Literature Review & Elaboration of Problem

Identification Of Risk Genes For Neurodegenerative Diseases

Version: [1.1]

Project Code	F23-335B
Internal Supervisor	Mr Shoaib Rauf
External Supervisor	-
Project Manager	-
Project Team	20K-0390 Robaisha Masood 20K-0266 Noor Afshan 20K-0374 Hira Tahir
Submission Date	04-Dec-2023

Document History

-Not Applicable.

Version	Name of Person	Date	Description of change

Distribution List

-Not Applicable.

Name	*Role*
	External Supervisor
	Project Manager

Document Sign-Off

-Not Applicable.

Version	*Sign-off Authority*	*Project Role*	*Sign-off Date*

Table Of Contents

1. Abstract

Our study targets neurodegenerative diseases that occur when neurons in the brain lose function and ultimately die. We aim to conduct research to identify key genes that govern the pathogens of such conditions and find proteins that are most vulnerable in each dataset. These genes form communities that cause malfunctioning. Two to three datasets will be used from NCBI GEO Data and DisGeNet. We will form individual networks of controlled and diseased from each dataset and find a key gene that has a major impact and find its similarities by making comparisons between each network. A Weighted Gene Co-expression Network (WGCNA) will be used to infer the hub genes and their role in such diseases. We will study different Machine learning algorithms that will help us to further tune and validate the results.

This research will give us a better understanding of ML algorithms and their implementation in the domain of bioinformatics which will surely help us to contribute to the community in the future. Our research aims to benefit the community on a large scale and raise awareness for neurodegenerative diseases that can lead to more effective treatments.

2. Background and Justification

Neurodegenerative diseases are chronic brain conditions when the neurons start to degenerate gradually due to abnormal accumulation and misfolding of specific proteins. Neurodegenerative diseases cause memory loss and behavioral abnormalities. The degeneration of neurons leads to neuronal damage and causes memory loss in the patient at different stages. Neurodegenerative disease syndrome includes Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic Lateral Sclerosis.

An estimated six million in the United States are afflicted with Alzheimer's Disease. This number is projected to double by 2050. Therefore our research will reveal the protein-to-protein interaction that will help in early diagnosis and drug development.

Recent research in the domain of neuroscience has discovered some factors that potentially cause the pathogenesis of neurodegenerative disease, especially Alzheimer's. One of the research hypotheses suggests that the mental levels of deregulation in the brain may play a very vital role in the development of neurodegenerative disorders. This hypothesis is supported by the observation of amyloid aggregation and oxidative stress, both of which are pathological factors associated with the disease. To further investigate the matter, our study will focus on identifying the potential hub genes that govern the pathogenesis of the disease. However, the main objective of our study is to identify the cancer-causing genes or the key genes. Identifying the key hub gene that plays a vital role in the pathogenesis of neurodegenerative disorders.

Our study will use computational biology for analyzing datasets from the NCBI GEO Database and DisGeNet. We will construct networks of control vs. disease conditions for each dataset. These networks will be further used to identify the hub genes significantly influencing the disease. These hub genes will further be evaluated using different algorithms, for instance, Weighted Gene Co-expression Network Analysis (WGCNA) and different community algorithms.

3. Problem Statement

What are the key genes that are responsible for neurodegenerative diseases? As neurodegenerative diseases like Alzheimer's and Parkinson's continue to rise which imposes societal challenges. For early diagnosis, it is very important to first identify the key genes that are causing the malfunction. Research is required to better understand these genetic interactions which in turn will contribute to drug development and enhance quality of life for the diseased community. Different Graph Learning, Community, and AI-based algorithms would help develop predictive models. In addition to this, our research will help us learn skills like data visualization and machine learning.

