2025-08-26	<1%
26	Publication	Xiaoshuang Chen, Yin Zheng, Peilin Zhao, Zhuxi Jiang, Wenye Ma, Junzhou Huang....	<1%
27	Internet	bpspubs.onlinelibrary.wiley.com	<1%
28	Internet	dspace.daffodilvarsity.edu.bd:8080	<1%
29	Internet	pmc.ncbi.nlm.nih.gov	<1%
30	Internet	sci-hub.red	<1%
31	Internet	www.oncotarget.com	<1%
32	Student papers	Islington College,Nepal on 2025-12-31	<1%
33	Student papers	Liverpool John Moores University on 2025-07-17	<1%
34	Publication	M. S. Vijaya Kumar, K. Srujan Raju, K. Rajakumar, S. Saravanakumar. "A Study on ...	<1%
35	Student papers	University of Strathclyde on 2022-08-20	<1%
36	Internet	academic.oup.com	<1%
37	Internet	df6sxcketz7bb.cloudfront.net	<1%
38	Internet	www.medrxiv.org	<1%

39

Internet

www.nature.com

<1%

Integrated Computational Analysis for Deep Learning-Based Neurodegenerative disorder diagnosis and Biomarker Identification

Abstract

Aim: This paper will develop an integrative computational framework to study neurodegenerative diseases in transcriptomic analysis of ALS and DL-based classification of AD by utilizing structural MRI. This aims at identifying the major molecular biomarkers and establishing a precise, interpretable diagnostic model.

Background: NDDs like ALS and AD are demonstrated by progressive neuronal loss and complex underlying mechanisms involving neuroinflammation, protein aggregation, and cellular dysfunction. Whereas transcriptomic research can reveal insights into non-invasive and early detectable disease dysregulation at the gene level, neuroimaging in combination with artificial intelligence can provide insights into how diseases can be detected at an early and non-invasive stage of disease progression. Nevertheless, the inability to integrate molecular and imaging methods restricts a full picture of disease pathology.

Methodology: Transcriptomic analysis was done on the GSE10953 dataset that was provided by the Gene Expression Omnibus. First, the differential gene expression was analyzed using the Limma package, then key hub genes were determined by functional enrichment and protein-protein interaction (PPI) network analysis. At the same time, a CNN was developed to classify the different Alzheimer's disease stages based on the publicly available datasets which were presented in the form of MR images. A model was used for preprocessing methods and data augmentation and evaluated using the standard performance metrics. Grad-CAM was used to make the models more interpretable, visualizing discriminative brain areas.

Findings: The transcriptomic analysis revealed that 1216 differentially expressed genes, including SKP1A, CALM13, RPS14, H3C7, and CD44, were associated with neuroinflammation, ubiquitin-proteasome dysfunction, epigenetic dysregulation, and ribosomal stress. The overall accuracy of the deep learning model was 94 per cent, characterized by high values in precision, recall, F1 score and ROC-AUC. Grad-CAM visualizations showed that the model targets clinically relevant brain areas that are related to neurodegeneration. An integrated approach to molecular and imaging discoveries reveals complementary insight into disease pathogenesis and the potential of complementary bioinformatics and artificial intelligence to enhance diagnosis and biomarker discovery in neurodegenerative diseases.

CHAPTER 1

INTRODUCTION

Neurodegenerative diseases are one of the major biomedical and public health burdens of the 21st century because of their progressive nature, increasing incidence, multifaceted pathogenesises and the absence of clear cures. These conditions are defined by progressive dysfunction and irreversible mislaying of neurons in the central or peripheral nervous system, leading to progressive cognitive, sensory and motor deterioration. The growing prevalence of neurodegenerative diseases due to the increasing life expectancy and the aged population is imposing a significant social, economic and medical burden [1], [4]. Despite the considerable amount of research into the mechanism of neural degeneration, these processes are still poorly understood due to the variety and complexity of these diseases. There is growing data to suggest that neurodegeneration is not caused by one pathogenic event but rather a complex interplay of genetic and environment risk factors, molecular alterations with age, protein misfolding, mitochondrial dysfunction, oxidative stress, and neuroinflammation and cellular homeostasis [1-5]. As a result, reductionist approaches are often not sufficient to fully explain disease progression, and the need for a more holistic computational and systems biology approach that can investigate neurodegenerative processes across different biological levels is evident. Muscle weakness atrophy paralysis and eventually death from respiratory failure are all consequences of ALS a neurodegenerative disease that progresses quickly and is fatal due to the selective loss of upper and lower motor neurons [1]. Most patients die within several years from the onset of symptoms, underscoring the need for better understanding, early diagnosis and treatment in this disease. ALS was previously considered a purely motor neuron disease; however, it is now known to be a complex multisystem disease with dysfunction in multiple pathways . Glutamate-mediated excitotoxicity oxidative stress mitochondrial dysfunction impaired axonal transport RNA processing defects dysfunctional autophagy proteostasis defects and chronic neuroinflammation are among the pathogenic processes that have been linked to the development of ALS [1-3]. Protein aggregation (SOD1, TDP-43, and FUS) and genetic mutations (C9ORF72, TARDBP, FUS, and SOD1) have been shown to be a prominent pathological event, and the disease is highly genetically heterogeneous with familial and sporadic forms [1, 2]. Several recent studies have highlighted the overlap of these mechanisms through interlinked pathogenic mechanisms such as immune dysfunction, metabolic dysfunction, stress granule pathology, and impaired protein trafficking [2, 3]. This has transformed the view of ALS from a motor neuron disease to a systems mechanism of disease and established diagnostic biomarkers. In a similar

30 vein, Alzheimer's disease (AD), the most prevalent cause of dementia worldwide, is another significant neurodegenerative condition marked by progressive memory loss, cognitive decline, altered behaviour and crippling functional impairment [4]. Extracellular amyloid-beta is one of the classic neuropathological characteristics of AD. plaques and hyperphosphorylated tau-containing intracellular neurofibrillary tangles; however, more comprehensive pathology has been revealed by contemporary research. Growing evidence has shown that neuroinflammation, microglial cell activation, synaptic dysfunction, mitochondrial dysfunction, vascular changes, metabolic dysfunction and inflammatory signalling all play important roles in disease onset and progression [4, 5]. Sustained inflammatory responses by activated microglia and astrocytes play significant roles in neuronal degeneration and disease spread. Likewise, mitochondrial dysfunction and metabolic disturbances have been linked to synaptic dysfunction and degeneration [4]. These findings are consistent with new systems-based hypotheses for Alzheimer's disease (AD) that predict disease evolution through dynamic interaction between multiple molecular and cellular processes rather than a simple amyloid cascade. The disease is multifaceted and quite diverse, making early detection challenging, and this has spurred efforts to develop strong molecular and imaging diagnostic markers that could identify disease-related changes in the earliest phases of disease development, prior to disease progression. The biomarkers have therefore been an important focus of research in neurodegenerative disease due to their potential applications in the diagnosis, monitoring, prognosis and therapeutic intervention. The development of technologies like high throughput omics has revolutionized the biomarker discovery field by offering a tool to extensively explore the molecular changes that occur in disease. Specifically, transcriptomic profiling and differential gene expression analysis have emerged as powerful approaches to discover disease-associated genes, pathways, and networks in diseases like ALS [6, 7]. Differentially expressed genes (DEGs) may be used to identify disease-specific molecular signatures and to discover altered biological functions associated with neurodegenerative processes. Functional enrichment analysis (e.g., Gene Ontology and pathway analyses) and protein-protein interaction (PPI) networks also allow interpretation of dysregulated biological processes and identification of key hub genes and modules involved in disease processes [8], [9]. Software such as STRING and Cytoscape have emerged as integral tools in systems biology approaches, allowing for visualisation and prioritisation of potential biomarkers at the network level [8], [9]. These methods have greatly improved understanding of disease processes and contributed to identification of new targets for drug development. Crucially, systems biology approaches highlight that disease phenotypes are not the result of single gene mutations but rather are due to altered molecular networks, making network-based computational analyses particularly pertinent to complex diseases like ALS and Alzheimer's disease [5], [12].

Together with the discovery of molecular biomarkers, computational intelligence and artificial intelligence have also played a growing role in neurodegenerative disease diagnosis and analysis. In recent years, machine learning and deep learning algorithms have demonstrated their great

potential to uncover the pertinent signals in complex biomedical data, including neuroimaging. Structural Magnetic Resonance Imaging (MRI) offers rich information about anatomical changes associated with Alzheimer's disease, such as cortical thinning, hippocampal atrophy and ventricular enlargement. Traditional diagnostic approaches could require manual interpretation or involve feature engineering and can be time-consuming and prone to errors. With the emergence of deep learning, particularly CNNs, these problems have been partly resolved by automatic hierarchical feature extraction [10]. In structural magnetic resonance imaging (MRI) and multimodal neuroimaging data these models have shown excellent performance in the classification of Alzheimer's disease, in the staging of the disease severity and in the prediction of disease progression [10, 11]. Further, the accuracy of classification has been enhanced, and the manual feature extraction has been reduced due to transfer learning and new neural network architectures. Explainable AI techniques (such as Grad-CAM) have also been used to improve the interpretability of deep learning models, making it easier to trust the use of AI tools in clinical settings [11]. All of these developments have made deep learning a disruptive technology in the diagnosis of neurodegenerative diseases and precision medicine.

More and more recently, research has recognised the importance of a combination of approaches to reflect the complex nature of neurodegenerative diseases. Molecular biomarkers reflect disease-related mechanisms at the genomic and pathway levels; imaging biomarkers reflect morphological and functional changes associated with disease. Consequently, integrated computational models that combine approaches from both bioinformatics and systems biology and machine learning promise a multi-scale systems-level understanding of a neurodegenerative disease [7, 10, 12]. These strategies allow for a variety of approaches to the study of disease mechanisms with the molecular processing linked to structural and functional changes. This multi-scale view is especially beneficial for diseases with significant inter- or intra-individual variability, where single-scale analyses do not reveal disease mechanisms. In this sense combining DL-based MRI analysis for AD detection with computational detection of differentially expressed genes for ALS biomarker discovery provides a computational viewpoint on the molecular and structural features of neurodegenerative diseases. This approach not only aids in the early diagnosis and discovery of biomarkers but also informs precision medicine strategies for understanding, prognostication and treatment of disease [3], [10], [12]. Thus, the combination of transcriptomic biomarker detection and artificial intelligence-based neuroimaging is the key driving force behind the current work and paves the way for computational precision neurology.

