

# Strategies to Faster Diagnosis of Alzheimer's Disease

Roman Studer

School of Engineering  
(Data Science)

FHNW University of Applied Sciences  
and Arts Northwestern Switzerland  
Brugg, Switzerland  
roman.studer1@students.fhnw.ch

Katarina Fatur

School of Engineering  
(Data Science)

FHNW University of Applied Sciences  
and Arts Northwestern Switzerland  
Brugg, Switzerland  
katarina.fatur@students.fhnw.ch

Bruno Baptista Kreiner

School of Engineering  
(Data Science)

FHNW University of Applied Sciences  
and Arts Northwestern Switzerland  
Brugg, Switzerland  
bruno.baptistakreiner@students.fhnw.ch

**Abstract**—The aim of our project was to examine possible methodologies for a more efficient diagnosis of Alzheimer's disease with the help of machine learning. For this purpose, we obtained neuropsychological tests and medical imaging scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We identified cognitive features which are the most informative for staging the participants, explored the data with clustering, and built an MRI image classifier using deep learning methods. The goal of our work was to support the clinicians and optimize the diagnostic process. The analysis of psychological tests has established, that the most valuable insights can be obtained from the *Mini Mental State Examination* and *Neurobattery Test Set*. Based on test features, the clustering algorithm identified four subgroups of subjects, which corresponds to cognitive impairment subtypes in literature. The results from our deep learning models suggest a link between diagnosis, hippocampus, and prefrontal cortex. Finally, we propose an example of how explainable AI on CNN models can support clinicians in diagnostic process.

**Keywords**— *Alzheimer's disease, magnetic resonance imaging, neuropsychological assessment, classification, deep learning, explainable AI, medical data visualization, ADNI*

## I. INTRODUCTION

### A. Alzheimer's Disease

Alzheimer's disease is currently the most common cause of dementia, affecting over 5% of the population aged 65 years and above [1]. The progressive neurodegenerative disease was first described by Alois Alzheimer in 1906, but no cure has been found yet, and only a handful of therapies have been known to slow down its progression [2]. In June 2021, FDA approved the first drug in the last 19 years to treat the prodromal Alzheimer's. However, its effectiveness is questionable, and the drug is applicable only in the very early stages of dementia [3][4].

The first clinical signs of cognitive impairment related to Alzheimer's are memory problems, but the initial damage could occur several decades prior to the onset of symptoms in the brain regions responsible for memory formation, such as the hippocampus and the entorhinal cortex [5]. The accumulation of amyloid plaques and tau proteins disturbs the normal neuronal function, which results in severe brain atrophy [6]. Cognitive deficits, such as forgetfulness, confusion, and disorientation become progressively worse, until the patients lose their independency completely. The memory problems can be measured with psychological tests, while neurodegenerative changes can be monitored with medical brain imaging.

Research and clinical practice have confirmed that early therapeutic intervention can slow down the disease progression and thus ensure a much higher quality of life for the patients [7]. Early diagnosis thus remains the most

important weapon we have in preventing the irreversible neurodegeneration in the later stages of the disease.

The purpose of our work was to provide tools that would help optimize the diagnosing and treatment of the disease. In average, it takes 2.8 years to reach a diagnosis, which in most cases translates to more than 25% of median survival time for the diagnosed person [8]. This time could be shortened though optimization of the diagnostic procedure. We stipulated, that determining the most informative cognitive tests, coupled with AI-assisted MRI image classification, would support this goal. We addressed the issue from different angles: we examined a subset of neuropsychological tests and identified the ones with the greatest predictive ability, we used machine learning to determine important features in an MRI scan, and we focused on the benefits of explainable AI in medical field to support the clinicians in their work.

### B. Data Source

We obtained de-identified demographics and cognitive assessments, as well as imaging data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database through the LONI Image and Data Archive (adni.loni.usc.edu). The study has been ongoing for the last 18 years and has since gone through different phases: ADNI1, ADNIGO, ADNI2, and currently ADNI3 [9].

### C. Data Description – Tabular Data

Information on demographical features of ADNI participants was obtained from the files PTDEMOG.csv and ADNIMERGE.csv for all four ADNI phases. The file DXSUM\_PDXCONV\_ADNIALL.csv provided information about the diagnosis. 2372 individuals have participated in the study, among them 1120 women and 1252 men. Racial composition was as follows: 5 American Indian/Alaskan participants, 53 Asian, 158 Black, 2 Hawaiian, 2116 White, 27 of mixed race, and 11 of an unknown race. 4.5% of the participants were of Hispanic ethnicity. Based on low correlation scores with the diagnostic outcome we dropped some of the demographical features, and examined only a selection: *age*, *gender*, *ethnicity*, and *education*. The cohort's mean age and years of education per diagnostic stage are shown in Table 1 below.

TABLE I. NEUROPSYCHOLOGICAL ASSESMENT, PARTICIPANTS' DEMOGRAPHICS

	Age and Education per Diagnostic Outcome		
	Normal controls, Mean $\pm$ SD	Mildly cognitively impaired, Mean $\pm$ SD	Alzheimer's disease patients, Mean $\pm$ SD
Age	73.2 $\pm$ 6.2	72.9 $\pm$ 7.4	74.4 $\pm$ 7.3
Education	16.5 $\pm$ 2.6	16.0 $\pm$ 2.8	15.5 $\pm$ 2.9

The number of participants per diagnostic stage changes as the disease progresses, so during the study we may encounter three different diagnoses per individual. For this reason, the diagnostic composition of the participants is presented as the number of exams that ended in a certain diagnosis. We have 5937 observations of normal controls (NC), 6490 observations of mildly cognitively impaired (MCI), and 3455 observations of participants with Alzheimer's disease (AD).

In our analysis we used the data from the following neuropsychological tests: *Clinical Dementia Rating* (CDR), *Geriatric Depression Scale* (GDS), *Mini Mental State Examination* (MMSE), *Montreal Cognitive Assessment Test* (MOCA) and *Neuropsychological Battery* (NEUROBAT) [10]. As the testing protocols differed for normal controls and dementia patients, the number of observations per test varies from individual to individual. The main advantage of the tests is, that they can detect cognitive deficits caused by Alzheimer's disease in its early stages.

CDR assesses six domains of cognitive performance: *Memory, Orientation, Judgment and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care*. Only the score of 0 is considered normal. The acquired score is the key criterium for inclusion into one of the subject groups in the study (NL, MCI, AD).

GDS identifies depression in the elderly. The test consists of 15 yes/no questions with maximum score of 25, where the score above 5 indicates different degrees of depression. We included it in the preliminary analysis because studies suggest that there is a link between depression and dementia [11].

MMSE evaluates orientation in time and space, attention, memory, language, and visual-spatial skills. It is fine-tuned for population over 55 years of age, and it is used as a test of cognitive function among the elderly. The total score is 30, a score of less than 21 indicates increased odds of dementia.

MOCA is a rapid screening test to detect deficits in several cognitive domains, from executive function to memory and orientation. Its maximum score is also 30. The average MOCA score indicating MCI is 22, for severe cognitive impairment it drops to less than 10.

NEUROBAT comprises several tests: 1) *Logical Memory*, 2) *Rey Auditory Verbal Learning Test*, 3) *Clock Drawing and Copying*, 4) *Fluency*, 5) *Trail Making Test*, 6) *Boston Naming Test/Multilingual Naming Test*, and the 7) *American National Adult Reading Test (ANART)*. Only a subset of these tests was administered at any given visit. While the first tests (CDR, GDS, MOCA, and MMSE) seem to be of diagnostic importance, the NEUROBAT test set is intended for repeated monitoring of cognitive abilities.

The ADNI diagnostic table contains information on diagnosis at the time of a specific visit (except for screening), as well as information whether the participant deteriorated or convalesced. Transitions of disease stage are NC to MCI, NC to AD, MCI to AD, as well as reverse transitions.

#### D. Data Description – Imaging Data

In addition to the structured, tabular data on subjects and their test results, up to three three-dimensional T1-weighted magnetic resonance imaging (MRI) scans are available per subject [12]. A total of 2294 MRI scans from 639 patients are included in the data set ADNI1. These were collected over a three-year period. MRI scans were obtained per patient over a

period of twelve months, where each patient received at least three MRI screenings. Patients in this dataset all underwent an MRI scan at the initial screening, and after six and twelve months [9]. The dataset is available on the ADNI website under the name "ADNI1: Complete 1Yr 1.5T". Images are divided into three groups: CN = Control Normal, MCI = Mild Cognitive Impairment, and AD = Alzheimer's Disease. The image data, available in Nifty format (.nii), shows a scan of the subjects' brain in the transversal, frontal, and sagittal plane (as seen in Fig. 1 below). This allows for a 3D reconstruction of the patient's head.

