Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

3D shearlet-based descriptors combined with deep features for the classification of Alzheimer's disease based on MRI data

Sadiq Alinsaif[a,b,*], Jochen Lang[a], the Alzheimer's Disease Neuroimaging Initiative[1]

[a] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, K1N 6N5, Canada
[b] College of Computer Science and Engineering, University of Hafr Al Batin, Al Jamiah, Hafar Al Batin, 39524, Saudi Arabia

ARTICLE INFO

Keywords:
Feature fusion
Deep learning
MRI Images
3D shearlets
Texture descriptors

ABSTRACT

Alzheimer's disease (AD) is a neurodegenerative disease that afflicts millions of people worldwide. Early detection of AD is critical, as drug trials show a promising advantage to those patients with early diagnoses. In this study, magnetic resonance imaging (MRI) datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and The Open Access Series of Imaging Studies are used. Our method for performing the classification of AD is to combine a set of shearlet-based descriptors with deep features. A major challenge in classifying such MRI datasets is the high dimensionality of feature vectors because of the large number of slices of each MRI sample. Given the volumetric nature of the MRI data, we propose using the 3D shearlet transform (3D-ST), but we obtain the average of all directionalities, which reduces the dimensionality. On the other hand, we propose to leverage the capabilities of convolutional neural networks (CNN) to learn feature maps from stacked MRI slices, which generate a very compact feature vector for each MRI sample. The 3D-ST and CNN feature vectors are combined for the classification of AD. After the concatenation of the feature vectors, they are used to train a classifier. Alternatively, a custom CNN model is utilized, in which the descriptors are further processed end to end to obtain the classification model. Our experimental results show that the fusion of shearlet-based descriptors and deep features improves classification performance, especially on the ADNI dataset.

1. Introduction

An age-related pathological condition that can be captured by magnetic resonance imaging (MRI) is Alzheimer's disease (AD), which afflicts millions of people around the world [1]. A consequence of this disease is the diminished quality of life for people who are afflicted [2]. AD can be confirmed through a postmortem (i.e., examining the brain after death) [3], yet advancements in radiology technologies, e.g, MRI and positron emission tomography, allow the identification of susceptibility to AD. Since MRI is noninvasive and less expensive [4] than other radiology modalities, it is extensively used. Thus, MRI scans have become an appealing approach for the early detection of AD and for observing the progression of the disease [5].

MRI samples in digital form allow for building computer-assisted diagnosis (CAD) systems. Given that an MRI sample includes an abundant amount of information, computational techniques that use machine learning (ML), coupled with image processing, can be exploited to create CAD systems [6]. Such CAD systems can be critical for AD identification and potentially lead to a proper course of treatment. According to the kind of descriptors that are extracted from the MRI samples, the methods for AD classification can be categorized into traditional ML and deep-learning (DL) techniques [7,8]. In traditional ML, robust descriptors have been hand-engineered for the automatic identification of AD [9,10]. On the other hand, a DL model, particularly a convolutional neural network (CNN), can be trained end to end.

A significant challenge is to obtain an early and accurate identification of AD; hence, a robust detection approach can assist medical experts in administering a proper course of treatment to slow the progression of the disease. The prediction of stable mild cognitive impairment (MCI) or MCI progressing to AD is an open issue [11]. Therefore,

[*] Corresponding author. School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, K1N 6N5, Canada.
E-mail addresses: salin025@uottawa.ca (S. Alinsaif), jlang@uottawa.ca (J. Lang).

[1] The 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 both to the design and the implementation of ADNI and provided data, but did not participate in the analysis or the writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.compbiomed.2021.104879
Received 4 April 2021; Received in revised form 15 September 2021; Accepted 15 September 2021
Available online 22 September 2021

the goal of this study is to propose a robust classification model. Our design is based on the combination of handcrafted descriptors in the 3D shearlet domain and deep features for AD classification.

In the following, we provide a literature review of AD classification techniques and outline our contributions in this work.

1.1. Related work

A significant application area of ML in biomedical image classification is MRI data classification [6]; in particular, the early diagnosis of AD from MRI data. In-depth reviews of existing techniques for AD are provided by Tanveer et al. [7] and Ebrahimighahnavieh et al. [8]. According to Ebrahimighahnavieh et al., MRI data for AD classification can be handled in one of the following ways: voxel-based methods, sliced-based methods, region of interest (ROI)-based methods, or patch-based methods. However, we divide our review of such CAD techniques for AD classification into two categories: traditional ML- and DL-based techniques.

1.1.1. Traditional ML-Based techniques

Traditional ML techniques aim at developing a pipeline starting with feature extraction (e.g., shape, color, texture, etc.) and feature selection, and then, effective classification based on these features. Various studies proposed sliced-based techniques from which informative knowledge is extracted. For example, Altaf et al. [12] proposed to combine both texture-based features and clinical data for AD classification. The bag of visual word techniques was utilized with a gray level co-occurrence matrix (GLCM), scale-invariant feature transform, local binary pattern (LBP), and a histogram of gradients. The texture-based features were extracted either from the whole image or segmented regions (i.e., gray matter, white matter, and cerebrospinal fluid). However, when the GLCM was obtained, four texture descriptors were computed for contrast, correlation, homogeneity, and entropy. Only a subset of slices was normally used, but if all slices were used, then the feature vectors were averaged.

Other studies compute features from slices and/or voxels. For example, Nanni et al. [9] proposed the fusion of texture-based and voxel-based features to train a support vector machine (SVM) classifier. Alternatively, Faturrahman et al. [13] proposed using gray matter maps to extract 1) mean and standard deviation of voxel values and 2) segmentation values of the voxel location. Such attributes were used to train a deep belief network.

In addition, Salvatore et al. [10] proposed training an SVM based on feature extraction and reduction of MRI samples using descriptors from principal component analysis (PCA). PCA descriptors were then sorted in a descending order based on their Fisher Discriminant Ratio (FDR) scores.

Other studies proposed to extract information from ROI. For example, Lee et al. [14] proposed using GLCM, but instead of applying this texture-based technique on the spatial domain slices, the authors computed the GLCM from the voxel pairing. Such attributes were computed from different regions of MRI scans, i.e., the hippocampus, precuneus, and posterior cingulate cortex.

Some studies proposed to use spectral methods with the MRI data for multi-scale and different orientation analysis. For instance, Jha et al. [15] proposed using a dual-tree complex wavelet transform (DTCWT) to transform each slice of an MRI. These wavelet coefficients were reduced using PCA. Finally, a feed-forward neural network was used to classify AD from cognitively normal (CN) samples. Similarly, Feng et al. [16] proposed using Daubechies wavelet transformation for each scan of an MRI sample. From the directional sub-bands, energy regions of interest were obtained to compute the energy descriptors, which were then fed to the nearest neighbor classifier for AD classification.

The shearlet transform, which is a truly anisotropic form of a wavelet, has been previously used for texture analysis in medical images. For instance, Zhou et al. [17] used the shearlet transform for the classification of breast tumor ultrasound images. As well, He et al. [18], Dong et al. [19], and Meshkini and Ghassemian [20] used the shearlet transform to construct feature representations for the classification of textured images. More specifically, Acharya et al. [21] proposed using the 2D shearlet transform from which features were extracted, and these features were fed to a nearest neighbor classifier for AD classification.

Spectral transformations have been used to analyze MRI data before, but only using a 2D transformations. Unlike Jha et al. [15], Feng et al. [16], and Acharya et al. [21], we propose to use the 3D shearlet transform (3D-ST) to study the MRI datasets. Similar to Altaf et al. [12], we also propose to utilize textural features. However, unlike Altaf et al., our textural features are computed from the 3D shearlet coefficients.

1.1.2. DL-based techniques

Unlike traditional ML techniques, DL is utilized to handle the complexity of MRI data using 2D and 3D DL models. For example, a sliced-based DL model was proposed by Jain et al. [22]. Jain et al. fine-tuned a pretrained VGG-16, where each MRI slice was considered as an input to the model. Each MRI sample was preprocessed, where a subset of slices was selected based on image entropy. Similarly, Hon and Khan [23] fine-tuned VGG-16 and Inception V4 models while using only a subset of slices that were selected based on image entropy. Each slice of an MRI was considered for AD classification as an independent, labeled example. Yagis et al. [24] also fine-tuned VGG-16 and ResNet-50 while using the same data samples that were used by Hon and Khan. Yagis et al. fine-tuned only the last three layers, and other layers were frozen.

The split of slices is often done randomly, where slices from the same patient can appear in both the training and the testing datasets. Arguably, subject-based splitting, where slices of a patient can only appear in either the training or testing split is more appropriate. Yagis et al. concluded that subject-based splitting, i.e., where slices of an MRI sample can appear either in the testing or training set exclusively, makes it more difficult to obtain high accuracy.

Instead of fine-tuning a DL model, Wang et al. [25] proposed training a shallow network of an 8-layer CNN with data augmentation. Prior to training the model, MRI scans were preprocessed for spatial normalization, smoothing, slice selection, and histogram stretching. Similarly, Abrol et al. [26] proposed training a ResNet model first to predict MCI subjects only. Afterwards, this model was fine-tuned to predict AD vs. CN samples. Lian et al. [27] proposed a patch-based hierarchical convolutional network, where various data augmentation techniques were applied, i.e., random flipping, distorting, and shifting.

Similar to Lian et al., Mendoza et al. [28] proposed a patch-based approach, but utilizing an autoencoder classification model. This autoencoder was trained on a single 2D slice, and three independent views were explored: axial, coronal, and sagittal anatomical planes. The final label of a 2D slice was obtained using a majority voting rule from the ensemble of patch classifications. As well, Pan et al. [29] proposed training a SENet (Squeeze-and-Excitation Network) with data augmentation while utilizing different views, i.e., sagittal, coronal, or transverse MRI slices. As such, a varying number of CNN models were trained on the aforementioned different views. Then, based on the particular binary classification task, the outputs of these various CNNs were ensembled.

Cao et al. [30] proposed first computing the gray matter images from the MRI data. Only a subset from such images was selected based on image entropy. Further, the authors proposed a DL framework that had two pipelines. In the first pipeline, the images were processed in 2D, and the other pipeline processed the images in 3D. Then, the feature maps from both pipelines were fed into a SoftMax layer.

