# Regression and Classification of Alzheimer's Disease Diagnosis Using NMF-TDNet Features From 3D Brain MR Image

Huan Lao and Xuejun Zhang

*Abstract*—**With the development of deep learning and medical imaging technology, many researchers use convolutional neural network(CNN) to obtain deep-level features of medical image in order to better classify Alzheimer's disease (AD) and predict clinical scores. The principal component analysis network (PCANet) is a lightweight deep-learning network that mainly uses principal component analysis (PCA) to generate multilevel filter banks for the centralized learning of samples and then performs binarization and generates blockwise histograms to obtain image features. However, the dimensions of the extracted PCANet features reaching tens of thousands or even hundreds of thousands, and the formation of the multilevel filter banks is sample data dependent, reducing the flexibility of PCANet. In order to solve these problems, in this paper, we propose a data-independent network based on the idea of PCANet, called the nonnegative matrix factorization tensor decomposition network (NMF-TDNet). Specifically, we use nonnegative matrix factorization (NMF) instead of PCA to create multilevel filter banks for sample learning, then uses the learning results to build a higher-order tensor and perform tensor decomposition (TD) to achieve data dimensionality reduction, producing the final image features. Finally, our method use these features as the input of the support vector machine (SVM) for AD classification diagnosis and clinical score prediction. The performance of our method is extensively evaluated on the ADNI-1, ADNI-2 and OASIS datasets. The experimental results show that NMF-TDNet can achieve data dimensionality reduction and the NMF-TDNet features as input achieved superior performance than using PCANet features as input.**

*Index Terms*—**Alzheimer's disease (AD), deep learning, PCANet, NMF-TDNet, regression and classification.**

Huan Lao is with the Medical College, Guangxi University, Nanning, Guangxi 530004, P.R. China (e-mail: laohuanlh@qq.com).

Xuejun Zhang is with the Medical College, Guangxi University, Nanning, Guangxi 530004, P.R. China, and also with the School of Computer, Electronics and Information, Guangxi University and the Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning, Guangxi 530004, P.R. China (e-mail: xjzhang@gxu.edu.cn).

Digital Object Identifier 10.1109/JBHI.2021.3113668

## I. INTRODUCTION

ALZHEIMER'S disease (AD) is a neurodegenerative disease commonly observed in elderly adults that has a long incubation period. Over time, the disease gradually worsens, the patient's neurons are gradually destroyed, and his or her memory and cognitive abilities gradually decline, eventually leading to death [1]. There are 50 million people suffering from AD worldwide. With the aging of the global population, the number of AD patients are expected to double by 2050 [2], [3]. Although many drugs are currently available for treating AD, the efficacy of these drugs can only slow its development; none of them can completely treat AD or prevent the progression of the disease [4]. Many studies have shown that in the early stage of the disease, the patient's cognitive impairment will be at a level between the cognitive normal state (CN) and the AD state, which is called the mild cognitive impairment state (MCI). At present, many researchers hope that patients can be identified when they are in the MCI stage and then take effective measures to prevent further progression of the disease [5]. Therefore, the early diagnosis of AD is particularly important, and determining the stage of the disease has become the focus of current research.

In recent years, medical imaging technology has developed rapidly. Its purpose is to provide doctors and researchers with another perspective for disease diagnosis by analyzing medical images, further verifying the correctness of doctors' diagnosis and providing additional data to encourage further research and analysis. Different medical imaging techniques obtain medical images via different modalities, such as magnetic resonance imaging (MRI) [6], [7], single-photon emission computed tomography (SPECT) [8], [9] and positron emission tomography (PET) [10]. These different imaging modalities can help with the noninvasive examination of changes in brain structure and metabolism and the determination of biomarkers of AD. Many studies have established MRI as one of the most standardized and widely used imaging techniques in clinical practice. The most obvious feature of AD pathology is the loss of neurons, followed by brain atrophy, from the characteristic areas of AD (such as the hippocampus and amygdala) to the entire cortical area; these changes can be identified on MRI [11]–[15]. These observable structural changes occur before the first signs of substantial cognitive ability reduction are observed. Therefore, the use of MRI-based computer-aided diagnosis (MRI-CAD) for identifying the AD stage of patients has become the focus of much current research.

The application of deep learning in medical imaging has attracted widespread attention in academia, and the classification of medical images is one of the main research directions of scholars at home and abroad. Therefore, the combination of

deep learning and MRI-CAD has become the focus of much research. Traditional deep learning models include convolutional neural networks (CNNs), deep belief networks (DBNs), stacked autoencoders (SAEs), etc. CNNs can better learn the deep-level features of images, which has become the focus of studies on medical image diagnosis [16]–[18]. Suk *et al.* [19], [20] developed an autoencoder (AE) network-based model for AD diagnosis and used several complex support vector machine kernels for classification. Payan *et al.* [21] used an AE and a 3D-CNN to construct a method that can predict the disease state of AD. The authors of [22] also proposed a 3D-CNN method to achieve AD classification by extracting AD disease biomarkers. Choi and Jin [23] predicted cognitive decline in the process of AD conversion by using an in-depth method based on a CNN. With the continuous development of CNN technology for CAD, image analysis has transitioned from traditional manual analysis to computer-aided analysis. The key factor for the success of a CNN is the ability to automatically discover and learn the abstract representation of data established in multiple stages, where each stage represents an intermediate-level representation developed from the previous stage. Despite the success of CNNs in medical imaging, certain aspects of their feature learning mechanisms and optimal network configurations still pose a great challenge to researchers [24], such as the design of the appropriate network architecture and the choice of the correct configuration and parameters (such as the number of layers, filter size, and pooling function). To overcome the weaknesses of CNNs, Chan *et al.* [25] proposed a lightweight unsupervised deep learning network called the principal component analysis network (PCANet). The network structure of PCANet is very simple. Firstly, use the principal component analysis (PCA) filters to process the input image by layer-by-layer convolution, then performs binarization and calculates block-by-block histograms, and finally obtains a long histogram feature vector as the image features. However, there are two problems with PCANet. The first is that the dimensions of the extracted PCANet features reaching tens of thousands or even hundreds of thousands. With the development of the Big Data era, the practical application of this technique must contend with high storage costs and low calculation efficiency. Second, the formation of multilevel PCA filter banks depends on the sample data, reducing the flexibility of PCANet. In this paper, we propose a network structure combining nonnegative matrix factorization (NMF) and tensor decomposition (TD) based on the network structure of PCANet called the nonnegative matrix factorization tensor decomposition network (NMF-TDNet). Specifically, in this network, we use NMF instead of PCA to construct multilevel filter banks to process the input image by layer-by-layer convolution, and then use convolution results to build a higher-order tensor, finally, TD to achieve data dimensionality reduction to obtain the final image features. NMF-TDNet has the following advantages:

a) NMF is an efficient data dimensionality reduction technique that has better local learning representation than PCA. It reduces high-dimensional random patterns to low-dimensional random patterns while keeping the information unchanged as much as possible. The basis of this simplification is the estimation of the essential structure in the data. Therefore, we use the base matrix obtained from NMF as the multilevel filter banks to convolve the image and obtain its basic features.

b) A tensor is used to store the convolution results, retaining the structural information of the image data.

c) TD is essentially a higher-order generalization of matrix decomposition, decomposing a higher-order tensor into a core tensor and three orthogonal factor matrices. The factor matrices can effectively reflect the topological structure of the original tensor. In this study, we subject the convolution results to Tucker decomposition, which can well perform image dimensionality reduction processing, sparse data filling and implicit relationship mining.

d) The 3D MR images of the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies (OASIS) datasets are used as data sources to verify the experimental performance of our method. The experimental results show that NMF-TDNet can reduce the dimensionality of the data (resulting in features with only a few hundred dimensions, far less than the hundreds of thousands of dimensions obtained from PCANet), which reduces storage costs and time consumption. At the same time, in the two tasks of AD/CN/MCI classification and clinical score prediction, the performance of using NMF-TDNet features as support vector machine (SVM) input is better than PCANet features as input.

The rest of the paper is organized as follows. In Section II, we review the theoretical basis and algorithm of PCANet. In Sections III and IV, we introduce the studied datasets and our proposed NMF-TDNet method, respectively. In Section V, our proposed NMF-TDNet method is evaluated (including disease classification experiments, clinical scoring regression experiments and parameter analyses) and compared with state-of-the-art methods. In Section VI, we discuss the difference between NMF and PCA and analyze the main limitations of the current study. The paper is finally concluded in Section VII.

## II. PRELIMINARY

Chan *et al.* [25] proposed PCANet, which is based on unsupervised learning. It uses the basic PCA filters as convolutional layer filters and then uses binary hash coding and block histogram statistics at the output layer to obtain the final features of the image. PCANet is divided into three steps: first stage, second stage and output layer.

### A. First Stage

Suppose there are $N$ input 3D MR images $I_i(i = 1, 2, \ldots, N)$ with a size of $m \times n \times d$. Each slice $x_{i,j}$ of $I_i$ is divided into small overlapping patches of size $k_1 \times k_2$, and then all patches are mapped into vectors

$$x_{i,j} = [x_{i,j,1}, x_{i,j,2}, \ldots, x_{i,j,p}] \in R^{k_1 k_2 \times p}$$
$$p = 1, 2, \ldots, (m - k_1 + 1)(n - k_2 + 1) \tag{1}$$

In addition, the patch mean is removed from each patch to obtain

$$\overline{x}_{i,j} = [\overline{x}_{i,j,1}, \overline{x}_{i,j,2}, \ldots, \overline{x}_{i,j,p}] \in R^{k_1 k_2 \times p} \tag{2}$$

All slices of $I_i$ are subjected to the above procedure and then concatenated to produce

$$\overline{x}_i = [\overline{x}_{i,1}, \overline{x}_{i,2}, \ldots, \overline{x}_{i,d}] \in R^{dk_1 k_2 \times p} \tag{3}$$

For all input 3D MR images, each image can be used to construct the same matrix, which are then combined into one matrix to obtain

$$X = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N] \in R^{dk_1k_2 \times Np} \quad (4)$$

Then, PCA is used to learn the filters in the first stage. Assuming that the number of filters in this stage is $L_1$, the filters are denoted as $W_l^1$

$$V = XX^T \in R^{dk_1k_2 \times dk_1k_2} \quad (5)$$

$$W_l^1 = mat(q_l(V)) \in R^{dk_1k_2 \times L_1}, l = 1, 2, \ldots, L_1 \quad (6)$$

where $q_l(V)$ denotes the $l$th principal eigenvectors of covariance matrix $V$. After convolution learning with the first stage filters $W_l^1$, the $N$ input images can be expressed as

$$y_i = (W_l^1)^T * \overline{x}_i, i = 1, 2, \ldots, N \quad (7)$$

where * denotes 2D convolution.