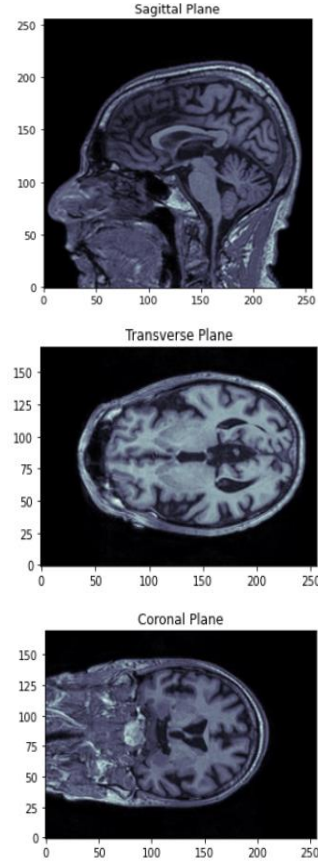


Fig. 1. MRI imaging example over all three planes

All scans included in the data set were quality controlled and meet the condition of desired spatial coverage, spatial resolution ( $1.0 \times 1.0 \times 1.2 \text{ mm}^3$ ) and a maximum scan time of 9-10 minutes. Most images underwent one or more forms of preprocessing. For example, most were adjusted with gradient non-linearity and intensity inhomogeneity correction (label "GW\_N3"). Furthermore, scans with distortion correction were provided. These scans contain the suffix "scaled" or "scaled\_2" in the file name [12].

## II. METHODS

### A. Preparation of Tabular Data

The data we received was incomplete. We examined it for inconsistencies and missing values. Possible duplicate entries were observed in the transition from phase ADNI1 to ADNI2, which we deleted. We validated birth dates and test scores, as well as removed outliers, where appropriate. Since a subset of tests was not administered at every visit, this resulted in systematic missing values. Where possible, we

reconstructed a participant's data by copying the information from the dataset which contained it.

In some columns the amount of missing data was more than 70%, but we decided against imputing to avoid introducing bias. Completely empty rows and unused columns were removed. We also removed rows, for which information in the ADNIMERGE or PTDEMOG file was missing, since we had no way of inferring it. In the absence of test scores, we calculated them with functions, where sufficient input was provided.

In terms of missing values, the NEUROBAT dataset with 80 variables and a sparse data dictionary was the most challenging to process. Due to adjustments of the protocol between different ADNI phases, unifying the data was difficult. We also discovered that the missing values ensued due to a special testing protocol: the test tasks were distributed between two visits, where some tests took place only during certain exams. Given the crisscross pattern of missing values, we decided to separate the test into 7 subtests: *Logical Memory, Rey Auditory Verbal Learning Test, Clock, Fluency, Trail Making, Naming Test, and Reading Test*.

To understand the data better, we used data visualization, by means of which we explored variable distributions and correlations. In the pre-clustering transformation, varying maximum values among different test scores suggested that we should transform them. We scaled the features to a given range with MinMaxScaler, which made them more suitable for comparison. Since most of the variables exhibited non-standard distribution, we attempted to rectify this by using StandardScaler.

Due to higher-than-desired dimensionality of our data, we preprocessed it with Principal Component Analysis. The technique is implemented to reduce the dimensionality and increase interpretability with minimal information loss. The method calculates uncorrelated new variables that are linear functions of those in the original dataset, and successively maximizes variance to retain as much original variability as possible [13].

To identify subgroups with unsupervised machine learning, we implemented k-means clustering [14]. The algorithm groups the data into k number of clusters, where the optimal k needs to be defined with the within-cluster sum of squares (the “elbow method”). To find the elbow point, we drew an inertia plot (sum-of-squared-error (SSE)). The point where SSE begins to decrease in a linear manner is the elbow point and determines the number of clusters for the algorithm.

### B. Image Preprocessing

To improve model performance, we preprocess and augment the input images. During training and testing, all images are transformed in the same manner, except for image preprocessing on the test set, which refrains from all forms of random transformations.

a) *Greyscale Transformation*. In the data loader, the images are transformed from a 3-channel RGB format (extracted from .nii-files) to 1-channel grayscale images. As the MRI Images do not contain information encoded in color, only in pixel intensity (as in a black and white image), we can reduce the image to a grayscale image without losing information. Additionally, no fourth channel (i.e., transparency) is available.

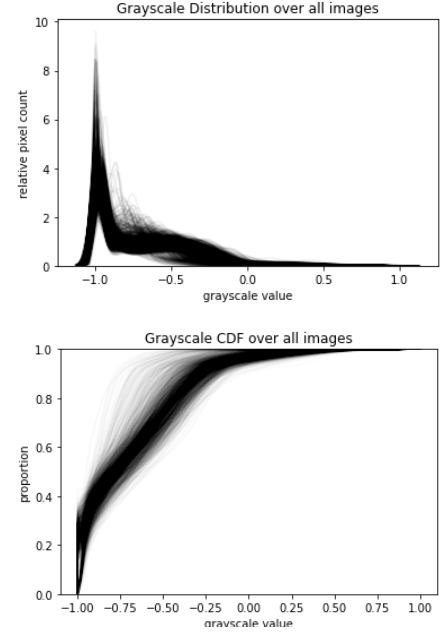


Fig. 2. Histogram (above) and ECDF (below) of grayscale pixel values over all images (center slice of sagittal plane)

b) *Histogram Normalization*. In the next step, the images are normalized using histogram normalization. Fig. 2 shows the distribution of pixel intensity in the grayscale images after standardizing the image into range  $[-1,1]$  with the mean of 0. We see a right skewed distribution above, with a high relative frequency at the left end of the distribution. This is due to the high number of background pixels in the images, which are predominantly black. By performing a histogram standardization over all images, we reduce the noise over all densities. Normalizing the intensity over all images can greatly improve model performance as stated in [15].

c) *Resizing*. In the concluding step, the images are resized to an appropriate square size (either 256px or 512px).

### C. Data Augmentation

Using image augmentation, we can artificially expand the size of our dataset. It has been shown that the performance in image classification models on a test set can be greatly improved by training with augmented data. By effectively increasing the train set with an infinite set of adapted images we not only decrease the generalization gap of our model but make it more robust to further, unseen data. MRI images are particularly suitable for some data augmentation methods.

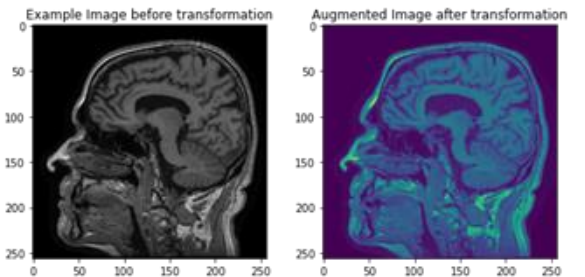


Fig. 3. Data Augmentation with visible deformation (right) based on reference image (left)

For example, soft tissue like brain can be spatially transformed. By means of a random zoom, the images are expanded, but their structure is not adjusted. A spatial deformation (see Fig. 3) is an additional possibility to augment the data in a basically unlimited way. The effect on the model must be well considered. A deformation of the tissue can destroy important diagnostic indicators. Additional random adjustments like ghosting, blur, bias field, or noise can further improve the training. Unsuitable are transformations like flipping and random motion effects.

#### D. Models

##### 1) Convolutional Neural Network (CNN)

This is a neural network that is especially suited for image data. The simplest CNN models typically consist of three layers. A convolution layer, pooling layer, and fully connected layer. The convolution layer performs a dot product between two matrices (a part of the input image and a kernel matrix). The kernel is used as a sliding window and executes the dot product on the overlapping part of the input matrix. The resulting activation map contains extracted features of the input. In the early layers, features such as edges or curves are extracted. In deeper layers, more complex patterns are extracted. The advantage of CNNs, in contrast to conventional neural networks (MLPs), is the sparse interaction with the input. Instead of mapping single pixels to an output element, the kernels can map multiple pixels to an output at once. This makes it possible to extract rough structures in the image (like lines etc.). During training, we utilized multiple known CNN model architectures such as ResNets and DenseNets. Especially DenseNets are suited for dealing with MRI data and are extensively referred to in the literature [16]. Thus, versions pre-trained on MRI are available as well.

##### 2) Evaluation Metric

Evaluating model performance during runtime is achieved by closely monitoring an evaluation metric. The distribution of classes in the target variable is in balance. Since we are not operating on an unbalanced dataset, well-known metrics such as accuracy are applied. The Accuracy Score is the relative accuracy over all classes and represents the ratio of True Positives and True Negatives over the number of total predictions. In a subsequent evaluation of the models (partly by visual inspection through Grad-CAM), additional metrics such as sensitivity, specificity and F1-score were obtained.

##### 3) Cross Entropy Loss

The Cross Entropy Loss, also known as Log Loss, is the negative log-likelihood and defined for a multiclass problem. It is used to evaluate a probabilistic outcome of a model. Such as a logistic regression, or a neural network with a sigmoid output activation. The multiclass Cross Entropy Loss is defined as seen in formula 1 [17].

$$\text{CrossEntropy} = L_{\log}(Y, P) = -\log \Pr(Y|P) \quad (1)$$

$$= -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} y_{i,k} \log p_{i,k}$$

##### 4) Focal Loss

A similar loss function that is specifically suited for a classification task on image data is Focal Loss. This loss is especially useful for image classification, as it deals with the foreground-background class imbalance and prevents the vast number of easy negatives from overwhelming the

detector. The loss expands cross entropy by adding the factor  $(1 - p_t)^\gamma$  to cross entropy (see formula 2). This factor reduces the relative loss for well-classified examples (e.g.,  $p_t > .5$ ) focusing on hard, misclassified examples.

$$\text{FocalLoss} = FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log p_t \quad (2)$$

In the original paper  $\gamma = 2$  worked best. The  $\alpha_t$  stems from the balanced Cross Entropy and is in practice often the inverse relative frequency of the target class, or set as a hyperparameter where alpha is  $\alpha \in [0,1]$  [18].

