****

**FAST NATIONAL UNIVERSITY OF COMPUTER AND EMERGING SCIENCES**

**(KARACHI CAMPUS)**

**Department of Computer Science**

**Spring 2024**

**CS4091 Final Year Project**

**Batch 2020**

**“Identification Of Risk Genes For Neurodegenerative Disease”**

**Progress Report**

**F23-335 B**

**Group Members:**

* Robaisha Masood 20K-0390
* Noor Afshan 20K-0266
* Hira Tahir 20K-0374

**Internal Supervisor:**

* Mr Shoaib Rauf

### 

### 

### 

### INTRODUCTION

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 hypothesis suggests that the metal levels 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.

### 

### 

### 

### TIMELINE

* Work done until task 4, Jan 2024.

|  | **Sep** | **Oct** | **Nov** | **Dec** | **Jan** | **Feb** | **Mar** | **Apr** | **May** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***1.Study and Research*** | **Task 1** | **Task 1** |  |  |  |  |  |  |  |
| ***2. Data Collection and Data Preprocessing*** |  |  | **Task 2** |  |  |  |  |  |  |
| ***3. Construction Of GRN*** |  |  |  | **Task 3** |  |  |  |  |  |
| ***4. ML Algorithm Research*** |  |  |  |  | **Task 4** |  |  |  |  |
| ***5. Construction Of WCGNA*** |  |  |  |  |  | **Task 5** |  |  |  |
| ***6. Functional Enrichment Analysis*** |  |  |  |  |  |  | **Task 6** |  |  |
| ***7.Validation of results by ML and different algorithms*** |  |  |  |  |  |  |  | **Task 7** |  |
| ***8.Research paper writing*** |  |  |  |  |  |  |  |  | **Task 8** |

### 

### PROGRESS

#### MILESTONE ACHIEVED TILL FYP 1

1. Data collection :

* Dataset GSE118553 downloaded.

1. Data PreProcessing :

* Normalization of dataset

1. Data Transformation

* Extract insights from data
* Construction of Views

1. Network Construction

* Downloading Trust Alias
* Performing Join
* Construction of GRN

1. ML Algorithm research for Key Gene Identification

#### FUTURE WORK

1. Construction of WGCNA: Weighted correlation Network Analysis is a data mining technique for studying biological networks based on pairwise correlation between variables (Genes).
2. Validation Of Results by ML:Training and testing of different ML models ( SVM, Random Forest , Logistic Regression, AdaBoost).
3. Research Paper Writing:

### REFERENCES

[1] Lagisetty Y;Bourquard T;Al-Ramahi I;Mangleburg CG;Mota S;Soleimani S;Shulman JM;Botas J;Lee K;Lichtarge O; (no date) *Identification of risk genes for alzheimer’s disease by Gene Embedding*, *Cell genomics*. Available at: https://pubmed.ncbi.nlm.nih.gov/36268052/ (“[Identification of risk genes for Alzheimer's disease by gene embedding”](https://pubmed.ncbi.nlm.nih.gov/36268052/)) (Accessed: 28 February 2024).

[2] *(PDF) a network dysfunction perspective on neurodegenerative diseases*. Available at: <https://www.researchgate.net/publication/6745022_A_Network_dysfunction_perspective_on_neurodegenerative_diseases>

[3] *Neuropathology, Volume 145 - 1st Edition | Elsevier Shop*. Available at: <https://www.elsevier.com/books/neuropathology/kovacs/978-0-12-802395-2> (Accessed: 28 February 2024).

[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. <https://pubmed.ncbi.nlm.nih.gov/32669685/>

[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. <https://pubmed.ncbi.nlm.nih.gov/17502481/>

[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. <https://www.nejm.org/doi/full/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. <https://jamanetwork.com/journals/jamapsychiatry/article-abstract/191033>

[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. <https://pubmed.ncbi.nlm.nih.gov/24162737/>

[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. <https://pubmed.ncbi.nlm.nih.gov/25778476/>

[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. <https://www.sciencedirect.com/science/article/pii/S0896627314004851>

[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. <https://pubmed.ncbi.nlm.nih.gov/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. <https://fjps.springeropen.com/articles/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. <https://academic.oup.com/brain/article/138/12/3673/411685?login=false>

[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. <https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-023-04411-0>

[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 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9581494/>

[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. <https://www.sciencedirect.com/science/article/abs/pii/S2214785321041018>

[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. <https://www.sciencedirect.com/science/article/abs/pii/S2214785320357709>

[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. <https://www.mdpi.com/2504-4990/6/1/24>

[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. Nagasaki Kamesh, S. H. Basha, S. Santosh Kumar, and Mrs. S. Sri Durga Kameswari, "Alzheimer's Disease Detection Using Different Machine Learning Algorithms,"

\*International Journal of Research in Advent Technology (IJRASET)\*, DOI: 10.22214/ijraset.2022.46937. <https://www.ijraset.com/research-paper/alzheimers-disease-detection-using-different-ml-algorithms>

[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. <https://www.frontiersin.org/articles/10.3389/fncom.2022.1001546/full>

[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>. <https://bmcbioinformatics.biomedcentral.com/articles/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. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4840697/>

[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.

\*NeurobiologyofAging\*,35(4),757-768.[[https://doi.org/10.1016/j.neurobiolaging.2013.10.081](https://doi.org/10.1016/j.neurobiolaging.2013.10.081)](https://doi.org/10.1016/j.neurobiolaging.2013.10.081%5D(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 gep 13, 6817 (2023). [https://doi.org/10.1038/s41598-023-34021-1ene-based diagnostic model. Sci R](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>.