Applied Soft Computing 106 (2021) 107371

Contents lists available at ScienceDirect

Applied Soft Computing

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

FAF-DRVFL: Fuzzy activation function based deep random vector functional links network for early diagnosis of Alzheimer disease

Rahul Sharma [a,*], Tripti Goel [a], M. Tanveer [b], Shubham Dwivedi [a], R. Murugan [a]

[a] Biomedical Imaging Lab, National Institute of Technology Silchar, Assam, 788010, India
[b] Department of Mathematics, Indian Institute of Technology Indore, Simrol, Indore, 453552, India

ARTICLE INFO

Article history:
Received 30 September 2020
Received in revised form 16 February 2021
Accepted 25 March 2021
Available online 8 April 2021

Keywords:
Alzheimer's disease
Transfer Learning
Mild Cognitive Impairment
Neuroimaging
Random vector functional link

ABSTRACT

Alzheimer's disease (AD) is a degenerative neural condition marked by gradual memory loss and cognitive impairment. It is irreversible in nature and leads to progressive cerebral cortex atrophy. Therefore, structural Magnetic Resonance Imaging (sMRI) is an important tool that can be used for early-stage prediction of AD. Currently, deep learning networks are used for the diagnosis of AD, but it suffers from the limitations of gradient descent training of deep networks like local minima, slow learning speed, and overfitting. Also, there is a need to select hyperparameters like learning rate, momentum, number of epochs, and regularization coefficient. This paper proposes a deep non-iterative random vector functional link (RVFL) neural network. First, the MRI images' features are extracted using transfer learning, and the classification of the extracted features is done using a non-iterative random vector initialized RVFL network. At the hidden layer of the RVFL classifier, the fuzzy activation function (FAF), is used to calculate the hidden layer's output. The proposed algorithm has been evaluated and compared with the state-of-the-art methods on the ADNI dataset consisting of Cognitive Normal (CN), AD, converter Mild Cognitive Impairment (cMCI) and non-converter Mild Cognitive Impairment (ncMCI) MRI images. The performance achieved for CN vs AD diagnosis includes accuracy (86.67%), sensitivity (83.33%), specificity (88.89%), precision (83.33%), recall (83.33%) and F-score(86.07%) as well as Receiver Operating Characteristics shows that proposed method outperforms over several compared methods.

1. Introduction

Alzheimer's disease (AD) is a progressive, untreatable, and neuro-degenerative condition most often seen in the elderly persons and it progressively affects brain efficiency. It is irreversible in nature and is characterized by structural changes in the cortex region. AD is the prevailing form of dementia, as it accounts for 60% to 80% of the cases. In 2019, it was estimated that the global economic costs for dementia treatment were around 1 trillion dollars. By 2030, it is predicted that this economy will surge to double its present value [1]. Prodome diagnosis of AD, especially at its transition state, commonly known as Mild Cognitive Impairment (MCI), is especially crucial so that one can take therapeutic measures and delay the disease progression. MCI consist of two subclasses commonly known as converter MCI (cMCI) and non-converter MCI (ncMCI). When patient acquires the state of AD from MCI state, this type of MCI is known as cMCI. When the patient improves the lifestyle and does not attain the AD state even though MCI state was acquired, that comes under the category of ncMCI.

Current medical imaging systems render a good amount of detail about the subject under study, making image analysis a robust AD diagnosis tool. Computer-aided diagnosis (CAD) systems learn patterns associated with cerebral neurodegeneration by exploiting the images' information. The brain changes with MCI and AD progression could be quantified with clinical neuroimaging techniques such as Positron-Emission Tomography (PET) and Magnetic Resonance Imaging (MRI). For instance, high-resolution information of brain structure is rendered by T1-weighted MRI, which accurately quantifies the structural metrics such as shape, volume of hippocampus [2], gray matter, white matter [3], as well as cortical thickness [4]. Likewise, the regional metabolic rate of glucose in cerebral region is indicated by 18-Fluoro-DeoxyGlucose PET (18F-FDG-PET or FDG-PET), makes it feasible to quantify the metabolic activity of the tissues. By analyzing these medical images, assistance in the diagnosis and prediction of AD is provided.

Recently, with the rapid advancement in computer science domain and the advent of medical analysis, Machine Learning

* Corresponding author.
E-mail addresses: rahul_rs@ece.nits.ac.in (R. Sharma), triptigoel@ece.nits.ac.in (T. Goel), mtanveer@iiti.ac.in (M. Tanveer), shubham_pg@ece.nits.ac.in (S. Dwivedi), murugan.rmn@ece.nits.ac.in (R. Murugan).

https://doi.org/10.1016/j.asoc.2021.107371
1568-4946/

(ML) and Deep Learning (DL) are gaining recognition in the area of medical imaging. Support Vector Machine (SVM) and its variants has shown promising results in disease classification [5–7]. Popular ML approaches for the classification of the AD is presentented in [8]. ML algorithms are used not only in medical imaging but also in many forecasting applications. One such application is forecasting of crude oil price [9] using support vector regression along with heuristic optimization techniques for obtaining optimal solution. Another application where ML outperforms is financial exchange rates forecasting [10] uses SVM for classification and provides how different kernel functions help improve performance. Solar radiation emission forecasting is yet another domain application of ML proposed in [11] which employed a state-of-the-art classifier i.e. Random Forest regression, and a feature selection techniques. Random Forest is also famous in classification problems when we are dealing with randomization. Ardekani et al. [12] selected neuropsychological scores and hippocampal volumetric integrity of MRI as features for a random forest classifier. Lebedeva et al. [13] took mini-mental state examination (MMSE) as a cognitive measure and extracted structural features of MRI to implement the random forest-based scheme. ML algorithms suffer from limitations because of the use of backpropagation for parameter tuning, network initialization. The training time will also increase because of backpropagation training.

Besides ML, DL also contributes to a wide variety of forecasting applications. Digital currency forecasting proposed in [14] implemented a Deep Neural Network (DNN) i.e. LSTM. Another most popular neural networks is, Convolutional Neural Network. These networks learn from a set of convolutional kernels. The most recent application is Covid detection [15], which uses an EfficientNet B0 model, a variant of the ResNet model. In [16], researchers implemented a 3D Convolutional network based on a 3D auto-encoder model to predict AD vs MCI vs Cognitive Normal (CN). Potential of AD diagnosis with DL based schemes is promising, as evident in previous studies [17]. However, DL's main issue is the use of backpropagation (BP) approach for tuning the parameter, which is very time-consuming and suffers from limitations of slow convergence, overfitting, and local minima. BP may lead to failure on convergence to single global maxima. Also, we have to manually choose the values of many hyperparameters such as learning rate, number of epochs, batch size, momentum, and regularization coefficient.

Pao et al. [18] proposed the randomized functional link network with a single hidden layer. Non Learnable, randomly initialization of bias and weight is done between input and hidden layers of the network [19]. Evolution from single network to deep links in Random Vector Functional Link (RVFL) has been proposed in [20]. RVFL not only has application in the medical field, but it outperforms on various other fields such as fault detection [21], electricity load prediction for a short span [22], Wind speed and the power generation predictions [23], where RVFL is used as a classifier and the model outperforms when compared to several ML techniques. Extreme Learning Machine (ELM) [24] is another non-iterative fast learning model but RVFL outperforms ELM. Therefore, in the present paper, the classification of the features extracted from DNN are done using improved RVFL classifier. At the hidden layer of RVFL classifier, Fuzzy Activation function (FAF) is also used to trace the features to the output space.

