



International Journal of Biomedical Engineering and Technology

ISSN online: 1752-6426 - ISSN print: 1752-6418

<https://www.inderscience.com/ijbet>

Alzheimer's disease classification using the fusion of improved 3D-VGG-16 and machine learning classifiers

Priyanka Gautam, Manjeet Singh

DOI: [10.1504/IJBET.2025.10068503](https://doi.org/10.1504/IJBET.2025.10068503)

Article History:

Received:	23 April 2024
Last revised:	14 June 2024
Accepted:	26 June 2024
Published online:	06 January 2025

Alzheimer's disease classification using the fusion of improved 3D-VGG-16 and machine learning classifiers

Priyanka Gautam* and Manjeet Singh

ECE Department,

Dr B.R. Ambedkar National Institute of Technology,

Jalandhar, 144008, Punjab, India

Email: priyankag.ec.19@nitj.ac.in

Email: singhm@nitj.ac.in

*Corresponding author

Abstract: Alzheimer's disease is a significant cause of mortality in older adults. It is characterised by the accumulation of amyloid- β (A β) and hyperphosphorylated tau in the brain. Structural magnetic resonance imaging (MRI) can detect these progressive brain changes in the early stages. This work presented an improved 3D-VGG-16 deep neural network model fused with three machine learning classifiers: GPC, SVM, and RF. The improved 3D-VGG-16 incorporated the dropout technique, removed fully connected layers, and replaced them with new ones. To train and test T1-weighted 3D MRI images were utilised, a total of 445 images where 239 (AD-69, CN-170) from ADNI and 206 (AD-136, HC-70) from MIRIAD. The models achieved accuracies of 98.75%, 97.5%, and 93.75% (ADNI), and 96.25%, 93.75%, and 91.25% (MIRIAD) for 3D-VGG-16+GPC, 3D-VGG-16+SVM, and 3D-VGG-16+RF. The proposed model with dropout, 3D data adaption, and simultaneous data evaluation demonstrated accurate AD classification, especially with the GPC classifier.

Keywords: deep learning; Alzheimer's disease; magnetic resonance imaging; MRI; machine learning; ADNI; MIRIAD.

Reference to this paper should be made as follows: Gautam, P. and Singh, M. (2025) 'Alzheimer's disease classification using the fusion of improved 3D-VGG-16 and machine learning classifiers', *Int. J. Biomedical Engineering and Technology*, Vol. 47, No. 1, pp.1–27.

Biographical notes: Priyanka Gautam is a PhD Research Scholar at Dr. B.R. Ambedkar National Institute of Technology Jalandhar. She obtained her MTech degree from the Dr. B.R. Ambedkar National Institute of Technology Jalandhar, Punjab, India in 2018. She obtained her BTech degree from U.I.E.T C.S.J.M University Kanpur, Uttar Pradesh, India in 2015. Her research interests include image processing, biomedical signal processing and machine learning.

Manjeet Singh is an Assistant Professor at Dr. B.R. Ambedkar National Institute of Technology Jalandhar. Previously, he was an Assistant Professor at Jawaharlal Nehru Government Engineering College Sundernagar, Himachal Pradesh. He obtained his MTech and PhD degree from the National Institute of Technology Hamirpur and BE degree from GGITS Jabalpur. His professional experience and research areas are wireless sensor networks, wireless communication systems, and machine learning applications in wireless networks.

1 Introduction

The human brain is considered one of the body's most vital and intricate organs. The brain serves numerous crucial functions, including generating ideas, solving problems, thinking, making decisions, imagining, and retaining memories. The process of retaining and recalling information or experiences is commonly called memory. Our physical memory system retains a comprehensive archive of our lives and shapes our personality and identity. Memory loss resulting from dementia and the incapacity to identify our surroundings can be an alarming and distressing experience. Alzheimer's disease (AD) is the most widespread type of dementia (AlSaeed and Omar, 2022). AD is a neurodegenerative disorder that tends to occur more frequently in older individuals. This disease slowly and progressively destroys areas of the brain responsible for memory, learning, thinking, and behaviour, ultimately leading to cognitive decline. Unfortunately, there is currently no known cure for AD, and it is irreversible (Weiner et al., 2015). Dementia affects about 50 million individuals globally, and the expense of offering healthcare and social services to them is comparable to that of the 18th largest economy in the world (Christina, 2018). It is anticipated that the number of new cases of AD and other types of dementia will triple by the year 2050, resulting in a total of 152 million individuals affected worldwide. This projection suggests that a new case of dementia will develop every 3 seconds on average.

