








Interpretable Weighted Siamese Network to Predict the Time to Onset of Alzheimer's Disease from MRI Images

Misgina Tsighe Hagos^{1,2} , Niamh Belton^{1,3} , Ronan P. Killeen^{3,4,5} ,
Kathleen M. Curran^{1,3} , Brian Mac Namee^{1,2} ,
and for the Alzheimer's Disease Neuroimaging Initiative

¹ Science Foundation Ireland Centre for Research Training in Machine Learning,
Dublin, Ireland

misgina.hagos@ucdconnect.ie

² School of Computer Science, Dublin, Ireland

³ School of Medicine, University College Dublin, Dublin, Ireland

⁴ Department of Radiology, Dublin, Ireland

⁵ UCD-SVUH PET CT Research Centre, St Vincent's University Hospital,
Dublin 4, Ireland

Abstract. Alzheimer's Disease (AD) is a progressive disease preceded by Mild Cognitive Impairment (MCI). Early detection of AD is crucial for making treatment decisions. However, most of the literature on computer-assisted detection of AD focuses on classifying brain images into one of three major categories: healthy, MCI, and AD; or categorizing MCI patients into (1) *progressive*: those who progress from MCI to AD at a future time, and (2) *stable*: those who stay as MCI and never progress to AD. This misses the opportunity to accurately identify the trajectory of progressive MCI patients. In this paper, we revisit the AD identification task and re-frame it as an ordinal classification task to predict *how close a patient is to the severe AD stage*. To this end, we construct an ordinal dataset of progressive MCI patients with a prediction target that indicates the time to progression to AD. We train a Siamese network (SN) model to predict the time to onset of AD based on MRI brain images. We also propose a Weighted variety of SN and compare its performance to a baseline model. Our evaluations show that incorporating a weighting factor to SN brings considerable performance gain. Moreover, we complement our results with an interpretation of the learned embedding space of the SNs using a model explainability technique.

Keywords: Alzheimer's Disease · Mild Cognitive Impairment ·
Computer Assisted Diagnosis · Siamese Networks

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
M. Bramer and F. Stahl (Eds.): SGAI 2023, LNAI 14381, pp. 391–403, 2023.
https://doi.org/10.1007/978-3-031-47994-6_35

1 Introduction

Although it has been more than a century since Alois Alzheimer first described the clinical characteristics of Alzheimer’s Disease (AD) [3,4], the disease still eludes early detection. AD is the leading cause of dementia, accounting for 60%-80% of all cases worldwide [20,34], and the number of patients effected is growing. In 2015, Alzheimer’s Disease International (ADI) reported that over 46 million people were estimated to have dementia worldwide and that this number was expected to increase to 131.5 million by 2050 [29]. Since AD is a progressive disease, computer assisted early identification of the disease may enable early medical treatment to slow its progression.

Methods that require intensive expert input for feature collection, such as Morphometry [13], and more automated solutions based on deep learning [5,8,24] have been utilized in the computer assisted diagnosis of AD literature. These automated detection methods usually classify patients as belonging to one of three stages: Normal (patients exhibiting no signs of dementia and no memory complaints), Mild Cognitive Impairment (MCI) (an intermediate state in which a patient’s cognitive decline is greater than expected for their age, but does not interfere with activities of their daily life), and full AD.

A participant’s progression from one of the stages to the next, however, can take more than five years [30]. This can mean that when automated disease classification systems based on these three levels are used, patients at a near severe stage do not receive the required treatment because they are classified as belonging to the pre-severe stage. This is illustrated in Fig. 1(a) for five participants from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) study [28]. Using the typical classification approach (See Fig. 1(b)), for example, even though participant five is only a year away from progressing to the severe AD stage, they would be classified as MCI in the year 2009. To address this issue we focus on the clinical question “how far away is a progressive MCI patient on their trajectory to AD?” To do this we propose an ordinal categorization of brain images based on participants’ level of progression from MCI to AD as shown in Fig. 1(c). Our approach adds ordinal labels to MRI scans of patients with progressive MCI indicating how many years they are from progressing to AD, and we construct a dataset of 444 MRI scans from 288 participants with these labels and share a replication script.