4. Literature Review

S.No.	Study	Findings
1	Network dysfunction perspective on neurodegenerative diseases (Year: 2017)	- Understanding the diseases by brain & neural networks - Targeted Diseases: Parkinson
2	Concepts and classification of neurodegenerative (Year: 2018)	- Detailed description on prion diseases - Targeted Diseases: Alpha-synucleinopathies
3	Application of machine learning to diagnosis and treatment of neurodegenerative diseases (Year: 2020)	- How machine learning helps in early diagnosis and its treatment - Targeted Diseases: Neuronal degenerative
4	A cell biological perspective of mitochondrial dysfunction in Parkinson diseases (Year: 2018)	- It caters with mutation of Parkinson diseases - Targeted Diseases: Parkinson
5	Neurodegenerative diseases and prions (Year: 2016)	- Prion diseases and accumulation of neural proteins - Targeted Diseases: Prion
6	Alzheimer's disease and other neurodegenerative diseases in elderly people (Year: 2021)	- It tells how schizophrenia works with Alzheimer's - Targeted Diseases: Schizophrenia + Alzheimer

7	Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease (Year: 2013)	- Meta-analysis identifies 11 new susceptibility loci for Alzheimer's - Targeted Diseases: Alzheimer's Disease
8	Identification of an Uncharted Alzheimer's Disease Locus near the Tau Gene (Year: 2016)	- Novel Alzheimer disease locus near the tau gene - Targeted Diseases: Alzheimer's Disease
9	Alzheimer's disease genetics: from the bench to the clinic (Year: 2014)	- Discusses Alzheimer's disease genetics from bench to clinic - Targeted Diseases: Alzheimer's Disease
10	Neurodegenerative disorders associated with genes of mitochondria (Year: 2021)	- Mitochondrial genes have provided compelling evidence that mitochondria are involved in the initiation as well as progression of diseases. - Targeted Diseases: Parkinson’s disease (PD), Alzheimer’s disease (AD), Huntington's disease (HD), and Friedreich ataxia (FA).
11	Examining the Polygenic Variation Enhancing Alzheimer's Disease Risk Prediction (Year: 2015)	- Examines how common polygenic variation enhances Alzheimer's disease risk prediction. - Targeted Diseases: Alzheimer's Disease
12	Identification of risk genes for Alzheimer’s disease by gene embedding (Year: 2022)	- The study investigates the role of network dysfunction in neurodegenerative diseases, specifically focusing on Parkinson's disease. - Targeted Diseases: Parkinson’s Disease

13	A network dysfunction perspective on neurodegenerative diseases	-This research paper discusses the varying abilities seen in patients with neurodegenerative disorders (like Alzheimer's) within a day. Researchers suggest that these fluctuations don't come from losing or gaining nerve cells but possibly from changes in the activity of neural networks - think of it like changing traffic patterns in a city.
14	Weighted gene coexpression network analysis and machine learning reveal oncogenome associated microbiome plays an important role in tumor immunity and prognosis in pan-cancer (Year: 2021)	-The methodology used in the study conducted by Guan et al. is related to our project on identifying coexpression networks of genes for neurodegenerative diseases like Alzheimer's in terms of using weighted gene coexpression network analysis (WGCNA) to analyze genomic data.
15	Co-expression Network Analysis Reveals Novel Genes Underlying Alzheimer’s Disease Pathogens (Year: 2020)	- The study involved functional enrichment analysis, protein-protein interaction network analysis, and gene set enrichment analysis for the hub genes. The results suggest that the identified genes are mainly involved in axonogenesis, synaptic transmission, and ion transport, among other processes.
16	Machine learning and deep learning algorithms used to diagnosis of Alzheimer’s (Year: 2021)	- This article discusses the role of AI algorithms in medical diagnosis , specifically focusing on their contribution to classifying Alzheimer's disease. This study aims to consolidate relevant knowledge about machine learning models used in Alzheimer's identification, presenting results from different algorithms employed for diagnosis.
17	Detection and analysis of Alzheimer’s disease using various machine learning algorithms (Year: 2021)	- This paper explores the potential of artificial intelligence (AI) and machine learning (ML) algorithms to enhance Alzheimer's detection. Recognizing the limitations of conventional approaches, the study employs big data processing to gather information from diverse sources, considering the evolving nature of the disease.
18	Alzheimer's Disease Prediction using Machine Learning and Transfer Learning Models (Year: 2021)	- This paper focuses on the detection of Alzheimer's using two datasets: the Longitudinal dataset with text values and the OASIS dataset containing MRI images.