Distinguishing AD from the normal ageing process can be difficult due to the similarity in symptoms between the two (Huang et al., 2020; Castellazzi et al., 2020). The specific factors accountable for the development of AD are still unknown. Although the precise factors responsible for AD are still unidentified, there are numerous local and global brain parameters that impact cognitive abilities, and biomarkers like Cerebrospinal fluid (CSF) can serve as indicators of an individual's condition (Zimmermann et al., 2018). One can identify these characteristics by utilising advanced neuroimaging techniques such as magnetic resonance imaging (MRI), functional MRI (fMRI), and positron emission tomography (PET). Various neuropsychological evaluation tests, such as the mini-mental state examinations (MMSE), Alzheimer's disease assessment scale cognitive (ADAS-Cog), and clinical dementia rating (CDR), are utilised for identifying individuals with AD (Kurlowicz and Wallace, 1999; Salmon and Bondi, 2009). These techniques allow tracking small variations in specific brain structure regions. However, it can be challenging for doctors to predict AD patients since generalising the biomarker variations is difficult (Zimmermann et al., 2018). When facing such situations, utilising a machine learning (ML) model trained with multiple multi-modal features can prove to be helpful in the classification of AD (Zhang and Shen, 2012; Zhang and Shen, 2011).

Numerous ML approaches were employed for AD classification, and the results of these models exhibited a considerable degree of efficacy. Conventional learning-based methods typically involve three phases: identifying brain region of interests (ROIs), selecting features from them, and building classification models. The key challenge in traditional learning techniques is feature engineering, where manual feature selection and extraction impact model performance. Many researchers used structural MRI feature extraction and selection for AD classification through various methods, including multiview learning using grey matter (GM) (Liu et al., 2016), ROIs with statistical test (t-test) (Krashenyi et al., 2016), fuzzy-possibilistic tissue segmentation (Lazli et al., 2018), and speeded up robust features (SURF) with grey level co-occurrence matrix (GLCM) (Thulasi and Varghese, 2018). Deep learning (DL) methodology has brought a

revolutionary change in recent decades, significantly distinguishing it from traditional ML techniques (Ji et al., 2019). DL automated the feature extraction process by learning directly from images through neural networks, eliminating the need for human experts to extract features manually. Various DL techniques were used for automatic feature extraction and selection, including stacked auto encoder (SAE) (Liu et al., 2015), 3D-CNN [volumetric CNN (VoxCNN) and residual neural network (ResNet)] (Korolev et al., 2017), convolutional neural network (CNN) (Gunawardena et al., 2017), CNN (AlexNet) (Lin et al., 2018), 3D-ConvNet (Backstrom et al., 2018), ResNet50+Softmax (AlSaeed and Omar, 2022), FreeSurfer (Rallabandi et al., 2020), and convolutional neural network (ConvNet) (Ji et al., 2019). In recent times, CNN-based techniques have achieved exceptional accuracy and precision in classifying images (Ji et al., 2019). Recent advancements in AD classification involved combining CNNs with transformers or using different versions of transformers. Several papers in this field included the cascaded multi-modal mixing transformer (Liu et al., 2023), CNN and Swin-transformer (Xin et al., 2023), optimised vision transformer (OViTAD) (Sarraf et al., 2023), joint transformer architecture (Alp et al., 2024), and longitudinal transformer (Chen et al., 2024). Classifying AD with traditional ML is challenging due to unstructured data like images. These models struggle with insufficient data, class imbalance, and labour-intensive manual feature extraction, often resulting in overfitting and poor generalisability. While capable of automatic feature extraction, DL models require large datasets and significant computational resources. They are also prone to overfitting with limited data and are less interpretable due to their black-box nature. Both ML and DL models face difficulties in integrating into clinical workflows. Although DL architectures perform well, selecting the most suitable architecture is a persistent challenge due to problem-specific requirements.

Considering this challenge, the hypothesis of this study is to present a fusion architecture integrating DL and ML with optimised dropout regularisation to accurately classify AD patients. The study will utilise renowned publicly available brain MRI image datasets, specifically the Alzheimer's disease neuroimaging initiative (ADNI) and minimal interval resonance imaging in Alzheimer's disease (MIRIAD). To fine-tune the parameters of the classifiers, a grid search approach will be implemented. A detailed investigation will be conducted to evaluate the performance of various classifiers in conjunction with the DL model. Data from both repositories will be employed concurrently in the proposed architecture. In this regard, a fusion architecture of an improved 3D-VGG-16 neural network and three ML classifiers is proposed. Compared to the original Visual Geometry Group-16 (VGG-16) architecture, the improved version is designed for Three-Dimensional (3D) data and uses a dropout regularisation technique. In this version, fully connected layers at the end of the model are removed and replaced with new ones. This improved 3D-VGG-16 architecture is used to extract essential features from input images. These extracted features are then fed into the ML classifiers. This work uses three ML classifiers: Gaussian process classifier (GPC), support vector machine (SVM), and random forest (RF). The performance of each classifier is evaluated and compared to determine the optimal pairing with the improved 3D-VGG-16 network. The key contributions of this work are given below:

