Signal Processing and Renewable Energy
March 2023, (pp. 57-67)
ISSN: 2588-7327
eISSN: 2588-7335

A Denoising Autoencoder Stacked Deep Learning Method for Clinical Trial Enrichment and Design Applied to Alzheimer's Disease

Aref Safari [1*]

[1] Department of Computer Engineering, Islamic Azad University, Shahr-e-Qods Branch, Tehran, Iran.

Received: 07-Jul-2022, Revised: 03-Dec-2022, Accepted: 03-Dec-2022.

Abstract

In this research, we first present some background on the sample size estimation for conducting clinical trials, discussing the necessity of a computational enrichment criterion. The Denoising Autoencoder Stacked Deep Learning (DASDL) design and development are directly motivated by the optimal enrichment design. Although there are many types of deep architectures in the literature, we focus our presentation using two of the most widely used models stacked denoising autoencoders and fully-supervised dropout. The ideas presented here are applied to any such architectures used for learning problems in the small-sample regime. In this work, we propose a novel, scalable, deep learning method that is applicable for learning problems in the small sample regime and obtains reliable performance. The results show via extensive analyses using imaging, cognitive, and other clinical data alongside a ROC curve analysis. When used as trial inclusion criteria, the new computational markers result in cost-efficient clinical Alzheimer's disease trials with moderate sample sizes.

Keywords: Pattern Analysis, Medical Imaging, Deep Learning, Alzheimer's Disease.

1. INTRODUCTION

Alzheimer's disease (AD) affects over 20 million people worldwide [1], and in the last decade, efforts to identify AD biomarkers have intensified. There is now a broad consensus that the disease pathology manifests in the brain images years before the onset of AD. Various groups have adapted sophisticated machine learning methods to learn patterns of pathology by classifying healthy controls from AD subjects. The

*****Corresponding Authors Email: safari.aref@gmail.com

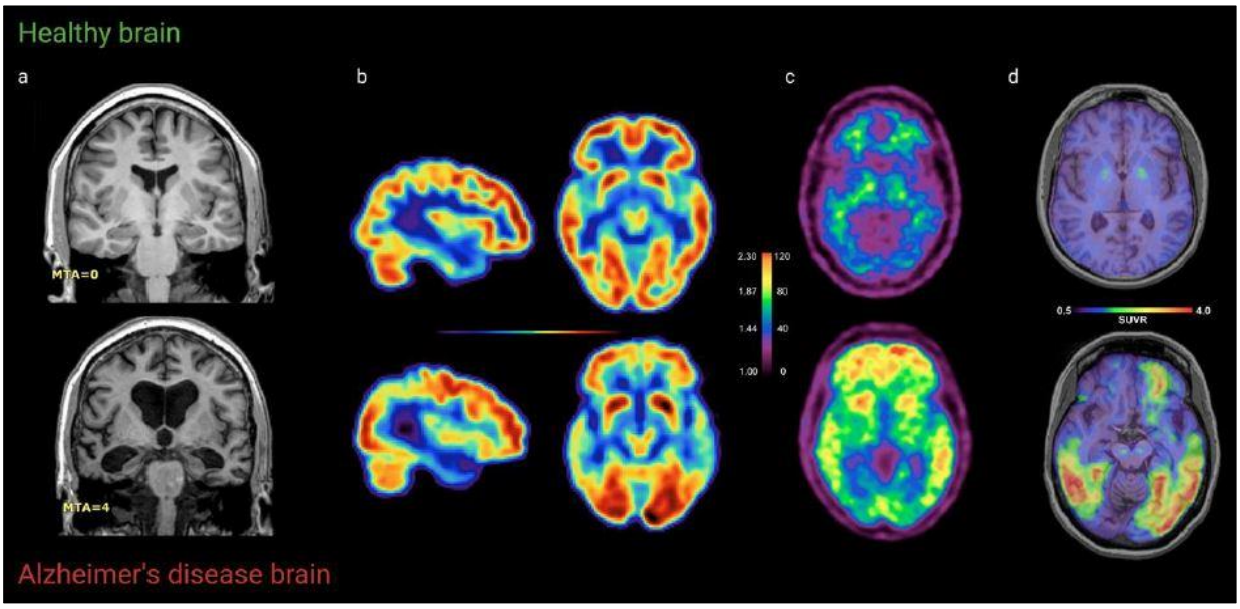

Fig. 1. Neuroimages of the healthy versus the Alzheimer's Disease (AD) brain. Neuroimaging with (a) structural MRI, (b) FDG-PET, (c) amyloid-PET with PiB, and (d) tau PET with 18F-AV1451 in both healthy and AD brains [8].

