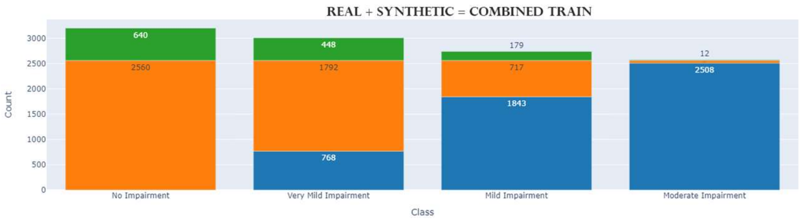
Alzheimer's disease diagnosis with different ResNet sizes

1. Project Introduction	Alzheimer's disease (AD) is a neurodegenerative disease that causes 60-70% of dementia cases. AD is ranked as the seventh leading cause of death worldwide. AD's financial burden on society is large, with an estimated global annual cost of US$1 trillion. AD is usually clinically diagnosed based on a person's medical history, observations from friends or relatives, and behavioral changes. Advanced medical imaging (computed tomography (CT) or magnetic resonance imaging (MRI)) can be used with diagnosis. On MRI or CT, Alzheimer's disease usually shows a generalized or focal cortical atrophy, which may be asymmetric. Atrophy of the hippocampus is also commonly seen. Advanced imaging may predict conversion from prodromal stages (mild cognitive impairment) to Alzheimer's disease. Accurate and timely medical diagnosis is crucial for early detection and better treatment outcomes, and using AI to assist provides numerous benefits, including faster and more precise diagnoses, early disease detection and personalized treatments. It also automates routine tasks, easing the workload for healthcare professionals and detects subtle patterns in medical data that humans might miss. Objective of this project is to test feasibility of different sized Residual Neural Networks (ResNet), a type of Convolutional Neural Network (CNN), for AD classification tasks.

2. Literature Review	In Convolutional Neural Networks to Classify Alzheimer's Disease Severity Based on SPECT Images: A Comparative Study (Lien et al., 2023, p.) accuracy of 65.37 - 68.51% was achieved on transverse, sagittal, and coronal section SPECT image data against the test set with ResNet-model. In Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging (Bae et al., 2020, p.2) accuracy of 88 – 94% per class was achieved against the test set with modified Inception-v4 architecture CNN-model on 3D ADNI-data. In Classification of Alzheimer's Disease Using Convolutional Neural Networks (Samhan et al., 2022, p.22) accuracy of 100% was achieved against the test set with VGG16 CNN-model on 2D JPG image data. In Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation (Wen et al., 2020, p.1) it was discovered that more than half of the surveyed papers may have suffered from data leakage and thus reported biased performance. They state that "the classification performance is difficult to compare across studies due to variations in components such as participant selection, image preprocessing or validation procedure. Moreover, these studies are hardly reproducible because their frameworks are not publicly accessible and because implementation details are lacking. Lastly, some of these papers may report a biased performance due to inadequate or unclear validation or model selection procedures."
3. Materials and Methods **3.1 Data**	The dataset for this project can be found in: https://www.kaggle.com/datasets/lukechugh/best-alzheimer-mri-dataset-99-accuracy/data The dataset consists of JPG images labeled to 'No Impairment', 'Very Mild Impairment', 'Mild Impairment' and 'Moderate Impairment' categories. Each category has 2560 images. The images can not be formatted to 3D images and must be considered and handled as individual 2D MRI-image layers.

The dataset has been preprocessed two ways by the dataset distributor:

- Skull (and other soft tissue besides brain) has been removed from images
- Severe class imbalance of the dataset has been handled by creating synthetic data to account for the size differences of classes



Bar graphs show the relation of real images (in orange) and synthetic images (in blue) in the dataset across target classes.

3.2 Training and Evaluation

ResNets used in this project are: ResNet18, ResNet34 and ResNet50.

20 % of the dataset was divided into a test set with equal (stratified) distribution of the target classes. The rest of the dataset was divided into 5 folds with target classes equally distributed. From each of the folds 80 % was used for training and 20 % for validation.

The model was trained with Adam optimizer with varying learning rates, batch sizes and epoch counts.

Data transformations methods tested during training for better model performance where:

- Normalize
- RandomVerticalFlip
- RandomHorizontalFlip
- RandomResizedCrop
- RandomAffine
- GaussianBlur
- RandomRotation

During the training process, model training and performance was tracked by following training loss, validation loss, balanced accuracy (across target classes), ROC AUC (multi class) and average precision.

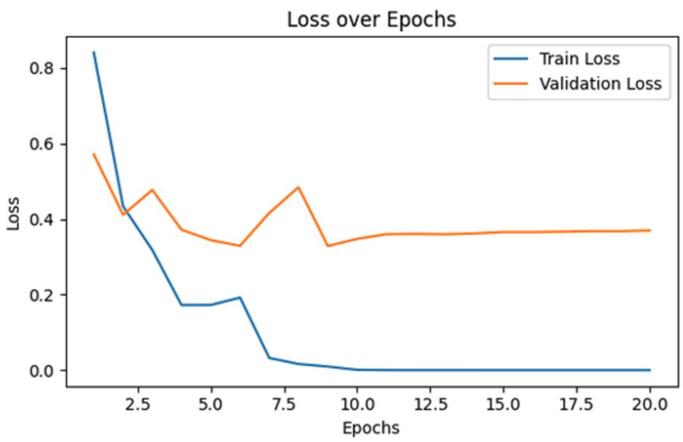
From differences in these metrics the performance of models across different target classes can be observed. The results presented are calculated averages of trained folds tested on the test dataset.