Chapter 2

LITERATURE REVIEW

2.1 Overview of Neurodegenerative Diseases

NDDs are a group of diseases that lead to progressive damage of the neurons, with selective vulnerability, and then to dysfunction of the synapses and death of the neuronal cells. They are mainly central nervous system (CNS) diseases that lead to permanent destruction of the structure and function of neurons leading to severe clinical manifestations such as cognitive, motor and behavioural impairments [13, 14]. There are four major types of NDDs, namely AD, ALS, Parkinson's disease (PD) and Huntington's disease (HD), which exhibit distinct pathological and clinical characteristics and share common molecular mechanisms. AD is the most frequent type of dementia characterised by progressive memory, cognitive and behavioural impairments. ALS, in contrast, is a fatal motor neuron disease, affecting both upper and lower motor neurons, that leads to muscle weakness, wasting and paralysis [15]. But there is increasing evidence that NDDs have similar pathological processes which include protein aggregation, mitochondrial dysfunction, failure of axonal transport and inflammation [16, 17]. Neurodegeneration is not a single molecular defect, but a combination of genetic, epigenetic and environmental defects at a systems level. Several genes and pathways have been identified using the genome-wide association studies (GWAS) and gene expression profiles that are associated with disease development [18]. Further, the diversity of patients suggests the need for integrated approaches that integrate molecular information with computational models to gain insights into disease processes and to enhance treatment options.

2.2 Molecular Mechanism Underlying Neurodegeneration

2.2.1 Protein Misfolding, Aggregation, Proteostasis

Protein homeostasis (proteostasis) is critical for normal cell function, especially in neurons that are vulnerable to protein quality control (PQC) dysfunction. Neurodegenerative diseases are characterised by misfolded proteins evading degradation and forming toxic aggregates that lead

to dysfunction and death of neurons [19]. β -secretase (BACE1) and γ -secretase misleave the amyloid precursor protein (APP) in AD, producing amyloid-beta ($A\beta$) peptides that clump together to form oligomers and plaques that lead to neurotoxicity and synaptic dysfunction [20]. Phosphorylated tau protein forms neurofibrillary tangles (NFTs) which damage microtubules and transport [21]. TAR DNA-binding protein 43 (TDP-43), fused in sarcoma (FUS), and superoxide dismutase 1 (SOD1) comprise aggregates in ALS mutations, causing mislocalisation and aggregation in the cytoplasm, impairing RNA metabolism and function [22]. Misfolded proteins must be removed by the ubiquitin-proteasome system (UPS) and the autophagy-lysosomal pathways. When these systems malfunction, toxic protein aggregates are produced which are the root cause of these illnesses [23].

2.2.2 Neuroinflammation and Immune Response

In many neurodegenerative disorders, there is a heightened inflammatory response and the activation of glial cells (astrocytes and microglia) leads to the secretion of pro-inflammatory cytokines, chemokines and reactive oxygen species (ROS) that generate a neurotoxic environment [24][25]. Amyloid-beta aggregates activate microglial cells with the release of pro-inflammatory cytokines, including tumour necrosis factor-alpha (TNF- α), interleukin-1 beta (IL-1 β) and interleukin-6 (IL-6), leading to inflammation and eventually to the death of the neurons in Alzheimer's disease [25, 26]. Similarly, in ALS, neuroinflammation is a key contributor to motor neurone degeneration and includes increased microglial activation and recruitment of peripheral immune cells [27]. Recent studies have also shown that neuroinflammation is not just a bystander of disease but actively contributes to its progression and that there are opportunities for therapeutic intervention via inflammatory processes [26].

2.2.3 Oxidative Stress and Mitochondrial Dysfunction

Oxidative stress happens when there is a lack of antioxidants compared to the levels of the Reactive Oxygen Species (ROS). Oxidative stress is specifically damaging to neurons as they have high energy demands and are non-dividing cells [28]. Mitochondria are involved in ATP synthesis, calcium homeostasis and apoptosis. Dysfunctional mitochondria in neurodegenerative diseases cause energy deprivation, ROS production and cell death [29]. In ALS, superoxide radicals are not scavenged by mutant SOD1, and in Alzheimer's disease, synaptic dysfunction and cell death is caused by mitochondrial dysfunction [30].

2.3 Transcriptomics in ALS: From Gene Expression to Biomarkers

Transcriptomics is the analysis of the mass scale of RNA transcripts in cells or tissues. Technologies such as microarrays and RNA sequencing are available to quantify the expression

of thousands of genes [31]. The RNA-seq technique is more sensitive and specific for assessment of gene expression changes, alternative splicing and non-coding RNA changes and is widely adopted for ALS studies [32]. Differential gene expression (DGE) analysis is available to identify genes that are significantly expressed in disease and control conditions. The software programs used are DESeq2, edgeR and limma [33]. DGE has found changes in the expression of genes involved in RNA processing, immune regulation and the cytoskeleton in ALS. Global changes in the transcriptomic profile occur in neurons following impairment of RNA-binding proteins, like TDP-43 [34]. Protein-protein interaction (PPI) networks show protein interactions. Databases like STRING [35] and software tools like Cytoscape [36] can be used to generate and analyse PPI networks. Highly connected "hub" genes play an important role in maintaining the integrity of networks. The hub genes in ALS are the SOD1, TARDBP, FUS and C9orf72 genes that play several roles [37]. The differentially expressed genes are mapped onto biological processes and pathways (Gene Ontology (GO) and KEGG) using pathway analysis maps [38, 39]. ALS is a complex disease with enriched pathways associated with RNA processing, neuroinflammation, apoptosis and oxidative stress [39].

2.4 Deep learning in Alzheimer's Disease Diagnosis

AD is a great application of deep learning for the automatic diagnosis of AD, as it has enabled the automatic extraction of complex information from both neuroimaging and clinical data, avoiding the need for feature extraction [39]. Specifically, CNNs have demonstrated outstanding accuracy with regard to the processing of sMRI, functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) images and are able to learn hierarchical spatial features associated with early neurodegenerative processes such as hippocampal atrophy, cortical thinning and ventricular enlargement [40]. Recent advances in architectures like 3D-CNNs and CNN-recurrent neural network (RNN) hybrids further reinforce the learning of representations of spatial and temporal information which enables the precise classification of different stages of AD, including mild cognitive impairment (MCI) and conversion to AD [41]. The use of pre-trained networks and transfer learning has also been widely used to overcome the small medical data problem, which improves the generalization and accelerates learning [42]. Furthermore, multimodal data such as imaging, genetic (e.g., apolipoprotein E; APOE) and other clinical biomarkers have been shown to clearly improve the diagnostic performance and reliability [43]. A key development in this area is the use of explainable artificial intelligence (XAI) methods, such as Grad-CAM, which produce heatmaps of the brain regions that are most significant for model predictions, thus improving trust, interpretability, and the potential for clinical use [44]. Although these have shown to be effective, there are still challenges such as data heterogeneity, class imbalance, overfitting and interpretability that remain to be solved, with the aim of clinical implementation [45]. Despite this, deep learning remains crucial in early detection of AD, disease evolution prediction, and the development of personalized treatment strategies, thus driving precision neurology and AI-integrated healthcare [46].

2.5 System Biology and Network Biology

Advances in systems biology and network medicine have enabled researchers to move beyond single gene/pathway studies to study the complex biological networks involved in disease processes for neurodegenerative diseases (NDDs) [47]. Systems biology is the application of multi-omics data (including genomics, transcriptomics, proteomics, metabolomics and epigenomics) to model the dynamic interactions between molecular components of cells and tissues. This method is particularly helpful for NDDs because they often have multiple dysfunctional pathways that contribute to the disease, such as AD and ALS [48]. Systems biology approaches using the computational modelling and regulatory networks can be used to identify critical molecular regulators, feedback loops and emergent behaviour that can drive neuronal dysfunction and degeneration [49]. Network medicine is a further extension of this, which applies graph theory and network science to explain and analyse interactions among the components of a biological system such as genes, proteins, and metabolites [50]. The networks are made up of nodes and edges, where nodes are molecular entities and edges are interactions. Protein-protein interaction networks, gene regulatory networks and metabolic networks are frequently employed to investigate disease processes. One concept in network medicine is that of hub genes or hub proteins, which are highly connected hubs that play an important role in ensuring network stability and cellular functionality [51]. The failure or dysfunction of these hubs may have a ripple effect on the network resulting in disease. Hub genes in ALS such as SOD1, TARDBP, FUS and C9orf72 belong to a number of interconnected pathways, such as RNA processing, protein quality control and stress response [52]. The other important concept is that of a disease module, a collection of interrelated genes or proteins in a network that cooperate to produce a disease phenotype [53]. Discovering these modules helps to identify key pathways involved in NDDs such as neuroinflammation, oxidative stress, mitochondrial dysfunction and protein imbalance. Additionally, network-based methods help in drug repurposing to identify possible drug targets within these disease modules [54]. Combining network medicine with high-throughput omics approaches makes it possible to identify hidden links and more accurately predict disease-related biomarkers [55]. Crucially, systems biology and network medicine also offer support for personalized medicine through the generation of patient-specific networks. This allows researchers to integrate individual genetic and molecular data to determine disease-specific networks and design patient-specific treatments [56]. Despite the benefits, there are challenges such as integration of data sources, data variability, and the need for robust computational models [57]. However, integration of systems-level approaches and network modelling is a promising strategy to dissect the complexity of neurodegenerative diseases and guide the development of effective disease-modifying therapies [58].

