A Parallel Machine Learning Framework for Detecting Alzheimer's Disease

Sean A Knox[1], Tianhua Chen[1], Pan Su[2], and Grigoris Antoniou[1]

[1] University of Huddersfield, Queensgate, Huddersfield HD1 3DH, United Kingdom
{sean.knox2,t.chen,g.antoniou}@hud.ac.uk
[2] School of Control and Computer Engineering, North China Electric Power University, China supan@ncepu.edu.cn

Abstract. This paper proposes a parallel machine learning framework for detecting Alzheimer's disease through T1-weighted MRI scans localised to the hippocampus, segmented between the left and right hippocampi. Feature extraction is first performed by 2 separately trained, unsupervised learning based AutoEncoders, where the left and right hippocampi are fed into their respective AutoEncoder. Classification is then performed by a pair of classifiers on the encoded data from the AutoEncoders, to which each pair of the classifiers are aggregated together using a soft voting ensemble process. The best averaged aggregated model results recorded was with the Gaussian Naïve Bayes classifier where sensitivity/specificity achieved were 80%/81% respectively and a balanced accuracy score of 80%.

Keywords: Machine Learning · Alzheimer's Disease · AutoEncoder · Multi-layer Perceptron · Support Vector Machine · Gaussian Naïve Bayes.

1 Introduction

Dementia is a syndrome that causes a more than normal deterioration in cognitive function to that caused by the usual affects of aging. Dementia is chronic and/or progressive in nature. Globally an estimated 50 million people have some form of dementia ca. 2020, according to the WHO (World Health Organisation) [24].

Alzheimer's Disease (AD) is the most common form of dementia contributing around $60-70\%$ of all dementia cases globally [24]. AD is predominantly found in the aged population (65+), with the risk of AD doubling every 5 years after [1]. While AD is not currently curable, the symptoms of AD can be lessened with the aid of prescription drugs: Acetylcholinesterase (AChE) inhibitors (early- to mid-stage), Memantine (late-stage), antipsychotics (behavioural and psychological symptoms of dementia (BPSD)) [17]. There are also therapies and activities that are beneficial to patients diagnosed with AD [17], therefore it is pivotal that patients are diagnosed accurately and as early, in the stages of their AD, as possible. The earliest clinical stage to detect progression to either dementia,

in general or specifically AD, is Mild Cognitive Impairment (MCI) [14]. MCI is the "transitional stage between age-related cognitive decline and AD" [15].

AD is defined by the "observation of specific pathological lesions" [16], these pathological lesions include: intracellular neurofibrillary tangles, extracellular β-amyloid senile plaques and blood vessel deposits, and synaptic degradation and neuronal atrophy of the brain [16]. As these pathological lesions are found in a specific pattern associated with AD, it is possible with the use of Magnetic Resonance Imaging (MRI) to detect the signs, or the absence of signs, of AD in a patient. "MRI has attracted a significant interest in AD related studies because of its completely non-invasive nature, high availability, high spatial resolution and good contrast between different soft tissues" [15].

In recognition of the challenges diagnosing dementia and the potentials of utilising recent machine learning techniques to facilitate clinical decision support, this paper proposes a parallel machine learning architecture where feature extraction and classification are trained in pairs (left/right hippocampi). Decisions from these pairs are then aggregated together using a soft voting ensemble algorithm. In particular, the feature extraction is performed by an unsupervised learning AutoEncoder and classification is performed by 3 different classifiers (Multi-layer Perceptron, Support Vector Machine, and Gaussian Naïve Bayes). Experimental results suggest this research is able to achieve comparable performance when compared to other proposed solutions which require more data points, than what is required for the proposed solution in this paper.

The reminder of this paper is organised as follows. Section II introduces the related works. Section III describes data set and the proposed pipeline. Section IV presents and discusses the experimental outcomes. Section V concludes the paper and outlines ideas for further development.

2 Related Works

Recent advances in machine learning (ML) have lead to many successes in the healthcare domain [3,22,4]. Numerous models [21,7,8,9,18,11,12] have been proposed for ML driven AI for the detection of signs of AD, ranging from the binary style classification of AD or CN (Cognitive normal), to multi-class classification of AD, MCI (progressive MCI(pMCI), stable MCI(sMCI)), and CN. Models that concentrate on multi-class classification introduce the capability of predicting the likelihood of a patient transitioning through MCI-to-AD in a given time frame.