In addition to constructing the dataset, we also develop a computer assisted approach to identifying a participant’s (or more specifically, their MRI image’s) progression level. Accurately identifying how far a patient is from progressing to full AD is of paramount importance as this information may enable earlier intervention with medical treatments [2]. Rather than using simple ordinal classification techniques, we use Siamese networks due to their ability to handle the class imbalance in the employed dataset [21,33]. We use a Siamese network architecture, and a novel Weighted Siamese network that uses a new loss function tailored to learning to predict input MRI image’s likelihood of progression. Furthermore, we complement results of our Siamese network based method with interpretations of the embedding space using an auxiliary model explanation technique, T-distributed Stochastic Neighbor Embedding (t-SNE) [26]. t-SNE

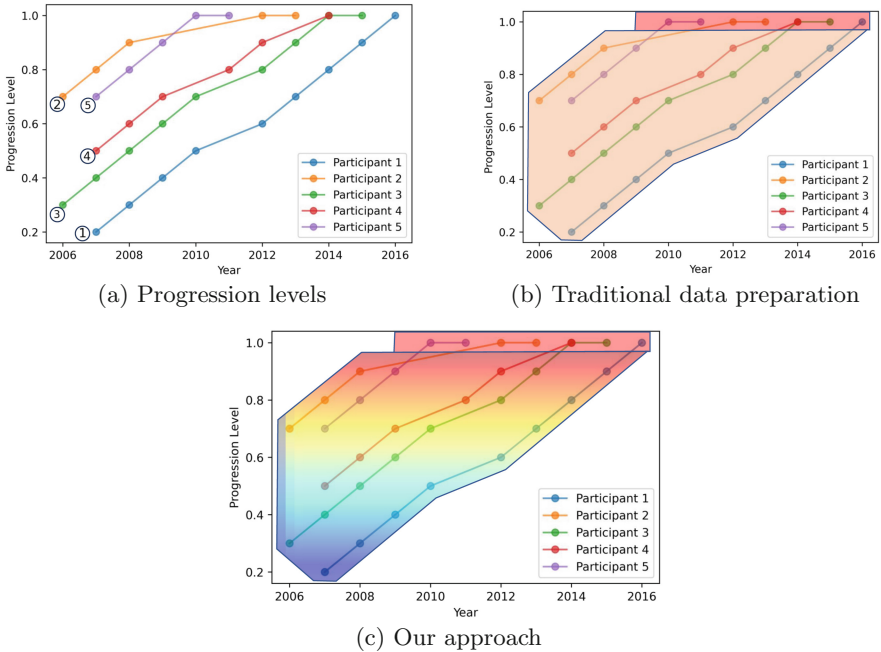


Fig. 1. (a) Progression levels of five sample MCI participants where each dot represents an MRI image during an examination year. (b) shows the typical approach of organizing images for identification of progressive MCI or classification of MCI and AD. The images in the lower light orange section are categorized as MCI when preparing a training dataset (this includes those images from patients nearing progression to AD—near progression level 1.0). The images in the upper red section are categorized as AD. In (c) our approach to organizing brain images is illustrated using a Viridis map. Images are assigned ordinal progression levels $\in [0.1, 0.9]$ based on their distance in years from progressing to AD stage. (Color figure online)

condenses high dimensional embedding spaces learned by a Siamese network into interpretable two or three dimensional spaces [9].