19	Early diagnosis of Alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach (Year: 2022)	- This study addresses the crucial need for early and accurate diagnosis of Alzheimer's disease (AD) using a multi-diagnostic and generalizable approach. Leveraging machine learning (ML) and structural MRI, the study develops classifiers for mild cognitive impairment (MCI) and AD diagnosis, demonstrating transparency, interpretability, and generalizability across datasets and MRI protocols.
20	Alzheimer's Disease Detection Using Different Machine Learning Algorithms (Year: 2021)	This paper explores the application of machine learning algorithms, including Decision Trees, SVM, Logistic Regression, and Naive Bayes, to identify Alzheimer's disease at an early stage.
21	Identification of feature genes and pathways for Alzheimer's disease via WGCNA and LASSO regression (Year: 2022)	- This study aimed to identify feature genes associated with Alzheimer's disease (AD) by analyzing data from three Gene Expression Omnibus (GEO) databases: GSE122063, GSE15222, and GSE138260.
22	WGCNA: an R package for weighted correlation network analysis (Year: 2008)	-This approach involves identifying clusters (modules) of highly correlated genes, summarizing these clusters, and relating them to external sample traits.
23	Co-expression network-based analysis of hippocampal expression data associated with Alzheimer's disease using a novel algorithm (Year: 2021)	-The study explores the shared pathophysiological mechanisms of Alzheimer's Disease (AD) and Type 2 Diabetes Mellitus (T2DM) through co-expression network analysis.
24	Alzheimer's disease: connecting findings from graph theoretical studies of brain networks (Year: 2013)	-The study explores the interrelationships between pathological processes and clinical phenotypes in Alzheimer's disease (AD) using graph theory.
25	Functional connectivity and graph theory in preclinical Alzheimer's disease (Year: 2014)	-This study investigated graph theory metrics in the context of Alzheimer's disease (AD). Resting-state functional connectivity MRI was utilized to assess functional integration, functional segregation, and functional distinctness.

Neurodegenerative diseases are a significant challenge for public health. The study of neurodegenerative disorders demands a needful understanding of the genetic foundations. This literature review highlights the key findings from the recent research in neurodegenerative disease genetics.

The authors of the study introduce a novel approach called GeneEMBED, which aims to identify gene interactions associated with complex diseases such as Alzheimer's. Through the application of GeneEMBED on multiple datasets related to Alzheimer's, researchers successfully discovered previously unidentified genes (namely PLEC, UTRN, TP53, and POLD1) that play a role in this disease. Additionally, it was observed that two out of these four genes are targeted by approved pharmaceutical drugs. **[1]**

The "Network dysfunction perspective" on neurodegenerative diseases, as proposed by Jorge J. Palop, Jeannie Chin, and Lennart Mucke, focuses on understanding these diseases with the help of examination of broader neural networks and connection within the brain, rather than working with a single individual area. **[2]**

The paper "Chapter 21 - Concepts and Classification of Neurodegenerative Diseases" by Gabor G. Kovacs discusses neurodegenerative diseases and follows the classification by providing a detailed description of the neuropathology of Alzheimer's disease, alpha-synucleinopathies, tauopathies, FTLD with TDP-43 or FUS/FET proteinopathies, trinucleotide repeat disorders, and prion diseases. **[3]**

" Applications of machine learning to diagnosis and treatment of neurodegenerative diseases" by Monika A. Myszczynska, Poojitha N. Ojamies, …Laura Ferraiuolo explains how machine learning is helping in diagnosis in the early stage and interpretation of medical images as well as the discovery and development of new therapies. This also helps in multiple high-dimensional sources of data that provide different views on different diseases. **[4]**

Wim Mandemakers 1, Vanessa A Morais, and Bart De Strooper's article "A cell biological perspective on mitochondrial dysfunction in Parkinson's disease and other Neurodegenerative diseases"; focus on mutations of reasons causing Parkinson's disease.
It tells the study of these mutations and it causes other diseases that indicate mitochondrial dysfunction which becomes the main contributor to neurodegenerative processes. **[5]**

According to Stanley B. Prusiner, M.D.'s "Neurodegenerative Diseases and Prions," it is obvious that prion or neurodegenerative diseases are caused by anomalies in the course of these illnesses, which further cause the accumulation of particular neuronal proteins. Laboratory search results led to the discovery of prions, yielding findings like infectious pathogens and central nervous system degeneration. **[6]**

"Alzheimer Disease and Related Neurodegenerative Diseases in Elderly Patients With Schizophrenia" by Dushyant P. Purohit, Daniel P. Perl, Vahram Haroutunian, et al. Clinical studies suggest that severe cognitive impairment is common in older people with schizophrenia who reside mostly in psychiatric people and that its result is in conflict with Alzheimer's disease when combined with schizophrenia. **[7]**