While binary classifiers offer less granularity in their classification (i.e. MCI patients are usually coupled with AD), they do typically offer greater accuracy than those found with multi-class classification, around $80-95\%$ according to [20]. The popular OASIS dataset classifies any patient with greater than 0 Clinical Dementia Rating (CDR) as having AD, whereas the alternative ADNI repository would classify a patient with a CDR of 0.5 as having MCI not AD [13]. This stipulation would mean that the OASIS dataset would have a generally higher sensitivity than other datasets, as it is more likely to predict AD than CN.

The capability of predicting when/if a patient is transitioning through MCI-to-AD, allows for the patient to take quality of life improvements/preparations for when they do transition into AD. While preventative measures are not proven to work, cognitive stimulation therapy, has been proven to improve memory and problem-solving skills [17]. As such, the multi-class classification is one of the most researched areas in AI detection of AD, with it's added granularity that enables the ability to predict when a patient is going through the transition of MCI-to-AD.

The popular machine learning methods used in AD classification include Convolutional Neural Networks (CNN), which was originally designed for objective detection, and is able to learn image features of most importance that allow for correctly classifying the image as an image of a patient with AD or CN. In extension to the multiple types of neural networks, deep learning neural networks are also used in AD classification, these deep neural networks have achieved high performances e.g. some GoogleNet and ResNet models have achieved accuracies in the 97.9% on 3 way class classification tasks [18]. In regards to a binary classification task, AD vs CN, as used in this paper, deep neural network-based utilizing a sparse AutoEncoder and CNN for subjects only 75 and over have achieved 89.47% [18,19].

Another machine learning method showing promise in AD classification is the utilisation of an AutoEncoder to perform unsupervised learning on feature extraction. An AutoEncoder's architecture is very similar to that of a CNN except that it performs encoding (takes the data and encodes it to reduce it's size) and then decodes (takes the reduced data and tries to reconstruct the data to be as close to the original input as possible).

Although AutoEncoders may have a vanishing gradient problem, which can be mitigated by only allowing for the k highest hidden units to be active in any given hidden layer, an AutoEncoder is promising choice for AD classification, with its dimensionality reduction - it reduces the chance of data in the image, that is not important to AD classification, from influencing the classifier as much. This also inspires the underlying research to utilise AutoEncoders as a feature extraction technique, followed by the use of powerful classifiers to construct predictive models.

3 Method

3.1 Data Subjects, Acquisition, and Pre-processing

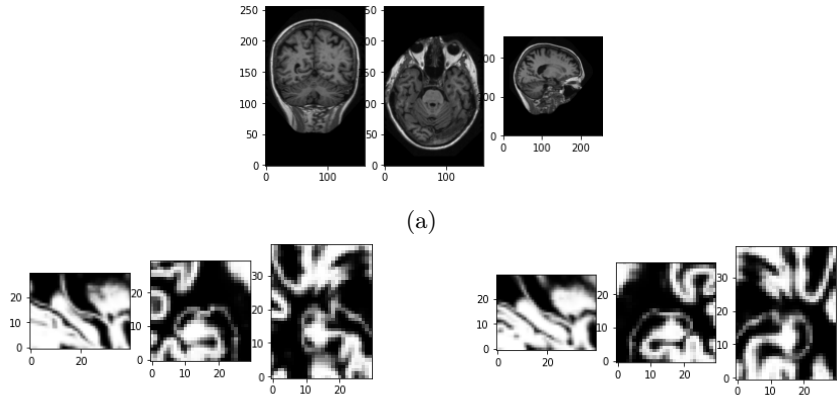
All data used during the experimentations relating to this paper were acquired from the OASIS-1: Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults dataset [13]. OASIS-1 contains 416 unique subjects, age ranging from 18-96, with 100 of those subjects being older than 60 and have been clinically diagnosed with very mild to moderate Alzheimer's disease. OASIS-1 contains age, education level (if present), CDR, MRI T1-weighted scans, and sex for each unique subject.