Major contributions of the proposed Fuzzy Activation Function based Deep RVFL (FAF-DRVFL) neural network is as follows:

1. Non-iterative feed-forward random weight initialized neural network is proposed to diagnose AD at the early stage that will be helpful to the clinicians for automatic diagnosis.
2. Features of the MRI images are extracted using Transfer Learning (TL), which provides robust features related to the disease.
3. Fuzzy activation function is used at the RVFL neural network's hidden layer to convert the features into non-linear space and remove the outliers from the features.
4. Diagnosis outcomes are contrasted with the state-of-the-art networks for CN versus AD, cMCI versus ncMCI, and CN versus cMCI diagnosis in terms of accuracy, sensitivity, specificity, precision, recall, F-score and ROC curve.

The motivation of the proposed FAF-DRVFL neural network are as follows:

1. Traditional DL algorithms are based on gradient descent learning, which has the limitations like improper learning rate, local minima, slow convergence because of tuning of weights and biases, overfitting issues, etc. Therefore, to avoid these issues, feed-forward RVFL is used to classify weights in which weights are randomly initialized and give better accuracy with faster speed.
2. Feature extraction is the main step for disease diagnosis; therefore, pretrained DL model, ResNet-50, is implemented to extract the features from MRI images.
3. MRI images may contain some outliers like motion artifacts, bias field correction, noise, etc. To address this issue, s-membership Fuzzy Activation Function (s-FAF) is used that will map the outliers to a crisp range of membership values.

The rest of the paper is organized as follows: Section 2 provides the literature review related to the AD diagnosis and FAF-DRVFL. Section 3 provides information about the dataset and the methodology proposed in the paper. Section 4 provides the results and discussion, and finally, Section 5 concludes the paper.

2. Literature review

Traditional ML algorithms observed the functional and anatomical neural lesions [25,26] for AD diagnosis. For instance, Zhang et al. [27] proposed a novel scheme that builds a latent representation, which helps explore the complex correlation between the feature and the label. Lei et al. [28] recommended the use of a discriminative sparse learning method. This method leads to the classification of AD stages using multimodal features and renders the clinical score's prediction jointly with relational regularization. Low generalization error and acceptable performance with limited training samples are the major reason behind SVM's use as a classifier. Rabeh et al. [29] suggested SVM-based classification of the AD in early-stage diagnosis. The proposed scheme used three brain sections, yielding a high accuracy on the OASIS MRI database. Bron et al. [30] proposed a comparative study on the feature selection method based on expert knowledge, t-statistics, SVM weight vector directly, and SVM weights as a significance map (p-map). Classification evaluations were obtained, and a fair accuracy was observed on the ADNI MRI database.

In most machine learning approaches, the negligence of strong texture descriptions is generally observed while the segmentation is being emphasized. This loss of detail is not appealing. Nonetheless, the need for image segmentation can be achieved only if compelling features and characteristics are retrieved from a complete image. DL can serve this purpose. Recently, DL based schemes are being used for AD diagnosis. Typically, DL uses a nonlinear multilayer networks which process the information to recognize feature quantities from data like color, shapes, text, images, etc. Without knowing the detailed internal process, an unerring mathematical model can be devised from huge database. This serves as the major advantage of DL.

Due to its semantic interpretation ability from the input data and the generation of optimized high-level features, DL in medical image analysis is encouraging. The DL field is thus advancing swiftly in delivering an accurate and early diagnosis of AD. Islam et al. [31] proposed an AD diagnosis scheme based on 2D

DenseNet in which the MRI data was sliced in three directions. Then for final diagnostic results, analysis and fusion of three parallel 2D DenseNet were done. A deep multitask multichannel learning approach was proposed by Liu et al. [32] in which brain disease classification via MRI data and deduction of clinical score regression was performed and the performance achieved was better than some existing state-of-the-art approach.

Suk et al. [33] proposed model combines two conceptually different model i.e. sparse regression schemes and DL based network. The prediction and diagnosis of AD and MCI were achieved using this Deep Ensemble Sparse Regression Network. The model showed high accuracy in the classification of the ADNI cohort. Maqsood et al. [34] also proposed a TL based model. In the proposed method, the model accessed pre-processed OASIS dataset as an input to AlexNet, and then classification was made. The parameters are fine-tuned, and the Monte-Carlo method was implemented to obtain simulations. Castro et al. [35] developed a model that uses sagittal MRI from OASIS and ADNI datasets and performs TL. The model used SVM as a classifier. The model proposed only used a single plane from 3-D MRI scans.

Zhang et al. [36] proposed a novel approach named sparse pre-trained RVFL. For a particular learning task, a sparse autoencoder is used to learn superior network parameters iteratively. Zhang et al. showed the advantage of using RVFL over the 16 diverse benchmark datasets. Another approach of transforming shallow RVFL into a deep network is proposed by Katuwal et al. [37]. In this approach, authors proposed a framework for RVFL inspired by stacked autoencoders. Also, adopted the approach of feature reuse by directly connecting the previous layer output to forecoming layers. Though the model attains a good performance but still matrix inversion is involved due to stacked autoencoders and RVFL classifier. When the dataset is huge and the feature dimensions are very large, then matrix inversion will be a costly deal with trade-off with time and memory over performance.

Cecotti [38] also proposed the Deep RVFL based model. Since RVFL is generally a single hidden layered model, the author increased the number of hidden layers and termed it as Deep RVFL. Model output was based on a multiclassifier approach in which the performance evaluation is done by combining the decision scores of the different classifiers. Here author used ELM and RVFL as multiple classifier. The model may look similar to the proposed approach, but the feature extraction part in the proposed algorithm is executed using a Deep Neural Network (DNN) i.e. ResNet-50, and the classification is performed using standard single layered RVFL. Also, we proposed s-FAF as the activation function in RVFL in the evaluation of output in the model proposed.

Since the weight allocation of the layer is random, Zhang et al. [39] proposed a technique to locate the ideal span for the arbitrary weights by a range scaling calculation. This paper gives the idea for parameter randomization and the variation in performance due to the direct link and effect of bias in the output. Another interesting approach was adopted in [40]. In this paper, authors adopted the multiple hidden layers in the RVFL based model and the output is extracted from each of the hidden layers. Then ensembles the output from the hidden layer through majority voting and hence the final output is calculated. 13 different real-world datasets were taken into consideration from the various domain to calculate output. We can say that ensembling increases the computational cost as compared to shallow RVFL. In this paper, the method performance remains outstanding when compared to other non-ensemble methods. Suganthan [41] compared various model which are non-iterative in nature with a closed form solution and concludes that RVFL outperforms the Extreme Learning Machine (ELM) is many aspects. Katuwal et al. [42] proposed a new ensemble classifier with the combination of RVFL and Decision Tree, and classification was done on 65 multiclass datasets. The decision tree implemented was both univariate and multivariate. The proposed model dominates the performance when compared to the existing state-of-the-art methods such as Random Forest and others.

This section gives us an idea of why we have selected DNN for feature extraction and RVFL as the classifier and provides some details on how we can improvise the model's performance parameter.

