

# Alzheimer Disease Classification Using CNN



NATIONAL UNIVERSITY  
OF COMPUTER & EMERGING SCIENCES  
KARACHI CAMPUS

## Abstract

Alzheimer's disease is the most common cause of dementia. Some inherited traits called Single Nucleotide Polymorphisms (SNPs) play a role in the development of Alzheimer's disease (AD). Hence, it is crucial to identify these SNPs. However, conventional machine learning algorithms like XG Boost and Random Forest are ineffective for such complex and high-dimensional data. To address this problem, our research introduces a deep learning approach using Convolutional Neural Networks (CNNs). We have incorporated modern deep learning models such as Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and the Attention mechanism, each playing a vital role in enhancing the performance of our approach. This study helps improve our understanding of the genetic causes of Alzheimer's and supports future work in personalized medicine.

## Objectives

**Objective:** To develop a deep learning approach to identify Alzheimer's disease-related SNPs using whole-genome data.

**Model:** To utilize CNNs for Alzheimer's disease classification based on associated SNPs.

**Improvement:** Enhance SNP classification compared to traditional models like XGBoost and random forest.

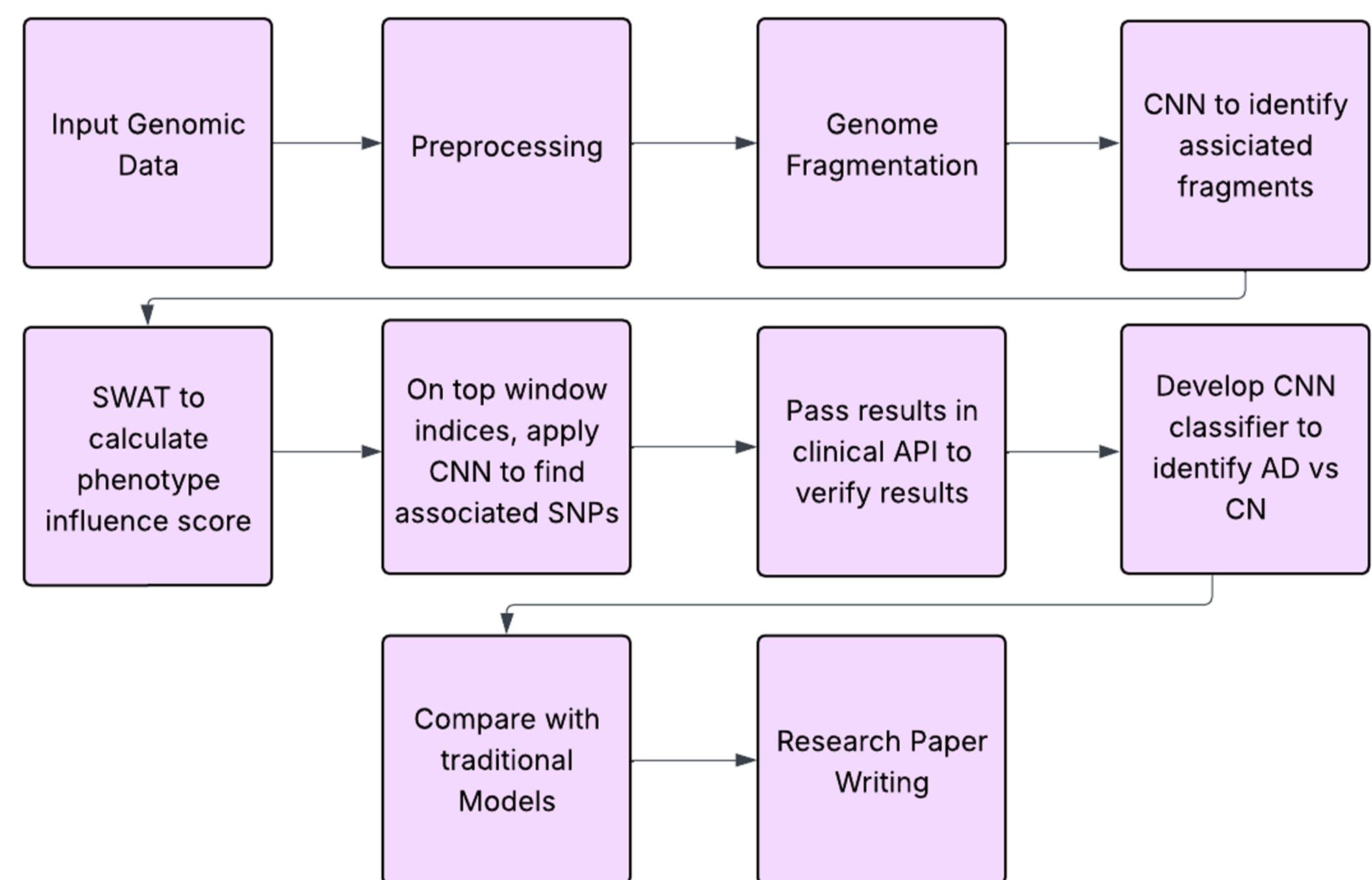
**Goal:** Discover novel genetic variants linked to Alzheimer's disease.

## Methods

Our method consists of three steps:

- Breaking down the genetic data into smaller, non-overlapping parts.
- Calculating the impact of each SNP on the disease (called the Phenotype Influence Score).
- Finally training a CNN-LSTM model to identify the target SNPs, which would further assist in diagnosis.

## Model Architecture



## Tools



Github



Python



Tensorflow



Visual Studio  
Code



Keras



Google  
Colaboratory

## Performance

The CNN model, which we used for Alzheimer's disease (AD) classification achieved an accuracy of 78% and an AUC of 0.8157, significantly more than that of the traditional machine learning models like XGBoost (71.15% accuracy, AUC = 0.7412) and Random Forest (68.19% accuracy, AUC = 0.7263). It was very effective in handling high-dimensional SNP data. Key SNPs identified include APOE, APOC1, TOMM40, SNX14, and SNX16, with APOE 04 being the strongest predictor, while novel SNPs in SNX genes suggest some roles in Alzheimer's disease. The model achieved its best results with a fragment size of 40 SNPs and, according to the Phenotype Influence Score (PIS), it was among the top 4,000 SNPs.

## Team Members:

Insha Javed 21k-3279

Muhammad Tahir 21K-4503

Hasan Iqbal 21k-3297

Supervisor:  
**Shoaib Rauf**