Select and Explain Three SSL Methods (Task-3)

Group-9

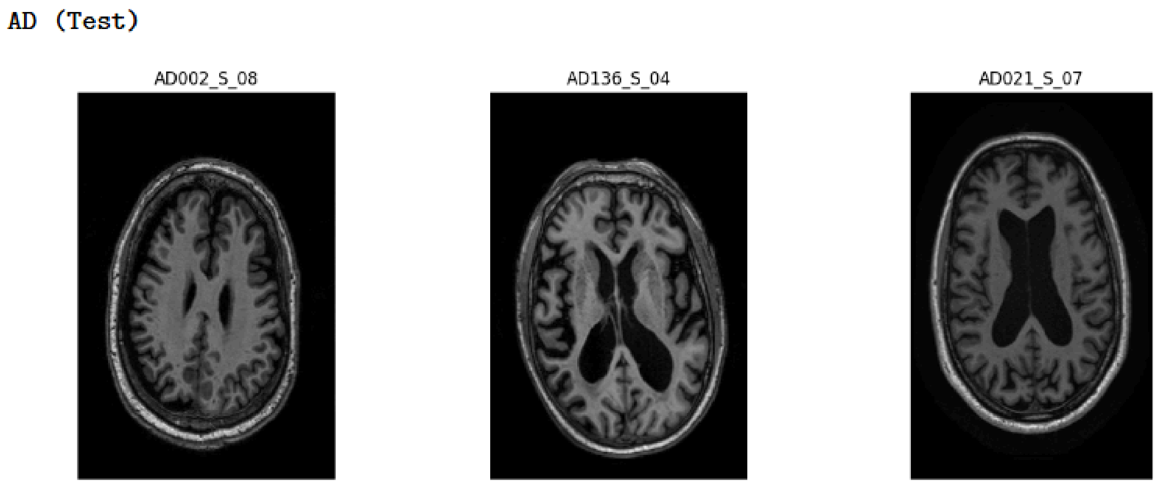
Group Members

Name	ID
Farhan Ibtesham Joy	2022-3-60-150
Miftha Ahona Majid	2022-1-50-015
Mithila Akter Aka	2022-1-50-010
Eshrat Jahan Eshi	2020-1-61-383

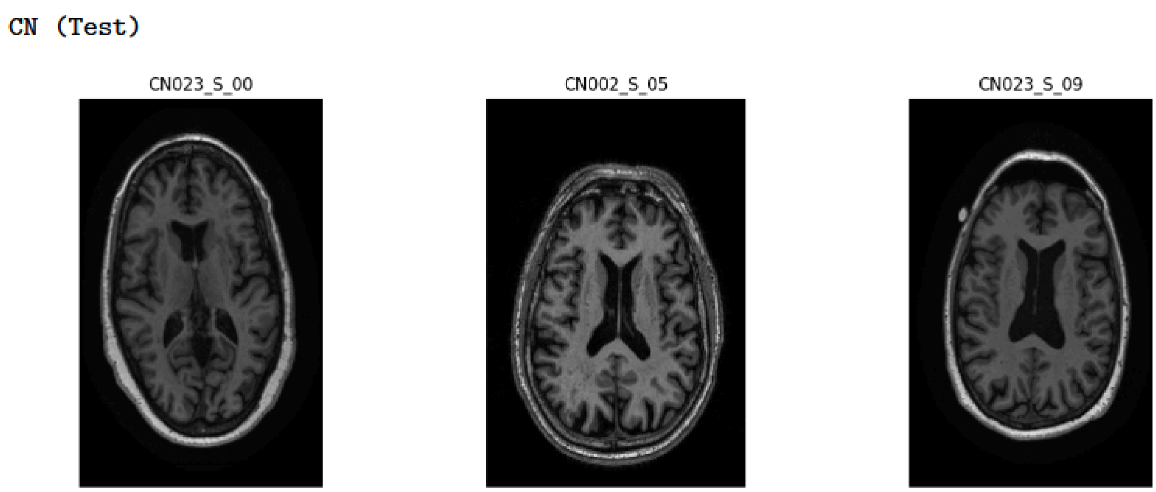
Submitted To:

Dr. Raihan Ul Islam
Associate Professor
Dept. Of Computer Science & Engineering
East West University

Test Dataset Samples (AD, CN, MCI)

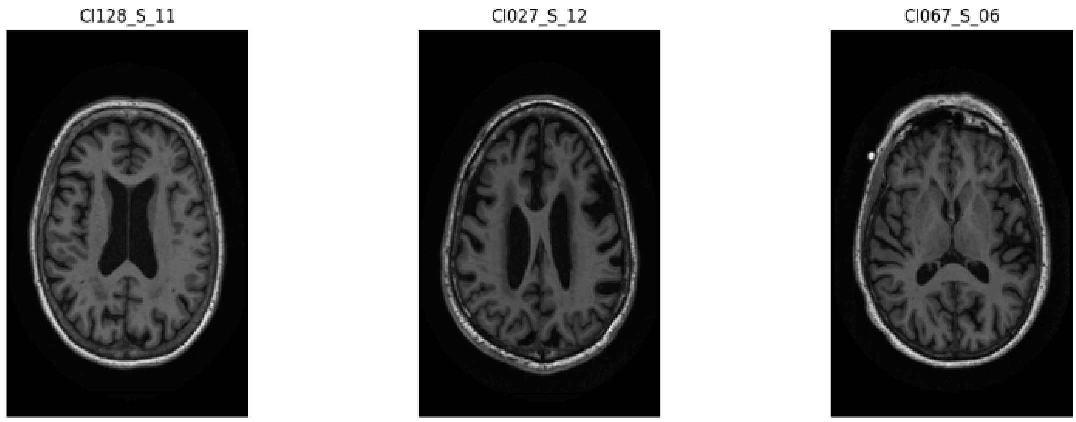


This image represents a brain scan from the test dataset labeled as **Alzheimer's Disease (AD)**.
It typically shows structural abnormalities such as cortical thinning and enlarged ventricles.
Used to evaluate model performance in identifying AD cases.



A test image labeled as **Cognitively Normal (CN)**.Displays a healthy brain structure with no visible signs of degeneration.Serves as a baseline for comparison with AD and MCI scans.

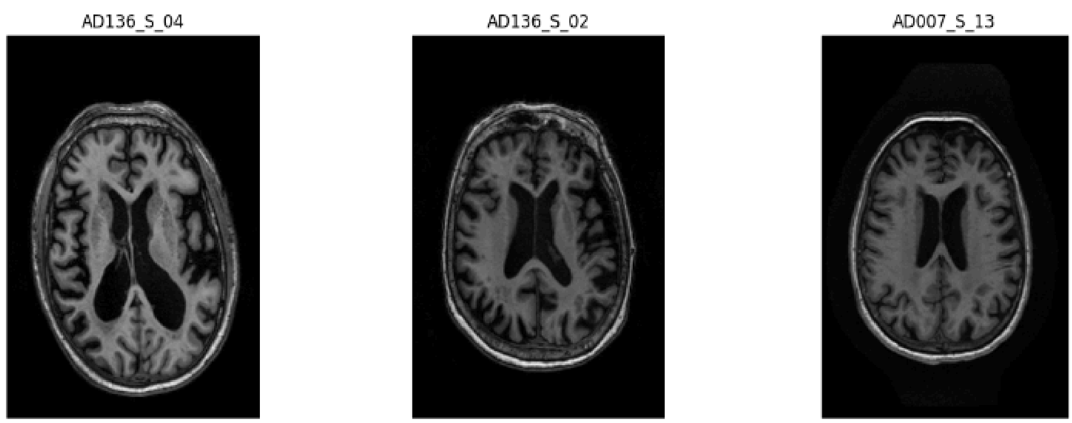
MCI (Test)



This test image is labeled as **Mild Cognitive Impairment (MCI)**.Shows subtle changes in brain structure, intermediate between CN and AD.Helps assess the model's sensitivity to early-stage cognitive decline.

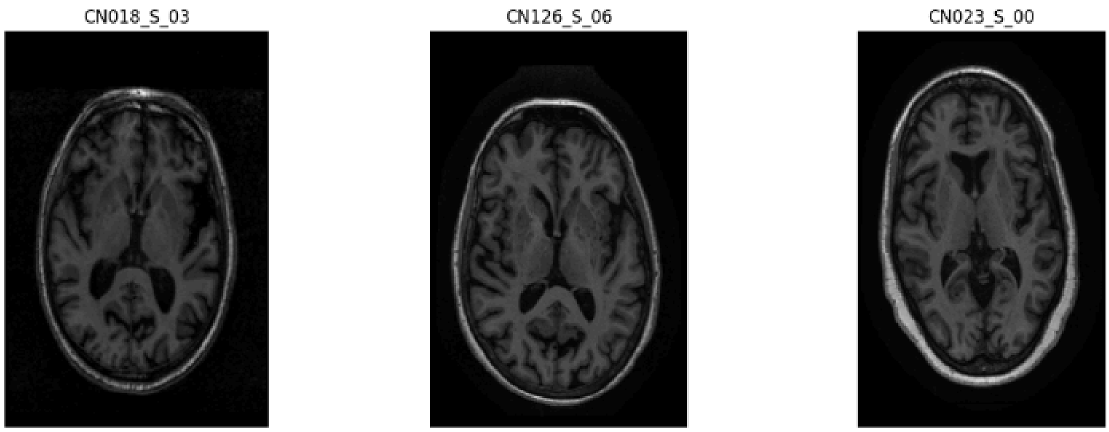
Train Dataset Samples (AD, CN, MCI)

AD (Train)

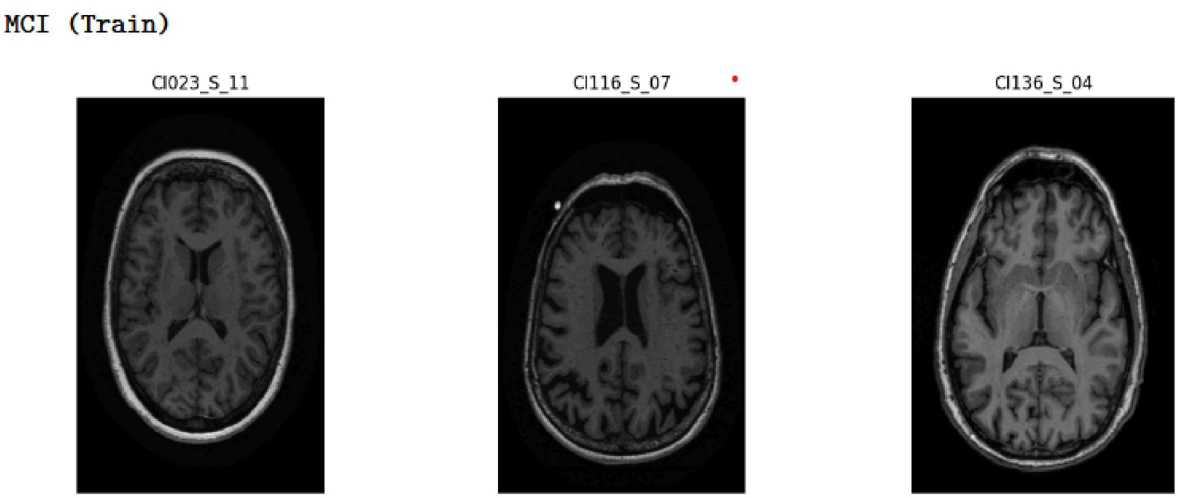


A training image labeled as **Alzheimer's Disease (AD)**.Used to teach the model to recognize patterns associated with AD.Features similar to the AD test image, aiding in model generalization.