- The study proposes a fusion architecture that combines an improved 3D-VGG-16 network with three ML classifiers. This work uses GPC, SVM, and RF classifiers. Using Google Colab cloud services reduces the need for graphics processing unit (GPU) computing hardware.
- This study uses an improved 3D-VGG-16 to extract essential input features for subsequent classification. The dropout regularisation used in 3D-VGG-16 optimises its performance. To further enhance the generalisability of the model, it is trained and tested with two datasets: ADNI and MIRIAD.
- This work proposes three approaches: 3D-VGG-16+GPC, 3D-VGG-16+SVM, and 3D-VGG-16+RF. Results show that 3D-VGG-16+GPC outperforms the other two for both datasets.
- Parameter tuning is done through grid search to optimise the performance of classifiers further.

2 Materials

This research utilised two renowned publicly available brain MRI datasets known as ADNI (2003) and MIRIAD (2012). The details of both the datasets related to the data distribution are listed in Table 1. Both datasets are comprised of two categories, number of subjects, gender distribution of these subjects, age range of the selected subjects, and the number of MRI images corresponding to each category in these two different repositories. Data consist of two binary categories known as AD and cognitively normal (CN) or healthy control (HC). Here, CN and HC are analogous to each other in respect of different repositories.

Table 1 Demographic information of dataset from ADNI and MIRIAD

<i>Dataset</i>	<i>Category</i>	<i>Subjects</i>	<i>Gender</i>	<i>Age</i>	<i>Number of MRI images</i>
ADNI	AD	34	13M, 21F	57–90	69
	CN	99	53M, 46F	60–91	170
MIRIAD	AD	46	19M, 27F	62–76	136
	HC	23	12M, 11F	62–76	70

In the ADNI dataset, approximately 26% of subjects belong to the AD category and 74% to the CN category. In the MIRIAD dataset, about 67% of subjects belong to the AD category and 33% to the HC category. In ADNI, there are 48% more CN subjects than AD subjects, while in MIRIAD, 34% more AD subjects than HC subjects. In the AD category, female subjects outnumber male subjects by 24% in ADNI and 18% in MIRIAD. In the CN/HC category, male subjects outnumber female subjects by 6% in ADNI and 4% in MIRIAD. The ages of all subjects from both datasets range from 57 to 91. Overall, 445 MRI images were collected, with 239 from ADNI and 206 from MIRIAD. A further, more detailed description of this dataset is presented in subsequent subsections.