### B. Second Stage

This stage repeats the same processes as the first stage with two differences. First, the input images to this stage are the $L_1 \times N$ images from the first stage filter convolution results. Second, the $L_1 \times N$ images are first zero-padded prior to the creation of the overlapping patches to ensure they are the same size as the original 3D MR image slices. Similar to the first stage, after convolution learning with the second stage filters, the $L_1 \times N$ input images can be expressed as

$$O_i^l = (W_l^2)^T * \overline{y}_i, i = 1, 2, \ldots, N \times L_1 \quad (8)$$

### C. Output Layer

Each 3D MR image has $L_2 \times L_1$ convolution results after the first and second stages. First, a Heaviside-like step function $H(.)$ is used to convert the $L_2$ output of the second stage $O_i^l$ back into a single integer in the range $[0, 2^{L_2-1}]$, and each pixel is encoded by the following function:

$$T^l = \sum_{l=1}^{L_2} 2^{l-1} H(O_i^l) \quad (9)$$

The mathematical representation of the function $H(.)$ is

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (10)$$

Then, $T^l$ is partitioned into $p$ nonoverlapping blocks $B$, and the histogram of each block $B$ is computed with $2^{L_2}$ values. Then, the histograms of all blocks $B$ are concatenated into one vector $Bhist(T^l)$. Finally, $L_1$ histograms can be obtained and placed into a vector as an image feature $F$; the formula is shown in (11).

$$F = [Bhist(T^1), Bhist(T^2), \ldots, Bhist(T^l)] \quad (11)$$

According to these procedures, we can see that like most CNN models, when using PCANet to learn the features of the training data, the network parameters must be provided beforehand, such as the number of layers, the block size and the number of filters. However, the number of parameters is much smaller than that required for a CNN model, and the filter learning in PCANet does not involve regularization parameters or require a numerical optimization solver. Therefore, the training complexity is much

lower than that of the CNN model, so PCANet is called a CNN-based simplification deep-learning model. However, the learning of the PCANet filters also has some shortcomings: among them, PCANet relies on training data to learn the multistage filters (which are constructed by calculating the sum of the covariances of all input images). Thus, the formation of the multistage PCA filters is data dependent, reducing the flexibility of PCANet. Detailed experiments demonstrating this henomenon are given in the discussion section.

## III. MATERIALS

Three datasets consisting of 1022 subjects are used in this study, including 1) the ADNI-1 dataset, 2) the ADNI-2 dataset, and 3) the OASIS dataset.

*1) Adni:* The ADNI was launched in 2003 to connect researchers with research data (adni.loni.usc.edu). The ADNI dataset is divided into four stages, ADNI-1, ADNI/GO, ADNI-2 and ADNI-3, and collects a large amount of MRI and PET images, genetic data, blood biochemical indicators and CSF data. The primary goal of the ADNI is to verify and determine the relationship between the collected data, determine the progression of AD, and provide a basis for the early diagnosis and treatment of AD. ADNI's research protocol was approved by the local institutional review board. The study protocols corresponding to the ADNI-1 and ADNI-2 datasets are described as follows. The ADNI-1 stage recruited more than 800 adults, including CN subjects, AD subjects and MCI subjects, to participate in the study, with an age range of 55-90 years (all subjects signed written informed consent). Specifically, 200 CN subjects were observed for 3 years, and 400 MCI subjects and 200 AD subjects were followed up for 3 years and 2 years, respectively. The ADNI-2 stage included 650 newly enrolled subjects. In addition, ADNI-2 followed approximately 700 subjects from ADNI-1 and ADNI/GO for five years.

In our method, the baseline images of 778 subjects from the ADNI-1 and ADNI-2 stages are mainly selected for the experiments. To ensure the independence of samples, subjects that appear in both ADNI-1 and ADNI-2 were removed from ADNI-2. Among them, the subjects selected from the ADNI-1 dataset consist of 95 AD subjects, 171 MCI subjects (including 108 progressive MCI (pMCI) and 63 stable MCI (sMCI) subjects), and 158 CN subjects, and the ADNI-2 dataset consists of 80 AD subjects, 156 MCI subjects (including 59 pMCI and 97 sMCI subjects) and 118 CN subjects. The all subjects list can be found in Table S1 and Table S2 in the Supplementary Materials. The T1-weighted (T1w) MR images of these subjects were acquired using MPRAGE or equivalent protocols of different resolutions with a slice thickness of 1.2 mm and have been subjected to several preprocessing steps by the research groups belonging to the ADNI. In detail, these MR images were preprocessed in steps. First, geometric distortions caused by gradient models were corrected, as were image intensity B1 nonuniformities. Finally, an N3 histogram peak sharpening algorithm was applied to reduce image intensity nonuniformities. Four types of clinical scores are employed for subjects in both ADNI-1 and ADNI-2, including Clinical Dementia Rating Sum of Boxes (CDRSB), classic Alzheimer's Disease Assessment Scale Cognitive subscale (ADAS-Cog) with 11 items (ADAS-11), modified ADAS-Cog with 13 items (ADAS-13), and Mini-Mental State Examination (MMSE) [26]. Table I shows the detailed statistics of all

TABLE I
CHARACTERISTICS OF THE ADNI-1 AND ADNI-2 DATASET SUBJECTS
USED IN THIS STUDY

|  | Diagnosis | Gender(M/F) | Age (Mean ± SD) | MMSE(Mean ± SD) | ADAS-11(Mean ± SD) | ADAS-13(Mean ± SD) |
|---|---|---|---|---|---|---|
| ADNI-1 | AD | 57/38 | 75.91 ± 7.38 | 22.78 ± 2.65 | 19.37 ± 7.13 | 29.87 ± 8.19 |
|  | sMCI | 41/22 | 76.49 ± 7.08 | 27.76 ± 1.86 | 9.228 ± 3.41 | 15.32 ± 5.55 |
|  | pMCI | 72/36 | 75.68 ± 7.54 | 26.44 ± 2.63 | 11.70 ± 3.94 | 19.25 ± 5.57 |
|  | CN | 85/73 | 77.06 ± 4.92 | 29.04 ± 1.27 | 6.236 ± 2.99 | 9.616 ± 4.20 |
| ADNI-2 | AD | 49/31 | 76.44 ± 8.03 | 21.95 ± 3.01 | 20.59 ± 6.72 | 31.03 ± 8.02 |
|  | sMCI | 55/42 | 71.91 ± 7.65 | 27.85 ± 1.92 | 8.639 ± 3.78 | 14.07 ± 5.83 |
|  | pMCI | 30/29 | 73.51 ± 7.20 | 26.14 ± 2.12 | 13.54 ± 4.89 | 21.58 ± 6.74 |
|  | CN | 59/59 | 73.75 ± 6.49 | 28.92 ± 1.14 | 5.864 ± 3.30 | 9.051 ± 4.86 |

Note. Abbreviations: MMSE=Mini-Mental State Examination; SD=Standard Deviation.

TABLE II
CHARACTERISTICS OF THE OASIS DATASET SUBJECTS
USED IN THIS STUDY

| Diagnosis | Gender(M/F) | Age(Mean ± SD) | MMSE(Mean ± SD) |
|---|---|---|---|
| AD | 53/61 | 75.65 ± 8.01 | 24.07 ±4.07 |
| CN | 49/81 | 67.01 ± 8.27 | 29.19 ± 1.06 |

Note. Abbreviations: MMSE=Mini-Mental State Examination; SD=Standard Deviation.

research subjects included in our method, who are divided into three different classes:

CN: The normal control group collected by the ADNI, who did not have a diagnosis of depression, MCI, or other dementia. The MMSE score of these subjects is 24 to 30, with a Clinical Dementia Rating (CDR) score of 0 [27], [28].

MCI: These subjects have no other substantial cognitive impairments and are able to maintain their daily activities. The MMSE score is between 24 and 30, and the CDR is 0.5. MCI is a prodromal stage of AD; MCI that converts to AD is called pMCI, and MCI that remains stable is called sMCI.

AD: The group of ADNI-collected subjects identified as having AD who met the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA criteria) for possible AD [29]. The MMSE score of AD is between 20 and 26, and the CDR is between 0.5 and 1.

*2) Oasis:* The OASIS dataset is a series of MRI datasets, including OASIS-1, OASIS-2 and OASIS-3, which can be used publicly by researchers. OASIS-3 is a dataset used to classify and diagnose CN and AD, collecting longitudinal neuroimaging sequences, clinical indicators, cognitive scores, and biomarkers. The dataset includes 609 CN and 489 AD subjects, with an age ranging from 42 to 95 years. In addition, it collects more than 2000 MR imaging sessions containing multiple sequences, such as T1w, T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), resting-state blood-oxygen-level dependent (BOLD) and diffuse-tensor imaging (DTI), and provides segmentation files generated from MR session images processed by FreeSurfer software. All T1w MR images are available via www.oasis-brains.org. In our method, the T1w MR baseline images of 244 subjects in the OASIS-3 dataset are mainly selected for the experiments, including 114 AD subjects and 130 CN subjects. Note that in this OASIS dataset, only MMSE score is available for all subjects. The detailed statistics of all research subjects are shown in Table II. All subjects list can be found in Table S2 in the Supplementary Materials.

## IV. METHOD

In this paper, 3D MR images are used as the data source, and a method named NMF-TDNet is constructed based on the structural framework of PCANet to extract features to classify the



Fig. 1. Block diagram of our method.

disease type and predict the clinical score of AD. NMF-TDNet mainly includes three steps: the first stage of NMF convolution, the second stage of NMF convolution and the output layer (high-order tensor construction and Tucker decomposition). The algorithm block diagram of this paper is shown in Fig. 1. The detailed steps are explained in the following sections.

### A. MR Image Preprocessing

Many studies have shown that the main morphological and structural abnormalities of AD occur in the gray matter (GM) of the brain [6], [11]–[15]. Therefore, the accuracy of the MRI-CAD system is largely dependent on the segmentation of brain structures or tissues, such as GM or white matter (WM) tissue sections. In this study, all of the original 3D MR images, saved in the NIFTI format, are segmented using the CAT12 (dbm.neuro.uni-jena.de/cat/) toolkit running on MATLAB (mathworks.cn) software. CAT12 is a MATLAB toolkit based on SPM12 (fil.ion.ucl.ac.uk/spm/) that was developed by Ph.D. Christian Gaser and Ph.D. Robert Dahnke of the Department of Psychiatry and Neurology at Jena University Hospital, Germany. The tissue segmentation procedure can be implemented via the "Segment Data" module. This step mainly registers all 3D MR images into the MNI space (MNI152 T1 1.5 mm brain) via Dartel registration to achieve spatial standardization [30], [31]. Finally, the skull tissue of each MR image is removed, and a GM MR image of size $121 \times 145 \times 121$ voxels is obtained.