The dataset used had been pre-processed according to the methods set out by [23], in which bias field correction was performed to reduce any bias field

Table 1: OASIS-1 dataset [13] after stage 1 pre-processing [23]

	N	Age	%SexF	Education	MMS	CDR=0	CDR=0.5	CDR=1	CDR=2
AD	73	77.5 ± 7.4	63.0	2.7 ± 1.3	22.7 ± 3.6	0	45	26	2
CN	304	44.0 ± 23.3	62.2	3.5 ± 1.2	29.7 ± 0.6	124	0	0	0

corruption in the images, each image is then registered to the MNI space using a linear algorithm. Once the MNI space is registered the images were then cropped to remove the background, finally each image went through quality control where a probability score was produced on how accurate the MNI registration was, any images below .5 were removed automatically and images below .7 were manually reviewed. During this quality control phase 39 non-usable data subjects were removed as their MRI scans were not viable for use in ML (Table 1), during this pre-processing stage the dataset was split into separate training and validation sets (80/20 percentage split). After the pre-processing performed by [23], another pre-processing stage was undertaken which primary goal was to separate and crop the left and right hippocampus (Fig. 1), this allowed for the Neural Network to train only on the hippocampi and effectively doubled the dataset, allowing for a parallel computing approach where the left and right hippocampus are fed into the two separate models with an aggregation process at the end which will then calculate the prediction.



(a)

(b) (c)

Fig. 1: MRI Slices of subject sub-OASIS10003 after stage 1 and stage 2 pre-processing.
Where a shows the centre slices after stage 1 pre-processing on MRI image of subject sub-OASIS10003 and b and c shows the left and right hippocampus slices, respectively, after stage 2 pre-processing on MRI image of subject sub-OASIS10003

3.2 Feature Extraction

Feature extraction is performed by unsupervised learning with an AutoEncoder, two separate AutoEncoders are trained, left hippocampus and right hippocampus models. The AutoEncoder performs feature extraction by reducing the size of the data, it does this by removing detail in the image that are not of importance to the classification layer.

Both AutoEncoders used in this paper were trained using the Mean Squared Error loss function with reduction set to summation, i.e., Sum Squared Error:

$$\mathcal{L}(x, y) = \sum_{i=1}^{n} (\mathcal{Y}_i - \widehat{\mathcal{Y}}_i)^2 \tag{1}$$

where $\mathcal{Y}_i$ is the target image which in the case of AutoEncoder's is the input image, as an AutoEncoder's performance is evaluated by how closely the recreated image is to the original image. Adam was used as the algorithm optimiser. Each AutoEncoder had a starting learning rate of 10^{-3}, each AutoEncoder was trained using mini-batch gradient descent with an epoch of 50.

Since this is an unsupervised learning algorithm it is possible that the AutoEncoder will fixate on the wrong area of the image. To ensure that the AutoEncoder has fixated on the important details of the respective left and right hippocampus, the AutoEncoders and all the classifiers were trained 10 times with the average of the aggregation layer being recorded.

Both AutoEncoders had the same architecture, with the encoder starting with a 3D convolutional layer, a 3D batch normaliser, leaky ReLU process followed by a 3D pad max pool (custom operation that ensures no loss of data), repeating 3 times (Fig. 2).

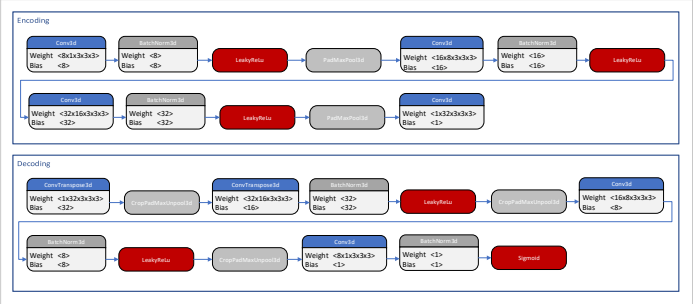


Fig. 2: AutoEncoder architecture used for feature extraction

3.3 Classification

There were 3 different classification methods used during the course of experimentation, these are Gaussian Naïve Bayes (GNB), Multi-layer Perceptron (MLP), and SVM. Each classification method had left and right hippocampus pairs, this was to complement the left and right AutoEncoder pairs.

The 3 different classifiers, GNB, MLP, and SVM, were selected based on preliminary testing and their general success/high utilization in certain classification tasks [18,6].