success of these methods [2] has led to attempts at more fine-grained classification tasks, such as separating controls from Mild Cognitively Impaired (MCI) subjects and even identifying which MCI subjects will go on to develop AD [3,4]. Current methods have reported over 75% accuracy even in this challenging setting. This work is in the context of designing methods for AD trials that are deployable in practice and cost-effective. While accurate classifiers are undoubtedly desirable, one may ask if they address a real practical need. In [5,6] the authors show the utility of computational methods beyond diagnosis and prognosis; based on devious efficient clinical trials for AD.

2. RESEARCH BACKGROUND

Recent clinical trials designed to evaluate new treatments and interventions for AD at the mild to moderate dementia stage have largely been unsuccessful. There is growing consensus that trials should focus on the earlier stages of AD, including MCI or even the pre-symptomatic stage [7,8], if such stages can be accurately identified in individual subjects [9-11]. However, MCI is a clinical syndrome with heterogeneous underlying etymology that may not be readily apparent from a clinical work-up, posing a major challenge in identifying the most likely beneficiaries of a putative effective treatment [12]. For instance, MCIs may have clinical but not biomarker evidence of incipient AD, may have biomarker evidence in some modalities, or, despite biomarker presence, may not show symptomatic progression during the trial period. An efficient MCI trial would ideally include only those patients that are most likely to benefit from treatment; who possess AD pathology based on a

constellation of amyloid, tau, and neural injury biomarker assessments, and who are most likely to progress clinically to symptomatic AD. The typical annual conversion rate to dementia among MCI due to AD is 3–20% across several studies [13]. However, these markers are unimodal, while several studies have shown the efficacy of multimodal data [14]. Furthermore, CSF cannot be used in practice as a screening instrument because assays typically need to be performed in a single batch and are highly lab specific [15]. Several recent studies have used computational multimodal markers derived from support vector machines (SVMs) and other machine learning models [16]. The strategy here uses imaging data from two-time points (i.e., TBM or hippocampus volume change) and derives a machine learning-based biomarker. Based on this marker, say, the top (strongest decliners) one-third of quantile subjects may be selected to be included in the trial. The drug effect can be detected with higher statistical power in the enriched cohort, making the trial more cost-effective and far more accessible to setup/conduct. Most such approaches use longitudinal data; however, a practical enrichment criterion should only use baseline (trial start-point) data. We argue that existing approaches to enrichment, including state-of-the-art computational techniques, cannot guarantee this optimal enrichment behavior – optimally correlate with dementia spectrum with high confidence having access only to the baseline data while simultaneously ensuring slight intra-stage variance.

2.1. Clinical Trials and Enrichment

Consider a randomized clinical trial (RCT) designed to test the efficacy of some treatment for an underlying disease condition. The population under study is randomly assigned to either treatment (or intervention) or non-treatment (or placebo) groups. If the drug improves, the two groups should show this change when measured using some outcome measure summarizing the disease status. Such change would generally correspond to reducing the disease progression to a certain extent, referred to as the effect size [14], in the treatment group compared to the placebos. The outcome measure is, in general, a reliable disease marker. Given such an outcome, the trial efficacy is measured by estimating the Type-II error between the two groups after inducing the drug or intervention. The Type-II error and the effect size would be influenced by the choice of the outcome and the trial population's size (and demographics). Hence, in practice, one would want to "estimate" the trial's efficacy ahead of time to ensure that the effect size is good enough, the population is reasonably large and diverse, the outcome is appropriately chosen, and ensuring that the trial makes sense. In such a hypothetical RCT, the drug is induced by fixing the effect size ahead of time and computing the resulting Type-II error for the given outcome and population. Let δ denote the difference in mean outcome (the standard change) between the trial start and end points (e.g., two years) in the placebos. Let σ be the standard deviation of the outcome, and the effect size be η. δ and σ are a priori known (reported in alternate studies on the disease). The treatment group is then expected to have the change in the outcome decreased to $(1 - \eta)\,\delta$,