2.6 Integration of Transcriptomics and Deep Learning

This paper takes an integrative multi-modal computational viewpoint of transcriptomic profiling of Amyotrophic Lateral Sclerosis (ALS) and deep learning-based neuroimaging analysis of Alzheimer Disease (AD) to identify common molecular and structural neurodegeneration signatures. The ability of deep learning to learn non-linear relationships, as well as the ability to learn hierarchical representations, has made it a powerful means of analyzing such complex data sets [60]. Neural network architectures like deep neural networks (DNNs), autoencoders and graph neural networks (GNNs) are increasingly being used to discover latent features and to draw out hidden structure in transcriptomic data [61]. One of the primary applications of this integration is the discovery of biomarkers. By training deep learning models on RNA-seq or microarray data, it is possible to identify gene expression signatures that distinguish between disease states and healthy controls. As an example, it is possible to learn valuable genes associated with disease using autoencoders that reduce high-dimensional data on gene expression to lower-dimensional forms, preserving important biologically relevant information [62]. Similarly, transcriptomic profiles can be used to forecast the survival of patients in a comparable supervised learning model to predict disease progression and disease subtyping [63]. Such techniques have revealed the dysregulated signalling of RNA metabolism, immune response and synaptic activity in ALS and Alzheimer disease, which provides insights into the mechanisms of disease [64]. Another important development is the use of transcriptomic data in deep learning models and other modalities, such as clinical data with neuroimaging and genomics. Multimodal deep learning models combine various sources of data to increase accuracy and stability of predictions. As an example, there are models that can be used to correlate molecular changes with structural changes in the brain by integrating data on gene expression with MRI features, and this leads to improved disease characterization [65]. Protein-protein interaction (PPI) networks and other biological networks are added to graph-based deep learning techniques to further improve this. This allows the model to learn the degree of gene expression as well as the functional relationships between the genes. [66]. Explainability remains essential when applying deep learning in biomedical research. By estimating the genes or pathways with the greatest contribution to model predictions, attention mechanisms, feature importance scoring and saliency maps, among others, help to increase interpretability and biological relevance [67]. It is especially relevant for translating research into clinical practice, where the ability to win the confidence of the medical fraternity is pegged on the understanding of logic in model decisions. Despite its potential, there are several challenges to consider, such as the lack of large, well-annotated datasets, batch effects in transcriptomic data, overfitting issues in deep learning models and computational resource needs. Regardless of its potential, several challenges must be considered: lack of large, well-annotated datasets, batch effects in transcriptomic data, overfitting in deep learning models and computational resource needs. Additionally, careful normalization

and alignment techniques are needed to ensure consistency and reliability when merging different types of data. The developing of more resilient multimodal structures, federated learning mechanisms of data exchange and preserving privacy, and the combination of longitudinal data to trace the progression of the diseases over time are among the future directions [69]. In general, the combination of transcriptomics and deep learning has high potential to enhance precision medicine by enabling the early diagnosis of neurodegenerative diseases and individualized treatment plans and better insight into their molecular mechanisms [70].

2.7 Research Gap and Future Directions

Although considerable progress has been made in understanding neurodegenerative diseases (NDDs) and the application of advanced computational tools, several key gaps in research that limit the use of recent findings in clinical practice exist [71]. Lack of thorough integration of multi-omics data such as transcriptomics proteomics metabolomics and genomics is one of the biggest obstacles. The absence of common pipelines to combine these disparate types of data constrains the possibilities of constructing a unified and systems-level view of disease processes despite the fact that individual research provides informative data [72]. Moreover, most of the studies are based on relatively small cohort-specific studies, which leads to issues such as population bias, batch effects and limited generalisability, reducing the strength of discovered biomarkers and predictive models [73]. A second shortcoming is the lack of high-quality annotated longitudinal data, which tracks the course of disease over time. Although neurodegenerative diseases such as AD and ALS are dynamic, cross-sectional data cannot be used to capture these changes in gene expression, brain structure and clinical symptoms over time [74]. This restricts the growth of the early diagnosis instruments and precise prognostic models. Moreover, another component of complexity that is often not adequately considered in existing research is the diversity of patients, which is induced by genetic variations, environmental exposure and lifestyles [75]. Deep learning models are black-box systems due to their interpretability issues in spite of their power in a computational perspective. This transparency is a hindrance to their incorporation in clinical settings where explainability and reproducibility are crucial [76]. Furthermore, models trained on large-dimensional biological data sets containing small sample sizes are very vulnerable to overfitting. The difference in the features (e.g.) also makes it more difficult to validate and train the model [77]. There also needs to be improved benchmarking frameworks and standardized evaluation metrics to enable comparisons between studies of models. One underdeveloped aspect is the incorporation of biological information into machine learning models. Most of the existing approaches process

information as the purely numerical input without considering the prior biological information like pathway interactions or gene regulatory networks. This limits biological interpretability and would lead to false correlations [78]. In the same manner, a disparity still exists between the prediction of *in silico* and biological significance in the real world because of the poor translation of computational results into experimental outcomes. Further research is essential to improve understanding of disease progression and mechanisms; however, it is necessary to focus on the development of powerful multimodal models that can incorporate different types of data, including imaging, omics and clinical data [79]. The limitations on data-sharing can be eliminated by using federated learning methods that enable the model training to be performed collaboratively by institutions without the need to disclose data [80]. The improvements in explainable artificial intelligence (XAI) are expected to improve transparency and reliability of the model, allow researchers and clinicians to interpret predictions more accurately and identify significant attributes that lead to disease [81]. Also, modelling disease trajectories and addressing early intervention strategies will be enhanced through the integration of real-world and longitudinal data, including electronic health records [82]. Recent technologies offering a new level of resolution in the analysis of cellular heterogeneity and microenvironmental interactions involve single-cell RNA sequencing and spatial transcriptomics [83]. These technologies, in conjunction with deep learning models, can identify new biomarkers and therapeutic targets. In addition, interdisciplinary collaboration of biologists, physicians and data scientists will be crucial in addressing current challenges and gaps and accelerating the creation of personalized medicine approaches [84]. The entire landscape will be key to solving these issues to convert computational and molecular understanding into practical diagnostics and therapeutic approaches for neurodegenerative disease.

CHAPTER 3

METHODOLOGY

3.1 Study Design and Computational Framework

This paper takes an integrative multi-modal computational viewpoint of transcriptomic profiling of Amyotrophic Lateral Sclerosis (ALS) and deep learning-based neuroimaging analysis of Alzheimer Disease (AD) to identify common molecular and structural neurodegeneration signatures. Contrary to traditional studies, which focus on single domains, this framework will be tailored to connect the molecular-level dysregulation to the phenotypic expression in brain imaging, thus complying with the principles of systems biology.

The pipeline of methods is broken down into two large analytical modules:

- Biomarker discovery (ALS) based on transcriptomics.
- Neuroimaging classification (AD) based on deep learning.

Network-based interpretation and biological correlation analysis are then used to integrate these modules.

The rationale of such a design is founded on the hypothesis that:

- Proteostasis imbalance and neuroinflammation are some of the common molecular pathways found in neurodegenerative diseases.
- Imaging can identify these molecular alterations which are structural alterations in the brain.

3.2 Transcriptomic Analysis for ALS

3.2.1 Data Collection

Transcriptomic data presented in this study was acquired through the publicly available Gene Expression Omnibus (GEO) repository, which is a large repository of high-throughput genomic data. GSE10953 dataset was chosen according to its relevance to Amyotrophic Lateral Sclerosis (ALS), sample size consisting of diseased and healthy control samples, and the availability of high-quality raw expression data. The data is in the form of microarray based gene expression profiles constructed out of human neural or patient biological samples and will give continuous intensity data that can be subjected to subsequent computational analysis. The selection criteria were sufficient to ensure that the sample size, biological relevance and reproducibility were sufficient to allow the dataset to identify transcriptional changes related to ALS pathology.

3.2.2 Preprocessing and Quality control of Data

Preprocessing was done to remove the technical noise and increase the biological validity of the dataset. Background correction was done to eliminate non-specific signal variation as part of microarray experiments. The distribution of the expression values of all samples was then standardized through quantile normalization to reduce batch effects and technical variability. A $\log_2$ transformation was done to stabilize the variance and approximate normal distribution of the expression data. The missing values were dealt with by either eliminating the unreliable probes or by use of statistical imputation methods. Annotation files were used to map probe identifiers to gene symbols to make them biologically interpretable. Furthermore, Principal Component Analysis (PCA) was used to identify outlier samples; the samples that exhibited a great deal of deviation in cluster patterns were filtered out to enhance the strength and accuracy of downstream analysis.

3.2.3 Differential Gene Expression Analysis

The limma (Linear Models to Microarray Data) package in R was used to perform the analysis of the differential gene expression (DGE), as this package combines the use of linear modelling and empirical Bayes statistical approaches to the analysis. The data on the expression was modelled by using the equation $Y = X\beta + e$, where Y is the data on the expression of the genes, X is the design matrix, β are the estimated coefficients, and e is the error term. Genes were deemed differentially expressed significantly with a threshold of adjusted p-value (False Discovery Rate, FDR) < 0.06 and an absolute $\log_2$ fold change of 1. This discussion led to the discovery of 1216 differentially expressed genes (DEGs), which showed significant transcriptional alteration related to ALS progression and underlying the molecular pathology.

3.2.4 Construction of Protein Network (PPI)

15 A protein-protein interaction (PPI) network was built in the STRING database with a high-confidence interaction score threshold (≥ 0.7) to investigate the functional relationships between the identified DEGs. The data interaction was visualised and analysed using the software Cytoscape. This network model represents the proteins as nodes, and the interactions between them as edges, creating a complex network of connections. Adjacency matrix representation indicates if there is an interaction or not between proteins. This network-based methodology allows recognizing functional modules, central regulatory proteins, and biologically important pathways, which can give information about the molecular interactions underlying the ALS pathogenesis.

3.2.5 Hub Gene Identification

The CytoHubba plugin of Cytoscape (<http://cytoscape.org>) was used to find the hub genes; the hub genes are the genes in the center of a network that play a key role in the network. The centrality of the degree was taken as the main measure, as the significance of a node is calculated based on the number of its direct neighbours (C). $D(v) = \deg(v)$. The number of connections was considered as an indicator of the regulatory importance of the genes, and therefore they were referred to as hub genes. The most well-known hub genes that play key roles in fundamental cellular functions such as protein degradation, calcium signalling and neuroinflammation were the most well-defined: SKP1A, CALM13, CD44, RPS14, and H3C7. Likely, these hub genes are involved in the regulation of ALS and may prove to be potential markers or targets for drug therapy.

3.2.5 Functional Enrichment Analysis

22 The biological meaning of the DEGs was analyzed via functional enrichment analysis on STRING based on Gene Ontology (GO) and KEGG pathway databases. The analysis was carried

21 out in the categories of Biological Processes (BP), Molecular Functions (MF), and Cellular Components (CC) to get the overall picture of the role of genes. RNA metabolism, neuroinflammation, apoptosis, and oxidative stress were found to be key enriched pathways, which have been shown to be strongly related to the pathology of ALS. These findings are consistent with the known disease's pathophysiology, and highlight the multifactorial aspects of ALS, in terms of genetic regulation, cellular stress response and neuronal degeneration.