##### 5) Optimizer

All models are optimized using Adam or AdamW, respectively. Adam can intuitively be seen as a combination of Gradient Descent and Root Mean Square Propagation (RMS). On one hand, Adam uses Momentum, which uses the exponentially weighted mean of the gradients to convert more quickly to a local minimum. In addition, with RMS, the exponential running mean of the gradients is computed. It can be shown that Adam achieves better results than, for example, SGD or AdaGrad [19].

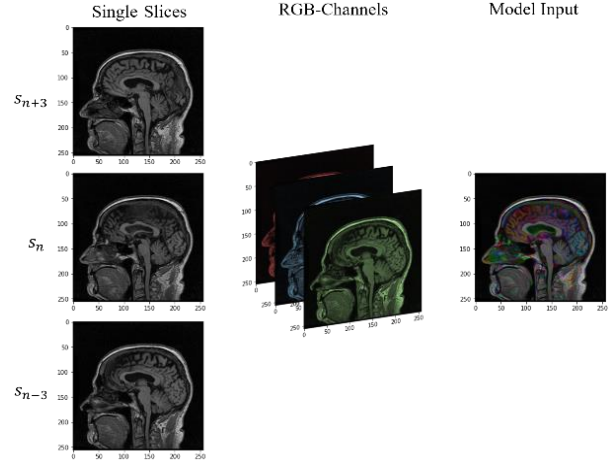


Fig. 4. Example of stacking of multiple slices of a single plane (sagittal) to a multichannel model input

### III. EXPERIMENTS

#### A. Single Image Slice per Plane

Initial models (DenseNet and ResNet) were trained on single slices per plane. We extracted all middle slices per plane from each available MRI scan. This gave us three images that individually act as model input (See Fig. 1, above). We trained multiple models per slice and continuously adapted pre-processing and hyperparameters based on the current results. By training models at each level, we hope to identify regions that have a strong effect on prediction.

#### B. Skull Stripping.

Skull stripping refers to the removal of tissue and bone in MRI scans that do not belong to the brain. The method uses image segmentation to divide MRI scans into *brain* and *non-brain* regions. By using only gray matter as input, noise is reduced, which forces the model to base its decision on regions of the brain and not the surrounding tissue [20]. Again, several models were trained per level, each of them receiving the extracted center slices of the skull-stripped scans as input.



The differences in model performance and visual differences in Grad-CAM are discussed in more detail in the results section.

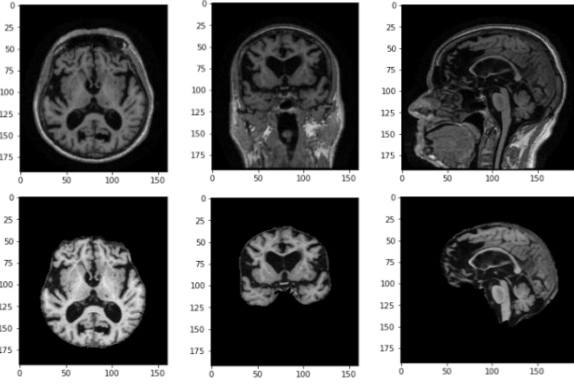


Fig. 5. Example of skull stripping with raw input image (top) and stripped image (bottom)

### C. Stacked Images

To further improve model performance, we made more information available to the model by stacking multiple slices from the same plane as seen in Fig. 6. Per scan we extracted the center slice  $s_n$ , and obtained the neighboring slices  $s_{n-3}$  and  $s_{n+3}$ . By stacking these images into a multi-channel image (visualized as an RGB-Image in Fig. 3) we create a multichannel input for the 2D CNN Models.

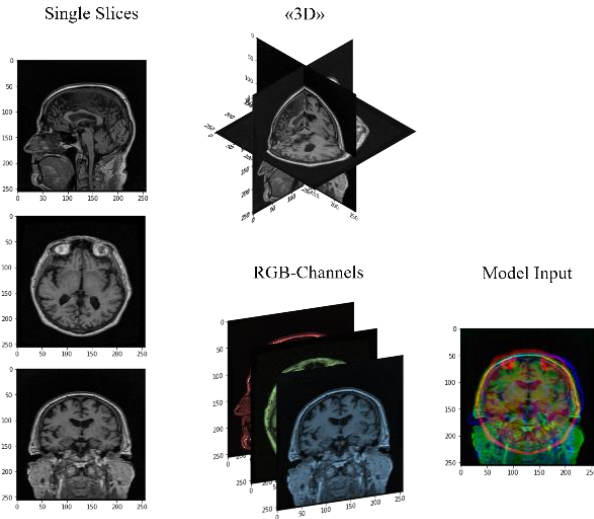


Fig. 6. Example of stacking slices from multiple planes into a multichannel model input

Finally, we trained models which use one slice per plane instead of several slices of the same plane. This allows us to create a two-dimensional representation of all three dimensions of the MRI scan. The model input shown in Fig. 6, here represented as an RGB-Image, is a three channel, two-dimensional tensor. A channel represents a plane of the scan. By doing so we further increase the amount of information available to the model.

## IV. RESULTS

The findings are preliminary and are intended as a steppingstone in further research. The most important results are cleaned test datasets, identification of the most valuable cognitive tests, definition of the four subgroups in the

participants' base (instead of three), and a prototype of a CNN model, upon which further improvements are planned.

### A. Tabular Data – Cleaning and Explorative Analysis

#### 1) Data Cleaning.

During data preparation we identified several inconsistencies within the data. Current tables with demographic data (ADNIMERGE, PTDEMOG) would need to be updated, because there is a discrepancy in the number of participants: there are 1008 individuals, who are recorded in PTDEMOG, but not in ADNIMERGE, and further 52 participants, whose information can be found in ADNIMERGE, but not in PTDEMOG. This results in an unnecessary discarding of entries for which no reliable demographic information is available. However, with an informed reconstruction of missing entries we were in some cases able to retain more than 99% of data. This was not true for the NEUROBAT data, where we retained only complete rows, which resulted in a 58.4% loss of entries.

#### 2) Demographic Data Insights.

Explorative analysis revealed that certain features carry no correlation with the diagnosis, and thus *handedness*, *marital status*, *type of home*, and *work* were discarded. The variables *education*, *gender*, *age*, *race*, and *ethnicity* were retained.

The age when AD is detected was in our dataset roughly normally distributed, with 50% of patients having been diagnosed between the ages 66 and 78. For early onset AD, the age was much lower, with the first symptoms occurring already around the age of 50.

In our sample population we also observed that the ratio between AD and NC participants was roughly 1:9 across all races, except for Native Hawaiian, where the ratio was 1:2. We were not able to confirm a higher incidence rate of Alzheimer's among the Black ethnic group [21], nor did we observe a higher incidence or earlier onset among the Hispanic people [22]. However, the study group was predominately white, comprising less than 11% of other races, so no reliable conclusions could be drawn, and the findings are not statistically significant. The underrepresentation of non-white races has important negative consequences on the quality of data and research potential.

#### 3) Test Scores.

The scores in CDR and GDS values were related to the extent of participants' cognitive deficits. NC participants had a CDGLOBAL score of 0.04 with standard deviation of 0.13, MCI had  $0.47 \pm 0.16$ , and AD patients had  $1.01 \pm 0.54$ . This is line with the diagnostic criteria: the accepted CDGLOBAL score for NC is 0, for MCI 0.5, and for AD 1.0-3.0 [23].

Contrary to our expectation, slight depression was observed only among MCI, while NC and AD exhibited normal scores. A correlation with the diagnosis, that was greater than 0.5, was observed in 15 variables belonging to the tests NEUROBAT, MMSE, MOCA, and CDR. MOCA score and MMSE score had a moderately strong negative correlation (-0.62 and -0.61). The highest negative correlation was exhibited by the NEUROBAT test segment *Logical Memory – Delayed Recall* and *Logical Memory – Immediate*, with correlations of -0.67 and -0.64, respectively. Another variable measuring memory, which also bore the highest positive correlation, was CDMEMORY from the CDR test, with a coefficient of 0.80. *Ray Auditory Learning Test* from NEUROBAT contributed seven significant variables, all with the coefficients between -0.51 to -0.60.

#### 4) PCA Dimensionality Reduction.

The dataset with the 15 most strongly correlated variables was transformed, and its dimensions were reduced with the Principal Component Analysis (PCA). Observing the scree plot, we determined that we would retain just the first two principal components (PC).

Combined, they explain 74.67% of the total variance in the data (Fig. 7). NEUROBAT features contributed the most to the principal component 1 (PC1), mostly these were the trials' variables from the *Rey Auditory Learning Test*. CDGLOBAL projection contributed the most to the principal component 2 (PC2). (Please refer to the Loadings Matrix in Appendix A.1 for more details.)