Table 1
Demographics details of the available subjects for AD diagnosis.

	AD	CN	MCI
No. of subjects	137	162	210
Average age	76	69	64

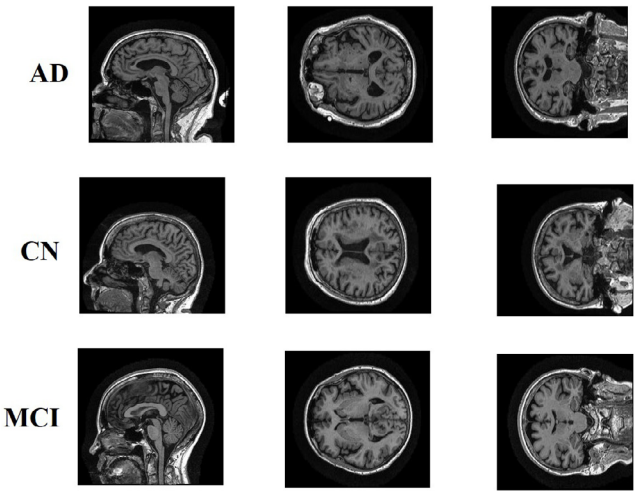

Fig. 1. 3 Plane view of MRI middle slices.

3. Methodology

A comprehensive description of the proposed method is presented in the following subsections.

3.1. Dataset

The MRI images related to AD diagnosis are taken from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [43]. More than 500 participants, including patients with MCI, AD, and CN, are involved in the model. Most of the patient ages range from 70–90. All the neuro scans used are T1 weighted MRI scans. MCI patients are further sub-classified into two categories, non-converter MCI (ncMCI) and converter MCI (cMCI), out of which 134 subjects are ncMCI, and 76 subjects are cMCI. The dataset has been split in a ratio of 70:30 as training and testing, respectively. The demographics details of the datasets are mentioned in Table 1.

3.2. Pre-processing

The 3D scans used has a slice count of 160 in each scan. Out of the 160 slices, we extract two pictures of each plane i.e., sagittal, coronal and axial, from entire MRI scans. The three-plane view of MRI is shown in Fig. 1. The Fig. 1 is just a plane extraction for one single slice. In the proposed model, we extracted two slices from each plane of the MRI scan. Therefore, total six slices set was extracted from each MRI scan. The slice gap used was two

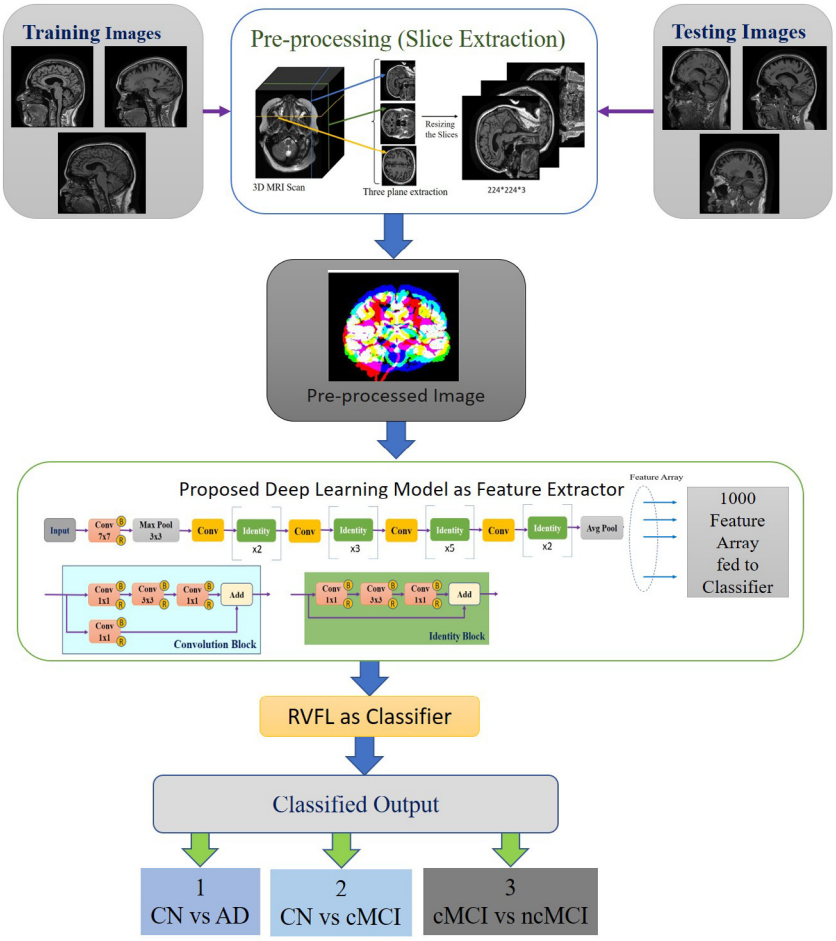

Fig. 2. Proposed Model.

i.e., after two slices from the middle next image will be extracted. Since these slices is fed to a deep network, the slice is resized into 224*224 dimensions.

3.3. Network architecture

In the proposed method, TL is adopted to extract the features from the pretrained network, ResNet-50 and FAF-RVFL is used as classifier. ResNet consists of an array of multiple convolutional and pooling layers that will extract the MRI images' high-level features [44]. A brief explanation of the DNN and RVFL is given in the proceeding section. The basic architecture of the proposed model is shown in Fig. 2.

In the proposed method, ResNet-50, a pretrained DNN is used for feature extraction. The convolutional layers with the combination of skip connections of DNN play the role of efficient feature extractor as represented in Fig. 3. In pretrained ResNet-50, there is no need to tune the parameters of the network. Batch normalization is done to normalize the parameter. Rectified Linear Unit (ReLU) is used as an activation function because of its threshold. Max pooling is used in the network to reduce the complexity by reducing the size of the feature extracted by the convolutional layer, thereby reducing the computational requirements.

After passing through all the layers, these features will be fed to a single hidden layer feed-forward neural network known as RVFL. Deep-RVFL (DRVFL) provides the network's validness by modifying linear non-separable data into lower dimension space. RVFL's main concept is to initialize the random parameter for the enhancement layer(also known as a hidden layer), as shown in Fig. 4. After the random parameter assignment, the parameter values are fixed and do not change during the training phase. It systematically calculates the parameter of the output layer generally by using the least square method. For calculating output weights, both the original data and the learned non-linear feature by the hidden layer are considered. The hidden layer's fixed weight diminishes the problem raised by iterative learning techniques, as in the iterative method, the model gets trapped in local minima, monotonous convergence, and overfitting. For MRI scans, which are a bit more complex in nature, BP based classifier may take a much longer training time [45]. RVFL is much faster than other classifiers because of the single hidden layer in the network and non iterative processing.