Gaussian Naïve Bayes (GNB) is an efficient, supervised, probabilistic learning algorithm applying a Gaussian variation on Bayes' theorem (Eq. (2)). The Gaussian variation applies the assumption of normal distribution as with the dataset's features being continuous, making GNB the most appropriate for the dataset used.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \tag{2}$$

Multi-layer Perceptron (MLP) is a feedforward artificial neural network used for classification and regression tasks. MLPs consist of multiple hidden layers with multiple hidden nodes in each layer, they are often trained using back-propagation, back-propagation was chosen to train the model as this enabled supervised learning. During the training of the MLP the Adam optimizer and the ReLu activation function were used.

Adam optimizer is seen as the best of both worlds as it combines parts of RMSProp and Stochastic Gradient Descent (SGD) with momentum optimizers, from RMSProp it uses squared gradients to scale the learning rate, and from SGD with momentum it uses the moving average of the gradient instead of the gradient itself [10].

Support Vector Machine (SVM) is a supervised learning algorithm for classification and regression tasks. SVMs are highly efficient, simple classifiers, however the efficiency is reduced when used in this papers classification task as prediction probabilities are required for the aggregation layer, to allow for this an expensive five-fold cross validation process is applied.

3.4 Aggregation

The aggregation layer consists of a soft voting process where each classification method pair is combined to achieve a final prediction, allowing for the left and right hippocampi to have equal influence on the final prediction. The soft voting process is the average of each of the models probabilities with the class with the highest probability being chosen (Eq. (3)).

$$\widehat{y} = \arg\max_i \sum_{j=1}^{m} w_j p_{ij} \tag{3}$$

4 Results and Discussions

4.1 Results

Each classification method was run 10 times with the average of each methods being recorded. As can be seen in Table 2, the accuracy of the models are fairly consistent at $83\% \pm 9\%$ over both training and validation sets, while sensitivity

(true positive rate) is concerning with 2 of the classification methods, MLP and SVM, leaving an average percentage error of $53\% \pm 33\%$ over both training and validation sets.

As can be seen in Table 3 the AUC values for each classification methods are extremely good for both the left and right hippocampus, this shows that the models, before aggregation, are highly capable of differentiating between both classes.

Table 2: Average results over 10 runs foreach of the classification methods

Model	**MLP**		**SVM**		**GNB**	
	Train	**Valid**	**Train**	**Valid**	**Train**	**Valid**
Acc (Avg)	.82	.87	.85	.90	.74	.81
Sen (Avg)	.20	.52	.26	.59	.81	.80
Spe (Avg)	.96	.95	.99	.98	.72	.81
BAcc (Avg)	.58	.73	.63	.78	.76	.80

Table 3: Area Under Curve (AUC) Snapshot of 10^{th} run of the classification methods

Model	**MLP**		**SVM**		**GNB**	
	Train	**Valid**	**Train**	**Valid**	**Train**	**Valid**
AUC (Left)	.93	.90	.99	.89	.93	.88
AUC (Right)	.94	.91	.99	.92	.93	.89

The performance of the aggregated MLP model is highly biased to specificity (true negative rate) with an average of 96%, while sensitivity suffers at 20% on the training set, the model does perform much better on the validation set with sensitivity/specificity at 52%/95%. The balanced accuracy - a metric that is the average of sensitivity and specificity to adjust for the imbalance in the dataset - achieves 53%/73% in the training/validation sets.

The performance of the aggregated SVM model are slightly better with accuracy reaching 85%/90% on the train/validation set. Sensitivity/specificity also peak at 59%/98% respectively on the validation set. Balanced accuracy, influenced by the increased sensitivity compared to MLP, achieved 63%/78% on the training/validation sets.

The aggregated GNB out performed MLP and SVM with sensitivity/specificity reaching and average 80%/81% on the validation set. Accuracy at an average of 74%/81% has been sacrificed compared to MLP and SVM, whereas balanced accuracy, influenced by the more higher sensitivity and specificity scores, is higher at 76%/80% training/validation sets.

4.2 Discussions

Feature extraction has been successful using the unsupervised learning based AutoEncoders, this is evident by the AUC scores in Table 3 where each of the

models seem to have, almost, excellent performance in differentiating between the AD and CN classes. The AutoEncoders never once fixated on the wrong features in the MRI images supplied this is most likely due to the highly localised cropping of the images before the unsupervised learning process (Figs. 1b and 1c).

