Prediction of Alzheimer's Disease from Magnetic Resonance Imaging using a Convolutional Neural Network

Kevin de Silva
University College London
Holger Kunz (h.kunz@ucl.ac.uk)
University College London

Research Article

Keywords:

Posted Date: April 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1419743/v1

Abstract

Objectives: The primary goal of this study is to examine if a convolutional neural network (CNN) can be applied as a diagnostic tool predicting Alzheimer's Disease (AD) from magnetic resonance imaging (MRI) using the MIRIAD-dataset (Minimal Interval Resonance Imaging in Alzheimer's Disease).

Methods: The MIRIAD-dataset contains patients represented by a set of MRI scans of the brain and further diagnostic data. Hyperparameter and configurations of CNNs were optimized to determine the best-performing model. The CNN was implemented in Python with the deep learning library 'Keras' using Linux/Ubuntu as the operating system.

Results: This study obtained the following best performance metrics on predicting Alzheimer's Disease from MRI: Matthew's Correlation Coefficient (MCC) of 0.77; accuracy of 0.89; F1-score of 0.89; AUC of 0.92. The computational time for training of a CNN takes less than 30 seconds with a GPU (graphics processing unit).

Conclusions: The study suggests that an axial MRI scan can be used to diagnose if a patient has Alzheimer's Disease with a performance of 0.92 AUC.

1 Introduction

1.1 Alzheimer's Diseases

Alzheimer's Disease (AD) is associated with progressive accumulation of abnormal proteins in the brain, which leads to progressive synaptic, neuronal, and axonal damage [1]. ICD-11 (eleventh revision of the International Classification of Diseases) from the WHO codifies Alzheimer with 6D80* and 8A20* as a disorder with neurocognitive impairment as a major feature [2]. Clinical symptoms include loss of memory, linguistic and cognitive degradation, personality and mood changes [3]. The number of people with AD worldwide is estimated at 50 million in 2017, growing to 132 million by 2050, while the total cost associated with AD worldwide as of 2018 is estimated at 1 trillion dollars [3]. Although these costs and prevalence numbers appear high, they may represent a substantial underestimate of the true figures since undiagnosed AD can be as high as 80% of all cases worldwide [3].

Although currently, no drugs can cure AD early diagnosis and treatment of AD has substantial benefits, both in terms of personal wellbeing and societal cost [3]. A class of drugs, cholinesterase inhibitors, are effective at slowing down the progression of AD [3]. Given the advantages of early-stage diagnosis of Alzheimer's disease, any methodology that improves early detection is beneficial. There is no specific biomarker for AD and diagnosis relies on a range of tests which include one or more of the following: cognitive assessment tests, blood tests, Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET)[4]. Of particular relevance to this study are MRI and the cognitive assessment test – Mini-Mental State Examination (MMSE). MRI scans show atrophy of certain brain regions that are indicative of

Alzheimer's [5][6]. MMSE is a quick, inexpensive test, scoring from 0–30, where higher scores are indicative of better cognitive functioning [7].

1.2 Convolutional Neural Networks to detect Alzheimer

Studies have been using CNNs to diagnose Alzheimer's [8–19, 19–25] using data from the ADNI or OASIS-dataset. The Alzheimer's Disease Neuroimaging Initiative (ADNI) has 1455 participants with five diagnosis groups [26]. The OASIS dataset is composed of 193 participants aged 62 years or more [26]. The primary goal of this study is to examine if convolutional neural networks can also be applied as a diagnostic tool using the MIRIAD-dataset (Minimal Interval Resonance Imaging in Alzheimer's Disease).

2 Methods

2.1 Data and material

MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) is a series of longitudinal volumetric T1-MRI scans of mild-moderate Alzheimer's subjects and controls [27]. An overview of the MIRIAD demographics and publications is published in Malone et [27]. The dataset consists of scans with the same scanner with accompanying information on gender, age, and Mini-Mental State Examination (MMSE) scores [27]. The data used in this study classifies subjects as AD if they have an MMSE score of 26 or under at baseline while a healthy control (HC) has an MMSE of 27 or above [27]. This is also the cutoff point to describe the class label for each feature vector. Each patient has multiple MRI scans from different time points. Many scans were collected of each participant at intervals from two weeks to two years, the study was designed to investigate the feasibility of using MRI as an outcome measure for clinical trials of Alzheimer's treatments [27]. Table 1 shows the demographics of the included patients.