3.3 Deep Learning Based MRI Analysis of Alzheimer's Disease

3.3.1 Data Curation

34 The MRI data employed in this paper has been taken in the Open Access Series of Imaging Studies (OASIS), which is a public database containing structural brain imaging data on the study of Alzheimer disease. The dataset comprised of 6400 two-dimensional MRI slices divided into four classes ushering various stages of cognitive impairment; Non-Demented (3200 images), Very Mild Demented (2240 images), Mild Demented (896 images), and Moderate Demented (64 images). The data showed a large imbalance in the number of classes especially in the category of Moderate Demented. In order to guarantee high-quality model estimation and maintain the balance of classes, stratified sampling was utilized to divide the data into training (80 percent), validation (10 percent) and test (10 percent) parts. In this method, the proportionality of each class was preserved in all subsets. Also, class weighting was used in the training of the model to reduce the bias on the majority classes and enhance the performance of the classification on the underrepresented classes.

3.3.2 Image Preprocessing and Data Augmentation

26 The MRI data was preprocessed to standardized the data and increase the performance of the model. All images were brought to a pixel intensity range of [0, 1] so that the scaling of the images is equally comparable throughout the dataset. To fit the input demands of the convolutional neural network (CNN) and decrease the complexity of the computations, the images were rescaled to a constant size of 128×128 pixels. The training set was the only set to which data augmentation techniques were used to augment data variability and prevent overfitting. These methods involved random rotations that were small (less than 10 degrees), horizontal flips, and slight zoom effects. These augmentations model realistic real-world changes in imaging conditions, and allow the model to learn more robust and generalized feature representations.

3.3.3 Model Architecture

The Keras Functional API was used to create a two-dimensional Convolutional Neural Network (2D-CNN) to help extract features and classify multiple classes automatically. The architecture comprised 4 successive convolutional layers with 3×3 kernel learning the spatial information in the input MRI images. Convolutional layers were then succeeded by max-pooling layers to decrease the spatial dimensions and computing power but maintain key features. A dropout layer was added after the dense layer to avoid overfitting, with a rate of 0.5. The connection layer was fully connected with 512 neurons, which allowed the high-level feature learning and classification. The last output layer had four neurons that represented the four disease classes, and the softmax activation function was employed to provide probability distributions of each disease class. This architecture successfully models the hierarchical trends in brain imaging data and can be used to correctly classify the stages of Alzheimer's disease.

3.3.4 Training Protocol

The model of CNN was trained on the Adam optimizer and an initial learning rate of 0.001 was selected due to its adaptive learning and efficient convergence properties. The 50-epoch training process was done with a batch size of 32. Early stopping was used to avoid overfitting and achieve the best generalization; the stopping criterion was the validation loss and the patient was five epochs. The training was done to save the model with the least validation loss to promote reproducibility and further analysis. This training plan ensures the stability of the model, its efficiency and improved prediction behaviour.

3.3.5 Model Assessment and Computational Efficiency

The trained model was tested on a separate test set with various performance measures, such as accuracy, precision, recall, F1-score, confusion and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) performance metrics. These measures offer an overall assessment of classification performance, especially when there is an imbalance in classes. It was highlighted that macro-averaged F1-score and ROC-AUC are better measures than overall accuracy. The proposed CNN architecture was computationally efficient, which enabled training and inference in typical systems with GPUs. The mean prediction time per image of the MRI was under a second, which implies that the model could be used in large-scale screening and, possibly, in clinical decision-support processes.

3.3.6 Explainability using Grad-CAM

Grad-CAM, which is an explainable artificial intelligence (XAI) technique, was used to improve the interpretability of the models. Grad-CAM calculates the gradients of the target class in respect to the feature maps of the final convolutional layer, which produces class-specific

heatmaps indicating the most significant regions in the input image. These heatmaps were upsampled with the OpenCV and overlaid with the original MRI images to be visualized with Matplotlib. This was implemented with TensorFlow's GradientTape. GradientTape automatic differentiation. The method allows visualizing of discriminating brain areas related to Alzheimer's disease, verifying model predictions, and helping to determine misclassification trends. Grad-CAM increases the reliability of deep learning models in medical imaging tasks by promoting transparency and clinical interpretability.

3.4 Integration of Transcriptomics and Deep Learning

This paper takes an integrative approach and integrates transcriptomic analysis and deep learning-based neuroimaging to gain a more in-depth insight into the mechanisms of neurodegenerative diseases. The identified differentially expressed genes (DEGs) and hub genes of the ALS transcriptomic analysis were employed to intercept latent molecular changes, such as neuroinflammation, RNA metabolism, oxidative stress, and apoptosis pathways. At the same time, the convolutional neural network (CNN) model was applied to the high-level spatial features, extracted from the MRI scans, to classify the Alzheimer's disease. It was done at the systems level by mapping the gene-level insight to features derived from imaging, and identifying common biological pathways and cross-disease signatures. Such a multimodal method enables the correlating of molecular dysregulation with structural brain variations, enhancing the predictiveness and strength of the predictive model. Furthermore, the integration of this kind allows for the identification of potential biomarkers and early ability to diagnose and create a customized treatment plan. The proposed approach is a data-driven transcriptomic approach along with feature learning through imaging, which helps in having a holistic view to bridge the gap between genotype-to-phenotype correlations in neurodegenerative diseases.

Chapter 4

RESULTS

4.1 Transcriptomic Analysis Results

4.2.1 Identification of Differentially Expressed Genes in ALS

The results of comparative transcriptomic analysis of the GSE10953 dataset led to the identification of a total of 1216 differentially expressed genes (DEGs) in ALS spinal cord tissue. Of these, 613 genes were significantly upregulated and 603 genes were downregulated compared to control samples, which were healthy. A volcano plot (Fig. 1) has been used to show the global distribution of these DEGs, with the strongly upregulated genes in red, the strongly downregulated ones in blue, and those that were not significantly regulated in grey. A subset of genes exhibited significant differential expression at the specified threshold of $|\log_2 \text{fold change}| > 1$ and $p\text{-value} < 0.06$. At the set of $\log_2$ fold change of above 1.0 and $p\text{-value}$ of below 0.06, a subset of genes had significant differences in their expression. Nevertheless, with more stringent statistical cutoffs, smaller numbers of genes were found as strongly dysregulated, which requires a rather liberal $p\text{-value}$ cutoff to identify more biologically relevant candidates. Moreover, box plot visualization (Fig. 2) also showed the definite difference of distribution of gene expression between the samples of ALS and control, which helps to prove reliability of the dataset. Because of the exploratory character of this analysis, the identification of genes with relaxed thresholds was further ranked on the basis of pathway enrichment and network-based methods to understand their biological importance.

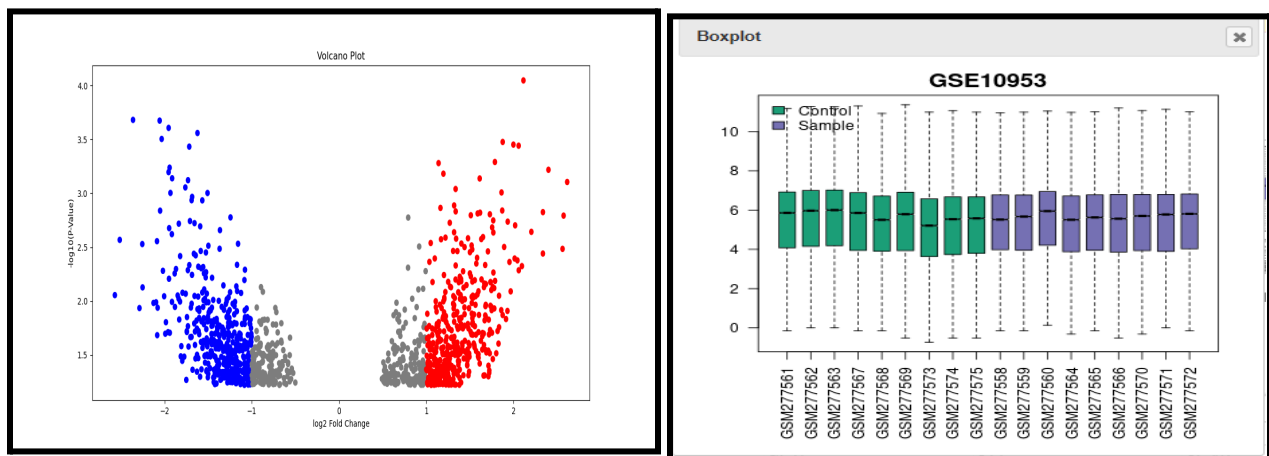


Fig. 1. Significantly upregulated (red) and downregulated (blue), not significant (grey), plotted in a volcano using $\log_2$ fold change and $p\text{-value}$ thresholds. **Fig. 2.** A Box plot of expression values for GSE10953, showing distribution across control and sample.

4.2.2 Gene Ontology and Pathway Enrichment Analysis

The functional enrichment analysis was performed on the upregulated and downregulated sets of genes in isolation to comprehend the biological implication of the analysis. The upregulated genes analysis demonstrated that there was a significant enrichment of the immune response and

neuroinflammatory pathways. The analysis of Gene Ontology (GO) revealed that the most notable biological processes were immune system processes and inflammatory responses, which implied that immune dysregulation was actively involved in the ALS pathology (Fig. 3a). On the molecular functional level, the activity of cytokines and cell adhesion molecule binding were highly enriched (Fig. 3b), and cellular component analysis revealed an engagement with extracellular and immune-related structures (Fig. 3c). Further involvement of immune-related pathways was further confirmed by additional analyses, such as Mammalian Phenotype Ontology (Fig. 3d) and tissue-specific expression profiling (Fig. 3e). These findings were supported by KEGG pathway analysis, which revealed that cytokine-cytokine receptor interaction was an enriched pathway.

Conversely, enrichment analysis of the downregulated genes revealed a perturbation of basic neuronal and cellular maintenance processes (Fig. 4). Protein ubiquitination, cell-cycle, and synaptic signaling biological processes were greatly impacted (Fig. 4a). The analysis of molecular functions showed enrichment in ubiquitin-protein ligase activity and ion channel activity (Fig. 4b), and cellular component analysis highlighted activity of the synaptic structure and proteasome complex (Fig. 4c). The analysis of KEGG further revealed that there was a significant enrichment in the ubiquitin-mediated proteolysis and synaptic vesicle cycle pathways. These findings underscore the functional relationship between the DEGs.