### B. NMF-TDNet

In this section, we will briefly review NMF and TD and then introduce the architecture of NMF-TDNet in detail.

*1) NMF:* With the advent of the era of Big Data, researchers have begun focusing on transformation methods that can better describe multidimensional data. In general, a good transformation method should have two basic characteristics: (1) it should reveal some of the potential structures of the data, and (2) it should reduce the dimensionality of the data to a certain extent. NMF and PCA are two commonly used transformation methods. Between them, NMF has a better local learning representation ability than PCA. The classic NMF algorithm was published in the well-known international journal Nature in 1999 [32]. NMF has attracted considerable attention from researchers and is widely used in many applications, such as image coding [33], facial recognition [34], image analysis [35], and data clustering [36]. In general, regard a nonnegative matrix $P = (P_{i,j})_{t \times n}$ as consisting of $n$ vectors of size $t \times 1$. NMF can decompose this matrix into two nonnegative matrices, $B = (B_{i,j})_{t \times r}$ and $C = (C_{i,j})_{r \times n}$, where $r$ is called the rank of the NMF, which must satisfy the inequality $r(t + n) < tn$. The nonnegative matrices $B$ and $C$ are called the basis matrix and coefficient matrix, respectively, which can approximately represent the nonnegative matrix $P$ as follows

$$P \approx BC \qquad (12)$$

Researchers have proposed many NMF implementation technologies [37], [38]. In this paper, the basis matrix $B$ and coefficient matrix $C$ of the NMF are calculated by using the multiplication update rule introduced in [37].

$$\begin{cases} B_{i,k} \leftarrow B_{i,k} \frac{\sum_{j=1}^{n} C_{k,j} P_{i,j}/(BC)_{i,j}}{\sum_{j=1}^{n} C_{k,j}} \\ C_{k,j} \leftarrow C_{k,j} \frac{\sum_{i=1}^{t} B_{i,k} P_{i,j}/(BC)_{i,j}}{\sum_{i=1}^{t} B_{i,k}} \\ i = 1, 2, \ldots, t; j = 1, 2, \ldots, n; k = 1, 2, \ldots, r \end{cases} \quad (13)$$

The minimum cost function corresponding to the update rule is called the generalized KL-distance, and its specific calculation formula is as follows

$$F = \sum_{i=1}^{t} \sum_{j=1}^{n} \left[ P_{i,j} \log \frac{P_{i,j}}{(BC)_{i,j}} - P_{i,j} + (BC)_{i,j} \right] \quad (14)$$

It is worth noting that NMF does not allow negative elements in the factor matrices $B$ and $C$. This nonnegative constraint also provides NMF with a reliable local learning representation ability [34]. NMF can be applied to 3D MR image data, reducing a high-dimensional random pattern to a low-dimensional random pattern $C$ while keeping the information as unchanged as possible. The basis of this simplification is the estimation of the essential structure $B$ in the data. Therefore, we use the basis matrix $B$ obtained by NMF as a set of multilevel filter banks to convolve the image and obtain its basic features.

*2) TD:* TD [39], [40] is an effective data analysis technique that plays an important role in data mining, image analysis, signal processing, computer vision and other fields. One commonly used TD technique is Tucker decomposition [41], which has played a large role in facial recognition [42], image quality evaluation [43], noise reduction [44] and data analysis [45].

For a three-order tensor $X \in R^{I \times J \times K}$, Tucker decomposition represents X as a core tensor multiplied by a matrix along each mode. The mathematical formula of Tucker decomposition of X is as follows:

$$X \approx G \times_1 A \times_2 B \times_3 C = [|G; A, B, C|]$$

$$= \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{p,q,r}(a_p \circ b_q \circ c_r) \quad (15)$$

where $A \in R^{I \times P}$, $B \in R^{J \times Q}$ and $C \in R^{K \times R}$ are called factor matrices, which are usually orthogonal and regarded as the principal components of each mode. $a_p$, $b_q$ and $c_r$ are the column vectors of matrices A, B and C, respectively. Each element of the core tensor $G \in R^{P \times Q \times R}$ is represented by $g_{p,q,r}$, the symbol "∘" represents the inner product operation of two vectors, and the symbol "$[--]$" represents a concise representation of Tucker decomposition [39]. The element representation formula of Tucker decomposition is

$$x_{i,j,k} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{p,q,r}(a_{i,p} b_{j,q} c_{k,r})$$

$$(i = 1, 2, \ldots, I; j = 1, 2, \ldots, J; k = 1, 2, \ldots, K) \quad (16)$$

in which $P$, $Q$ and $R$ are the number of components of the corresponding factor matrices A, B and C (such as the number of column vectors). If $P$, $Q$ and $R$ are less than $I$, $J$ and $K$, respectively, the core tensor G can be regarded as a compressed version of X. In some cases, the space required for generating

the compressed version is much smaller than that for the original tensor. In formula (16), $x_{i,j,k}$, $a_{i,p}$, $b_{j,q}$ and $c_{k,r}$ represent the elements of tensor X and orthogonal matrices A, B and C, respectively.

The matrix form of the three-order tensor Trucker decomposition can be expanded as follows:

$$\begin{aligned} X_{(1)} &\approx AG_{(1)}(C \otimes B)^T \\ X_{(2)} &\approx BG_{(2)}(C \otimes A)^T \\ X_{(3)} &\approx CG_{(3)}(B \otimes A)^T \end{aligned} \quad (17)$$

In general, the Tucker decomposition can be solved by well-known methods called higher-order singular value decomposition (HOSVD) and iterative alternate least-squares (ALS) [39], [45], [46]. The factor matrices obtained by Tucker decomposition can effectively reflect the topological structure of the original tensor. In this study, a tensor is used to store the convolution results, retaining the structural information of the image data and well performing dimensionality reduction processing, sparse data filling and implicit relationship mining of the image data.

*3) NMF-TDNet Architecture:* NMF-TDNet uses NMF instead of PCA to construct multilevel filter banks to process the input image by layer-by-layer convolution, then uses the convolution results to build a higher-order tensor, and finally uses TD to reduce the data dimensionality to obtain the final image features, which solves the data dependency and computational efficiency problems of PCANet. NMF-TDNet mainly consists of three steps: the first stage of NMF convolution, the second stage of NMF convolution and the output layer. Each stage of NMF-TDNet is described in detail as follows:

(a) First stage: As in the first stage of PCANet, suppose there are $N$ input 3D MR images $I_i (i = 1, 2, \ldots, N)$ of size $m \times n \times d$. Then, we divide each slice $x_{i,j}$ of $I_i$ into $p$ small overlapping patches of size $k_1 \times k_2$ and map all patches into vectors

$$\begin{aligned} x_{i,j} &= [x_{i,j,1}, x_{i,j,2}, \ldots, x_{i,j,p}] \in R^{k_1 k_2 \times p}, \\ p &= 1, 2, \ldots, (m - k_1 + 1)(n - k_2 + 1) \end{aligned} \quad (18)$$

We subtract the patch mean from each patch and obtain the mean-removed patch vector

$$\overline{x}_{i,j} = [\overline{x}_{i,j,1}, \overline{x}_{i,j,2}, \ldots, \overline{x}_{i,j,p}] \in R^{k_1 k_2 \times p} \quad (19)$$

For all slices of $I_i$, we can perform the same mean patch removal and then concatenate the adjusted slices

$$X_i = [\overline{x}_{i,1}, \overline{x}_{i,2}, \ldots, \overline{x}_{i,d}] \in R^{dk_1 k_2 \times p} \quad (20)$$

Then, we use the NMF implementation techniques of formula (13) to calculate the basis matrix $B$ and the coefficient matrix $C$ of $X_i$ to learn the filters in the first stage. Assuming that the number of filters in this stage is $L_1$, the NMF filters of the first stage are expressed as

$$W(i)_l^1 = mat(B) \in R^{dk_1 k_2 \times L_1}, l = 1, 2, \ldots, L_1 \quad (21)$$

After convolution learning with the first stage filters, the $N$ input images can be expressed as

$$y_i = (W(i)_l^1)^T * X_i, i = 1, 2, \ldots, N \quad (22)$$

(b) Second stage: Similar to the second stage of PCANet, we set the boundary of the input images $y_i$ and perform zero-padding to ensure that the size of $y_i$ is $m \times n$. Then, we can obtain the filters of the second stage by simply repeating the

processes of the first stage. Specifically, we obtain all the overlapping patches of $y_i$

$$y_{i,p} = [y_{i,1}, y_{i,2}, \ldots, y_{i,p}] \in R^{k_1 k_2 \times p}, p = 1, 2, \ldots, mn \quad (23)$$

We subtract the patch mean from each patch and obtain the mean-removed patch vector

$$Y_i = [\overline{y}_{i,1}, \overline{y}_{i,2}, \ldots, \overline{y}_{i,p}] \in R^{k_1 k_2 \times p} \quad (24)$$

Then, we also use the NMF implementation techniques of formula (13) to calculate the basis matrix $B$ and the coefficient matrix $C$ of $Y_i$ to learn the filters in the second stage. Finally, for each input image of the second stage, the NMF filters at this stage are obtained through the basis matrix $B$

$$W(i)_l^2 = mat(B) \in R^{dk_1 k_2 \times L_1}, l = 1, 2, \ldots, L_2 \quad (25)$$

After convolution learning with the second stage filters, the $N \times L_1$ input images can be expressed as

$$O_i^l = (W(i)_l^2)^T * Y_i, i = 1, 2, \ldots, N \times L_1 \quad (26)$$

(c) Output layer: After both the first and second stages are implemented, each input 3D image ultimately yields $L_2 \times L_1$ convolution results. First, the $L_2 \times L_1$ convolution results are stacked together to form a three-order tensor $T_i$ of size $m \times n \times (L_2 L_1)$. The $T_i$ can be expressed as

$$T_i = tensor[O_i, O_{i+1}, O_{i+2}, \ldots, O_{i+L_1 L_2}] \quad (27)$$

Then, the Tucker decomposition (HOSVD method) of $T_i$ is performed

$$[A, B, C] = HOSVD(T_i) \quad (28)$$

Where $A \in R^{m \times 1}, B \in R^{n \times 1}, C \in R^{L_1 L_2 \times 1}$.