In this study, our proposed pipeline in the context of DL is similar to Jain et al. [22] and Yagis et al. [24]. However, we propose to adjust a DL model to learn deep features from stacked MRI slices instead of using each slice as an independent example. In this manner, we avoid the “information leakage” issue discussed by Yagis et al. Unlike Wang et al. [25], we do not apply any data augmentation while fine-tuning a pretrained CNN model. Similar to Cao et al. [30], we propose to combine different types of features. Our classification model is based on a combination of shearlet-based descriptors with deep features.

1.2. Contributions

Many computational techniques have been developed for the identification of AD patients using traditional and DL approaches. In this paper, we propose to investigate the combination of the 3D-ST and DL for the classification of AD. Unlike earlier studies that have proposed the use of texture-based techniques [12,14], our handcrafted technique is based on the 3D-ST. The 3D-ST considers all slices while transforming the volumetric MRI sample and is expected to better capture volumetric features. GLCM texture descriptors [31,32], LBP [33], local oriented statistics information booster (LOSIB) [34], and segmentation-based fractal texture analysis (SFTA) [35] are utilized to summarize the 3D shearlet coefficients [36]. Such textural features are suffcient to classify the MRI data with an SVM [37] or a decision tree bagger (DTB) [38]. Our starting point in the context of DL is a pretrained CNN that is used for weight initialization. We adopt a 2D stacked MRI slices approach to be able to fine-tune a CNN model that is adjusted to fit the MRI data dimensions. We investigate four fine-tuned DL models as a feature extractor to train a traditional ML model. Our major contribution is a robust classification model that combines both shearlet-based descriptors and deep features. Two strategies are investigated for training a classification model after concatenating the feature vectors. In one strategy, the concatenated descriptors are fed to a traditional ML model; in the other, a custom CNN model is used, which further processes the attributes and then obtains the classification.

The summary of our main contributions are:

- We propose a combined feature representation of shearlet-based descriptors and deep features for the classification of AD.
- To the best of our knowledge, we are the first to apply the 3D shearlet transform for AD classification based on MRI data.
- Our shallow CNN model combines shearlet-based descriptors with deep features and demonstrates competitive classification results on two AD datasets: the Alzheimer's Disease Neuroimaging Initiative (ADNI) and The Open Access Series of Imaging Studies (OASIS).
- We demonstrate the generalization of a classification model to other time-points in the ADNI repository.

2. Proposed system

The proposed system consists of two pipelines. In the first pipeline of our work, an MRI sample is first preprocessed and then transformed using the 3D shearlet, and after that, the coefficients of the shearlet transform are summarized using various texture-based methods. These methods are GLCM, LBP, LOSIB, and SFTA. In this study, two different datasets (ADNI and OASIS) are utilized. Prior to transforming the ADNI dataset, each MRI sample is preprocessed using CAT12 [39]. The following pre-processing techniques are applied: skull-stripping, image normalization, and cropping (as shown in Fig. 2). However, when the OASIS dataset is used, only cropping is applied, since skull removal and intensity correction were applied by the data provider [40].

In the second pipeline, a variable number of CNN models are fine-tuned based on the MRI samples, and then those fine-tuned models are used as feature extractors. Unlike the shearlet-based approach, the MRI volumes are not preprocessed. The CNN model, rather, has full control to learn relevant features. Furthermore, Rajinikanth et al. [41] demonstrated that using an SVM classifier rather than a SoftMax layer can boost the classification performance. Therefore, in our study, instead of using SoftMax, each set of deep features is fed to another classifier, either SVM or DTB.

Finally, we investigate concatenating our proposed descriptors in two different ways: 1) shearlet-based and deep features are concatenated and classified, or 2) shearlet-based descriptors and deep features are combined using a custom CNN model. Each component of our system is described next.

2.1. Shearlet-based descriptors for AD classification

The wavelet transform has shortcomings in dealing with multidimensional data, such as MRI data or videos, but it can work efficiently with approximating a 1D signal. Multidimensional data, in general, are governed by curvilinear singularities, i.e., anisotropic features. Typical wavelet transforms are not capable of effectively localizing or detecting such anisotropic features. Subsequently, the shearlet transform, which is multi-scale and multidirectional, has been proposed to overcome such limitations [36]. Typical wavelets use isotropic scaling, which is not efficient for non-isotropic features. However, one of the aspects of the shearlet transform is its ability to capture anisotropic features. Parabolic scaling is used to control the length and width of the anisotropic element to capture the orientation, but a translation lattice is used to capture the location of discontinuities.

Given that we are dealing with volumetric data in our study, each volume of the MRI sample is transformed using the 3D-ST. The 3D-ST [36,42] is an extension of the 2D shearlet transform. Hence, the 3D shearlet coefficients are also obtained using parabolic scaling, shearing, and translation, forming an affine system as in the 2D domain. Similar to the 2D domain, a "cone-adapted variant" in the 3D domain is used. This variant is called a pyramid-adapted shearlet system. Such systems treat directions in a uniform scheme as in the cone-adapted 2D shearlet transform. These pyramid-adapted discrete shearlet systems can be generated by the aforementioned affine system.

Comparable to the cone-adapted 2D shearlet transform, the 3D-ST partitions the Fourier domain into a rectangular central area and six pyramids. Such partitioning in the frequency domain will generate well-localized directional filter banks that tile the complete frequency space. Therefore, those well-localized waveforms raise the ability of the 3D-ST to handle the complex geometric discontinuities and potentially can lead to generating enhanced descriptors for MRI data classification. For a further description of the shearlet transform and its mathematical representation, readers are referred to Refs. [36,42].

2.1.1. 3D shearlet application to MRI data

ShearLab [36] is used in our application of 3D shearlets. An MRI sample of size $h \times w \times c$ is decomposed with S scales and K orientations per scale. The dimension of the three-dimensional filters is $h \times w \times z$, which is used for convolution with the input MRI sample. This setting generates a structure of size $h \times w \times c \times z$, where h, w, c, and z represent the height, width, number of slices, and number of filters, respectively.

Having the shearlet magnitude coefficients of each slice with various orientations, z, the average presence of these coefficients across this channel is obtained; hence, the cth slice is represented by only one averaged shearlet sub-band (i.e., $h \times w \times c \times 1$). Such an aggressive reduction of shearlet coefficients is necessary; otherwise, a very large number of descriptors per slice is produced while computing textural features from every directional sub-band. Concatenating all descriptors of all slices to form the feature vector of an MRI sample will lead to a very large number of descriptors. With a very large number of descriptors and a low number of training samples, the classification performance will suffer from overfitting; hence, it will lead to low classification results on the testing samples.

2.1.2. Texture-based descriptors in the 3D shearlet domain

In previous work [43,44], we developed texture features in the 2D shearlet domain. These texture features were based on GLCM textural features [32], LBP [33], LOSIB [34], or SFTA [35]. Each one of the four texture-based techniques is also applied in this study, but independently. Given that an MRI slice is represented by only one averaged shearlet sub-band, a set of descriptors, g, for this cth slice is computed as one vector, $G_c = (g^1, g^2, \ldots, g^j)$, where j represents the number of descriptors. The final feature vector, E, for an MRI sample can be denoted as $E = (G_1, G_2, \ldots, G_q)$, where q represents the number of slices of an MRI sample.

The texture-based methods are applied on the averaged shearlet sub-band that has dimension of $h \times w$, which allows the applicability of such methods for the 3D shearlet coefficients. Next, a brief description of each texture method is provided.

2.1.2.1. Gray-level Co-occurrence matrix (GLCM). The concept of GLCM is owed to Haralick et al. [32]. Other studies [31,45] proposed further statistics that can be extracted from the GLCM. Denote a and b to be two specific coefficients that are separated by a distance, D, then an image content can be summarized with a relative frequencies matrix, F_{ab}. To compute the GLCM, the orientation or the distance to obtain the relative frequencies matrix can be adjusted. Hence, the GLCM contains relative frequencies of neighboring pixels for quantized orientation and distance. Thus, the value of D can be adjusted to reflect either local or global relative frequencies of coefficients.

In this study, for a given averaged shearlet sub-band, the orientation is changed to $= (0^\circ, 45^\circ, 90^\circ, \text{and } 135^\circ)$ for a given D. Thus, four GLCMs are obtained, but we calculate the mean of the four GLCMs from which rotation invariant statistics [46] are computed. Next, from this averaged GLCM, the following 20 textural features are obtained: contrast, correlation, energy, autocorrelation, cluster prominence, cluster shade, dissimilarity, entropy, homogeneity, maximum probability, sum of squares, variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation, inverse difference normalized, and inverse difference moment normalized.

2.1.2.2. Local binary pattern (LBP). The textural features that are computed from the LBP [33] method are rotationally invariant of the local occurrence of an image.

In our study, LBP textural features are computed from each averaged shearlet sub-band that represents an MRI slice. For a given shearlet (SH) coefficient at the position (u, v), the LBP of the SH coefficient is calculated as follows:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(SH_p - SH_l)2^p, s(x) = \begin{cases} 1, & x \geqslant 0. \\ 0, & x < 0. \end{cases} \quad (1)$$

SH_l and SH_p denote the central magnitude coefficients, and P denotes the neighboring magnitude coefficients in the circular neighborhood.

2.1.2.3. Local oriented statistic information booster (LOSIB). The first step of the LOSIB [34] method is to obtain the absolute difference d_p. The absolute difference is computed for every central shearlet coefficient, l, with P surrounding coefficients, in the following manner: $d_p(u_l, v_l) = |SH_l - SH_p|$ where $p \in 0, 1, \ldots, (P-1)$.

Next, the average of the differences across the same orientation is calculated, in the following manner:

$$\mu_p = \frac{\sum_{x_l=1}^{M} \sum_{y_l=1}^{N} d_p(x_l, y_l)}{M \cdot N} \quad (2)$$

N and M denote the height and width of an averaged shearlet sub-band, respectively.

2.1.2.4. Segmentation-based fractal texture analysis (SFTA). The SFTA [35] technique decomposes an averaged shearlet sub-band into binary images using a two-threshold binary decompositions (TTBD). As such, the following features are calculated from the binary images: the dimension of the fractal boundaries, the average gray level, and the count of pixels belonging to the region.