CN (Train)

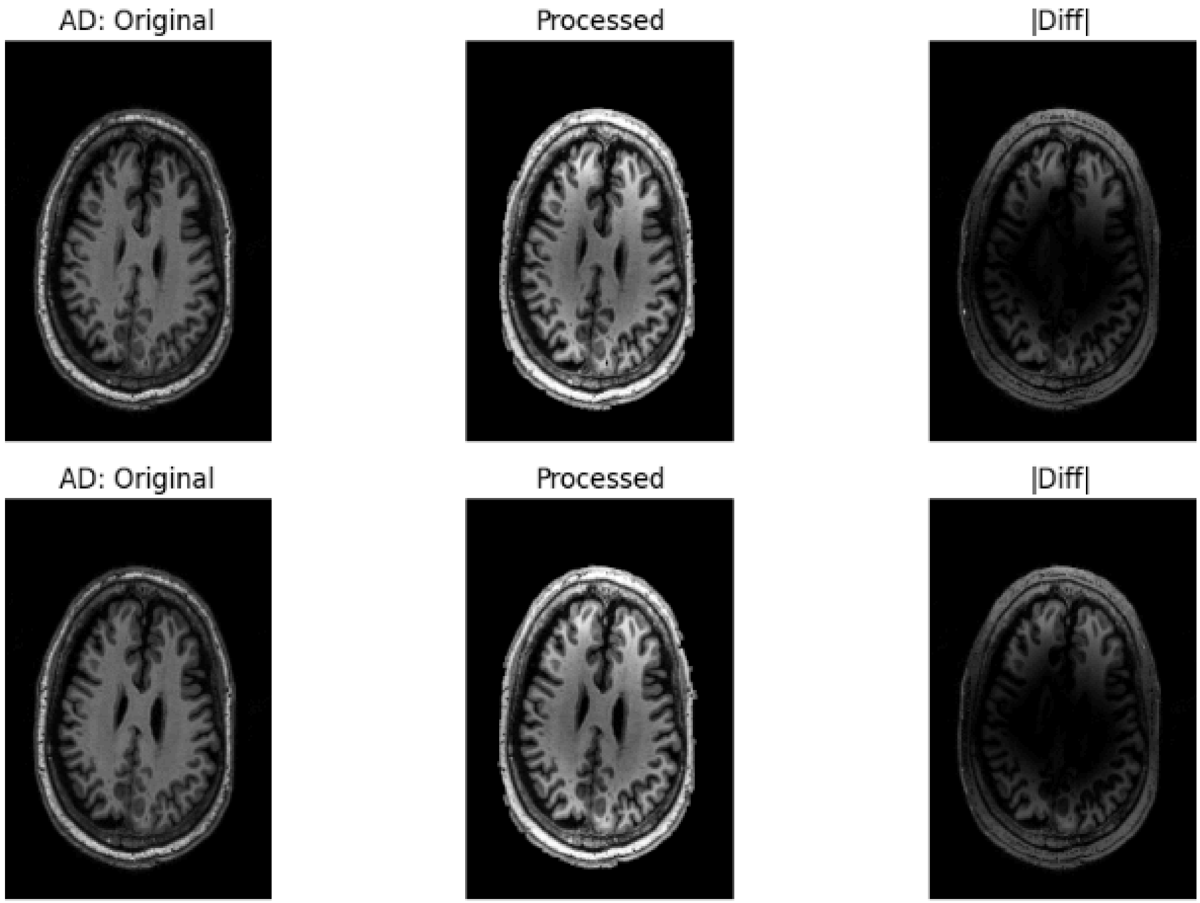


A training image labeled as **Cognitively Normal (CN)**.Represents normal brain anatomy, helping the model learn what healthy scans look like.Crucial for distinguishing CN from MCI and AD.



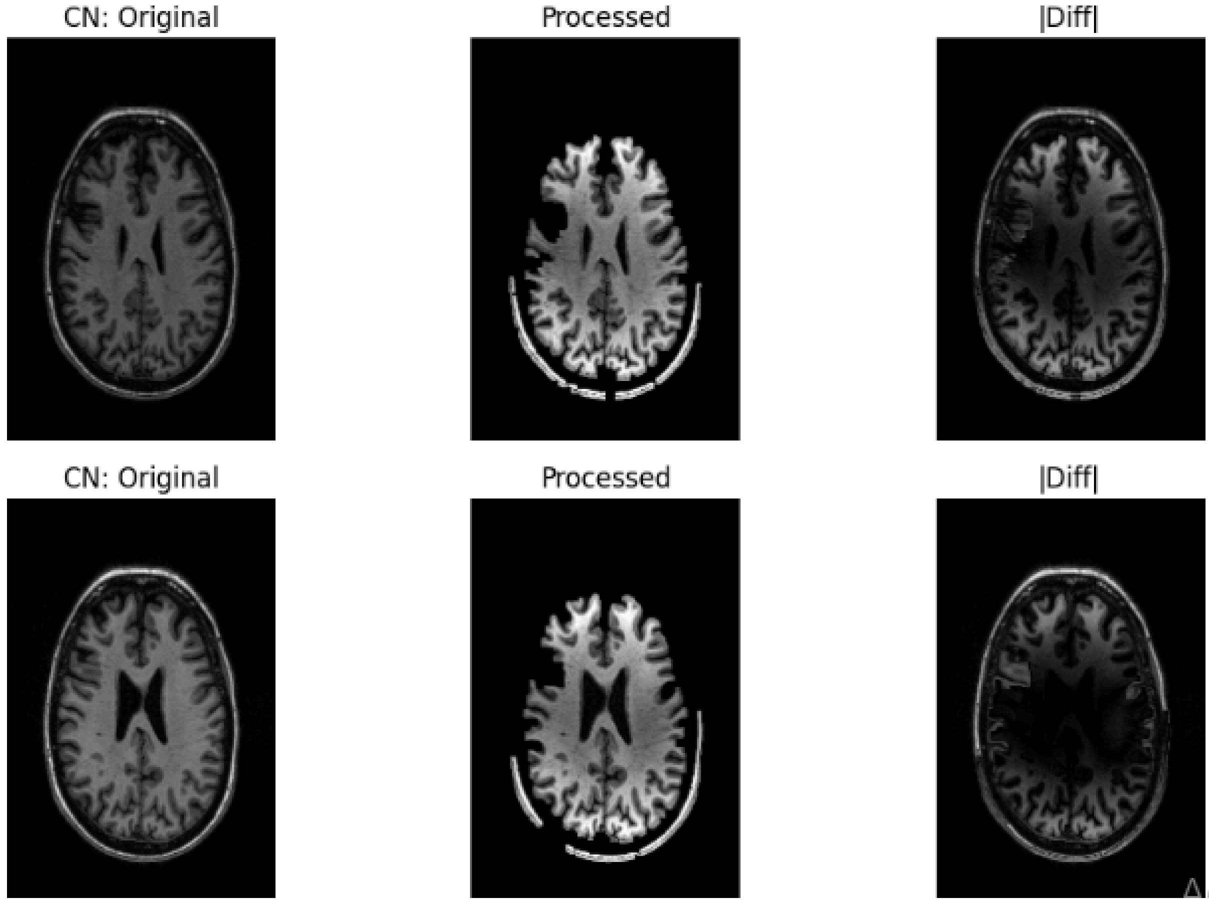
A training image labeled as **Mild Cognitive Impairment (MCI)**.Provides examples of early cognitive decline for model learning.Important for improving classification accuracy between CN and AD.

AD Original- Processed- Difference



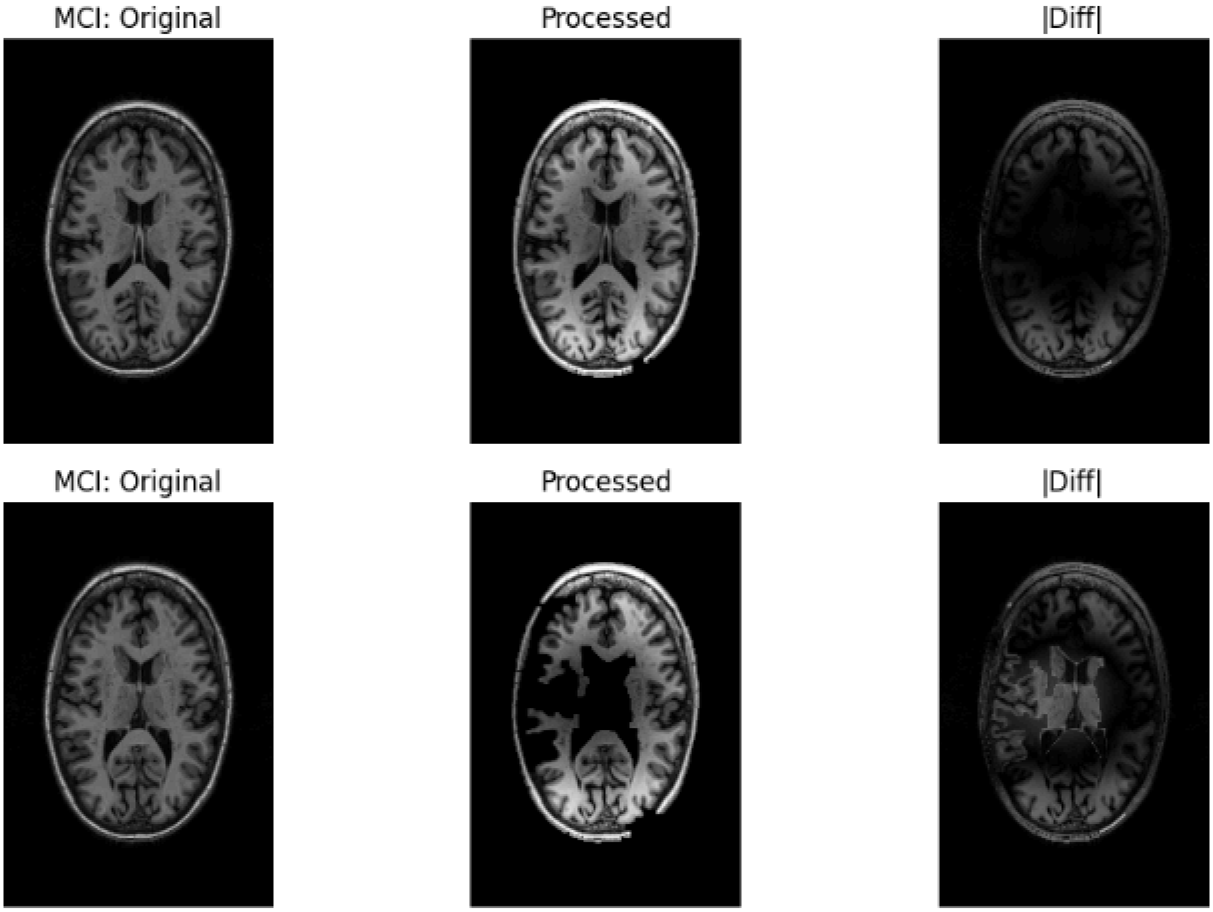
Original: Displays the raw brain scan of a patient diagnosed with Alzheimer's Disease.
Processed: Shows the image after preprocessing steps like normalization, resizing, or filtering.
Difference: Highlights the changes between the original and processed image, often emphasizing regions affected by AD such as hippocampal shrinkage or cortical thinning.