The study, “Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.” by Lambert, J. C., et al. (2013); was conducted on 74,046 individuals, which led to 11 discoveries associated with Alzheimer’s disease. This study served a great purpose in highlighting the importance of large-scale collaborative research in deciphering the complex genetic landscape of this condition. **[8]**

Jun et al. (2016) presented their opinion in the novel by identifying an uncharted Alzheimer's disease locus near the tau gene. This discovery essentially shows that there could be a potential link between tau protein pathology and susceptibility to Alzheimer’s disease. **[9]**

Karch, C. M., et al. (2014) presented their study, " Alzheimer's Disease Genetics: from the Bench to the Clinic.". The key discussion in this paper was the need to build the gap between genetic discoveries and clinical applications in Alzheimer's disease genetics. Their work highlights the genetic factors involved in Alzheimer's disease and it also underscores the importance of applying this knowledge in clinical practice to potentially improve the diagnosis and the treatments. **[10]**

John Hardy and Dennis J. Selkoe's paper, "Alzheimer's disease: The amyloid cascade hypothesis," examines emphasis on the causal link between mitochondrial gene mutations and neurodegenerative disorders such as Parkinson's and Alzheimer's diseases. It reviews structural and functional studies, highlighting mitochondrial dysfunction's role. In the study, animal models provide evidence for mitochondria's involvement in disease initiation and progression. **[11]**

"Neurodegenerative disorders associated with genes of mitochondria" by Vaibhav S. Marde et al. (2021); emphasizes the causal link between mitochondrial gene mutations and neurodegenerative disorders such as Parkinson's and Alzheimer's diseases. It reviews structural and functional studies, highlighting mitochondrial dysfunction's role. In the study, animal models provide evidence for mitochondria's involvement in disease initiation and progression. **[12]**

Common polygenic variations have a significant impact on Alzheimer's disease research, according to Escott-Price, who highlights how these genetic variations improve risk assessment, early diagnosis, and stage prediction. Microglial-Mediated Innate Immunity. **[13]**

The methodology used in the study conducted by Guan et al. is related to our project on identifying coexpression networks of genes for neurodegenerative diseases like Alzheimer's in terms of using weighted gene coexpression network analysis (WGCNA) to analyze genomic data. WGCNA is a powerful tool for identifying modules or clusters of genes that are highly correlated and may play a role in disease mechanisms. By constructing separate networks for tumor microbial data and messenger RNA (mRNA) data, the study identified the oncogenome associated microbiome module (OAM), which was found to be tumor-specific enriched and had a better prognostic value when combined with the oncogenome. In our project, we aim to apply similar techniques to analyze gene expression data in neurodegenerative diseases like Alzheimer's, with the goal of identifying coexpression networks that could potentially reveal novel biomarkers or therapeutic targets. **[14]**

This open-access article presents a study on Alzheimer's disease pathogenesis, in which researchers used co-expression network analysis to identify novel genes related to AD. The study involved functional enrichment analysis, protein-protein interaction network analysis, and gene set enrichment analysis for the hub genes. The results suggest that the identified genes are mainly involved in axonogenesis, synaptic transmission, and ion transport, among other processes. The study provides new insights into the molecular mechanisms underlying AD and may contribute to the development of new treatments for this disease.
In this study, the researchers used the Weighted Gene Co-expression Network Analysis (WGCNA) method to construct a co-expression network for a gene dataset with a large number of samples from the brain tissues of AD patients and normal controls. They then screened the hub genes that were related to AD pathogenesis and validated the robustness of the expression of hub genes using an independent validated cohort. They also analyzed the pathways and clinical significance of the hub genes.
Overall this study provides a useful example of how co-expression network analysis can be used to identify novel genes related to AD pathogenesis, and may provide insights that can be applied to your research on identifying risk genes for AD. **[15]**

This article discusses the role of AI algorithms in medical diagnosis , specifically focusing on their contribution to classify Alzheimer's disease. This study aims to consolidate relevant knowledge about machine learning models used in Alzheimer's identification, presenting results from different algorithms employed for diagnosis. This paper tells further about work done in the medical research field that is related to Alzheimer's disease, comparing the efficiency and error rates. The article highlights the shift from methods relying solely on individual features from sMRI images for classification, which often resulted in low accuracy. Instead, recent multimodal studies demonstrate improved classification accuracy by combining multiple features from different sMRI analysis techniques. This approach enhances the understanding and diagnosis of Alzheimer's disease and its stages. **[16]**