2.2. Fine-tuning pretrained CNN models for AD classification

In this study, we examine pretrained models on nonmedical images, i.e., on ImageNet [47]. Four pretrained CNN architectures are fine-tuned for AD diagnosis, as described in this section. Then these pretrained CNN models are leveraged for their capability as feature extractors.

2.2.1. Fine-tuning a pretrained model

Sharma and Mehra [48] demonstrated that fine-tuning a pretrained model coupled with logistic regression as a classifier in the task of breast cancer histology image classification led to good classification performance. In this study, we propose to use four pretrained CNN models in MATLAB® 2020b that can handle the complexity of MRI datasets: SqueezeNet-v1.1 [49], MobileNet-v2 [50], Xception [51], and Inception-v3 [52]. Sharma and Mehra [48] fine-tuned a pretrained model using histology images that have the same number of channels as the input layer of the pretrained model. In our case, the MRI sample consists of a different channel number as the slices are stacked in depth. Consequently, each adopted CNN model needs to be adjusted in the following manner:

- The input layer needs to be changed to fit the dimensions of an MRI sample, i.e., $h \times w \times c$, where the number of channels of an MRI sample is $c > 3$. In the case of ADNI and OASIS datasets, an MRI sample has a dimension of $150 \times 150 \times 90$ and $150 \times 150 \times 32$, respectively. Since the input layer is pretrained on RGB images, the convolution layer which follows the input layer has a pretrained weights matrix with a channel size $= 3$. Therefore, this convolution layer must be replaced and trained from scratch while keeping the same number of filters as the pretrained model. As a result, a new weight tensor with an input depth that fits the channel size of the new input layer (90 or 32, respectively) is created.
- The specialized layers prior to the classification layer, typically convolution and pooling layers or dense layers, are replaced and trained from scratch using new filter sizes to fit the size of the preceding feature maps. Finally, given a binary problem, our classification layer, which is also trained from scratch, must have an output size of two.

The remaining pretrained weights of a CNN model are used as weight initializers.

2.2.2. CNN as feature extractor

A pretrained deep model can be used to extract a feature vector per MRI slice, and then concatenate all feature vectors of all slices to form a single feature vector representing an MRI sample. Such a process would generate a very large number of descriptors per MRI sample. However, necessary changes are incorporated to a pretrained DL model for MRI classification, as described in the previous section. We make use of state-of-the-art deep networks to learn the deep features of a stacked MRI sample simultaneously.

However, the selected feature maps in SqueezeNet-v1.1 is module “Fire9.” The mean of each feature map is computed, generating a feature

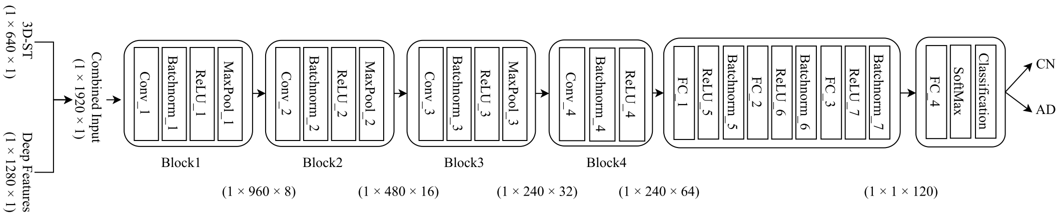


Fig. 1. Custom 1D CNN model which is used to transform the combined shearlet-based descriptors with deep features. Dimensions in the figure correspond to Fusion #2 (See Section 4.3) descriptors computed from the OASIS dataset. Details about the activations and parameters are provided in Appendix B.

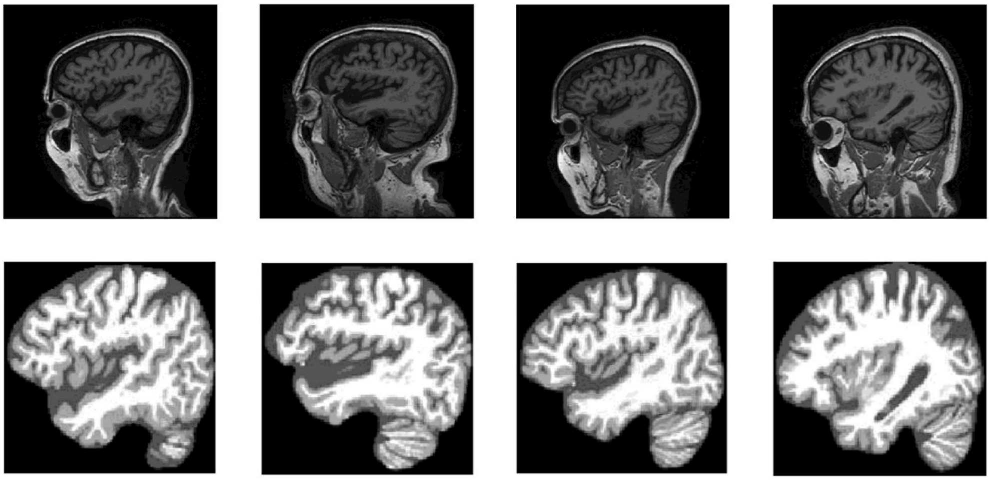


Fig. 2. Examples of MRI slices before (top) and after (bottom) processing. Left to Right: CN, sMCI, AD, and pMCI, respectively.

vector of size 1 × 1 × C, which is finally flattened to form the feature vector of an MRI sample.

In MobileNet-v2, Xception, and Inception-v3, the output of the average pooling (Avg Pool) layer is utilized, which is the layer just prior to the classification layer.

Thereafter, a standard classifier is trained in the context of nested cross-validation (see Section 3.3). We experiment with SVM and DTB as classifiers.

2.3. Custom CNN

Although the proposed 3D-ST approach does not make full use of the multiple scales and directionalities, it still uses averaged shearlets as a reduced form of the coefficients. This reduced averaged shearlet subband is used to obtain low-level image structures, e.g., utilizing SFTA, from which local and global statistics are computed. Alternatively, one can obtain high-level semantic deep features that are computed from different CNN models. We then investigate whether combining shearlet-based descriptors and deep features enhances the classification model, and if fusing such attributes provides a more discriminative and effective classifier. A typical and simple approach is to feed the new combined feature vector to a classifier. We will compare our method with this simple approach.

Alternatively, our main idea is to feed and transform the combined feature vectors using a customized CNN model, as shown in Fig. 1. The building block of our custom CNN model consists of the following operations: convolution (Conv), batch normalization, rectified linear unit (ReLU), and, finally, max pooling. For the convolution layers, a filter size of 7 is used for each convolution layer, but the number of filters is 8, 16, 32, and 64, respectively. The convolution layer is similar to a feature extraction step in the traditional pipeline, but the parameters of this layer can be learned during back propagation. The filter size for the max-pooling layers is 1 with a stride of 2. A pooling layer is an important element of our CNN model for feature reduction and speeding up the training process. Furthermore, to avoid overfitting to a dataset while training our model, batch normalization is utilized. This building block is repeated four times and then followed by three fully connected layers that consist of 120 neurons and then one fully connected layer of 2 neurons corresponding to a binary classification problem.

Our custom CNN model is derived empirically, such that it can generalize on the validation data for the ADNI and OASIS datasets.

3. Experiments

3.1. Datasets

3.1.1. Alzheimer's Disease Neuroimaging Initiative (ADNI) data

We use the same data samples as Salvatore et al. [53]. The data are archived and can be accessed from the ADNI data repository[2]. The dataset consists of 200 patients. The distribution of samples in this dataset is as follows: 50 patients with a stable diagnosis of CN during the 24-month follow-up, 50 patients with a stable diagnosis of mild cognitive impairment (sMCI), 50 patients with a stable diagnosis of AD, and 50 patients with an MCI diagnosis that may progress to AD (pMCI). All patients had three serial MRI examinations at multiple points after the baseline: after 6, 12, and 24 months. However, since our target is to examine the capability of our techniques to distinguish Alzheimer's

[2] The ADNI data archive can be accessed via adni.loni.usc.edu.

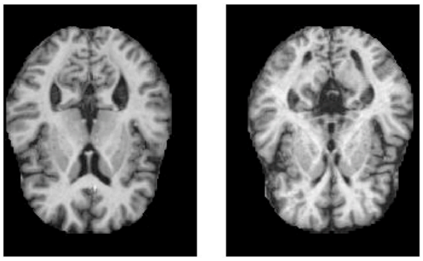


Fig. 3. Examples of the OASIS MRI slices. CN on the left and AD on the right.

Table 1
ADNI classification performance using shearlet-based descriptors for (CN + sMCI) vs. (pMCI + AD).

Method		ACC	Sen	SP	GM
GLCM	SVM	0.6800 ± 0.0542	0.6700 ± 0.1037	0.6900 ± 0.0418	0.6778 ± 0.0570
	DTB	0.6850 ± 0.0335	0.7100 ± 0.0742	0.6600 ± 0.0418	0.6831 ± 0.0332
LBP	SVM	0.6700 ± 0.0570	0.7000 ± 0.0791	0.6400 ± 0.0822	0.6674 ± 0.0571
	DTB	0.6300 ± 0.0671	0.6300 ± 0.0758	0.6300 ± 0.0908	0.6284 ± 0.0678
LOSIB	SVM	0.6450 ± 0.0326	0.6700 ± 0.0447	0.6200 ± 0.0570	0.6436 ± 0.0324
	DTB	0.6400 ± 0.0379	0.6400 ± 0.0652	0.6400 ± 0.0652	0.6382 ± 0.0389
SFTA	SVM	**0.7050 ± 0.0447**	**0.7300 ± 0.1151**	0.6800 ± 0.0570	**0.7009 ± 0.0458**
	DTB	0.6750 ± 0.0433	0.6300 ± 0.0837	**0.7200 ± 0.0447**	0.6718 ± 0.0460

Best performance is highlighted.

patients from non-Alzheimer's patients using MRI data, our focus in this study is on the baseline MRI scans of the patients. Examples of this dataset are shown in Fig. 2.