While Pre-aggregation classification full results were not recorded a snapshot of AUC were recorded, through the AUC scores some partial assessments can be made of the performance of the models at this point. It is evident that the left and right hippocampi models pre-aggregation perform extremely well in their classification task, whereas once the aggregation layer has performed it's task the MLP and SVM aggregated models started to decline in accuracy.

Post-aggregation, the MLP and SVM aggregated models started to produce a high bias towards CN classification, it is probable that this happened due to the imbalance in the dataset where the training/validation sets have 4.17/4.13 times as many CN records as AD records, respectively. It is also probable that the pre-aggregation SVM models were mis-reporting the prediction probabilities, as SVM does not natively support prediction probabilities and this functionality was artificially introduced to the SVM. With regards to the GNB aggregated models, they didn't seem to have the same issues as MLP or SVM, as they kept both high sensitivity and specificity scores.

5 Conclusion

This paper has proposed a parallel machine learning framework where feature extraction (by AutoEncoder) and classification (by Multi-layer Perceptron, Support Vector Machine, and Gaussian Naïve Bayes) are trained in pairs through the left/right hippocampi of an image, before aggregating decisions through a soft voting ensemble algorithm.

Overall the performance of the best averaged post-aggregation GNB model is able to match, and succeed in some cases, that of other models proposed [7,6,8,19] in the same area while only taking into account the single data point of T1-weighted MRI scans localised to the hippocampi and not discounting training data of those 74 and under. With the post-aggregation GNB model proposed in this paper achieving exceedingly high sensitivity score, it is highly beneficial to the medical field where an incorrectly predicted positive result is much more detrimental than an incorrectly predicted negative result.

Future work may considering incorporating fuzzy logic system [2,5] to enhance the capability of working with imprecision and uncertainly that commonly exist in the medical imaging.

6 Acknowledgements

The authors are grateful to the reviewers for their constructive comments, which have helped elevate this piece of work significantly. The first author would also like to thank the University of Huddersfield for providing the PhD scholarship in support of their research. Part of this piece of work is also supported by the Santander International Research Funding Scheme and National Natural Science Foundation of China under grant 61906181.

References

1. Alzheimer's Association: What Causes Alzheimer's Disease? | Alzheimer's Association, https://www.alz.org/alzheimers-dementia/what-is-alzheimers/causes-and-risk-factors
2. Chen, T., Shang, C., Yang, J., Li, F., Shen, Q.: A new approach for transformation-based fuzzy rule interpolation. IEEE Transactions on Fuzzy Systems pp. 1–1 (2019). https://doi.org/10.1109/TFUZZ.2019.2949767
3. Chen, T., Antoniou, G., Adamou, M., Tachmazidis, I., Su, P.: Automatic diagnosis of attention deficit hyperactivity disorder using machine learning. Applied Artificial Intelligence (2019)
4. Chen, T., Shang, C., Su, P., Keravnou-Papailiou, E., Zhao, Y., Antoniou, G., Shen, Q.: A decision tree-initialised neuro-fuzzy approach for clinical decision support. Artificial Intelligence in Medicine **111**, 101986 (2021)
5. Chen, T., Shang, C., Su, P., Shen, Q.: Induction of accurate and interpretable fuzzy rules from preliminary crisp representation. Knowledge-Based Systems **146**, 152–166 (2018)
6. Cho, Y., Seong, J.K., Jeong, Y., Shin, S.Y.: Individual subject classification for Alzheimer's disease based on incremental learning using a spatial frequency representation of cortical thickness data. NeuroImage **59**, 2217–2230 (2012). https://doi.org/10.1016/j.neuroimage.2011.09.085, http://adni.loni.ucla.edu/wp-content/uploads/how
7. Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O.: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus **19**(6), 579–587 (jun 2009). https://doi.org/10.1002/hipo.20626, http://doi.wiley.com/10.1002/hipo.20626
8. Coupé, P., Eskildsen, S.F., Manjón, J.V., Fonov, V.S., Collins, D.L.: Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. NeuroImage **59**(4), 3736–3747 (feb 2012). https://doi.org/10.1016/j.neuroimage.2011.10.080, https://linkinghub.elsevier.com/retrieve/pii/S1053811911012444
9. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H.: BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease. PLoS ONE **8**(6), e67346 (jun 2013). https://doi.org/10.1371/journal.pone.0067346, https://dx.plos.org/10.1371/journal.pone.0067346
10. Kingma, D.P., Ba, J.L.: Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR (dec 2015), https://arxiv.org/abs/1412.6980v9
11. Mahmud, M., Kaiser, M.S., McGinnity, T.M., Hussain, A.: Deep Learning in Mining Biological Data. Cognitive Computation **13**(1), 1–33 (jan 2021). https://doi.org/10.1007/s12559-020-09773-x, http://link.springer.com/10.1007/s12559-020-09773-x
12. Mahmud, M., Kaiser, M.S., Hussain, A., Vassanelli, S.: Applications of Deep Learning and Reinforcement Learning to Biological Data. IEEE Transactions on Neural Networks and Learning Systems **29**(6), 2063–2079 (jun 2018). https://doi.org/10.1109/TNNLS.2018.2790388, https://ieeexplore.ieee.org/document/8277160/

13. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults. Journal of Cognitive Neuroscience **22**(12), 2677–2684 (dec 2010). https://doi.org/10.1162/jocn.2009.21407, https://direct.mit.edu/jocn/article/22/12/2677-2684/4983
14. Markesbery, W.R.: Neuropathologic alterations in mild cognitive impairment: a review. Journal of Alzheimer's disease : JAD **19**(1), 221–8 (jan 2010). https://doi.org/10.3233/JAD-2010-1220, http://www.ncbi.nlm.nih.gov/pubmed/20061641
15. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J.: Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. NeuroImage **104**, 398–412 (jan 2015). https://doi.org/10.1016/j.neuroimage.2014.10.002
16. Mosconi, L., Brys, M., Glodzik-Sobanska, L., De Santi, S., Rusinek, H., de Leon, M.J.: Early detection of Alzheimer's disease using neuroimaging. Experimental Gerontology **42**(1-2), 129–138 (jan 2007). https://doi.org/10.1016/j.exger.2006.05.016
17. National Health Service (NHS): Alzheimer's disease - Treatment - NHS (2018), https://www.nhs.uk/conditions/alzheimers-disease/treatment/
18. Noor, M.B.T., Zenia, N.Z., Kaiser, M.S., Mamun, S.A., Mahmud, M.: Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia. Brain Informatics **7**(1), 11 (dec 2020). https://doi.org/10.1186/s40708-020-00112-2, https://braininformatics.springeropen.com/articles/10.1186/s40708-020-00112-2
19. Payan, A., Montana, G.: Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks (feb 2015), http://arxiv.org/abs/1502.02506
20. Samper-González, J., Burgos, N., Bottani, S., Fontanella, S., Lu, P., Marcoux, A., Routier, A., Guillon, J., Bacci, M., Wen, J., Bertrand, A., Bertin, H., Habert, M.O., Durrleman, S., Evgeniou, T., Colliot, O.: Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data. NeuroImage **183**, 504–521 (dec 2018). https://doi.org/10.1016/j.neuroimage.2018.08.042, https://linkinghub.elsevier.com/retrieve/pii/S1053811918307407
21. Stirling, J., Chen, T., Bucholc, M.: Diagnosing alzheimer's disease using a self-organising fuzzy classifier. In: Fuzzy Logic: Recent Applications and Developments. Springer, Cham (2020)
22. Su, P., Chen, T., Xie, J., Zheng, Y., Qi, H., Borroni, D., Zhao, Y., Liu, J.: Corneal nerve tortuosity grading via ordered weighted averaging-based feature extraction. Medical Physics **47**(10), 4983–4996 (2020)
23. Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-Gonzalez, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O.: Convolutional Neural Networks for Classification of Alzheimer's Disease: Overview and Reproducible Evaluation. Medical Image Analysis **63** (apr 2019). https://doi.org/10.1016/j.media.2020.101694, http://arxiv.org/abs/1904.07773http://dx.doi.org/10.1016/j.media.2020.101694
24. World Health Organisation: Dementia (2020), https://www.who.int/news-room/fact-sheets/detail/dementia