# SSL for Alzheimers

## Motivation

One of the primary challenges in medical imaging is the scarcity of high-quality, expertly annotated data, as labelling MRI scans requires significant clinical expertise, time, and cost. This limitation often restricts the performance and generalizability of fully supervised deep learning models. This Hack4Health project is motivated by the need to overcome this bottleneck by leveraging Self-Supervised Learning (SSL). Specifically, we employ the SimCLR framework to pre-train a robust feature encoder on a large corpus of unlabelled brain MRI data, enabling the model to learn rich and invariant visual representations without relying on manual annotations. These learned representations are then fine-tuned on smaller, task-specific labelled datasets for neurological disease classification, including Alzheimer's disease, brain tumors, multiple sclerosis, and Parkinson's disease. By decoupling representation learning from downstream supervision, this approach aims to improve diagnostic accuracy, enhance data efficiency, and promote better model generalization across diverse neurological disorders and imaging conditions.

## Data Sources and Exploratory Data Analysis (EDA)

The dataset consists of 100,947 brain MRI images categorized into four Alzheimer's disease stages:

- Non-Demented: 67,222 (66.6%)
- Very Mild Dementia: 13,725 (13.6%)
- Mild Dementia: 10,000 (9.9%)
- Moderate Dementia: 10,000 (9.9%)

All images were resized to 496 × 248 pixels for consistent model input. Exploratory analysis of pixel intensity distributions revealed distinguishable patterns across disease stages, supporting the feasibility of MRI-based classification.

The dataset exhibits notable class imbalance, with Non-Demented samples dominating. To mitigate this, data augmentation including random rotations, horizontal flips, and color jitter were applied to the Mild and Moderate Dementia classes, improving class balance and training robustness.

## Data Preparation and Splitting

To simulate real-world medical data scarcity, the dataset was divided into large unlabelled and small labelled subsets. For the downstream fine-tuning stage, only 450 labelled images per class were used for supervised training, effectively modelling a low-data clinical setting and evaluating the data efficiency of the self-supervised approach.

# Methodology and Model Development

## Model Architecture: Custom ResNet50

The project employs a custom ResNet50 as the backbone encoder for feature extraction from brain MRI images. The architecture is composed of residual blocks with bottleneck layers, batch normalization, and ReLU activations, enabling efficient learning of deep hierarchical representations. The network begins with an initial 7×7 convolutional layer, followed by four residual stages, and concludes with a global (adaptive) average pooling layer. Depending on the training phase, the pooled features are forwarded either to a projection head for self-supervised learning or to a linear classifier for downstream disease classification.

### Self-Supervised Learning with SimCLR

The core innovation of this work is the adoption of the SimCLR framework for self-supervised representation learning. SimCLR enables the model to learn discriminative features from unlabelled MRI data through contrastive learning.

Augmentation Pipeline: Each MRI image is transformed into two distinct augmented views using random resized cropping, horizontal flips, color jittering, Gaussian blur, and grayscale conversion. These augmentations encourage invariance to appearance-level transformations while preserving anatomical structure.

Encoder and Projection Head: A ResNet50 encoder extracts feature embeddings, which are then passed through a multi-layer MLP projection head (2048 → 1024 → 512 → 128). The projection head maps representations into a contrastive space where similarity learning is performed.

Contrastive Objective: Training is guided by the NT-Xent (Normalized Temperature-scaled Cross Entropy) loss with a temperature parameter of 0.07, which maximizes agreement between different augmented views of the same image while separating representations of different images.

### Training Protocol

The SimCLR model was pre-trained for **50 epochs** on the unlabelled MRI dataset, with a total training time of approximately **7,470 seconds**. Following pre-training, the projection head was discarded, and a **linear classifier** was attached to the encoder for supervised fine-tuning on the labeled Alzheimer's dataset. This two-stage training strategy enables effective transfer of learned representations to low-data medical classification tasks.

# Evaluation, Explainability, and Impact

### Model Performance (Alzheimer's Disease)

The fine-tuned **SimCLR-based self-supervised model** achieved an overall **test accuracy of 87.41%** on the Alzheimer's disease classification task, demonstrating strong performance in a low-label training regime.

Class-wise evaluation highlights robust detection across disease stages:

- **Moderate Dementia:** Achieved the highest performance with **100% recall** and an **F1-score of 0.97**, indicating excellent sensitivity for advanced disease.
- **Mild Dementia:** Recorded an **F1-score of 0.88** with strong recall, reflecting reliable early-stage detection.
- **Non-Demented:** Achieved an **F1-score of 0.86**, showing effective separation between healthy and pathological cases.

- **Very Mild Dementia:** Obtained an **F1-score of 0.80**, representing the most challenging class due to subtle structural changes.

These results demonstrate the effectiveness of self-supervised pretraining in learning discriminative MRI representations, particularly under limited labelled data conditions.

Cross-Disease Generalisation: To evaluate robustness and transferability, the pre-trained SSL encoder was fine-tuned on additional neurological datasets. The model demonstrated strong generalization across diseases:

- **Brain Tumours (Glioma, Meningioma, Pituitary): 92.59% accuracy**
- **Multiple Sclerosis (MS vs. Normal): 87.30% accuracy**
- **Parkinson's Disease (Parkinson vs. Normal): 99.26% accuracy**

These results confirm that the SimCLR-pretrained encoder learns **disease-agnostic structural representations**, enabling effective transfer to diverse neurological classification tasks with minimal labeled data.

## Explainability and Clinical Interpretability (Grad-CAM)

To ensure clinical trust and transparency, **Grad-CAM** was applied to visualize class-discriminative regions influencing model predictions. For a **Moderate Dementia** case, the model achieved a **confidence score of 1.0000** and highlighted an activation region covering **28.10% of the brain scan**, corresponding to areas associated with structural atrophy. Similar activation patterns were observed for **brain tumor (meningioma)** and **multiple sclerosis** cases, indicating that the model consistently focuses on clinically relevant pathological regions rather than spurious features.

## Conclusion and Impact

This project demonstrates that **Self-Supervised Learning using SimCLR** can effectively bridge the gap posed by scarce labeled medical imaging data. By learning universal representations from unlabelled MRIs, the model achieves high accuracy, strong cross-disease generalization, and meaningful interpretability. The proposed framework offers a scalable and clinically relevant solution for **multi-disease neurological screening**, with potential applicability across a wide range of medical imaging tasks.