We carry out the following classification task while using the ADNI dataset as did Salvatore et al.: (CN + sMCI) vs. (pMCI + AD)[3].

To evaluate the generalization of our proposed method, we retrieve a separate set of MRI data from the ADNI data repository. As such, 10 MRI volumes per category are retrieved (i.e., CN, sMCI, AD, and pMCI), but these MRI volumes are of different time-points (i.e., not necessarily of baseline scans). Such a dataset of new subjects is unseen during the training process from start to finish, and it will be used to test our proposed method at the very end.

3.1.2. *The Open Access Series of Imaging Studies (OASIS) data*

To further examine our proposed techniques, we retrieve another set of cross-sectional MRI data from the OASIS data repository[4] for AD classification [40]. Examples of AD and CN slices are shown in Fig. 3. In this study, we retrieve 100 samples each of CN and AD patients. We retrieve the same 200 subjects[5] that were used by Hon and Khan [23]. However, 10 examples out of each of the 100 samples from each category (i.e., CN and AD) are kept hidden from beginning to end. These hidden examples are then used at the very end to examine the generalization of our proposed method. Hence, only 90 examples of each category are used to train and validate our proposed techniques.

The OASIS data repository provides MRI scans as individual slices for a patient. However, we deal with the data as a volume for each patient's slices by converting them into the Neuroimaging Informatics Technology Initiative (NIfTI) format. Again, a binary classification CN vs. AD is applied to this data.

3.2. *Classification algorithms*

Different classifiers were compared by Thanh and Kappas [54], who demonstrated that the highest classification results were obtained by an SVM model, which were then followed by decision trees. In this study, SVM [37], with radial basis function as a kernel, and DTB [55] are used as implemented in MATLAB® 2020b. In the case of SVM, in an attempt to choose the best value of the regularization parameter (C), trials of values between [1,5] with an increment of 1 are conducted. In the case of DTB, the number of grown trees in an ensemble are between [50, 750] with an increment of 50 trees. These hyperparameters are determined in the context of nested cross-validation.

3.3. *Nested cross-validation (CV)*

We utilize five-fold nested CV for hyperparameter optimization. While using a five-fold nested CV, the data is split into five subsets of equal sizes. Therefore, four out of five subsets are used for training and validation for hyperparameter optimization in an inner loop. The last subset is then used for testing the performance of a classifier in an outer loop.

For each optimization round, the set of hyperparameters is tuned to maximize the geometric mean of the classifier in the inner loop. For each optimization round, the classification performance is estimated in the outer loop in terms of accuracy, sensitivity, specificity, and geometric mean.

The number of samples that are used in our nested CV is as follows:

- Given that the number of MRI volumes from the ADNI data is 200 (i.e., 100 CN + sMCI, and 100 pMCI + AD) for each round of the nested CV, the number of MRI samples that are used to train a classifier is 128; the number of samples that are used in the inner loop for optimization is 32; the number of MRI samples that are used to test the performance in the outer loop is 40.
- Given that the number of MRI volumes from the OASIS data is 180 (i.e., 90 CN and 90 AD) for each round of the nested CV, the number of MRI samples that are used to train a classifier is 115; the number of samples that are used in the inner loop for optimization is 29, except for one round, where a classifier is trained with 116 samples and validated with 28 samples. However, the number of samples that are used to test the performance in the outer loop is 36.

We use five-fold nested CV, i.e., the above procedure is repeated five times, and then, across the five rounds, the average of common classification performance in the outer loop is reported. We use the following metrics:

- Accuracy (*ACC*) is the ratio of correct classifications over total number of samples, or $ACC = (TP + TN)/(TP + TN + FP + FN)$, where *TP*, *TN*, *FP*, *and FN* denote the number of true positives, true negatives, false positives, and false negatives, respectively.
- Sensitivity (Sen) is the ratio of true positives over actual positives, or $Sen = TP/(TP + FN)$.
- Specificity (SP) is the ratio of true negatives over actual negatives, or $SP = TN/(TN + FP)$.
- Geometric mean (GM) is the square root of the product of sensitivity and specificity, or $\text{GM} = \sqrt{Sen \times SP}$.

[3] More information about the dataset and patients IDs can be accessed at https://github.com/christiansalvatore/Salvatore-200Longitudinal.

[4] The OASIS data repository can be accessed via http://www.oasisbrains.org.

[5] More information about the patients can be found at: https://github.com/marciahon29/Ryerson_MRP.

Table 2
OASIS classification performance using shearlet-based descriptors for CN vs. AD

Method		**ACC**	**Sen**	**SP**	**GM**
GLCM	SVM	**0.6889 ± 0.1028**	0.6444 ± 0.1394	0.7333 ± 0.0724	**0.6860 ± 0.1066**
	DTB	**0.6889 ± 0.0602**	0.6333 ± 0.1152	**0.7444 ± 0.0843**	0.6827 ± 0.0613
LBP	SVM	**0.6889 ± 0.0819**	0.6556 ± 0.1542	0.7222 ± 0.0786	0.6823 ± 0.0917
	DTB	0.6667 ± 0.0810	0.6333 ± 0.1152	0.7000 ± 0.0930	0.6628 ± 0.0847
LOSIB	SVM	0.6556 ± 0.0421	0.5889 ± 0.0843	0.7222 ± 0.1039	0.6479 ± 0.0380
	DTB	0.6278 ± 0.0824	0.6444 ± 0.1152	0.6111 ± 0.1179	0.6229 ± 0.0850
SFTA	SVM	0.6222 ± 0.0505	0.6444 ± 0.1009	0.6000 ± 0.0724	0.6185 ± 0.0506
	DTB	0.6444 ± 0.0819	**0.6667 ± 0.1303**	0.6222 ± 0.0609	0.6418 ± 0.0815

Best performance is highlighted.

Table 3
ADNI classification performance using deep features for (CN + sMCI) vs. (pMCI + AD).

Method		**ACC**	**Sen**	**SP**	**GM**
SqueezeNet	SVM	0.8400 ± 0.0859	0.8569 ± 0.0878	0.8218 ± 0.1102	0.8378 ± 0.0831
	DTB	0.8950 ± 0.0570	0.8774 ± 0.0602	0.9132 ± 0.0909	0.8939 ± 0.0543
MobileNet	SVM	0.9050 ± 0.0411	0.8948 ± 0.0898	0.9063 ± 0.0628	0.8987 ± 0.0443
	DTB	**0.9100 ± 0.0379**	**0.9161 ± 0.0649**	0.8964 ± 0.0563	**0.9052 ± 0.0381**
Xception	SVM	0.9000 ± 0.0586	0.8744 ± 0.1016	0.9217 ± 0.0645	0.8963 ± 0.0659
	DTB	0.9050 ± 0.0737	0.8839 ± 0.1124	**0.9241 ± 0.0846**	0.9021 ± 0.0789
Inception	SVM	0.8650 ± 0.0652	0.8861 ± 0.0779	00.8410 ± 0.0572	0.8631 ± 0.0656
	DTB	0.8800 ± 0.0818	0.8839 ± 0.1018	0.8727 ± 0.0822	0.8776 ± 0.0848

Best performance is highlighted.

Table 4
OASIS classification performance using deep features for CN vs. AD.

Method		**ACC**	**Sen**	**SP**	**GM**
SqueezeNet	SVM	0.8944 ± 0.0497	0.8603 ± 0.0492	0.9313 ± 0.0714	0.8942 ± 0.0422
	DTB	0.8833 ± 0.0534	0.8637 ± 0.0899	0.9146 ± 0.0885	0.8861 ± 0.0480
MobileNet	SVM	0.9167 ± 0.0196	0.9034 ± 0.0688	0.9367 ± 0.0579	0.9184 ± 0.0225
	DTB	**0.9278 ± 0.0248**	**0.9146 ± 0.0703**	**0.9467 ± 0.0542**	**0.9291 ± 0.0266**
Xception	SVM	0.9000 ± 0.0541	0.9034 ± 0.0688	0.9035 ± 0.1113	0.9010 ± 0.0562
	DTB	0.8889 ± 0.0481	0.8909 ± 0.0500	0.8929 ± 0.1169	0.8894 ± 0.0513
Inception	SVM	0.8889 ± 0.0589	0.9091 ± 0.0539	0.8723 ± 0.1071	0.8889 ± 0.0622
	DTB	0.9056 ± 0.0639	0.8966 ± 0.0529	0.9129 ± 0.1266	0.9023 ± 0.0653

Best performance is highlighted.

Table 5
The configurations of the texture-based methods.

Dataset	GLCM (D)	LBP (R, P)	LOSIB (R, P)	SFTA (n_t)
ADNI	$D = 1$	(2, 16)	(5, 16)	$n_t = 1$
OASIS	$D = 2$	(3, 16)	(3, 16)	$n_t = 4$

Distance (D), radius (R), neighborhood (P), number of thresholds (n_t).

Table 6
ADNI classification performance while combining shearlet-based descriptors with deep features for (CN + sMCI) vs. (pMCI + AD).

Method		**ACC**	**Sen**	**SP**	**GM**
(a): Combining shearlet-based descriptors with deep features to train/validate a classifier					
Fusion #1	SVM	0.9050 ± 0.0274	0.9100 ± 0.0418	0.9000 ± 0.0500	0.9044 ± 0.0276
	DTB	0.8750 ± 0.0306	0.8900 ± 0.0418	0.8600 ± 0.0548	0.8742 ± 0.0311
Fusion #2	SVM	**0.9350 ± 0.0285**	**0.9400 ± 0.0418**	0.9300 ± 0.0447	**0.9345 ± 0.0283**
	DTB	0.8950 ± 0.0597	0.9100 ± 0.0224	0.8800 ± 0.1151	0.8933 ± 0.0618
(b): Combining shearlet-based descriptors with deep features to train/validate the custom CNN					
Fusion #1	CNN	0.9000	0.8500	**0.9500**	0.8986
Fusion #2	CNN	0.9000	0.9000	0.9000	0.9000

Best performance is highlighted. SVM and DTB classifiers are trained in the context of 5-fold nested CV, but CNN is trained while using 80%–20% for training and validation, respectively.

Table 7
Using unseen data samples of ADNI to test the generalization of our proposed techniques.