2.1 ADNI

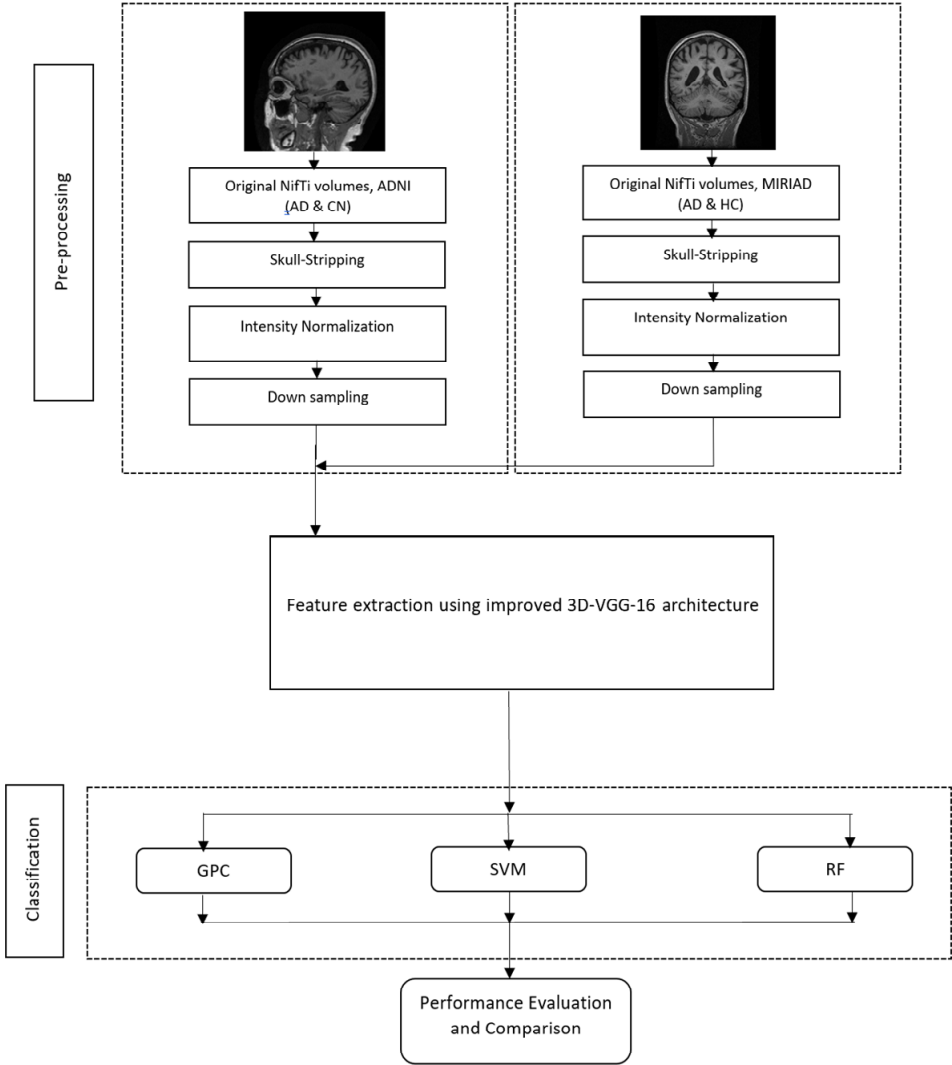
This study utilises structural MRI data sourced from the publicly available ADNI database (ADNI, 2003). ADNI researchers had collected and integrated multi-modal data covering various stages of AD. Their primary aim was to explore the potential fusion of MRI, fMRI, PET, genetic data, and clinical assessments to investigate the human brain's structural and functional aspects. This was ultimately aimed at potentially curing AD. The ADNI dataset was acquired to illustrate the effectiveness of the proposed architecture. This dataset includes T1-weighted 3D MRI images with dimensions of $192 \times 192 \times 160$, with the first two dimensions representing a two-dimensional (2D) image's dimensions and the third dimension representing the total slices in an MRI scan. It encompasses a total of 133 subjects, including 69 AD scans and 170 CN scans.

2.2 MIRIAD

The publicly available MIRIAD dataset comprises 206 MRI brain scans, including 46 AD and 23 HC subjects. This dataset consists of 3D T1-weighted MRI images sized at $256 \times 256 \times 124$. The images were acquired using a 3D T1-weighted inversion recovery prepared fast spoiled gradient Recalled (IR-FSPGR) sequence with a 24 cm field of view and 1.5 mm coronal partitions. The scans employed a repetition time (TR) of 15 ms, an echo time (TE) of 5.4 ms, a flip angle of 15 degrees, and an inversion time (TI) of 650 ms (MIRIAD, 2012). Notably, the datasets did not specify the AD stage in the AD patient images. In these experiments, multiple images from a single patient were treated as independent, similar to images from different patients.

3 Proposed methodology

This section presented the methodology for developing an automatic AD classification system using a fusion architecture. Figure 1 illustrates the proposed system framework. The process started with the pre-processing of 3D T1-weighted brain MRI images from two datasets, ADNI and MIRIAD, which included AD and CN/HC images. Pre-processing involved skull-stripping, intensity normalisation, and down-sampling. Next, features were extracted from these pre-processed images using an improved 3D-VGG-16 neural network architecture. These features were then classified using three different algorithms: GPC, SVM, and RF. Finally, the performance of these classifiers was evaluated and compared to determine the most effective fusion architecture for AD classification.

