Regular Article

Enhancing Feature Selection in MCI Diagnosis using FDG-PET Images: Leveraging Multiple Simple Autoencoder Architectures

Pham Minh Tuan[1,2,3], **Mouloud Adel**[1,2,4], **Nguyen Linh Trung**[3], **Eric Guedj**[1,2]

[1] CNRS, Centrale Med, Institut Fresnel, Aix-Marseille University, Marseille, France
[2] Institut Marseille Imaging, Marseille, France
[3] University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
[4] Department of Computer Engineering, Galatasaray University, Ciragan Cad. No: 36, Ortakoy, 34349, Istanbul, Turkey

Correspondence: Nguyen Linh Trung, linhtrung@vnu.edu.vn
Communication: received 19 April 2024, revised 15 June 2024, accepted 20 June 2024
Online publication: 24 June 2024, Digital Object Identifier: 10.21553/rev-jec.374

Abstract– **Alzheimer's Disease (AD) is the most common type of neurodegenerative brain disease in elderly people. Early diagnosis of AD is crucial for providing suitable care. Positron Emission Tomography (PET) images and machine learning can be used to support this purpose. In this paper, we present a method for ranking the effectiveness of brain regions of interest (ROIs) to distinguish between stable mild cognitive impairment (sMCI) from progressive mild cognitive impairment (pMCI) in brain PET images based on AutoEncoder (AE). Experiments on the ADNI dataset show that our proposed method significantly improves classifier performance when compared to other popular feature ranking methods such as Fisher score, t-score, and LASSO. Our results suggest that instead of focusing on designing a complex AE structure, we can also use simple-but-multiple AEs for feature ranking. The proposed method could be easily applied to any image dataset where a feature selection is needed.**

Keywords– **AutoEncoder, feature ranking, Alzheimer's disease, mild cognitive impairment, positron emission tomography.**

1 Introduction

Alzheimer Disease (AD) is the most common form of neurodegenerative brain disease in elderly people. According to a recent study, 50 million people worldwide were living with AD in 2018, and this number is expected to triple by 2050 [1]. Unfortunately, AD is incurable, and there are no vaccines or effective treatments currently available. However, early detection of AD is crucial to providing appropriate care for patients. Biomarkers such as Fluorodeoxyglucose Positron Emission Tomography (FDG-PET), a nuclear medicine functional imaging that uses 18F-FDG as a radiotracer, have been found to be effective in the early diagnosis of AD. In recent years, a variety of studies have been interested in using machine learning and FDG-PET to distinguish AD from non-AD group (Normal Control – NC) [2] or to predict Mild Cognitive Impairment (MCI) conversion [3–5]. While it is easy to observe differences between AD and NC in FDG-PET brain images, distinguishing between stable MCI (sMCI) and progressive MCI (pMCI) is a more challenging task since typical changes are subtle. Nevertheless, diagnosing AD at an early stage is still a crucial task in order to provide a more reliable diagnosis, and early interventions to slow down AD progression.

When using FDG-PET images for AD diagnosis, previous studies used either voxel-based (whole brain) approach or region-based approach. The first approach uses all voxels as features, as studied in [6, 7]. Meanwhile the second one segments the brain into different regions and then extracts features from those regions, as studied in [8, 9]. The region-based approach is preferable, since it can help reduce the dimension of data and the effect of noise as well as facilitate further image analysis. When working with an ROI-based approach and a classification task, in order to achieve a good performance, a classifier should be trained on informative features which capture different patterns of structural degeneration [10]. To obtain such an informative feature set, either feature extraction or feature selection can be used. In clinical applications, feature selection is preferable since its ability to retain the physical meanings of original features.

The literature on feature selection shows a variety of approaches, which can be categorized into: filter, wrapper and embedded methods, depending on their selection strategies [11]. Filter methods rely on a ranking criterion to measure the importance of each feature. Wrapper methods usually select different subsets of features, measure their performance by a classifier to find the best subset, while embedded methods integrate feature selection into the learning process of the model. Recently, with developments of deep learning, there have been a lot of studies using neural networks for feature selections, particular the Autoencoder (AE). In [12], authors have incorporated autoencoder regression and $\ell_{2,1}$ regularization to select a subset of most useful features. In [13], authors proposed the Agnostic

