scientific reports

Check for updates

OPEN

Deep learning-based CAD system for Alzheimer's diagnosis using deep downsized KPLS

Syrine Neffati[1]✉, Kawther Mekki[1] & Mohsen Machhout[2]

Alzheimer's disease (AD) is the most prevalent type of dementia. It is linked with a gradual decline in various brain functions, such as memory. Many research efforts are now directed toward non-invasive procedures for early diagnosis because early detection greatly benefits the patient care and treatment outcome. Additional to an accurate diagnosis and reduction of the rate of misdiagnosis; Computer-Aided Design (CAD) systems are built to give definitive diagnosis. This paper presents a novel CAD system to determine stages of AD. Initially, deep learning techniques are utilized to extract features from the AD brain MRIs. Then, the extracted features are reduced using a proposed feature reduction technique named Deep Downsized Kernel Partial Least Squares (DDKPLS). The proposed approach selects a reduced number of samples from the initial information matrix. The samples chosen give rise to a new data matrix further processed by KPLS to deal with the high dimensionality. The reduced feature space is finally classified using ELM. The implementation is named DDKPLS-ELM. Reference tests have been performed on the Kaggle MRI dataset, which exhibit the efficacy of the DDKPLS-based classifier; it achieves accuracy up to 95.4% and an F1 score of 95.1%.

Keywords Alzheimer's disease, AD, Deep learning, KPLS, DDKPLS, Feature reduction, CAD system

Alzheimer disease (AD), a form of serious brain illness afflicting elderly patients, is one of the four major leading causes of death around the globe, following in line after heart disease, cancers, and brain hemorrhages[1]. Approximately 50 million people suffering from dementia are distributed throughout the world, almost 60% in low and middle-income countries. Dementia affects 5–8% of the population aged 60 and above at any given time. By 2030, projections show that people with dementia will grow to 82 million, increasing to 152 million by 2050[1]. This overall increase is greatly influenced by the growing number of dementia patients in lower to middle-income countries, with nearly ten million new cases reported each year. Neurologically, AD is a chronic neurodegenerative disorder that gradually damages brain tissue and ultimately leads to the death of nerve cells in the brain. It gradually leads to the deterioration of the patient's cognition and memory, commonly referred to as senile dementia. Besides this, Alzheimer's disease results in a decline in the patient's performance of activities of daily living, including handwriting, talking, and reading, in addition to identifying relatives and friends[2,3]. AD is classified into three stages: early, moderate cognitive, and late stage. Patients with moderate cognitive impairment exhibit aggression, while those in the late stage are also at risk for complications such as heart failure and may even develop terminal respiratory failure. It is not easy to diagnose dementia, therefore routine clinical follow-up on AD is essential at all three stages[2], including general practitioner consultation, comprehensive neuropsychiatric evaluation, and MRI or PET scans.

With the development of computer technology, scientists are paying increased attention to its applications in medical and biological fields. This includes the development of Computer-Aided Diagnostic (CAD) systems, which mainly help specialists with the early diagnosis of diseases using information extracted from various imaging modalities. Noninvasive image-based CAD systems, particularly MRI-based analysis for the brain disease, greatly improved diagnostic performance and provided substantial clinical gain[4]. Recently, MRI-based CAD systems show significant potential for distinguishing AD subjects from elderly controls[3]. CAD systems were developed so as to make use of the analytical and computational power of computers to deal with the intricacy of medical images. This technology has been very effective in separating benign from malignant diseases or lesions[5]. The development of a mature CAD system has to be built on two key principles in order to yield optimum results: dimensionality reduction, that is, feature selection or extraction, and classification. These are the key principles that grow the performance of the systems[6].

[1]Department of Computer Engineering, College of Computer Science and Engineering, University of Ha'il, 2440 Ha'il, Saudi Arabia. [2]Laboratory of Electronics and Microelectronics, Faculty of Sciences of Monastir, University of Monastir, 5019 Monastir, Tunisia. ✉email: s.hnaien@uoh.edu.sa

Research in dimensionality reduction, especially when applied to clinical diagnosis, involves the crucial minimization of patient misdiagnosis[7]. Partial Least Squares (PLS)[8] and Principal Component Analysis (PCA)[9] have been among the most widely used linear features extraction methods in image classification. However, conventional PCA and PLS techniques have difficulty producing accurate feature extraction from nonlinear relationships among variables. Besides, a series of nonlinear approaches to PLS and PCA - including neural networks[10], local methods[11], and the kernel trick[12] - have been implemented[12]. Kernel methods have become quite popular due to their ability to solve various problems both linearly and non-linearly[13]. Techniques which have gained very high popularity are Support Vector Machines (SVM)[14], Kernel PCA (KPCA)[15], Kernel Partial Least Square (KPLS)[16], among others; these are extremely efficient approaches with good performance on real-life problems, rendering fantastic results. Generally, kernel methods map the data to an Hilbert space to map the data into a higher-dimensional feature space, dimensionality is indeed reduced during projection to lower-dimensional subspaces to capture the maximum variance of data. This now ensures that there is no need to explicitly consider this higher-dimensional space via the kernel trick as it increases dimensionality to reduce it[3]. Whereas all kernel methods use the kernel trick, they differ slightly among themselves. For instance, in KPCA, the projections are made to enhance the variance of the input data within the feature space; while in KPLS the projections are made keeping in view of both the target data and the transformed input data[15]. Today's KPLS has emerged as a faster and more elegant technique for feature extraction from Magnetic Resonance Images (MRIs). One of its main advantages is that it relies solely on the development of a calibration that does not require nonlinear optimization with a kernel function or the problem of stability, keeping it as straightforward as traditional PLS[17]. Through dimensionality reduction of the data while maintaining critical features, the approach is made computationally efficient and thus improves the classification accuracy. The method proposed not only tackles the problem of handling MRI data in high dimension but also adds strength and accuracy to the Alzheimer's disease classification, thus providing a potential tool for early diagnosis and better patient care.

The organization of this paper is as follows: a review of the related work is covered in section "Methods and approach". In section "Suggested DDKPLS for dimensionality reduction", the methods and approach are outlined. Identification of the deep learning technique used for feature extraction and proposed DDKPLS method is in section "Optimization with Tabu search algorithm". Section "ELM-based classifier" explains the Tabu search algorithms utilized for optimization. Section "Experiments" covers the classifier, ELM. Section "Clinical integration and deployment considerations" outlines the experimental setup and presents the results, while section "Conclusion and future work" concludes and explores future work.

Related work

Alzheimer's is a medical term that indicates the progressive degradation of body functions caused by neurological problems found mostly in older adults. However, the disease progresses towards dementia with time. Commonly, first symptoms include cognitive decline, loss of memory, loss of ability to execute daily activities, and lack of recognition of familiar people. As the disorder progresses, the person may get bedridden and lose control of the body parts and eventually loses the ability to speak[18]. Brain MRIs of the patient are the most accurate means of diagnosing the different stages of AD. This part gives a brief overview of the recent approaches based on traditional features and deep features for detection purposes of AD. To summarize, the current methodologies of AD detection can be classified into two groups: traditional feature-based and deep learning-based approaches.

Approaches based on traditional features

Traditional methods for diagnosing Alzheimer's Disease mainly rely on automated extraction techniques. However, they often overlook texture-based approaches, which examine patterns in MRI scans to detect abnormalities in specific brain regions. Another statistical based texture and outline analysis could thus mean that approach-haulting method is based on histogram data for feature extraction. One of the foremost would mean histogram of oriented gradients[19], which calculates the pixels by computing gradients[20]. Although the HOG captures a significant amount of the local shape information, it is weak in terms of extraction of spatial relations. An improvement to HOG, endowing with pyramid structures, is the pyramid histogram of oriented gradients.

Similarely to histogram of oriented gradients, pyramid histogram of oriented gradients exhibits the extraction of local shape features and development of spatial relationship between pixels of the image. Additionally, there is also some attention by researchers to another statistical method dealing with analyzing second-order characteristics from gray-level image matrices[21]. The Gray Level Co-occurrence Matrix approach is used for feature extraction by using a combination of direction, distance, and frequency. The five major properties derived from this method include correlation, energy, contrast, entropy and homogeneity[22]. A study was performed by Sorensen et al.[23] on texture analysis through different pixel feature extraction from neighboring pixels and local shape regions. This study introduced the Local Binary Pattern (LBP) as a widely adopted method for creating a binary pattern according to its threshold value. A study conducted by Ahonen et al.[24] on gradient binary patterns which extend binary patterns through relative gradients in all directions. Gradients are determined, and transformed images are histogrammed. A texture feature based on Fourier transform was generated by extracting the real and imaginary elements of the image by local phase quantization. The image texture map is subjected to bit plane analysis using the ISO feature according to[25]. Feature selection using EEG signals for early detection of AD has also been suggested in[26].