Method		**ACC**	**Sen**	**SP**	**GM**
(a): Combining shearlet-based descriptors with deep features to test the generalization of a classifier					
Fusion #2	SVM	0.6450 ± 0.0371	0.4800 ± 0.0671	0.8100 ± 0.0418	0.6221 ± 0.0425
(b): Combining shearlet-based descriptors with deep features to test the generalization of the custom CNN					
Fusion #2	CNN	0.7000	0.6000	0.8000	0.6928

Best models are determined based on the highest geometric mean that is computed in Table 6.

In the following Section 4, we first discuss results classifying for AD while using shearlet-based descriptors alone, and report results while using two classifiers: SVM and DTB. Similarly, we report the performance of the SVM and DTB classifiers when being trained based on the various deep features computed from different fine-tuned CNN architectures. Our classical ML classifiers are trained and validated in the context of nested CV.

However, given the fact that training a DL model is time consuming, our custom CNN models are trained from scratch while using 80% of the data for training and 20% for validation.

4. Results and analysis

4.1. Classification results using shearlet-based descriptors

In this section, results are reported when shearlet-based techniques are used for feature extraction. Each MRI sample in the ADNI dataset is transformed using the 3D shearlet with $S = 2$ and $K = 49$ orientations per scale. Since a fewer number of slices are available for the OASIS dataset, the number of scales is set to $S = 2$ and $K = 13$ per scale. As described in Section 2.1, the average of all directionalities of an MRI slice is computed. Our averaging strategy yields dramatically fewer descriptors. For example, if 20 textural features were computed while using the

Table 8
OASIS classification performance while combining shearlet-based descriptors with deep features for CN vs. AD.

Method		ACC	Sen	SP	GM
(a): Combining shearlet-based descriptors with deep features to train/validate a classifier					
Fusion #1	SVM	0.8944 ± 0.0412	0.9000 ± 0.0724	0.8889 ± 0.0878	0.8922 ± 0.0422
	DTB	0.9000 ± 0.0421	0.9000 ± 0.0724	0.9000 ± 0.0824	0.8981 ± 0.0424
Fusion #2	SVM	0.9278 ± 0.0541	0.9111 ± 0.0497	**0.9444 ± 0.0680**	0.9274 ± 0.0539
	DTB	0.8944 ± 0.0692	0.8889 ± 0.0680	0.9000 ± 0.0724	0.8944 ± 0.0691
(b): Combining shearlet-based descriptors with deep features to train/validate the custom CNN					
Fusion #1	CNN	0.8611	0.9130	0.7692	0.8381
Fusion #2	CNN	**0.9444**	**0.9565**	0.9231	**0.9397**

Best performance is highlighted. SVM and DTB classifiers are trained in the context of 5-fold nested CV, but CNN is trained while using 80%–20% for training and validation, respectively.

Table 9
Using unseen data samples of OASIS to test the generalization of our proposed techniques.

Method		ACC	Sen	SP	GM
(a): Combining shearlet-based descriptors with deep features to test the generalization of a classifier					
Fusion #2	SVM	0.8000 ± 0	0.7000 ± 0	0.9000 ± 0	0.7937 ± 0
(b): Combining shearlet-based descriptors with deep features to test the generalization of the custom CNN					
Fusion #2	CNN	0.8000	0.7000	0.9000	0.7937

Best models are determined based on the highest geometric mean that is computed in Table 8.

GLCM technique for each of the 98 directionalities of the 3D shearlet sub-bands, we would end up having 1, 960 descriptors per slice. In total, 90 (no. of slices) × 1, 960 = 176, 400 descriptors per MRI sample in the ADNI dataset. As a result, a problem of overfitting the model to the descriptors may arise due to a very large number of descriptors and the small number of training samples.

The parameter setting of the texture-based techniques can be adjusted. Various parameters are tested for each texture-based technique to build a classification model. The hyperparameters that maximize the performance on the validation split of the dataset are selected. Therefore, the parameters used for the ADNI dataset are not necessarily the same as for the OASIS dataset. The results reported in this paper for the two datasets are based on the hyperparameters shown in Table 5.

The performance of two classifiers, i.e., SVM and DTB, are reported in Table 1 for the ADNI dataset and Table 2 for the OASIS dataset based on five-fold nested CV. When using the ADNI dataset, the proposed descriptors extracted from the 3D shearlet coefficients can achieve reasonable accuracy between 68% and 70% when fed into an SVM or DTB classification model. We notice that the highest accuracy is 0.7050, obtained while classifying this dataset utilizing SFTA descriptors coupled with an SVM classifier with a corresponding sensitivity of 0.7300, specificity of 0.6800, and geometric mean of 0.7009.

Similarly, our 3D shearlet technique is applied to the OASIS dataset. On the OASIS dataset, our technique achieves accuracies between 62% and 68%. The highest obtained accuracy is 0.6889, obtained while classifying the dataset utilizing GLCM textural features coupled with an SVM classifier, with a corresponding sensitivity of 0.6444, a specificity of 0.7333, and a geometric mean of 0.6860. From Tables 1 and 2, performance values of GLCM, LBP, and LOSIB textural features for both datasets are comparable. On the other hand, SFTA shows better performance for the ADNI data samples. The reason for this may be the different number of slices in the ADNI and OASIS datasets, and, hence, the different number of descriptors computed by each method.

4.2. *Classification results using fine-tuned pretrained CNNs*

In this section, the ADNI and OASIS datasets are used to fine-tune a CNN model. In order to be able to fine-tune a CNN model based on an MRI volume, the model needs to be adjusted, as detailed in Section 2.2. To fine-tune a CNN model, the dataset used in this study is divided into 80% for training and 20% for validation. The purpose of such a division is to select and monitor the hyperparameters and the model performance. A momentum value of 0.9 and a small learning rate of 0.0001 are chosen, while utilizing stochastic gradient descent [56] for fine-tuning a model. Fine-tuned pretrained CNN models are trained for a maximum of 60 epochs with mini-batches of size 10. In this study, various CNN models are re-trained: SqueezeNet, MobileNet, Xception, and Inception. After fine-tuning a model, each model is then used to generate a feature vector for each MRI volume. Those extracted feature vectors of each CNN model are used to train two classifiers, i.e., SVM and DTB, in the context of five-fold nested CV, as described in Section 3.

Tables 3 and 4 show results of various CNN models when used as feature extractors after being fine-tuned using either the ADNI or the OASIS dataset. The baseline results for each CNN model for feature extraction are obtained without any fusion with shearlet-based descriptors. An SVM or DTB classifier is used with only the deep features. We observe that for both datasets, using fine-tuned MobileNet deep features leads to the highest classification results. In particular, the highest performance is obtained when these deep features are used to train a DTB model in the five-fold nested CV fashion.

4.3. *Classification results while combining shearlet-based descriptors with deep features*

In the two previous sections, we provided a comprehensive examination of our shearlet-based descriptors and deep features. Each set of descriptors is used to train an SVM or DTB classifier. The underlying principle of each method differs in generating a meaningful set of descriptors for an MRI sample, and that can influence the performance of the classifier.

Thus, we propose to combine both handcrafted and deep features together. The following fusions are compared:

- **Fusion #1**: All of our extracted handcrafted descriptors are aggregated as one data matrix (X), then aggregated with all deep features (Y). This fusion is general to all MRI brain scans. Thus, the number of descriptors for an MRI sample in the ADNI dataset is [GLCM (1800) + LBP (1620) + LOSIB (1440) + SFTA (270)] + [SqueezeNet (512) + MobileNet (1280) + Xception (2048) + Inception (2048)] = 11018 descriptors; the number of descriptors for an MRI sample in the OASIS dataset is [GLCM (640) + LBP (576) + LOSIB (512) + SFTA (672)] + [SqueezeNet (512) + MobileNet (1280) + Xception (2048) + Inception(2048)] = 8288 descriptors.

Table 10
Comparative analysis of related work for AD Classification.

Authors	Training/Validation Protocol	Separate Set for Testing	Subjects for Training/Validation	[a](RD/SbD)	ACC	Sen	SP
Reported performance of methods on the ADNI dataset for (sMCI vs. pMCI) classification							
Nanni et al. [9]	20-fold nested CV.	×	134 sMCI + 76 pMCI	SbD	0.671	0.355	0.865
Salvatore et al. [10]	20-fold nested CV.	×	134 sMCI + 76 pMCI	SbD	0.66	–	–
Feng et al. [16]	10-fold CV.	×	160 sMCI + 120 pMCI	SbD	0.7448	0.7111	0.7664
Lian et al. [27]	2-fold CV.	×	465 sMCI + 205 pMCI	SbD	0.81	0.53	0.85
Pan et al. [29]	5-fold CV.	✓	134 sMCI + 76 pMCI	SbD	0.62	–	–
Ours	80% of the data used for training and 20% used for validation.	×	50 CN + 50 sMCI + 50 pMCI + 50 AD	SbD	0.9000	0.9000	0.9000
[b]Ours	80% of the data used for training and 20% used for validation.	✓	50 CN + 50 sMCI + 50 pMCI + 50 AD	SbD	0.7000	0.6000	0.8000
Reported performance of methods on the OASIS dataset for (CN vs. AD) classification							
Jha et al. [15]	10-fold CV.	×	98 CN + 28 ADs	SbD	0.9006	0.9200	0.8778
Wang et al. [25]	49 subjects of each class were used for training, and the other 49 subjects from each class were used for validation.	×	98 CN + 28 AD + **70 AD (privately collected)**.	SbD	0.9765	0.9796	0.9735
Yagis et al. [24]	5-fold CV.	×	100 CN + 100 AD	RD	0.925	–	–
Yagis et al. [24]	5-fold CV.	×	100 CN + 100 AD	SbD	0.671	–	–
Hon and Khan [23]	5-fold CV.	×	100 CN + 100 AD	RD	0.9625	–	–
Mendoza et al. [28]	134 samples were used for training and 40 samples were used for validation.	×	87 CN + 87 AD	SbD	0.90	0.95	0.85
Ours	80% of the data used for training and 20% used for validation.	×	90 CN + 90 AD	SbD	0.9444	0.9565	0.9231
[b]Ours	80% of the data used for training and 20% used for validation.	✓	90 CN + 90 AD	SbD	0.8000	0.7000	0.9000

[a] Random division (RD) of slices that potentially the same patient slices can appear in both training and testing datasets. Subject-based division (SbD) is more appropriate because slices of a patient could appear only in the training or testing split.
[b] Our results reported in this raw are based on a separate set for testing, i.e., unseen MRI data.