After that, the three orthogonal factor matrices obtained by the decomposition are taken as the feature vectors of the image

$$F = [A, B, C] \in R^{(m+n+L_1 L_2) \times 1} \quad (29)$$

## C. Classification Model and Regression Model

In this study, we use LIBSVM to build a support vector classification (SVC) model and a support vector regression (SVR) model for AD classification and clinical MMSE score prediction [47]. LIBSVM is an open source library based on a SVM [48]–[52] and developed by Professor Chih-Jen Lin of Taiwan University. It is mainly used for classification (supporting binary classifications and multiple classifications) and regression. LIBSVM is characterized by its simplicity of operation, ease of use, speed and efficiency, and relatively few adjustments to the parameters involved in SVM.

## V. EXPERIMENTS

### A. Validation Methods and Evaluation Metrics

To obtain a reliable classification and regression evaluation for our method, we perform 10-fold cross-validation 10 times for the classification and regression experiments to obtain the final performance metrics. In the regression experiment, we calculate the correlation coefficient (CC) and root mean square error (RMSE) between the predicted and target clinical scores. Accuracy (ACC), sensitivity (SEN) and specificity (SPE) are used to illustrate the classification performance of the proposed method. In addition, many studies use the receiver operating

TABLE III
COMPARISON OF COMPUTATIONAL TIME COSTS FOR FEATURE EXTRACTION
AND CLASSIFICATION

| Method | Feature | Number of features | Feature Extraction Times(s) | | Classification | Times(s) |
|---|---|---|---|---|---|---|
| | | | ADNI | OASIS | | |
| Volume-based method | MRI | 90 decimals | 10.3 | 9.942 | AD vs. CN | 0.394 |
| | | | | | MCI vs. CN | 0.257 |
| PCANet | MRI | 32,768 decimals | 20.47 | 19.96 | AD vs. CN | 50.53 |
| | | | | | MCI vs. CN | 78.22 |
| NMF-TDNet | MRI | 282 decimals | 12.44 | 13.8 | AD vs. CN | 0.508 |
| | | | | | MCI vs. CN | 0.674 |

Note. The classification time corresponds to the time required to perform 10-fold cross-validation.

characteristic (ROC) curve to evaluate the classification of brain diseases; therefore, in this paper, we also give the area under the ROC curve (AUC) to prove the overall classification performance of the algorithm [53]. In general, a method has a low classification performance if its AUC value is 0.5-0.7, a moderate performance if the value is 0.7-0.9, and a high performance if the value is greater than 0.9.

### B. Experimental Results

*1) Computational Time Costs:* In the era of Big Data, the number of features can be too large, which will result in high storage costs and long-term consumption problems during practical applications. Therefore, in this experiment, we compare the number of features, feature extraction times and classification times between PCANet and our proposed NMF-TDNet. For an additional comparison, we calculate the computational time for the Volume-based method, defined as follows. First, each MRI is used for tissue segmentation to obtain the GM tissue. Then, the AAL template is used to extract 90 ROIs. Finally, the GM tissue volumes of the 90 ROIs are calculated and used as input features. It is worth noting that both PCANet and Volume-based methods also use LIBSVM to build a SVC model and a SVR model for AD classification and clinical score prediction. The computational time costs for feature extraction and classification are summarized in Table III. In the experiments, all compared algorithms are implemented in MATLAB R2014a, running on a laptop with a 2.70 GHz Intel Core i7-7500 U CPU, 8.0 GB RAM, and the Windows 10 (64-bit version) operating system.

As seen from Table III, the Volume-based method has the least number of features, with only 90 decimals in total. The number of features from NMF-TDNet is 282 decimals (the overlapping patch size is $5 \times 5$ and the number of NMF filters in each stage is 4; the parameter selection is discussed in the following section), which is far less than the 32,768 decimals of PCANet (with an overlapping patch size is $5 \times 5$ and 8 PCA filters in each stage). In terms of image feature extraction time, the difference between the Volume-based and NMF-TDNet methods is small, but both are faster than PCANet. Specifically, the feature extraction times for the ADNI and OASIS datasets are respectively 10.3 s and 9.942 s for the Volume-based methods, 12.44 s and 13.8 s for NMF-TD Net, and 20.47 s and 19.96 s for PCANet. However, the construction of the PCA filters in PCANet depends on all input images, while for NMF-TDNet, it does not. Therefore, as the amount of data increases, the time required for PCANet to extract the features will gradually increase. Since the number of features from PCANet reaches tens of thousands of dimensions, its time consumption for the two classification tasks is hundreds of times that of the Volume-based method and NMF-TDNet. Specifically, the classification time of PCANet is 50.53 s for AD vs. CN, which is 128 times (0.394 s) that of the Volume-based method

TABLE IV

THE CLASSIFICATION RESULTS OF THREE METHODS USING 10-FOLD CROSS-VALIDATION IN THE TASKS OF AD VS. CN AND MCI VS. CN. ACC, SEN, SPE AND AUC ARE LISTED AS MEAN ± STANDARD DEVIATION. VALUES IN BRACKETS ARE 95% CONFIDENCE INTERVALS
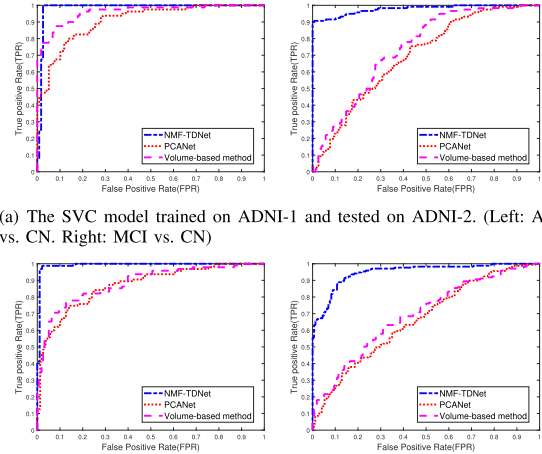
| | Methods | AD vs. CN | | | | MCI vs. CN | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ACC(%) | SEN(%) | SPE(%) | AUC(%) | ACC(%) | SEN(%) | SPE(%) | AUC(%) |
| (a) | Volumes-based | 89.24 ± 0.72 [84.93, 93.55] | 87.75 ± 1.94 [80.59, 94.91] | 90.25 ± 0.72 [84.9, 95.6] | 96.01 ± 0.14 [93.29, 98.73] | 64.05 ± 1.48 [58.37, 69.73] | 76.09 ± 2.42 [69.3, 82.88] | 48.14 ± 4.17 [39.17, 57.11] | 73.45 ± 0.87 [68.22, 78.68] |
| | PCANet | 82.42 ± 0.75 [77.12, 87.72] | 76 ± 2.49 [66.66, 85.34] | 86.78 ± 1.28 [80.68, 92.88] | 90.49 ± 0.45 [90.49, 90.49] | 63.47 ± 1.15 [57.76, 69.18] | 80.51 ± 1.49 [74.13, 86.89] | 40.93 ± 3.53 [32.04, 49.82] | 69.16 ± 1.01 [63.69, 74.63] |
| | NMF-TDNet | **95.61 ± 0.63 [92.76, 98.46]** | **100 ± 0 [100, 100]** | **92.63 ± 1.06 [87.93, 97.33]** | **98.26 ± 0.07 [96.44, 100]** | **89.16 ± 1.22 [85.5, 92.82]** | **100 ± 0 [100, 100]** | **74.83 ± 2.83 [67.04, 82.62]** | **97.81 ± 0.29 [96.08, 99.54]** |
| (b) | Volumes-based | 83.08 ± 0.87 [78.46, 87.7] | 66.32 ± 2.33 [56.83, 75.81] | **93.16 ± 0.65 [89.23, 97.09]** | 88.69 ± 0.25 [84.79, 92.59] | 62.95 ± 1.18 [57.73, 68.17] | 62.75 ± 2.88 [55.59, 69.91] | 63.16 ± 4.6 [55.59, 70.73] | 68.15 ± 1.02 [63.11, 73.19] |
| | PCANet | 80.16 ± 0.61 [75.25, 85.07] | 72.32 ± 1.49 [63.33, 81.31] | 84.87 ± 1.05 [79.29, 90.45] | 87.15 ± 0.17 [83.03, 91.27] | 60.36 ± 1.32 [55.07, 65.65] | 84.15 ± 2.96 [78.7, 89.6] | 34.62 ± 5.25 [27.28, 41.96] | 67.08 ± 0.83 [62, 72.16] |
| | NMF-TDNet | **94.11 ± 1.26 [91.23, 96.99]** | **99.05 ± 0.33 [97.56, 100]** | 85.89 ± 3.55 [78.96, 92.82] | **99.2 ± 0.03 [98.1, 100]** | **85.78 ± 0.85 [82, 89.56]** | 81.99 ± 1.4 [76.25, 87.73] | **89.87 ± 1.16 [85.14, 94.6]** | **94.46 ± 0.32 [91.99, 96.93]** |
| (c) | Volumes-based | 79.92 ± 1 [74.89, 84.95] | 76.58 ± 1.6 [68.82, 84.34] | 82.85 ± 1.15 [76.34, 89.36] | 86.72 ± 0.28 [82.46, 90.98] | - | - | - | - |
| | PCANet | 79.71 ± 0.62 [74.68, 84.74] | 64.74 ± 1.29 [56, 73.48] | **92.85 ± 0.37 [88.38, 97.32]** | 88.76 ± 0.14 [84.8, 92.72] | - | - | - | - |
| | NMF-TDNet | **94.06 ± 0.78 [91.06, 97.06]** | **100 ± 0 [100, 100]** | 88.85 ± 1.46 [83.39, 94.31] | **99.77 ± 0.03 [99.17, 100]** | - | - | - | - |
| (d) | Volumes-based | 79.41 ± 0.68 [74.4, 84.42] | 74.42 ± 2.33 [65.71, 83.13] | 82.41 ± 1.36 [76.38, 88.44] | 87.56 ± 0.4 [83.49, 91.63] | - | - | - | - |
| | PCANet | 77.19 ± 1 [72, 82.38] | 86.21 ± 2.4 [79.35, 93.07] | 71.77 ± 2.28 [64.72, 78.82] | 87.23 ± 0.43 [83.12, 91.34] | - | - | - | - |
| | NMF-TDNet | **94.66 ± 0.94 [91.85, 97.47]** | **98.95 ± 0 [96.9, 101]** | **92.09 ± 1.5 [87.82, 96.36]** | **99.42 ± 0.06 [98.49, 100]** | - | - | - | - |

The Boldface Denotes the Best Performance in Each Metric.

and 99 times (0.508 s) that of NMF-TDNet. In classifying MCI from CN, the classification time of PCANet (78.22 s) is 304 times (0.257 s) that of the Volume-based method and 116 times (0.674 s) that of NMF-TDNet. This indicates that, compared with PCANet, using our proposed NMF-TDNet method to extract image features can effectively improve computational efficiency, reduce storage costs and time consumption.