Given n samples with labels $s = (a_i, b_i)$; $a_i \in R^d$, $b_i \in R^c$, $i = 1, 2, \ldots, n$; where every samples is defined as $a_i = [a_i^1, a_i^2, \ldots, a_i^d]^T$ and $b_i = [b_i^1, b_i^2, \ldots, b_i^c]^T$. For RVFL network having E enhancement nodes, the output is given by:

$$C.O = Y \tag{1}$$

"C" represents a matrix formed by the concatenation of random features from the input and the hidden layer. "Y" matrix consists of "n" target vectors. "O" is the matrix for the output weights. The weight and biases between the input and the enhancement nodes are set random with values between $\{-x, x\}$ and $\{0, x\}$, respectively, "x" indicate a positive real scaling factor. The concatenate matrix formed consist of hidden parameter denoted by $h_i(a_i)$ will have the value for first instance will be $h_1(a_1) = T(O_1^h.a_1 + b_1)$, $h_E(a_1) = T(w_E^h.a_1 + b_E)$ and similar for "n" samples. Here $T(.)$ represents the activation function used in the RVFL, which will be non-linear in nature.

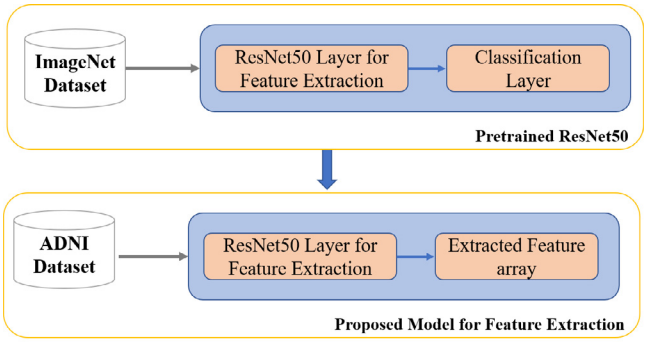

Fig. 3. Model for Feature extraction.

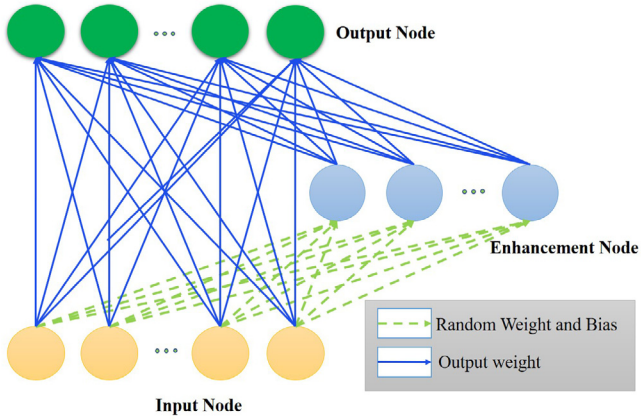

Fig. 4. Schematic for RVFL.

3.4. s- Fuzzy activation function

In the proposed algorithm, we implemented *s*-FAF [46] as the activation function because *s*-FAF maps the non-linear data from the input vector to the feature vector. In order to reduce the computational complexity, *s*-FAF uses only three linguistic variables; low, medium and high. Another advantage of *s*-FAF is that it compresses the outliers due to motion artifacts and others to the range of membership value of either near to zero or one. For the normal intensity values, the values are almost linear. For example, if the inputs are negative values, then it will be mapped to zero and inputs having values more positive will be mapped near to one. Inputs that contains average grayscale values means having more information will be mapped linearly. The *s*-FAF is defined by Eq. (2).

$$T(i,j,k,m) = \begin{cases} 0, & i \le j, \\ 2^{f-1}\left(\dfrac{i-j}{k-j}\right)^{f}, & j \le i \le m, \\ 1 - 2^{f-1}\left(\dfrac{i-j}{k-j}\right)^{f}, & m \le i \le k, \\ 1, & i \ge k. \end{cases} \tag{2}$$

Eq. (2) is the functional equation representation of the *s*-FAF. The slope of the *s*-FAF is controlled by "f", called fuzzifier. The function has two crossover points, *j*, and *k*, at 0 (minima) and 1 (maxima), respectively and one center point at *m*. Fig. 5 shows the s-shaped *s*-FAF used in RVFL network.

From Eq. (2), one can calculate the weight for the output node of RVFL by an approach called ridge regression, which helps in solving the problem:

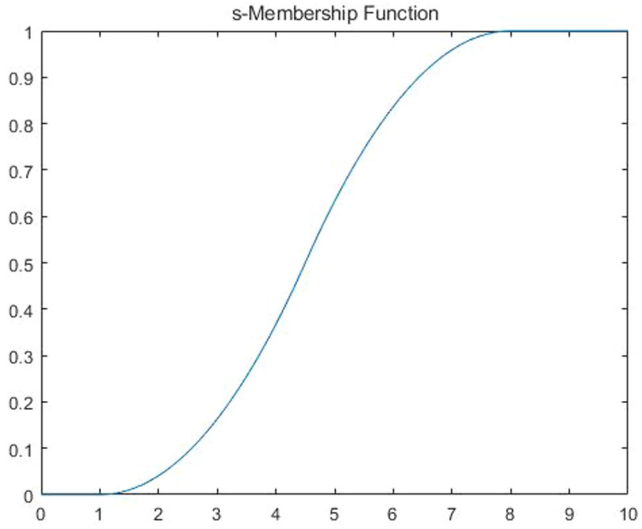

Fig. 5. s-Membership Function.

$$\sum_{i}^{l} (b_i - h_i * O) + r\|O\|^2, i = 1, 2, \ldots\ldots, n, \tag{3}$$

where h_i represents a vector having both input and random feature. The solution of the above equation is done by using regularization parameter r as

$$O = C^T \left((CC^T + 1/r)^{-1} Y\right). \tag{4}$$

This provides a RVFL based model with *s*-FAF to classify the disease into three classes. As *s*-FAF is proposed in the model, but we also evaluated the output with other standard activation function such as sigmoid, hard limit (hardlim), triangular basis (tribas), radial basis (radbas), sine function and sign function, just to compare the output with the proposed *s*-FAF.

In a summarized manner, we pre-processed the data and extracted the slices from the MRI scans. The images are then fed to the DNN based network, which extracts the features, and the extracted feature is fed to the FAF-DRVFL for fast and accurate classification. The models' main aspects are the parameters involved in the classification network are randomly initialized due to which the tuning of learnable parameters is almost negligible and hence the computational time gets reduced.

4. Results and discussion

The result and discussion section is further divided into eight subsection. Each section shows the comparative discussion related to the output generated from the proposed model and the compared models.

4.1. Experiments

In this section, we will discuss and compare the output from the proposed FAF-DRVFL model with the state-of-the-art DNN. The classifications are made in 3 classes, i.e., CN vs AD, CN vs cMCI, and cMCI vs ncMCI. The performance metric on which the models will be evaluated is accuracy, sensitivity, specificity, precision, recall, F-score and ROC curve. The flowchart of the proposed algorithm is shown in Fig. 6.

The next subsections deals with the tools used for implementing the models and generating the output.

Table 2
Algorithm comparison of proposed FAF-DRVFL model with different Activation Function (in %).