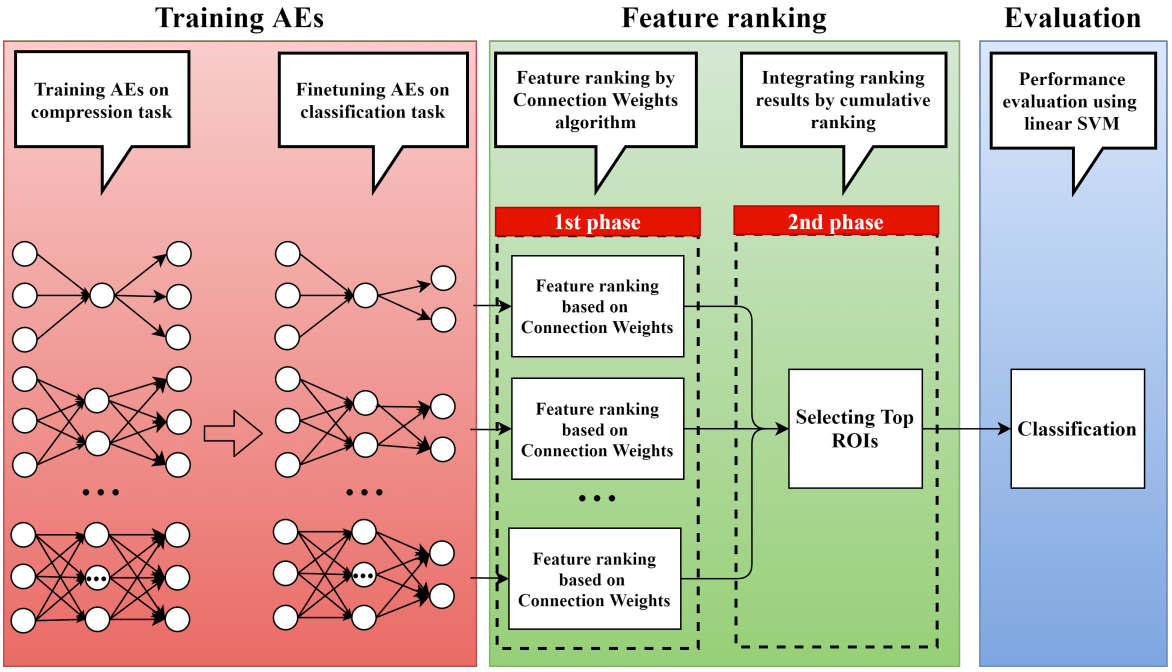

Figure 1. Flow chart of the proposed method.

feature selection algorithm, combining an AE with structural regularizations to perform feature selection. In [14], authors have replaced the first hidden layer of AE with a "concrete selector" layer, which is the relaxation of a discrete distribution, and features are selected based on their probability of connecting to the nodes of the concrete layer. Instead of embedding feature selection in the learning process of AE, in a recent work [15], a two-phase AE-based feature ranking (AEFR) method was proposed, in which feature ranking is performed based on learning capabilities of different simple AE structures. While the obtained results were quite interesting, the proposed methods still need to be validated on other datasets. Moreover, it also has limitations in the second phase of the ranking algorithm, as one needs to specify several parameters.

To overcome these limitations and verify the efficiency of the proposed method, in this work, we propose several improvements for our feature ranking method, and apply it to a more challenging task – MCI conversion diagnosis. More specifically, our contributions are:

- Improving the second phase of the ranking algorithm of [15] (parameter-free version).
- Improving the way for training AE, making the learning process more dependent on AE structures.
- Applying our method to MCI classification, which is a more challenging task.

Part of this work was presented in [16].

The remainder of this paper is organized as follows. Section 2 describes the proposed method and the used dataset. Section 3 presents experimental results and comparisons. Conclusions and future work are then given in Section 4.

2 Proposed Method

The flowchart of the proposed method is shown in Figure 1. This flowchart can be divided into several blocks:

- Training Autoencoder models (in red).
- Two-phase feature ranking method (in green).
- Performance Evaluation (in blue).

In the following subsections, we introduce the proposed method in detail.

2.1 Dataset

Data used in the experiment is 18F-FDG PET images downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In the ADNI dataset, participants can take several scans at different time points, the first time of scan refers to a baseline scan. Time points after baseline are considered as follow-up time. In this work, we only focus on predicting MCI conversion, data is selected based on following criteria:

- sMCI - subjects diagnosed as MCI at the baseline and did not convert to AD in the follow-up period at least 24 months.
- pMCI - subjects diagnosed as MCI at the baseline and progressed to AD in the available scan time.

After downloading the data, images are spatially normalized to MNI template using SPM12 [17] with $2 \times 2 \times 2$ mm^3 voxel size, $91 \times 109 \times 91$ tensor dimension. The intensity normalization is then performed through dividing each voxel intensity by the average value of the cerebellum region. After that, images are further smoothed by a Gaussian kernel with a full width at half maximum of 8 mm. All the procedures above are implemented with SPM12. After the preprocessing phase, we have a total 569 FDG-PET images for later experiments, including 209 pMCI

Table I
Demographic and Clinical Information

Characteristic	sMCI	pMCI
Number of subjects	360	209
Female/Male	153/207	87/122
Age (Mean ± std.)	71.73 ± 7.66	73.89 ± 6.88
MMSE (Mean ± std.)	28.20 ± 1.59	27.13 ± 1.71

subjects, and 360 sMCI subjects. The demographic and clinical information of subjects is provided in Table I, in which MMSE stands for the Mini-Mental State Examination. After that, each subject image is segmented into 120 ROIs using AAL atlas [18]. Each region is then characterized by mean value. Since cerebellum is used for normalizing data, we will remove those regions from our data, after this step for each image, we obtain a vector with 94 features, corresponding to the mean intensity of 94 ROIs.

2.2 Two-phase Feature Ranking Method

2.2.1 Autoencoder Training:
An autoencoder [19] is a type of neural network used for learning efficient representation of unlabelled data, in compressed or sparse form. An AE usually consists of an input layer, an output layer and at least one hidden layer. The complexity of the AE network depends on the number of hidden layers and learning techniques. In our work, we only use simple AEs with one hidden layer. Instead of designing a single complex AE structure, we focus on single-but-multiple AEs. This approach allows us to evaluate the stability of the features within minor changes in the network architecture during the feature ranking phase. Specifically, we train 94 AE models with the number of hidden units varying from 1 to 94 on compression tasks. After that, we replace the output layer of AEs by softmax layer and fine-tuned autoencoder models on the sMCI/pMCI classification task.

In comparison to the previous work in [15], to ensure that the ranking results are more heavily dependent on the structure of AE models, we fix the connection weights between the hidden-output layers during the AE training phase for the compression task. This constraint ensures that the learning process for compression tasks only happens between input-hidden layers. Therefore, in the finetuning phase, we can replace the output layer of AEs with a softmax layer without losing any learned information from the compression phase. By this approach, we expect that the learned weights will be more dependent on AE structures, as well as more adapted for ranking problems.