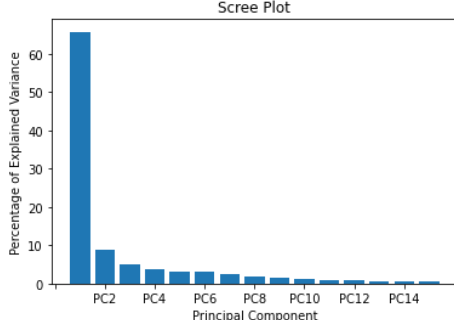


Fig. 7. Cummulative variability

#### 5) K-Means Clustering

We ran clustering on the principal components PC1 and PC2. To determine the number of clusters, we referred to the elbow method. Since the point where SSE begins to decrease in a linear manner was ambiguous, we tested clustering with three and four clusters.

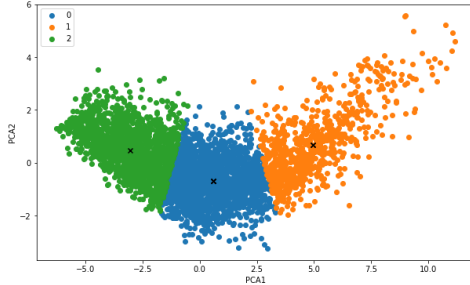


Fig. 8. K-Means with 3 clusters

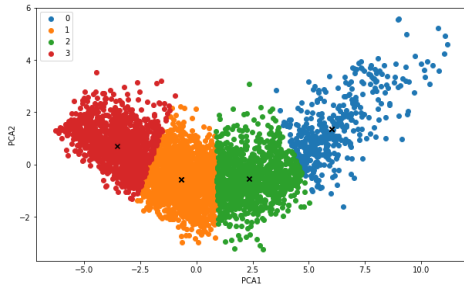


Fig. 9. K-Means with 4 clusters

Using 4 clusters (Fig. 8) reduced the average distance from centroids, and thus decreased the variability within cluster, which was the desired result. Rand index for four clusters was 0.65. Since true labels are available as patients' diagnosis, we also measured the accuracy score, which was 0.48.

#### B. MRI Images

##### 1) Single Slice and Skull Stripping Results

During training, we achieved on the train set for the sagittal plane a maximum accuracy of 45.31 by using a ResNet18 architecture without pretrained weights. We observed severe overfitting on most runs despite extensive data augmentation. One possible cause is, that the alignment of the extracted slices differs significantly from scan to scan, which would result in a high difference in the distribution of the test set to the training set. Table 2 shows the mean accuracy and loss per plane. We can observe very similar performance on the transversal and frontal plane. The high standard deviation on the sagittal plane confirms the suspicion of alignment problems in the MRI scans. Individual training results are available in the Appendix A.2.

TABLE II. MEAN ACCURACY AND LOSS PER PLANE ON TRAIN SET

Plane	Mean Accuracy	STD	Mean Loss	STD
Sagittal	39.75	4.98	0.350	0.120
Frontal	41.63	1.56	0.230	0.016
Transversal	41.47	1.89	0.200	0.013

##### 2) Single Slice Comparison to Skull Stripping

In Fig. 9 and Fig. 10, respectively, we compare the result of trained models on raw MRI slices to skull stripped MRI slices using Grad-CAM. Grad-CAM (Gradient-weighted Class Activation Map) is a technique used to visualize the decision of a convolutional neural network. It takes the gradients of the target flowing into the final layer of the model. It calculates the gradients of the output of the model (before the softmax function) with respect to the convolutional layer that we want to analyze. Here we can choose which class (CN/MCI/AD) to base the gradient on. Normally, the last convolutional layer is analyzed. The gradients are then global-averaged-pooled and multiplied with the convolutional layer's activation. Finally, ReLU is applied to the product to focus on features that have a positive influence on the class of interest.

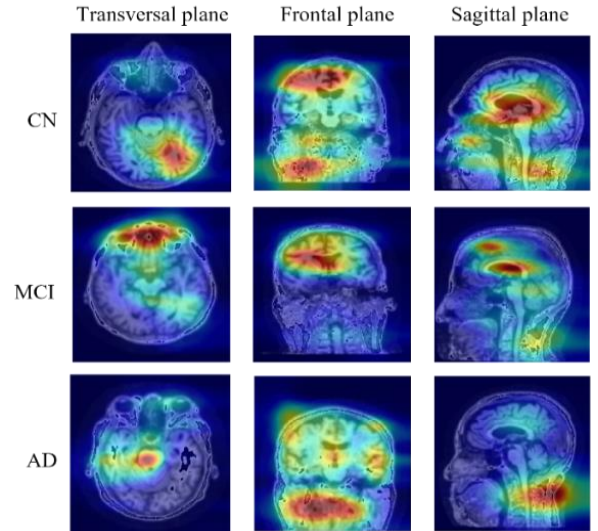


Fig. 10. Examples of Grad-Cam Activations per plane and class

The result is a heatmap of the size of the convolutional feature maps, highlighting the important sections in the image used for predicting the class of interest [24]. We observe a noisy picture in Fig. 10, where the model often focuses on tissue that is not part of the brain. Especially in the sagittal and

frontal plane, where the ratio to gray matter and surrounding tissue is the highest, a high variance can be observed where a large portion of the neck has high activation on the Grad-CAM. By removing the tissue around the brain (as discussed in section III. B.), we force the model to concentrate on the brain tissue. We observe a more focused activation per class and plane in Fig. 4, with a notable focus on the hippocampus and prefrontal cortex. We observed an increase in accuracy on the train set of 10.55% on average, using skull stripped images in comparison to raw images.

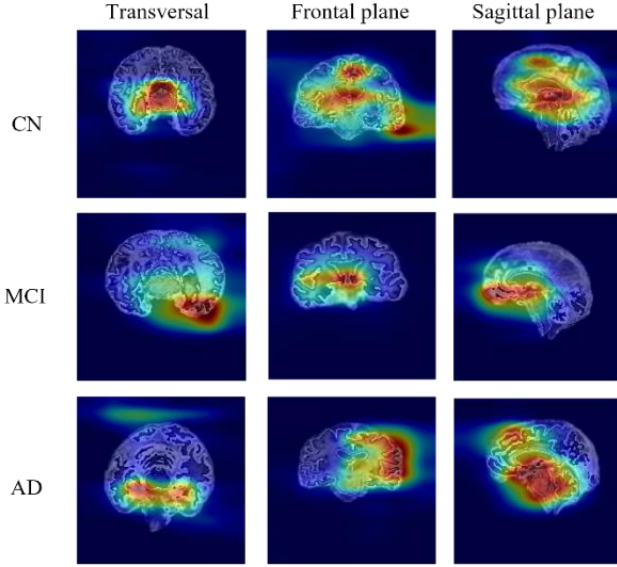


Fig. 11. Examples of Grad-CAM Activations per plane and class for skull stripped MRI scans

### 3) Multi-Channel Results

After training on single slices, we trained models using multichannel input as discussed in section III. C. We observe an improvement in performance in comparison to single slice models. A maximal accuracy of 50% has been achieved, which marks an improvement of +5% on single slice models using a triplanar approach, meaning using a slice per available plane as input. No significant improvement has been observed using stacked images of the same plane.

## V. DISCUSSION

### A. Clustering Results and Implications

Considering the k-means clustering findings, we would need to revisit the cognitive test features and true labels of our dataset. Since the literature suggests, that there are two subtypes of MCI [25], four clusters seem appropriate for grouping the participants, which is confirmed after mapping true labels to clusters. However, the utility of cognitive tests must be re-interpreted. The Loadings Matrix in Appendix 1 demonstrates, which test features contributed the most to which principal components (PC). The resulting clusters were thus based mainly on the participants' differences in RAVLT and MMSE tests, since these variables contributed the most to the chosen PC1 and PC2. Information from the features that contributed to further 13 PCs was less represented (see A.1 Loadings Matrix).

Our data was noisy, and the features exhibited very different distributions. It is still not clear, how many actual subtypes of MCI and AD there are. For this reason, it might be advantageous to investigate other clustering algorithms,

which do not require a specified number of clusters a priori. Density-based spatial clustering (DBSCAN) is one such algorithm, which is robust to outliers and is capable of isolating nested clusters as well – something that we could use due to multicollinear effects of age, gender, and ethnicity on diagnosis.

Dementia is a multifactorial disease; hence sufficiently accurate segmentation or prediction might not be possible based on limited selection of variables. Considering our unpublished preliminary observations and in line with the research, a need for additional data might be warranted. It has been shown that active social life, healthy diet and exercising diminish the risk of developing MCI, as well as slow down the progress of AD [26]. In our work we focused solely on cognitive features, but to build a better predictive model additional data could be collected on these factors as well.

### B. Explainable AI using CNNs

The explainability of machine learning models is particularly important in medical data analysis due to the legal issues that could arise. For example, the GDPR (General Data Protection Regulation) states in article 22(1) that “the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her”. [27] This can be bypassed by laying down suitable measures or by explicit consent from the data subject. “Suitable measures” are subjective and require transparent machine learning models. In the end, the final decision in a healthcare procedure should always lay in human hands, and automating every process with the help of AI should not be the end goal. Only a subset of machine learning which carries a significant legal or similarly important effect will ever be fully automated [28].