Activation Function	Class	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Score
s-FAF	**CN vs AD**	**86.67**	**83.33**	**88.89**	**83.33**	**83.33**	**86.07**
	CN vs cMCI	**81.81**	**1**	**33.33**	**80**	**100**	**88.89**
	cMCI vs ncMCI	**80**	**100**	**33.33**	**77.78**	**100**	**87.5**
sigmoid	CN vs AD	80	72.72	100	100	72.73	84.21
	CN vs cMCI	63.64	75	33.33	75	75	75
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54
sine	CN vs AD	66.67	100	44.44	54.55	100	70.59
	CN vs cMCI	72.73	100	0	72.73	100	84.21
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54
hardlim	CN vs AD	73.33	100	20	71.43	100	83.33
	CN vs cMCI	63.64	62.50	66.67	83.33	62.5	71.43
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54
tribas	CN vs AD	80	81.82	75	90	81.82	85.71
	CN vs cMCI	63.64	75	33.33	75	75	75
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54
radbas	CN vs AD	80	90.91	50	83.33	90.91	86.96
	CN vs cMCI	72.73	75	66.67	85.71	75	80
	cMCI vs ncMCI	80	85.71	66.67	85.71	85.71	85.71
sign	CN vs AD	80	100	25	78.57	100	88
	CN vs cMCI	70	85.71	33.33	75	85.71	80
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54

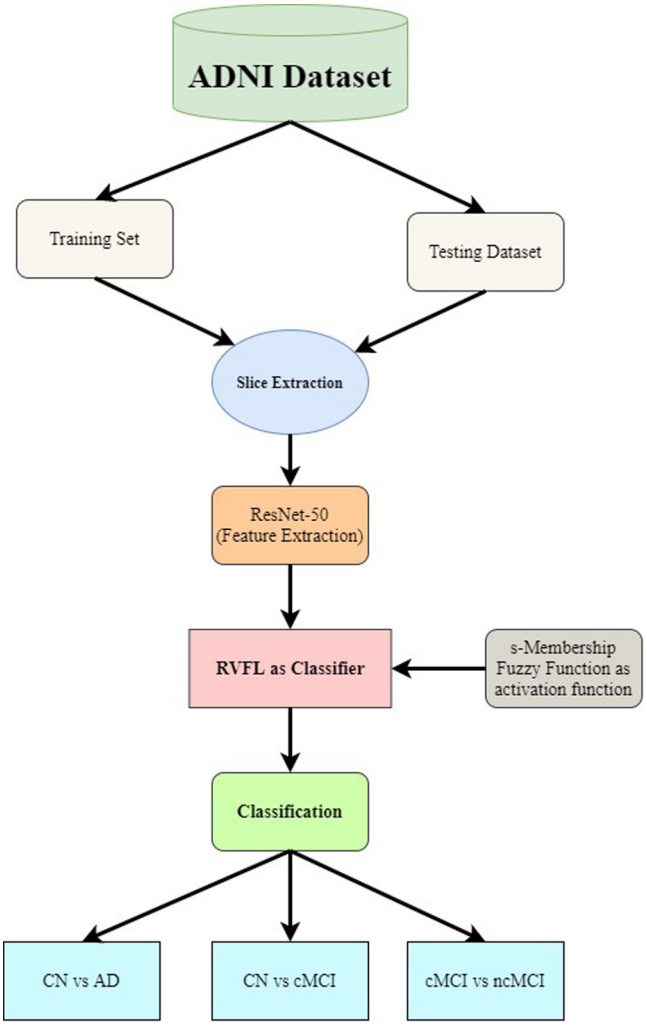

Fig. 6. Flowchart representation for the model implemented.

4.2. Implementation details

The experiments are executed using MATLAB 2020a software tool. All the training and testing of network along with dataset preparation is done on a machine with an Intel(R) Xenon(R) W-2133 CPU @3.60 GHz with a 64 GB RAM. The GPU used is NVIDIA Quadro P2200. All the experiments are done using ADNI dataset which includes structural MRI images of CN, AD, cMCI and ncMCI patients.

The next subsection deals with the comparison of the performance parameters of various activation function available for the proposed model.

4.3. Comparison with different activation functions

First of all, the model performance is tested with the different activation functions and *s*-FAF in RVFL network for classification in Table 2. Performance is compared in terms of accuracy, sensitivity, specificity, precision, recall and F-score. Activation functions help in classification by adding non-linearity to the system so that the system can handle non-linearity with much improved efficiency. Out of the various activation function used, the proposed activation function i.e., *s*-FAF outperforms in results with a maximum accuracy of 86.67% in classifying CN vs AD and 81.81% CN vs cMCI. The least performer was sine activation with the lowest accuracy in all the classes and zero specificity for one of the classes because of the miss classification of CN as cMCI and cMCI as CN. Specificity resembles the negatives that are identified positive. For the case of sine function, all the negative cases were classified as positive hence, the specificity tends to zero.

The next subsection shows the comparison of the proposed model with the similar single layered neural networks along with their various activation functions.

4.4. Comparison with state-of-art non-iterative classifier

Furthermore, the proposed method is contrasted with the state-of-the-art non-iterative classifiers. In comparisons, first the features are extracted using pretrained ResNet-50 DL network and then state-of-the-art non-iterative classifiers with different activation functions are used for classification. Classification performance of sF activation function based RVFL is compared with

Table 3
Comparison of the proposed FAF-DRVFL model with State-of-art non-Iterative Classifier (in %).

Classifier used	Class	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Score
ELM using sigmoid function	CN vs AD	73.33	63.64	100	100	63.64	77.78
	CN vs cMCI	63.64	75	33.33	75	75	75
	cMCI vs ncMCI	60	71.43	33.33	71.43	71.43	71.43
ELM using radial basis function	CN vs AD	73.33	100	65.56	60	100	75
	CN vs cMCI	72.73	100	0	72.73	100	84.21
	cMCI vs ncMCI	60	85.71	0	66.67	85.71	75
ELM using sine function	CN vs AD	40	16.67	55.56	20	16.67	18.18
	CN vs cMCI	72.73	87.5	33.33	77.78	87.5	82.35
	cMCI vs ncMCI	40	28.57	66.67	66.67	28.57	40
KELM using sigmoid kernel	CN vs AD	73.33	100	0	73.33	100	84.62
	CN vs cMCI	72.73	100	0	72.73	100	84.21
	cMCI vs ncMCI	70	100	0	70	100	82.35
KELM using polynomial kernel	CN vs AD	80	72.73	100	100	72.73	84.21
	CN vs cMCI	63.64	75	33.33	75	75	75
	cMCI vs ncMCI	50	57.14	33.33	66.67	57.14	61.54
KELM using radial basis kernel	CN vs AD	60	54.55	75	85.77	54.55	66.67
	CN vs cMCI	36.63	37.50	33.33	60	37.50	46.15
	cMCI vs ncMCI	40	42.86	33.33	60	42.86	50
KELM using wavelet kernel	CN vs AD	60	54.55	75	85.77	54.55	66.67
	CN vs cMCI	36.63	37.50	33.33	60	37.50	46.15
	cMCI vs ncMCI	40	42.86	33.33	60	42.86	50
Proposed FAF-DRVFL model	**CN vs AD**	**86.67**	**83.33**	**88.89**	**83.33**	**83.33**	**86.07**
	CN vs cMCI	**72.73**	**87.5**	**33.33**	**77.78**	**87.5**	**82.35**
	cMCI vs ncMCI	**80**	**100**	**33.33**	**77.78**	**100**	**87.5**