- **Fusion #2**: Alternatively, the best performing set of handcrafted descriptors with the best performing set of deep features are fused. This fusion is specific to an MRI brain scans dataset. In the case of the ADNI dataset, SFTA led to the highest classification performance, while GLCM textural features led to the highest classification performance for the OASIS dataset. MobileNet deep features on the ADNI or OASIS datasets achieved the highest classification performance. Thus, the total number of descriptors in this fusion is 1550 and 1920 descriptors for the ADNI and OASIS datasets, respectively.

In Table 6 and Table 8, we explore the combination of shearlet-based features with deep features when being fed to an SVM or DTB classifier and not using our custom 1D CNN fusion model. When we use all of our handcrafted descriptors with deep features in Fusion #1 to classify the ADNI dataset, an accuracy of 0.9050 is obtained with an SVM classifier. However, a higher classification accuracy of 0.9350 is achieved in Fusion #2 with an SVM model while selecting only one of our shearlet-based descriptors (i.e., SFTA) and one of our deep features based on MobileNet. Furthermore, the generalization of this high performing model is examined on an unseen dataset that was hidden from start to end (as shown in Table 7). As such, an accuracy of 64.50% is obtained with a corresponding sensitivity of 0.4800, a specificity of 0.8100, and a geometric mean of 0.6221.

Similarly, our proposed combination is applied to classify the OASIS dataset. The highest classification accuracy of 90.00% is obtained while using a DTB with Fusion #1. However, classification accuracy of 92.78% is obtained by Fusion #2. Therefore, this model is selected to be tested on unseen data, and the accuracy on the unseen data is 80.00% with a corresponding sensitivity of 0.7000, a specificity of 0.9000, and a geometric mean of 0.7937 (as shown in Table 9).

In an attempt to reach a better classification model, we have built a custom CNN model (see Section 2.3). Such a model takes a feature vector as an input which will be further processed end to end to obtain a prediction, and hence, does not require a separate classifier such as an SVM or DTB. This custom CNN model is trained from scratch while utilizing 80% of the dataset for training and 20% for validation. The purpose of such a division is to select and monitor the hyperparameters and the model performance. As a result, a weight decay of 0.005 and a small learning rate of 0.001 are chosen. Our custom CNN model is optimized using adaptive moment estimation (Adam) [57] while being trained for 200 epochs with mini-batches of size 10.

Similar to before, our custom CNN model is examined on the concatenation of all shearlet-based descriptors with deep features in Fusion #1. Alternatively, our custom CNN model is examined while using only the best performing approach of each domain, i.e., one shearlet-based descriptor and one deep features. In Table 6, we observe that the validation geometric mean for classifying the ADNI dataset with the custom CNN for Fusion #2 is 0.9000. Therefore, we select Fusion #2 as the best model and test this model on the unseen dataset, which leads to an accuracy of 70.00% with a corresponding sensitivity of 0.6000, a specificity of 0.8000, and a geometric mean of 0.6928, as shown in Table 7. Similarly, in Table 8, when we use only GLCM textural features that are computed in the shearlet domain combined with MobileNet deep features in Fusion #2 to train and validate our custom CNN model on the OASIS scans, we obtain a validation geometric mean of 0.9397. This model is then tested on the unseen data to examine its generalization. The achieved accuracy on the unseen data is 80.00%, with a corresponding sensitivity of 0.7000, a specificity of 0.9000, and a geometric mean of 0.7937, as shown in Table 9.

Our custom CNN provides higher classification results on the unseen dataset than using SVM or DTB. This is an indication that it is less susceptible to overfitting. There is a gap between the validation results and the test results, especially on the ADNI datasets. As such, the validation accuracy and testing accuracy while using the ADNI dataset are 90.00% and 70.00%, respectively. However, as has been indicated in the dataset section, the ADNI MRI volumes that are used for training/validation are of the baseline visits of the patients, but our unseen data are of patients at different time-points. Thus, one can view our approach as effective in classifying patients' AD scans that are not necessarily of the same time-point.

4.4. Discussion

Many studies have been developed for AD classification using traditional ML and DL. In Table 10, we provide a summary of several state-of-the-art classification performances of methods using the ADNI and OASIS datasets. Although we applied the classification task of (CN + sMCI) vs. (pMCI + AD) while using the ADNI dataset, it can be compared with the classification task of (sMCI vs. pMCI) [53].

In particular, we compare the performance of our method on the ADNI dataset with results that were reported by Salvatore et al. [53]. We conducted our experiments for training a classifier while using the same patients' data samples of the ADNI dataset as Salvatore et al. When using the baseline of the ADNI dataset, Salvatore et al. reported the highest accuracy of 85%, with a corresponding sensitivity of 0.83, and a specificity of 0.87. This performance was achieved while using (PCA + FDR) extracted from the MRI samples with neuropsychological data. However, when only PCA features were used, the classification accuracy, sensitivity, and specificity were 0.72, 0.69, and 0.75, respectively. The classification performance improved with the neuropsychological data, but the authors did not report the performance while using neuropsychological data only. In comparison to Salvatore et al., our highest achieved accuracy was 93.50% while combining shearlet-based descriptors and deep features that were extracted from the MRI samples, coupled with an SVM model. When we examined our combined descriptors with a custom CNN model, it led to an accuracy of 90.00%. Additionally, we did not use neuropsychological data. It is also worth noting that we examined the generalization of our best-performing custom CNN model (i.e., achieved the highest geometric mean) on unseen data. Such a model led to an accuracy of 70.00%.

Furthermore, we also compare with the performance of a study done by Hon and Khan [23]. Our experiments were conducted in a similar manner for training a classifier while using the same patients' data samples of the OASIS dataset. Hon and Khan fine-tuned two CNN models while using each slice of an MRI as an independent example. When VGG16 was fine-tuned, they achieved an accuracy of 92.3%. However, a better classification accuracy of 96.25% was achieved while using Inception V4. Hon and Khan applied random split of slices, which might cause data leakage.

Similarly, Yagis et al. [24] fine-tuned VGG-16 and ResNet-50 models. Each slice of an MRI was considered as an independent labeled example from the OASIS dataset. When the split of slices was done randomly, where the same patient's slices can appear either in the training or the testing dataset, the highest reported classification accuracy was 92.5% while using ResNet-50. When ResNet-50 was trained in a subject-based split, where slices of a patient could appear only in the training or testing split exclusively, the highest classification accuracy was 67.1%. Hence, the accuracy drop of over 25% points gives an excellent indication of the importance of patient-based splitting in order to avoid data leakage. In contrast, our method achieves an accuracy of 94.4% with patient-based splitting.

Unlike Yagis et al. [24], and Hon and Khan [23], we used each MRI sample of the OASIS data as a volume. Our highest classification accuracy was 92.78% while combining shearlet-based descriptors and deep features, coupled with an SVM model. Also, we examined our combined descriptors with a custom CNN model, which achieved an accuracy of 94.44%. Furthermore, our custom CNN model, trained on the OASIS dataset that achieved the highest geometric mean on the validation set, was examined for its generalization on unseen data. The model accuracy was 80.00%.

In Table 10, only the study by Pan et al. [29] used a separate set for testing the generalization of a model.

When the OASIS dataset was used, some authors reported the performance based on the random split of slices. This will lead to over-optimistic results, because of possible "information leakage." As has been indicated earlier, Yagis et al. [24] demonstrated that a subject-based split is more appropriate, but a classifier can be expected to perform poorly. The studies by Mendoza et al. [28] and Wang et al. [25] selected a single slice to train and validate their respective models. Arguably, one slice might not be sufficient for predicting AD, since the practice in the hospitals by a radiologist is to use multiple slices [15]. In addition, Wang et al. utilized additional MRI samples of AD patients that were privately collected. The authors mentioned in their study that their eight-layer CNN model was empirically driven, but the CNN model was adjusted to maximize the classification performance on the validation split. Therefore, the chosen model and hyperparameters were biased to maximize the performance on the validation set.

The main advantages of our proposed techniques can be summarized as:

- A 3D shearlet transform is used to arrive at a compact feature representation per MRI sample.
- A 2.5D approach is used by stacking MRI slices in depth and, because of transfer learning, a compact and high-level semantic deep feature representation is obtained that leads to an effective classifier.
- A shallow CNN model is proposed to classify AD using a combination of shearlet-based descriptors and deep features. Such a model outperformed the simple concatenation approach while using a traditional ML classifier.
- We train and validate our model with two data repositories. The generalization of the model is tested on unseen data, and the performance is promising. Particularly, Fusion #2 led to the best classification results.

Although the proposed system is not computationally intensive, the main drawback of this work is the requirement for a couple of steps to ensure building a model that is less susceptible to overfitting.

5. Conclusion and future work

Although a cure for AD does not exist, the drugs in a trial show that patients who are diagnosed early benefit the most [58]. Thus, a robust technique for classification of AD must be advanced. In this paper, we have proposed a combination of 3D shearlet-based descriptors with deep features for the classification of AD. We rigorously evaluate our proposed technique utilizing a commonly used training and validation protocol. The performance of our proposed technique is highly ranked in comparison with previously published studies for AD classification. Most importantly, we further verify the capability of our technique to generalize on new data samples; i.e., a separate testing set of MRI volumes is utilized. Particularly, we showed that our customized CNN model is less susceptible to overfitting and can effectively generalize on unseen MRI data. In the future, we plan to examine our approach on other datasets with a similar nature of complexity; for example, the Parkinson's Progression Markers Initiative (PPMI) data repository [59]. Although our techniques are examined for the classification task, it would be interesting to extend our shearlet descriptors and DL strategies to segmentation tasks. For example, the multimodal Brain Tumor Image Segmentation Benchmark (BRATS)[6] is worth exploring.