Approaches based on deep features

Different deep learning techniques have been developed to categorize stages of AD. Most require pre-processing steps such as image registration or denoising, but some methods may perform these steps directly from raw images, without pre-processing and feature selection. A deep learning technique is reported by Ngiam et al.[27],

which uses stacked autoencoders to process several image types for the classification of AD and Cognitive Normal (CN) cases. Subsequently, Li et al.[28] developed a 3D CNN to train samples using both PET and MRI modalities. This method successfully took advantage of the non-linear relationships between these modalities and PET data were predicted from known MRI patterns. Cheng et al.[29] employed two separate 3D CNNs to train on PET and MRI modalities. The extracted features from these techniques were then integrated using a 2D CNN for an improved accuracy. Korolev et al.[30] presented a method that combines residual neural networks (NN) with two distinct 3D CNNs for an automated diagnosis of AD. In addition to two-dimensional and three-dimensional CNNs, autoencoders (AE) have been used to learn encoded features that can successfully differentiate between AD and CN modalities. Zhang et al.[31] made use of sparse autoencoder (SAE) for very high-level feature learning, followed by deep neural network (DNN) at the final stage, attaining a 84.6% classification accuracy. Shifting the focus to unsupervised learning for handling real-time data, Suresha[32] dealt with unsupervised learning in AD detection. MRI images are processed using the histogram of oriented gradients (HOG), and a deep neural network (DNN) is then developed for learning HOG features. The DNN was then used for classifying AD and MCI patients by applying it to images using a corrected Adam optimizer. An efficient accuracy was achieved using the deep neural network (DNN) with a corrected Adam optimizer for classifying patients with Alzheimer disease and mild cognitive impairment as compared to other earlier state-of-the-art models. The authors Shankar et al.[33] found the group grey wolf optimization technique applicable for the detection of AD. After denoising their MRI images, the conventional features such as texture, transformation, and histogram-based properties were extracted. The features were made use of by GGWO classifiers to improve the performance of K-Nearest Neighbor (KNN), Decision Trees (DT) and CNN. Prabha et al.[34] have developed a framework that can classify healthy and affected brains from MRIs. The first stage of this process is essentially optimized segmentation for our white and grey matter in MRIs. Following segmentation, the most relevant features were selected using PCA based on GLCM features and an improved version of the Cuckoo search method. Finally, images of healthy brains were identified using a support vector machine classifier. The researchers Lu et al.[35] implemented the bat algorithm for classifying normal and pathological brain MRIs by optimizing the ELM classifier. Some of the final layers of AlexNet were replaced with ELM layers, and the chaotic bat method was used to optimize the number of layers required in the ELM. Within early diagnosis classification as being used in the ADNI dataset, deep learning methodologies tend to be associated more with end-to-end CNN architecture than classical machine learning[36]. Abdul-Azeem et al.[36], for example, apply adaptive thresholding to handle light variations in images and augment the data for larger dataset size. In this case, the multiple class AD biomarkers are considered as a target for the end-to-end deep 3D CNN, which uses the whole image as input. The accuracy of classifying AD features without prior knowledge is by far end-to-end fast and precise. This approach may be used and improved further by designing RNNs as a diagnostic for AD. For evaluating the images, Cui et al.[37], considered longitudinal features and fed them as inputs to a combined RNN with CNN model. The recent focus in research has centered around transfer learning in order to address the issues of limited amounts of data in the AD domain. Accordingly, Dubey[38] has researched a multi-class problem based on transfer learning using Kaggle's AD dataset.

For predicting Alzheimer's disease, various pre-trained nets such as VGG, ResNet, AlexNet, Inceptionv3, DenseNet, and GoogleNet have been used in assessing performance. They discover a much better result with the pre-trained weights in VGG for predicting AD. As for current research, using all the extracted features in a deep learning model would require a very complicated and time-consuming structure. The work of Saraswathi et al.[39] describes the feature selectionby utilizing genetic algorithm and particle swarm optimization optimizers to extract features from voxel-based morphometry. The performance of the ELM model increases the multiclass classification accuracy to 94.57% using all selected features. Additionally, various pre-trained models were adapted for AD detection, as demonstrated by Shanmugam et al.[40]. Deep feature extraction has proven to be beneficial in the early diagnosis of certain diseases in its several recent studies. Azzali et al.[41] used a vectorial genetic algorithm-based deep feature extractor for handwriting pattern identification of early-stage AD patients. Another related research has been citied for using geometry group in feature extraction employing GConv module and RFSM in image feature extraction to detect AD. This work is indicated by Wang et al.[42]. Other studies have also reported employing deep feature extractors for early detection of various diseases[43–45].

Methods and approach

The objective of this work is to improve the classification accuracy of MRI images for early AD diagnosis using deep learning and convolutional neural networks. This paper proposes developing and evaluating a disease diagnosis framework that employs CNNs for feature extraction. Subsequently, these extracted features are reduced using a Downsized Kernel Partial Least Squares (DKPLS) technique to obtain a significantly downsized set of features. The combination of these feature extraction and feature reduction techniques forms the proposed Deep Downsized Kernel Partial Least Squares (DDKPLS) method. The outline of the design of this work is given in the block diagram presented in Fig. 1, summarizing the overview of the methodology. The sequential flow of the process is also illustrated in Fig. 2.

The suggested research is divided into three phases. The first phase consists of acquisition and preprocessing, during which MRI data is collected and prepared for analysis. Proper preprocessing of data is essential in ensuring the reliability and reproducibility of findings in the sequential steps. In the second phase, the DDKPLS technique is used for dimensionality reduction. A deep learning approch is used first for feature extraction then the suggested DKPLS method is used as a feature reduction method. Here, an approach is devised to obtain a compact version of the MRI data that preserves the maximum information needed for AD detection. The DDKPLS method helps identify important patterns and key features that are needed to classify MRI data into specific groups. Last, but not least, the final phase will apply classification using Extreme Learning Machines (ELMs). ELM is selected for its competence in capturing fast learning speed and strong generalization power.

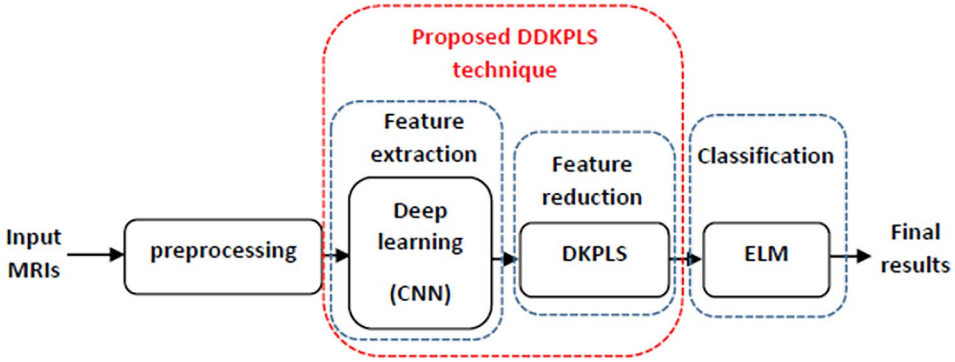

Fig. 1. Block diagram for proposed work based on DDKPLS for dimensionality reduction.

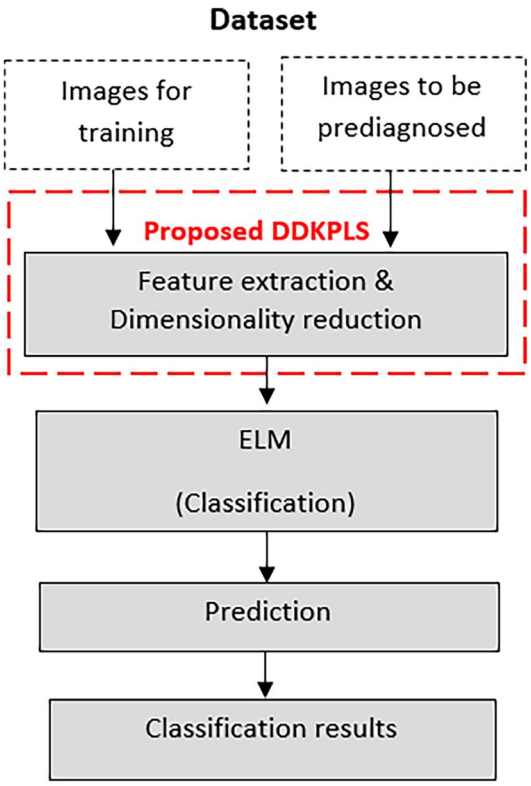

Fig. 2. Flowchart depiction of the proposed analysis framework.

The classifier exploits these extracted features and assigns MRI data into various classes for accurate and effective diagnosis.

The proposed DDKPLS method finds the right balance between accuracy and efficiency. It outperforms other dimensionality reduction techniques, like PCA, KPCA, and KPLS, by offering better classification while keeping computations low. KPCA and KPLS handle non-linear patterns well, but DDKPLS achieves higher accuracy without the extra complexity. This makes it a great choice for early AD diagnosis especially when using large datasets. More details and experiments are covered in the "Experiments" section.