which will correspond to a hypothetical improvement of η induced by the drug/treatment. Within this setting, the number of samples s required per arm (treatment and placebo) is given by [31]:

$$s = \frac{2\left(Z_\alpha - Z_{1-\beta}\right)^2 \sigma^2}{(1-\eta)^2 \delta^2} \quad (1)$$

where *(1−β)* denotes the desired statistical power at a significance level of *α*. This expression directly follows from applying a difference of means *t*-test, where the means are computed from the two distributions of interest – outcome change in treatment and placebo groups. The null hypothesis is that the mean change in the outcome is the same for the treatment and placebo groups. Suppose the population under study has a less standard change in the outcome δ. In that case, the required sample *s* for achieving a given power $1 - \beta$ will be very large.

3. METHODS AND MATERIALS

Denoising Autoencoders (DA) and Stacked DA (SDA)

An autoencoder is a single-layer network that learns robust distributed representations of the input data. Given inputs X_i ,the autoencoder learns hidden/latent representations $h_i = \sigma(Wx_i + b)$, such that the reconstructions $\hat{x}_i = \sigma(W^T h_i + c)$ are as close as possible to X_i. It minimizes the following input reconstruction error:

$$\mathcal{Z}_a(\{\mathrm{x}_i\}_1^n, \theta) \coloneqq \arg\min_{\mathrm{W},b,c} \sum_{i=1}^{N} \ell\big(\mathrm{x}_i, \sigma(\mathrm{W}^T \sigma(\mathrm{W}\mathrm{x}_\mathrm{i} + \mathrm{b}) + c)\big) \quad (2)$$

The main novelty of designing the proposed method is to model a stochastic gradient scheme which can be used to