CN Original- Processed- Difference



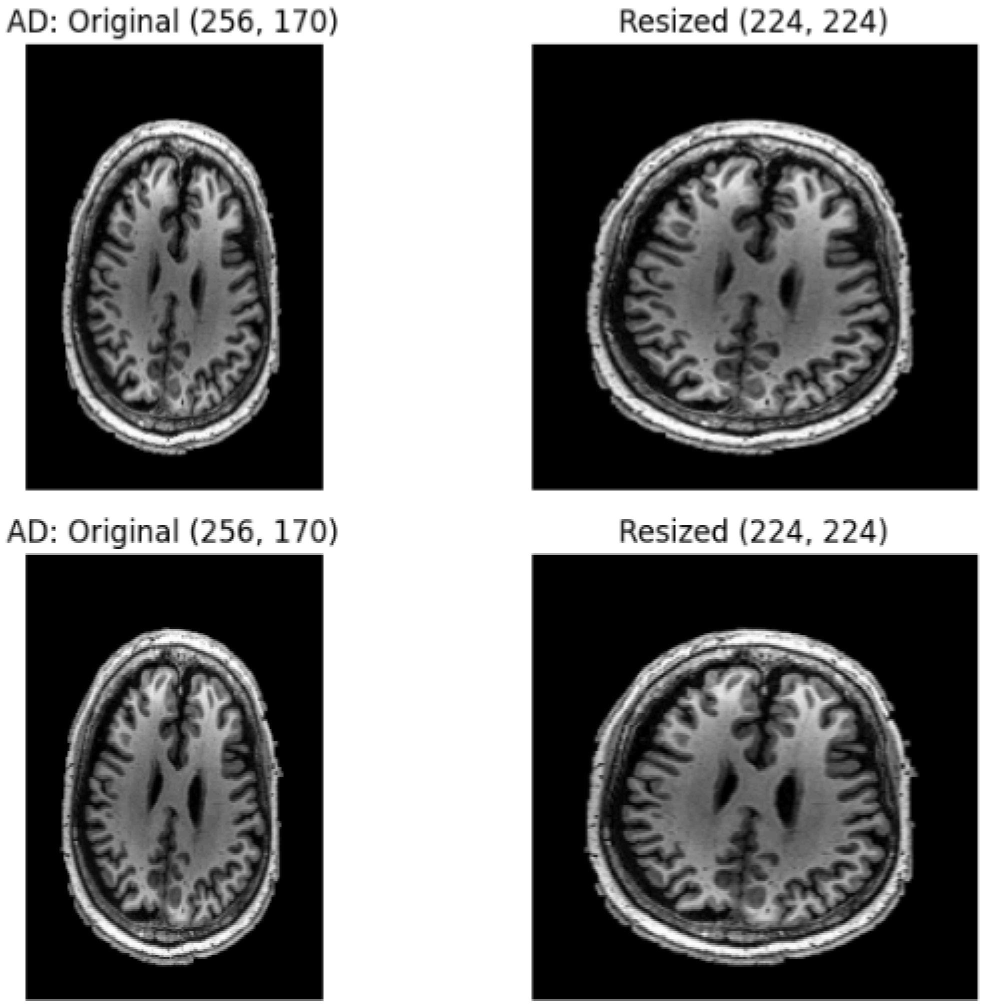
Original: A clean brain scan of a cognitively normal individual.
Processed: The image after enhancement and standardization for model input.
Difference: Minimal changes are visible, indicating stable brain structure with no signs of degeneration.

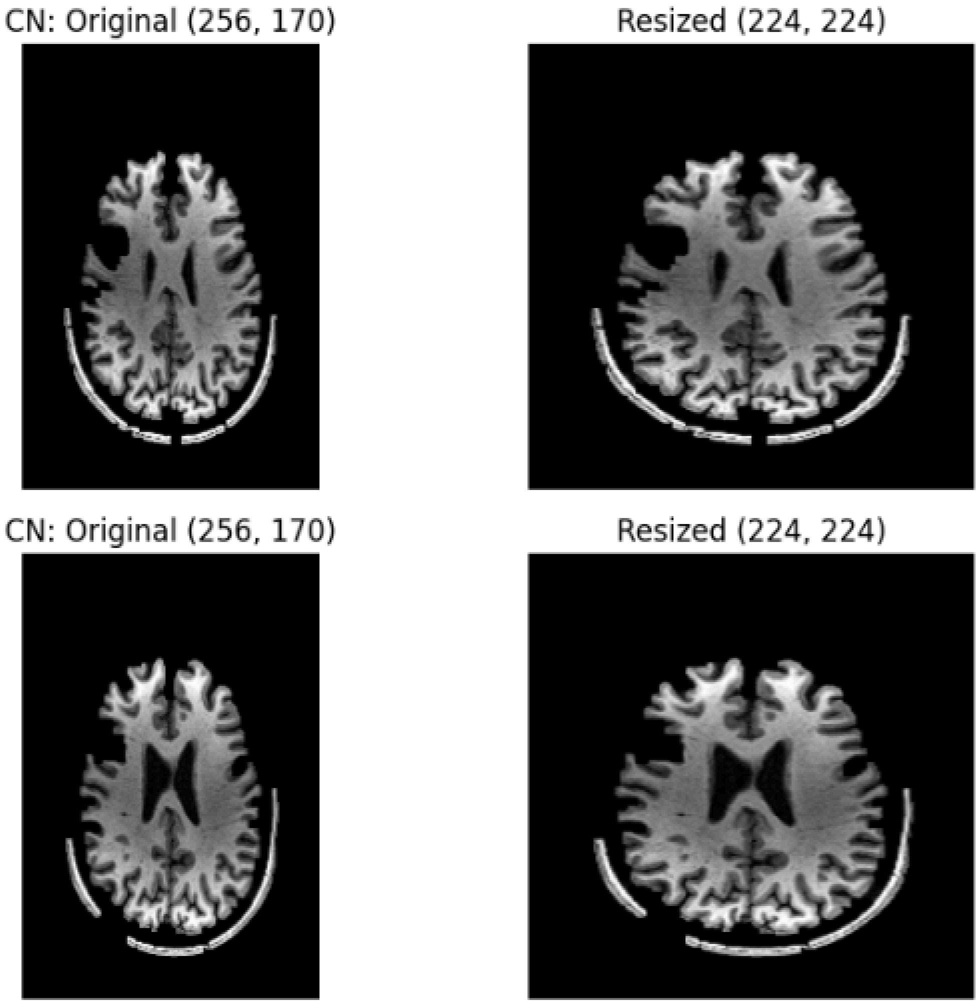
MCI Original- Processed- Difference

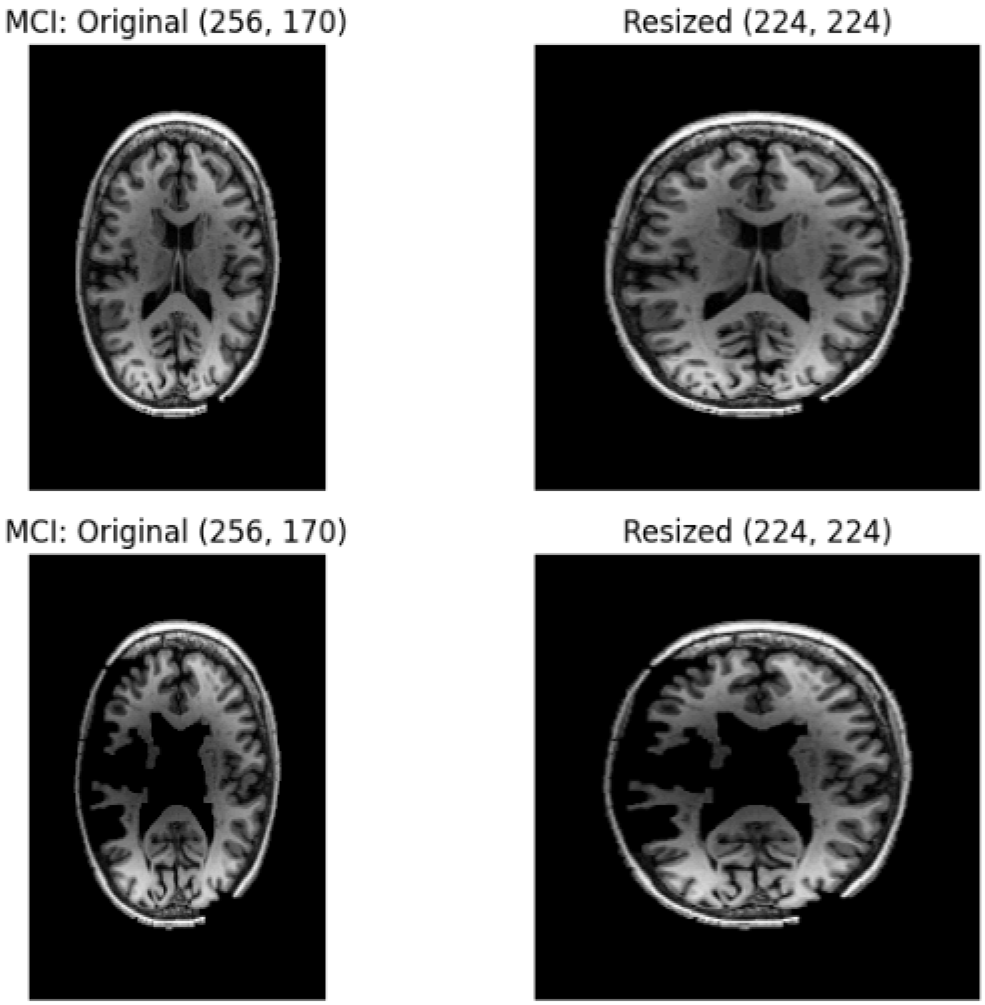


Original: Brain scan of a patient with Mild Cognitive Impairment.
Processed: Preprocessed version optimized for feature extraction.
Difference: Shows subtle structural changes, such as mild atrophy, which are crucial for early detection and classification.

Original (256x170) vs. Resized (224x224) images for AD, CN, and MC



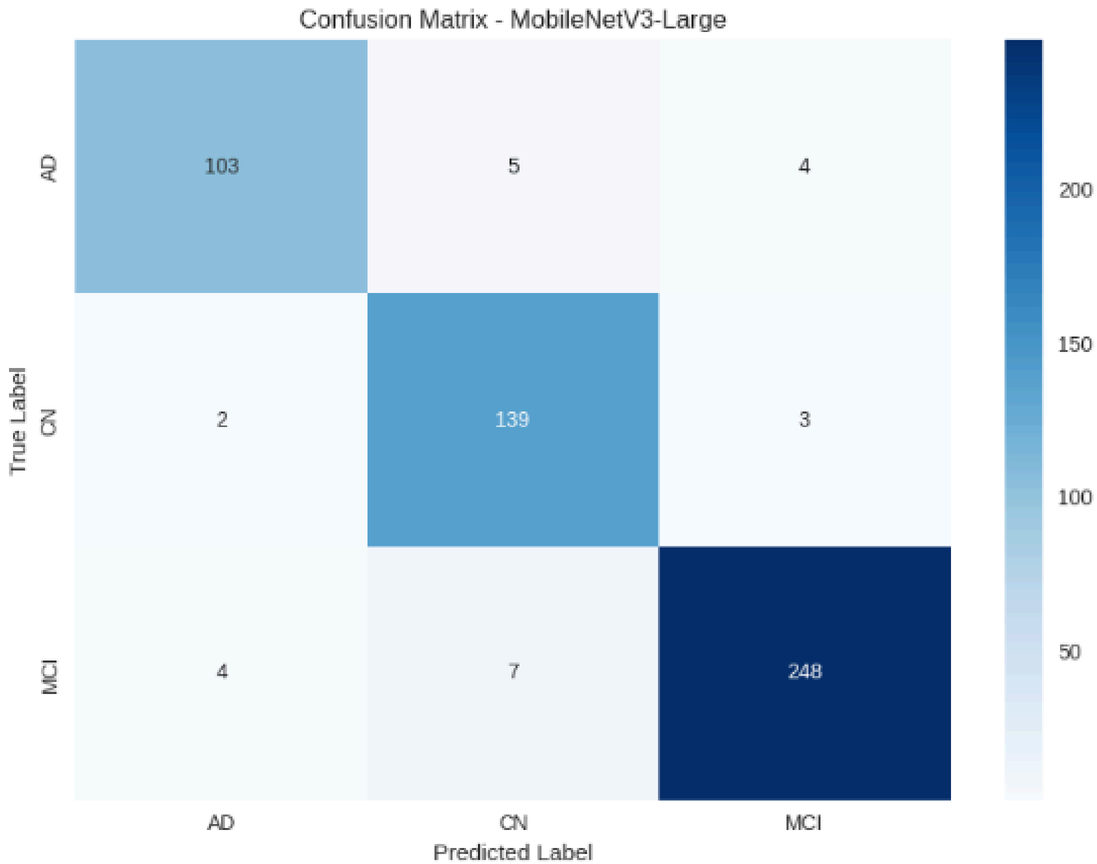

CN: Original (256, 170)
Resized (224, 224)
CN: Original (256, 170)
Resized (224, 224)