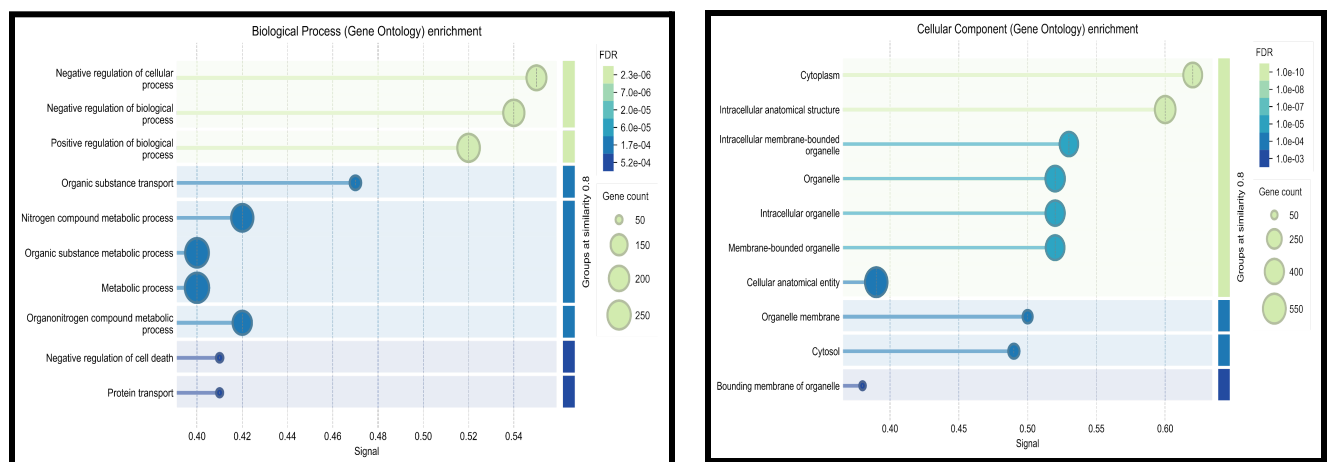


Fig. 3 (a). Gene ontology (Biological Process)

enrichment analysis of differentially expressed genes (DEGs) in ALS. The plot highlights significantly enriched biological processes, which have strong emphasis on immune response, inflammatory pathways, and cellular regulatory mechanisms contributing to disease pathology.

Fig. 3 (b). Gene Ontology (Cellular Component) enrichment analysis of DEGs in ALS. These results show enrichment in cellular structure, including synapses, membrane components, and protein complexes, reflecting structural and functional disruptions in neuronal cells.

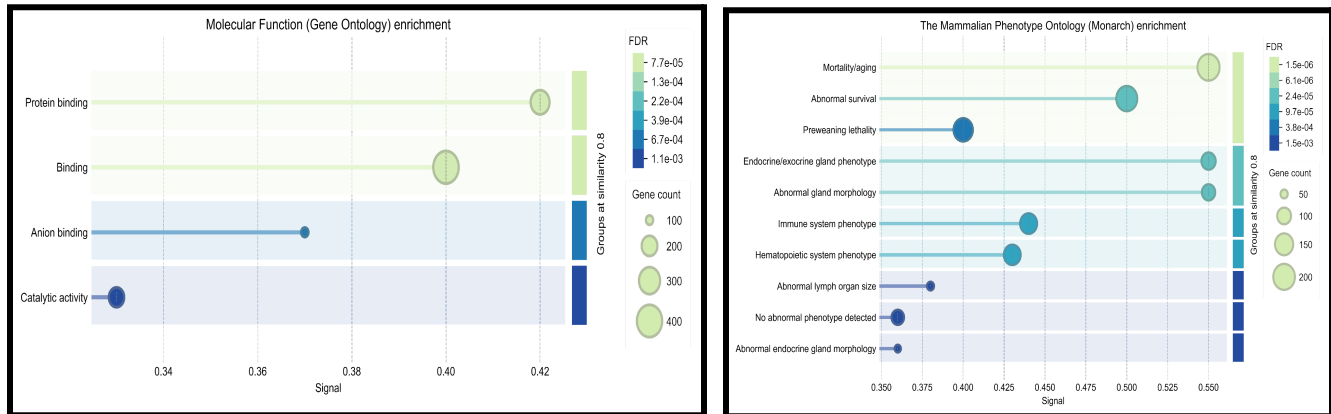


Fig. 3 (c) Gene Ontology (Molecular Function) DEGs analysis in ALS. Important molecular processes like protein binding, cytokine activity and ion channel activity are depicted in this figure, which also shows changes in signaling and molecular interactions.

Fig. 3(d) Mammalian Phenotype Ontology (MONARCH) enrichment analysis of ALS-related upregulated genes (URGs). Significantly enriched phenotype terms linked to the upregulated gene set are highlighted in the plot suggesting connections to aberrant neurological functions, immune system changes and disease-related phenotypic manifestations pertinent to ALS pathology.

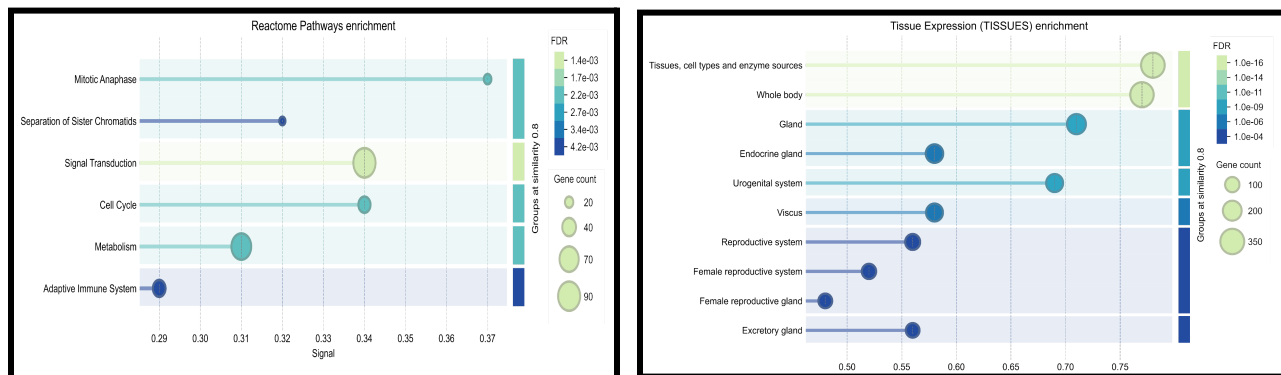


Fig. 3(e) Reactome pathway enrichment. The figure highlights the role of dysregulated immune processes in the progression of ALS by showing significantly enriched biological pathways especially those connected to immune signaling cytokine-mediated interactions and inflammatory responses.

Fig. 3 (f) Analysis of upregulated genes (URGs) in ALS using tissue-specific expression. The findings show that these genes are expressed preferentially in particular tissues, with strong

enrichment seen in tissues related to the nervous system and immune system, indicating their significance to disease-affected areas and systemic involvement in ALS.

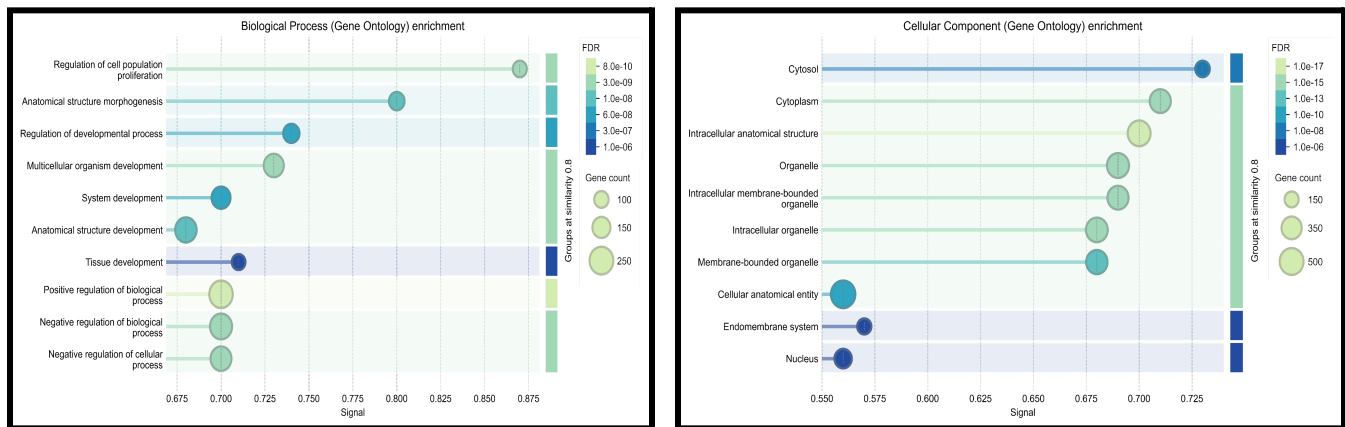


Fig. 4. (a) Downregulated genes (DRGs) in ALS are analyzed for enrichment using Gene Ontology (Biological Process). The plot demonstrates a marked enrichment in processes associated with cellular maintenance, synaptic transmission and neuronal signalling, suggesting decreased synaptic activity and compromised neuronal function in ALS pathology.

Fig. 4. (b) Enrichment analysis of downregulated genes (DRGs) in ALS using Gene Ontology (Cellular Component). Synapses, cytoskeletal structures and proteasome complexes are among the enriched components which represent the structural and functional decline of neuronal cells.

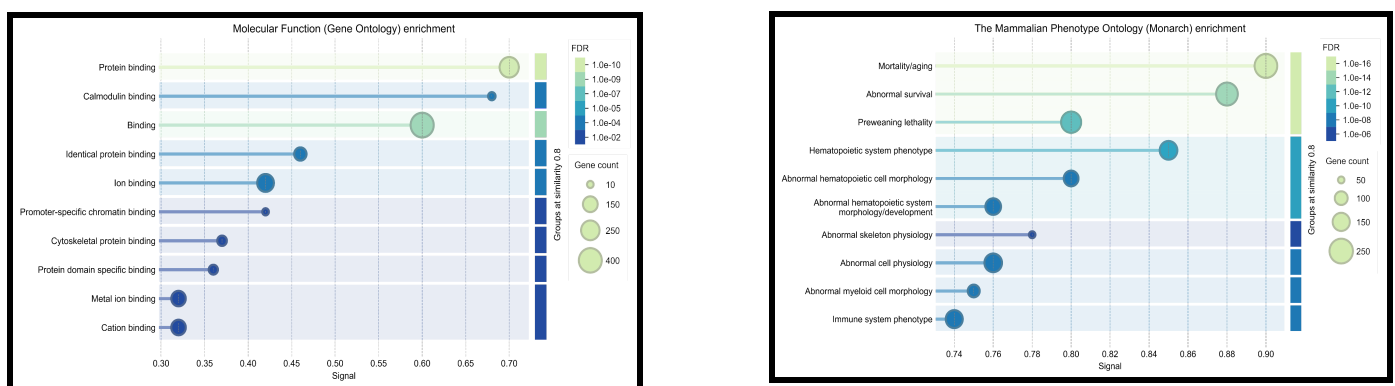


Fig. 4 (c): Enrichment analysis of downregulated genes (DRGs) in ALS using Gene Ontology (Molecular Function). The findings show reduced activity in molecular processes like protein binding, ion channel activity and enzymatic regulation, indicating a disturbance in crucial cellular interactions and signaling pathways.