Suggested DDKPLS for dimensionality reduction

The proposed DDKPLS technique for dimensionality reduction can be divided into two phases. First, we begin by constructing and validating a deep learning model (CNN) for deep feature extraction. Second, the validated model will be utilized as input to the DKPLS to form the final DDKPLS approach. In the experiments to evaluate the performance of this technique by analyzing the features extracted by the CNN (ResNet) from the fully connected layer, and then reduced by DKPLS, we use the widely-recognized conventional ML classifier: ELM. Figure 3 depicts the detailed steps of the proposed DDKPLS and Fig. 4 presents workflow diagram of DDKPLS for dimensionality reduction. The different steps will be explained further in the following subsections.

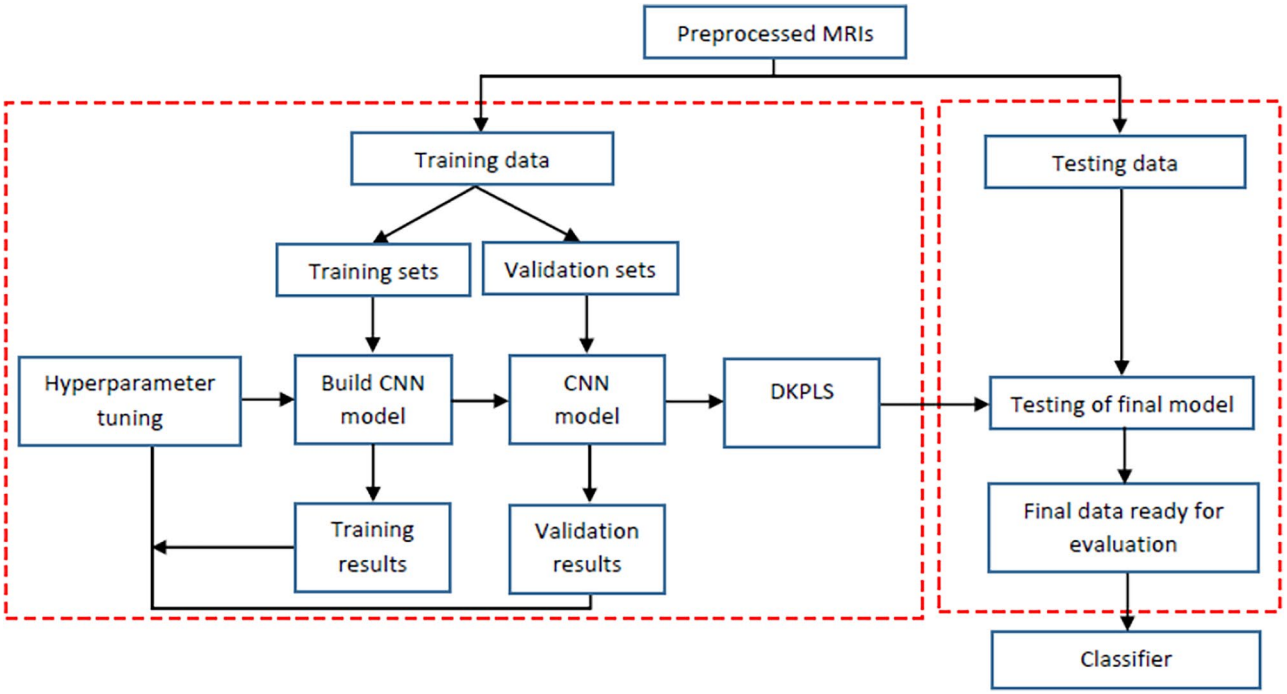

Fig. 3. Block diagram of DDKPLS.

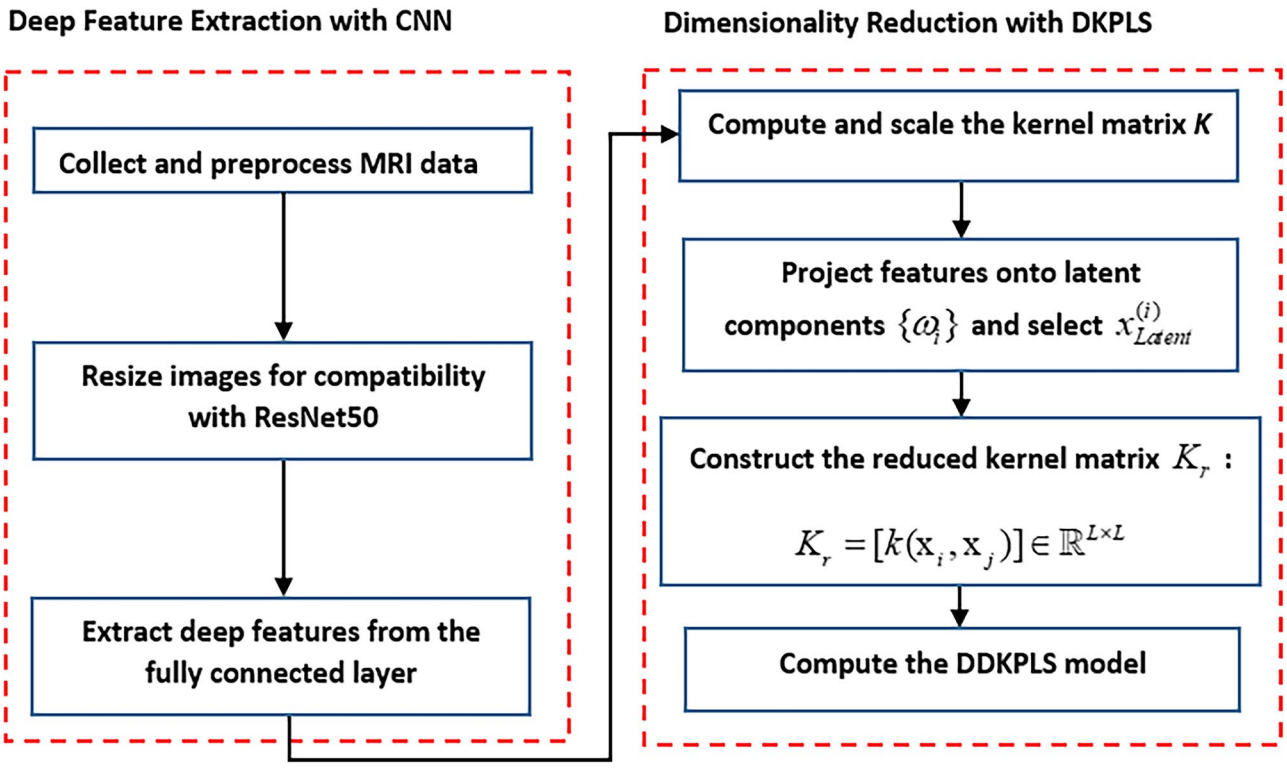

Fig. 4. Block diagram of DDKPLS for dimensionality reduction.

CNN for deep feature extraction

The methodology for building an AD diagnosis approach involves the following steps: first, MRI data collection. In the second stage, image preprocessing involves resizing each MRI image to an appropriate size for the CNN model. Following this, we employ the pre-trained convolutional neural network ResNet50 to extract MRI image

features for use in the subsequent classification stage. In this step, we will use ELM to classify. At the last stage, we will compare results, analyze efficiency, effectiveness for metrics for evaluation, and contrast with recent studies.

The proposed pre-trained CNN model (ResNet-50[46]) comprises five convolution blocks followed by pooling layers and a fully connected (FC) layer. The convolutional and pooling layers work together in extracting features from MRIs. This process uses local connections for feature detection and pooling to merge similar features into one. The FC layer produces the output for each MRI input at the classification stage. In addition to classifying tasks, the FC layers can be replaced by those of other classifiers such as ELM, Random Forests (RF) or Support Vector Machine (SVM).

The whole set of data after collection and image preprocessing is separated into three portions: training, validation and testing. The training set, being a labeled dataset, is used to train the CNN model for tasks such as feature extraction, where the CNN model is capable of performing theorical tasks such as feature extraction-hence generalizing MRI feature vectors for fully connected layers. These feature vectors are used as input for the classifier. The validation set is being used to evaluate the quality of training fit for the model while fine-tuning it. This work utilizes a pre-trained ResNet-50 CNN[46], implemented with TensorFlow[47] and Keras[48], for MRI analysis.

Convolutional layer

Convolutional layers[49] are the most basic unit and fundamental component of deep learning Convolutional Neural Networks (CNNs). These layers mainly perform the feature extraction process, releasing a set of 2D matrices known as feature maps. Each convolutional layer contains a set number of filters that function as feature extractors, identifying patterns by applying convolution operations to the input image. ResNet50 has (7×7), (1×1), and (3×3) filter sizes. During training, a filter learns to catch a low-level feature for images, like colors, edges, blobs, and corners.