The main contributions of this paper are:

1. We provide a novel approach that interpolates ordinal categories between existing MCI and AD categories of the ADNI dataset based on participants' progression levels.
2. We apply the first Siamese network approach to predict interpolated progression levels of MCI patients.
3. We propose a simple and novel variety of triplet loss for Siamese networks tailored to identifying progression levels of MCI patients.
4. Our experiments demonstrate that using our version of the triplet loss is better at predicting progression level than the traditional triplet loss. Code is shared online¹.

¹ <https://github.com/Msgun/WeightedSiamese>.

2 Related Work

Before the emergence of deep learning, and in the absence of relevant large datasets, computer-assisted identification of AD relied on computations that require expensive expert involvement such as Morphometry [13]. However, the release of longitudinal datasets, such as ADNI [28], inspired research on automated solutions that employed machine learning and deep learning methods for the identification of AD.

Most of the approaches proposed for AD diagnosis perform a classification among three recognized stages of the disease: Normal, MCI, and AD [23]. Some examples in the literature distinguish between all three of the categories [36], while others distinguish between just two: Normal and MCI [19], Normal and AD [32], or MCI and AD [31].

Patients at the MCI stage have an increased risk of progressing to AD, especially for elderly patients [30]. For example, in the Canadian Cohort Study of Cognitive Impairment and Related Dementia [17] 49 out of a cohort of 146 MCI patients progressed to AD in a two-year follow-up. In general, while healthy adult controls progress to AD annually at a maximum rate of 2%, MCI patients progress at a rate of 10%-25% [15]. This necessitates research on identifying MCI subjects at risk of progressing to AD. In a longitudinal study period, participants diagnosed with MCI can be categorized into two categories: (1) Progressive MCI, which represents participants who were diagnosed with MCI at some stage during the study but were later diagnosed with AD, and (2) Stable MCI, patients who stayed as MCI during the whole study period [16]. This excludes MCI participants with chances of reverting back to healthy, since they were also reported to have chances of progressing to AD [30]. There are some examples in the literature of using machine learning techniques such as random forest [27] and CNNs [16] to classify between stable and progressive MCI. While feature extraction is used prior to model training towards building relatively simpler models [23, 35], 3D brain images are also deployed with 3D CNNs to reduce false positives [5, 25].

The brain image classification task can also be transformed to ordinal classification to build regressor models. For example, four categories of AD: healthy, stable MCI, progressive MCI, and AD were used as ordinal labels to build a multivariate ordinal regressor using MRI images in [14]. However, the output of these models gives no indication of the likelihood of a patient to progress from one stage to another. Furthermore, this does not provide prediction for interpolated inter-category progression levels. Albright et al. (2019) [2] used a longitudinal clinical data including ADAS13, which is a 13-item Alzheimer's Disease Assessment Scale, and Mini-Mental State Examination (MMSE) to train multi-layer perceptron and recurrent networks for AD progression prediction. This work, however, uses no imaging data and it has been shown that brain images play a key role in improving diagnostic accuracy for Alzheimer's disease [18].

Siamese networks, which use a distance-based similarity training approach [10, 12], have found applications in areas such as object tracking [7] and anomaly detection [6]. Although it does not focus on AD detection, we found [22] to be the closest approach to our proposed method in the literature. Li et al. [22]

report that a Siamese network’s distance output could be translated to predict disease positions on a severity scale. Although this approach takes the output as severity scale without any prior training on disease severity, only deals with existing disease stages, and does not interpolate ordinal categories within, it does suggest Siamese networks as a promising approach for predicting ordinal progression levels.

3 Approach

In this section, we describe the datasets (and how they are processed), model architectures, model training and evaluation techniques used in our experiments, as well as our proposed triplet loss for Siamese networks.

3.1 Dataset Preparation

The data used in the experiments described here was obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database. ADNI was launched in 2003, led by Principal Investigator Michael W. Weiner, MD (<https://adni.loni.usc.edu>). For up-to-date information, see <https://adni.loni.usc.edu/>.

Table 1. Image distribution across progression levels, ρ .