This paper explores the potential of artificial intelligence (AI) and machine learning (ML) algorithms to enhance Alzheimer's detection. Recognizing the limitations of conventional approaches, the study employs big data processing to gather information from diverse sources, considering the evolving nature of the disease. Previous research using Support Vector Machine (SVM) showed low accuracy, prompting the need for improved precision. The paper presents alternative algorithms to enhance efficiency, revealing that the Support Vector Machine with a linear kernel outperforms other models. The existing method utilizing SVM for prediction faces challenges such as time inefficiency, particularly in handling large, heterogeneous datasets. The paper highlights the drawbacks of traditional data mining tools and proposes a new approach using algorithms like Linear Regression, SVM, Decision Tree, Random Forest, and Naïve Bayes. The study recognizes the complexity of big data processing and emphasizes the role of supercomputing platforms in overcoming obstacles at the data, model, and system levels. Advantages of the proposed method over the conventional SVM approach include faster business value realization, early detection of Alzheimer's, identification of optimal treatment modes across different age groups, and insights driving growth and profitability. **[17]**

The brain, a remarkable organ overseeing various physical and cognitive functions, undergoes structural and functional changes in Alzheimer's Disease (AD), a chronic condition contributing to the majority of dementia cases. Early diagnosis is crucial for effective intervention, as AD worsens over time. Traditional testing methods are time-consuming, prompting the exploration of machine learning models for efficient and early detection. This paper focuses on the detection of Alzheimer's using two datasets: the Longitudinal dataset with text values and the OASIS dataset containing MRI images. Utilizing 14 machine learning algorithms, including the Random Forest Algorithm with a peak accuracy of 92.1385% and the KNN Algorithm with a baseline accuracy of 47.1910%, the OASIS Longitudinal dataset demonstrates efficient detection capabilities. For MRI images, various Transfer learning models are employed, with the InceptionV3 model and ADAM optimizer yielding the highest accuracy. The study emphasizes the potential of machine learning in achieving accurate and early Alzheimer's detection, presenting a multifaceted approach by considering diverse datasets and models. **[18]**

This study addresses the crucial need for early and accurate diagnosis of Alzheimer's disease (AD) using a multi-diagnostic and generalizable approach. Leveraging machine learning (ML) and structural MRI, the study develops classifiers for mild cognitive impairment (MCI) and AD diagnosis, demonstrating transparency, interpretability, and generalizability across datasets and MRI protocols. The classifiers, trained and tested on the AD Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) databases, achieve impressive results. The "healthy controls (HC) vs. AD" classifier attains a balanced accuracy of 90.6% and a Matthew's correlation coefficient (MCC) of 0.811. The "HC vs. MCI vs. AD" classifier, trained on the ADNI dataset, achieves a balanced accuracy of 62.1%, with hippocampal features identified as significant contributors. The study's diagnostic tool is unique in its multi-diagnostic approach, considering HC, MCI, and AD simultaneously, providing a more practical solution for clinicians. It also addresses the challenge of generalizability by evaluating performance across independent datasets and various MRI protocols. The

study presents a promising ML-based diagnostic tool for MCI and AD, offering a comprehensive and clinically applicable solution. The multi-diagnostic approach, generalizability across datasets and protocols, and transparent reporting contribute to the robustness and potential clinical utility of the proposed classifier. **[19]**

This paper explores the application of machine learning algorithms, including Decision Trees, SVM, Logistic Regression, and Naive Bayes, to identify Alzheimer's disease at an early stage. Utilizing datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Investigations (OASIS), which include longitudinal MRI data and demographic information, the study aims to enhance diagnostic efforts. The datasets encompass factors such as age, gender, mini mental status, and Clinical Dementia Rating (CDR). Precision, F1 Score, Recall, and specificity are considered in evaluating each method. The results reveal that the Decision Tree Algorithm achieves a maximum accuracy of 93.7%, showcasing its effectiveness in early Alzheimer's disease detection. Alzheimer's disease is highlighted as a progressive condition with initial signs including difficulty recalling recent events. The study emphasizes the challenges in clinical diagnosis, leading to the exploration of automated systems for medical decision support. **[20]**