Table 1
MIRIAD demographic information

	Alzheimer's Disease (N = 46, Total MRI-scans = 465)	**Healthy Controls (N = 23, Total MRI-scans: 243)**
Age at study entry	69.4 ± 7.1	69.7 ± 7.2
Men	41%	52%
Mean (SD) baseline MMSE	19.2 ± 4	29.4 ± 0.8

Each scan is provided in *NIfTI*-format (Neuroimaging Informatics Technology Initiative) [28]. It is an open file format for volumetric images with a size of 256 x 256 x 124. Figure 1 shows a sample of the MRI dataset. An axial, sagittal, and coronal view is displayed. The raw dataset still contains bone structures. The bone structures are not relevant for the diagnosis of Alzheimer's and are getting removed in the pre-processing.

2.2 Feature engineering and pre-processing

Pre-processing is an important step to prepare the dataset for the following training of the classification algorithm. The MIRIAD dataset is pre-processed by applying spatial normalization, bias correction, and grey matter segmentation. Spatial normalization is the process of mapping images from different scans onto a single template. There are two steps to this: linear transformation (e.g. translation, rotation, shear) and non-linear transformation (e.g. warping). This results in all images referencing the same coordinate space [29] and should adjust, for example, for different subject positioning when the MRI was recorded.

The ratio of MRI scans of AD subjects to healthy controls is approximately 2:1. To mitigate this imbalance, data augmentation is performed by creating copies and flipping them. This results in almost the same number of instances labeled for AD and non-AD subjects. This can also be considered as a specific type of oversampling in medical imaging.

Finally, grey matter segmentation is performed and grey matter is extracted from the raw data. This excludes features that are unlikely to be discriminative in the classification task e.g. skull tissue (skull-stripping). The Python 'Nipype' library interface is used, allowing all processing to be done in Python [30]. An axial MRI scan of the central part of the brain for each patient was used as an input for the following classification algorithm.

2.3 Convolutional Neural Network

Convolutional Neural networks are a specialized kind of neural network for processing data that has a grid-like topology [31]. A CNN consists of several layers: convolutional, pooling, and fully connected layers. Each convolutional layer consists of a certain number of trainable parametric filters. Each convolutional layer is typically followed by a pooling layer which reduces the feature space. Finally, the data is passed to one or more fully connected layers and the predicted output is produced. A further description of the basic ingredients of a convolutional neural network can be derived from a textbook in deep learning [31] and are not further explained.

The applied CNN to distinguish between Alzheimer's and non-Alzheimer patients is used as a classification algorithm. Classification is to learn a mapping from inputs x to output y, where $y \in \{1,.., C\}$ with C being the number of classes [32]. If $C = 2$, this is called binary classification [32]. In our study, a binary classification task is performed to distinguish between patients with Alzheimer's and patients who do not show signs of Alzheimer's.

Loss function and optimization

As a loss function for the convolutional neural network, the binary cross-entropy was chosen [33][34]. Every training epoch of the CNN has the aim to reduce the loss function (binary cross-entropy). RMSprop is a gradient-based optimization technique used in training neural networks. It has also been applied in deep learning for MR-images by Medina et al [35].

Convolutional filter size and Max-pooling.

For a two-dimensional image *I* as our input (from an MRI scan), a two-dimensional kernel *K* can be used. In this study, the convolutional filter size was set to (3,3). In convolutional network terminology, the output is referred to as a feature map [31]. The convolutional operation can be described as follows [31]:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n)$$

1

A pooling function replaces the output with a summary statistic. For example, the max-pooling operation reports the maximum output within an area [31]. The Max-pooling filter size of the final configuration after hyperparameter tuning was set to (2,2).

Dropout layer

Dropout provides a computationally inexpensive method for regularizing a model and to prevent overfitting [31][36]. During training, units get randomly get removed [36]. The randomly selected unit is removed from the network, along with all its incoming and outgoing connections [36]. It prevents overfitting and provides a way of approximately combining exponentially many different neural network architectures efficiently [36]. Dropout introduces an extra hyperparameter—the probability of retaining a unit [36]. A value of $p = 1$ implies no dropout, and low values of p mean more dropout [36]. The dropout rate was set to 0.4 in our configuration to avoid overfitting.