ρ	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Number of images	4	4	6	10	24	56	172	273	467

From the ADNI dataset we identified MRI brain images of 1310 participants who were diagnosed with MCI or AD. 288 participants had progressive MCI, 545 had stable MCI, and the rest had AD. We used MRI images of the 288 progressive MCI participants to train and evaluate our models. We labeled the progressive MCI participants based on their progression levels towards AD, $\rho \in [0.1, 1.0]$ with a step size = 0.1, where for a single participant, P , $\min(\rho) = 0.1$ represents the first time P was diagnosed with MCI and P transitions to stage AD at $\max(\rho) = 1.0$. This transforms the binary MCI and AD labels to 10 ordinal labels. An example of the data organization based on progression level is plotted in Fig. 1. The distribution of the constructed ordinal regression levels is shown in Table 1 (where $\rho = 1.0$ represents AD) where the imbalance between the different labels is clear. Within the ADNI dataset, the maximum number of MRI scans that the progressive MCI participants have had until they progressed to AD ($\rho = 1.0$) is 9, which means that the smallest ρ is 0.2. We took advantage of Siamese networks robustness to class imbalance to circumnavigate the imbalance in the ordinal labels. By sub-sampling from the majority classes, we selected 444 3D MRI images (shape = $160 \times 192 \times 192$) for the negative, anchor, and positive datasets (each holding 148 images) required when training a Siamese network

using triplet loss. We used 80% of the images for training and the rest for testing. AD images were randomly separated to the anchor and positive dataset. We ensure that there is no participant overlap between sets when performing the data splitting between training and testing dataset, and between anchor and positive dataset.

3.2 Weighted Siamese Network

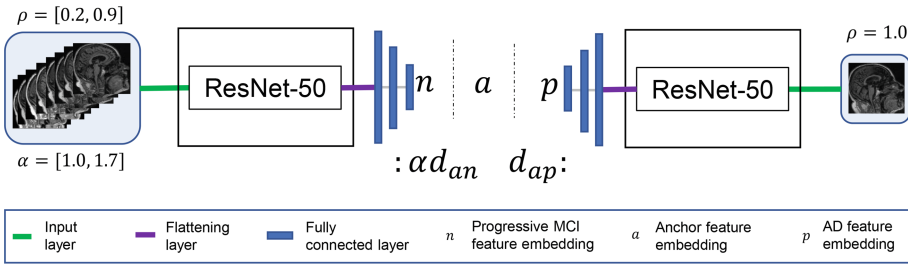


Fig. 2. Weighted Siamese network. The text ResNet-50 here refers to the base layers of the ResNet-50 architecture which are trained from scratch for extracting image embeddings, i.e. excluding the fully connected classifier layers. While we used 3D MRI images for model training and evaluation, sagittal plane is used here only for visualization purposes.

Siamese networks are usually trained using a triplet loss or its variants. While a traditional triplet loss teaches a network that a negative instance is supposed to be at a larger distance from the anchor than a positive instance, we propose a Weighted triplet loss that teaches a network that instances, which can all be considered to be in the negative category, are not at the same distance from the anchor and that their distance depends on their progression level, ρ . So that lower progression levels have larger distance from an anchor instance, we transform ρ to a weighting coefficient $\alpha = 1.9 - \rho$, excluding $\rho = 1.0$, as shown in Fig. 3. The architecture of our proposed Weighted Siamese network is shown in Fig. 2.

We used two different loss functions to train our Siamese networks. The first is a traditional triplet loss, which we refer to as Unweighted Siamese:

$$L_u = \max(d_{ap} - d_{an} + margin, 0) \tag{1}$$

where $margin = 1.0$, d_{ap} is the Euclidean distance between anchor and positive embeddings, and d_{an} is the distance between the embeddings of anchor and negative instances.