Fig. 4(d) Enrichment Analysis of downregulated genes (DRGs) in ALS using Mammalian Phenotype Ontology (MONARCH). Significantly enriched phenotype terms linked to abnormal synaptic transmission impaired neuronal function and cellular homeostasis defects are highlighted in the plot .

4.2.3 Protein-Protein Interaction Network and Hub Gene Identification

The protein-protein interaction (PPI) networks of the upregulated and downregulated sets of genes were built in order to explore further the function dysfunction of protein degradation systems and neuronal communication processes in ALS, which play a key role in neurodegeneration.

The upregulated gene PPI network (Fig. 5) presented a dense and interconnected network, which showed that there were strong functional associations among the upregulated genes. Topological analysis was used to identify the key hub genes according to the degree centrality, and SKP1A (degree = 12) and CALM13 (degree = 10) turned out to be the most important nodes. The location of these genes at the centre indicates their important regulatory functions in the pathways related to immune response and cellular signalling. The PPI network downregulated genes (Fig. 6), likewise, had a complicated interaction space. RPS14 (degree = 10), H3C7 (degree = 10), and CD44 (degree = 12) were the most significant hub genes. These genes relate to such fundamental cellular activities as protein synthesis, chromatin organization, and cell adhesion. These hub genes findings suggest that the pathology of ALS is characterized not only by the stimulation of inflammatory pathways but also by the inhibition of important cell repair mechanisms. In general, network analysis shows the significance of regulatory genes that are central to the process of disease progression and offers possible targets to be studied further.

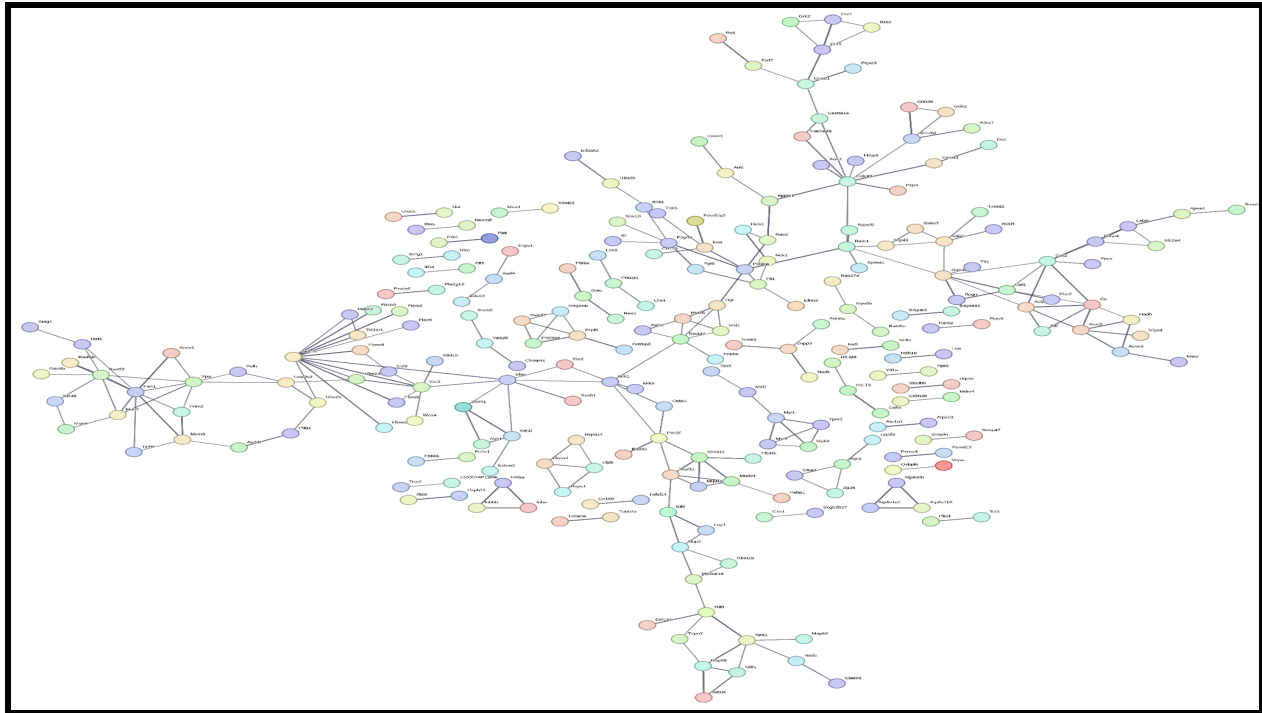


Fig. 5 The STRING database was used to create the Protein-Protein Interaction (PPI) network of upregulated genes (URGs) in ALS which was then displayed in Cytoscape. Proteins are represented by nodes and interactions that have been predicted or confirmed through experimentation are indicated by edges.

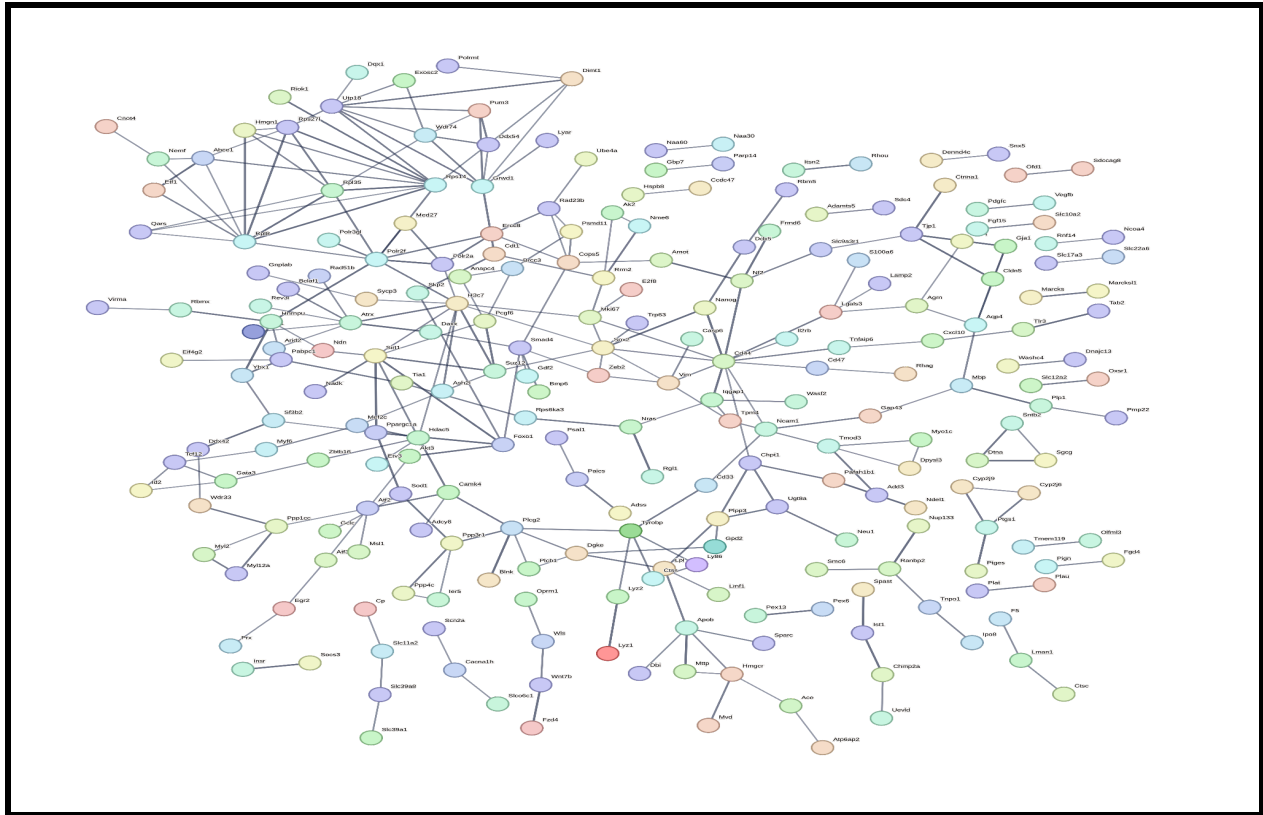


Fig. 6: The STRING database was used to create the protein-protein interaction (PPI) network of downregulated genes (DRGs) in ALS. The network shows how different proteins related to cellular maintenance synaptic activity and neuronal function interact. Important hub nodes with increased connectivity point to important regulatory functions and show that vital biological processes like synaptic signalling and protein degradation are disrupted in ALS.

4.2 Deep Learning MRI Results

4.2.1 Quantitative Performance Evaluation

The proposed Convolutional Neural Network (CNN) model demonstrated good classification performance on the independent test dataset, with an overall classification accuracy of 94%. But since there is a class imbalance (especially the under-representation of the moderate demented class), accuracy is not adequate to completely assess model performance. Thus, additional measures of precision, recall, and F1-score were also taken into account to give a more reliable and clinically significant measure. The model was observed to be highly sensitive in terms of detection of disease stage, particularly in the mild demented category, where the value of recall is 0.98, implying that the likelihood of false negatives is very low. It is also vital in clinical practice, whereby timely and appropriate diagnosis may be critical in the treatment of the disease. The Non-Demented condition had an ideal precision (1.00) and a recall of 0.90, which

meant that the model was efficient in minimizing the number of times a false-positive diagnosis was made whilst the model was effective in maximizing the number of times a false-positive diagnosis was made. The model used in the case of the moderately demented class had a 1.00 preciseness and a 0.888 recall. In spite of the fact that these values point out the high predictive power, they have to be regarded with some reservations due to the low size of the sample ($n = 8$) which can introduce some variability and reduce the level of the statistical reliability. The model recall was 0.99 and precision was 0.88, indicating the ability of the model to detect the faint early-stage features of the disease. Overall, the average F1-score of the model of 0.94 indicates that the model is balanced as far as its performance in all the classes is concerned, even though the data is imbalanced. It implies that the specified model is not biased towards a particular class and is consistent in different phases of Alzheimer's disease.