Activation function

Neurons in the activation map pass through a non-linear function [37]. There are different activation functions. For example, the sigmoid function, the rectified linear unit (ReLU), and the leaky rectified linear unit (leaky ReLU). The logistic sigmoid function can be defined as following [20]:

$$f_{sigmoid}(x) = \frac{1}{1 + \exp(-x)}$$

2

Another activation-function is the ReLu-function [20]:

$$f_{ReLu}(x) = \max(0, x) = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases}$$

3

Whenever the activation values are zero, the ReLu-function cannot learn in a gradient-based learning method [20]. Therefore, a leaky ReLu-function can be used.

$$f_{LeakyRelu}(x) = \begin{cases} x, x \geq 0 \\ \alpha x, x < 0 \end{cases}$$

4

In our study, the parameter alpha was set to 0.1 and a leaky rectified linear unit was used.

Regularization

To prevent overfitting a regularization method can be used to train the neural network. L1-Regularization is also known as Lasso-Regularization [38]. L2-Regularization, also known as Ridge Regularization [38]. L1 + L2 Regularization is also known as Elastic Net Regularization [38]. A small value for the regularization parameters for $L1 = 0.001$ and $L2 = 0.002$ has been added to prevent overfitting.

Table 2 contains the final settings of the CNN model. The number of layers and convolutional filters per layer were varied. The hyperparameter tuning is used either with a 3-layer or 4-layer setting.

Table 2
Configuration of the applied CNN

Setting/Parameter	Values in Keras
Loss Function	binary_crossentropy
Optimiser Function	RMSprop(lr = 0.001)
Convolutional filter (kernel) size	(3, 3)
Max-pooling filter size	(2, 2)
Activation function for all layers	Leaky ReLU (alpha = 0.1)
Weight regularisation added to all models to mitigate overfitting	L1 = 0.001 L2 = 0.002
Dropout layer added to all models to mitigate overfitting	0.4
Batch size	100

2.4 Performance metrics

The evaluation of model performance is an essential step in understanding and developing a machine learning algorithm. Definitions of conventional performance metrics such as accuracy, precision, specificity, recall, and F1-score are not further described. The definition can be obtained from textbooks in machine learning such as Goodfellow et al [31], Murphy [32], and Hastie et al [38]. This study used as an additional metric Matthew's Correlation Coefficient (MCC) [39]. The following abbreviations have been used TP = True Positives, TN = True Negatives, FP = False Positive, FN = False Negative.

The MCC is defined according to [39] as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

5

The MCC metric is more balanced than metrics like accuracy and F1-score because its score is high only if the classifier is good on both positive and negative predictions [40]. The MCC is calibrated so that it ranges from − 1 to + 1. A value of 0 indicates a result close to chance, the closer to + 1 the score is, the better the result [40]. Receiver Operating Characteristic (ROC) curves have also been plotted for the best outcome.

The data are split into training, validation, and test-dataset. Approximately, 20% of each category is randomly allocated to the validation dataset and 10% to the validation to the validation dataset. The best configuration of the CNN was determined with the highest MCC on unseen medical images of a set of AD and non-AD patients. The training of the CNN used 20 epochs per instance. Table 3 shows the split of the data for the binary classification with a CNN.

Table 3
data augmentation, training, validation, test

Class Label	Number of MRI-scans	Total slices after data augmentation	Slices in training split	Slices in validation split	Slices in test split
AD	465	465	326	39	100
non-AD	243	486	342	42	102

2.5 Implementation

The implementation used as hardware an Intel Core i9 with 64 Gb of memory, hard disk: Samsung 1 Terabyte SSD, GPU: NVIDIA GE Force RTX 2080 TI. The implementation used as software: Statistical Parametric software v. 12 (SPM 12), Matlab v. 2019a, Python v. 3.7.3, Keras v. 2.2.4, operating system Ubuntu 18.04.02.

The authors confirm that all methods were carried out in accordance with relevant guidelines and regulations.

3 Results

3.1 Pre-processed medical images

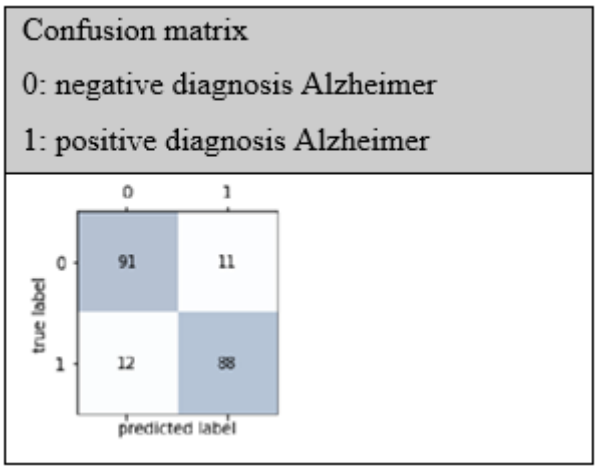
Figure 2 shows the axial, sagittal, and coronal views as well as a 3D surface rendering of a typical pre-processed scan. It can be seen that many extraneous features (for example the skull and bone structures)

have been removed.