*2) Classification Results:* In this paper, the results extracted by NMF-TDNet are used as the input features of an SVM classifier for AD, CN and MCI classification. It is worth noting that the performance evaluation concerns the feature extraction method rather than the design of the classifier. To provide a more reliable classification evaluation, the SVC model was trained on one complete dataset, and then tested on the other independent dataset. Specifically, we conducted four group of experiments for verification. (a) The SVC model was trained on ADNI-1 and tested on ADNI-2. (b) The SVC model was trained on ADNI-2 and tested on ADNI-1(reversed the training and testing sets of (a)). (c) SVC model was trained on ADNI-1 and tested on OASIS. (d) The SVC model was trained on OASIS and tested on ADNI-1(reversed the training and testing sets of (c)). The specific classification results of experiments (a)-(d) are shown in Table IV. We report the mean and standard deviation of 10-fold cross-validation results in each metric with the 95% confidence interval in this table.

As seen from Table IV, although the number of NMF-TDNet features (282 decimals) is far less than that PCANet features (32,768 decimals), our method achieves superior classification performance that PCANet for the various classification tasks, and also far higher than the Volume-based method. Specifically, as can be seen from Table IV(a), for the classification of AD and CN, NMF-TDNet achieved an ACC of 95.61%, SEN of 100%, SPE of 92.63% and AUC of 98.26%, which are higher than the classification results of the PCANet (ACC was 82.42%, SEN was 76%, SPE was 86.78% and AUC was 90.49%) and Volume-based method (ACC was 89.24%, SEN was 87.75%, SPE was 90.25% and AUC was 96.01%). For classifying MCI from CN, compared with those of the PCANet and Volume-based method, the ACC, SEN, SPE and AUC of NMF-TDNet improved by 25.69% *vs.* 25.11%, 19.49% *vs.* 23.91%, 33.9% *vs.* 26.69% and 28.65% *vs.* 24.36%, respectively. To study the influence of the training data as well as the generalizability of our proposed method, we also reversed the training and testing sets of (a). Specifically, we first train the SVC model using all the data from the ADNI-2 dataset and then test the model performance with the ADNI-1 dataset. The classification results are summarized in Table IV(b). As shown in Table IV(b), we can observe that our proposed method still outperforms the competing method in this scenario. In addition, by comparing the results achieved by



(a) The SVC model trained on ADNI-1 and tested on ADNI-2. (Left: AD vs. CN. Right: MCI vs. CN)



(b) The SVC model trained on ADNI-2 and tested on ADNI-1. (Left: AD vs. CN. Right: MCI vs. CN)

Fig. 2. Comparison of the ROC curves with three methods for classifying (a) AD vs. CN classification, and (b) MCI vs. CN classification.

NMF-TDNet on Table IV(a) and (b), we can have the following observations. 1) the diagnostic results are comparable (e.g., in AD vs. CN, 95.61% *vs.* 94.11% for ACC, and 98.26% *vs.* 99.2% for AUC. In MCI vs. CN, 89.16% *vs.* 85.78% for ACC, and 97.81% *vs.* 94.46% for AUC). 2) the model trained on ADNI-1 dataset is slightly better, possibly due to the fact that more training subjects are available in ADNI-1 dataset than in ADNI-2 dataset. In conclusion, these experiments suggest that our proposed method has good generalization capacity in sMRI-based AD diagnosis.

In the task of AD vs. CN classification, we also used ADNI-1 and OASIS datasets to do two group of experiments to verify. In the first group of experiments, we first train the SVC model using all the data from the ADNI-1 dataset and then test the model performance with the OASIS dataset. In the second group of experiments, we also reversed the training and testing sets. The specific classification results in Table IV(c) and (d). Similarly, it can be seen from the experimental results that our proposed method has good generalization capacity in sMRI-based AD diagnosis (e.g., NMF-TDNet achieved an ACC of 94.06% *vs.* 94.66%, SEN of 100% *vs.* 98.95%, SPE of 88.85% *vs.* 92.09% and AUC of 99.77% *vs.* 99.42%).

In addition, we also plot the ROC curves of three different methods for classification between AD vs. CN and MCI vs. CN, as shown in Figs. 2 and 3. From the results, it is clear that the NMF-TDNet method outperforms the PCANet and Volume-based methods in all experiments.
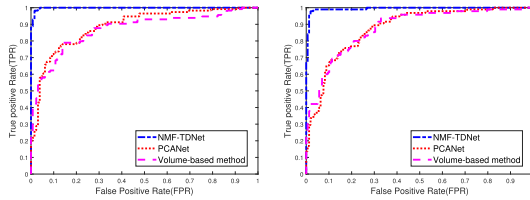
**Fig. 3.** Comparison of the ROC curves with three methods in the tasks of AD vs. CN classification. Left: The SVC model trained on ADNI-1 and tested on OASIS. Right: The SVC model trained on OASIS and tested on ADNI-1.

TABLE VI

CLASSIFICATION RESULTS OF NMF-TDNET METHOD USING LEAVE-ONE-OUT CROSS-VALIDATION IN THE TASKS OF AD VS. CN, MCI VS. CN AND SMCI VS. PMCI

| | Diagnosis | ACC(%) | SEN(%) | SPE(%) |
|---|---|---|---|---|
| (a) | AD vs. CN | 95.95 ± 0.16 [93.20, 98.70] | 100 ± 0 [100, 100] | 93.20 ± 0.28 [88.60, 97.74] |
| | MCI vs. CN | 90.51 ± 0.45 [87.00, 94.20] | 100 ± 0 [100, 100] | 77.97 ± 1.05 [70.43, 85.51] |
| | sMCI vs. pMCI | 79.49 ± 0.58 [73.20, 85.78] | 73.20 ± 0.60 [64.39, 82.01] | 89.83 ± 1.17 [82.42, 97.24] |
| (b) | AD vs. CN | 96.91 ± 1.81 [94.86, 98.96] | 94.34 ± 5.59 [90.23, 98.45] | 98.46 ± 0.54 [96.61, 100] |
| | MCI vs. CN | 86.63 ± 0.41 [82.87, 90.39] | 82.46 ± 1.03 [76.61, 88.31] | 91.14 ± 0.62 [86.67, 95.61] |
| | sMCI vs. pMCI | 79.53 ± 0.99 [73.32, 85.74] | 69.84 ± 0.90 [58.51, 81.17] | 85.19 ± 1.65 [78.06, 92.32] |
| (c) | AD vs. CN | 93.03 ± 0.22 [90.01, 96.05] | 100 ± 0 [100, 100] | 86.92 ± 0.41 [81.42, 92.42] |
| (d) | AD vs. CN | 93.67 ± 0.39 [90.70, 96.64] | 98.94 ± 0 [96.89, 100] | 90.51 ± 0.62 [85.99, 95.03] |

ACC, SEN and SPE Are Listed as Mean ± Standard Deviation. Values in Brackets Are 95% Confidence Intervals.

TABLE V

CLASSIFICATION RESULTS OF NMF-TDNET METHOD USING 10-FOLD CROSS-VALIDATION IN THE TASKS OF SMCI VS. PMCI ON THE ADNI-2 DATASET(ADNI-1 AS THE TRAINING SET) AND THE ADNI-1 DATASET(ADNI-2 AS THE TRAINING SET)

| Dataset | ACC(%) | SEN(%) | SPE(%) | AUC |
|---|---|---|---|---|
| ADNI-2 | 79.17 ± 1.46 [72.8, 85.54] | 73.20 ± 2.06 [64.42, 81.98] | 88.98 ± 1.65 [80.87, 97.09] | 0.899 ± 0.01 [0.852, 0.946] |
| ADNI-1 | 77.6 ± 1.91 [71.4, 83.8] | 69.84 ± 3.17 [58.61, 81.07] | 82.13 ± 2.05 [74.96, 89.3] | 0.828 ± 0.03 [0.773, 0.883] |

ACC, SEN, SPE and AUC Are Listed as Mean ± Standard Deviation. Values in Brackets Are 95% Confidence Intervals.

Studies have found that MRI findings shows a high correlation with the progression from MCI to AD [54]. However, the pathological variations between sMCI and pMCI that can be detected by MRI are subtle. This subtle difference is also accompanied by large intersubject variability and age-related changes, which makes MRI-based predictions of sMCI and pMCI challenging tasks. In this study, we also conducted an experiment to distinguish sMCI from pMCI. Firstly, we still use the data from ADNI-1 as the training set, while the data from ADNI-2 are used as an independent testing set. Thus, we have a total of 171 MCI subjects from ADNI-1 and 156 MCI subjects from ADNI-2. The experimental results are shown in Table V. In most research methods, the classification accuracy in discriminating sMCI and pMCI is in the range of 56-82% [55]–[58]. In the sMCI vs. pMCI classification task, Our method NMF-TDNet achieved ACC of 79.17%, SEN of 73.2%, SPE of 88.89% and AUC of 0.899, indicating its potential capability of diagnosing sMCI vs. pMCI. Similarly, in order to evaluate the generalization ability of our method, we also switched the training and testing sets in the sMCI and pMCI classification tasks, which can be seen that the diagnostic results are comparable (e.g., 79.17% vs. 77.60% for ACC, 73.2% vs. 69.84% for SEN, 88.98% vs. 82.13% for SPE, and 0.899 vs. 0.828 for AUC).

In the above experiments, we perform 10-fold cross-validation 10 times to obtain the final classification results. In order to give more reliable classification evaluation, we apply another verification method (leave-one-out cross-validation) to verify the classification results. The classification results are summarized in Table VI((a) The SVC model was trained on ADNI-1 and tested on the ADNI-2. (b) Reversed the training and testing sets of (a). (c) The SVC model was trained on ADNI-1 and tested on the OASIS. and (d) Reversed the training and testing sets of (c)). Comparing the results of Tables IV, V and VI, it can also be verified that our proposed NMF-TDNet method has good generalizability for AD diagnosis.