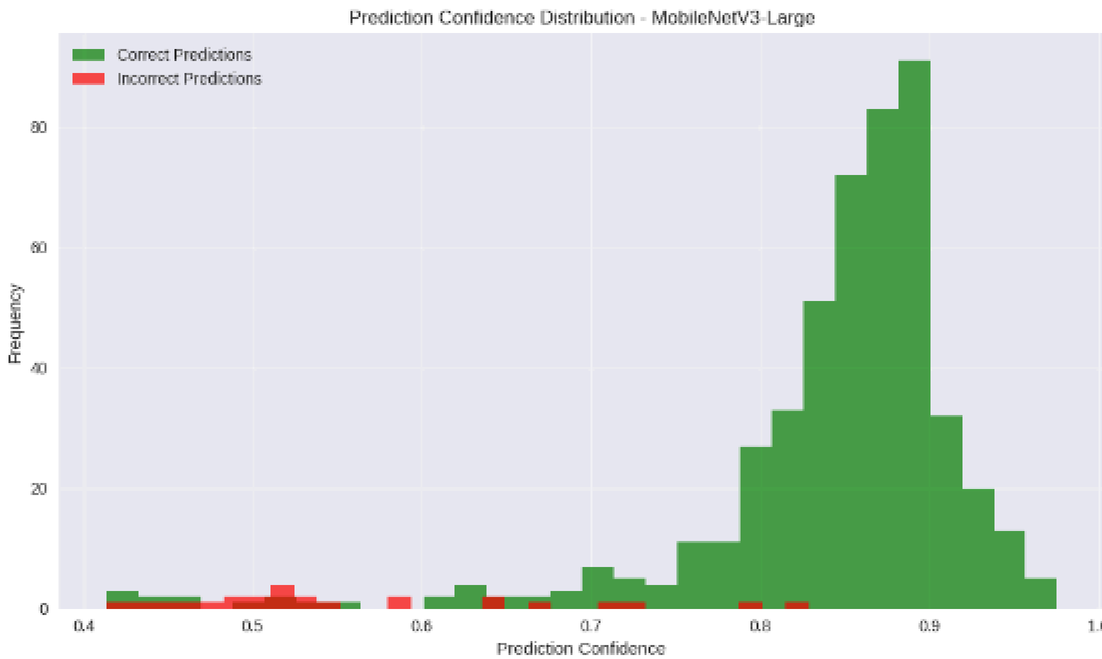
These images demonstrate standardization of input dimensions for deep learning models.

MobileNetV3-large Model

```
=== MobileNetV3-Large Detailed Results ===
Accuracy: 0.9515
Precision: 0.9520
Recall: 0.9515
F1-Score: 0.9515

Classification Report:
              precision    recall  f1-score   support

          AD       0.94      0.92      0.93       112
          CN       0.92      0.97      0.94       144
         MCI       0.97      0.96      0.96       259

    accuracy                           0.95       515
   macro avg       0.95      0.95      0.95       515
weighted avg       0.95      0.95      0.95       515

Average Prediction Confidence: 0.8301
```



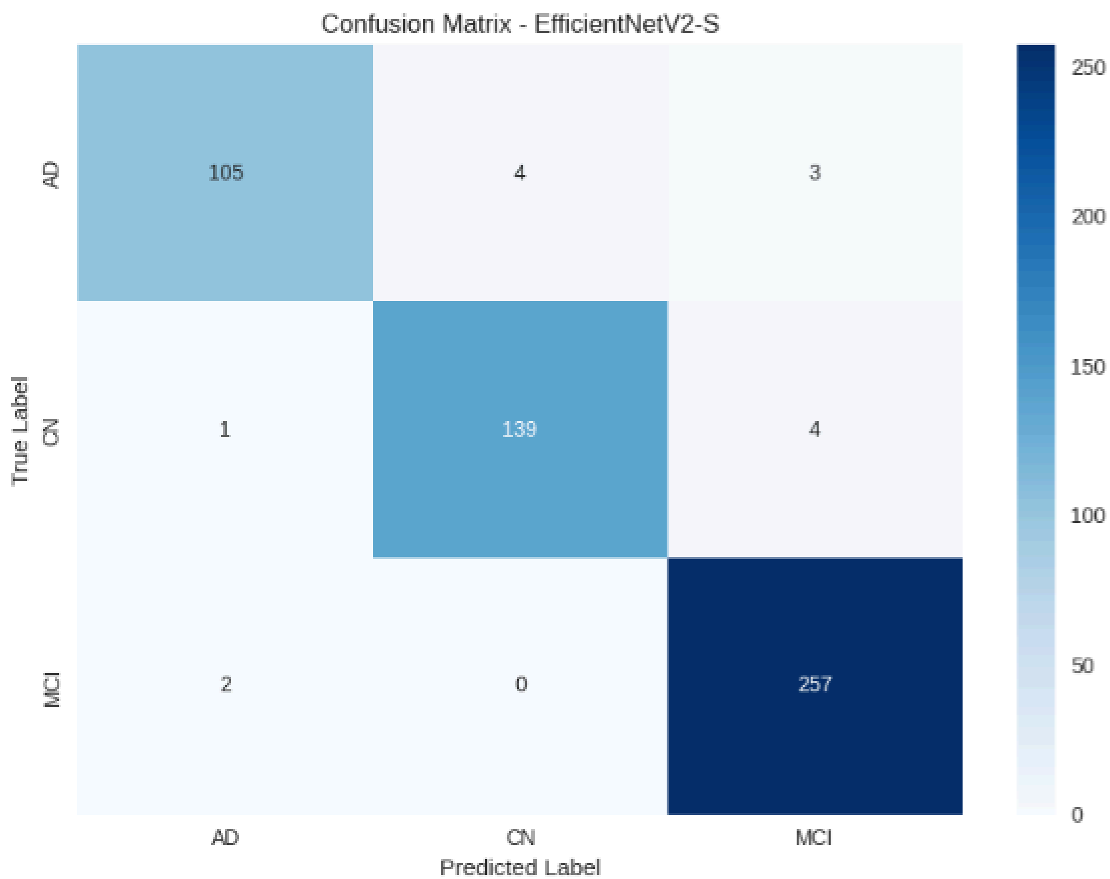
This graph A 3x3 matrix showing true vs. predicted labels.Most predictions are correct, especially for CN and MCI. AD had 103 correct predictions, with some misclassified as CN and MCI. Indicates good but not perfect classification.



This Bar graph comparing confidence of correct vs. incorrect predictions.Correct predictions cluster around high confidence (0.8–1.0).Incorrect ones are more spread out, often with lower confidence.Shows the model is more confident when it’s correct.

EfficientNet V2-S

```
=== EfficientNetV2-S Detailed Results ===
Accuracy: 0.9728
Precision: 0.9728
Recall: 0.9728
F1-Score: 0.9727

Classification Report:
              precision    recall  f1-score   support

          AD       0.97      0.94      0.95       112
          CN       0.97      0.97      0.97       144
         MCI       0.97      0.99      0.98       259

    accuracy                           0.97       515
   macro avg       0.97      0.97      0.97       515
weighted avg       0.97      0.97      0.97       515

Average Prediction Confidence: 0.8627
```



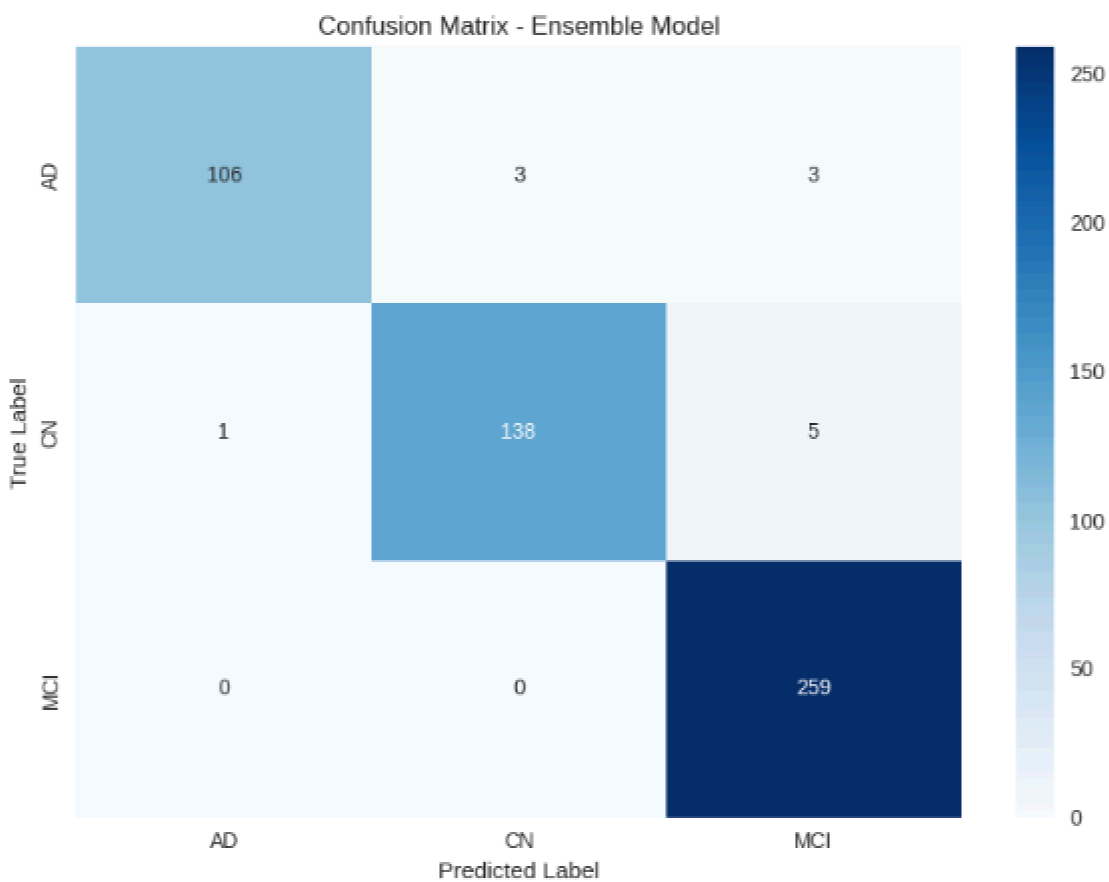
This graph shows fewer misclassifications than MobileNet.MCI class has 257 correct predictions out of 259. AD and CN also show improved classification accuracy. Demonstrates stronger performance overall.

This graph shows Higher frequency of confident predictions (0.8 -- 1.0 range).Correct predictions dominate the higher confidence bins.Fewer low-confidence predictions compared to MobileNet.Reflects improved model certainty.

