



National University
of computer and emerging sciences

CS492 Final Year Project

IDENTIFICATION OF NEUROLOGICAL DISEASES USING DEEP LEARNING

Submitted By

Muhammad Tahir (21k-4503)

Insha Javed (21k-3279)

Hasan Iqbal (21k-3297)

Project Progress Report

**Department of Computer Science National University
of Computer and Emerging Sciences FAST Karachi Campus**

1. Introduction

Alzheimer's disease (AD) is the most common dementia and affects millions of people worldwide. With its neurodegeneration, memory loss, and cognitive decline, it shares pathological features of amyloid- β plaques and neurofibrillary tau tangles. Though environmental factors are involved in AD, there exists genetic susceptibility—with genome-wide association studies (GWAS) characterizing many single nucleotide polymorphisms (SNPs) associated with disease risk [1, 2]. The most powerful recognized genetic risk allele for AD is the Apolipoprotein E (APOE) ϵ 4 allele, yet additional genetic variants that determine susceptibility and disease progression remain to be explored [3, 4]. Standard GWAS methodologies use statistical association techniques to map disease-causing genetic variants [5].

However, these methods examine individual SNPs separately, which limits their ability to determine complex interactions between genetic variations [6]. This shortcoming has been addressed through the use of machine learning models like XGBoost and Random Forest, which offer feature selection and improved classification accuracy [7, 8]. Despite their advantages, these classic machine learning algorithms rely on excessive manual feature engineering and do not fully leverage the spatial relationships among SNPs in the genome [9].

2. TIMELINE

The project timeline outlines the key phases and milestones for developing the deep learning-based system to identify neurological diseases, with a particular focus on Alzheimer’s disease. The accompanying Gantt chart provides a visual representation of the schedule, and highlighting deliverables.

Months	Activity	Plan Start	Plan Duration	Actual Start	Actual Duration	Percent Complete
1	Study and Research	1	1	1	1	100%
2	Study and Research (Abstract & Methodologies)	2	1	2	1	100%
3	Data Collection & Preprocessing	3	1	3	1	100%
4	Genome Fragmentation & SNP Identification	4	1	4	1	100%
5	Phenotype Influence Score (PIS) & SWAT	5	1	5	1	100%
6	CNN Classification Model Development	6	1	0	0	50%
7	Comparison with Traditional Models	7	1	0	0	0%
8	Research Paper Writing	8	1	0	0	40%

Figure 1: Gantt Chart for Project Timeline

FYP-I (Months 1–4): Foundation Phase

In the initial phase, the focus was on establishing the research groundwork. Key activities included an extensive literature review on neurodegenerative disorders, identifying and preprocessing a comprehensive SNP dataset for Alzheimer's classification, and experimenting with various model architectures to determine the optimal genomic fragment size (found to be 40). Preliminary deep learning models i.e CNN, LSTM, and hybrid approaches, were developed and validated.

Deliverables Achieved:

- Literature Review Report
- Preprocessed SNP Dataset for Alzheimer's Classification
- Identified Alzheimer's Disease associated SNPs
- Performance Evaluation Report

FYP-II (Months 5–8): Implementation & Deployment Phase

Month 5 & 6: Model Enhancement & System Optimization

- Refine and fine-tune the deep learning models based on insights from FYP-I.
- Optimize hyperparameters to boost classification accuracy and reduce error rates.
- Update the system's architecture to incorporate advanced feature extraction methods.

Deliverable:

- AD vs CN Classification using Deep Learning Model
- Report on comparative analysis between the developed CNN model and traditional models.

Month 7 & 8: Final Reporting & Project Completion

- Complete Alzheimer's Disease vs Cognitive Normal classification model.
- Finalize the comprehensive research paper detailing findings, challenges, and innovations.
- Prepare a detailed presentation summarizing project outcomes, future recommendations, and lessons learned.

Deliverables:

- Complete AD vs CN Classification Model
- Final Research Paper
- Final Project Report and Presentation Slides

3. Project Progress

Milestone 1: Study and Research

- **Planned Start:** Month 1 | Planned Duration: 1 month
- **Actual Start:** Month 1 | Actual Duration: 1 month
- **Status:** 100% Complete

Overview: Comprehensive research was conducted, establishing the foundational knowledge required for the project.

Milestone 2: Study and Research (Abstract & Methodologies)

- **Planned Start:** Month 2 | Planned Duration: 1 month
- **Actual Start:** Month 2 | Actual Duration: 1 month
- **Status:** 100% Complete

Overview: The abstract and methodologies were developed, summarizing the core research strategies and theoretical framework.

Milestone 3: Data Collection & Preprocessing

- **Planned Start:** Month 3 | Planned Duration: 1 month
- **Actual Start:** Month 3 | Actual Duration: 1 month
- **Status:** 100% Complete

Overview: Data was collected and preprocessed, ensuring quality and consistency for subsequent analysis.