*3) Regression Results:* We evaluated the regression performance through the estimation of clinical scores (i.e., ADAS-11, ADAS-13 and MMSE) of the 3D MR images in the three datasets. To evaluate the generalization ability of model, we use subjects from ADNI-1 as the training data, while subjects from ADNI-2 and OASIS as independent testing data. In the first group of experiments, we train a model for three clinical scores (i.e., MMSE, ADAS-11 and ADAS-13) regression on ADNI-1, and test this model on ADNI-2. In the second group of experiments, we train a model for MMSE score regression on ADNI-1, and test it on OASIS. The CC and RMSE values for the first group of experiments are show in Table VII(a). Specifically, in the prediction of MMSE, ADAS-11 and ADAS-13 scores for AD vs. CN, our NMF-TDNet method obtained the RMSEs of 2.854, 6.483 and 8.411, respectively, and the CCs of 0.761, 0.725 and 0.789, respectively. It can be seen from the experimental results, although NMF-TDNet yields far fewer features than PCANet, the prediction results of NMF-TDNet outperforms that of PCANet. For MCI vs. CN, NMF-TDNet and Volume-based method obtained superior CCs (0.343 vs. 0.330(MMSE), 0.413 vs. 0.429(ADAS-11), and 0.439 vs. 0.467(ADAS-13)) and RMSEs (1.943 vs. 2.08(MMSE), 4.416 vs. 4.668(ADAS-11), and 6.769 vs. 7.245(ADAS-13)) than the PCANet. As seen from Table VII(b), in the prediction of the MMSE score for AD vs. CN, NMF-TDNet and PCANet obtained CCs of 0.611 and 0.603, respectively, and RMSEs of 3.239 and 3.268, respectively. Both methods are superior to the Volume-based method (CC and RMSE of 0.527 and 3.399, respectively). Similarly, we also compared the time required to perform clinical scores prediction by all methods. As seen from Table VII, the time costs of PCANet, which are much higher than those of the Volume-based and NMF-TDNet methods. Furthermore, we performed a *t*-test to compare the CCs and RMSEs between our proposed method and the other compared methods (Volume-based and PCANet) separately, as shown in Table VII. We can see that most of the *p*-values for the regression tasks are less than 0.05. These results also validate the efficacy of our proposed method in estimating the clinical scores.

*4) Parameter Analysis:* The important parameters of NMF-TDNet include the overlapping patch size and the number of NMF filters in each stage. Therefore, we conduct experiments in this section to evaluate the impact of these important parameters on the performance of NMF-TDNet for various classification and regression tasks. For the classification of AD and CN, we again first train the model using all the data from the ADNI-1 dataset and then test the performance with the OASIS and ADNI-2 dataset. For MCI vs. CN and sMCI vs. pMCI, the model is trained on the ADNI-1 dataset and tested on the ADNI-2 dataset. Fig. 4(a)–(d) shows the results of the various classification task experiments using different parameters. According to the classification results, when each stage includes 4 filters and the patch size is $5 \times 5$, NMF-TDNet obtains the best classification performance for both classification tasks.

| | Diagnosis | Method | MMSE | | | ADAS-11 | | | ADAS-13 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | RMSE | CC | Times(s) | RMSE | CC | Times(s) | RMSE | CC | Times(s) |
| (a) | AD vs. CN | Volumes-based | 2.885(0.049) | 0.703(0.008)* | 0.072 | **6.142(0.079)*** | 0.715(0.009)* | 0.057 | 8.658(0.147)* | 0.728(0.012)* | 0.049 |
| | | PCANet | 3.275(0.021)* | 0.686(0.007)* | 16.68 | 6.632(0.071)* | 0.72(0.005) | 19.57 | 9.144(0.071)* | 0.753(0.005)* | 21.7 |
| | | NMF-TDNet | **2.854(0.014)** | **0.761(0.002)** | 0.039 | 6.483(0.077) | **0.725(0.009)** | 0.054 | **8.411(0.074)** | **0.789(0.004)** | 0.061 |
| | MCI vs. CN | Volumes-based | 2.08(0.043)* | 0.33(0.012) | 0.222 | 4.668(0.116)* | **0.429(0.017)*** | 0.099 | 7.245(0.194)* | **0.467(0.013)*** | 0.083 |
| | | PCANet | 2.152(0.012)* | 0.029(0.013)* | 24.24 | 4.841(0.020)* | 0.059(0.021)* | 29.66 | 7.435(0.037)* | 0.057(0.035)* | 35.5 |
| | | NMF-TDNet | **1.943(0.010)** | **0.343(0.017)** | 0.1 | **4.416(0.021)** | 0.413(0.013) | 0.126 | **6.769(0.030)** | 0.439(0.015) | 0.127 |
| (b) | AD vs. CN | Volumes-based | 3.999(0.057)* | 0.527(0.013)* | 0.122 | - | - | - | - | - | - |
| | | PCANet | 3.268(0.024)* | 0.603(0.005)* | 30.85 | - | - | - | - | - | - |
| | | NMF-TDNet | **3.239(0.008)** | **0.611(0.004)** | 0.078 | - | - | - | - | - | - |

The Boldface Denotes the Best Performance in Each Metric.



(a) AD vs. CN(Model Trained on ADNI-1 and Tested on OASIS)



(b) AD vs. CN(Model Trained on ADNI-1 and Tested on ADNI-2)



(c) MCI vs. CN(Model Trained on ADNI-1 and Tested on ADNI-2)



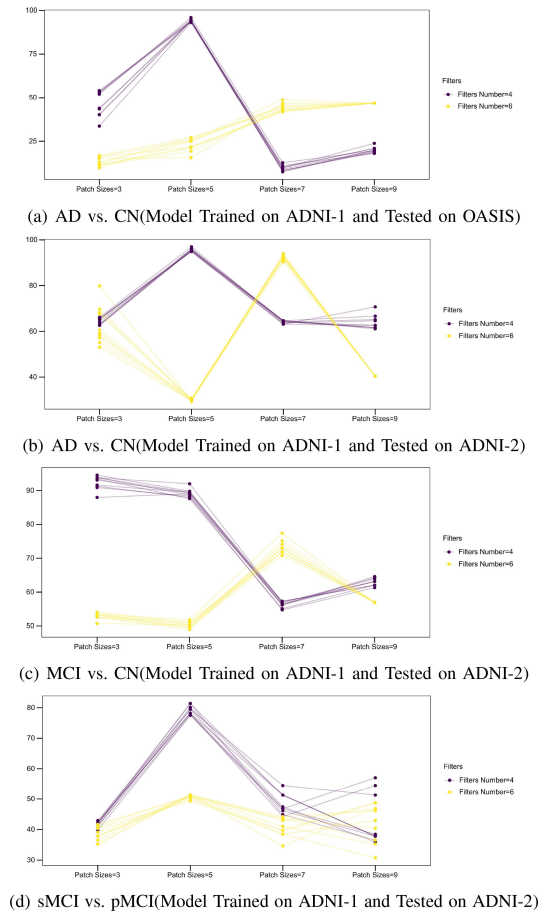(d) sMCI vs. pMCI(Model Trained on ADNI-1 and Tested on ADNI-2)

Fig. 4. Comparison of classification accuracy with different parameters for various classification tasks. The $x$-axis represents the patch sizes, and the $y$-axis represents the classification accuracy.

*5) Comparison With State-of-The-Art Methods:* With the development of machine learning and deep learning, conventional learning-based and deep learning-based methods have been applied to AD diagnosis. Among them, the conventional learning-based method uses image features (such as GM density map, cortical thickness or hippocampal shape measurement) as the input of the classifier, so it is divided into two independent stages, namely the feature extraction stage and the classification stage. Deep learning-based methods is to extract the deep features of the image in a data-driven way for brain disease diagnosis, which

is end-to-end learning. In Table VIII, our method is compared with the results of several methods, including six conventional learning-based and six deep learning-based methods. Like many studies [56], [58], [59], [67], in order to ensure the classification performance of each method, the classification results of each method in the table are the best results obtained through experiments on its original selected dataset(ADNI). That is, these methods were performed with the varying number of subjects, and also the varying partition of training and testing samples, and the definition of pMCI/sMCI may be partially different as well. Although the image data selected by all experimental methods are not exactly the same, the ADNI research group has carried out quality control and preprocessing on the image data. In addition, compared with studies using only partial sMRI data of ADNI-1 [6], [7], [60], [61], [64], [67], our method and the method proposed by Lian *et al.* [56] used more subjects for evaluation (sMRI data from ADNI-1 and ADNI-2), And used a more challenging evaluation protocol (i.e., independent training and testing sets), which should be more challenging but fairer. Therefore, although the results in Table VIII are may not completely comparable, we can roughly comparing our study with these state-of-the-art methods to verify the efficacy of our proposed method.

Firstly, the classification accuracy of our method is compared with the results of conventional learning-based methods that use sMRI data as the research subject and SVM as the classifier to classify AD versus CN, MCI versus CN and sMCI versus pMCI. Specifically, in AD vs. CN, our method achieved a classification accuracy of 95.61%, an improvement of 11.48% (*vs.* Hu *et al.* [7]), 7.12% (*vs.* Khedher *et al.* [62]), 1.24% (*vs.* Liu *et al.* [63]), 1.81% (*vs.* Zhu *et al.* [6]) and 10.81% (*vs.* Zhang *et al.* [64]). Compared with the methods of Peng *et al.* [59], our classification accuracy is slightly lower by 0.49%. The reason possibly due to the fact that the imaging mode (MRI and PET images) and SNP used by this methods at the same time, while our method only has MRI images. For discriminating MCI from CN, we achieved a classification accuracy of 89.16%, an improvement of 8.86% (*vs.* Peng *et al.* [59]), 7.27% (*vs.* Khedher *et al.* [62]), 10.36% (*vs.* Liu *et al.* [63]), 9.46% (*vs.* Zhu *et al.* [6]) and 8.37% (*vs.* Zhang *et al.* [64]). Finally, for classifying sMCI vs pMCI, the accuracy of NMF-TDNet improved by 2.48% (*vs.* Hu *et al.* [7]), 11.34% (*vs.* Liu *et al.* [63]), 8.37% (*vs.* Zhu *et al.* [6]) and 17.17% (*vs.* Zhang *et al.* [64]). In addition, we noticed that Padilla *et al.* also used NMF to extract image features for AD disease diagnosis [8], [9]. Specifically, Padilla *et al.* proposed a novel CAD technology, which is based on functional brain images (SPECT