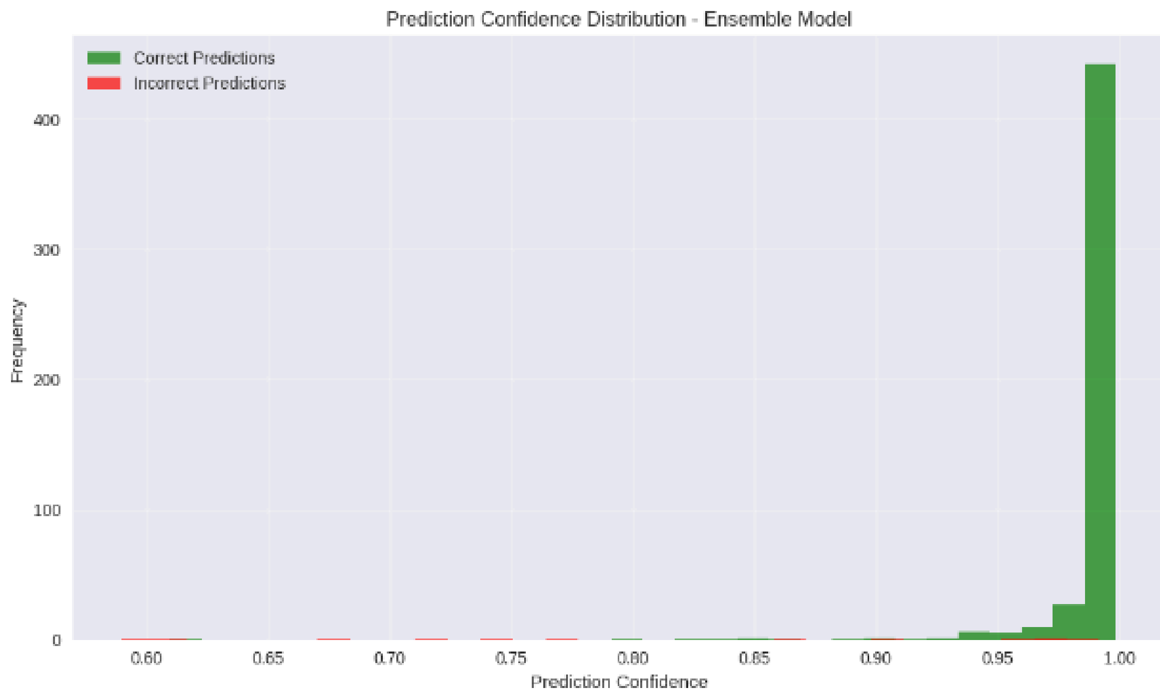
Ensemble Model

```
=== Ensemble Model Detailed Results ===
Accuracy: 0.9767
```

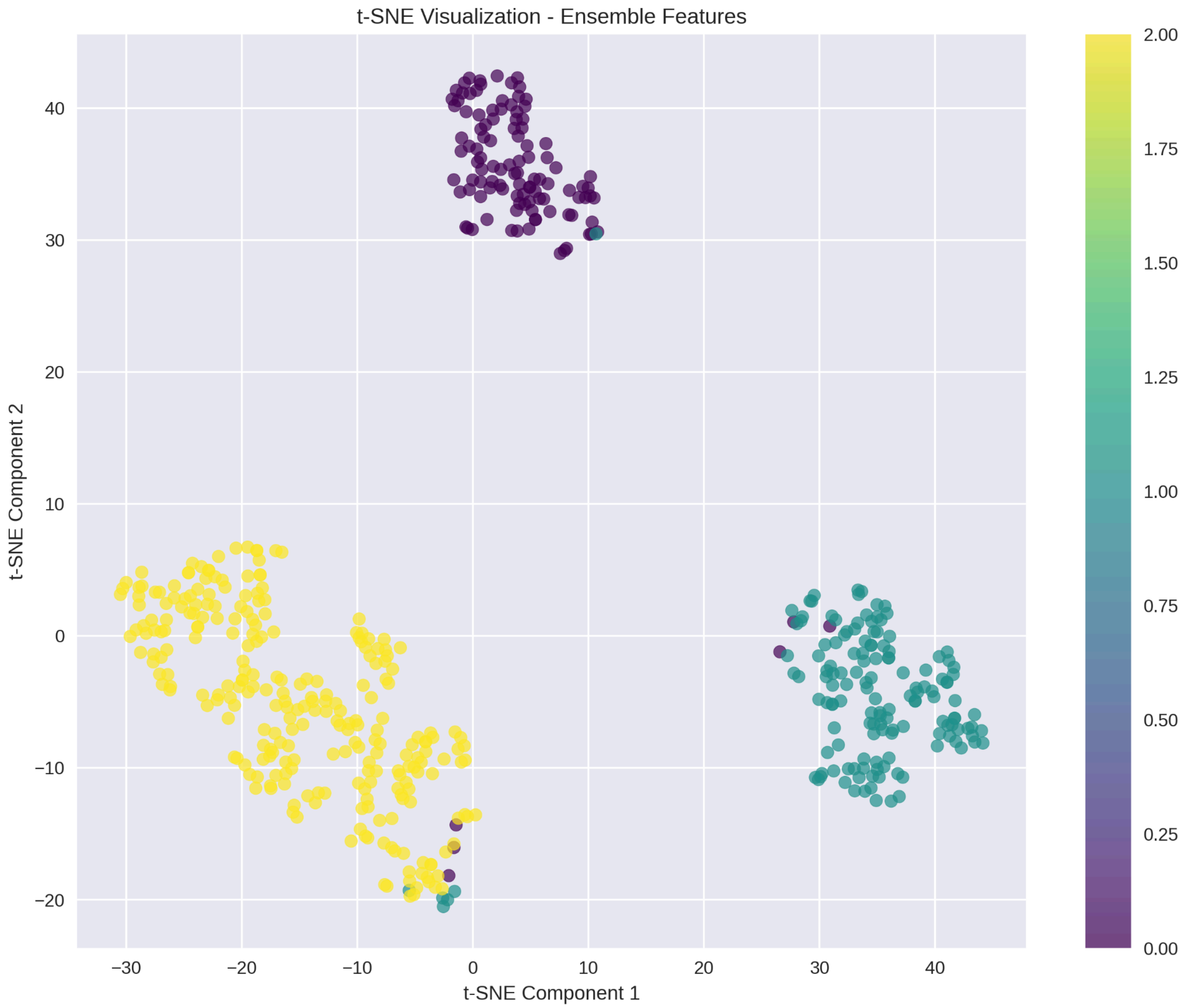
```
Precision: 0.9769
Recall: 0.9767
F1-Score: 0.9766

Classification Report:
              precision    recall  f1-score   support

          AD       0.99      0.95      0.97       112
          CN       0.98      0.96      0.97       144
         MCI       0.97      1.00      0.98       259

    accuracy                           0.98       515
   macro avg       0.98      0.97      0.97       515
weighted avg       0.98      0.98      0.98       515

Average Prediction Confidence: 0.9842
```



This graph shows near-perfect classification across all classes. Only 6 total misclassifications out of 515 samples. MCI class has 100% accuracy. Demonstrates the power of ensemble learning.



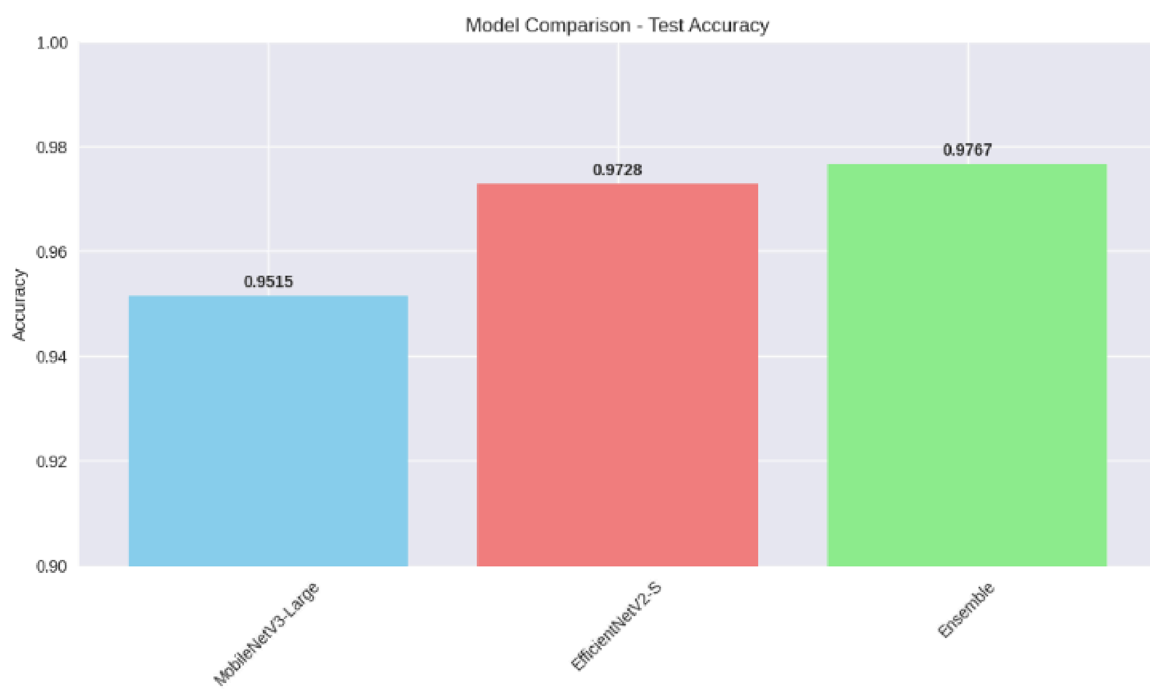
This graph given most predictions has confidence above 0.9. Very few incorrect predictions, mostly at lower confidence. Shows the ensemble model is both accurate and confident. Best distribution among all models.



This displays a 2D scatter plot of high-dimensional features. Clear separation between AD, CN, and MCI clusters. Indicates strong feature learning and class distinction. Supports the ensemble model's high performance.

Final Ensemble Results Summary:-

```
MobileNetV3-Large:
  Accuracy: 0.9515
  F1-Score: 0.9515
  Avg Confidence: 0.8301

EfficientNetV2-S:
  Accuracy: 0.9728
  F1-Score: 0.9727
  Avg Confidence: 0.8627

Ensemble:
  Accuracy: 0.9767
F1-Score: 0.9766
Avg Confidence: 0.9842
```



This Bar chart compares MobileNet, EfficientNet, and Ensemble. Ensemble outperforms both in accuracy and F1-score. Highlights the benefit of combining models. Visual summary of performance metrics.

SSL Methods

We've chosen three SSL methods: BYOL, Barlow Twins, and SimSiam. These are great for image data because they don't need huge batches of "negative" examples (unlike some other methods), making them efficient for smaller datasets like our Alzheimer's MRI collection (which has limited scans). They focus on making the AI learn consistent features from twisted versions of the same image, which helps in spotting subtle brain differences without labels.

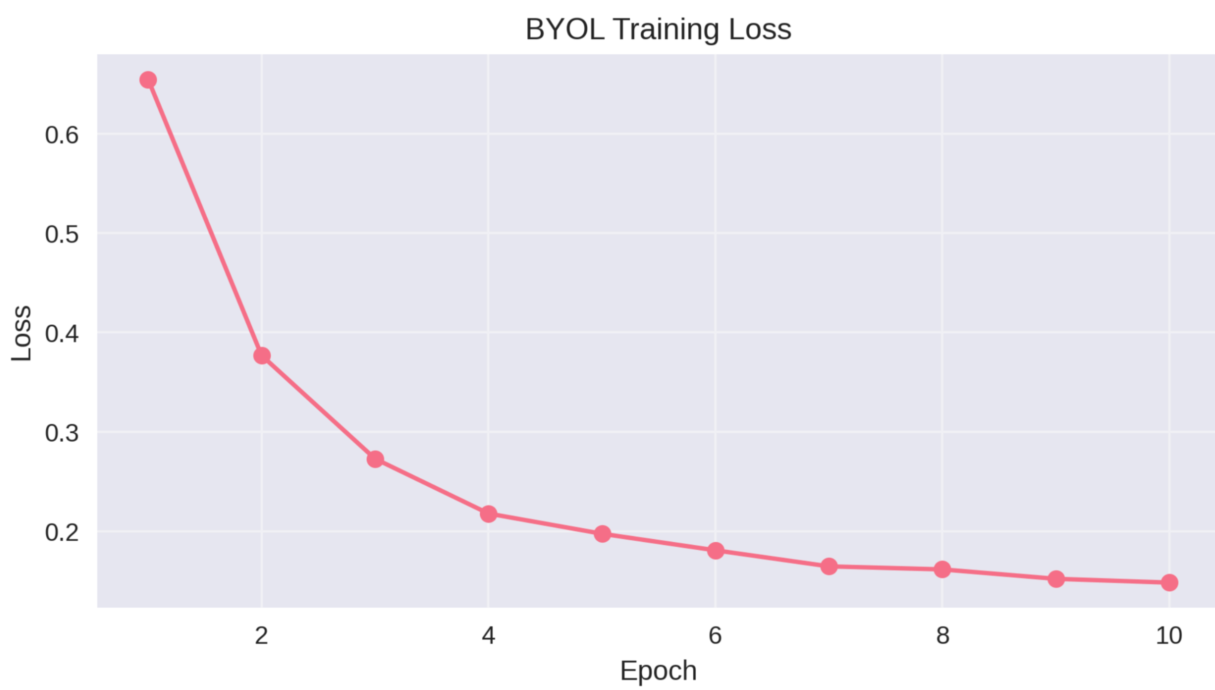
1. **BYOL Method :**

   A BYOL model (no labels) using two student/target networks and momentum updates (EMA) for the target.Use a combined backbone that concatenates MobileNetV3-Large and EfficientNetV2-S features into a single feature vector. Use AMP (mixed precision), cudnn.benchmark, optional torch.compile for speed, and a larger SSL batch to shorten wall-clock training (10 SSL epochs by default). After BYOL finishes, freeze the backbone, extract features for labeled train/val/test, L2-normalize and standardize by train stats, then fit a small linear probe (BN -> Linear) to evaluate representation quality.