This study aimed to identify feature genes associated with Alzheimer's disease (AD) by analyzing data from three Gene Expression Omnibus (GEO) databases: GSE122063, GSE15222, and GSE138260. The datasets were filtered based on AD-related keywords, Homo sapiens as the selected species, and a sample size greater than 20 for each dataset, including both normal and AD groups. GSE15222 and GSE138260 were combined as a training group to build a model, while GSE122063 served as a test group for model verification. Weighted gene co-expression network analysis (WGCNA) on the combined dataset identified AD-related module genes. The intersection of differential and AD-related module genes resulted in AD key genes. LASSO regression further filtered these genes, leading to the identification of three AD-related feature genes: SST, MLIP, and HSPB3. Differential expression analysis in the combined datasets revealed 111 common differential AD genes. Gene Ontology (GO) analysis highlighted terms related to cognition, learning, and memory. Kyoto Encyclopedia of Genes and Genome Pathways (KEGG) analysis identified enrichment in pathways such as neuroactive ligand-receptor interaction, cAMP signaling, and Calcium signaling. **[21]**

Weighted Gene Co-expression Network Analysis (WGCNA) is a widely used method in bioinformatics for exploring correlation patterns among genes in microarray samples. This approach involves identifying clusters (modules) of highly correlated genes, summarizing these clusters, and relating them to external sample traits. The WGCNA R software package, presented in this paper, offers a comprehensive set of functions for performing various aspects of weighted correlation network analysis. These functions include network construction, module detection, gene selection, topological property calculations, data simulation, visualization, and integration with external software. The package is user-friendly and can be applied not only to gene expression data but also to various biological contexts, such as cancer, genetics, and brain imaging data. The accompanying tutorials enhance the accessibility of the software, making it a valuable tool for researchers in the field. **[22]**

The study explores the shared pathophysiological mechanisms of Alzheimer's Disease (AD) and Type 2 Diabetes Mellitus (T2DM) through co-expression network analysis. Microarray data of AD and T2DM were obtained from the Gene Expression Omnibus (GEO) database, and co-expression networks were constructed using Weighted Gene Co-Expression Network Analysis (WGCNA). Gene Ontology (GO) and pathway enrichment analyses were performed on common genes related to AD and T2DM modules. The results revealed significant modules for both AD and T2DM, enriched in pathways such as circadian entrainment, phagosome, glutathione metabolism, and synaptic vesicle cycle. Protein-protein interaction network analysis identified 10 hub genes (CALM1, LRRK2, RBX1, SLC6A1, TXN, SNRPF, GJA1, VWF, LPL, AGT) shared

between AD and T2DM. The findings suggest common pathogenesis and shared pathways, providing insights for further mechanistic studies and potential therapeutic targets for AD and T2DM. **[23]**

The study explores the interrelationships between pathological processes and clinical phenotypes in Alzheimer's disease (AD) using graph theory. Neuroimaging techniques reveal disruptions in brain networks, with findings pointing to a loss of highly connected areas in AD. However, variability in reported group differences suggests non-isometric brain graphs, complicating interpretation. The use of graph theory provides quantitative measurements for understanding the integrated nature of local brain activity, highlighting the significance of hubs in information processing and disease propagation within the network. Confounding factors, such as differences in graph construction methods, are identified, and recommendations for future research are provided. **[24]**

This study investigated graph theory metrics in the context of Alzheimer's disease (AD). Resting-state functional connectivity MRI was utilized to assess functional integration, functional segregation, and functional distinctness. The study found reductions in clustering coefficient and modularity, indicative of large-scale disconnection in symptomatic AD. Cognitively normal participants with preclinical AD biomarkers also exhibited similar, albeit smaller, reductions in these graph measures. The impact on modularity was influenced by age, and AD was observed to affect hub-like regions in the brain. These findings suggest significant brain changes in preclinical AD, emphasizing the relevance of large-scale disconnection even before symptomatic onset. **[25]**

In summary, this literature review offers a rich tapestry of insights into neurodegenerative diseases, with a particular emphasis on identifying risk genes. The GeneEMBED approach introduces an innovative method for uncovering genes associated with Alzheimer's, showcasing the potential for targeted gene identification. The exploration of network dysfunction, disease classification, and the integration of machine learning not only broadens our understanding but also provides practical avenues for identifying crucial risk genes. The diverse perspectives on mitochondrial dysfunction, prions, and disease co-occurrence contribute to a holistic comprehension of the genetic landscape. As our research focuses on pinpointing risk genes, this review lays a robust foundation, offering a nuanced and interdisciplinary perspective for advancing our understanding of neurodegenerative diseases at the genetic level.