Declaration of competing interest

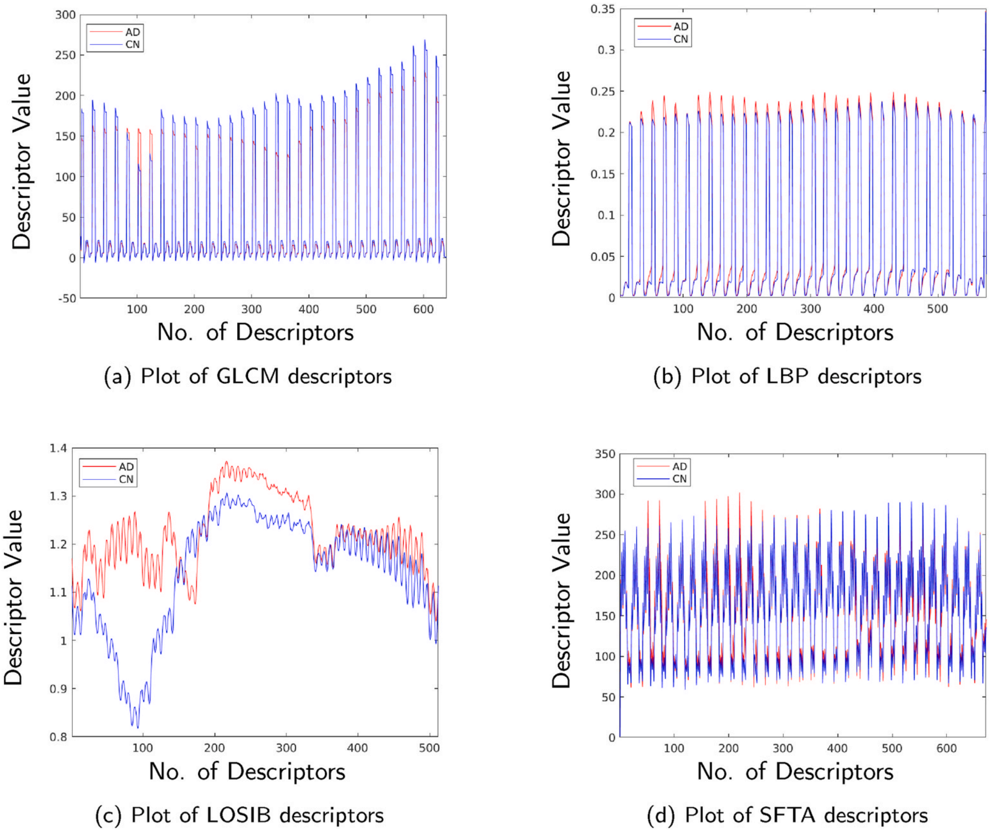
None Declared.

Acknowledgment

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Appendix A

In Section 2.1, we described the process of reducing the shearlet sub-bands to only one averaged sub-band for each MRI slice. A set of descriptors from each slice are extract. For each texture-based method, all sets of descriptors of all slices are concatenated to form the feature vector of an MRI sample. In Fig. 4, an example of different texture descriptors in the 3D shearlet transform of a CN and an AD sample is shown. In this example, the descriptor values are showing only subtle differences except for the LOSIB descriptor.



(a) Plot of GLCM descriptors

(b) Plot of LBP descriptors

(c) Plot of LOSIB descriptors

(d) Plot of SFTA descriptors

Fig. 4. Descriptors extracted from one CN and another from AD MRI sample using the OASIS dataset.

[6] The BRATS benchmark can be accessed via http://www.braintumorsegmentation.org/.

Appendix B

The details of our custom CNN activations and parameters are provided in this appendix (as shown in Table 11).

Table 11
Our custom 1D CNN activations and parameters details.

Name	Activations	Parameters
Combined Input	1 × 1920 × 1	–
Conv_1	1 × 1920 × 8	Weights 1 × 7 × 1 × 8 Bias 1 × 1 × 8
Batchnorm_1	1 × 1920 × 8	Offset 1 × 1 × 8 Scale 1 × 1 × 8
ReLU_1	1 × 1920 × 8	–
MaxPool_1	1 × 960 × 8	–
Conv_2	1 × 960 × 16	Weights 1 × 7 × 8 × 16 Bias 1 × 1 × 16
Batchnorm_2	1 × 960 × 16	Offset 1 × 1 × 16 Scale 1 × 1 × 16
ReLU_2	1 × 960 × 16	–
MaxPool_2	1 × 480 × 16	–
Conv_3	1 × 480 × 32	Weights 1 × 7 × 16 × 32 Bias 1 × 1 × 32
Batchnorm_3	1 × 480 × 32	Offset 1 × 1 × 32 Scale 1 × 1 × 32
ReLU_3	1 × 480 × 32	–
MaxPool_3	1 × 240 × 32	–
Conv_4	1 × 240 × 64	Weights 1 × 7 × 32 × 64 Bias 1 × 1 × 64
Batchnorm_4	1 × 240 × 64	Offset 1 × 1 × 64 Scale 1 × 1 × 64
ReLU_4	1 × 240 × 64	–
FC_1	1 × 1 × 120	Weights 120 × 15360 Bias 120 × 1
ReLU_5	1 × 1 × 120	–
Batchnorm_5	1 × 1 × 120	Offset 1 × 1 × 120 Scale 1 × 1 × 120
FC_2	1 × 1 × 120	Weights 120 × 120 Bias 120 × 1
ReLU_6	1 × 1 × 120	–
Batchnorm_6	1 × 1 × 120	Offset 1 × 1 × 120 Scale 1 × 1 × 120
FC_3	1 × 1 × 120	Weights 120 × 120 Bias 120 × 1
ReLU_7	1 × 1 × 120	–
Batchnorm_7	1 × 1 × 120	Offset 1 × 1 × 120 Scale 1 × 1 × 120
FC_4	1 × 1 × 2	Weights 2 × 120 Bias 2 × 1
SoftMax	1 × 1 × 2	–

Appendix C

In this appendix, we provide the performance of each fold of our best-performing technique in Tables 12–15. Also, the training/validation loss plot of our customized CNN model is provided in Fig. 5.

Table 12
The performance of each fold while training/validating an SVM model in the context of nested 5-fold CV with **Fusion #2** descriptors extracted from the ADNI data.

Fold No.	ACC	Sen	SP	GM
1	0.9000	0.9000	0.9000	0.9000
2	0.9250	0.9000	0.9500	0.9247
3	0.9500	1.0000	0.9000	0.9487
4	0.9750	0.9500	1.0000	0.9747
5	0.9250	0.9500	0.9000	0.9247

Table 13
The performance of each fold while using unseen ADNI data to test the SVM model generalization that obtained in Table 12.

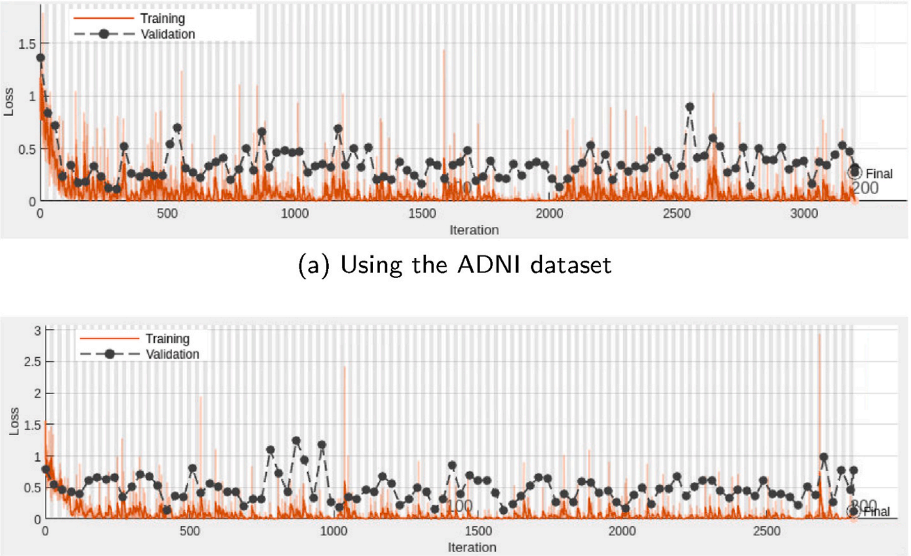
Fold No.	ACC	Sen	SP	GM
1	0.7000	0.6000	0.8000	0.6928
2	0.6250	0.4500	0.8000	0.6000
3	0.6000	0.4500	0.7500	0.5809
4	0.6500	0.4500	0.8500	0.6185
5	0.6500	0.4500	0.8500	0.6185

Table 14
The performance of each fold while training/validating an SVM model in the context of nested 5-fold CV with **Fusion #2** descriptors extracted from the OASIS data.

Fold No.	ACC	Sen	SP	GM
1	0.9444	0.9444	0.9444	0.9444
2	0.9444	0.8889	1.0000	0.9428
3	0.9722	0.9444	1.0000	0.9718
4	0.9444	0.9444	0.9444	0.9444
5	0.8333	0.8333	0.8333	0.8333

Table 15
The performance of each fold while using unseen OASIS data to test the SVM model generalization that obtained in Table 14.

Fold No.	ACC	Sen	SP	GM
1	0.8000	0.7000	0.9000	0.7937
2	0.8000	0.7000	0.9000	0.7937
3	0.8000	0.7000	0.9000	0.7937
4	0.8000	0.7000	0.9000	0.7937
5	0.8000	0.7000	0.9000	0.7937



(a) Using the ADNI dataset

(b) Using the OASIS dataset

Fig. 5. The training and validation loss values of the proposed custom CNN while using **Fusion #2** descriptors for each corresponding dataset.

Appendix D

In this appendix, a breakdown of the required time for each step in our proposed system is provided. Table 16 shows the required time to build the whole system using 200 MRI samples from the ADNI data repository. It only requires ≈ 37 min for feature extraction and training/validating a custom CNN while using Fusion #2 descriptors. Similarly, Table 17 shows the required time for building the whole system using 180 MRI samples from the OASIS data repository. While using Fusion #2 descriptors for training/validating a custom CNN, it only requires ≈ 10 min.

All experiments were conducted on a computer with an Intel Core i7-9700 3.00 GHz CPU, and 8 GB NVIDIA GeForce RTX 2080 graphics card.