The GDPR article provides insight into how regulations are adapting to automated systems based on machine learning. Regulations such as the “Individual’s Right under HIPAA to Access their Health Information” have been put into effect, which gives the patient the right to access and receive a copy of their protected health information (PHI), file a complaint, and even request an explanation or summary of their records [29]. Furthermore, the decision of an AI needs to be explainable to ensure trust and transparency, be it between clinician and patient, or humans and AI models. Developing AI models with ethical implications in mind might also make the product more accessible and appealing. Deep learning models are susceptible to bias and can be even trained to maximize profitability [20]. The prediction of a machine learning model is inherently dependent on the data it has been trained on. In the ADNI dataset, patients are mainly Caucasian. We do not know whether a patient with a different ethnicity is susceptible to less accurate predictions by the deep learning model. Making the models as transparent as possible can help fight such problems.

All of that emphasizes the importance of developing explainable AI methods. Our goal was to combine our results from previous chapters with an investigation of how the models can be explained using explainable AI (XAI) techniques, without being the deciding voice in a diagnosis. Major uncertainties remain in how and when the AI needs to be explainable. We propose an example of how explainable AI can help clinicians in the scope of ADNI data and deep learning models.



### 1) What do we want to explain?

In this project, we've analyzed tabular data in the form of cognitive tests, and image data in the form of MRIs. Tabular data can achieve high accuracies in the classification of ADNI patients using decision tree algorithms [30]–[32]. These algorithms are inherently more transparent than deep learning algorithms that we used for the image data due to their nature. Decision trees generate decision leaves which can be consulted when making statements about the classification tasks such as: “If the volume of the hippocampus is less than  $x$ , then the patient is in AD stage” [32]. For CNNs we have looked at Grad-CAM, but their explainability remains nontrivial for clinicians. Still, there are techniques, which can help clinicians make better decisions.

a) *Proposition.* To propose a workflow that helps clinicians in their diagnostic process, we need to look at model continuity, which expects the model to generate similar activation maps for similar input images.[33] This property, alongside the model's raw output in the form of probabilities for each possible diagnosis, is the basis for our explainability proposition for clinicians. We can compare each activation map that highlights important areas of the predicted diagnosis of one patient to an average of test subjects whose prediction was correct (True Positives). The resulting difference would be:

$$C_{diff} = C(x) - \frac{1}{n} \sum_{i=1}^n C(y_i) \quad (3)$$

where  $C(x)$  returns an activation map with an image as input and  $y$  is a stack of True Positives. The resulting difference ( $C_{diff}$ ) can be plotted and analyzed to observe where the patient differs the most from other True Positives. We can sum up all the values inside the difference  $C_{diff}$ :

$$sum_{diff} = \sum_{i=1}^n \sum_{j=1}^n |C_{diff_{i,j}}| \quad (4)$$

and compare it to the average sum of differences by taking the average of multiple True Positives. With minimal effort, variance and standard deviation can be calculated to estimate the average difference distribution and place the patient on the distribution curve.

These calculations can also be done on patches of images or regions of interest (ROI) that the clinician deems important. The clinician can zoom into an ROI and retrieve the information given by equations (3) and (4) only for that specific area, such as the hippocampus, or the entorhinal cortex. This gives the clinician the freedom to explore the prediction of the model by adding a layer of interactivity.

### 2) Further discussion

The methods discussed above highlight areas where a patient with ambiguous diagnosis differs from the average True Positive prediction of the machine learning model. The clinician also sees the probabilities for each class that are automatically given by the model's output. To offer even more insight into the prediction, the calculations can be extended to compare the patient to False Positives (patients who have been classified as AD but are not), and to other classes, giving insights into how the patient's activation maps differ from them.

Clinicians themselves can help contribute to better machine learning models by identifying areas in the activation map that influence the prediction. Such areas could be mistakenly considered insignificant in the diagnosis of Alzheimer's disease. As we have observed when training the models, some of these areas are the upper neck, the palate, as well as the olfactory system. This can be seen in Fig. 10 on the Grad-CAM images of the sagittal plane. These areas show higher activation over all three planes, but most prominently in the sagittal plane. Further research needs to be done to confirm that such areas are significant and not simply noise in the training data. (The regression of the upper neck area and the area around the nose might as well be natural due to aging.)

## VI. FUTURE OUTLOOK

The data provided by ADNI offers enormous opportunities for research on MCI and AD. In this work, we laid a foundation that could be of significant value for further projects. By building on the gained insights, we can continue to contribute to the fight against Alzheimer's disease – a complex, multifactorial pathology, which does not allow for any generalizations.

In terms of the Deep Learning aspect of the work, we propose to tackle not only the classification task but also the prediction of the disease risk. A natural next step would be to concatenate cognitive features with the anatomical ones, and closely examine subjects whose diagnosis deteriorated (NC to MCI or AD, MCI to AD).

We propose to investigate prognosis and classification using 3D CNN models, or voxel-based models. In addition to skull stripping, it seems useful to identify regions of interest that further reduce noise in MRI scans. The trained models and the resulting outcomes are useful for medical professionals. By summarizing the results in an accessible tool (e.g., a dashboard) and including explainable AI through interactive visualizations, we can present the findings of a patient in a coherent way, and thus support the clinicians in the decision-making process.

## APPENDIX

- A.1 Loadings Matrix
- A.2 Deep Learning Result Table
- A.3 Grad-CAM Examples Transversal Plane
- A.4 Grad-CAM Examples Frontal Plane
- A.5 Grad-CAM Examples Sagittal Plane

## ACKNOWLEDGMENT

In accordance with the Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy, we wish to state that the data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). While the investigators within the ADNI contributed the data, they did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

Led by Principal Investigator Michael W. Weiner, MD, the ADNI was launched in 2003 as a public-private partnership. Its primary goal has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), biological markers, and clinical and neuropsychological assessment can be combined to measure



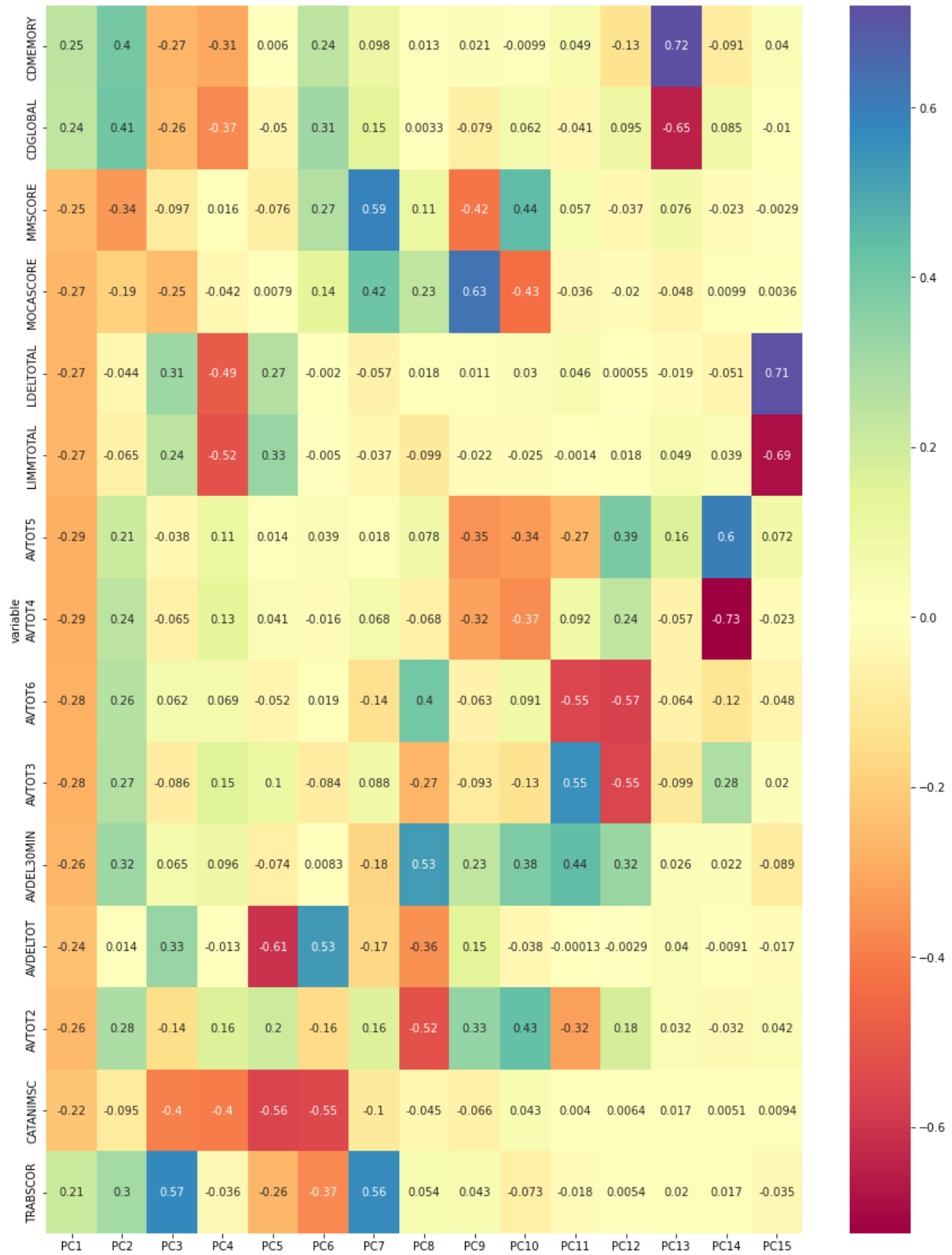
the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from many others. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