The second loss is a newly proposed Weighted triplet loss—Weighted Siamese which introduces a coefficient $\alpha \in [1.0, 1.8]$ to d_{an} in L_u :

$$L_w = \max(d_{ap} - \alpha d_{an} + margin, 0) \tag{2}$$

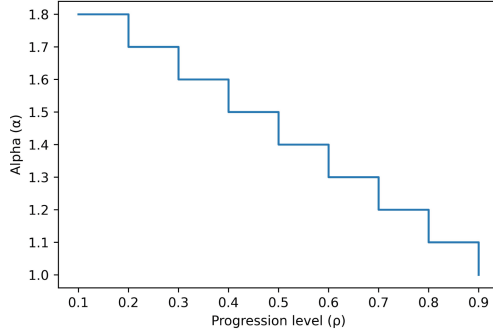


Fig. 3. Transforming progression level, ρ , to α .

3.3 Training and Evaluation

We implemented all of our experiments using TensorFlow [1] and Keras [11]. After comparing performance between different architectures and feature embedding size, we chose to train a 3D ResNet-50 model from scratch by adding three fully connected layers of sizes 64, 32, and 8 nodes with ReLU activations, taking the last layer of size 8 as the embedding space. We used an Adam optimizer with a decaying learning rate of $1e-3$. We trained the model with five different seeds for 150 epochs, which took an average of 122 min per a training run on an NVIDIA RTX A5000 graphics card.

For model evaluation on training and testing datasets, we use both the Unweighted Siamese and Weighted Siamese losses as well as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). MAE and RMSE are presented in Eqs. 3 and 4 respectively, for a test set of size N where y_i and Y_i hold the predicted and ground truth values for instance i , respectively. We turn the distance outputs of the Siamese networks into y by discretizing them into equally spaced bins, where the number of bins equals the number of progression levels.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - Y_i| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - Y_i)^2} \quad (4)$$

We make use of t-SNE to explain the 8 dimension embedding space learned by the Weighted Siamese network by condensing it two dimensions. For presentation purposes and in order to fit the t-SNE well, we drop underrepresented progression levels; while we dropped progression level 0.2 from the training dataset, progression levels 0.2, 0.3, and 0.5 were removed from the testing dataset. The t-SNE was fitted over a 1000 iterations using Euclidean distance metric with a perplexity of 32 and 8 for the training and testing datasets, respectively.

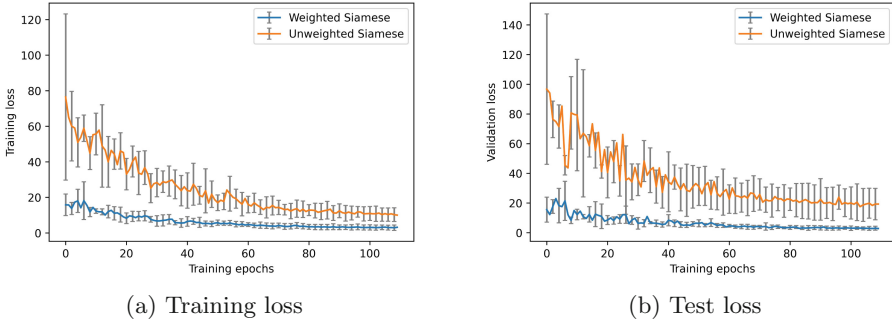


Fig. 4. Training and testing losses of the Weighted and Unweighted Siamese models. The first 40 epochs are cropped out for easier visualization. Bars represent std. errors over five runs.

4 Results and Discussion

In this section, we present training and testing losses, MAE and RMSE metrics of evaluation, a plot showing comparison between predicted and ground-truth progression levels, as well as interpretation of the results.

Training and testing losses over five runs of model training for both the Unweighted and Weighted Siamese networks are shown in Fig. 4. While the average training and testing losses of the Weighted Siamese network are 2.92 and 2.79, the Unweighted Siamese achieves 10.02 and 17.53, respectively. We were able to observe that the Unweighted Siamese network had a hard time learning the progression levels of all the ordinal categories. However, our proposed approach using Weighted loss was better at fitting to all the levels. We accredit this to the effects of adding a weighing factor using ρ .