2.2.2 Ranking Methods:
In order to rank all features, we follow a ranking pipeline that involves two phases: (i) performing ranking process on each AE model, and (ii) integrating ranking results from multiple AE models. Specifically, after training AE models on compression tasks and finetuning them on MCI classification problems, we perform the first phase in our ranking algorithm. For each finetuned AE model, we use their learned connection weights and modified version of connection weight algorithm [15] to rank all features, given by

$$c_i = \sum_{j=1}^{p} |W_{ij}^1| \times \sum_{k=1}^{m} |W_{jk}^2|, \quad (1)$$

where c_i is the contribution of the i-th feature made to the output layer, n, p and m are the number of units in the input layer, hidden layer, and output layer, respectively, W_{ij}^1 denotes the parameter of the connection between the i-th neuron in the input layer and the j-th neuron in the hidden layer, W_{jk}^2 denotes the parameter of the connection between the j-th neuron in the hidden layer and the k-th neuron in the output layer. The higher value of c_i indicates the higher importance of the i-th feature. After this step, for 94 AE models, we obtained a ranking matrix $\mathbf{R}$ in which its element r_{ij} corresponds to the j-th most important feature ranked by the finetuned model AE_i.

In order to combine results from multiple AEs into a single outcome, we continue with the second phase of our ranking algorithm. We make several improvements to this phase compared to the previous work in [15], resulting in a parameter-free version of our algorithm. Specifically, we propose a cumulative feature ranking algorithm, which is described in Algorithm 1. Given the ranking matrix $\mathbf{R}$ of size $n \times n$, where n is the total number of features, we extract the first j columns of $\mathbf{R}$ to form the submatrix $\mathbf{R}_j$. Firstly, we compute the frequency matrix $\mathbf{F}$ in which its element f_{ij} equals to frequency of the i-th feature in the submatrix $\mathbf{R}_j$. After this step, we compute the cumulative frequency for each feature, $\mathbf{C} = \{c_i\}$, where $c_i = \sum_{j=1}^{n} f_{ij}$. Then, we rank the features in the descending order according to their corresponding cumulative frequency. A higher value of the cumulative frequency indicates a higher rank of the feature.

Algorithm 1 Parameter-free version of the 2nd phase in the ranking algorithm

Input: Matrix $\mathbf{R}_{n\times n}$ corresponds to ranking order of n features given by n AutoEncoder models, submatrix $\mathbf{R}_j$ contains first j columns of $\mathbf{R}$
Output: Feature ranking list

Algorithm:
1. Compute frequency matrix $\mathbf{F}_{n\times n} = \{f_{ij}\}$ where f_{ij} is the frequency of i-th feature in the submatrix $\mathbf{R}_j$
2. Compute cumulative frequency $\mathbf{C} = \{c_i\}$ where $c_i = \sum_{j=1}^{n} f_{ij}$
3. Ranking n features in descending order according to their cumulative frequency.

The second phase of the ranking algorithm is based on the concept of cumulative information, designed to aggregate information from multiple autoencoder models. The motivation is to make the ranking results more accurate and robust. By leveraging cumulative frequency concepts, our algorithm evolves into a parameter-free approach. The complexity of the algorithm is less than $\mathcal{O}(n^2)$, since we only need to compute the frequency of features in every submatrix $\mathbf{R}_j$.

Table II
Performance on sMCI/pMCI Classification Task

	Original		Fisher		t-score		LASSO		AEFR		Proposed method	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
k = 5	63.27	64.50	69.68	74.53	69.82	74.59	66.64	76.45	67.63	75.13	70.19	75.30
k = 10	63.27	65.00	70.93	74.76	70.95	74.80	68.60	77.12	69.79	75.35	70.92	76.33
k = 20	63.28	64.33	71.52	76.23	71.52	76.12	71.16	77.73	71.57	76.86	73.01	78.57
k = 30	63.31	62.38	71.77	76.64	71.80	76.58	71.93	77.79	73.02	78.12	73.93	79.21
k = 40	64.55	68.31	71.80	76.55	71.76	76.51	72.23	77.96	74.27	79.07	74.13	79.39
k = 50	65.73	71.57	71.76	76.56	71.92	76.60	72.56	77.90	74.29	79.24	74.11	79.32
k = 60	66.51	72.32	72.09	76.56	72.08	76.60	72.94	77.91	74.45	79.30	74.24	79.21
k = 70	73.12	78.28	72.33	76.64	72.38	76.68	73.21	78.08	74.37	79.18	74.32	79.23
k = 80	73.70	78.66	73.28	77.40	73.29	77.37	74.04	78.60	74.39	79.16	74.22	79.11
k = 90	73.83	78.60	74.19	78.70	74.16	78.70	74.26	78.83	74.36	79.00	74.20	78.98
k = 94	74.26	78.95	74.26	78.95	74.26	78.95	74.26	78.95	74.26	78.95	74.26	78.95

2.3 Experimental Design and Evaluation Metrics

To evaluate the effectiveness of our proposed method, we compare the performance between our method and other feature ranking methods on the ADNI dataset. We choose Fisher, t-score, and LASSO [11] as the comparative methods, since they are well-known methods for feature ranking. After the ranking phase, for each ranking method, the top k ROIs are selected as candidate features for the classification task. We choose the linear Support Vector Machine (SVM) [20] as the classifier, since SVM is a well-known and widely used technique for AD diagnosis in previous studies. Given by the classifier are the true positive (TP), true negative (TN), false positive (FP) and false negative (FN). The performance is measured in terms of accuracy (ACC), specificity (SPE), sensitivity (SEN) and the area under the Receiver Operating Characteristic curve (AUC), where

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}, \tag{2}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{3}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{4}$$

and AUC is determined by SEN and $1-\mathrm{SPE}$. The higher value of these metrics indicates better performance.