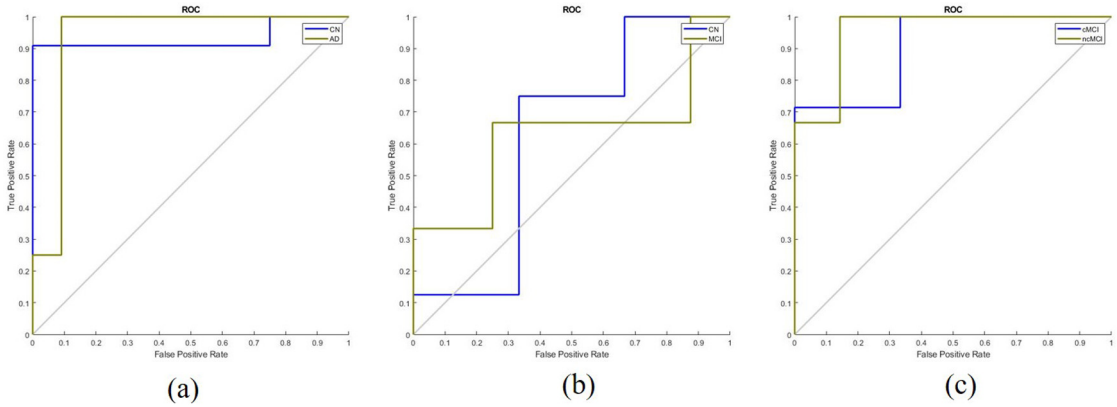

Fig. 7. ROC for the classification of (a) CN vs AD (b) CN vs cMCI (c) cMCI vs ncMCI for the proposed FAF-DRVFL model.

ELM [24] and Kernel ELM (KELM) [47] classifier. The performance matrices are compared in Table 3. ELM is a feed forward network with single hidden layer that calculates the output by matrix inversion. The choice of activation function at the hidden layers of ELM is still unanswered. Therefore, the authors in [24] proposed kernel function at the hidden node to calculate the output of the hidden layer to improve the stability and generalization of the algorithm. In this comparison, the proposed model, which is based on sF activation function, outperforms with the highest classification accuracy for all three classes when compared to the ELM and KELM with different activation functions. Almost similar result were shown for KELM with kernel function as Radial Basis and Wavelet for all the three classes. However KELM with polynomial kernel shows a good results when compared to other KELM kernels implemented. This is due to the non-linear complexity in the model. The lowest accuracy was noticed when ELM is used with sine kernel for CN vs AD and cMCI vs ncMCI and for CN vs cMCI, radial basis shows relatively poor results. All the performance parameter values are low with a huge difference for sine function. This is because the sine function is symmetric in nature, thereby cannot handle non-linear features. Overall, we can clearly state that the proposed model shows outstanding results when compared to ELM and KELM for all the classes.

The upcoming subsections deals with the comparison of the proposed model with popular pretrained DL algorithms available.

4.5. Comparison with the state-of-art deep architectures

Another comparison was made with the popular five pre-trained DNN networks: Alexnet, ResNet, SqueezeNet, VGG19, GoogleNet, with the proposed method. The proposed FAF-DRVFL model outperforms in CN vs AD and cMCI vs ncMCI classification tasks with the highest accuracy, sensitivity, specificity, precision, recall and F-score. However, for CN vs cMCI, although the accuracy and sensitivity were the same compared to the several models, the specificity is high in the proposed model and specificity means the degree to which a diagnostic test for a specific condition is specific. Hence the FAF-DRVFL model shows descent specificity when compared to the remaining state-of-the-art deep models. The classification performance is mentioned in detail in Table 4. Figs. 7 and 8 shows the ROC curve of the FAF-DRVFL algorithm and the other compared state-of-the-art Deep Learning Architectures respectively.

The next subsection represents the complexity analysis for the proposed model.

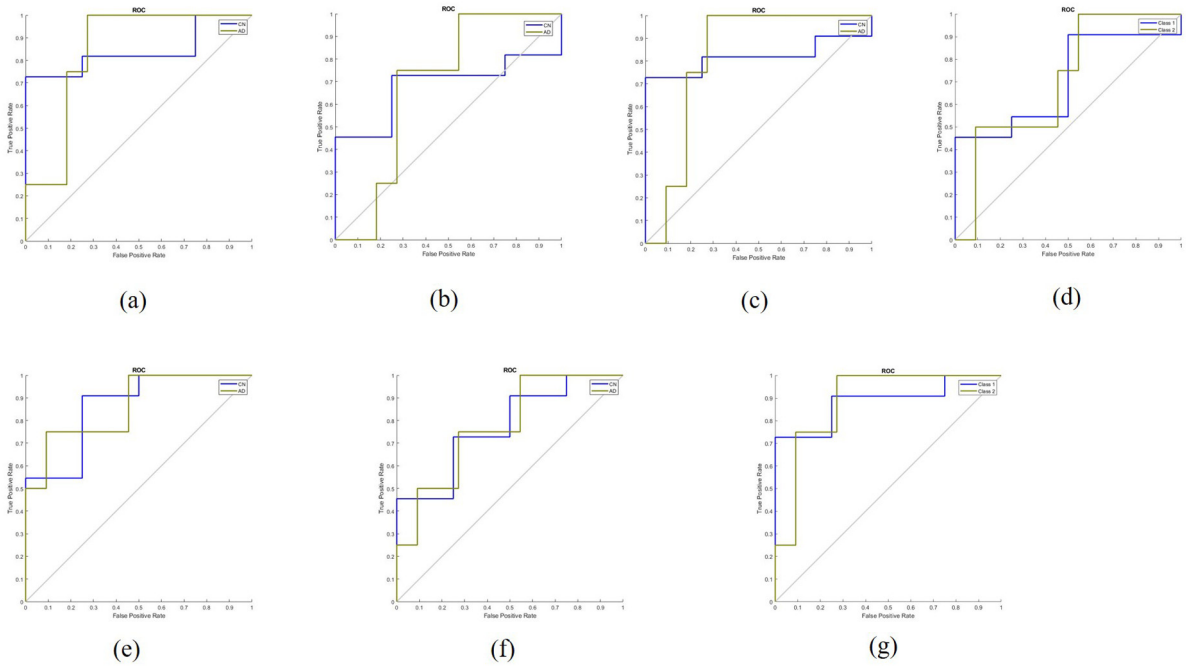

Fig. 8. ROC curve for classification of CN vs AD for (a) AlexNet (b) GoogleNet (c) ResNet-50 (d) KELM (e) SqueezeNet (f) VGG-19 (g) ELM.

Table 4
Comparison with the State-of-the-Art Deep Architectures (in %).