5. Appendices

-Not Applicable

6. References

[1] Y. Lagisetty et al., “Identification of risk genes for Alzheimer’s disease by gene embedding,” Cell Genomics, vol. 2, no. 9, p. 100162, Sep. 2022, doi: 10.1016/j.xgen.2022.100162.

[2] J. J. Palop, J. Chin, and L. Mucke, “A network dysfunction perspective on neurodegenerative diseases,” Nature, vol. 443, no. 7113, pp. 768–773, Oct. 2006, doi: 10.1038/nature05289.

[3] G. G. Kovacs, “Concepts and classification of neurodegenerative diseases,” in Handbook of Clinical Neurology, 2018, pp. 301–307. doi: 10.1016/b978-0-12-802395-2.00021-3.

[4] M. A. Myszczynska et al., “Applications of machine learning to diagnosis and treatment of neurodegenerative diseases,” Nature Reviews Neurology, vol. 16, no. 8, pp. 440–456, Jul. 2020, doi: 10.1038/s41582-020-0377-8.

[5] W. Mandemakers, V. A. Morais, and B. De Strooper, "A cell biological perspective on mitochondrial dysfunction in Parkinson disease and other neurodegenerative diseases," Journal of Cell Science, vol. 120, no. 10, pp. 1707–1716, May 2007, doi: 10.1242/jcs.03443.

[6] S. B. Prusiner, "Neurodegenerative diseases and prions," The New England Journal of Medicine, vol. 344, no. 20, pp. 1516–1526, May 2001, doi: 10.1056/nejm200105173442006.

[7] D. P. Purohit, D. P. Perl, V. Haroutunian, P. Powchik, M. Davidson, and K. L. Davis, "Alzheimer disease and related neurodegenerative diseases in elderly patients with schizophrenia," Archives of General Psychiatry, vol. 55, no. 3, p. 205, Mar. 1998, doi: 10.1001/archpsyc.55.3.205.

[8] J.-C. Lambert et al., "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease," Nature Genetics, vol. 45, no. 12, pp. 1452–1458, Oct. 2013, doi: 10.1038/ng.2802.

[9] G. Jun et al., "A novel Alzheimer disease locus located near the gene encoding tau protein," Molecular Psychiatry, vol. 21, no. 1, pp. 108–117, Mar. 2015, doi: 10.1038/mp.2015.23.

[10] C. M. Karch, C. Cruchaga, and A. M. Goate, "Alzheimer's Disease Genetics: From the bench to the clinic," Neuron, vol. 83, no. 1, pp. 11–26, Jul. 2014, doi: 10.1016/j.neuron.2014.05.041.

[11] J. Hardy and G. A. Higgins, "Alzheimer's Disease: The Amyloid Cascade hypothesis," Science, vol. 256, no. 5054, pp. 184–185, Apr. 1992, doi: 10.1126/science.1566067.

[12] V. S. Marde et al., "Neurodegenerative disorders associated with genes of mitochondria," Future Journal of Pharmaceutical Sciences, vol. 7, no. 1, Mar. 2021, doi: 10.1186/s43094-021-00215-5.

[13] V. Escott-Price et al., "Common polygenic variation enhances risk prediction for Alzheimer's disease," Brain, vol. 138, no. 12, pp. 3673–3684, Oct. 2015, doi: 10.1093/brain/awv268.

[14] Guan et al. "Weighted gene coexpression network analysis and machine learning reveal oncogenome associated microbiome plays an important role in tumor immunity and prognosis in pan-cancer." Journal of Translational Medicine 21.537 (2023): 1-21.