Concerning data allocation, we divide data randomly into two subsets corresponding to two phases: half for feature ranking phase the other half for performance evaluation. Therefore, these two phases are independent learning processes. 5-folds cross validation is used for training and evaluating a classifier. To avoid bias, experiments are repeated for 50 times, and the reported results are averaged values of these different runtimes.

Finally, all experiments are implemented using scikit-learn [21] and Keras. To train the Autoencoders, we utilized the Adam optimizer with a mini-batch size of 32 and conducted training over 1500 epochs. For other experimental parameters, default values set by the libraries were used if not specified.

3 Results and Discussion

3.1 Results

The use of ranking methods significantly improved classifier performance, as shown in Table II and Figure 2. When comparing accuracy of linear SVM classifiers with and without using ranking methods, we observe that feature ranking methods gave a better average accuracy. This is most noticeable in the range from 20 to 40 features, the use of the ranking methods improved ACC by about 8% on average compared to no ranking. Fisher and t-score gave very similar performance and they were the best methods in the range from 1 to 20 features, with an average ACC improvement of approximately 7% compared to no ranking. However, with more than 20 features, LASSO was the better one. Therefore, we chose t-score and LASSO as the baseline when comparing cases with less than 20 features, and more than 20 features, respectively. Here, we expect that a good ranking method should give either similar or better performance than t-score or LASSO when using less or more than 20 features, respectively.

The AE-based ranking methods significantly improved classifier performance, as shown in Figure 3 and Table II. In many cases, ACC was improved by up to 10% as compared to no ranking (e.g., the case with 30 features). AE-based ranking methods also gave better results when compared to the baseline methods, especially for AE models with fixed weights. Interestingly, after more than 30 features, the performance of the classifiers seemed to reach a stable state with AE-based ranking methods, adding features almost did not change the performance much. It means that the AE-based ranking methods showed better convergence results as compared to the baseline ones. Moreover, in the range from 40 to 60 features, there were many cases in which our proposed ranking methods gave better results than using either all 94 features or voxel-based approaches.

The ranking method with fixed AEs gave better performance as compared to normal AEs, and outperformed the baseline methods. More specifically, the

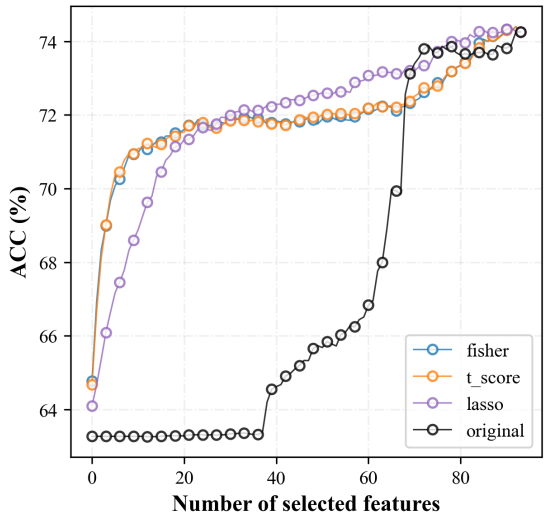

Figure 2. Accuracy of the linear SVM classifiers with and without ranking methods.

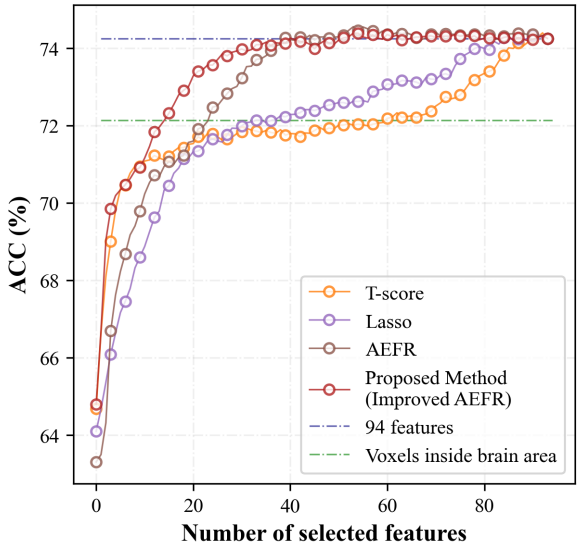

Figure 3. Accuracy of the linear SVM classifiers between our proposed methods and the baseline ones.

top-1 feature selected by the fixed AE models gives a similar performance compared to t-score, and much better than LASSO, and the normal AE method. The performance of the fixed AE models was always approximately or better than t-score in the range from 1 to 20 features. With more than 20 features the obtained results are much better as compared to LASSO. At the threshold of 30 features, the result converged faster than all the baseline methods. In other words, by using fixed AE models, we can achieve remarkable results with the number of features being only 1:3 of the total number of features, and approximately 1:1850 of features when compared to the voxel based approach.