Deep learning model	Class	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Score
Alexnet	CN vs AD	71.67	71.88	28.41	47.92	71.88	57.50
	CN vs cMCI	64.77	41.67	26.56	37.04	41.67	39.22
	cMCI vs ncMCI	63.75	20.83	17.86	33.33	20.83	25.64
ResNet-50	CN vs AD	66.67	54.55	100	100	54.55	70.59
	CN vs cMCI	72.73	87.5	33.33	77.78	87.5	82.5
	cMCI vs ncMCI	70	100	0	70	100	82.35
SqueezeNet	CN vs AD	66.67	54.55	100	100	54.55	70.59
	CN vs cMCI	72.73	100	0	72.73	100	84.21
	cMCI vs ncMCI	70	85.71	33.33	75	85.71	80
VGG19	CN vs AD	73.33	81.82	50	81.82	81.82	81.82
	CN vs cMCI	72.73	87.5	33.33	77.78	87.5	82.35
	cMCI vs ncMCI	70	100	0	70	100	82.35
GoogleNet	CN vs AD	53.33	45.45	75	83.33	45.45	58.52
	CN vs cMCI	72.73	100	0	72.73	100	84.21
	cMCI vs ncMCI	80	100	33.33	77.78	100	57.74
Proposed FAF-DRVFL model	**CN vs AD**	**86.67**	**83.33**	**88.89**	**83.33**	**83.33**	**86.07**
	CN vs cMCI	**72.73**	**87.5**	**33.33**	**77.78**	**87.5**	**82.35**
	cMCI vs ncMCI	**80**	**100**	**33.33**	**77.78**	**100**	**87.5**

4.6. Complexity analysis

In this section, the complexity of the proposed FAF-DRVFL network is explained. TL is used for feature extraction using pretrained DL network. So, the complexity of the network in the proposed FAF-DRVFL model mainly depends on the number of the hidden neurons in the RVFL network. In RVFL network, the matrix inversion of size n hidden neurons needs $O(n^3)$ time. Direct link calculation also takes time. For complexity analysis, the training time of the proposed algorithm is compared with the DL algorithms as well as state-of-the-art single hidden layer feed forward networks in Table 5. For single hidden layer feed forward networks: ELM and KELM, the number of hidden neurons require are more, which will increase the training time. In DL networks, tuning of the parameters is done using backpropagation algorithm which will increase the training time.

The next sections deals with uncertainty analysis in the proposed model.

Table 5
Execution time comparison.

Deep learning model	Time taken (in sec)
ELM	46.70
KELM	44.34
AlexNet	372
ResNet-50	2186
SqueezeNet	118.41
VGG-19	1678.12
GoogleNet	156.08
Proposed FAF-DRVFL model	**33.38**

4.7. Parameter uncertainty analysis

In the proposed network, there is no need for any tuning of the parameters of the network and selection of hyperparameters, therefore the performance depends only on one parameter: number of hidden neurons. Therefore, experiments are done to find

Table 6
Comparison of proposed model with various number of neuron in the model for CN vs AD classification (in %).

s-Membership Fuzzy activation function based RVFL							
Number of neurons	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Score	Time taken (in sec)
500	53.33	100	22.22	46.15	100	63.16	33.81
1000	60	54.55	75	85.71	54.55	66.67	58.51
5000	**86.67**	**83.33**	**88.89**	**83.33**	**83.33**	**86.07**	**33.38**
10 000	80	72.73	100	100	72.73	84.21	45.62
50 000	40	100	0	40	100	57.14	38.77
100 000	73.33	100	0	73.33	100	84.61	35.02

the optimized value of the number of the hidden neurons for maximum performance results using grid search strategy. Table 6 shows the results for different values of number of hidden neurons for CN vs AD classification. From Table 6, an optimum value of hidden neurons is 5000. If we increase the hidden neurons from 5000, the network may face over fitting issue. Therefore, performance is decreasing if we are increasing the number of hidden nodes.

The next sections highlights the remarks for the results obtained.

4.8. Discussion/remarks

The primary goal of this paper is to early diagnose the AD so that care can be taken by proper medications and by changing the lifestyle of the patients to slow down the progression of the disease. Therefore, classification is done using three categories: CN vs AD, CN vs cMCI and cMCI vs ncMCI. Based on the experimental results calculated on the proposed FAF-DRVFL model and comparing with state-of-the-art classification model, we can state that the proposed model outperforms in all the performance parameter. The model achieves the highest accuracy of 86.67% for CN vs AD, 72.73% for CN vs cMCI and 80% for cMCI vs ncMCI classification. The proposed fuzzy function i.e., s-membership fuzzy function, outperforms with other activation functions. Due to its random weight assignment feature, the classification model leads to fast and much accurate classification results. Therefore, the proposed FAF-DRVFL algorithm can be used to alleviate the burden of the clinicians for automatic diagnosis of the disease at the early stage.

5. Conclusion

This paper presents a novel, non-iterative and fuzzy activation function-based deep random vector functional link neural network for early AD diagnosis. The features of MRI images are extracted using pretrained ResNet-50 DNN. Classification of the extracted features is done using a single hidden layer FAF-DRVFL classifier. Weights and biases between the input layer and enhancement nodes of the RVFL network are randomly initialized. The fuzzy activation function is used at the hidden layer to map the features into non-linear space. Then the enhanced and original features are combined and fed to the output layer for classification. The proposed network's main advantage is a non-iterative feed-forward network in which weights and biases are generated randomly. Results show the best performance of the proposed network than the state-of-the-art methods for CN vs AD, cMCI vs CN, and cMCI vs ncMCI classification with faster speed.

In future, the proposed algorithm can be applied to larger datasets and used to diagnose other neural disorders. Also, in present scenario, we are only considering the structural changes occurred in the cerebral region, which can be further extended to consider metabolic changes as well. Also, one can approach for Multimodal dataset i.e. inclusion of features from PET, Diffusion Weighted Imaging, Quantitative susceptibility mapping along with sMRI to improve the diagnosis results.

CRediT authorship contribution statement

Rahul Sharma: Methodology, Data curation, Writing - original draft. **Tripti Goel:** Conceptualization, Algorithm implementation. **M. Tanveer:** Visualization, Investigation, Reviewing and editing. **Shubham Dwivedi:** Reviewing and editing. **R. Murugan:** Reviewing and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by Department of Science and Technology, INDIA under Ramanujan fellowship grant no. SB/S2/RJN-001/2016 and under Interdisciplinary Cyber Physical Systems (ICPS) Scheme grant no. DST/ICPS/CPS-Individual/2018/276.. It is also supported by Council of Scientific & Industrial Research (CSIR), New Delhi, INDIA under Extra Mural Research (EMR) Scheme grant no. 22(0751)/17/EMR-II.

References

[1] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y. Wu, M. Prina, Alzheimer's disease international: World alzheimer report 2015: The global impact of dementia: an analysis of prevalence, incidence, cost and trends. 2015, Alzheimer's Dis. Int.: London (2019).
[2] F. Liu, L. Zhou, C. Shen, J. Yin, Multiple kernel learning in the primal for multimodal Alzheimer's disease classification, IEEE J. Biomed. Health Inf. 18 (3) (2013) 984–990.
[3] M. Liu, D. Zhang, D. Shen, Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment, IEEE Trans. Med. Imaging 35 (6) (2016) 1463–1474.
[4] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Démonet, V. Duret, M. Puel, I. Berry, J.-C. Fort, P. Celsis, et al., Early diagnosis of Alzheimer's disease using cortical thickness: impact of cognitive reserve, Brain 132 (8) (2009) 2036–2047.
[5] B. Richhariya, M. Tanveer, A.H. Rashid, ADNI, Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE), Biomed. Signal Process. Control 59 (2020) 101903.
[6] R.U. Khan, M. Tanveer, R.B. Pachori, Alzheimer's Disease Neuroimaging Initiative (ADNI), A novel method for the classification of Alzheimer's disease from normal controls using magnetic resonance imaging, Expert Syst. 38 (1) (2021) e12566.
[7] B. Richhariya, M. Tanveer, ADNI, Least squares projection twin support vector clustering (LSPTSVC), Inform. Sci. 533 (2020) 1–23.
[8] M. Tanveer, B. Richhariya, R. Khan, A. H. Rashid, P. Khanna, M. Prasad, C. Lin, Machine learning techniques for the diagnosis of Alzheimer's disease: A review, ACM Trans. Multimedia Comput. Commun. Appl. 16 (1s) (2020) 1–35.
[9] S. Karasu, A. Altan, S. Bekiros, W. Ahmad, A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series, Energy 212 (2020) 118750.
[10] A. Altan, S. Karasu, The effect of kernel values in support vector machine to forecasting performance of financial time series, J. Cogn. Syst. 4 (1) (2019) 17–21.