[15] Hu, Y., Yu, X., Zhou, Y., Yin, X., Hu, J., Lu, X., & Hu, C. (2020). Identification of novel genes related to Alzheimer's disease pathology using genome-wide co-expression network analysis and supervised machine learning. Frontiers in Aging Neuroscience, 12, 605961. doi: 10.3389/fnagi.2020.605961

[16] S. Balne and A. Elumalai, "Machine learning and deep learning algorithms used for the diagnosis of Alzheimer's" in Materials Today: Proceedings, vol. 47, no. 15, pp. 5151-5156, 2021, doi: 10.1016/j.matpr.2021.05.499.

[17] P. Kishore, Ch. Usha Kumari, M.N.V.S.S. Kumar, and T. Pavani, "Detection and analysis of Alzheimer's disease using various machine learning algorithms," in Materials Today: Proceedings, vol. 45, no. 2, pp. 1502-1508, 2021, doi: 10.1016/j.matpr.2020.07.645.

[18] A. Amrutesh, G. B. C G, A. A, A. R. KP and G. S, "Alzheimer's Disease Prediction using Machine Learning and Transfer Learning Models," 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/CSITSS57437.2022.10026365.

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

*[20] S. Harika, T. Yamini, T. Nagasaikamesh, S. H. Basha, S. Santosh Kumar, and Mrs. S. Sri DurgaKameswari, "Alzheimer's Disease Detection Using Different Machine Learning Algorithms," *International Journal of Research in Advent Technology (IJRASET)*, DOI: 10.22214/ijraset.2022.46937.*

[21] H. Sun, J. Yang, X. Li, Y. Lyu, Z. Xu, H. He, X. Tong, T. Ji, S. Ding, C. Zhou, P. Han, J. Zheng, "Identification of feature genes and pathways for Alzheimer's disease via WGCNA and LASSO regression," Frontiers in Computational Neuroscience, vol. 16, p. 1001546, Sep. 2022, doi: 10.3389/fncom.2022.1001546.

[22] Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008), https://doi.org/10.1186/1471-2105-9-559

*[23] Hong Yue, Bo Yang, Fang Yang, Xiao-Li Hu, Fan-Bin Kong, "Co-expression network-based analysis of hippocampal expression data associated with Alzheimer's disease using a novel algorithm," *Experimental and Therapeutic Medicine*, vol. 11, no. 5, pp. 1707-1715, 2016.*

[24] B. M. Tijms, A. M. Wink, W. de Haan, W. M. van der Flier, C. J. Stam, P. Scheltens, and F. Barkhof, "Alzheimer's disease: connecting findings from graph theoretical studies of brain networks," Neurobiology of Aging, vol. 34, no. 8, pp. 2023-2036, Aug. 2013. https://doi.org/10.1016/j.neurobiolaging.2013.02.020

*[25] Brier, M. R., Thomas, J. B., Fagan, A. M., Hassenstab, J., Holtzman, D. M., Benzinger, T. L., Morris, J. C., & Ances, B. M. (2014). Functional connectivity and graph theory in preclinical Alzheimer's disease. *Neurobiology of Aging*, 35(4), 757-768. https://doi.org/10.1016/j.neurobiolaging.2013.10.081*

[26]Q. Jing, H. Zhang, X. Sun, Y. Xu, S. Cao, Y. Fang, X. Zhao, and C. Li, "A Comprehensive Analysis Identified Hub Genes and Associated Drugs in Alzheimer's Disease," vol. 2021, Article ID 8893553, 2021, https://doi.org/10.1155/2021/8893553.

[27]C. Zou, L. Su, M. Pan, L. Chen, H. Li, C. Zou, J. Xie, X. Huang, and M. Lu, "Exploration of novel biomarkers in Alzheimer's disease based on four diagnostic models," Front. Aging Neurosci., vol. 15, p. 1079433, Feb. 16, 2023. [Online]. Available: https://doi.org/10.3389/fnagi.2023.1079433

[28]Zhang, Y., Kiryu, H. Identification of oxidative stress-related genes differentially expressed in Alzheimer's disease and construction of a hub gene-based diagnostic model. Sci Rep 13, 6817 (2023). https://doi.org/10.1038/s41598-023-34021-1

[29]A. Sharma and P. Dey, "A machine learning approach to unmask novel gene signatures and prediction of Alzheimer's disease within different brain regions," Genomics, vol. 113, no. 4, pp. 1778-1789, 2021, ISSN 0888-7543, https://doi.org/10.1016/j.ygeno.2021.04.028.

7. Bibliography

-Not Applicable