Milestone 4: Genome Fragmentation & SNP Identification

- **Planned Start:** Month 4 | Planned Duration: 1 month
- **Actual Start:** Month 4 | Actual Duration: 1 month
- **Status:** 100% Complete

Overview: The genome was fragmented and key SNPs were identified, setting the stage for detailed analysis.

Milestone 5: Phenotype Influence Score (PIS) & SWAT

- **Planned Start:** Month 5 | Planned Duration: 1 month
- **Actual Start:** Month 5 | Actual Duration: 1 month
- **Status:** 100% Complete

Overview: Calculation of the Phenotype Influence Score and application of the Sliding Window Association Test (SWAT) were successfully completed, highlighting critical genomic segments.

Milestone 6: CNN Classification Model Development

- **Planned Start:** Month 6 | Planned Duration: 1 month
- **Actual Start:** Month 6 / Partial Work Completed
- **Status:** 50% Complete

Overview: Development of the CNN-based classification model is in progress. Work has advanced to half completion, focusing on refining the model's architecture and initial training results.

Milestone 7: Comparison with Traditional Models

- **Planned Start:** Month 7 | Planned Duration: 1 month
- **Actual Start:** Not Yet Started
- **Status:** 0% Complete

Overview: A comparative analysis between the developed CNN model and traditional models is pending. This phase will assess performance metrics to benchmark the deep learning approach.

Milestone 8: Research Paper Writing

- **Planned Start:** Month 8 | Planned Duration: 1 month
- **Actual Start:** Not Yet Started / Partial Work Completed
- **Status:** 40% Complete

Overview: The research paper is underway, with 40% of the writing completed. This document will comprehensively detail the methodologies, experiments, results, and implications of the project.

References

1. Wang X, Zhang Y, Liu H, et al. Predicting early Alzheimer's with blood biomarkers and clinical features. *Alzheimers Res Ther* 2024;16:19–27.
2. Wang Z, Li J, Chen X, et al. Wide and deep learning-based approaches for classification of Alzheimer's disease using genome-wide association studies. *IEEE Trans Biomed Eng* 2023;70:2546–55.
3. Zhang Q, Zhang L, Zhao Y, et al. AD-Syn-Net: systematic identification of Alzheimer's disease-associated mutation and co-mutation vulnerabilities via deep learning. *Nat Commun* 2023;14:567–78.
4. Huang Y, Zhou X, Yang H, et al. Transfer learning for classification of Alzheimer's disease based on genome-wide data. *Bioinformatics* 2023;39:2993–3001.
5. Lee H, Wong M, Kim C, et al. An Alzheimer's disease gene prediction method based on the ensemble of genome-wide association study summary statistics. *Neurogenetics* 2022;23:145–56.
6. Zhang H, Wang R, Zhao Z, et al. A machine learning method to identify genetic variants potentially associated with Alzheimer's disease. *Front Genet* 2021;12:735248.
7. Tan Y, Xu X, Liu X, et al. Use of deep-learning genomics to discriminate healthy individuals from those with Alzheimer's disease or mild cognitive impairment. *J Alzheimers Dis* 2021;82:1403–15.
8. Liu X, Yang H, Zhang L, et al. Early detection of Alzheimer's disease based on single nucleotide polymorphisms (SNPs) analysis and machine learning techniques. *J Clin Neurosci* 2020;77:222–30.
9. Lee J, Cho K, Park S, et al. A deep learning-based approach to detect neurodegenerative diseases. *IEEE Access* 2020;8:142005–15.
10. Zhang W, Zhao L, Zhang Q, et al. Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and k-means. *Alzheimer's Dement* 2011; 7: 115–23.
11. Wainberg M, Merico D, Delong A, et al. Deep learning in biomedicine. *Nat Biotechnol* 2018;36:829–38.
12. Yamada M, Jitkrittum W, Sigal L, et al. High-dimensional feature selection by feature-wise Kernelized lasso. *Neural Comput* 2014;26:185–207.
13. Im HK, Min S, Song M, et al. Deep learning improves the prediction of CRISPR–Cpf1 guide RNA activity. *Nat Biotechnol* 2018;36: 239–41.
14. Zhang Z, Park CY, Theesfeld CL, et al. An automated framework for efficiently designing deep convolutional neural networks in genomics. *Nature Machine Intelligence* 2021;3:392–400.
15. Zhang S, Hu H, Jiang T, et al. TITER: predicting translation initiation sites by deep learning. *Bioinformatics* 2017;33:i234–42.
16. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74. 13
17. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80.
18. Lautrup S, Sinclair DA, Mattson MP, et al. NAD⁺ in brain aging and neurodegenerative disorders. *Cell Metab* 2019;30:630–55.