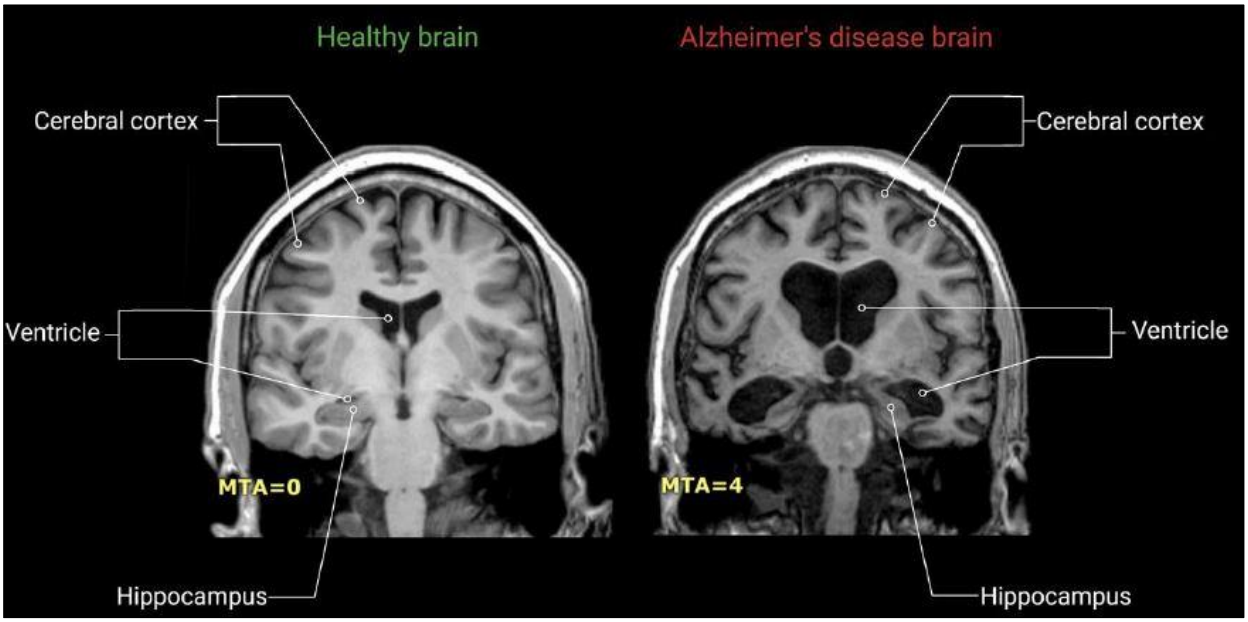

Fig. 2. AD leads to hippocampal atrophy and ventricle enlargement. Healthy brain (left) versus AD brain (right). AD leads to decreased hippocampal volume, shrinkage of the cerebral cortex, and ventricle enlargement. MTA: medial temporal lobe atrophy; MTA = 0: no atrophy in medial temporal lobe; MTA = 4: severe volume loss of hippocampus.

perform this minimization where ℓ denotes a suitable loss function, e.g., squared loss. Without any other constraints, the above minimization could potentially learn identity mappings, i.e., h_is will be identical to the inputs, making the autoencoding setup useless. Several approaches have been suggested to avoid identity mappings and instead learn valuable representations. In this research, we consider the approach where the inputs X_i are corrupted stochastically. It is referred to as the minimization of the error as follows:

$$\mathcal{Z}_{da}(\{\mathrm{x}_i\}_1^n, \theta) \coloneqq \arg\min_{\mathrm{W},b,c} \sum_{i=1}^{n} \mathbb{E}_{\tilde{\mathrm{x}}\sim\gamma(\tilde{\mathrm{x}}|\mathrm{x})} \ell\big(\mathrm{x}_i, \sigma(\mathrm{W}^T\sigma(\mathrm{W}\tilde{\mathrm{x}}_i + b) + c)\big) \quad (3)$$

where γ is a stochastic corruption function, and X_i represents the corrupted form of the $X_i.\gamma(x_{ij}) = x_{ij}$ with some (given) probability ζ and 0 elsewhere. Denoising Autoencoder (DA) is a stochastic autoencoder whose learning procedure seeks to undo the input corruptions (hence it is called denoising). The corruption forces the transformations to correspond to some properties of input data since the reconstruction error decreases only if the transformations pick out the most informative data dimensions. Multiple Das can then be concatenated to construct a stacked DA (SDA), where the hidden representations of lth DA are the uncorrupted inputs to $(l+1)$th. The objective of SDA is:

$$\mathcal{Z}_{sda}(\{\mathrm{x}_i\}_1^n, L, \theta) \coloneqq \sum_{l=0}^{L-1} \mathcal{Z}_{da}\left(\{\mathrm{h}_i^l\}_1^n, \theta\right) \quad (4)$$

$$\mathrm{h}_i^l = \sigma\big(\mathrm{W}^l\mathrm{h}_i^{l-1} + p^l\big); \mathrm{h}_i^0 = \mathrm{x}_i \quad (5)$$

3.1. Small Samples and Multiple Modalities

Although the pretraining idea in tandem with the dropout learning addresses the issue of non-convexity to a certain extent, the earlier works have shown extensive evidence that one of the main reasons for the success of deep learning is the availability of a large number of unsupervised and/or supervised training instances. Simply put, the non-convexity, together with the stochasticity that comes from the corruption in DA or the dropout process, demands a large number of gradient search iterations and data instances to search through the solution space in computing generalizable solutions effectively. With increasing data dimensionality, the dataset size required to ensure that sufficient combinations of corrupted/dropped dimensions are passed to the objective also increases. The fundamental difference between vision type domains and medical imaging and bioinformatics is the lack of such large datasets. In the vision, one can access enormous datasets (on the orders of millions of images), including many unlabeled and supervised instances. On the other hand, a typical voxel-wise imaging study, for instance, will have n < 500 subjects while the number of voxels/features (d) will exceed a million – the classical small-sample regime. The problems in bioinformatics involve data from multiple acquisition types/domains (e.g., brain imaging data including Magnetic Resonance images (MRI), Positron Emission Tomographic

(PET) images, several types of cognitive and neuropsychological scores, lists of vascular and blood perfusion data, and genetic single nucleotide polymorphisms). Most applications in these areas would require efficient statistical models for "fusing" such multimodal data, mainly because they provide specific information about the underlying disease.