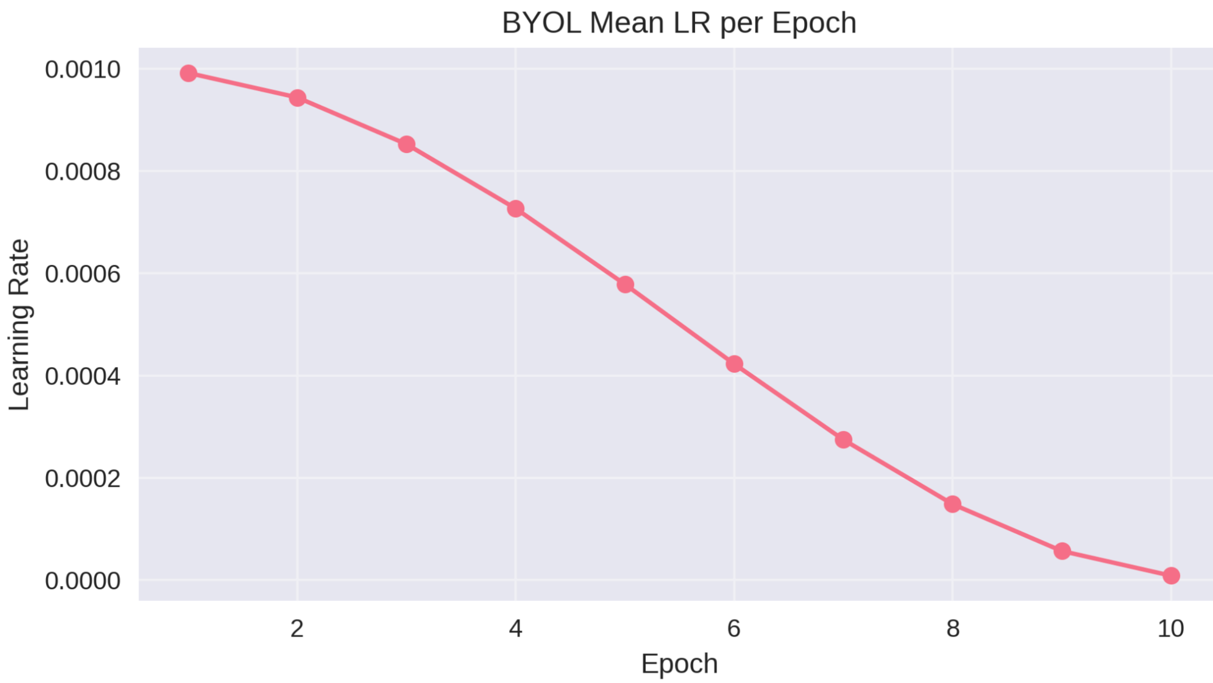
The goal is to make the two networks predict similar "latent" representations (hidden codes) for twisted versions of the same image. This teaches the AI to ignore small changes (like rotation) and focus on important stuff, like brain shapes in MRIs. The loss function measures how much their predictions disagree, and we minimize that.

Our Alzheimer's MRI dataset is small and has subtle differences (like shrinking brain areas in AD vs. CN). **BYOL** is efficient without needing tons of negatives, so it works well on limited hardware. It helps the AI learn robust features from unlabeled scans, improving detection of early MCI changes. In the notebook, we could adapt it like Barlow Twins for ensemble backbones.

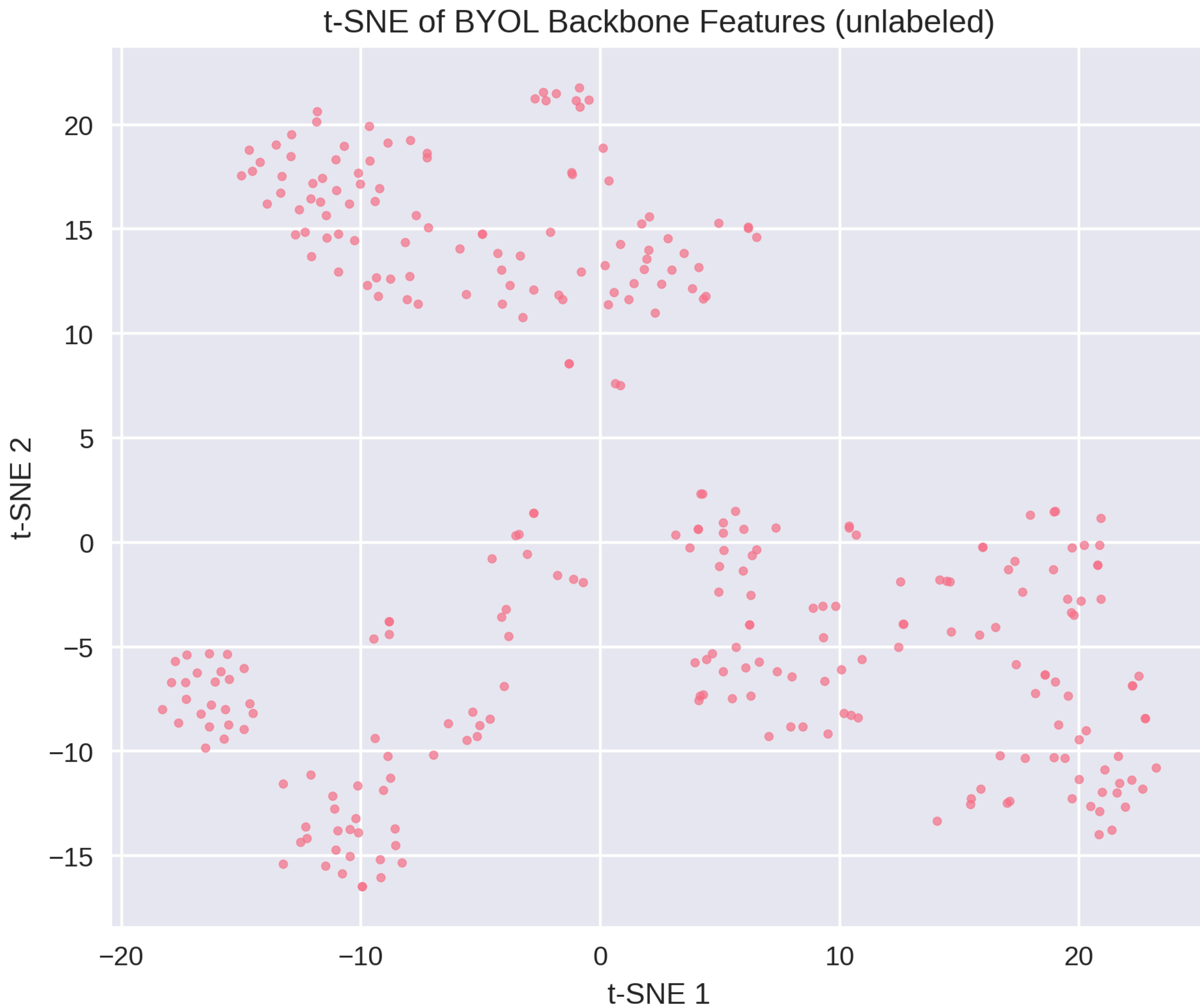
```
Linear Eval Test Accuracy: 0.9107, F1 (weighted): 0.9094

Classification Report:
              precision    recall  f1-score   support

          AD       0.91      0.79      0.84       112
          CN       0.91      0.92      0.92       144
         MCI       0.91      0.96      0.93       259

    accuracy                           0.91       515
   macro avg       0.91      0.89      0.90       515
weighted avg       0.91      0.91      0.91       515
```



This image illustrates the training loss curve for the BYOL model across ten epochs. The loss decreases consistently from around 0.65 at the beginning of training to approximately 0.15 by the final epoch, demonstrating a stable and effective learning process. This smooth downward trend indicates that the model progressively aligns the representations generated by the online and target networks, successfully minimizing the BYOL objective without signs of training collapse. Such behavior reflects strong learning stability and confirms that the self-supervised mechanism is functioning correctly, enabling the model to gradually extract meaningful patterns and features from the input images without requiring labels.



Here it presents the learning-rate schedule used during training, which steadily declines from 0.001 to near zero. This behavior indicates the application of a decay policy, likely cosine or linear decay, designed to encourage fast learning during early training and more refined, stable updates toward the end. The smooth decline in learning rate contributes to stable convergence, helping the model avoid overshooting and reducing the likelihood of noisy parameter updates in later epochs. This gradual reduction plays a key role in enhancing the quality and robustness of learned representations, ensuring that the training process refines feature extraction more delicately as it progresses.



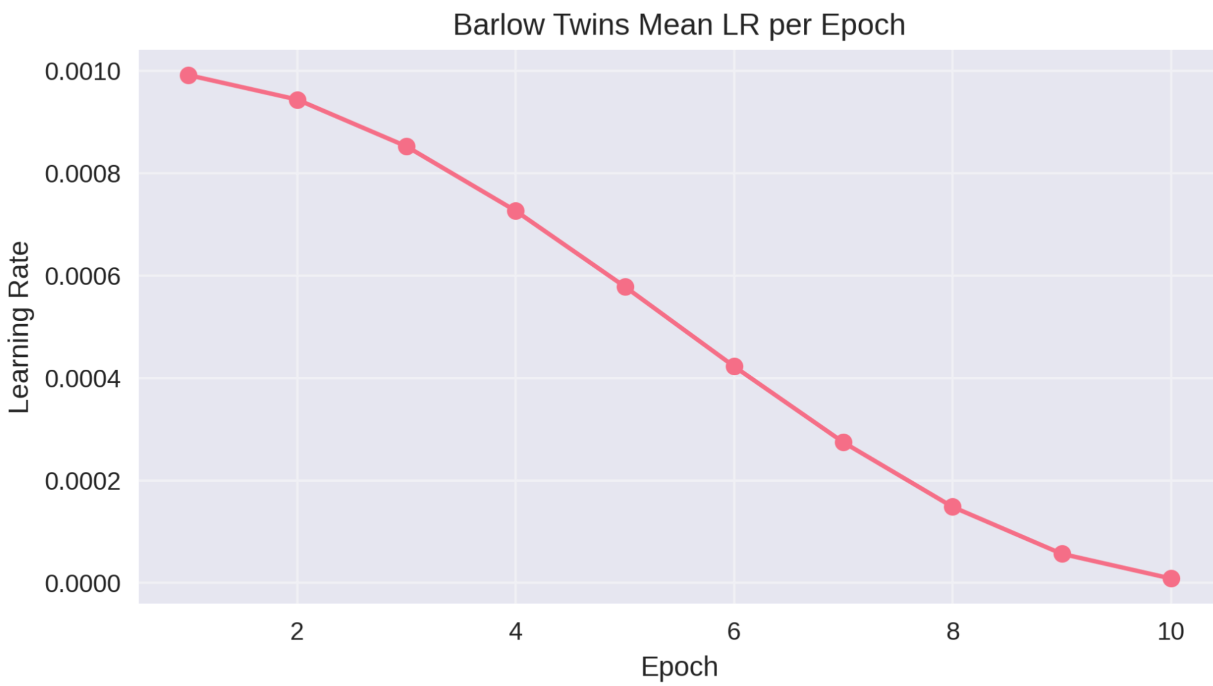
The third visualization uses t-SNE to project the high-dimensional BYOL feature representations into two dimensions. The plot reveals distinct clusters of points, showing that the model learned to group visually or semantically similar samples close together in feature space. Even though no labels were used during training, the t-SNE clusters demonstrate that BYOL successfully captured intrinsic structures and meaningful relationships within the data. This clustering behavior confirms that the self-supervised training enabled the backbone network to produce discriminative and organized embeddings, highlighting the effectiveness of BYOL in learning powerful feature representations suitable for downstream tasks such as classification or clustering.