Table 16
The required time starting from feature extraction to training/validating the custom CNN model based on Fusion #2, extracted from the ADNI dataset.

Step	Time (Minutes)
SFTA in 3D-ST	8.18
MobileNet fine-tuning	26.56
Deep feature extraction	0.50
Custom CNN training and validation	1.51
Total time	36.75

Table 17
The required time starting from feature extraction to training/validating the custom CNN model based on Fusion #2, extracted from the OASIS dataset.

Step	Time (Minutes)
GLCM in 3D-ST	2.0
MobileNet fine-tuning	5.67
Deep feature extraction	0.16
Custom CNN training and validation	2.27
Total time	10.1

References

[1] Alzheimer's Disease International, The global impact of dementia: an analysis of prevalence, incidence, cost and trends, 2015, 24–24.
[2] R.G. Stefanacci, "The costs of Alzheimer's disease and the value of effective therapies, Am. J. Manag. Care 17 (2011) S356–S362.
[3] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group* under the auspices of department of health and human services task force on Alzheimer's disease, Neurology 34 (7) (1984), 939–939.
[4] Y. Gandon, D. Olivie, D. Guyader, C. Aube, F. Oberti, V. Sebille, Y. Deugnier, Non-invasive assessment of hepatic iron stores by MRI, Lancet 363 (9406) (2004) 357–362.
[5] A. Coimbra, D.S. Williams, E.D. Hostetler, The role of MRI and PET/SPECT in Alzheimer's disease, Curr. Top. Med. Chem. 6 (6) (2006) 629–647.
[6] A.S. Lundervold, A. Lundervold, An overview of deep learning in medical imaging focusing on MRI, Z. Med. Phys. 29 (2) (2019) 102–127.
[7] M. Tanveer, B. Richhariya, R. Khan, A. Rashid, P. Khanna, M. Prasad, C. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: a review, ACM Trans. Multimed Comput. Commun. Appl 16 (1s) (2020) 1–35.
[8] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: a systematic literature review, Comput. Methods Progr. Biomed. 187 (2020) 105242.
[9] L. Nanni, S. Brahnam, C. Salvatore, I. Castiglioni, A.D.N. Initiative, et al., "Texture descriptors and voxels for the early diagnosis of Alzheimer's disease, Artif. Intell. Med. 97 (2019) 19–26.
[10] C. Salvatore, A. Cerasa, P. Battista, M.C. Gilardi, A. Quattrone, I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach, Front. Neurosci. 9 (2015) 307.
[11] M.S. Albert, S.T. DeKosky, D. Dickson, B. Dubois, H.H. Feldman, N.C. Fox, A. Gamst, D.M. Holtzman, W.J. Jagust, R.C. Petersen, et al., "The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's Dementia 7 (3) (2011) 270–279.
[12] T. Altaf, S.M. Anwar, N. Gul, M.N. Majeed, M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features, Biomed. Signal Process Contr. 43 (2018) 64–74.
[13] M. Faturrahman, I. Wasito, N. Hanifah, R. Mufidah, Structural MRI classification for Alzheimer's disease detection using deep belief network, in: *2017 11th International Conference on Information & Communication Technology and System (ICTS)*, IEEE, 2017, pp. 37–42.
[14] S. Lee, H. Lee, K.W. Kim, Magnetic resonance imaging texture predicts progression to dementia due to Alzheimer disease earlier than hippocampal volume, J. Psychiatry Neurosci.: JPN 45 (1) (2020) 7.
[15] D. Jha, J.-I. Kim, and G.-R. Kwon, "Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network," Journal of Healthcare Engineering, vol. 2017, 2017.

[16] J. Feng, S.-W. Zhang, L. Chen, A. D. N. I. ADNI, et al., Identification of Alzheimer's disease based on wavelet transformation energy feature of the structural MRI image and NN classifier, Artif. Intell. Med. 108 (2020) 101940.

[17] S. Zhou, J. Shi, J. Zhu, Y. Cai, R. Wang, Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image, Biomed. Signal Process Contr. 8 (6) (2013) 688–696.

[18] J. He, H. Ji, X. Yang, Rotation invariant texture descriptor using local shearlet-based energy histograms, IEEE Signal Process. Lett. 20 (9) (2013) 905–908.

[19] Y. Dong, D. Tao, X. Li, J. Ma, J. Pu, Texture classification and retrieval using shearlets and linear regression, IEEE Transactions on Cybernetics 45 (3) (2015) 358–369.

[20] K. Meshkini, H. Ghassemian, Texture classification using Shearlet transform and GLCM, in: *Electrical Engineering (ICEE), 2017 Iranian Conference on*, IEEE, 2017, pp. 1845–1850.

[21] U.R. Acharya, S.L. Fernandes, J.E. WeiKoh, E.J. Ciaccio, M.K.M. Fabell, U.J. Tanik, V. Rajinikanth, C.H. Yeong, Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques, J. Med. Syst. 43 (9) (2019) 1–14.

[22] R. Jain, N. Jain, A. Aggarwal, D.J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images, Cognit. Syst. Res. 57 (2019) 147–159.

[23] M. Hon, N.M. Khan, "Towards Alzheimer's disease classification through transfer learning, in: 2017 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2017, pp. 1166–1169.

[24] E. Yagis, A.G.S. De Herrera, L. Citi, "Generalization performance of deep learning models in neurodegenerative disease classification," in, in: 2019 *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2019, pp. 1692–1698.

[25] S.-H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang, H. Cheng, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling, J. Med. Syst. 42 (5) (2018) 85.

[26] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, V. Calhoun, A.D.N. Initiative, et al., "Deep residual learning for neuroimaging: an application to predict progression to alzheimer's disease, J. Neurosci. Methods (2020) 108701.

[27] C. Lian, M. Liu, J. Zhang, D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, in: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018.

[28] R. Mendoza-Léon, J. Puentes, L.F. Uriza, M.H. Hoyos, "Single-slice Alzheimer's disease classification and disease regional analysis with supervised switching autoencoders, Comput. Biol. Med. 116 (2020) 103527.

[29] D. Pan, A. Zeng, L. Jia, Y. Huang, T. Frizzell, and X. Song, "Early detection of Alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning," Front. Neurosci., vol. 14, 2020.

[30] P. Cao, J. Gao, Z. Zhang, Multi-view based multi-model learning for MCI diagnosis, Brain Sci. 10 (3) (2020) 181.

[31] L.-K. Soh, C. Tsatsoulis, Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices, IEEE Trans. Geosci. Rem. Sens. 37 (2) (1999) 780–795.

[32] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics 6 (1973) 610–621.

[33] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 7 (2002) 971–987.

[34] O. García-Olalla, E. Alegre, L. Fernández-Robles, V. González-Castro, Local oriented statistics information booster (LOSIB) for texture classification, in: *2014 22nd International Conference on Pattern Recognition*, IEEE, 2014, pp. 1114–1119.

[35] A.F. Costa, G. Humpire-Mamani, A.J.M. Traina, An efficient algorithm for fractal analysis of textures, in: *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference on*, IEEE, 2012, pp. 39–46.

[36] G. Kutyniok, D. Labate, Introduction to Shearlets, *Shearlets*, 2012, pp. 1–38.

[37] T. Hastie, R. Tibshirani, Classification by pairwise coupling, in: Advances in Neural Information Processing Systems, 1998, pp. 507–513.

[38] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.

[39] C. Gaser, F. Kurth, Manual computational anatomy toolbox-CAT12, in: Structural Brain Mapping Group at the Departments of Psychiatry and Neurology, University of Jena, 2017.

[40] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, R.L. Buckner, Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults, J. Cognit. Neurosci. 19 (9) (2007) 1498–1507.

[41] V. Rajinikanth, A.N. Joseph Raj, K.P. Thanaraj, G.R. Naik, A customized VGG19 network with concatenation of deep and handcrafted features for brain tumor detection, Appl. Sci. 10 (10) (2020) 3429.

[42] G. Kutyniok, W.-Q. Lim, G. Steidl, Shearlets: theory and applications, GAMM-Mitteilungen 37 (2) (2014) 259–280.

[43] S. Alinsaif, J. Lang, Texture Features in the Shearlet Domain for Histopathological Image Classification, *BMC Medical Informatics and Decision Making*, 2020.

[44] S. Al-Insaif, Shearlet-based Descriptors and Deep Learning Approaches for Medical Image Classification," Ph.D. Dissertation, Université d'Ottawa/University of Ottawa, 2021.

[45] D.A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Rem. Sens. 28 (1) (2002) 45–62.

[46] J.N. Kather, C.-A. Weis, F. Bianconi, S.M. Melchers, L.R. Schad, T. Gaiser, A. Marx, F.G. Zöllner, Multi-class texture analysis in colorectal cancer histology, Sci. Rep. 6 (2016) 27988.

[47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, " ImageNet, A large-scale hierarchical image database, in: 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 248–255.

[48] S. Sharma, R. Mehra, Breast cancer histology images classification: training from scratch or transfer learning? ICT Express 4 (4) (2018) 247–254.

[49] F.N. Iandola, S. Han, M.W. Moskewicz, K. Ashraf, W.J. Dally, K. Keutzer, SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size, arXiv preprint arXiv:1602.07360 (2016).

[50] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[51] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[52] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[53] C. Salvatore, A. Cerasa, I. Castiglioni, "MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia 24 months before probable diagnosis, Front. Aging Neurosci. 10 (2018) 135.

[54] P. Thanh Noi, M. Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery, Sensors 18 (1) (2018) 18.

[55] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.

[56] K.P. Murphy, Machine Learning: a Probabilistic Perspective, MIT Press, 2012.

[57] D. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv: 1412.6980 (2014).

[58] J. Sevigny, P. Chiao, T. Bussière, P.H. Weinreb, L. Williams, M. Maier, R. Dunstan, S. Salloway, T. Chen, Y. Ling, et al., The antibody aducanumab reduces aβ plaques in Alzheimer's disease, Nature 537 (7618) (2016) 50–56.

[59] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, et al., The Parkinson progression marker initiative (PPMI), Prog. Neurobiol. 95 (4) (2011) 629–635.