A plot of predicted vs. ground truth MCI to AD progression levels is presented in Fig. 5. Our proposed Weighted Siamese network outperforms the Unweighted Siamese network at predicting progression levels(Fig. 5 and Table 2).

We observed that the simple modification of factoring the distance between an embedding of anchor and negative instances by a function of the progression level brought considerable performance gain in separating between the interpolated categories between MCI and AD.

In Fig. 5, although the Weighted Siamese outperforms the Unweighted Siamese, it also usually classifies the input test images with lower progression levels as if they are on a higher progression levels. This would mean brain images

Table 2. Average MAE and RMSE over five runs.

Method	MAE	RMSE
Unweighted Siamese	2.30	2.94
Weighted Siamese	2.00	2.40

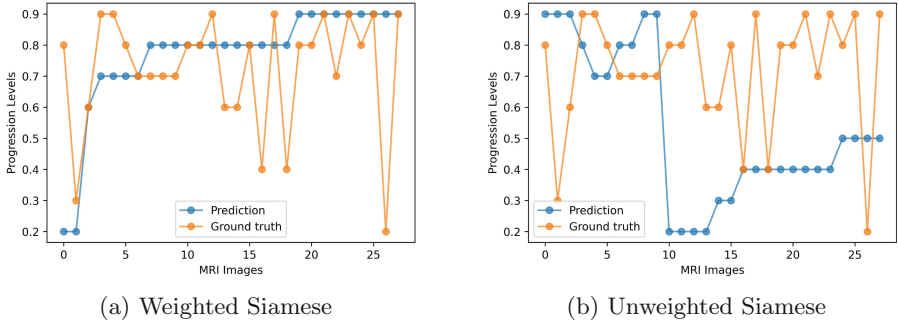


Fig. 5. Predicted progression levels of test MRI images against ground truth levels.

of patients that are far away from progressing to AD would be identified as they are close to progressing. While it’s important to correctly identify these low risk patients, we believe it’s better to report the patients at lower risk as high risk and refer them for expert input than classifying high risk patients as low risk.

An interpretation of the results of the proposed Weighted Siamese method using t-SNE is displayed in Fig.6. The clustering of the embedding of input instances according to their progression levels, especially between the low-risk and high-risk progression levels assures us that the results represent the ground truth disease levels.

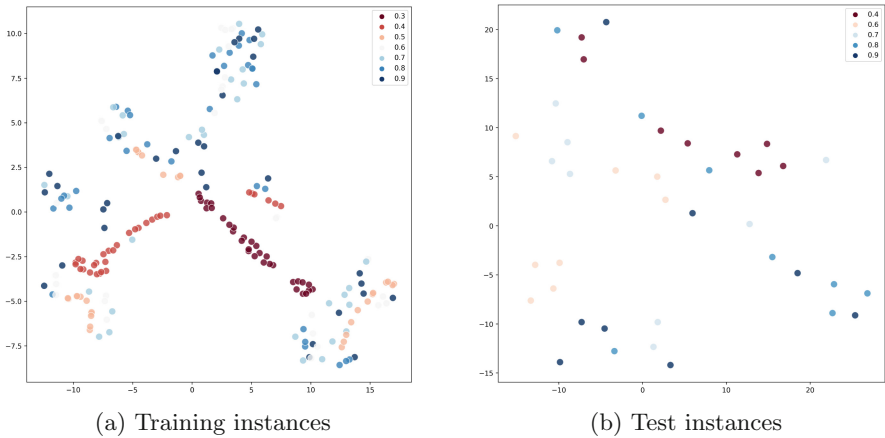


Fig. 6. Visualization of t-SNE of the embedding spaces of training and test instances.