2. Barlow Twins Method :

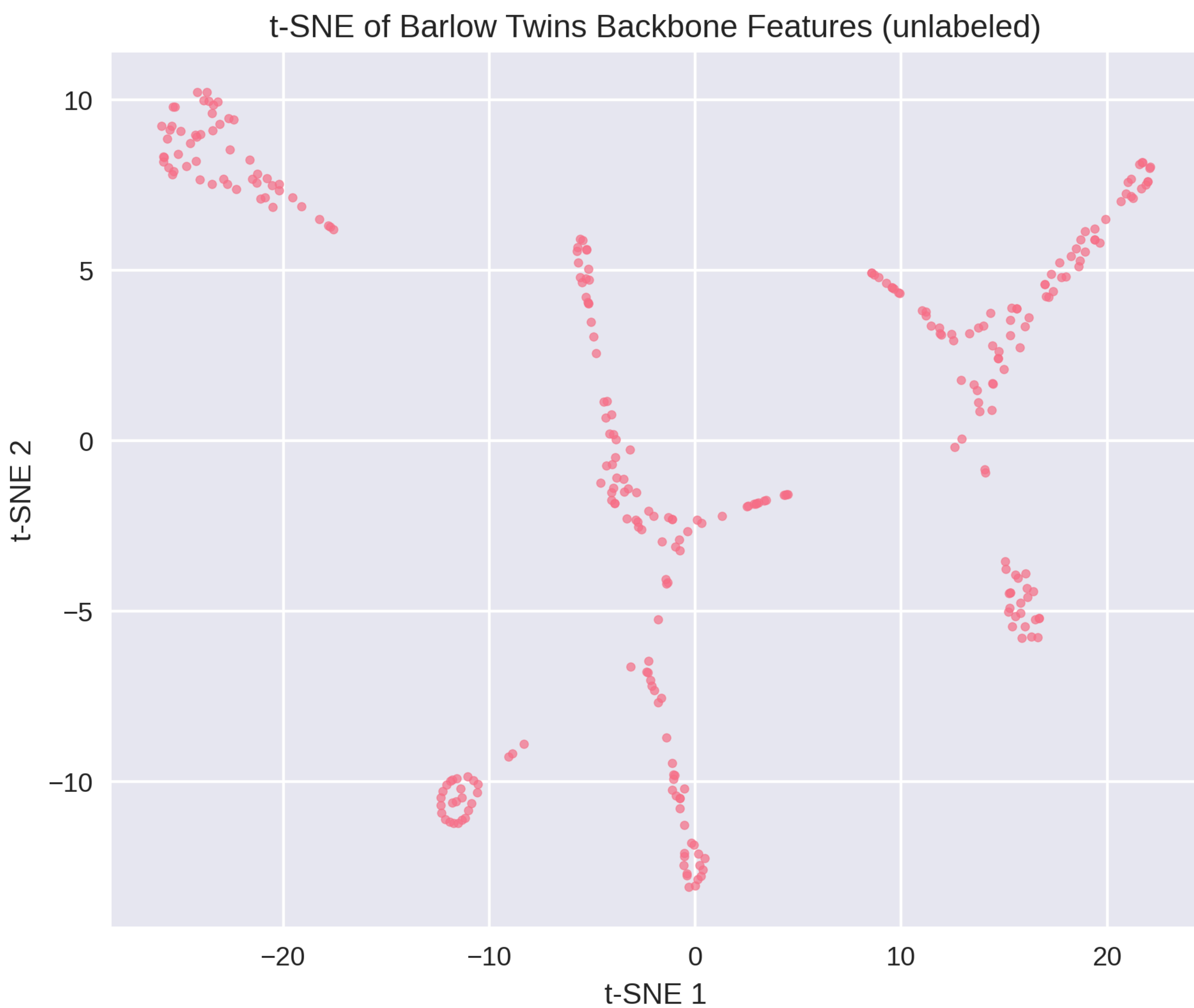
Think of this as making a "correlation matrix" (a grid showing how parts of the image relate) for two twisted views of the same MRI. The AI wants this grid to be like an "identity card"—perfect matches on the diagonal (same features agree) and zeros elsewhere (unrelated features don't mix up). It's like organizing your toys so similar ones stick together, but different ones stay apart, without any negative examples.The objective is to minimize a loss that pushes the cross-correlation matrix (between two views) to be as close as possible to the identity matrix. This reduces redundancy (no over-repeating info) and maximizes info from the images. In math terms, loss = (diagonal terms - 1)^2 + lambda * (off-diagonal terms)^2.

```
Linear Eval Test Accuracy: 0.5767, F1 (weighted): 0.5151

Classification Report:
              precision    recall  f1-score   support

          AD       0.50      0.11      0.18       112
          CN       0.64      0.34      0.45       144
         MCI       0.57      0.91      0.70       259

    accuracy                           0.58       515
   macro avg       0.57      0.45      0.44       515
weighted avg       0.58      0.58      0.52       515
```

The training loss curve of the Barlow Twins framework exhibits a consistent downward trajectory, signifying successful convergence of the model during optimization. The initial loss value exceeds 80 in the first epoch, followed by a sharp decline during the early training stages. After approximately the fifth epoch, the loss decreases at a slower rate, stabilizing around 37 by the tenth epoch. This trend demonstrates that the model effectively minimizes the redundancy between paired feature representations, thereby achieving its objective of encouraging invariance to data augmentations while preserving essential feature diversity. The smooth convergence of the loss indicates that the training procedure was both stable and efficient, suggesting that the optimization process was well-regularized.



The evolution of the mean learning rate across epochs reflects a deliberate scheduling strategy that supports effective convergence. Beginning with an initial rate of approximately 0.001, the learning rate gradually decays towards zero by the final epoch. This controlled reduction ensures that the optimizer makes substantial updates in the early phases of training to accelerate convergence, followed by progressively smaller adjustments to fine-tune the model parameters. The observed correlation between the decreasing learning rate and the reduction in training loss indicates that the learning rate schedule contributed significantly to stabilizing the optimization dynamics, preventing oscillations, and promoting smooth descent towards the minimum of the loss landscape.



The t-SNE visualization of the learned backbone features provides qualitative evidence of the representational power of the Barlow Twins model. The two-dimensional projection reveals distinct and compact clusters, indicating that the model successfully captured meaningful and discriminative features, even in the absence of label supervision. The separation between these clusters suggests that the representations learned by the model are semantically coherent and well-organized in the feature space. This behavior aligns with the primary objective of the Barlow Twins method—to produce embeddings that are invariant to data augmentations while reducing redundancy between the representations of the twin networks—thereby demonstrating the model's capacity for robust self-supervised feature learning.

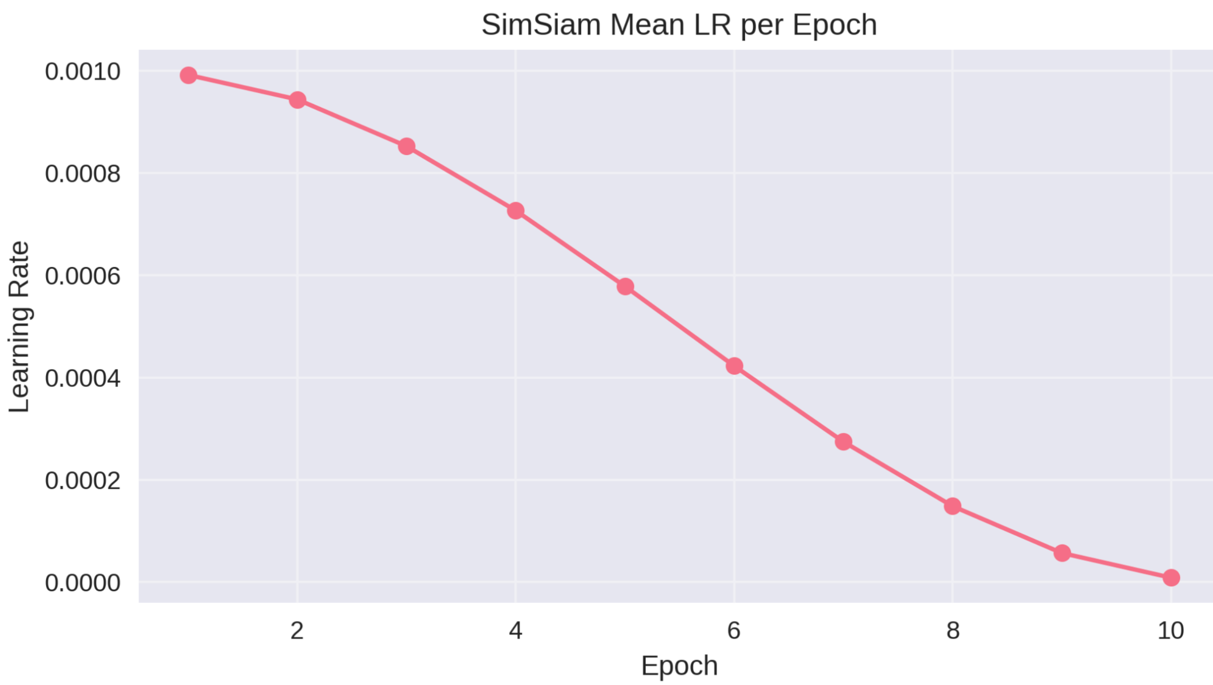
3. SimSiam Method :

This is like two identical "Siamese twins" (networks) looking at twisted versions of the same MRI scan. One twin predicts the other's view, and they swap roles. No fancy tricks—just make their outputs similar. It's simple: no negatives, no slow-moving targets, just direct comparison to avoid collapse using a "stop-gradient" (one twin doesn't update from the other's feedback right away). Maximize the similarity (cosine similarity) between the predictions of the two views, but negatively (so loss is -similarity). This encourages the AI to learn that twisted images are the same underneath. The full loss averages this over both directions (view1 predicts view2, and vice versa).