Pooling layer

The purpose of this subsampling layer (pooling layer[49]) is to reduce the size of feature maps resulted from the convolutions. The most popular pooling operation is max pooling, which minimizes feature map by picking maximum from small regions in the image. Max-pooling takes (2×2), non-overlapping image areas and extracts the highest value from each. This (2×2) pooling layer decreases the area of the featured map to a quarter. Max pooling helps reduce overfitting by simplifying the representation of image regions. It also decreases computational costs by cutting down the number of parameters. Average pooling, which is another form of pooling, works in a similar way to max pooling but calculates the averages of (2×2) blocks to produce a subsampled image instead of selecting the maximum value.

Batch normalization layer

It adjusts the output of the convolution layer[49]. It normalizes the batch by adjusting the mean to zero and the variance to one. This method accelerates the training process by allowing the use of higher learning rates. It also stops gradients from vanishing during backpropagation. Furthermore, deep learning that use these layers can better handle poor weight initialization.

Dropout layer

This layer is designed to reduce overfitting by randomly deactivating neurons during the training process[49]. The dropout rate parameter specifies how many neurons are dropped, which affects the chances of neuron removal. This technique is applied only during training.

Fully connected layer

It serves as the final layer in the ResNet50 network[49]. It functions as a classifier, connecting the various layers in the network and delivering the final classification outcome. Typically, it is succeeded by a final layer that employs a Softmax function. This layer has been adjusted to optimize ResNet50 for classifying AD. These layers can be substituted with other classifiers. In our project, we will test how well the CNN performs when we use ELM classifier.

DKPLS for feature reduction

KPLS

Principle Kernel methods address the limitations of linearity in PLS approaches by mapping data into a high-dimensional nonlinear feature space[50]. These techniques estimate the nonlinear latent variables with an approximately linear computational cost. Thus, the kernel trick can be applied to linear systems in combination with PLS theory. Defined in the feature space, forms a nonlinear kernelized version of linear PLS[51]. This approach therefore performs a nonlinear transformation of the input variables x_i, $i = [1; 2; \ldots; N]$, mapping them in a feature space *F*, as expressed in Eq. 1:

$$\Phi : x_i \in R^N \rightarrow \Phi(x_i) \in F \tag{1}$$

That being said, the *curse of dimensionality* makes all the nonlinear mappings of the unfolded data samples impossible to compute. The resolution to this problem is the characterization of the Mercer kernel *k(.,.)* as the inner product between two mapped samples[52]:

$$k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle = \Phi(x_i)\Phi(x_j)^T \tag{2}$$

where $\Phi(x_i) \in R^{1\times D}$, [i=1;2;..;N], and D being the dimension of the feature space. The kernel function must comply with Mercer's theorem. As stated in Eq. 2, the Gram matrix $K \in R^{N\times N}$ is evaluated according to the following equation:

$$K = \Phi(X)\Phi(X)^T \tag{3}$$

with $\Phi(X) = [\varphi(x_1)^T, \varphi(x_2)^T, \ldots, \varphi(x_N)^T]$.
There are multiple kernel functions available, but mostly used ones are Radial Basis Function (RBF) kernel and polynomial kernel as illustrated in Table 1[52].

Algorithm Lindgren et al. provided the kernelized PLS techniques that take data matrix optimization into account[53]. Firstly, mean centering is done in a high-dimensional space. Next, the following formula is used to get the Gram matrix K:

$$K \leftarrow (I_n - \frac{1}{n}1_n 1_n^T)K(I_n - \frac{1}{n}1_n 1_n^T) \tag{4}$$

Where 1_n refers to a vector of ones of length N and I_n represents an identity matrix with N dimensions. In the present case, it is to examine a modified version of the PLS-algorithm where, instead of scaling the weight vectors W and C, we have scaled the score vectors T and U to unit vector norms according to[54]. Furthermore, a new score vector T is required for ensuring a deflation phase based on the corresponding rank-one deflation of K and Y matrices. In fact, the following formula is involved in calculating K and Y:

$$K \leftarrow (I_n - tt^T)K(I_n - tt^T) = K - tt^T - Ktt^T + tt^T Ktt^T \tag{5}$$

$$Y \leftarrow Y - tt^T Y \tag{6}$$

where I_n represents an identity matrix of size N.
After computation of the kernel, predictions concerning training and testing instances are calculated. The prediction outcomes for the training instances are expressed as:

$$\hat{Y} = KU(T^T KU)^{-1}T^T Y \tag{7}$$

The prediction outcomes corresponding to the testing instances are written as:

$$\hat{Y}_t = K_t U(T^T KU)^{-1}T^T Y \tag{8}$$

where K_t indicates the kernel matrix of the test instances.
The standard phases of KPLS are summarized in the ten steps outlined in Algorithm 1.

Algorithm 1

1. Compute and center the kernel matrix.
2. Initialize variables: set $i = 1$, $K_1 = K$, and $Y_1 = Y$.
3. Randomly choose any column from Y_i as the initial value for u_i.
4. Compute $t_i = K_i^T u_i$, then normalize t_i by dividing it by its norm, i.e., $t_i = \frac{t_i}{\|t_i\|}$.
5. Compute $c_i = Y_i^T t_i$.
6. Update $u_i = Y_i c_i$, and normalize c_i by dividing it by its norm, i.e., $c_i = \frac{c_i}{\|c_i\|}$.
7. If t_i has converged, proceed to step 8. Otherwise, return to step 3 and repeat the process.
8. Perform deflation on K and Y.
9. Repeat steps 3 through 6 to extract additional latent variables.
10. Construct the cumulative matrices T and U.

Kernel name	Formula and parameters
RBF Kernel	$k(x, y) = \exp\left(-\sigma\|x - y\|^2\right)$
	$k(x, y) = \exp\left(\frac{-\|x - y\|^2}{2\sigma^2}\right)$
	σ: RBF kernel parameter
Polynomial Kernel	$k(x, y) = (x^T y + c)^d$
	$k(x, y) = (x^T y)^d$
	d: Degree of the polynomial kernel
	c: Positive constant

Table 1. Mathematical formulations and parameters of common kernel functions.

Suggested DKPLS

We propose in this study a DKPLS model that chooses a smaller subset of samples from the measurement variables in the initial information matrix. This set of samples can be used to construct a new data matrix.

Principle The proposed method simplifies MRIs by selecting the most relevant features. It extracts a subset from the feature matrix to create a reduced KPLS matrix based on the chosen latent components. Then, the smallest possible KPLS matrix with the most informative features is used to ensure to efficiency of the model. The suggested DKPLS matrix plays a key role in the DDKPLS-ELM algorithm.

The proposed method approaches each LC $\{w_i\}_{i=1\ldots P}$ by a modified input data $\phi(x^{(i)}_{Latent}) \in \phi\{x^j\}_{j=1\ldots M}$ that have the highest projections in w_i direction[50]. The projection $\phi(x^{(i)}_{Latent})$is expressed in Eq. 9 where L represents the number of LCs:

$$\phi(x^{(i)}_{Latent}) = \alpha_i \times k_i(x), i = 1 \ldots L \tag{9}$$

Thereafter, all input data vectors $\phi\{x^j\}_{j=1\ldots M}$ are projected on w_i LCs and only $x^{(i)}_{Latent} \in \phi\{x^j\}_{j=1\ldots M}$ satisfying Eq. 10 are retained.

$$\begin{cases} \phi(x^i_{Latent})_i = \max_{j\in[1,M]} \phi(x^j)_i \\ \phi(x^i_{Latent})_{i\neq j} < \zeta \end{cases} \tag{10}$$

where ζ is a given threshold.

The downsized matrix is obtained as follows once the downsized dataset has been determined:

$$X_r = [x^1_{Latent}; x^2_{Latent}; \ldots; x^L_{Latent}]^T \tag{11}$$

Furthermore, a new reduced kernel matrix K_r associated with the kernel function k is defined as shown in Eq. 15:

$$K_r = \begin{bmatrix} k(x_1, x_1) & \ldots & k(x_1, x_L) \\ \vdots & \ddots & \vdots \\ k(x_L, x_1) & \ldots & k(x_L, x_L) \end{bmatrix} \in R^{L\times L} \tag{12}$$

Algorithm Algorithm 2 presents the main steps of the proposed DKPLS algorithm, and Fig. 5 presents the corresponding organigram.

Algorithm 2

1. Gather the initial consistent blocks of information $\{x_i\}_{i=1..N}$ to be trained, and apply scaling.
2. Compute and scale the kernel matrix K.
3. Project $\{x_i\}_{i=1..N}$ onto the LCs $\{w_i\}$, and select the latent variables $x^{(i)}_{latent}$ that satisfy equation 10.
4. Construct the reduced kernel matrix $K_r \in \mathbb{R}^{L\times L}$ as given by equation 15.
5. Compute the DKPLS model.
6. Make forecasts using the Extreme Learning Machine (ELM).