3.2. Pattern Analysis

With the presence of multimodal data, the concept is far more complex than the unimodal setting. Using the classical version of deep architectures, including SDAs or dropout networks, for these multimodal problems with d>n issues will result in unreliable outputs, with no guarantee of generating stable or generalizable solutions. This is a direct consequence of under-sampling issues in statistical learning (like in VC-dimension or Nyquist sampling analyses), where the number of training instances cannot be below a certain pre-specified number for efficient estimation of the underlying concepts.

On the other hand, one can avoid the feature screening entirely by working with slices of the input data (e.g., 2D slices or smaller resolution images from a 3D image). Although this is lossless, working with one slice at-a-time will restrict interactions of voxels and brain regions from anatomically far apart regions. Allowing for arbitrary far-apart interactions in predicting the output label would be reasonable following the hypotheses that, in general, all brain regions have complex biological interactions in generating the final label (e.g., the disease status). An alternative to these extremes is to categorize (or tessellate) the entire set of voxels into multiple subsets (e.g., spatially contiguous blocks) and learn a network (e.g., SDA or dropout network) on each block separately while allowing for different blocks to interact with each other in some prescribed manner. Later, the individual block-wise networks, which interact with each other, can be combined in a meaningful manner yielding a better fit for the dependent variable.

4. RESEARCH EXPERIMENTS

Imaging data including [F-18] Florbetapir amyloid PET (AV45) singular uptake value ratios (SUVR), FDG PET SUVRs and gray matter tissue probability maps derived from T1-weighted MRI data, and several neuropsychological measures and CSF values from 516 individuals enrolled in Alzheimer's Disease Neuroimaging Initiative-II (ADNI2) were used in our evaluations. Of these 516 persons (age 72.46 ± 6.8, female 38%), 101 were classified as AD (age 75.5 ± 5.1), 148 as healthy controls (age 70.75 ± 7), and 131 and 136 as early and late MCI (age 74.3 ± 7.1 and 75.9 ± 7.7), respectively, at baseline. There was a significant age difference across the four groups, with $F > 10$ and $p < 0.001$. Among the MCIs, 174 had a positive family history (FH) for dementia, and 141 had at least one Apolipoprotein E (APOE) e4 allele. CSF measures were only available at baseline, and three-time point data (baseline, 12 months, and 24 months) was used for the rest. The imaging protocols follow the standards put forth by ADNI. MRI images are MP-RAGE/IR-SPGR from a 3T scanner. PET images are 3D scans consisting of four 5-min

frames2 from 50 to 70 min post-injection for [F-18] Florbetapir PET and six 5-min frames from 30 to 60 min post-injection for FDG PET. The segmented map was then normalized to Montreal Neurological Institute (MNI) space and smoothed using an 8 mm Gaussian kernel. The resulting map was thresholded at 0.25 to compute the final gray matter image. All PET images were first co-registered to the corresponding T1 images and then normalized to the MNI space. Manually constructed pons, vermis, and cerebellum masks were then used to scale these PET maps by the average intensities in pons and vermis (FDG PET SUVR) and cerebellum (Florbetapir PET SUVR). All preprocessing was done in SPM8.

4.1. Performance Analysis

In this section, a ROC curve analysis was conducted to have a reliable estimate of the performance of the proposed model, and the results were statistically verified as follows [17-19]:

$$\text{Precision} = \frac{TP}{(TP + FP)} \times 100\% \quad (6)$$

$$\text{Recall} = \frac{TP}{(TP + TN)} \times 100\% \quad (7)$$

$$\text{F-Measure} = \frac{2 \text{ "Precision } ^{*} \text{ Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (9)$$

$$\mu_i = \frac{1}{10} \sum_{k=1}^{10} AUCj \quad (10)$$

Where μ_i is the means of the accuracy of the ROC curve for the 10-fold cross-validation. A 10-fold cross-validation techniques [20-21] was applied, which randomly partitions the original sample into k equal sized subsamples.