## REFERENCES

- [1] R. Brookmeyer *u. a.*, «National estimates of the prevalence of Alzheimer's disease in the United States», *Alzheimers Dement.*, Bd. 7, S. 61–73, 2011, doi: 10.1016/j.jalz.2010.11.007.
- [2] «Alzheimer's Disease Fact Sheet». National Institute on Aging, 2019. [Online]. Verfügbar unter: <https://www.nia.nih.gov/health/alzheimers-disease-fact-sheet>
- [3] O. of the Commissioner, «FDA Grants Accelerated Approval for Alzheimer's Drug». 2021.
- [4] C. for D. E. Research, «Aducanumab (marketed as Aduhelm) Information», *FDA*, 2021.
- [5] A. R. Preston und H. Eichenbaum, «Interplay of Hippocampus and Prefrontal Cortex in Memory», *Curr. Biol.*, Bd. 23, S. R764–R773, 2013, doi: 10.1016/j.cub.2013.05.041.
- [6] S. Villegas, A. Roda, G. Serra-Mir, L. Montoliu-Gaya, und L. Tiessler, «Amyloid-beta peptide and tau protein crosstalk in Alzheimer's disease», *Neural Regen. Res.*, Bd. 17, S. 1666, 2022, doi: 10.4103/1673-5374.332127.
- [7] L. R. Trambaiolli, R. Cassani, D. M. A. Mehler, und T. H. Falk, «Neurofeedback and the Aging Brain: A Systematic Review of Training Protocols for Dementia and Mild Cognitive Impairment», *Front. Aging Neurosci.*, Bd. 13, 2021, doi: 10.3389/fnagi.2021.682683.
- [8] R. Brookmeyer, M. M. Corrada, F. C. Curriero, und C. Kawas, «Survival Following a Diagnosis of Alzheimer Disease», *Arch. Neurol.*, Bd. 59, S. 1764, Nov. 2002, doi: 10.1001/archneur.59.11.1764.
- [9] «ADNI | Study Design». Alzheimer's Disease Neuroimaging Initiative, 2017. [Online]. Verfügbar unter: <https://adni.loni.usc.edu/study-design/>
- [10] «ADNI General Procedures Manual». Alzheimer's Disease Neuroimaging Initiative, S. 166, 2006.
- [11] A. L. Byers und K. Yaffe, «Depression and risk of developing dementia», *Nat. Rev. Neurol.*, Bd. 7, S. 323–331, 2011, doi: 10.1038/nrneurol.2011.60.
- [12] «ADNI Standardized MRI Data Sets». <https://adni.loni.usc.edu/methods/mri-tool/standardized-mri-data-sets/> (zugegriffen 28. März 2022).
- [13] I. T. Jolliffe und J. Cadima, «Principal component analysis: a review and recent developments», *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, Bd. 374, S. 20150202, 2016, doi: 10.1098/rsta.2015.0202.
- [14] K. Arvai, «K-Means Clustering in Python: A Practical Guide – Real Python». Real Python. [Online]. Verfügbar unter: <https://realpython.com/k-means-clustering-python/>
- [15] X. Sun *u. a.*, «Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions», *Biomed. Eng. OnLine*, Bd. 14, Nr. 1, S. 73, Juli 2015, doi: 10.1186/s12938-015-0064-y.
- [16] M. Sethi, S. Ahuja, S. Rani, D. Koundal, A. Zaguia, und W. Enbeyle, «An Exploration: Alzheimer's Disease Classification Based on Convolutional Neural Network», *BioMed Res. Int.*, Bd. 2022, S. e8739960, Jan. 2022, doi: 10.1155/2022/8739960.
- [17] «3.3. Metrics and scoring: quantifying the quality of predictions», *scikit-learn*. [https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html) (zugegriffen 22. Juni 2022).
- [18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, und P. Dollár, «Focal Loss for Dense Object Detection», arXiv, arXiv:1708.02002, Feb. 2018, doi: 10.48550/arXiv.1708.02002.
- [19] L. Luo, Y. Xiong, Y. Liu, und X. Sun, «ADAPTIVE GRADIENT METHODS WITH DYNAMIC BOUND OF LEARNING RATE», S. 21, 2019.
- [20] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, und M. Hoffmann, «SynthStrip: Skull-Stripping for Any Brain Image», arXiv, 18. März 2022. Zugegriffen: 22. Juni 2022. [Online]. Verfügbar unter: <http://arxiv.org/abs/2203.09974>
- [21] S. I. Shiekh, S. L. Cadogan, L.-Y. Lin, R. Mathur, L. Smeeth, und C. Warren-Gash, «Ethnic Differences in Dementia Risk: A Systematic Review and Meta-Analysis», *J. Alzheimers Dis.*, Bd. 80, S. 337–355, 2021, doi: 10.3233/jad-201209.
- [22] K. A. Matthews *u. a.*, «Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years», *Alzheimers Dement.*, Bd. 15, S. 17–24, 2019, doi: 10.1016/j.jalz.2018.06.3063.
- [23] «ADNI General Procedures Manual». Alzheimer's Disease Neuroimaging Initiative, S. 166, 2006. [Online]. Verfügbar unter: [https://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI\\_GeneralProceduresManual.pdf](https://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf)
- [24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, und D. Batra, «Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization», *Int. J. Comput. Vis.*, Bd. 128, Nr. 2, S. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [25] G. Csukly *u. a.*, «The Differentiation of Amnesic Type MCI from the Non-Amnesic Types by Structural MRI», *Front. Aging Neurosci.*, Bd. 8, März 2016, doi: 10.3389/fnagi.2016.00052.
- [26] Q. Meng, M.-S. Lin, und I.-S. Tzeng, «Relationship Between Exercise and Alzheimer's Disease: A Narrative Literature Review», *Front. Neurosci.*, Bd. 14, März 2020, doi: 10.3389/fnins.2020.00131.
- [27] «Article 22 EU General Data Protection Regulation (EU-GDPR)», [www.privacy-regulation.eu](http://www.privacy-regulation.eu), 2018. <https://www.privacy-regulation.eu/en/article-22-automated-individual-decision-making-including-profiling-GDPR.htm>
- [28] J. Ordish und A. Hall, «A right to explanation? Black box medicine», PHG Foundation, 2019. 1. Januar 2022. [Online]. Verfügbar unter: <https://www.phgfoundation.org/media/487/download/A-right-to-explanation-sept-2019.pdf?v=1&inline=1>
- [29] HHS, «Individuals' Right under HIPAA to Access their Health Information», *HHS.gov*, 2016. <https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/access/index.html>
- [30] B. Bogdanovic, T. Eftimov, und M. Simjanoska, «In-depth insights into Alzheimer's disease by using explainable machine learning approach», *Sci. Rep.*, Bd. 12, Nr. 1, 2022, doi: 10.1038/s41598-022-10202-2.
- [31] S. El-Sappagh, J. M. Alonso, S. M. R. Islam, A. M. Sultan, und K. S. Kwak, «A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease», *Sci. Rep.*, Bd. 11, Nr. 1, 2021, doi: 10.1038/s41598-021-82098-3.
- [32] K. Achilleos, S. Leandrou, N. Prentzas, P. A. Kyriacou, A. Kakas, und C. S. Pattichis, «Extracting Explainable Assessments of Alzheimer's disease via Machine Learning on brain MRI imaging data», *2020 IEEE 20th Int. Conf. Bioinforma. Bioeng. BIBE*, S. 4–5, 2020, doi: 10.1109/bibe50027.2020.00175.
- [33] G. Montavon, W. Samek, und K.-R. Müller, «Methods for interpreting and understanding deep neural networks», *Digit. Signal Process.*, Bd. 73, S. 10–11, 2018, doi: 10.1016/j.dsp.2017.10.011.