Optimization with Tabu search algorithm

In this work, we use the Tabu Search (TS) algorithm to optimize the kernel parameter utilized in the DKPLS method, in order to ensure an efficient algorithm with improved generalization and reduced complexity. The primary goal of this research is to develop a fast and effective classifier. To achieve this, the first step involves selecting an appropriate kernel (in this case, the RBF kernel)[55]. The second step focuses on optimizing $_\sigma$, the RBF kernel parameter, as it plays a crucial role in the generalization capability of the proposed DKPLS technique. Actually, σ affects the outcome of the feature space partitioning. It leads to under-fitting if it is too small, and it leads to over-fitting if it is too large. If it is too small, it causes underfitting, while if it is too large, it leads to overfitting. The optimal value of σ is defined as the one that improves the accuracy by minimizing the error rate.

The initial solution of the TS method is set with the aim of optimizing the proposed DKPLS algorithm. The parameter σ is imposed in the limits defined by $[2^{-6}, 2^6]$, taking into account the constraints of parameter σ. The method identifies the best classification accuracy by sequentially evaluating the closest unused neighbor values of σ, continuing this process until all neighbors have been considered. Finally, the optimal value of σ is determined[55].

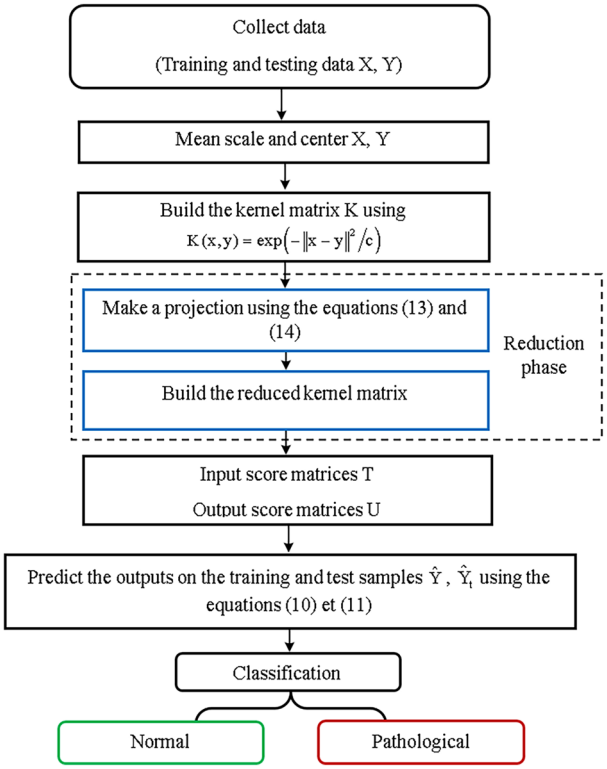

Fig. 5. Flowchart of DKPLS.

$w_i = [w_{i1}, w_{i2}, w_{i3}, \ldots, w_{in}]$	the weight vectors that associate the input
	neurons to the j^{th} hidden neuron
$\beta_i = [\beta_{i1}, \beta_{i2}, \beta_{i3}, \ldots, \beta_{im}]^T$	the weight vectors that associate the output
	neurons to the j^{th} hidden neuron
b_j	the bias of the j^{th} hidden neuron
$w_j x_i$	is the inner product of w_j and x_i
h	a sigmoid function

Table 2. Equation 14 parameters.

ELM-based classifier

ELM is a single hidden layer feed-forward network (SLFN) presented by Huang et al.[56]. The utilised structure in ELM is typically a SLFN structure composed of: (1) an input layer, (2) a hidden layer, and (3) an output layer. In training we consider a sample as follows:

$$\begin{cases} X = \{x_i, y_i\}_{i=1..n} \\ x_i = [x_{i1}, x_{i2}, \ldots, x_{in}^T] \in R^n \\ y_i = [y_{i1}, y_{i2}, \ldots, y_{im}^T] \in R^m \end{cases} \tag{13}$$

where the number of instances is N and the number of output nodes is m. In regression applications the number of output nodes is equal to 1, while in classification applications m is the number of labels, classes or categories. For an ELM with SLFN the output function with N hidden layer nodes is:

$$\begin{cases} f(x_i) = y_i = \sum_{j=1}^{k} \beta_j h(w_j x_i + b_j) \\ i \in [1, N] \end{cases} \tag{14}$$

Equation 14 parameters are summarized in Table 2.

Equation 14 can be reformulated as follows:

$$H = \begin{bmatrix} h(w_1 x_1 + b_1) & \dots & h(w_k x_1 + b_k) \\ \vdots & \ddots & \vdots \\ h(w_1 x_N + b_1) & \dots & h(w_K x_N + b_k) \end{bmatrix}_{N \times k} \tag{15}$$

with $\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_k^T \end{bmatrix}_{k \times m}$; $\beta = H^+ Y$ with $Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_k^T \end{bmatrix}_{N \times m}$ and H^+ is the Moore-Penrose pseudo-inverse[57] of H :

$$H^+ = (H^T H)^{-1}.H^T \tag{16}$$

Thus the output of the classifier is as follows

$$f(x) = h(x)\beta = h(x)(\frac{1}{c} + H^T H)^{-1} H^T Y \tag{17}$$

Introducing kernel to the basic ELM conception make the classifier more reliable[58]. If we suppose that $h(x)$ is unknown as an implicit function[59], we apply the Mercer's settings on ELM, and determine a kernel matrix for ELM as follows:

$$K_{ELM} = HH^T : K_{ELM}(i,j) = h(x_i)h(x_j) = K(x_i, x_j) \tag{18}$$

with $H = [h(x_1)^T, \dots, h(x_N)^T]$ is the output matrix of the hidden layer that maps data x_i from the input space to the hidden layer feature space, and it is unassociated to target value y_i and output nodes m. Kernel matrix is associated particularly to input data xi and training samples N.

The output function of ELM is presented by the following equation

$$f(x) = (\begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N)^2 \end{bmatrix}^T (\frac{1}{C} + K_{ELM})^{-1} Y) \tag{19}$$

Experiments

All preprocessing steps and model development were conducted using Python (Python 3.8;[48]). The code utilized libraries such as Keras[60], Scikit-learn[61], and TensorFlow[47] for the machine learning pipeline. The model was built using the Keras Sequential API, with the TensorFlow backend for training and evaluation.

Performance evaluation metrics

This research work assesses the evaluation of the proposed method against existing models on a number of recognized metrics that shed light on the model classification capability.

- **Accuracy:** measures the overall effectiveness of the classifier based on the ratio of all correctly classified instances to total instances measured. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{20}$$

- **F1 score:** it is a measure that balances precision and recall: a pseudo counterpart that in some ways compensates for the inherent yin and yang of varying degrees of precision and recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{21}$$

where Precision and Recall are calculated as follows:

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{22}$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

- **Sensitivity:** (otherwise recognized as recall or true positive rate) is the measure of the model's ability to correctly find true positive instances. It is calculated as follows:

$$Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{24}$$

– **Specificity:** also referred to as true negative rate, it is used to quantify the ability of the model to predict negative instances correctly. Mathematically, the specificity can be expressed as:

$$Specificity = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{25}$$

where *TP* refers to the true positive as the amount of images positively classified correctly, *TN* as true negatives refers to the amount of images negatively labeled correctly, and *FN* is defined by the misleading label of actually positive images, while *FP* is described by a positive label of an actually negative image. For multiclass classification, a one-against-all approach is used, which is a generalized formula for the binary classification problem.

Experiments on Kaggle dataset

The proposed DDKPLS-ELM framework was validated using the Kaggle MRI dataset[62], which consists of 6,400 MRI images in four classes: non-demented (ND), very mildly demented (VMiD), mildly demented (MiD), and moderately demented (MoD). The dataset consists of images with an original size of 176 × 208 pixels. Sample images representing the four classes are presented in Fig. 6.

The dataset was split into an 80(%) training set (5,121 images) and a 20(%) testing set (1,278 images) while ensuring a balanced distribution of classes. Initially, deep features were extracted using the pretrained model ResNet-50, well suited to capture structural abnormalities associated with Alzheimer's disease, such as hippocampal atrophy and cortical thinning. The extracted features were then subjected to reduction through DDKPLS, enhancing computational efficiency while preserving prominent discriminant information. Finally, the reduced feature set was classified using the ELM classifier, which is characterized by a fast learning speed with high accuracy.

The DDKPLS model is based on the RBF kernel, the expression of which can be found in Table 1. The optimal scaling factor σ corresponding to the kernel has been determined through application of TS algorithm as discussed before. To demonstrate the efficacy of the DDKPLS approach, ELM classifier was paired with DDKPLS for data classification. A 10-fold cross-validation (10-CV) procedure was adopted for our experiments. In our tests, partitioning of 4 separate classes was applied. We have used the one-against-all method while performing these experiments and defined *no dementia* as the reference group. The confusion matrix from the first run of CV is shown in Fig. 7.