Our proposed method not only significantly improved ACC but also AUC, as shown in Table II and Figure 4. We can observe that ranking methods such as t-score and LASSO improved AUC as compared to original order by an average of 4.6%. The averaged differences in AUC between t-score and LASSO did

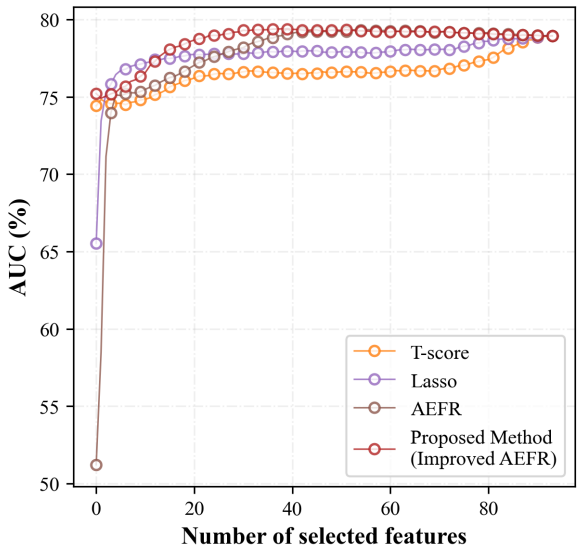

Figure 4. AUC of the linear SVM classifiers between our proposed methods and the baseline ones.

not differ too much. Meanwhile, the ranking methods using fixed AE and normal AE models, respectively, improved AUC approximately 1.5% and 1%, respectively, when compared to t-score. All these observations lead us to a conclusion that the AE-based ranking methods are suitable for feature selection on MCI classification tasks, and the ideas of fixed AE models can significantly improve the overall performance.

In order to explain the improvements of our proposed methods over the others, we consider the overlapped rate between obtained results by each ranking method. Figure 5 compares the overlapped rate between different ranking methods with various numbers of selected features. We only focused on the range from 5 to 60 features. With the range from 5 to 40 features, the overlapped rate between LASSO and t-score was around 35%, and these values increased to reach 60% within the range from 40 to 60 features. Obviously, these observations corresponded to the differences in performances between the two methods. With less than 15 features, the overlapped rates between fixed AE-based and the baseline methods were above 60% compared to t-score, and lower than 50% as compared to LASSO. As the number of features increased in the range from 1 to 15 features, we also observed that overlapped rates between fixed AE-based models and t-score decreased while the values between AE models and LASSO increased. In other words, the ranking results obtained by fixed AE-based modes are quite similar to t-score with a small number of features, and become similar to LASSO in the cases of much larger numbers of features. This observation confirms our early expectation about a good ranking method, which should give either similar or better performance than t-score or LASSO when using less or more than 20 features, respectively. This result also suggests that our fixed-AE method shows good qualities similar to those of both t-score and LASSO.

Moreover, the fixed AE method has its own capabili-

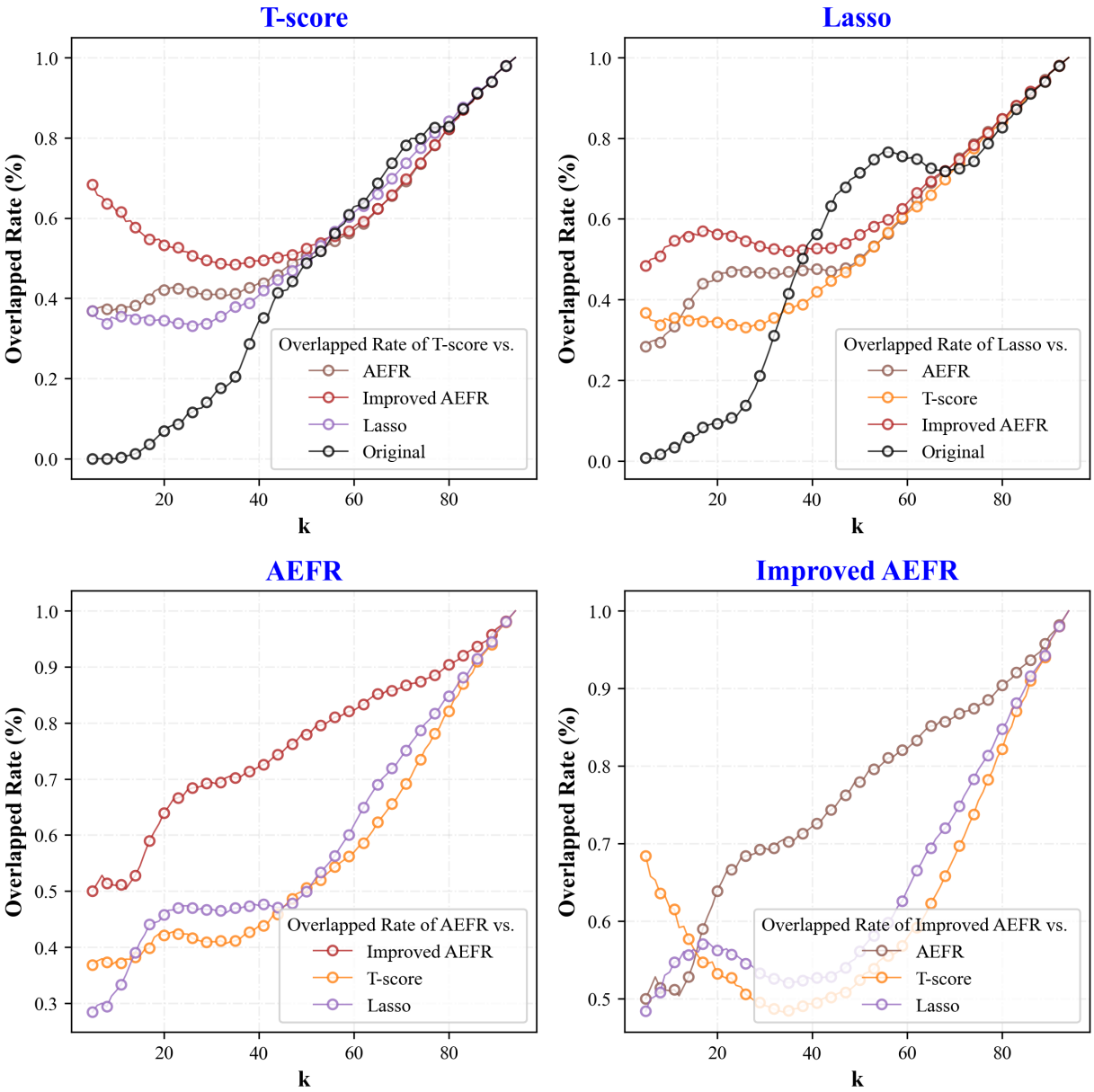

Figure 5. Overlapped rates for different ranking methods.