TABLE VIII
COMPARISON WITH STATE-OF-THE-ART METHODS

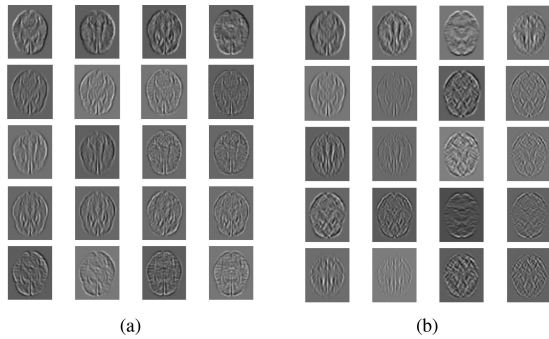| | Algorithm | Modalities (AD/CN/MCI) | Classification results (%) | | |
|---|---|---|---|---|---|
| | | | AD vs. CN | MCI vs. CN | sMCI vs. pMCI |
| Deep Learning-based | Liu et al. [67] | sMRI: 97/119/233 | 88.9 | 76.2 | - |
| | Lian et al. [56] | sMRI:ADNI-1:199/229/393(226sMCI/167pMCI) ADNI-2:159/200/277(239sMCI/38pMCI) | 90 | - | 81 |
| | Khvostikov et al. [60] | sMRI: 48/58/108 | 85.4 | 65.8 | - |
| | Korolev et al. [61] | sMRI: 50/61/120(43pMCI/77sMCI) | 80 | - | 52 |
| | Shi et al. [66] | sMRI+PET: 51/52/99(43pMCI/56sMCI) | 97.13 | 87.24 | 78.88 |
| | Liu et al. [68] | sMRI+PET: 93/100/204 | 93.26 | 74.34 | - |
| Conventional Learning-based | Peng et al. [59] | sMRI+PET+SNP: 49/47/93 | 96.1 | 80.3 | - |
| | Hu et al. [7] | sMRI: 188/228/133(71pMCI/62sMCI) | 84.13 | - | 76.69 |
| | Khedher et al. [62] | sMRI: 188/229/401 | 88.49 | 81.89 | - |
| | Liu et al. [63] | sMRI+PET: 51/52/99(43pMCI/56sMCI) | 94.37 | 78.8 | 67.83 |
| | Zhu et al. [6] | sMRI: 51/52/99(43pMCI/56sMCI) | 93.8 | 79.7 | 70.8 |
| | Zhang et al. [64] | sMRI: 45/50/91(43pMCI/48sMCI) | 84.8 | 73.9 | 62 |
| | Padilla et al. [9] | SPECT: 56/41 | 91.42 | - | - |
| | | PET: 53/52 | 86.59 | - | - |
| | Padilla et al. [8] | SPECT: 56/41 | 94.9 | - | - |
| | Proposed Method | sMRI:ADNI-1:95/158/171(108pMCI/63sMCI) ADNI-2:80/118/156(59pMCI/97sMCI) | 95.61 | 89.16 | 79.17 |



Fig. 5. Convolution results of different filters. (a) Convolution results of the NMF filters for the first stage (rows 1) and second stage (rows 2-5). (b) Convolution results of the PCA filters for the first stage (rows 1) and second stage (rows 2-5).

images and PET images) for early diagnosis of AD. Specifically, firstly, the functional brain image is analyzed by applying the Fisher discriminant ratio (FDR) for feature selection and NMF for feature extraction of relevant components of each subject. Then, the resulting NMF-transformed set of data is classified by SVM classifier. The difference between literature [8] and literature [9] is that the SVM classifier in literature [9] has bounds of confidence for decision. For AD vs. CN classification, the proposed NMF+SVM method achieved ACC of 94.9% (in the literature [8]), 91.42% and 86.59% (in the literature [9]), respectively.

In addition, our method is also compared with deep learning-based methods. From the table we can see that our method can achieve better classification accuracy in distinguishing various classification tasks, whether it is compared with the conventional learning-based method or the deep learning-based method.

## VI. DISCUSSION

To better compare the differences between the PCA filters and the NMF filters, we visualize the convolution results of the first and second stages of PCANet and NMF-TDNet in Fig. 5.

The images show that convolution results of the NMF filters retain more image detail information than those of the PCA filters. The most intuitive manifestation of this difference in the pictures is that the image becomes blurred. Compared with the corresponding panels in Fig. 5(a), the image details of the fourth, sixth, tenth and fourteenth convolution results in Fig. 5(b) are very blurry (from top to bottom and then from left to right).

Based on these observations, we also constructed a method with the same network structure as NMF-TDNet to classify and diagnose AD, referred to as PCA-TDNet. The first and second stages of PCA-TDNet use PCA filters, while the overall processes of the three stages are the same as those of NMF-TDNet. When performing the classification experiments with this PCA-TDNet, the classification model is also trained with ADNI-1 and tested with ADNI-2 and OASIS. The results are shown in Fig. 6, showing that in the classification of AD and CN, the average classification accuracy of PCA-TDNet for all sets of parameters reaches a minimum of 53.28% (57.53%) and a maximum of 59.96% (69.75%), which is far less than the corresponding classification accuracy of NMF-TDNet (ACC of 94.06% or 95.61%). For MCI vs. CN, the average classification accuracy of PCA-TDNet reaches a minimum of 52.63% and a maximum of 56.72%, which is also far less than the corresponding classification accuracy of NMF-TDNet (ACC of 89.16%).

In addition, another difference between the PCA filters and NMF filters is data dependence. Specifically, the PCA filters need to be learned by finding the covariance of all images, while the NMF filters do not need to be learned. For example, when using PCANet for the classification experiments, it is assumed that the PCA filters in the model training stage only use the data learned from the ADNI-1 dataset, and filters of the test stage use the data learned from the OASIS dataset. This reduces the classification accuracy for the AD and CN classification task from 79.71% to 53.28%. We believe that PCANet suffers from an underfitting problem.

Last but not least, we have also studied the brain regions associated with AD. Firstly, we calculate the weight of each NMF-TDNet feature value by ReliefF method [69], and the specific results are shown in Fig. 7. It can be seen from Fig. 7
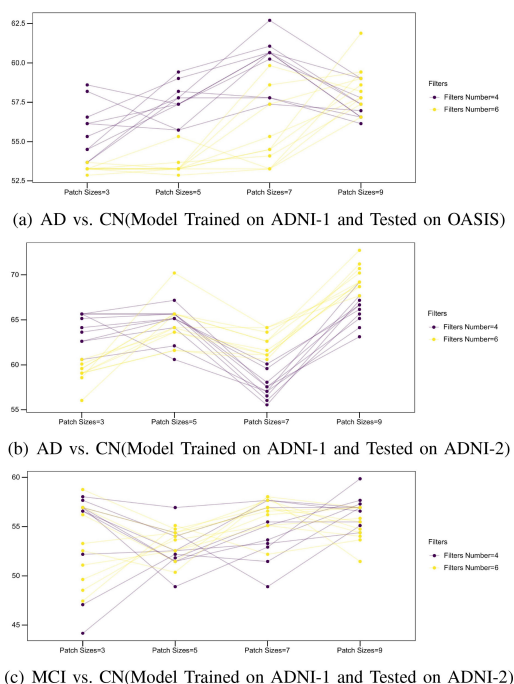
(a) AD vs. CN(Model Trained on ADNI-1 and Tested on OASIS)



(b) AD vs. CN(Model Trained on ADNI-1 and Tested on ADNI-2)



(c) MCI vs. CN(Model Trained on ADNI-1 and Tested on ADNI-2)

Fig. 6. Classification accuracy results of PCA-TDNet. The $x$-axis represents the patch sizes, and the $y$-axis represents the classification accuracy.
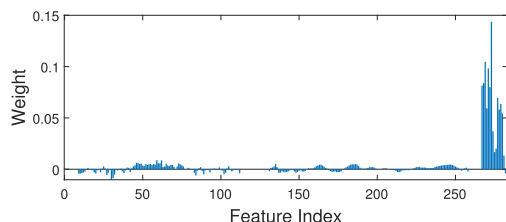


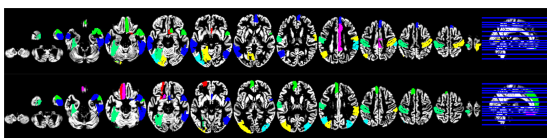Fig. 7. The weight of each NMF-TDNet feature obtained by the ReliefF method.



Fig. 8. The top 15 brain regions associated with AD estimated using the AoIs method in tasks of AD vs. CN classification (top) and MCI vs. CN classification (bottom).

that the $273\text{-}th$ value has the largest weight. Therefore, we can assume that this value contains relevant information about the structure of the AD. Then, we use the regional visualization technology-AoIs developed in the literature [65] to estimate the brain regions where the value is related to AD progression. Fig. 8 shows the top 15 brain regions associated with AD estimated using the AoIs method in the AD vs. CN and MCI vs. CN classification tasks. In Fig. 8, we can see that in the AD vs. CN classification, the relevant brain areas are mainly concentrated in the temporal lobe, frontal lobe and olfactory cortex. For MCI vs. CN, it is also concentrated in the temporal lobe, frontal lobe and olfactory cortex, but it is also more distributed in the caudate nucleus and occipital lobe. These brain areas have also been verified in the literature [63], [65]. At the same time,

we also use this feature to perform regression experiments on the ADAS-11 and ADAS-13 clinical scores. Specifically, when predicting ADAS-11 and ADAS-13 scores for AD vs. CN, this feature obtained RMSEs of 9.669 and 13.858, and CCs of 0.553 and 0.607, respectively. For MCI vs. CN, the CC and RMSE of ADAS-11 are 0.303 and 4.653, respectively, and the CC and RMSE of ADAS-13 are 0.355 and 7.002, respectively.

There are some aspects of our method that we can further improve in future work. The first improvement concerns the unimodality of the input features. When extracting features, we only consider the MR images from the ADNI/OASIS datasets. However, the images collected by the ADNI/OASIS dataset are multimodal, including, for example, PET as well as MRI. Therefore, we can combine images of multiple modalities and extract diverse input features. Second, we can also use this method to classify various forms of dementia, rather than AD alone. Last, we now extract the features of each 3D MR image by gradually processing and then combining all 2D slices. In future work, we can improve our method to more directly process 3D MR images.

## VII. Conclusion

In this work, to overcome the large number of features and the data dependence of the PCA filters that limit PCANet, we propose a method named NMF-TDNet based on the network structure of PCANet. NMF-TDNet uses NMF instead of PCA to construct multilevel filter banks to process the input image by layer-wise convolution, then uses the convolution results to build a higher-order tensor, and then uses TD to reduce the data dimensionality, producing the final image features. Finally, our method use these features as the input of the SVM for AD classification diagnosis and clinical score prediction. On the ADNI-1 and ADNI-2 datasets, we conducted clinical scores (MMSE, ADAS-11 and ADAS-13) prediction and category label discrimination experiments. And on the ADNI-1 and OASIS datasets, the clinical score (MMSE) prediction and category label distinction were performed. The experimental results show that although NMF-TDNet yields far fewer features than PCANet, using NMF-TDNet features as input achieved superior performance than using PCANet features as input.