3.2 Results from the optimized configuration of the CNN

Table 4 shows for the best model with optimized MCC the associated confusion metric (TN=91, FN=12, FP=11, TP=88

Table 4: Performance for the optimized CNN-model



The obtained best performance metric for MCC=0.77.

$$MCC = \frac{88 \times 91 - 11 \times 12}{\sqrt{(88+11)(88+12)(91+11)(91+12)}} = 0.77$$

Accuracy=0.89=(88+91)/(91+12+11+88), Precision=0.89=88/(11+88),Specificity=0.89=91/(91+12), Recall=0.88=88/(12+88),F1= 0.88=88/(88+0.5(11+12)), AUC=0.92 (obtained from Python).

Figure 3 shows AUC scores for different configurations and hyperparameters of the CNN. The final architecture for the model was guided by the best MCC-score of 0.77 in 3-layer CNN with 64 convolutional filters in each layer, represented as (64, 64, 64) and an associated AUC of 0.92.

The computational time for the training using a CNN takes less than 30 seconds with a GPU (graphics processing unit).

4 Discussion And Conclusion

This study shows that convolutional neural networks for pattern recognition of neurological conditions such as Alzheimer's can be used. A CNN was proposed to distinguish between patients having Alzheimer's Disease (AD) and patients who have not been diagnosed with AD. Medical images of the brain have been used as input for CNN. The CNN and the number of layers and convolutional filters per layer were varied and optimized based on Matthew’s Correlation Coefficient (MCC). This study obtained

the following performance metrics on predicting Alzheimer's Disease from MRI scans of the brain; MCC: 0.77, accuracy: 0.89, F1: 0.89, AUC: 0.92. An AUC > 0.90 is seen as an excellent diagnostic test [41]. A potential interpretation of AUC-values is 1.0 for a perfect test, 0.9–0.99 for an excellent test, 0.8–0.89 for a good test, 0.7–0.79 for a fair test, 0.51–0.69 for a poor test, and 0.5 of no value [41]. This suggests that potentially a diagnostic tool could be developed based on the provided methodology.

Minimizing the risk of overfitting was done with three techniques: Firstly, the dataset was randomly split into a training set, validation set, and test set. The general assumption is that the instances are randomly selected. This ensures that the trained CNN will be tested on previously unseen instances. Secondly, regularization parameters were set to non-zero values. This ensures that the loss function considers regularization. This is a common method in machine learning to avoid overfitting. Thirdly, a specified dropout rate is a safeguard that the neural network at the dense layer (fully connected deep neural network) does not overlearn presented instances. Random units of the dense layer get removed [36]. It can empirically be shown that a neural network that makes use of a dropout rate reduced the risk of overfitting [36]. These three methods 1) random split, 2) regularization, and 3) dropout helped to minimize the potential risk of overfitting and generally avoids that a complex function is perfectly fitted to the provided training set.

The underlying problem is a binary classification problem to diagnose Alzheimer vs. non-Alzheimer. The MCC-score is a more reliable statistical rate that produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false positives) [39]. For this reason, the optimization of our study used the MCC-score. The performance metrics such as F1-score or AUC are a by-product of this process. The MCC-score is useful for imbalanced datasets where the AUC might be less useful [40]. This imbalance was reduced by using methods of augmentation which can be considered as a type of oversampling in machine learning for medical images. The confusion matrix shows that overall few instances have been misclassified as false negatives or false positives.