ties that the baseline methods do not have. Specifically, overlapped rates between the fixed AE method and the baselines were always less than 60% within the range from 20 to 60 features. Meanwhile, the fixed AE method gave a faster convergence results at the threshold of 30 features. It suggests that the fixed-AE method could select some effective features that baselines are not able to do. These effective features not only improve the overall performance, but also contribute to stability and earlier convergence of results.

Fixing connection during training AE models significantly contributed to the improvement of obtained results. Comparing overlapped rates between fixed AE and normal AE models, we observe that the overlapped rate was around 50% in the range from 5 to 20 features. Meanwhile, with more than 20 features, the overlaps were always greater than 65% and continued to increase with the number of features. This observation not only explains the performance differences between the two methods within the range from 1 to 20 features, but also explains why the performance between them is quite similar and converges together with more than 30 selected features. This result seems to support our early expectation of using fixed connection when training AEs for compression tasks. By using fixed connection weights, the learned weights will be more dependent on the AE structures and more adapted for ranking problems.

3.2 Discussion

In this study, we proposed two improvements to our previous AE-based feature ranking work: (i) fixing connection weights during AE training for compression tasks and (ii) enhancing the second phase of the ranking algorithm. Our proposed methods outperformed the baseline methods such as t-score, Fisher,

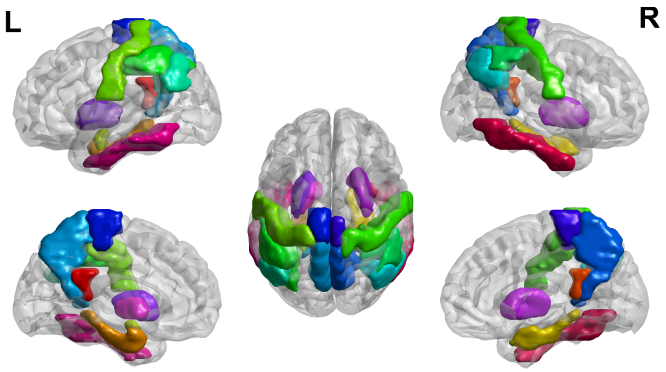

Figure 6. An illustration for the top-20 ROIs selected by our proposed method (with fixing connections).

and LASSO. In particular, fixing connection weights approach significantly improveed ACC and AUC, with early convergence and greater stability compared to both baseline methods and training AEs without fixing. This improvement confirms that restricting the learning process to happening only between input-hidden layers enables the connection weights between input-hidden layers to encode more information. As a result, more knowledge is transferred which leading to improved performance on the MCI classification task and ranking results.

Figure 6 and Table III show the top-20 regions considered most effective in differentiating sMCI from pMCI, this obtained result is consistent with previous studies. More specifically, the top-20 regions mainly belong to the temporal and parietal lobes. These lobes are considered as the standard FDG-PET finding, which are highly correlated to AD pathology [22, 23]. Top-ROIs such as Temporal_Inf, Angular, Precuneus are also selected in previous studies of Pan *et al.* on MCI conversion diagnosis [3], of Tuan *et al.* on AD [15]. Therefore, our results once again confirm the important role of these regions, and suggest that brain ROIs such as Angular, Precuneus can be considered as effective regions for tracking development and changes of AD in different stages.

In addition to similarity in the top-20 ROIs, there were also some significant differences between our obtained results and those of previous studies [3, 15]. Despite sharing several common ROIs, our method gives a different ranking order for each ROI. Specifically, regions such as Angular_L, Angular_R, and Cingulate_Post_L are ranked in the top-3 by our proposed method. However, Angular_L and Angular_R were ranked quite low in [3], while the Cingulate region did not appear in [15]. It is noteworthy that many studies have reported a high association between these brain ROIs and the conversion of MCI to AD [5, 24, 25]. More specifically, with Angular_L and Angular_R belong to the Parietal lobe, which are located in the posterior part of the inferior parietal lobule [26], in [23], authors have reported that hypometabolism in the inferior parietal lobules could be used to predict most reliably the progression from amnestic MCI to AD. Regarding Cingulate, in a recent study, the authors pointed out

Table III
Top 20 ROIs Selected by the Proposed Method

	Top ROIs
1	Angular_L
2	Angular_R
3	Cingulate_Post_L
4	Amygdala_L
5	Temporal_Inf_L
6	Precuneus_R
7	Parietal_Inf_R
8	Parietal_Inf_L
9	Putamen_R
10	ParaHippocampal_L
11	Precuneus_L
12	Pallidum_L
13	Cingulate_Post_R
14	Putamen_L
15	Postcentral_R
16	Paracentral_Lobule_L
17	ParaHippocampal_R
18	Temporal_Inf_R
19	Paracentral_Lobule_R
20	Postcentral_L

a correlation between Cingulate and apathy among MCI patients with and without apathy [27]. Meanwhile, apathy can be considered as a powerful risk factor for conversion to AD in MCI patients [28], especially when it occurs at the early stage of MCI [29]. Therefore, our obtained top-*k* regions seem consistent and coherent when compared to previous studies concerning MCI conversion.