A single subsample is retained as the validation data for testing the model from the ten sub-samples, and the remaining nine are used as training data. We train the deep networks based on imaging data from all three modalities, MRI, FDG PET, and AV45 PET with diseased (AD, labeled 0) and healthy (CN, cognitively normal, labeled 1) subjects. We only use baseline imaging data for training, thereby making the models deployable in practice. When testing on MCI subjects, these trained models output a multimodal rDA and rDr, which represent the

Table 1. ***Obtained results of the DASDL method.***

Type	AUC%	CI%	Recall	Precision	F-1
PET rDr	79.23	[77-81]	78.52%	79.67%	80.33%
PER rDr	82.07	[80-85]	81.23%	82.41%	82.77%
MRI rDA	90.92	[87-92]	91.79%	92.24%	90.33%
MRI rDA	92.05	[88-93]	91.76%	92.27%	91.17%

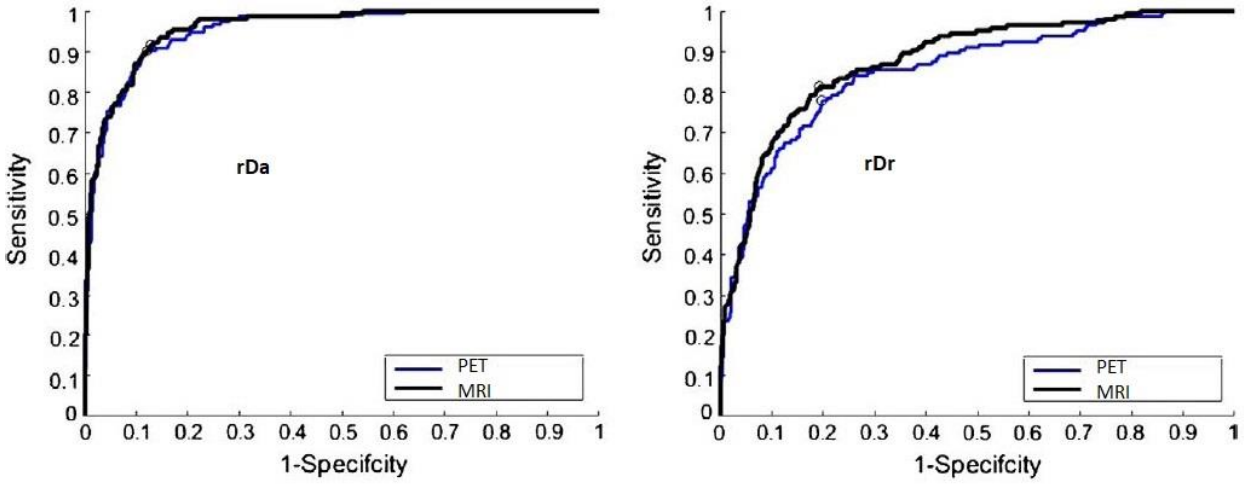

Fig. 3. ROC Curve analysis of the proposed method for different rDA and rDr.

Table 2. ***Left-tailed T-test of the DASDL, LSTM.***

Fold#	DASDL	LSTM
1	0.90	0.72
2	0.92	0.74
3	0.92	0.61
4	0.89	0.62
5	0.93	0.72
6	0.94	0.76
7	0.94	0.79
8	0.96	0.72
9	0.96	0.82
10	0.98	0.81
Mean	**0.96**	**0.82**

confidence of rDA and rDr that a given MCI subject is (or is not) likely to decline. At the same time, the predictions can be performed on MCIs at baseline and/or future time-points. After checking that multimodal markers are superior to unimodal markers, we evaluate whether the multimodal rDA and rDr markers are good disease progression markers. We demonstrate this by computing the dependence of this multimodal baseline rDA and rDr with well-known outcome measures, including the following presentations.