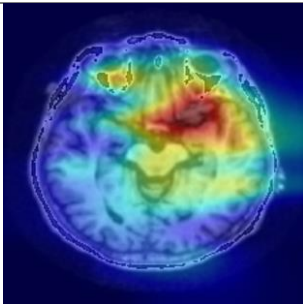
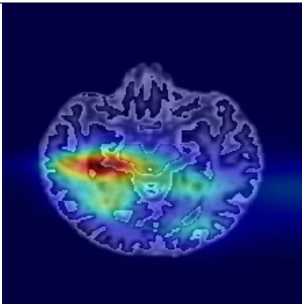
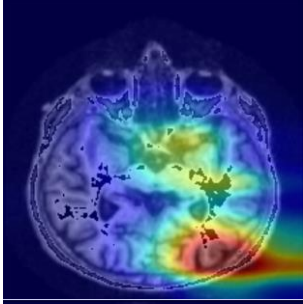
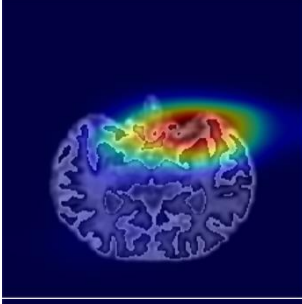
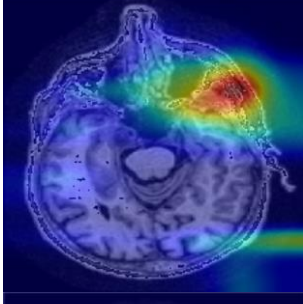
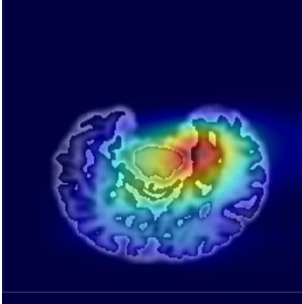
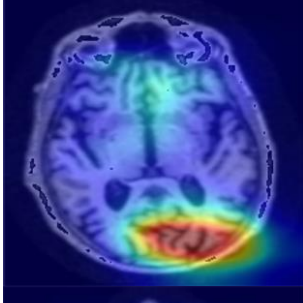
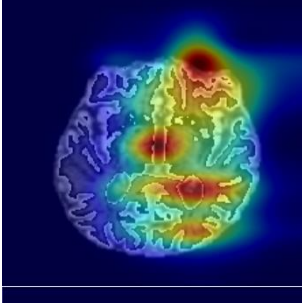
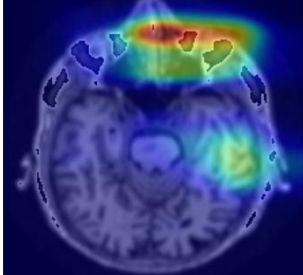
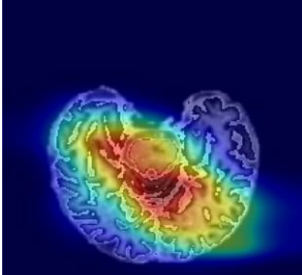
A.1 LOADINGS MATRIX



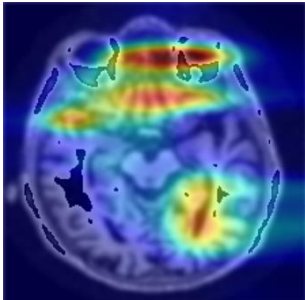
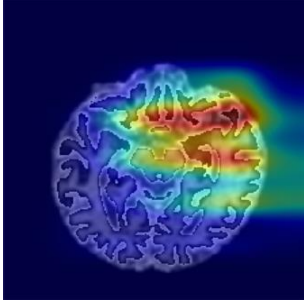
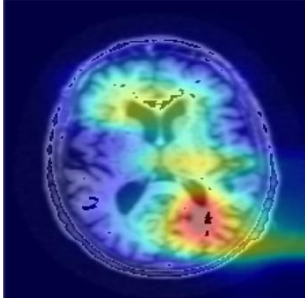
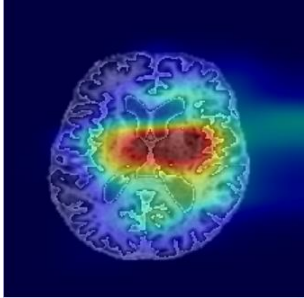
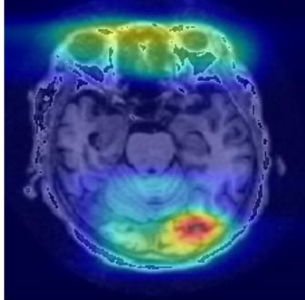
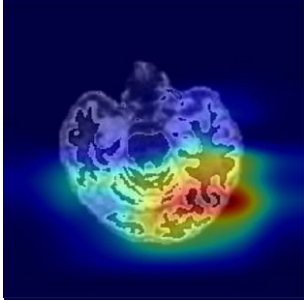
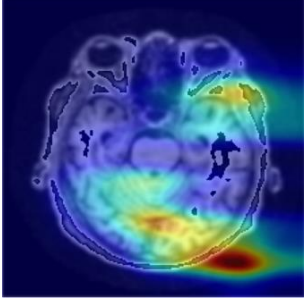
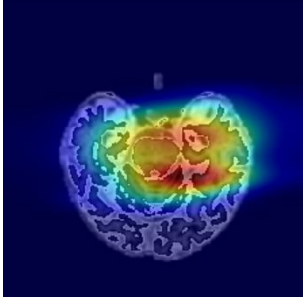
A.2 DEEP LEARNING RESULT TABLE

<i>Net</i>	<i>Train Accuracy</i>	<i>Train Loss</i>	<i>Test Accuracy</i>	<i>Test Loss</i>	<i>Data</i>	<i>Plane</i>	<i>Preprocessing</i>
<i>ResNet50</i>	48.11	0.12944962	50.00	0.134286259	MultiChannel	Triplan	Blurred
<i>ResNet50</i>	93.62	0.02364571	39.44	0.278103041	SingleSlice	Stacked	Blurred
<i>ResNet18</i>	93.50	0.01658179	45.31	0.491821622	SingleSlice	C3	Blurred
<i>ResNet18</i>	99.94	0.00055186	38.26	0.272357882	SingleSlice	C3	Blurred
<i>ResNet18</i>	100.00	0.00008648	35.68	0.296237832	SingleSlice	C3	Blurred
<i>ResNet18</i>	99.94	0.00044859	40.61	0.239650529	SingleSlice	C2	Blurred
<i>ResNet18</i>	100.00	0.00002255	43.43	0.211926839	SingleSlice	C2	Blurred
<i>ResNet18</i>	100.00	0.00014375	40.85	0.242737594	SingleSlice	C2	Blurred
<i>ResNet18</i>	99.94	0.00013100	39.44	0.195603388	SingleSlice	C1	Blurred
<i>ResNet18</i>	100.00	0.00000448	41.78	0.186810427	SingleSlice	C1	Blurred
<i>ResNet18</i>	99.94	0.00011633	43.19	0.212359639	SingleSlice	C1	Blurred
<i>DenseNet121</i>	100.00	0.00000278	42.25	0.058953611	SingleSlice	C3	MonaiAugment
<i>DenseNet121</i>	99.82	0.00013128	33.57	0.062065957	SingleSlice	C3	MonaiAugment
<i>DenseNet121</i>	99.88	0.00007097	42.02	0.059973445	SingleSlice	C3	MonaiAugment
<i>DenseNet121</i>	100.00	0.00000504	37.79	0.041306656	SingleSlice	C3	Blurred
<i>ResNet18</i>	100.00	0.00000050	35.21	0.053809441	SingleSlice	C3	MonaiAugment
<i>ResNet18</i>	100.00	0.00000085	42.25	0.037373781	SingleSlice	C3	MonaiAugment
<i>ResNet10</i>	100.00	0.00000205	43.43	0.030805032	SingleSlice	C3	MonaiAugment
<i>ResNet10</i>	100.00	0.00000148	37.79	0.032846155	SingleSlice	C3	MonaiAugment
<i>ResNet10</i>	100.00	0.00000161	40.38	0.035635391	SingleSlice	C3	MonaiAugment


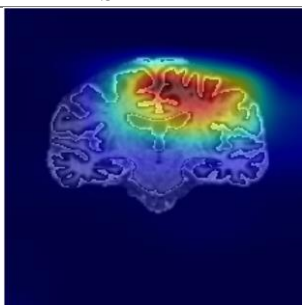

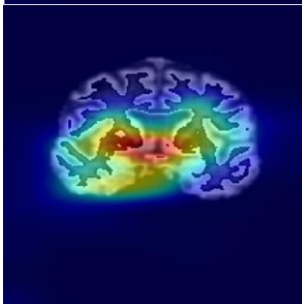

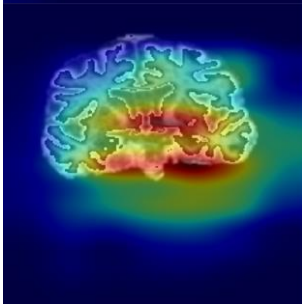
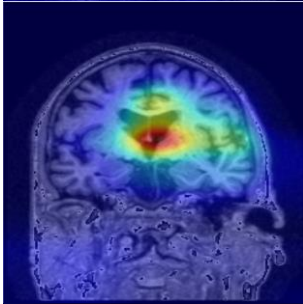
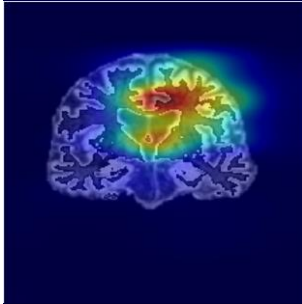
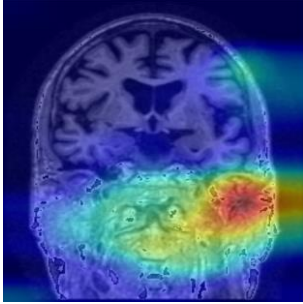
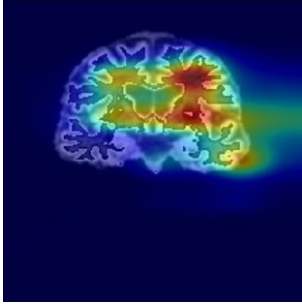
### A.3 GRAD-CAM EXAMPLES TRANSVERSAL PLANE