In this study, we used multiple AEs for feature ranking. Experimental results suggest that we can use simple-but-multiple AE models to perform feature selection instead of focusing on single-but-complex ones. By making several small changes in the AE architecture (such as changing the number of neurons, fixing connection weights), and taking these changes into account for the feature ranking, the classifier performance was significantly improved. The improvements can be explained as follows: (i) training AE through two steps, the first one on compression task (unsupervised learning), the next one on classification task (supervised learning), which gives AEs the capability to filter out features that effectively represent both internal properties of the data and external properties related knowledge about classification problem, and (ii) subtle changes between AEs are enough to learn similar knowledge while creating variations for selecting important features with high stability. Together with previous results on ranking using AEs for AD [15], the obtained results suggest: (i) AEs could be used for selecting the most effective ROIs for distinguishing between NC, MCI and AD, and (ii) instead of focusing on complex structures, we can also use simple-but-multiple AE models for feature selection with expectation that the obtained result is still very comparable to traditional feature ranking methods.

In comparison to our previous work, here, we proposed an enhancement to our ranking algorithm, resulting in a parameter-free version, making it easier to apply. While the obtained results are very promising, the approach has a limitation that may hinder its broad application, that is, multiple AE models need to be trained, which raises concerns regarding computational resources. Nonetheless, due to the simple architecture of the AE models, consisting of only one hidden layer, the computational demand is low. Furthermore, fixing connection weights reduces the number of parameters required for training, resulting in a significant decrease in training costs. Moreover, AE models training can be performed in parallel, thereby reducing the training time. So, in general, our proposed method still has a broad applicability.

4 Conclusions and Future Work

In this paper, we have made some improvements to our previous work on AE-based feature ranking. Specifically, we fixed the connection weights between hidden-output layers when training AEs on compression tasks, and improved the second phase of the ranking algorithm. We applied our method to a very challenging task – MCI conversion prediction. Experimental results on the ADNI dataset showed that our method outperforms baseline methods such as t-score and LASSO. For feature selection, these results suggest that instead of focusing on designing a single-but-complex AE architecture, we can also take advantage of the simple-but-multiple AE models to perform the same task. Our proposed method could be easily applied to a dataset where a feature selection is needed. For future work, it is important to apply our proposed method to different datasets to better understand the limitations and strengths of the method. Additionally, considering the influence of other factors such as dropout and activation functions on the obtained results would be an interesting research direction.

Acknowledgment

Pham Minh Tuan was supported by Excellence Initiative of Aix-Marseille University - A*MIDEX, a French "Investissements d'Avenir" program (AMX-19-IET-002) for his PhD scholarship. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2021.55.

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 andor provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

[1] C. Patterson, "World Alzheimer report 2018," Alzheimer's Disease International, Report, Sep. 2018.

[2] A. Alberdi, A. Aztiria, and A. Basarab, "On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey," *Artificial Intelligence in Medicine*, vol. 71, pp. 1–29, Jul. 2016.

[3] X. Pan, M. Adel, C. Fossati, T. Gaidon, and E. Guedj, "Multilevel feature representation of FDG-PET brain images for diagnosing Alzheimer's disease," *IEEE journal of biomedical and health informatics*, vol. 23, no. 4, pp. 1499–1506, 2018.

[4] L. Teng, a. Y. Li, Y. Zhao, T. Hu, Z. Zhang, Z. Yao, and B. Hu, "Predicting MCI progression with FDG-PET and cognitive scores: A longitudinal study," *BMC Neurology*, vol. 20, no. 1, Apr. 2020.

[5] M. Pagani, F. Nobili, S. Morbelli, D. Arnaldi, A. Giuliani, J. Öberg, N. Girtler, A. Brugnolo, A. Picco, M. Bauckneht, R. Piva, A. Chincarini, G. Sambuceti, C. Jonsson, and F. D. Carli, "Early identification of MCI converting to AD: A FDG PET study," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 12, pp. 2042–2052, Jun. 2017.

[6] B. Desgranges, V. Matuszewski, P. Piolino, G. Chételat, F. Mézenge, B. Landeau, V. de la Sayette, S. Belliard, and F. Eustache, "Anatomical and functional alterations in semantic dementia: A voxel-based MRI and PET study," *Neurobiology of Aging*, vol. 28, no. 12, pp. 1904–1913, Dec. 2007.

[7] C. Cabral, P. M. Morgado, D. Campos Costa, and M. Silveira, "Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages," *Computers in Biology and Medicine*, vol. 58, pp. 101–109, Mar. 2015.

[8] K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, and D. Rueckert, "Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease," *NeuroImage*, vol. 60, no. 1, pp. 221–229, Mar. 2012.

[9] I. Garali, M. Adel, S. Bourennane, and E. Guedj, "Histogram-based features selection and volume of interest ranking for brain PET image classification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 6, pp. 1–12, 2018.

[10] E. Varol, B. Gaonkar, G. Erus, R. Schultz, and C. Davatzikos, "Feature ranking based nested support vector machine ensemble for medical image classification," in *Proceedings of the 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, May 2012, pp. 146–149.

[11] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, Nov. 2018.

[12] K. Han, Y. Wang, C. Zhang, C. Li, and C. Xu, "Autoencoder inspired unsupervised feature selection," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2018, pp. 2941–2945.

[13] G. Doquet and M. Sebag, "Agnostic feature selection," in *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2019, pp. 343–358.

[14] A. Abid, M. F. Balin, and J. Zou, "Concrete autoencoders for differentiable feature selection and reconstruction," Jan. 2019.