The performance metrics presented in the confusion matrix presented in Fig. 7) and from Table 3, alongside the overall accuracy, sensitivity, and specificity values, highlight the effectiveness of the proposed DDKPLS-ELM method in classifying different stages of AD from MRIs. The confusion matrix shows that the model performs exceptionally well, especially when distinguishing "No dementia" (98.7%) and "Moderate AD" (100%). It accurately identifies all "Moderate AD" cases, with no misclassifications, and reliably detects "No dementia" with high precision. In addition, the model shows solid performance in the "Very mild AD" and "Mild AD" classes, with only a small percentage of misclassifications, indicating its potential to effectively differentiate between various stages of AD. These results highlight the model's strong ability to classify AD stages with high accuracy, demonstrating its promise for real-world clinical use.

The classification results indicate that the DDKPLS-ELM system that we propose performs well on all dementia classes. The Non-Demented (ND) class attains a precision of 95.0% and an extraordinary recall of 98.7%, thus giving rise to a strong F1 score of 96.8%. This implies that the model is maximizing the true positives with regard to non-demented individuals and minimizing false negatives, which is crucial in clinical diagnostics to safeguard against the misclassifications of healthy individuals. For Very Mildly Demented (VMiD), the precision is calculated at 91.3% and the recall at 90.5%, and hence the F1-score amounts to 90.9%. While these

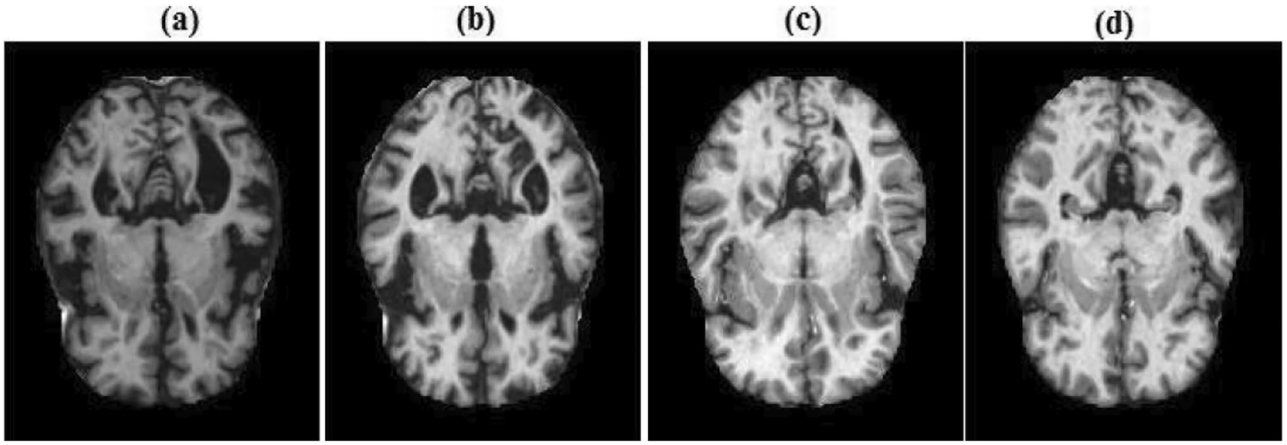

Fig. 6. Sample images of the four AD classes: (**a**) MID (**b**) MOD (**c**) ND (**d**) VMD.

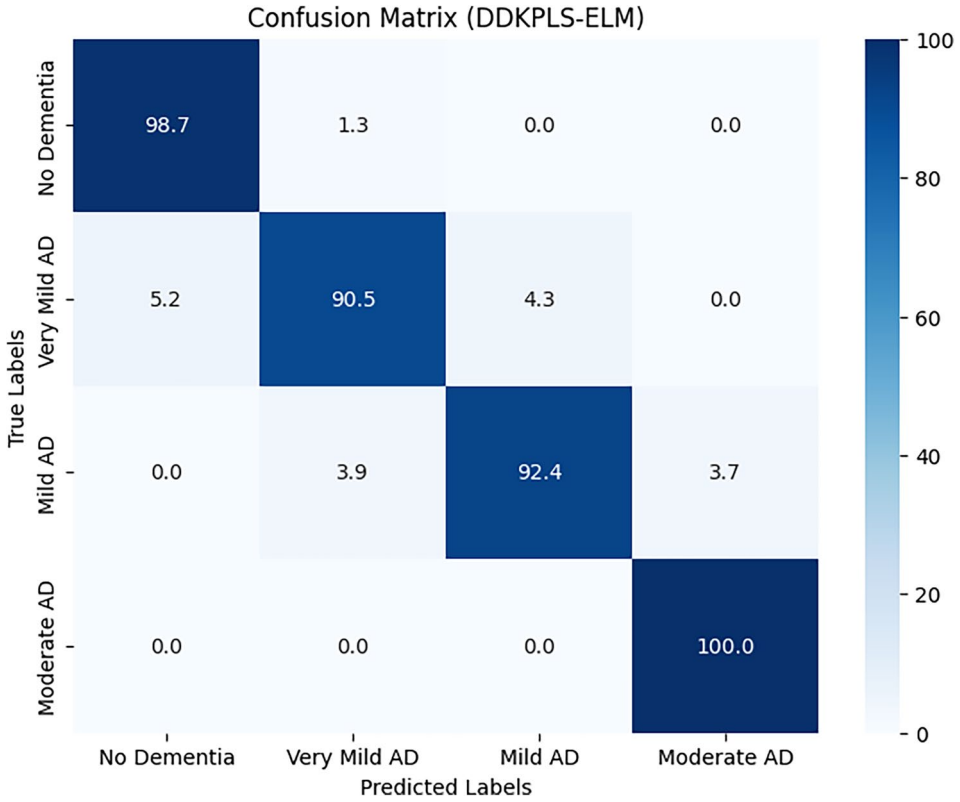

Fig. 7. Confusion matrix of DDKPLS-ELM.

Class	Accuracy (%)	Recall (%)	F1-Score (%)	Samples
ND	95.0	98.7	96.8	640
VMiD	91.3	90.5	90.9	448
MiD	95.5	92.4	93.9	179
MoD	100.0	100.0	100.0	12
Overall	95.4	94.8	95.1	1,278

Table 3. Classification metrics for dementia categories.

scores lie slightly lower than those for ND, they remain commendable and become critically important for early detection, where even slightest signs may escape casual observation. In the Mildly Demented (MiD) category, high precision (95.5%) and recall (92.4%) yield an F1-score of 93.9%: thus, the model shows great potential for distinguishing mild stages of dementia, which is vital for timely intervention and treatment planning. Interestingly, the Moderately Demented (MoD) class achieved immaculate scores (100%) for precision, recall, and F1-score. Despite the small support size (12 samples), it shows that the model can completely distinguish certain moderate dementia cases without making mistakes. The overall performance of the system shows 95.4% precision, 94.8% recall, and 95.1% F1-score, with a total of 1,278 samples. This verifies the effectiveness of the DDKPLS-ELM framework for higher diagnostic accuracy for Alzheimer's disease. The good performance of the model, even at very early dementia stages, indicates its potential for clinical usage by supporting early diagnosis and improving patient management.

The performance of the proposed DDKPLS method was evaluated in this work and compared with other kernel-based dimensionality reduction algorithms for AD classification on the Kaggle MRI dataset and results are presented in Table 4.

From the results shown in Table 4, we can see that DDKPLS performed better than traditional methods, achieving the best sensitivity (94.8%), specificity (97.9%), accuracy (95.4%), and F1-score (93.8%). Compared to PCA with low sensitivity (81.2%) but high specificity (92.0%) and PLS with good sensitivity (87.9%) but lowered specificity (82.6%), DDKPLS brings a balanced performance in identifying AD and non-AD cases. Kernel-based methods such as KPCA and KPLS improved on PCA and PLS, with KPLS achieving an accuracy score of 89.9%, demonstrating the merit of kernel transformations in feature extraction. Still, DDKPLS outperforms all earlier methods by effectively reducing the dimensionality of extracted features while retaining the most relevant information for classification. These results imply that the DDKPLS-based CAD system markedly improves

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1 Score (%)
PCA	81.2	92.0	87.1	78.4
PLS	87.9	82.6	85.3	80.5
KPCA	86.4	89.9	87.2	82.4
KPLS	88.4	91.3	89.9	84.8
DDKPLS	**94.8**	**97.9**	**95.4**	**93.8**

Table 4. Comparison of the proposed algorithm with other kernel-based methods using the Kaggle database.

Primary study	Dataset	Accuracy (%)	F1 score (%)
Khan et al.[63]	ADNI	86.84	–
Bangyal et al.[64]	Kaggle	94.61	–
Lin et al.[65]	ADNI	79.9	–
Marcus et al.[66]	ADNI	75.1	–
Ghazal et al.[67]	Kaggle	91.7	93.7
Sharma et al.[68]	Kaggle	90.4	90.4
DDKPLS	Kaggle	95.4	93.8

Table 5. Comparison of various AD detection approaches.

diagnostic accuracy by limiting misdiagnosis and serves as an excellent candidate for early and trustworthy AD detection in clinical uses.