4.2.2 Classification Summary

Table 4.2 shows the detailed classification metrics of each of the classes. The findings affirm that the model maintains high precision as well as recall in all categories, with slightly lower stability in the moderately demented category due to the limited availability of data.

	Precision	Recall	F1-score	Support
Mildly demented	0.96	0.98	0.97	90
Moderately demented	1.00	0.888	0.93	8
Non-Demented	1.00	0.90	0.95	320
Very Mild Demented	0.88	0.99	0.93	224
Accuracy	0.94	0.94	0.94	642
Macro avg	0.96	0.94	0.94	642
Weighted avg	0.95	0.94	0.94	642

Table 2 CNN classification performance for Alzheimer's disease: The suggested CNN models' performance metrics in four different classes. Despite class imbalance, the model achieved 94% accuracy with strong precision, recall and F1-scores, indicating dependable and balanced classification.

4.2.3 Analysis of Classification Errors

The confusion matrix (Fig. 4.2) gives a qualitative understanding of the classification behaviour of the model. A high diagonal means that most predictions are accurate, which proves that the trained CNN is reliable. The main observed misclassifications were between adjacent disease stages, with the most pronounced misclassifications being between Non-Demented and Very Mildly Demented classes. This pattern is not unexpected, given that the early stages of Alzheimer's disease can be described as a subtle progressive change in the structure that can hardly be detected even with advanced imaging techniques. Notably, there was a little confusion between remote classes (e.g., non-demented vs. moderately demented), and it suggests that this model is effective at capturing the progress of the severity of the disease. Such conduct proves that the model fits well with clinical knowledge, in which boundaries between initial stages of disease are always ambiguous.

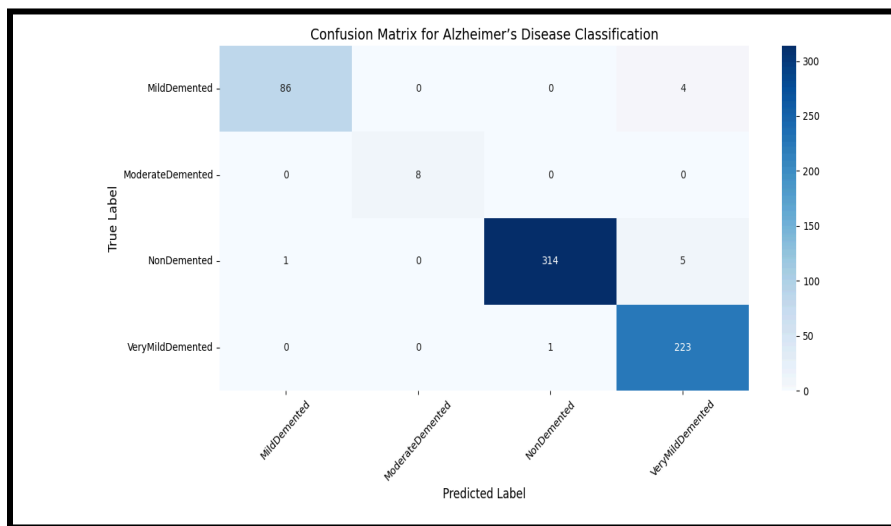


Fig. 7 The CNN model's confusion matrix for multi-class Alzheimer's disease stage classification. The distribution of true versus predicted labels for the four classes—non-demented, very mild demented, mild demented and moderate demented—is shown in the matrix. Strong diagonal values show high classification accuracy, whereas small off-diagonal values show little misclassification, mostly between stages of the disease.

4.2.4 ROC-AUC Analysis

The analysis of the receiver operating characteristic (ROC) curve was carried out in a one-vs-rest format to assess the discriminative power of the model on all the classes. The ROC curves (Fig. 4.3) are far above the diagonal reference line, implying that the classification performance is strong beyond that of random guessing.

The values of the Area Under the Curve (AUC) were extremely high:
Non-Demented: 0.999

Very Mildly Demented: 0.999

Mildly demented: 0.999

Moderately demented: 1.00

These findings show that there is good separability between the disease stages, and they indicate that the model is stable with variations in classification threshold. The ROC-AUC analysis also confirms the strength of the model and supplements the results achieved by values such as accuracy and F1-score.

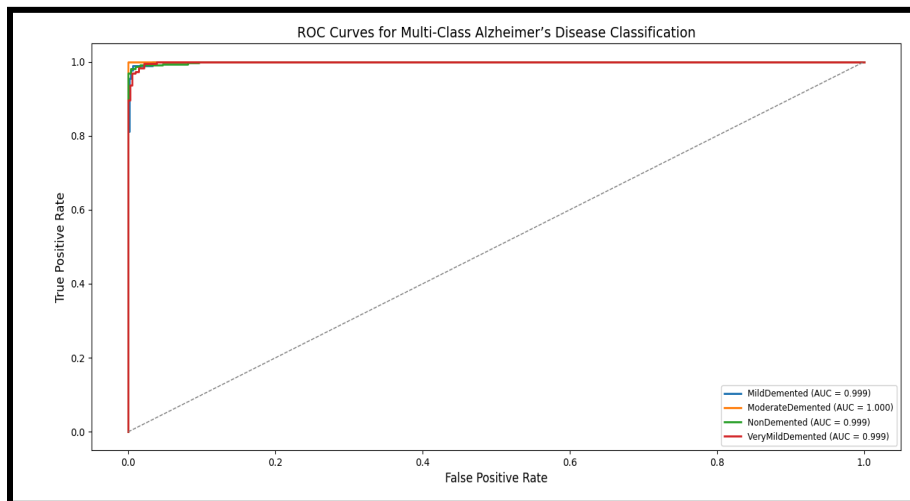


Fig. 8 Receiver Operating Characteristic (ROC) curves for a one-versus-rest method of multi-class Alzheimer's disease classification. Excellent discriminative performance is shown by the curves for all classes, which are located much above the diagonal reference line. Strong separability between disease stages and the robustness of the suggested CNN model are indicated by the high Area Under the Curve (AUC) values.

4.2.5 Explainability Analysis: Grad-CAM

Gradient-weighted Class Activation Mapping (Grad-CAM) was used to visualize parts of the model that contribute to model predictions. The heatmaps that were generated were overlaid on the original MRI scans to identify discriminatory brain regions. In the case of the non-demented group, the patterns of activation were spread across structurally normal brain areas, which means that no pathological characteristics were found. Conversely, the Mild and moderate demented classes exhibited focused activation in areas that are related to neurodegeneration, such as cortical thinning and structural atrophy. The patterns of activation of the very mildly demented class were more diffuse and less localized, which is reflected in the subtle anatomical changes typical of early-stage Alzheimer's disease. These results verify that the model is biologically and clinically relevant to the extent that the model is focused on irrelevant features. Grad-CAM is

further enhanced by transparency and visual justification of model predictions, thus raising the level of trust in its use as a clinical decision support system and diagnostic assistant.

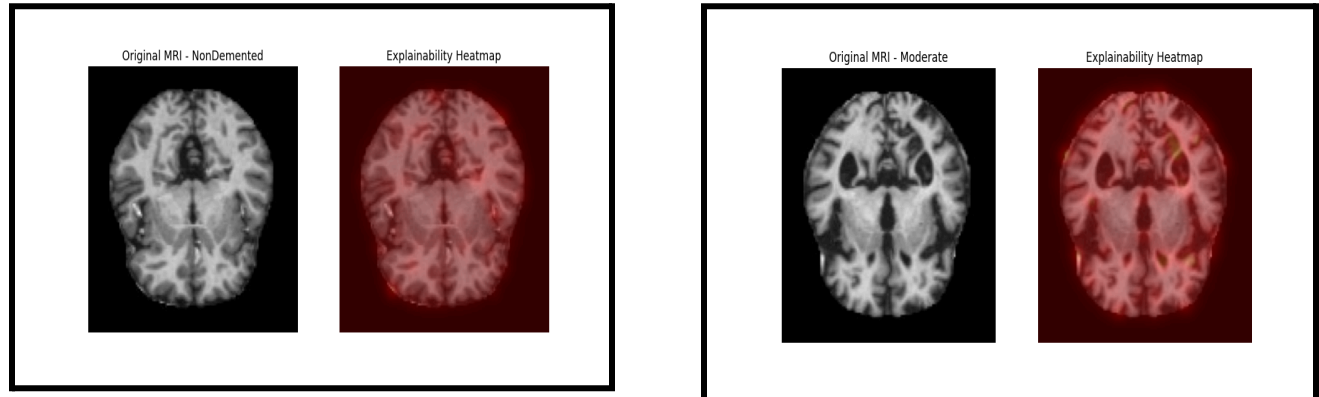


Fig. 9 (a). Visualization of Grad-CAM for the class without dementia. The absence of notable neurodegenerative patterns and the model's accurate classification are confirmed by the heatmaps' minimal and diffuse activation across structurally normal brain regions.

Fig. 9 (b): The very mildly demented class Grad-CAM visualization. Subtle and early-stage structural changes linked to the early stages of Alzheimer's disease progression are reflected in the activation maps' comparatively diffuse and less localized patterns.

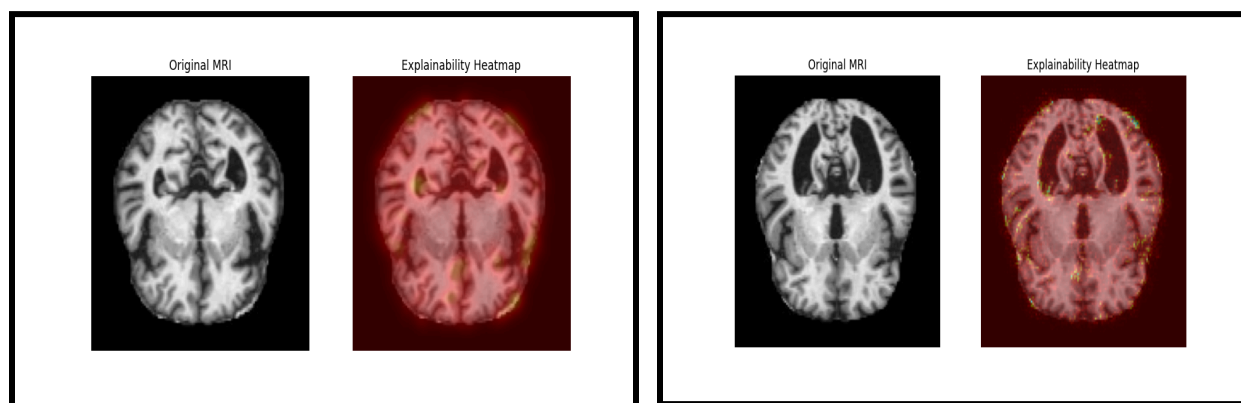


Fig. 9 (c) Grad-CAM visualization for the class with mild dementia. The model's prediction is supported by the heatmap, which indicates more concentrated activation in particular brain regions and suggests observable structural changes like cortical thinning and early atrophy.