5 Conclusion

Similarly to other image-based computer assisted diagnosis research work, the AD identification literature is heavily populated by disease stage classification. However, an interesting extra step can be taken to identify how far an input brain image is from progressing to a more severe stage of AD. We present a novel approach of interpolating ordinal categories in-between the MCI and AD categories to prepare a training dataset. In addition, we proposed and implemented a new Weighted loss term for Siamese networks that is tailored to such a dataset. With our experiments, we show that our proposed approach surpassed the performance of a model trained using a standard Unweighted loss term; and we show how the predicted levels translate to the ground truth progression levels by applying a model interpretability technique on the embedding space. We believe our approach could easily be transferred to other areas of medical image classification involving progressive diseases.

The diagnosis results taken in our study are bounded by the timeline of the ADNI study—meaning, even though based on extracted information a participant may have MCI during an examination year and they may progress to AD after some year(s), they could have had MCI before joining the ADNI study and their progression to AD might have taken longer than what we have noted. Future work should consider this limitation.

Acknowledgment. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

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 the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. 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) 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. Abadi, M., et al.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015). <https://www.tensorflow.org/>, software available from tensorflow.org
2. Albright, J., Initiative, A.D.N., et al.: Forecasting the progression of Alzheimer's disease using neural networks and a novel preprocessing algorithm. *Alzheimer's Dementia: Transl. Res. Clin. Interventions* **5**, 483–491 (2019)
3. Alzheimer, A.: Uber eine eigenartige erkrankung der hirnrinde. *Zentralbl. Nervenh. Psych.* **18**, 177–179 (1907)
4. Alzheimer, A., Stelzmann, R.A., Schnitzlein, H.N., Murtagh, F.R.: An english translation of alzheimer's 1907 paper," uber eine eigenartige erkankung der hirnrinde". *Clin. Anat. (New York, NY)* **8**(6), 429–431 (1995)
5. Basaia, S., et al.: Automated classification of alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical* **21**, 101645 (2019)
6. Belton, N., Lawlor, A., Curran, K.M.: Semi-supervised siamese network for identifying bad data in medical imaging datasets. arXiv preprint [arXiv:2108.07130](https://arxiv.org/abs/2108.07130) (2021)
7. Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., Torr, P.H.S.: Fully-convolutional Siamese networks for object tracking. In: Hua, G., Jégou, H. (eds.) *ECCV 2016*. LNCS, vol. 9914, pp. 850–865. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_56
8. Billones, C.D., Demetria, O.J.L.D., Hostallero, D.E.D., Naval, P.C.: Demnet: a convolutional neural network for the detection of Alzheimer's disease and mild cognitive impairment. In: 2016 IEEE Region 10 Conference (TENCON), pp. 3724–3727. IEEE (2016)
9. Borys, K., et al.: Explainable AI in medical imaging: an overview for clinical practitioners-beyond saliency-based XAI approaches. *Eur. J. Radiol.* 110786 (2023)
10. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a "Siamese" time delay neural network. In: *Advances in Neural Information Processing Systems*, vol. 6 (1993)
11. Chollet, F., et al.: Keras (2015). <https://keras.io>
12. Chopra, S., Hadsell, R., LeCun, Y.: Learning a similarity metric discriminatively, with application to face verification. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 539–546. IEEE (2005)
13. Dashjamts, T., et al.: Alzheimer's disease: diagnosis by different methods of voxel-based morphometry. *Fukuoka igaku zasshi= Hukuoka acta medica* **103**(3), 59–69 (2012)
14. Doyle, O.M., et al.: Predicting progression of Alzheimer's disease using ordinal regression. *PLoS ONE* **9**(8), e105542 (2014)
15. Grand, J.H., Caspar, S., MacDonald, S.W.: Clinical features and multidisciplinary approaches to dementia care. *J. Multidiscip. Healthc.* **4**, 125 (2011)
16. Hagos, M.T., Killeen, R.P., Curran, K.M., Mac Namee, B., Initiative, A.D.N., et al.: Interpretable identification of mild cognitive impairment progression using stereotactic surface projections. In: *PAIS 2022*, pp. 153–156. IOS Press (2022)
17. Hsiung, G.Y.R., et al.: Outcomes of cognitively impaired not demented at 2 years in the Canadian cohort study of cognitive impairment and related dementias. *Dement. Geriatr. Cogn. Disord.* **22**(5–6), 413–420 (2006)