Analysis of the literature identifies the usage of convolutional neural networks for AD-classification from MRI [24]. A direct comparison of the performance is limited as the studies used different datasets such as the ADNI or OASIS-dataset for Alzheimer's disease. For example, different demographics, origins, and sample sizes have an impact on the performance metric. Additionally, the hyperparameters were tuned on different performance metrics and the CNNs used different configurations regarding kernels and the number of layers. Another factor is different cross-validation methods across the studies. Since random seed values are used to split data in different folds or to split into training sets and test set a meaningful comparison is limited, too. The ADNI-dataset was used by Aderghal et al [8], Taqi [13], Cheng [21], Lian et [11], Farooq et al. [22], Senanayake [12], Gunawardena [23], Hosseini et al. [24], Korolev [10], Bäckström et al [9], Folego et al [18], Feng at al [17], Wu [15]. The OASIS-dataset has been used by Islam et al [25], Wang et al [20], Hon et al [42], Ebrahimi and Luo [43]. The models in Hosseini et al [24] and Bäckström et al [9] used accuracy as a performance metric for optimization. Accuracy, sensitivity, specificity have been

used in Farooq et al [22] and Wang et al [20]. Also, a multi-class classification process was used in Islam et al [25] as opposed to binary classification in this study.

A key question towards a potential clinical deployment is the reliability of such as clinical decision support system. Even though the AUC is above 0.90 it is crucial to indicate that the performance metrics must be seen in the context of the specific clinical application. The therapeutic consequences for false positive and false negative subjects must be carefully considered. As a rule of thumb, diagnostic tools having an AUC > 0.9 could potentially be candidates for a clinical decision support system. However, a decision for a potential deployment cannot be based on a fixed threshold but also need an extended qualitative study that considers the expert opinions of clinicians to provide further insights whether such a decision support system is fit for purpose. The practical advantage of the developed CNN lies in the fact that one axial scan provides sufficiently enough information to achieve high performance. One immediate potential usage of a deployed system could be in a low-resource setting or where clinical consultants are not readily available.

Declarations

*Ethical approval and consent to participa*te: Ethical approval for the data (and subsequently its release) was received from the local MIRIAD research ethics committee, and informed written consent obtained from all participants (see Malone *et al* [27]).

Consent for publication: Not applicable.

Availability of data and materials: The MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's Disease) dataset is publicly available. Data are here made publicly available as a common resource for researchers to develop, validate and compare techniques, particularly for measurement of longitudinal volume change in serially acquired MR (see [27]). By registering and agreeing to the data use agreement the data can be downloaded. Datasets are available in the MIRIAD database for research, which is accessible after registration from a public repository using the following URL: https://www.ucl.ac.uk/drc/research/research-methods/minimal-interval-resonance-imaging-alzheimers-disease-miriad.

Competing interest: No competing interests declared.

Funding: The original data collection was funded through an unrestricted educational grant from GlaxoSmithKline (Grant 6GKC) and funding from the UK Alzheimer's Society (Grant RF116) and the Medical Research Council (to Professor Fox). The Dementia Research Centre is an Alzheimer's Research UK (ARUK) Coordinating Centre [27]. The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust [grant number 091593/Z/10/Z] (see [27]). The implementation of the algorithms of this work makes use of the provided MIRIAD-data and has been conducted as a postgraduate dissertation without dedicated funding.

Authors' contribution: The first author (KdS) conducted the implementation of the algorithms and prepared the draft of the manuscript. The co-author (HK) in his role as senior researcher provided academic guidance, revised the manuscript, and verified the scientific robustness of the applied methods. The authors confirm that all methods were carried out in accordance with relevant guidelines and regulations. MIRIAD investigators did not participate in analysis or writing of this report.

Acknowledgement: The authors would like to acknowledge the contribution of Dr Ian Malone from the Dementia Research Centre (DRC) at UCL for the provision of the MIRIAD-dataset.

Author information: The first author (KdS) is a postgraduate student at the Institute of Health Informatics, UCL. The senior co-author (HK) is a lecturer at the Institute of Health Informatics and director of the MSc Health Data Science.