[11] S. Karasu, A. Altan, Recognition model for solar radiation time series based on random forest with feature selection approach, in: 2019 11th International Conference on Electrical and Electronics Engineering, ELECO, IEEE, 2019, pp. 8–11.
[12] B.A. Ardekani, E. Bermudez, A.M. Mubeen, A.H. Bachman, Prediction of incipient Alzheimer's disease dementia in patients with mild cognitive impairment, J. Alzheimer's Dis. 55 (1) (2017) 269–281.
[13] A.K. Lebedeva, E. Westman, T. Borza, M.K. Beyer, K. Engedal, D. Aarsland, G. Selbaek, A.K. Haberg, MRI-based classification models in prediction of mild cognitive impairment and dementia in late-life depression, Front. Aging Neurosci. 9 (2017) 13.
[14] A. Altan, S. Karasu, S. Bekiros, Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques, Chaos Solitons Fractals 126 (2019) 325–336.
[15] A. Altan, S. Karasu, Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique, Chaos Solitons Fractals 140 (2020) 110071.
[16] E. Hosseini-Asl, G. Gimel'farb, A. El-Baz, Alzheimer's disease diagnostics by a deeply supervised adaptable 3D convolutional network, 2016, arXiv preprint arXiv:1607.00556.
[17] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, in: Classification in BioApps, Springer, 2018, pp. 323–350.
[18] Y.-H. Pao, S.M. Phillips, D.J. Sobajic, Neural-net computing and the intelligent control of systems, Internat. J. Control 56 (2) (1992) 263–289.
[19] Y.-H. Pao, G.-H. Park, D.J. Sobajic, Learning and generalization characteristics of the random vector functional-link net, Neurocomputing 6 (2) (1994) 163–180.
[20] P.N. Suganthan, R. Katuwal, On the origins of randomization-based feedforward neural networks, Appl. Soft Comput. (2021) 107239.
[21] X. Li, Y. Yang, N. Hu, Z. Cheng, J. Cheng, Discriminative manifold random vector functional link neural network for rolling bearing fault diagnosis, Knowl.-Based Syst. 211 (2021) 106507.
[22] X. Qiu, P.N. Suganthan, G.A. Amaratunga, Ensemble incremental learning random vector functional link network for short-term electric load forecasting, Knowl.-Based Syst. 145 (2018) 182–196.
[23] Y. Ren, P.N. Suganthan, N. Srikanth, Ensemble methods for wind and solar power forecasting—A state-of-the-art review, Renew. Sustain. Energy Rev. 50 (2015) 82–91.
[24] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.
[25] B. Lei, S. Chen, D. Ni, T. Wang, Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion, Front. Aging Neurosci. 8 (2016) 77.
[26] B. Jie, M. Liu, J. Liu, D. Zhang, D. Shen, Temporally constrained group sparse learning for longitudinal data analysis in Alzheimer's disease, IEEE Trans. Biomed. Eng. 64 (1) (2016) 238–249.
[27] C. Zhang, E. Adeli, T. Zhou, X. Chen, D. Shen, Multi-layer multi-view classification for Alzheimer's disease diagnosis, in: Proceedings of the... AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence, Vol. 2018, NIH Public Access, 2018, p. 4406.
[28] B. Lei, P. Yang, T. Wang, S. Chen, D. Ni, Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis, IEEE Trans. Cybern. 47 (4) (2017) 1102–1113.
[29] A.B. Rabeh, F. Benzarti, H. Amiri, Diagnosis of Alzheimer diseases in early step using SVM (Support Vector Machine), in: 2016 13th International Conference on Computer Graphics, Imaging and Visualization, CGiV, IEEE, 2016, pp. 364–367.
[30] E.E. Bron, M. Smits, W.J. Niessen, S. Klein, Feature selection based on the SVM weight vector for classification of dementia, IEEE J. Biomed. Health Inf. 19 (5) (2015) 1617–1626.
[31] J. Islam, Y. Zhang, Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks, Brain Inf. 5 (2) (2018) 2.
[32] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.
[33] H.-I. Suk, S.-W. Lee, D. Shen, Alzheimer's Disease Neuroimaging Initiative, Deep ensemble learning of sparse regression models for brain disease diagnosis, Med. Image Anal. 37 (2017) 101–113.
[34] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, O.-y. Song, Transfer learning assisted classification and detection of Alzheimer's disease stages using 3D MRI scans, Sensors 19 (11) (2019) 2645.
[35] A.P. Castro, E. Fernandez-Blanco, A. Pazos, C.R. Munteanu, Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques, Comput. Biol. Med. (2020) 103764.
[36] Y. Zhang, J. Wu, Z. Cai, B. Du, S.Y. Philip, An unsupervised parameter learning model for RVFL neural network, Neural Netw. 112 (2019) 85–97.
[37] R. Katuwal, P.N. Suganthan, Stacked autoencoder based deep random vector functional link neural network for classification, Appl. Soft Comput. 85 (2019) 105854.
[38] H. Cecotti, Deep random vector functional link network for handwritten character recognition, in: 2016 International Joint Conference on Neural Networks, IJCNN, IEEE, 2016, pp. 3628–3633.
[39] L. Zhang, P.N. Suganthan, A comprehensive evaluation of random vector functional link networks, Inf. Sci. 367 (2016) 1094–1105.
[40] Q. Shi, R. Katuwal, P.N. Suganthan, M. Tanveer, Random vector functional link neural network based ensemble deep learning, Pattern Recognition (arXiv preprint arXiv:1907.00350) (2021).
[41] P.N. Suganthan, On non-iterative learning algorithms with closed-form solution, Appl. Soft Comput. 70 (2018) 1078–1082.
[42] R. Katuwal, P.N. Suganthan, L. Zhang, An ensemble of decision trees with random vector functional link networks for multi-class classification, Appl. Soft Comput. 70 (2018) 1146–1153.
[43] C.R. Jack Jr, M.A. Bernstein, N.C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P.J. Britson, J. L. Whitwell, C. Ward, The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods, J. Magn. Reson. Imaging: Offic. J. Int. Soc. Magn. Reson. Med. 27 (4) (2008) 685–691.
[44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[45] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, pp. 249–256.
[46] E. Soria-Olivas, J.D. Martín-Guerrero, G. Camps-Valls, A.J. Serrano-López, J. Calpe-Maravilla, L. Gómez-Chova, A low-complexity fuzzy activation function for artificial neural networks, IEEE Trans. Neural Netw. 14 (6) (2003) 1576–1579.
[47] G.-B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, Cogn. Comput. 6 (3) (2014) 376–390.