The comparison of various methods of Alzheimer's disease detection, based on accuracy and F1 score, is summarized in Table 5. It provides a broad set of performances of different methods on different datasets, highlighting the newly proposed DDKPLS-ELM method.

Table 5 compares the performance of varied approaches for detection of AD, particularly with respect to accuracy and F1 score, which are defined as important metrics in diagnosing systems. The studies listed show that AD detection using different datasets and methodologies were employed. The table indicates that the proposed DDKPLS-elm method, which has achieved an accuracy of 95.4% and an F1 score of 93.8%, has outperformed all studies mentioned. This shows that not only does the DDKPLS-based classifier exhibit superiority as far as accuracy is concerned, but also it balances precision and recall, hence being a fairly good approach for AD detection. The combination of the DDKPLS feature reduction followed by the ELM classification indicates an intention to optimize performance from both accuracy and computational efficiency perspectives. The novelty and efficiency of combining deep learning for feature extraction, DDKPLS for dimensionality reduction, and ELM for classification for early AD detection are a major advancement in comparison with existing methods.

The proposed DDKPLS-ELM technique has several strengths that make it adapt for larger and more diverse datasets. A major advantage of the DDKPLS model is its use of the RBF kernel, which captures complex non-linear patterns. This is especially useful for larger datasets with variations in MRI quality, resolution, and patient demographics. With increasing data, the ability of the proposed model to generalize is further supported by the application of 10-fold cross-validation, ensuring that the model is tested on multiple subsets of the data to reduce overfitting. The one-against-all method for multi-class classification, with a clear definition of "no dementia" as the reference group is applied which ensures a consistent and reliable classification even as more diverse data is introduced. In addition, the ELM classifier is ideal to handle larger datasets without sacrificing performance thanks to its fast learning and strong generalization ability. The model can also be further enhance by employing data augmentation techniques, which would help it adapt to different variations, making it more suitable to diverse datasets. This is especially useful when dealing with MRIs from different clinical settings, scanners, or patient groups.

To fully realize the potential of the DDKPLS-ELM framework, we need to explore how it performs in real-world clinical settings. The system can significantly improve early AD detection by giving clinicians a reliable and efficient tool for MRI analysis. Its fast and accurate processing of large imaging datasets can help streamline diagnostics, allowing healthcare professionals to make quicker, more informed decisions. In some places, there aren't enough specialized radiologists. This framework can help by providing support and ensuring patients get timely diagnoses, even with limited resources. By automating MRI analysis, the system helps reduce human error and ensures more consistent, accurate results. Integrating it into clinical practice could improve patient care, leading to earlier interventions and more personalized treatment for individuals affected by AD. The impact of this framework depends on healthcare institutions using it. This will require ongoing teamwork with clinical experts to make it easy to use and ensure it blends seamlessly into existing diagnostic routines.

Clinical integration and deployment considerations

The successful integration of the proposed DDKPLS-ELM framework into real-world clinical processes presents both opportunities and challenges. With its impressive accuracy of 95.4% and an F1 score of 93.8%, the system shows great promise in supporting clinicians with early AD detection. By delivering precise and efficient MRI

image analysis, it could facilitate the decision-making process of physicians and enable better outcomes for patients. By maintaining a strong balance between precision and recall without compromising efficiency, the system proves to be a valuable support, especially in resource-constrained healthcare environments. However, bringing this system into real-world clinical practice comes with important challenges. MRI scans vary between healthcare institutions, which could affect performance. Using techniques like transfer learning or domain adaptation can help ensure the model works well in different settings. Additionally, it's important for the system's decisions to be clear and understandable, especially for clinical use. Using explainable AI can improve transparency and help build trust with healthcare professionals. It's also important to ensure smooth integration with hospital information systems and electronic health records (EHR), so the model supports existing workflows rather than disrupting them. Moreover, it's vital to carefully address regulatory requirements and patient data privacy to ensure compliance with healthcare standards and guidelines. Despite these challenges, the impressive results of the proposed framework demonstrate its strong potential for clinical use. Future efforts will focus on enhancing its adaptability to different imaging environments, improving interpretability, and exploring practical deployment strategies to support its adoption in real-world clinical settings.

Conclusion and future work

In this paper, we propose a novel method for selecting relevant features from MRIs for use in CAD systems. The classification problem has been addressed so that the data are first modeled using the DDKPLS method and then the MRIs are classified using ELM. In the current study, we have dealt with the issue of the huge quantity of information that can be extracted from MRIs and which makes the computing time not acceptable and which causes the detoriation of the classifier. The DDKPLS approach resembles the KPLS one. It reduces the dimensionality of nonlinear process variables by projecting the data into a subspace with lower dimensions. The idea is to ameliorate the performance of the CAD systems by modeling the data using the DDKPLS method. The latter one has been put forward to deal with a nonlinear extension of PLS and provide better performances. First, we applied the TS algorithm to select the optimal kernel parameter, ensuring the construction of the best classifier for MRI classification. Once the parameter is selected, we build the DDKPLS model for feature selection. The designed method has been coupled with ELM. The experimental results of diversified MRIs from the Kaggle dataset have shown that the proposed DDKPL-ELM classifier presents the best accuracy rate and the minimal computational cost compared to other known kernel-based methods. The results demonstrate the effectiveness of the DKPLS-based classification methods over the linear PLS and PCA-based method and over nonlinear KPLS and KPCA-based methods. In future work, we will ameliorate the suggested approach, we will try to develop a classifier that focuses on detecting an early AD stage and we will investigate different medical image modalities for AD.

Data availability

The datasets used in this study are publicly available on Kaggle at https://www.kaggle.com/.

Received: 12 February 2025; Accepted: 19 May 2025
Published online: 27 May 2025

References

1. Zhang, F. et al. Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease. *Neurocomputing* **361**, 185–195 (2019).
2. Alayba, A. M., Senan, E. M. & Alshudukhi, J. S. Enhancing early detection of Alzheimer's disease through hybrid models based on feature fusion of multi-cnn and handcrafted features. *Sci. Rep.* **14**, 31203 (2024).
3. Neffati, S., Ben Abdellafou, K., Aljuhani, A. & Taouali, O. An enhanced cad system based on machine learning algorithm for brain mri classification. *J. Intell. Fuzzy Syst.* **41**, 1845–1854 (2021).
4. Ekmekyapar, T. & Taşcı, B. Exemplar mobilenetv2-based artificial intelligence for robust and accurate diagnosis of multiple sclerosis. *Diagnostics* **13**, 3030 (2023).
5. Hendrix, W. et al. Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest ct scans. *Commun. Med.* **3**, 156 (2023).
6. Wang, Z. et al. Computer-aided diagnosis based on extreme learning machine: A review. *IEEE Access* **8**, 141657–141673 (2020).
7. Yanase, J. & Triantaphyllou, E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst. Appl.* **138**, 112821 (2019).
8. Helmer, M. et al. On the stability of canonical correlation analysis and partial least squares with application to brain-behavior associations. *Commun. Biol.* **7**, 217 (2024).
9. Yi, S., Lai, Z., He, Z., Cheung, Y.-M. & Liu, Y. Joint sparse principal component analysis. *Pattern Recognit.* **61**, 524–536 (2017).
10. Lauriola, I., Gallicchio, C. & Aiolli, F. Enhancing deep neural networks via multiple kernel learning. *Pattern Recognit.* **101**, 107194 (2020).
11. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000).
12. Wang, D., Lu, H. & Yang, M.-H. Kernel collaborative face recognition. *Pattern Recognit.* **48**, 3025–3037 (2015).
13. Pani, A. K. Non-linear process monitoring using kernel principal component analysis: A review of the basic and modified techniques with industrial applications. *Braz. J. Chem. Eng.* **39**, 327–344 (2022).
14. Thangavel, K. D., Seerengasamy, U., Palaniappan, S. & Sekar, R. Prediction of factors for controlling of green house farming with fuzzy based multiclass support vector machine. *Alex. Eng. J.* **62**, 279–289 (2023).
15. Anowar, F., Sadaoui, S. & Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). *Comput. Sci. Rev.* **40**, 100378 (2021).
16. Talukdar, U., Hazarika, S. M. & Gan, J. Q. A kernel partial least square based feature selection method. *Pattern Recognit.* **83**, 91–106 (2018).
17. Nanni, L. et al. Comparison of transfer learning and conventional machine learning applied to structural brain mri for the early diagnosis and prognosis of Alzheimer's disease. *Front. Neurol.* **11**, 576194 (2020).
18. Roe, J. M. et al. Brain change trajectories in healthy adults correlate with Alzheimer's related genetic variation and memory decline across life. *Nat. Commun.* **15**, 10651 (2024).