References

1. Ramani A, Jensen JH, Helpern JA. Quantitative MR imaging in Alzheimer disease. Radiology. 2006. doi:10.1148/radiol.2411050628
2. WHO. ICD-11 Eleventh revision of the International Classification of Diseases. 2021;**2021**.
3. ADI. Alzheimer's Disease International. 2020;**2020**.https://www.alz.co.uk/
4. NHS. Overview - Alzheimer's disease. 2020;**2020**.https://www.nhs.uk/conditions/alzheimers-disease/
5. Frisoni GB, Fox NC, Jack Jr. CR, *et al.* The clinical use of structural MRI in Alzheimer disease. Nat Rev Neurol 2010;**6**:67–77. doi:10.1038/nrneurol.2009.215
6. Waldemar G, Dubois B, Emre M, *et al.* Recommendations for the diagnosis and management of Alzheimer's disease and other disorders associated with dementia: EFNS guideline. Eur J Neurol 2007;**14**:e1-26. doi:10.1111/j.1468-1331.2006.01605.x
7. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;**12**:189–98. doi:10.1016/0022-3956(75)90026-6
8. Aderghal, K., Benois-Pineau, J., Afdel, K., Gwenaëlle C. FuseMe: Classification of sMRI images by fusion of Deep CNNs in 2D + projections. In: *15th International Workshop on Content-Based Multimedia Indexing*. doi:10.1145/3095713.3095749
9. Backstrom K, Nazari M, Gu IYH, *et al.* An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images. In: *Proceedings - International Symposium on Biomedical Imaging*. IEEE Computer Society 2018. 149–53. doi:10.1109/ISBI.2018.8363543
10. Korolev S, Safiullin A, Belyaev M, *et al.* Residual and plain convolutional neural networks for 3D brain MRI classification. In: 2017 *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE 2017. 835–8. doi:10.1109/ISBI.2017.7950647
11. Lian C, Liu M, Zhang J, *et al.* Hierarchical Fully Convolutional Network for Joint Atrophy Localization and Alzheimer's Disease Diagnosis Using Structural MRI. IEEE Trans Pattern Anal Mach Intell

2020;**42**:880–93. doi:10.1109/TPAMI.2018.2889096

12. Senanayake U, Sowmya A, Dawes L. Deep fusion pipeline for mild cognitive impairment diagnosis. In: 2018 *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE 2018. 1394–997. doi:10.1109/ISBI.2018.8363832
13. Taqi AM, Awad A, Al-Azzo F, *et al.* The Impact of Multi-Optimizers and Data Augmentation on TensorFlow Convolutional Neural Network Performance. In: 2018 *IEEE Conference on Multimedia Information Processing and* Retrieval *(MIPR)*. IEEE 2018. 140–5. doi:10.1109/MIPR.2018.00032
14. Valliani A, Soni A. Deep Residual Nets for Improved Alzheimer's Diagnosis. In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics*. New York, NY, USA:: ACM 2017. 615–615. doi:10.1145/3107411.3108224
15. Wu C, Guo S, Hong Y, *et al.* Discrimination and conversion prediction of mild cognitive impairment using convolutional neural networks. Quant Imaging Med Surg 2018;**8**:992–1003. doi:10.21037/qims.2018.10.17
16. Ebrahimi A, Luo S, Disease Neuroimaging Initiative for the A. Convolutional neural networks for Alzheimer's disease detection on MRI images. J Med Imaging 2021;**8**. doi:10.1117/1.JMI.8.2.024503
17. Feng W, Halm-Lutterodt N Van, Tang H, *et al.* Automated MRI-Based Deep Learning Model for Detection of Alzheimer's Disease Process. Int J Neural Syst 2020;**30**. doi:https://doi.org/10.1142/S012906572050032X
18. Guilherme Folego, Marina Weiler, Raphael F. Casseb RP and A. Alzheimer's Disease Detection Through Whole-Brain 3D-CNN MRI. Front Bioeng Biotechnol 2020;**8**.
19. Basaia S, Agosta F, Wagner L, *et al.* Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. NeuroImage Clin 2019;**21**:101645. doi:10.1016/j.nicl.2018.101645
20. Wang SH, Phillips P, Sui Y, *et al.* Classification of Alzheimer's Disease Based on Eight-Layer Convolutional Neural Network with Leaky Rectified Linear Unit and Max Pooling. J Med Syst 2018;**42**:85. doi:10.1007/s10916-018-0932-7
21. Cheng D, Liu M. CNNs based multi-modality classification for AD diagnosis. In: *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE 2017. 1–5. doi:10.1109/CISP-BMEI.2017.8302281
22. Rehman AFSMAMAS. A deep CNN based multi-class classification of Alzheimer's disease using MRI. In: *IEEE International Conference on Imaging Systems and Techniques (IST)*. 2017. doi:10.1109/IST.2017.8261460
23. Gunawardena KANNP, Rajapakse RN, Kodikara ND. Applying convolutional neural networks for pre-detection of alzheimer's disease from structural MRI data. In: 2017 *24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*. IEEE 2017. 1–7. doi:10.1109/M2VIP.2017.8211486
24. Ehsan Hosseini Asl Ayman S El-Baz RSK. Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional Network. IEEE 2016. doi:10.1109/ICIP.2016.7532332

25. Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Inf 2018;**5**:2. doi:10.1186/s40708-018-0080-3
26. Wen J, Thibeau-Sutre E, Diaz-Melo M, *et al.* Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. Med Image Anal 2020;**63**. doi:10.1016/j.media.2020.101694
27. Malone IB, Cash D, Ridgway GR, *et al.* MIRIAD–Public release of a multiple time point Alzheimer's MR imaging dataset. Neuroimage 2013;**70**:33–6. doi:10.1016/j.neuroimage.2012.12.044
28. Larobina M, Murino L. Medical image file formats. J Digit Imaging 2014;**27**:200–6. doi:10.1007/s10278-013-9657-9
29. Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005;**26**:839–51. doi:10.1016/j.neuroimage.2005.02.018
30. Gorgolewski K, Burns CD, Madison C, *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform 2011;**5**:13. doi:10.3389/fninf.2011.00013
31. Ian Goodfellow, Yoshua Bengio AC. *Deep Learning*. Adaptive c. Cambridge, Mass.; London: : MIT Press 2016. https://lccn.loc.gov/2016022992
32. Murphy KP. *Machine learning: a probabilistic perspective*. Adaptive c. Cambridge, Mass.; London: : MIT Press 2012.
33. Chollet F. *Deep Learning with Python*. Manning Publications 2018.
34. Sánchez Fernández I, Yang E, Calvachi P, *et al.* Deep learning in rare disease. Detection of tubers in tuberous sclerosis complex. PLoS One 2020;**15**:e0232376. doi:10.1371/journal.pone.0232376
35. Medina G, Buckless CG, Thomasson E, *et al.* Deep learning method for segmentation of rotator cuff muscles on MR images. Skeletal Radiol 2021;**50**:683–92. doi:10.1007/s00256-020-03599-2
36. Nitish Srivastava; Hinton KS and S. Dropout: A Simple Way to Prevent Neural Networks from Over tting. J Mach Learn Res 2014;**15**:1929–58.
37. Li X, Pang T, Xiong B, *et al.* Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification. In: *Proceedings – 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017*. 2018. doi:10.1109/CISP-BMEI.2017.8301998
38. Friedman THRTJ. *The Elements of Statistical Learning*. New York, NY:: Springer New York 2009. doi:10.1007/b94608
39. Jurman DCG. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 2020;**21**.https://doi.org/10.1186/s12864-019-6413-7
40. Chicco D. Ten quick tips for machine learning in computational biology. BioData Min 2017;**10**:35. doi:10.1186/s13040-017-0155-3

41. Carter J V., Pan J, Rai SN, *et al.* ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. Surg (United States) 2016;**159**:1638–45. doi:10.1016/j.surg.2015.12.029
42. Hon M, Khan NM. Towards Alzheimer's disease classification through transfer learning. *Proc – 2017 IEEE Int Conf Bioinforma Biomed BIBM* 2017 2017;**2017-January**:1166–9. doi:10.1109/BIBM.2017.8217822
43. Ebrahimi A, Luo S, Disease Neuroimaging Initiative for the A. Convolutional neural networks for Alzheimer's disease detection on MRI images. J Med Imaging 2021;**8**:1–18. doi:10.1117/1.JMI.8.2.024503

Figures

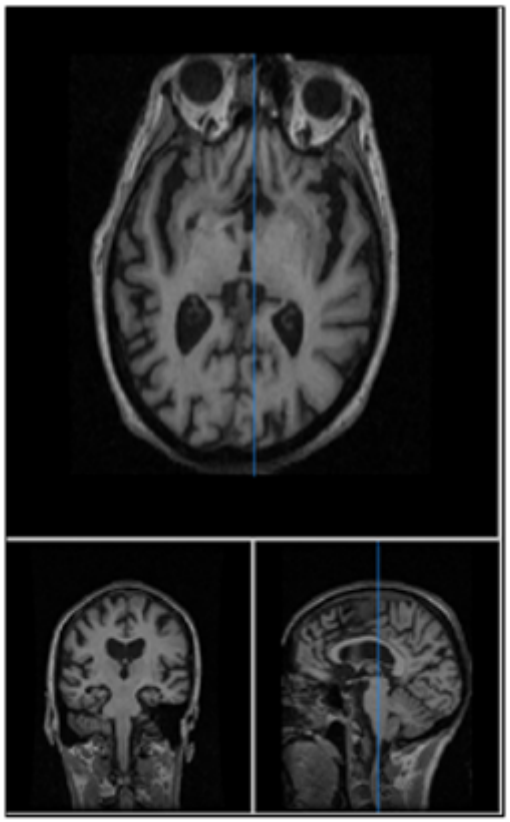


Figure 1

sample of raw MRI-data – axial, sagittal, and coronal views of the MIRIAD-dataset