18. Johnson, K.A., Fox, N.C., Sperling, R.A., Klunk, W.E.: Brain imaging in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**(4), a006213 (2012)
19. Ju, R., Hu, C., Zhou, P., Li, Q.: Early diagnosis of Alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform. (TCBB)* **16**(1), 244–257 (2019)
20. Kalaria, R.N., et al.: Alzheimer's disease and vascular dementia in developing countries: prevalence, management, and risk factors. *Lancet Neurol.* **7**(9), 812–826 (2008)
21. Koch, G., Zemel, R., Salakhutdinov, R., et al.: Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*, vol. 2. Lille (2015)
22. Li, M.D., et al.: Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging. *NPJ Digital Med.* **3**(1), 48 (2020)
23. Li, Q., Wu, X., Xu, L., Chen, K., Yao, L., Initiative, A.D.N.: Classification of Alzheimer's disease, mild cognitive impairment, and cognitively unimpaired individuals using multi-feature kernel discriminant dictionary learning. *Front. Comput. Neurosci.* **11**, 117 (2018)
24. Liu, M., Cheng, D., Wang, K., Wang, Y.: Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis. *Neuroinformatics* **16**(3), 295–308 (2018)
25. Liu, M., Zhang, J., Nie, D., Yap, P.T., Shen, D.: Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. *IEEE J. Biomed. Health Inform.* **22**(5), 1476–1485 (2018)
26. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(11) (2008)
27. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A.D.N., et al.: Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage* **104**, 398–412 (2015)
28. Mueller, S.G., et al.: Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimer's & Dementia* **1**(1), 55–66 (2005)
29. Prince, M.J., Wimo, A., Guerchet, M.M., Ali, G.C., Wu, Y.T., Prina, M.: World alzheimer report 2015-the global impact of dementia: An analysis of prevalence, incidence, cost and trends (2015)
30. Roberts, R.O., et al.: Higher risk of progression to dementia in mild cognitive impairment cases who revert to normal. *Neurology* **82**(4), 317–325 (2014)
31. Wegmayr, V., Aitharaju, S., Buhmann, J.: Classification of brain MRI with big data and deep 3D convolutional neural networks. In: *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575, p. 105751S. International Society for Optics and Photonics (2018)
32. Xiao, R., et al.: Early diagnosis model of Alzheimer's disease based on sparse logistic regression with the generalized elastic net. *Biomed. Signal Process. Control* **66**, 102362 (2021)
33. Yang, W., Li, J., Fukumoto, F., Ye, Y.: HSCNN: a hybrid-siamese convolutional neural network for extremely imbalanced multi-label text classification. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6716–6722 (2020)

34. Zhang, R., Simon, G., Yu, F.: Advancing Alzheimer's research: a review of big data promises. *Int. J. Med. Informatics* **106**, 48–56 (2017)
35. Zheng, C., Xia, Y., Chen, Y., Yin, X., Zhang, Y.: Early diagnosis of Alzheimer's disease by ensemble deep learning using FDG-PET. In: Peng, Y., Yu, K., Lu, J., Jiang, X. (eds.) *IScIDE 2018*. LNCS, vol. 11266, pp. 614–622. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02698-1_53
36. Zhou, T., Thung, K.H., Zhu, X., Shen, D.: Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Hum. Brain Mapp.* **40**(3), 1001–1016 (2019)