[15] P. M. Tuan, T.-L. Phan, M. Adel, E. Guedj, and N. L. Trung, "Autoencoder-based feature ranking for Alzheimer disease classification using PET image," *Machine Learning with Applications*, vol. 6, p. 100184, Dec. 2021.

[16] P. M. Tuan, N. L. Trung, M. Adel, and E. Guedj,

"Autoencoder-based feature ranking for predicting mild cognitive impairment conversion using fdg-pet images," in *Proceedings of the 2023 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2023, pp. 720–724.

[17] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, Apr. 2011.

[18] E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer, "Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas," *NeuroImage*, vol. 122, pp. 1–5, Nov. 2015.

[19] G. E. Hinton and R. S. Zemel, "Autoencoders, minimum description length and Helmholtz free energy," in *Proceedings of the Advances in Neural Information Processing Systems*. Advances in neural information processing systems, 1994, pp. 3–10.

[20] C. Cortes and V. Vapnik, "Support vector machine," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine Learning in Python," *Machine Learning in Python*, 2011.

[22] N. I. Bohnen, D. S. Djang, K. Herholz, Y. Anzai, and S. Minoshima, "Effectiveness and safety of 18F-FDG PET in the evaluation of dementia: A review of the recent literature," *Journal of Nuclear Medicine*, vol. 53, no. 1, pp. 59–71, Dec. 2011.

[23] J. M. Hoffman, K. A. Welsh-Bohmer, M. Hanson, B. Crain, C. Hulette, N. Earl, and R. E. Coleman, "FDG PET imaging in patients with pathologically verified dementia," *Journal of Nuclear Medicine*, vol. 41, no. 11, pp. 1920–1928, 2000.

[24] N. Smailagic, M. Vacante, C. Hyde, S. Martin, O. Ukoumunne, and C. Sachpekidis, "18F-FDG PET for the early diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI)," *Cochrane Database of Systematic Reviews*, Jan. 2015.

[25] G. Blazhenets, Y. Ma, A. Sörensen, G. Rücker, F. Schiller, D. Eidelberg, L. Frings, and P. T. Meyer, "Principal components analysis of brain metabolism predicts development of Alzheimer dementia," *Journal of Nuclear Medicine*, vol. 60, no. 6, pp. 837–843, Nov. 2018.

[26] M. L. Seghier, "The angular gyrus," *The Neuroscientist*, vol. 19, no. 1, pp. 43–61, Apr. 2012.

[27] J. Delrieu, T. Desmidt, V. Camus, S. Sourdet, C. Boutoleau-Bretonnière, E. Mullin, B. Vellas, P. Payoux, and T. L. and, "Apathy as a feature of prodromal Alzheimer's disease: An FDG-PET ADNI study," *International Journal of Geriatric Psychiatry*, vol. 30, no. 5, pp. 470–477, Jun. 2014.

[28] P. H. Robert, C. Berr, M. Volteau, C. Bertogliati-Fileau, M. Benoit, O. Guerin, M. Sarazin, S. Legrain, and B. Dubois, "Importance of lack of interest in patients with mild cognitive impairment," *The American Journal of Geriatric Psychiatry*, vol. 16, no. 9, pp. 770–776, Sep. 2008.

[29] R. Duara, D. A. Loewenstein, E. Potter, W. Barker, A. Raj, M. Schoenberg, Y. Wu, J. Banko, H. Potter, M. T. Greig, J. Schinka, and A. Borenstein, "Pre-MCI and MCI: Neuropsychological, clinical, and imaging features and progression rates," *The American Journal of Geriatric Psychiatry*, vol. 19, no. 11, pp. 951–960, Nov. 2011.

Pham Minh Tuan received the Bachelor's degree in information technology from University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam in 2016. He is currently pursuing the Ph.D. degree in Neuroscience as a Researcher with the Doctoral School of Life and Health Sciences, Aix-Marseille Université, France. His current research interests include computer-aided diagnosis for psychiatric and neurological disorders.

Mouloud Adel received his PhD degree in signal and image processing from Institut National Polytechnique de Lorraine in 1994. Since 2014 he is a Professor in Computer Science and Electrical Engineering at Aix-Marseille University, France. His research areas concern signal and image processing and machine learning applied to biomedical and industrial images. He has been involved in many international research programs and he is a member of the editorial board of Journal of Biomedical Engineering and Informatics. He has been an invited speaker at different universities and has been the co-organizer of various international conferences and workshops. He also served as a regular reviewer, associate and guest editor for a number of journals and conferences. Since April 2023, he is a visiting professor at Galatasaray University, Istanbul, Turkey.

Nguyen Linh Trung obtained his B.Eng. and Ph.D. degrees, both in Electrical Engineering, from Queensland University of Technology, Brisbane, Australia, in 1998 and 2005. He joined VNU University of Engineering and Technology, Vietnam National University, Hanoi (VNU), in 2006, where he is currently an associate professor of electronic engineering and the director of the Advanced Institute of Engineering and Technology. His broad technical interests are theory and methods of signal processing, including time-frequency signal analysis, blind source separation, blind system identification, compressed sensing, tensor-based signal analysis, machine learning, and application of signal processing in telecommunications and medicine, with a current focus on large-scale processing.

Eric Guedj is Professor of Biophysics and Nuclear Medicine at Timone hospital and Aix-Marseille University, Head of the Nuclear Medicine Department. He received a double training of Engineer and physician in Nuclear Medicine, with a PhD in Neuroscience. His field of medical/research interest is in brain molecular imaging for neurological diseases, psychiatric diseases, and tumors.