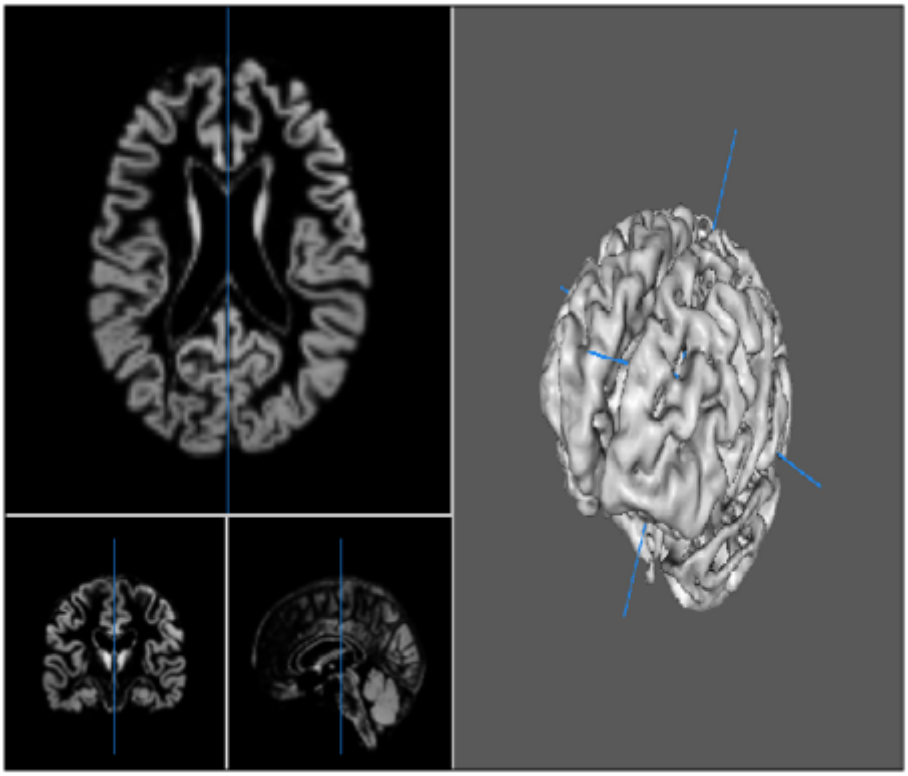


Figure 2

Pre-processed data – axial, sagittal, coronal, and 3D surface view

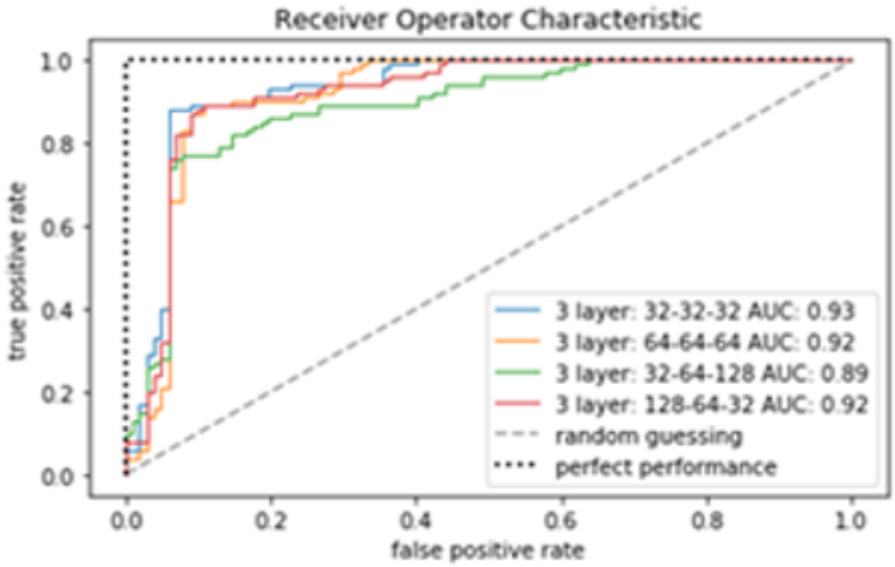


Figure 3

ROC-plot for different configurations of the CNN