4. Results

Different sized ResNets were first trained for 20 epochs (per fold) with normalization only. Chart below.

Model	Balanced accuracy	ROC-AUC	Average precisions
ResNet18	0.85	0.97	0.91
ResNet34	0.78	0.94	0.86
ResNet50	0.75	0.93	0.83

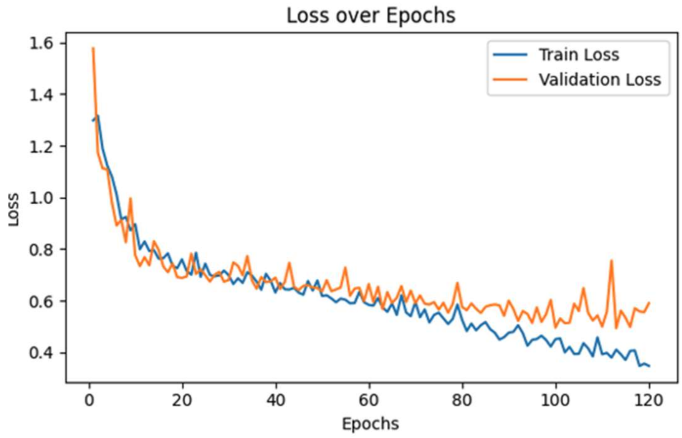
Initially the accuracy of ResNet18 looks very promising. However, the plotted training/validation loss showed overfitting and capping of learned knowledge. Picture below.



More transformations were applied to see if this would help the model to generalize better and to prevent overfitting. More transformations helped with training/validation loss, but also slowed the learning of the model. Epoch count was increased to give time for the model to learn. Transformations applied here are normalization, RandomRotation, RandomVerticalFlip and RandomAffine. Chart below.

Model / Epochs	Balanced accuracy	ROC-AUC	Average precisions
ResNet18 / 20 per fold	0.65	0.87	0.70
ResNet18 / 60 per fold	0.72	0.91	0.78
ResNet18 / 120 per fold	0.76	0.93	0.83

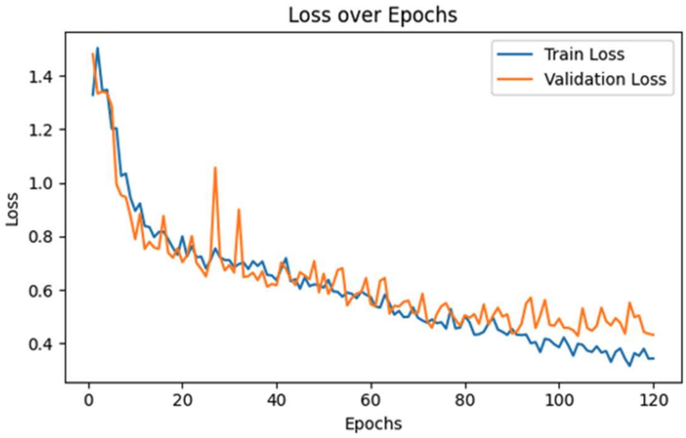
Although 120 epochs per fold during training helps the model a lot, training/validation loss still shows some tendency to overfit after 100 epochs. Picture below.



As increasing epochs resulted in better overall performance, deeper model (ResNet34) was trained with same transformations and 120 epochs per fold. Chart below.

Model / Epochs	Balanced accuracy	ROC-AUC	Average precisions
ResNet34 / 120 per fold	0.78	0.94	0.83

Training ResNet34 120 epochs per fold resulted in marginally better results than ResNet18. However, training/validation loss still shows some overfitting. Picture below.



Main result of this project is that from different ResNet sizes the ResNet18 shows the greatest potential for classification task of this magnitude with 4 different target classes. Resnet34 could possibly achieve the same level of performance as ResNet18 but considering that the training time of ResNet34 is multiple times longer, the likelihood to better results by iteratively training ResNet18 with finetuned parameters is higher.

5. Discussion

5.1 Limitations

Limitations with the test setup in this project include data availability, data quality and limitations in compute resources (mainly lack of time).

Availability of real medical imaging datasets is poor. Optimally the model training would be done to a dataset that is as close to the original data from imaging in DICOM or NIfTI format.

Data quality used in this project is limited due to huge class imbalance of the originating dataset. Without alternative data it is difficult to estimate the effect of synthetic data on the real-world prediction accuracy of models trained in this project.

5.2 Improvements

In addition to improving data quality, automated hyperparameter tuning could be utilized to programmatically find the best combination of different hyperparameters.

References:

https://en.wikipedia.org/wiki/Alzheimer%27s_disease#Diagnosis

https://www.mgma.com/articles/artificial-intelligence-in-diagnosing-medical-conditions-and-impact-on-healthcare

https://www.mdpi.com/2077-0383/12/6/2218