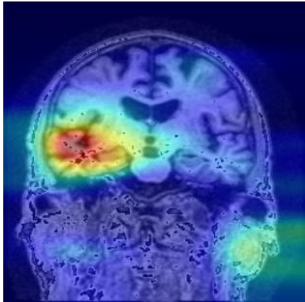
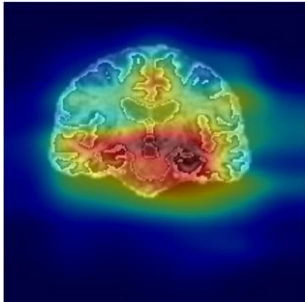
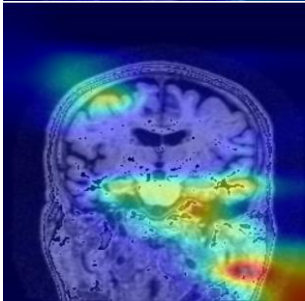
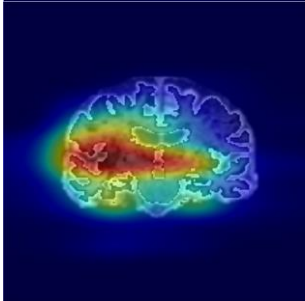
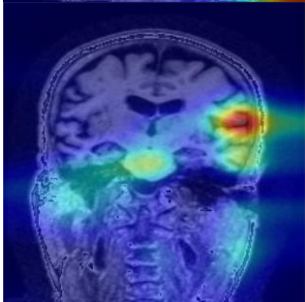
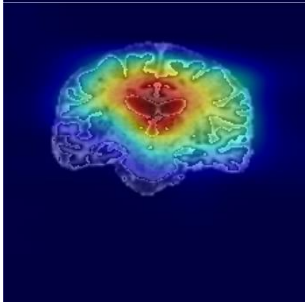
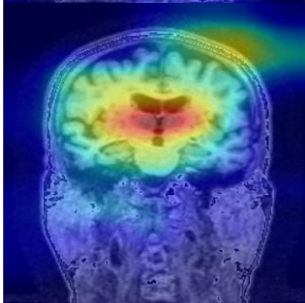
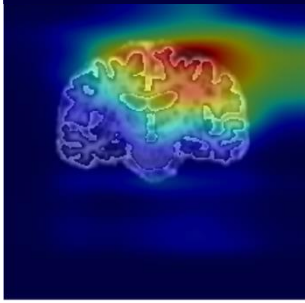
PATIENT	CLASS	PREDICTION	ACTIVATION RAW	PREDICTION STRIPPED	ACTIVATION SKULL STRIPPED
S22923	MCI	MCI		MCI	
S51307	MCI	MCI		MCI	
S35178	MCI	AD		CN	
S20274	CN	MCI		CN	
S32721	CN	MCI		MCI	



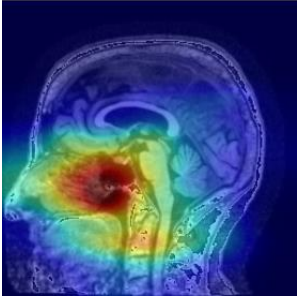
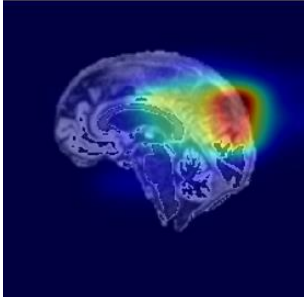
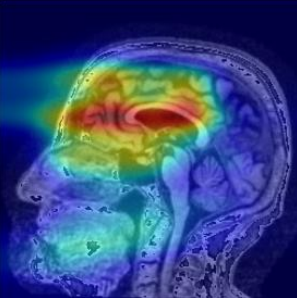
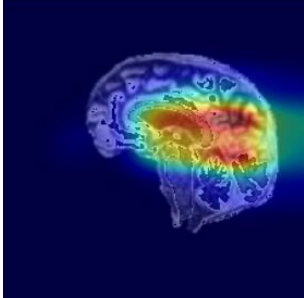
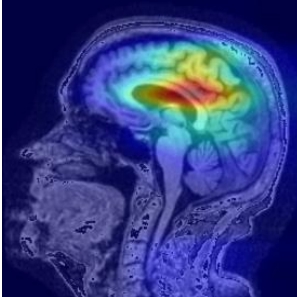
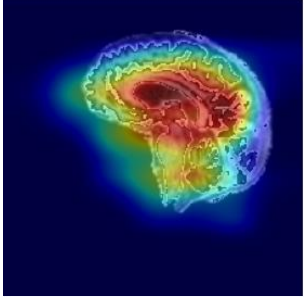

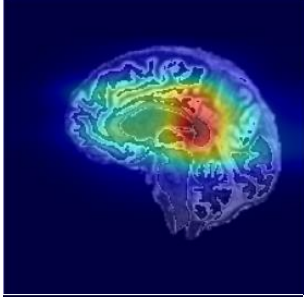
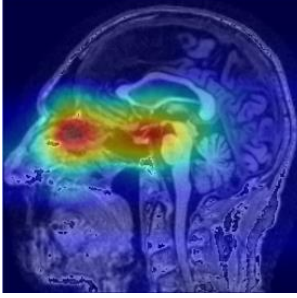
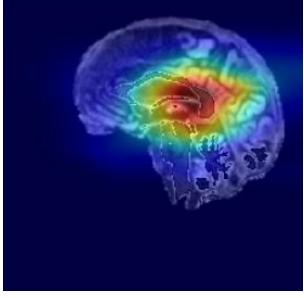
S50194	CN	MCI		MCI	
S18373	AD	MCI		MCI	
S40962	AD	MCI		CN	
S29153	AD	AD		AD	

# A.4 GRAD-CAM EXAMPLES FRONTAL PLANE


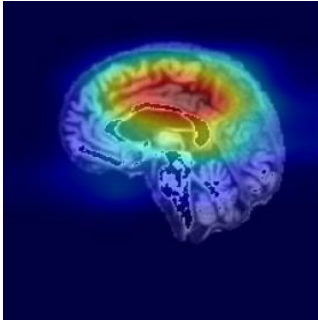
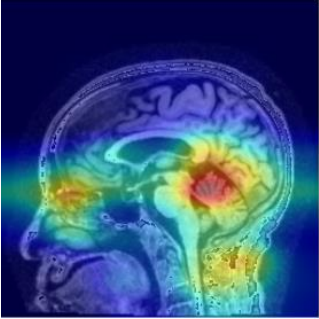
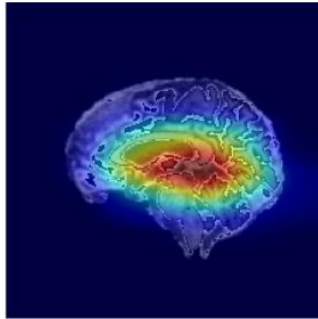
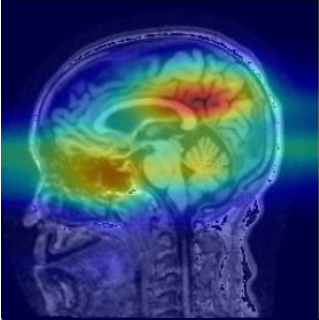
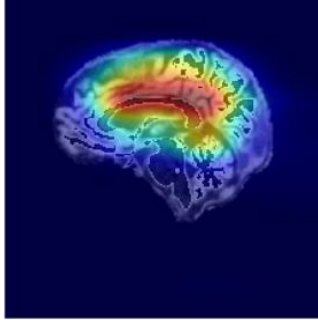

PATIENT	CLASS	PREDICTION	ACTIVATION RAW	PREDICTION STRIPPED	ACTIVATION SKULL STRIPPED
S22923	MCI	MCI		MCI	
S51307	MCI	MCI		MCI	
S35178	MCI	MCI		CN	
S20274	CN	MCI		MCI	
S32721	CN	MCI		MCI	

S50194	CN	MCI		AD	
S18373	AD	MCI		AD	
S40962	AD	MCI		MCI	
S29153	AD	MCI		AD	

# A.5 GRAD-CAM EXAMPLES TRANSVERSAL PLANE

PATIENT	CLASS	PREDICTION	ACTIVATION RAW	PREDICTION STRIPPED	ACTIVATION SKULL STRIPPED
S22923	MCI	MCI		MCI	
S51307	MCI	MCI		MCI	
S35178	MCI	MCI		MCI	
S20274	CN	MCI		MCI	
S32721	CN	MCI		MCI	



S50194	CN	MCI		MCI	
S18373	AD	MCI		MCI	
S40962	AD	MCI		MCI	
S29153	AD	MCI		MCI	