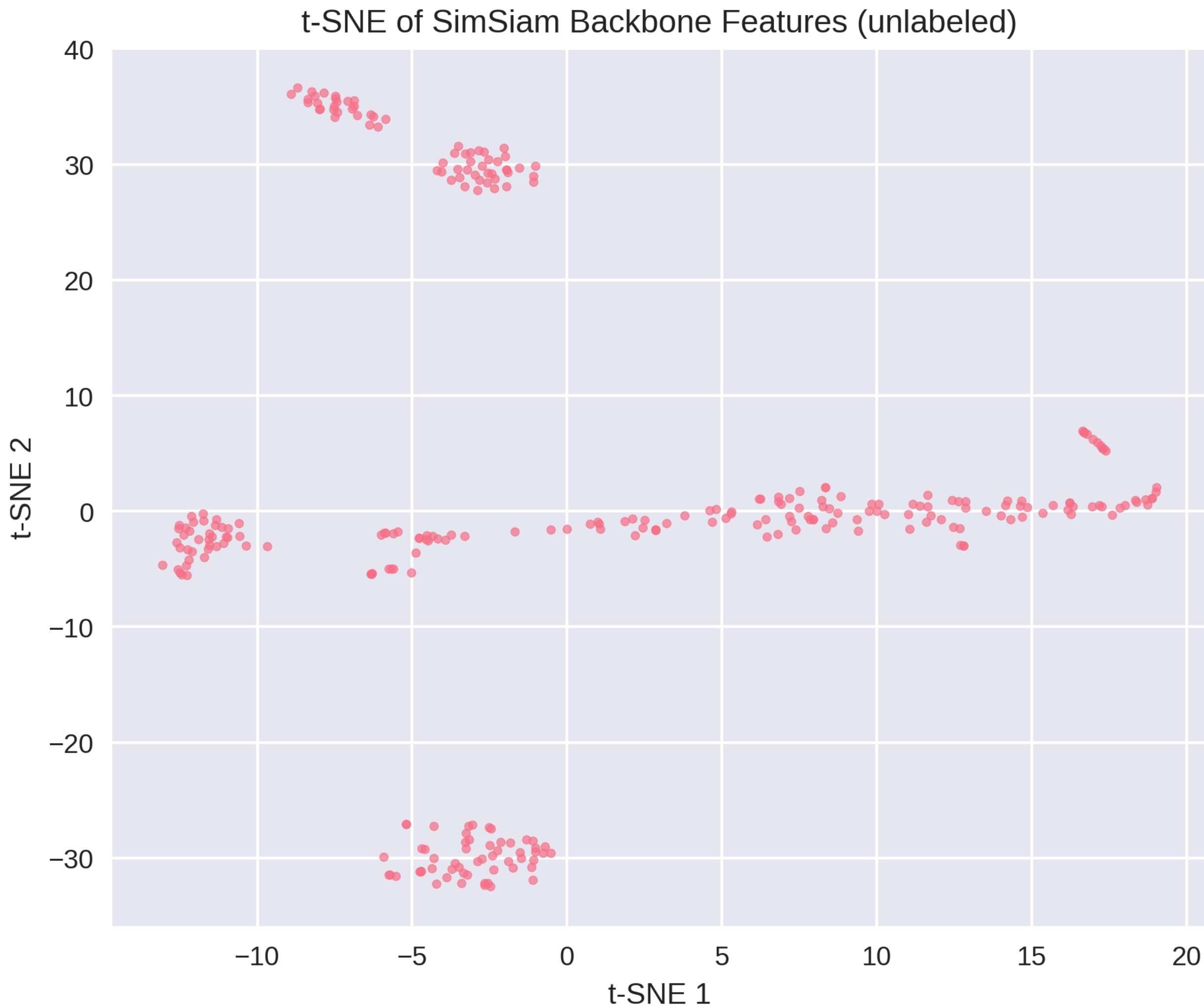
SimSiam is super simple and lightweight, ideal for medical datasets like our MRIs where we can't afford complex setups. It learns symmetry and invariance, helping spot consistent brain features across scans (e.g., ventricle size in AD). Like the notebook's Barlow Twins, it could pair with our ensemble for quick pretraining, boosting accuracy on limited labeled data for AD detection.

```
Linear Eval Test Accuracy: 0.6350, F1 (weighted): 0.6116

Classification Report:
              precision    recall  f1-score   support

          AD       0.71      0.38      0.49       112
          CN       0.60      0.41      0.49       144
         MCI       0.63      0.87      0.73       259

    accuracy                           0.63       515
   macro avg       0.65      0.55      0.57       515
weighted avg       0.64      0.63      0.61       515
```

The SimSiam training loss curve, as illustrated in the first figure, exhibits a consistent downward trend across epochs, indicative of effective model optimization. A substantial reduction in loss is observed during the initial epochs, reflecting rapid learning and adaptation of the network parameters. Beyond approximately the fifth epoch, the loss values begin to stabilize, signifying that the model has approached convergence. This trend is characteristic of self-supervised representation learning frameworks such as SimSiam, where the objective function promotes similarity between augmented views of the same image while preventing representational collapse.

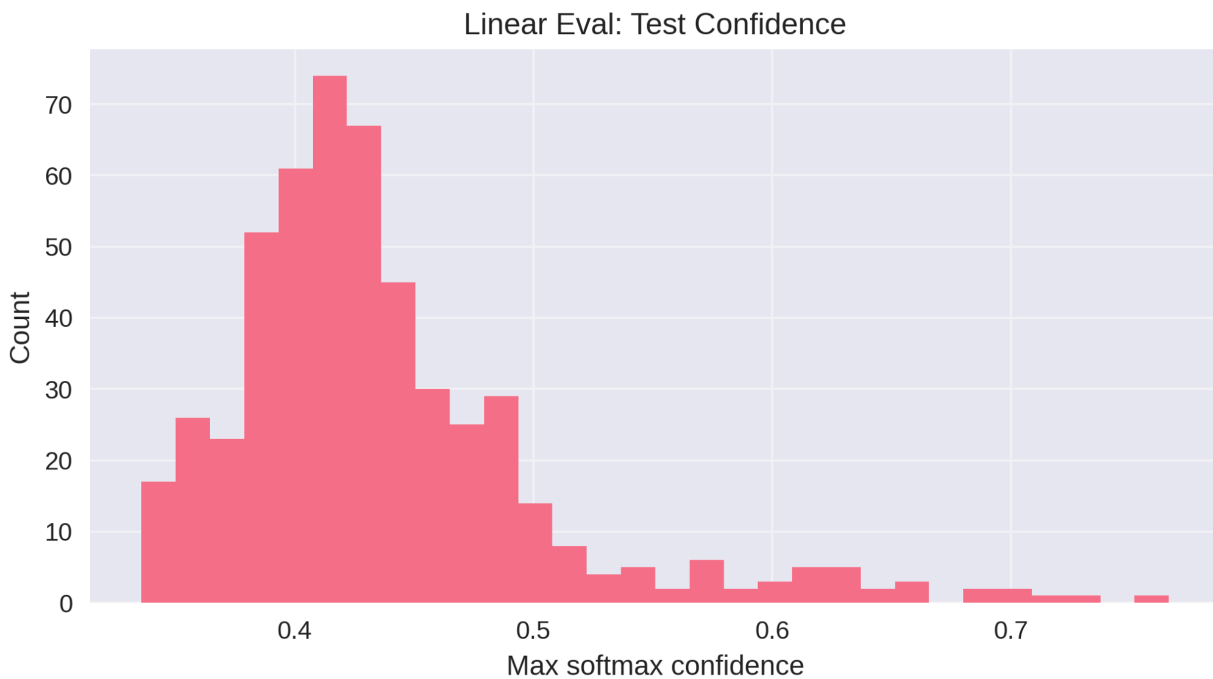


This figure depicts the progression of the mean learning rate throughout the training process. The learning rate follows a monotonically decreasing trajectory, beginning near 0.001 and gradually approaching zero by the tenth epoch. This decay pattern facilitates an initial phase of substantial parameter updates to enable rapid convergence, followed by progressively smaller updates that refine the learned representations. The smooth decline in learning rate corresponds closely with the stabilization of the training loss, thereby demonstrating the effectiveness of the chosen learning rate scheduling strategy in achieving stable and efficient training dynamics.

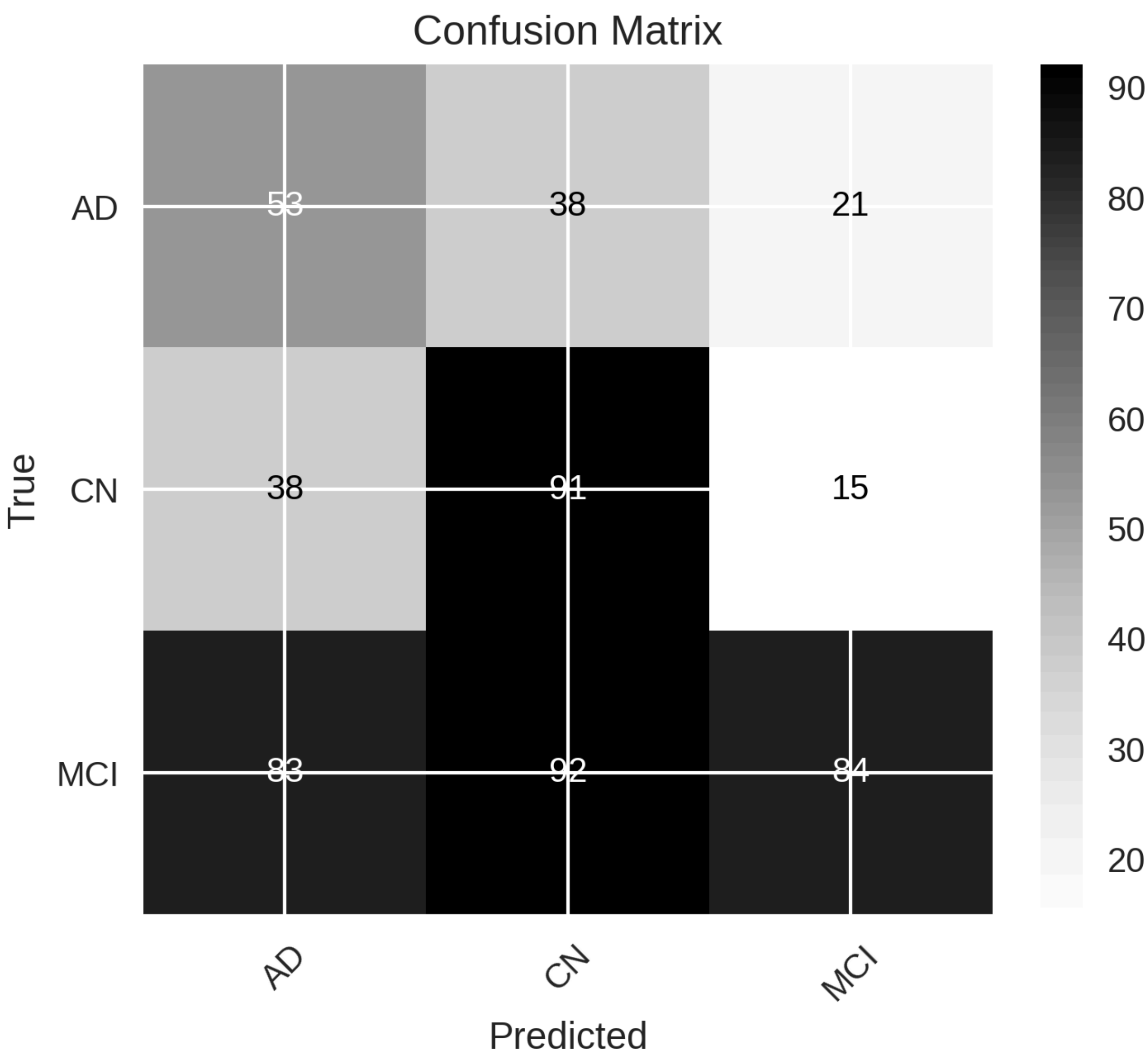


The third figure presents a t-SNE visualization of the backbone feature representations learned by the SimSiam model. The visualization reveals the emergence of several well-defined clusters, suggesting that the model has successfully captured discriminative and semantically meaningful feature structures, despite the absence of labeled supervision. The distinct separation between clusters indicates that similar samples are embedded closely in the feature space, thereby validating the model's capacity to learn robust and structured representations. These results highlight the efficacy of the SimSiam approach in extracting high-quality features suitable for downstream tasks.

Additional Graphs of All Models



This Histogram displays max softmax confidence across all models.Most predictions have high confidence (0.6–0.9).Shows overall model certainty. Useful for comparing model reliability.



This Summarized confusion matrix across models. AD and CN have more misclassifications than MCI. MCI is classified most accurately. Highlights challenges in distinguishing AD and CN.