Fig. 9 (d): The moderate demented class Grad-CAM visualization. Advanced stages of Alzheimer's disease are indicated by the activation maps' strong and localized regions of interest, which correspond to notable neurodegenerative alterations and substantial brain atrophy.

4.3 Bridging Transcriptomics and Deep learning in Neurodegenerative Disease

Transcriptomic analysis in a multi-level approach in the context of Amyotrophic Lateral Sclerosis (ALS) and MRI analysis based on deep learning in Alzheimer's disease (AD) gives a complementary multi-level understanding of neurodegenerative diseases. Transcriptomics and deep learning complement each other as they depict a comprehensive picture of the disease progression. Neuroinflammation, protein aggregation, oxidative stress and dysfunction of the ubiquitin-proteasome system (UPS) are shared pathological mechanisms, discussed in both studies. Impaired protein degradation and cellular regulation, which is also a key aspect of the AD pathology, has been identified as a major function of hub genes in ALS, such as SKP1A, CD44, and CALM13, in particular, amyloid-beta and tau deposits. To this end, CNN is able to identify structural changes, such as cortical atrophy, which are associated with these underlying molecular changes. Furthermore, as a whole, there is an overall link between the two approaches with the neuroinflammation emerging from the transcriptomic analysis, and the imaging results reflecting the structural consequences of the latter. This relationship is strengthened by applying Grad-CAM, which concentrates on clinically relevant parts of the brain in the classification of disease. Overall, this integrative approach demonstrates that a combination of bioinformatics and deep learning can enhance the understanding of diseases, produce a better diagnostic accuracy, and create a more effective strategy based on precision medicine to treat neurodegenerative disorders.

CHAPTER 5

DISCUSSION

In the present paper, an integrative and comprehensive approach to the study of neurodegenerative diseases is presented by combining transcriptomic analysis of Amyotrophic Lateral Sclerosis (ALS) with deep learning-based analysis of Alzheimer disease (AD) based on MRI. Combination of the molecular level knowledge and the structural neuroimaging results will enable the study to provide a multi-dimensional understanding of the disease mechanisms, progression and potential diagnostics.

Transcriptomic study of ALS spinal cord tissue revealed that there was a significant dysregulation of genes that relate to key biological pathways, such as the ubiquitin-proteasome system (UPS), neuroinflammation, and epigenetic regulation. The discovery of hub genes including SKP1A, CALM13, CD44, H3C7 and RPS14, demonstrates the importance of critical molecular drivers of disease pathology. SKP1A, a core adaptor of the SCF E3 ubiquitin ligase complex, was the most significantly upregulated hub gene. Seeing that impaired protein degradation is one of the hallmarks of ALS, the upregulation of SKP1A could be one of the forms of the compensatory cell response that is expected to help to clear the toxic protein aggregates like TDP-43. Nevertheless, even with this upregulation, the fact that protein aggregates persist is an indication that this compensatory action is not adequately responding to disease progression.

Likewise, the increase in the expression of CALM13, a calmodulin-like protein, highlights the importance of calcium dyshomeostasis and its subsequent effects on neuroinflammation. Calcium signalling is a highly important regulator of neuronal activity, and its deregulation can cause microglia and astrocyte activation to occur, contributing to non-cell-autonomous neuronal death. The CD44 downregulation, which is a major mediator of glial activation and inflammatory signalling, on the other hand, indicates a faulty or distorted glial response to disrupt neuroprotective mechanisms and accelerate the progression of the disease.

The downregulation of H3C7 and RPS14 added to the information as these are known to be the disruptions in epigenetic regulation and ribosomal activity, respectively. A decrease in the expression of H3C7 implies a disruption of extensive transcriptional reprogramming and epigenetic instability, which would be in line with the emerging view that ALS involves

widespread transcriptional reprogramming. Meanwhile, the down-regulation of RPS14 suggests the impairment of the process of ribosomal biogenesis and protein synthesis, which contributes to the cellular stress and accumulated misfolded proteins. All these findings are indicative of a highly stressed cellular environment in which compensatory mechanisms have been triggered but have ultimately failed as there are simultaneous dysfunctions in key cellular processes.

To complement these molecular findings, the MRI analysis of Alzheimer's disease using a deep learning-based algorithm revealed that a 2D Convolutional Neural Network (CNN) can effectively classify the different stages of disease progression with high accuracy (94%). The model demonstrated a high level of performance in distinguishing between non-demented and advanced disease stages, and the differentiation between very mild and mild stages of dementia was a challenge. This observation is consistent with the biological continuum of Alzheimer's disease, where subtle pathological changes are observed at the early stages, and in many cases, these pathological changes overlap each other in an anatomical manner.

Grad-CAM explainability offered valuable validation of the clinical relevance of the model. The activation map visualisation showed that the model is concentrated on significant brain areas related to neurodegeneration. Specifically, more localized and strong activations were observed in mild and moderate stages of dementia, which corresponded to regions of cortical atrophy, whereas diffuse patterns were observed in the very early stages. This shows that the model represents biologically relevant characteristics as opposed to dataset-specific artefacts, which improves model interpretability and potential clinical use.

Comparison of the proposed model with the existing studies shows that the proposed model performs competitively with higher accuracies on similar datasets reported heretofore. Nevertheless, the direct comparison is not completely possible because of the variations in datasets' composition, preprocessing methods, and assessment protocols. Also, the fact that it relies on 2D MRI slices is also a limitation since it does not fully capture the three-dimensional structural complexity of the brain. Future studies that integrate 3D CNN architectures may offer richer spatial representations and enhance the performance of classification.

In an integrative view, the two studies point to common pathological processes in the neurodegenerative diseases, especially the response of the cells to stressful situations. Although transcriptomic analysis directly measures the changes in molecular forms, deep learning-based imaging measures the structural implications of these changes. Such convergence supports the fact that the neurodegenerative diseases, though having clinical differences, are mediated by the overlapping biological processes. Although this study has strengths, a number of limitations should be mentioned. The reason behind the transcriptomic analysis is grounded in silico methods, which offer correlative information and need to be experimentally validated in cellular or animal models. On the same note, the deep learning model was trained and tested on the same dataset, which casts doubts on its ability to generalize to different populations and imaging conditions. Moreover, the absence of multimodal integration, i.e., the combination of imaging with genetic or clinical data, limits the extent of interpretation.

Future studies are encouraged to combine multi-omics information with more powerful deep learning models, allowing the creation of a more robust diagnostic framework that is generalizable. Incorporation of different datasets, longitudinal analysis and multimodal learning techniques might significantly contribute to the knowledge of disease progression and help to develop an individual approach to treatment.

To sum up, this paper has shown that transcriptomic and deep learning are two powerful tools to use to explore neurodegenerative diseases. It provides useful information about the mechanisms of diseases and emphasizes the possibilities of interdisciplinary approaches in the process of diagnosing, predicting, and treating diseases.

CHAPTER 6

CONCLUSION

In the present paper, an integrative and comprehensive approach to the study of neurodegenerative diseases is presented by combining transcriptomic analysis of Amyotrophic Lateral Sclerosis (ALS) with deep learning-based analysis of Alzheimer disease (AD) based on MRI. The combination of the molecular-level knowledge and the structural neuroimaging results will enable the study to provide a multi-dimensional understanding of the disease mechanisms, progression and potential diagnostics.

Transcriptomic study of ALS spinal cord tissue revealed that there was a significant dysregulation of genes that relate to key biological pathways, such as the ubiquitin-proteasome system (UPS), neuroinflammation, and epigenetic regulation. The discovery of hub genes, including SKP1A, CALM13, CD44, H3C7 and RPS14, demonstrates the importance of critical molecular drivers of disease pathology. These results support the multifactorial nature of ALS pathology and suggest possible molecular biomarkers in disease diagnosis and progression. Nevertheless, because this analysis is conducted based on *in silico* techniques and one dataset, the results are correlative and need to be further validated by conducting experimental studies, including but not limited to, qRT-PCR tests, protein-level tests, and functional tests using cellular and animal models.

To complement the molecular analysis, the deep learning-based framework to classify structural MRI data showed the usefulness of Convolutional Neural Networks (CNNs) to analyze structural MRI data. The proposed model has a high classification accuracy of 94% and has strong discriminative performance in various stages of the disease, as reflected by robust ROC-AUC values. Notably, Grad-CAM-based model interpretability highlights clinically relevant brain activity that is correlated with neurodegeneration, thus overcoming one of the most significant limitations of the black-box AI systems. Although these results are promising, a number of limitations still exist, including imbalance in classes (especially in the medium category of

dementia) as well as using 2D slices of images, resulting in the loss of 3D space information and lack of generalizability due to the use of a single dataset.

Another important contribution of this thesis is the conceptual combination of transcriptomics and deep learning methods. Whereas transcriptomic analysis can give insights into the underlying molecular mechanisms, deep learning can capture structural and phenotypic disease manifestations. This integrated view serves as a reminder of the need to integrate multi-level data to promote precision medicine in neurodegenerative diseases. These integrative mechanisms can potentially help enhance early diagnosis, biomarker discovery, and the development of targeted therapeutic interventions.

However, there are a number of challenges that need to be overcome in subsequent studies. These encompass the necessity of multi-omics integration (genomics, proteomics, and metabolomics), multimodal data (such as PET and functional MRI) and the creation of more interpretable and generalizable machine learning models. It further requires external validation of the work upon the use of several multicentred datasets to guarantee clinical applicability. Even better performance may be achieved by more sophisticated architectures that have access to more detailed spatial and contextual data.

To sum up, this article has shown that the combination of computational biology and artificial intelligence allows a potent paradigm to study complex neurodegenerative diseases. Although the outcomes of the study provide helpful information and a solid methodological framework, further interdisciplinary research is required to translate the techniques into clinical tools that can be validated and used in practice.

1.