4.2. Null hypothesis and left-tailed t-test

In this section, a two-sample t-test (left tailed) has been conducted. The null hypothesis was defined as $H_0 = \mu_i > \mu_j$ and, $H_1: \mu_i < \mu_j$, where μ_i and μ_j are the means of the area under the ROC curve (AUC) of DASDL and LSTM for ten different iterations of 10-fold cross-validation assessment [22-26]. The obtained results of the t-test analysis are shown in Table 2, which demonstrates the advantage of the proposed DASDL method compared to the LSTM. The results in Table 2 show that the *t-test* failed to reject the null hypothesis.

5. CONCLUSION

The ability to design clinical trials with smaller sample sizes but sufficient statistical

power will enable the implementation of affordable, tractable, and hopefully conclusive trials. Efficiency is seriously compromised in trials where there is poor biomarker specificity of disease progression and when the outcomes contain relatively high amounts of error variance. Determining whether promising treatments are effective in the MCI phase of AD requires accurate identification and inclusion of only those MCI participants most likely to convert to AD and selecting outcomes that are disease-related and possess optimal measurement properties. This research shows that the proposed inclusion strategy can substantially reduce the sample size required to detect a treatment effect. The central message of our empirical evaluations is that the multimodal markers based on our proposed stacked deep network learning models have good predictive power in identifying future disease progression, as shown in Table 1, Table 2, and Figure 3. An interesting extension would be to incorporate multimodal and multi-domain (e.g., ordinal, continuous, and nominal) information directly into the rDA or rDr construction leading to multi-variate randomized deep network models. These technical issues are of independent interest and will be investigated in future works.

REFERENCES

[1] T. J. Weuve, L.E. Hebert, P.A. Scherr, D.A. Evans, Deaths in the United States among persons with Alzheimer's disease, Alzheimer's Dement. 10 (2) (2014).

[2] D. Zhang, Y. Wang, L. Zhou, et al., Multimodal classification of Alzheimer's disease and mild cognitive impairment, NeuroImage 55 (3) (2011) pp856–867.

[3] S.J. Teipel, C. Born, M. Ewers, A.L. Bokde, M.F. Reiser, H.J. Möller, H. Hampel, Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment, NeuroImage 38 (2007) 13–24.

[4] Safari, A. (2021). Deep-Quantitative Medical Image Analysis Methods Applied on Brain Tumor Diagnosis. *Signal Processing and Renewable Energy*, 5(2), 1-13.

[5] C. Hinrichs, V. Singh, G. Xu, S.C. Johnson, Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population, NeuroImage 55 (2011) 574–589.

[6] C. Hinrichs, N. Dowling, S. Johnson, V. Singh, MKL-based sample enrichment and customized outcomes enable smaller ad clinical trials, in: MLINI, in: Lect. Notes Comput. Sci., vol. 7263, Springer, Berlin, Heidelberg, ISBN 978-3-642-34712-2, 2012, pp. 124–131.

[7] O. Kohannim, X. Hua, D.P. Hibar, S. Lee, Y.Y. Chou, A.W. Toga, C.R. Jack Jr., M.W. Weiner, P.M. Thompson, boosting power for clinical trials using classifiers based on multiple biomarkers, Neurobiol. Aging 31 (2010) 1429–1442.

[8] J.D. Grill, L. Di, P.H. Lu, C. Lee, J. Ringman, L.G. Apostolova, et al., Estimating sample sizes for predementia Alzheimer's trials based

on the Alzheimer's Disease Neuroimaging Initiative, Neurobiol. Aging 34 (2013) 62–72.

[9] Safari, A. (2019). MCSM-DEEP: A Multi-Class Soft-Max Deep Learning Classifier for Image Recognition. *Journal of Advances in Computer Research*, 10(4), 75-85.

[10] J.D. Grill, S.E. Monsell, Choosing Alzheimer's disease prevention clinical trial populations, Neurobiol. Aging 35 (3) (2014) 466–471.