## Acknowledgment

## References

[1] C. R. Martin, V. R. Preedy, and R. J. Hunter, *Nanomedicine and the Nervous System*, Boca Raton, FL, USA: CRC Press, 2012.

[2] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 14, no. 3, pp. 367–429, 2018.

[3] A. Khan, A. Corbett, and C. Ballard, "Emerging treatments for Alzheimer's disease for non-amyloid and non-tau targets," *Expert Rev. Neurotherapeutics*, vol. 17, pp. 683–695, 2017.

[4] K. G. Yiannopoulou, and S. G. Papageorgiou, "Current and future treatments for Alzheimer's disease," *Therapeutic Adv. Neuro. Disord.*, vol. 6, no. 1, pp. 19–33, 2012.

[5] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain MRI," *Med. Image Anal.*, vol. 18, no. 5, pp. 808–818, 2014.

[6] X. Zhu, H. Suk, D. Shen, "A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis," *NeuroImage*, vol. 100, pp. 91–105, 2014.

[7] K. Hu, Y. Wang, K. Chen, L. Hou, and X. Zhang, "Multi-scale features extraction from baseline structure MRI for MCI patient classification and AD early diagnosis," *Neurocomputing*, vol. 175, pp. 132–145, 2016.

[8] P. Padilla *et al.*, "Analysis of SPECT brain images for the diagnosis of Alzheimer's disease based on NMF for feature extraction," *Neurosci. Lett.*, vol. 479, no. 3, pp. 192–196, 2010.

[9] P. Padilla, M. Lopez, J. M. Gorriz, J. Ramírez, D. Salas-Gonzalez, I. Alvarez, " NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 207–216, Feb. 2012.

[10] A. Besga, M. Termenon, M. Graña, J. Echeveste, J. M. Pérez, and A. Gonzalez-Pinto, "Discovering Alzheimer's disease and bipolar disorder white matter effects building computer aided diagnostic systems on brain diffusion tensor imaging features," *Neurosci. Lett.*, vol. 520, no. 1, pp. 71–76, 2012.

[11] A. T. Du *et al.*, "Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia," *Brain*, vol. 130, pp. 1159–1166, 2007.

[12] V. Singh, H. Chertkow, J. P. Lerch, A. C. Evans, A. E. Dorr, and N. J. Kabani, "Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease," *Brain*, vol. 129, no. 11, pp. 2885–2893, 2006.

[13] B. C. Dickerson *et al.*, "Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area," *Neurobiol. Aging*, vol. 30, no. 3, pp. 432–440, 2009.

[14] C. Hutton, E. De Vita, J. Ashburner, R. Deichmann, and R. Turner, "Voxel-based cortical thickness measurements in MRI," *NeuroImage*, vol. 40, no. 4, pp. 1701–1710, 2008.

[15] Y. Li *et al.*, "Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features," *Neurobiol. Aging*, vol. 33, no. 2, pp. 427–e15, 2012.

[16] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 1097–1105, 2012.

[17] J. Ma, R. P. Sheridan, A. Liaw, G. E. Dahl, and V. Svetnik, "Deep neural nets as a method for quantitative structure-activity relationships," *J. Chem. Inf. Model.*, vol. 55, no. 2, pp. 263–274, 2015.

[18] T. N. Sainath, A. Mohamed, B. Kingsbury, and B. Ramabhadran, "Deep convolutional neural networks for LVCSR," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 8614–8618, 2013.

[19] H.-I. Suk, S.-W. Lee, and D. Shen, "Latent feature representation with stacked auto-encoder for AD/MCI diagnosis," *Brain Struct. Function*, vol. 220, no. 2, pp. 841–859, 2013.

[20] H.-I. Suk and D. Shen, "Deep learning in diagnosis of brain disorders," in *Recent Progress in Brain and Cognitive Engineering, Trends in Augmentation of Human Performance*, S. W. Lee, H. Bülthoff, and K. R. Müller Eds., Dordrecht, The Netherlands: Springer, 2015, vol. 5, pp. 203–213.

[21] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," in *Proc. Int. Conf. Pattern. Recognit. Appl. Methods*, 2015.

[22] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by adaptation of 3D convolutional network," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 126–130.

[23] H. Choi and K. H. Jin, "Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging," *Behav. Brain Res.*, vol. 344, pp. 103–109, 2018.

[24] J. Bruna, and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.

[25] T. H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[26] S. Duchesne, A. Caroli, C. Geroldi, G. B. Frisoni, and D. L. Collins, "Predicting clinical variable from MRI features: Application to MMSE in MCI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2005, pp. 392–399

[27] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state," *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189–198, 1975.

[28] J. C. Morris, "The clinical dementia rating (CDR): Current version and scoring rules," *Neurol.*, vol. 43, no. 11, pp. 2412–2412, 1993.

[29] G. M. McKhann *et al.*, "The diagnosis of dementia due to Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dement.*, vol. 7, no. 3, pp. 263–269, 2011.

[30] G. Grabner, A. L. Janke, M. M. Budge, D. Smith, J. Pruessner, and D. L. Collins, "Symmetric atlasing and model based segmentation: An application to the hippocampus in older adults," *Med. Image Comput. Comput. Assist. Interv.*, pp. 58–66, 2006.

[31] E. OHanlon, F. N. Newell, and K. J. Mitchell, "Combined structural and functional imaging reveals cortical deactivations in grapheme-color synaesthesia," *Front. Psychol.* vol. 4, pp. 755–767, 2013.

[32] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791,1999.

[33] S. Xie, Z. Yang, and Y. Fu, "Nonnegative matrix factorization applied to nonlinear speech and image cryptosystems," *IEEE Trans. Circuits Syst. I Regular Papers*, vol. 55, no. 8, pp. 2356–2367, Sep. 2008.

[34] J. Y. Pan and J. S. Zhang, "Large margin based nonnegative matrix factorization and partial least squares regression for face recognition," *Pattern Recognit. Lett.*, vol. 32, no. 14, pp. 1822–1835, 2011.

[35] R. Sandler and M. Lindenbaum, "Nonnegative matrix factorization with earth mover's distance metric for image analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1590–1602, Aug. 2011.

[36] Y. Chen, L. Wang, and M. Dong, "Non-negative matrix factorization for semisupervised heterogeneous data coclustering," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1459–1474, 2010.

[37] D. D. Lee, and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*," 2000.

[38] I. Kotsia, S. Zafeiriou, and I. Pitas, "A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 588–595, Sep. 2007.

[39] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[40] X. Cao, X. Wei, Y. Han, and D. Lin, "Robust face clustering via tensor decomposition," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2546–2557, Nov. 2015.

[41] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in Measuring Change*, C. W. Harris, Ed. Madison, WI, USA: Univ. Wisconsin Press, 1963, 122–137.

[42] L. Zaoralek, M. Prilepok, and V. Snael, "Recognition of face images with noise based on Tucker decomposition," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2015, pp. 2649–2653.

[43] C. Cheng and H. Wang, "Quality assessment for color images with Tucker decomposition," in *Proc. 19th IEEE Int. Conf. Image Process.*, 2012, pp. 1489–1492.

[44] A. Karami, M. Yazdi, and A. Zolghadre Asli, "Noise reduction of hyperspectral images using Kernel non-negative Tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.

[45] L. Li and D. Boulware, "High-order tensor decomposition for large-scale data analysis," in *Proc. IEEE Int. Congr. Big Data*, 2015, pp. 665–668.

[46] P. M. Kroonenberg and J. de Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms," *Psychometrika*, vol. 45, no. 1, pp. 69–97, 1980.

[47] F. T. Ulaby *et al.*, "Learning with Kernels: Support vector machines, regularization, optimization, and beyond," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 781–781, May 2005.

[48] Y. Fan, D. Shen, and C. Davatzikos, "Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM," in *Proc. Int. Conf. Med. Image Comput.-Assist Intervention*, Berlin, Heidelberg, Germany: Springer, 2005, pp. 1–8.

[49] J. Mourão-Miranda, A. L. W. Bokde, C. Born, H. Hampel, and M. Stetter, "Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data," *NeuroImage*, vol. 28, no. 4, pp. 980–99, 2005.

[50] R. C. Craddock, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "Disease state prediction from resting state functional connectivity," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1619–1628, 2009.

[51] S. Kloppel *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain*, vol. 131, no. 3, pp. 681–689,2008.

[52] C. C. Chang and C. J. Lin, "LIBSVM," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

[53] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.

[54] C. R. Jack *et al.*, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol.*, vol. 9, no. 1, pp. 119–128, 2010.

[55] T. Tong *et al.*, "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 155–165, Jan. 2017.

[56] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 880–893, Apr. 2020.

[57] B. Cheng, M. Liu, D. Zhang, and D. Shen, "Robust multi-label transfer feature learning for early diagnosis of Alzheimer's disease," *Brain Imag. Behav.*, vol. 13, no. 1, pp. 138–153, 2019.

[58] B. Cheng, M. Liu, D. Shen, Z. Li, and D. Zhang, "Multi-domain transfer learning for early diagnosis of Alzheimer's disease," *Neuroinformatics*, vol. 15, no. 2, pp. 115–132, 2016.

[59] J. Peng, X. Zhu, Y. Wang, L. An, and D. Shen, "Structured sparsity regularized multiple kernel learning for Alzheimer's disease diagnosis," *Pattern Recognit.*, vol. 88, pp. 370–382, 2019.

[60] A. Khvostikov, K. Aderghal, J. Benois-Pineau, A. Krylov, and G. Catheline, "3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies," 2018, *arXiv:1801.05968*.

[61] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, 2017, pp. 835–838.

[62] L. Khedher, J. Ramírez, J. M. Górriz, A. Brahim, and F. Segovia, "Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images," *Neurocomputing*, vol. 151, pp. 139–150, 2015.

[63] F. Liu, C. Y. Wee, H. F. Chen, D. G. Shen, and Adni, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's disease and mild cognitive impairment identification," *NeuroImage*, vol. 84, pp. 466–475, 2014.

[64] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[65] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 17–26, Jan. 2020.

[66] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.

[67] M. Liu *et al.*, "A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease," *NeuroImage*, vol. 208, 2019, Art. no. 116459.

[68] M. Liu, D. Cheng, K. Wang, and Y. Wang, "Multi-modality cascaded convolutional neural networks for Alzheimer's disease diagnosis," *Neuroinformatics*, vol. 16, pp. 295–308, 2018.

[69] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn. Mach. Learn.*, Berlin, Heidelberg, Germany: Springer, 1994, pp. 171–182.