

# Interpretable CNN Classification of Alzheimer's Disease via MRI Heatmaps

## Dataset Description and Source

This document provides a detailed description of the dataset selected for the proposed study on Alzheimer's disease classification using Convolutional Neural Networks (CNNs). The dataset is a crucial foundation for developing, training, and evaluating deep learning models aimed at improving early detection and diagnosis of Alzheimer's disease.

### Dataset Description

The dataset chosen for this project is the *Alzheimer's Multi-Class Dataset (Equal and Augmented)*, publicly available on Kaggle. It consists of MRI brain scan images that have already been pre-processed to remove noise and enhance structural features of the brain. The dataset is organized into four diagnostic categories, which represent different stages of cognitive decline:

- **Non-Demented:** Individuals without any symptoms of dementia.
- **Very Mild Demented:** Subjects showing subtle cognitive impairments.
- **Mild Demented:** Patients with noticeable but not severe dementia symptoms.
- **Moderate Demented:** Individuals in a more advanced stage of the disease.

The dataset is notable for being **balanced** across all four classes, a common challenge in medical datasets. To achieve this, data augmentation techniques such as rotation, flipping, and contrast enhancement were applied. This ensures that each class has a comparable number of samples, reducing the risk of class imbalance and improving the reliability of model evaluation.

Total Number of Images: 44,000

Image Format: MRI scans as .JPG files

Image Usage: Suitable for training and testing machine learning models focused on classifying Alzheimer's disease stages.

## **Data Source**

The dataset is hosted on Kaggle and can be accessed at the following link:

[Alzheimer's Multi-Class Dataset \(Equal and Augmented\)](#)

- **Uploader:** Aryan Singhal
- **Platform:** Kaggle Datasets
- **License:** Open for academic and research purposes

The dataset is provided openly for research and academic purposes, making it suitable for reproducibility and benchmarking. Its availability on a widely recognized platform such as Kaggle also ensures long-term accessibility to the research community.

## **Suitability for Research**

This dataset is particularly suitable for Alzheimer's disease classification using CNNs due to the following reasons:

- **Comprehensive Class Coverage:** It covers multiple stages of dementia, allowing researchers to move beyond simple binary classification and tackle more realistic multi-class problems.
- **Pre-processed Images:** MRI scans are already standardized and pre-processed, which reduces the preprocessing workload and enables direct application to deep learning architectures.
- **Balanced Dataset:** Augmentation ensures equal representation of all four classes, addressing the class imbalance issue often found in medical datasets.
- **Applicability to CNNs:** The dataset consists of 2D brain MRI slices that are well-suited for CNN-based models, which excel at extracting spatial and structural features.
- **Accessibility and Reproducibility:** As a publicly available dataset on Kaggle, it enables reproducibility of research results and facilitates benchmarking against other studies.
- **Educational Value:** The dataset has been designed to be accessible for both beginners and advanced researchers, making it an ideal choice for academic projects and thesis work.