19. Hussein, I. J. et al. Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (hog). *Expert. Syst.* **39**, e12789 (2022).
20. Chitalia, R. D. & Kontos, D. Role of texture analysis in breast mri as a cancer biomarker: A review. *J. Magn. Reson. Imaging* **49**, 927–938 (2019).
21. Waghere, S. S. & Shinde, J. P. A robust classification of brain tumor disease in mri using twin-attention based dense convolutional auto-encoder. *Biomed. Signal Process. Control* **92**, 106088 (2024).
22. Feng, F. et al. Radiomic features of hippocampal subregions in Alzheimer's disease and amnestic mild cognitive impairment. *Front. Aging Neurosci.* **10**, 290 (2018).
23. Sorensen, L., Shaker, S. B. & De Bruijne, M. Quantitative analysis of pulmonary emphysema using local binary patterns. *IEEE Trans. Med. Imaging* **29**, 559–569 (2010).
24. Ahonen, T., Rahtu, E., Ojansivu, V. & Heikkila, J. Recognition of blurred faces using local phase quantization. In *2008 19th international conference on pattern recognition*, 1–4 (IEEE, 2008).
25. Roomi, S. M. M., Raja, R. & Kalaiyarasi, D. Computing image texture by isopattern. *Int. J. Pattern Recognit. Artif. Intell.* **28**, 1454003 (2014).
26. Dogan, S. et al. Lattice 123 pattern for automated Alzheimer's detection using eeg signal. *Cogn. Neurodyn.* **2024**, 1–17 (2024).
27. Ngiam, J. et al. Multimodal deep learning. *ICML* **11**, 689–696 (2011).
28. Li, Y. et al. Joint embeddings of shapes and images via cnn image purification. *ACM Trans. Graph.* **34**, 1–12 (2015).
29. Cheng, D., Liu, M., Fu, J. & Wang, Y. Classification of mr brain images by combination of multi-cnns for ad diagnosis. In *Ninth international conference on digital image processing (ICDIP 2017)*, vol. 10420, 875–879 (SPIE, 2017).
30. Korolev, S., Safiullin, A., Belyaev, M. & Dodonova, Y. Residual and plain convolutional neural networks for 3d brain mri classification. In *2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017)*, 835–838 (IEEE, 2017).
31. Zhang, L., Lu, Y., Wang, B., Li, F. & Zhang, Z. Sparse auto-encoder with smoothed l 1 regularization. *Neural Process. Lett.* **47**, 829–839 (2018).
32. Suresha, H. S. & Parthasarathy, S. S. Alzheimer disease detection based on deep neural network with rectified adam optimization technique using mri analysis. In *2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, 1–6 (IEEE, 2020).
33. Shankar, K. et al. Alzheimer detection using group grey wolf optimization based features with convolutional classifier. *Comput. Electr. Eng.* **77**, 230–243 (2019).
34. Prabha, S., Sakthidasan Sankaran, K. & Chitradevi, D. Efficient optimization based thresholding technique for analysis of Alzheimer mris. *Int. J. Neurosci.* **133**, 201–214 (2023).
35. Lu, S., Wang, S.-H. & Zhang, Y.-D. Detection of abnormal brain in mri via improved alexnet and elm optimized by chaotic bat algorithm. *Neural Comput. Appl.* **33**, 10799–10811 (2021).
36. AbdulAzeem, Y., Bahgat, W. M. & Badawy, M. A cnn based framework for classification of Alzheimer's disease. *Neural Comput. Appl.* **33**, 10415–10428 (2021).
37. Cui, R. & Liu, M. Alzheimer's disease neuroimaging initiative: Rnn-based longitudinal analysis for diagnosis of Alzheimer's disease. *Comput. Med. Imaging Graph.* **73**, 1–10 (2019).
38. Dubey, S. Alzheimer's dataset (4 class of images). *Kaggle, Dec* **26** (2019).
39. Saraswathi, S., Mahanand, B., Kloczkowski, A., Suresh, S. & Sundararajan, N. Detection of onset of Alzheimer's disease from mri images using a ga-elm-pso classifier. In *2013 fourth international workshop on computational intelligence in medical imaging (CIMI)*, 42–48 (IEEE, 2013).
40. Shanmugam, J. V., Duraisamy, B., Simon, B. C. & Bhaskaran, P. Alzheimer's disease classification using pre-trained deep networks. *Biomed. Signal Process. Control* **71**, 103217 (2022).
41. Azzali, I. et al. Automatic feature extraction with vectorial genetic programming for Alzheimer's disease prediction through handwriting analysis. *Swarm Evol. Comput.* **87**, 101571 (2024).
42. Wang, Y. et al. A multi-scale feature selection module based architecture for the diagnosis of Alzheimer's disease on [18f] fdg pet. *Int. J. Med. Inform.* **191**, 105551 (2024).
43. Dai, Y. et al. Deep learning fusion framework for automated coronary artery disease detection using raw heart sound signals. *Heliyon* **10** (2024).
44. Ma, D. et al. Phe-sich-ct-ids: A benchmark ct image dataset for evaluation semantic segmentation, object detection and radiomic feature extraction of perihematomal edema in spontaneous intracerebral hemorrhage. *Comput. Biol. Med.* **173**, 108342 (2024).
45. Midasala, V. D. et al. Mfeuslnet: Skin cancer detection and classification using integrated ai with multilevel feature extraction-based unsupervised learning. *Eng. Sci. Technol. Int. J.* **51**, 101632 (2024).
46. Vidivelli, S. et al. Optimising deep learning models for ophthalmological disorder classification. *Sci. Rep.* **15**, 3115 (2025).
47. Tensorflow. https://www.tensorflow.org/?hl=ar. Accessed 30 Dec 2024.
48. Python libraries. https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c. Accessed 30 Dec 2024.
49. Wang, S.-H. et al. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Front. Neurosci.* **12**, 818 (2018).
50. Sharma, A. & Jacobs, D. W. Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In *CVPR 2011*, 593–600 (IEEE, 2011).
51. Ma, D. et al. Application of non-linear partial least squares analysis on prediction of biomass of maize plants using hyperspectral images. *Biosys. Eng.* **200**, 40–54 (2020).
52. Rostamzadeh-Renani, R. et al. Prediction of the thermal behavior of multi-walled carbon nanotubes-cuo-ceo2 (20–40-40)/water hybrid nanofluid using different types of regressors and evolutionary algorithms for designing the best artificial neural network modeling. *Alex. Eng. J.* **84**, 184–203 (2023).
53. Lindgren, F., Geladi, P. & Wold, S. The kernel algorithm for pls. *J. Chemom.* **7**, 45–59 (1993).
54. Zhou, D., Li, G. & Qin, S. J. Total projection to latent structures for process monitoring. *AIChE J.* **56**, 168–178 (2010).
55. Tang, J., He, L., Cao, Y. & Bai, H. Critical-edge based Tabu search algorithm for solving large-scale multi-vehicle Chinese postman problem. *Sci. Rep.* **14**, 12437 (2024).
56. Huang, G.-B., Zhu, Q.-Y. & Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **70**, 489–501 (2006).
57. Barata, J. C. A. & Hussein, M. S. The moore-penrose pseudoinverse: A tutorial review of the theory. *Braz. J. Phys.* **42**, 146–165 (2012).
58. Gurpinar, F., Kaya, H., Dibeklioglu, H. & Salah, A. Kernel elm and cnn based facial age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 80–86 (2016).
59. Huang, G.-B., Zhou, H., Ding, X. & Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B (Cybernetics)* **42**, 513–529 (2011).
60. Keras. https://keras.io.
61. Scikit-learn: Machine learning in python. https://scikit-learn.org/stable/ (2011).
62. Kaggle: Your home for data science. https://www.kaggle.com/. Accessed 2025-02-14.
63. Khan, A. & Zubair, S. An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease. *J. King Saud Univ. Comput. Inf. Sci.* **34**, 2688–2706 (2022).

64. Bangyal, W. H. et al. Constructing domain ontology for Alzheimer disease using deep learning based approach. *Electronics* **11**, 1890 (2022).
65. Lin, W. et al. Convolutional neural networks-based mri image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Front. Neurosci.* **12**, 777 (2018).
66. Marcus, D. S. et al. Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**, 1498–1507 (2007).
67. Ghazal, T. M. & Issa, G. Alzheimer disease detection empowered with transfer learning. *Comput. Mater. Continua* **70**, 5005–5019 (2022).
68. Sharma, S., Guleria, K., Tiwari, S. & Kumar, S. A deep learning based convolutional neural network model with vgg16 feature extractor for the detection of Alzheimer disease using mri scans. *Meas. Sensors* **24**, 100506 (2022).

Acknowledgements

This research has been funded by Scientific Research Deanship at University of Ha'il - Saudi Arabia through project number <<RG-24 163>>.

Author contributions

S.N. conceptualized and designed the study, implemented the proposed method, and drafted the manuscript. K.M. contributed to the acquisition, analysis, and interpretation of data and reviewed the manuscript for critical content. M.M. supervised the research, provided technical insights, and revised the manuscript substantively. All authors approved the submitted version, take responsibility for their contributions, and agree to address questions regarding the integrity or accuracy of the work.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.