[11] V. Jelic, M. Kivipelto, B.Winblad, Clinical trials in mild cognitive impairment: lessons for the future, J. Neurol. Neurosurg. Psychiatry 77 (4) (2006) 429–438.

[12] R.C. Petersen, Mild cognitive impairment: current research and clinical implications, in: Seminars in Neurology, vol. 27, 2007, pp. 22–31.

[13] P.S. Aisen, Clinical trial methodologies for disease-modifying therapeutic approaches, Neurobiol. Aging 32 (2011) S64–S66.

[14] Safari, A., Mazinani, M., Hosseini, R. (2017). A Novel Type-2 Adaptive Neuro-Fuzzy Inference System Classifier for Modelling Uncertainty in Prediction of Air Pollution Disaster (RESEARCH NOTE). International Journal of Engineering, 30(11), 1746-1751.

[15] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, A. Gamst D.M. Holtzman, W.J. Jagust, R.C. Petersen, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic for Alzheimer's disease, Alzheimer's Dement. 7 (3) (2011) 270–279.

[16] A.J. Mitchell, M. Shiri-Feshki, Rate of progression of mild cognitive impairment to dementia meta-analysis of 41 robust inception cohort studies, Acta Psychiatr. Scand. 119 (4) (2009) 252–265.

[17] M. Lorenzi, M. Donohue, D. Paternico, C. Scarpazza, S. Ostrowitzki, O. Blin, E. Irving, G. Frisoni, A.D.N. Initiative, et al., Enrichment through biomarkers in clinical trials of Alzheimer's drugs in patients with mild cognitive impairment, Neurobiol. Aging 31 (8) (2010) 1443–1451.

[18] J.M.S. Leoutsakos, A.L. Bartlett, S.N. Forrester, C.G. Lyketsos, Simulating effects of biomarker enrichment on Alzheimer's disease prevention trials: conceptual framework and example, Alzheimer's Dement. 10 (2) (2014) 152–161.

[19] N. Mattsson, U. Andreasson, S. Persson, M.C. Carrillo, S. Collins, S. Chalbot, N. Cutler, D. Dufour-Rainfray, A.M. Fagan, N.H. Heegaard, et al., CSF biomarker variability in the Alzheimer's Association quality control program, Alzheimer's Dement. 9 (3) (2013) 251–261.

[20] Aref Safari, Danial Barazandeh, Seyed Ali Khalegh Pour. A Novel Fuzzy-C Means Image Segmentation Mode for MRI Brain Tumor Diagnosis. J. ADV COMP ENG TECHNOL, 6(1) Winter 2020: 19-2

[21] Safari A, Hosseini R, Mazinani

M. Dynamic Type-2 Fuzzy Time Warping (DT2FTW): A Hybrid Model for Uncertain Time-Series Prediction. IJFIS 2021; 21:338 348.

[22] A.Safari, R. Hosseini and M. Mazinani, "A Type-2 Fuzzy Time Series Model for Pattern Similarity Analysis: A Case Study on Air Quality Forecasting," in IEEE Intelligent Systems, (2021).

[23] A.Safari, R. Hosseini, M. Mazinani, "A novel deep interval type-2 fuzzy LSTM (DIT2FLSTM) model applied to COVID-19 pandemic time-series prediction", Journal of Biomedical Informatics, Vol 123, (2021).

[24] Safari, A., Hosseini, R., Mazinani, M. (2022). An Adaptive Intelligent Type-2 Fuzzy Logic Model to Manage Uncertainty of Short and Long Time-Series in Covid-19 Patterns Prediction: A Case Study on Iran. Computational Intelligence in Electrical Engineering.

[25] Safari, A. (2022). A Technical Review on Unsupervised Learning of Graph and Hypergraph Pattern Analysis. Journal of Computer & Robotics, 15(2), 27-37.

[26] Safari, A., & Hosseini, R. (2022). An Interval Type-2 Fuzzy LSTM Algorithm for Modeling Environmental Time-Series Prediction. Anthropogenic Pollution, 6(2), 62-72. doi: 10.22034/ap.2022.1963124.1133