Figure 1 Framework of the proposed automatic system

3.1 MRI data pre-processing

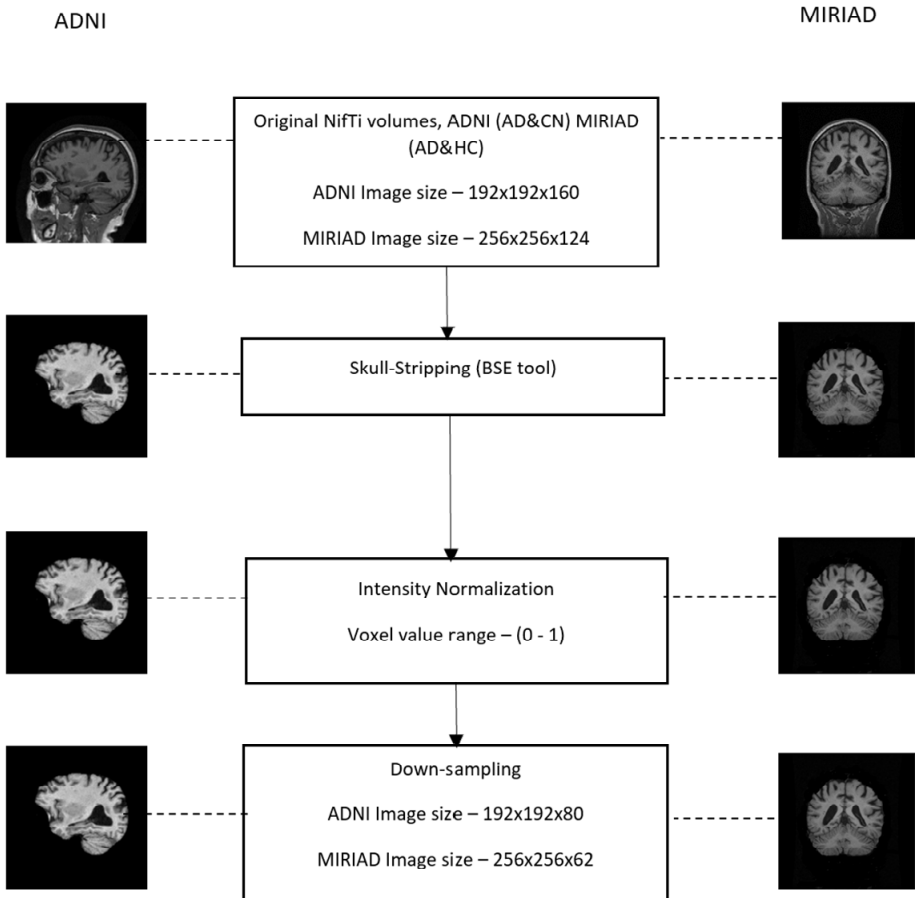
The 3D T1-weighted brain MRI images underwent three preprocessing stages: skull-stripping, intensity normalisation, and down-sampling. *Skull-stripping* utilised the Brain Surface Extractor (BSE) tool developed by Shattuck et al. (2001), which was integrated into BrainSuite (Shattuck and Leahy, 2002). This tool effectively distinguished brain and non-brain regions through diffusion filtering, Marr-Hildreth edge detection, and morphological operations, completing the extraction rapidly in approximately $3.5 (\pm 0.4)$ seconds (Rehman et al., 2020). Its efficiency has led to its widespread adoption for MRI skull-stripping in various studies (Hwang et al., 2019; Roy and Maji, 2015; Laha et al., 2018). Next, *intensity normalisation* reduced pixel value variations across different scans.

This normalisation provided various advantages, such as preventing larger numbers from eliminating smaller ones, mitigating the negative effects of outliers, and facilitating faster neural network convergence. To achieve this, the following formula was utilised:

$$V_{x,y,z} = \frac{OV_{x,y,z} - \min(OV)}{\max(OV) - \min(OV)} \quad (1)$$

where $V_{x,y,z}$ denoted the voxel value at coordinates (x, y, z) in the resulting image, and $OV_{x,y,z}$ represented the original image with its complete set of pixel values (Fan et al., 2021). *Down-sampling* was then performed to reduce the MRI scan size. This entailed removing slices lacking sufficient information after skull-stripping. The initial and ending slices generally contained very little information, with the most essential information found in the middle slices. Therefore, 40 slices from the beginning and end of the ADNI dataset were removed. Similarly, 31 slices from the beginning and end of the MIRIAD dataset were removed. This process reduced the dimensions of MRI scans to $192 \times 192 \times 80$ and $256 \times 256 \times 62$ for ADNI and MIRIAD, respectively. See Figure 2 for a visual overview of the MRI scan pre-processing.

Figure 2 Pre-processing framework of MRI scans



3.2 Improved 3D-VGG-16 architecture

Pre-processed MRI images were then ready for crucial feature extraction using the improved 3D-VGG-16 architecture. This section explained the proposed method's architecture and how it differed from the original VGG-16 architecture.

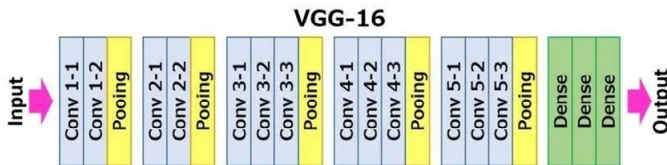
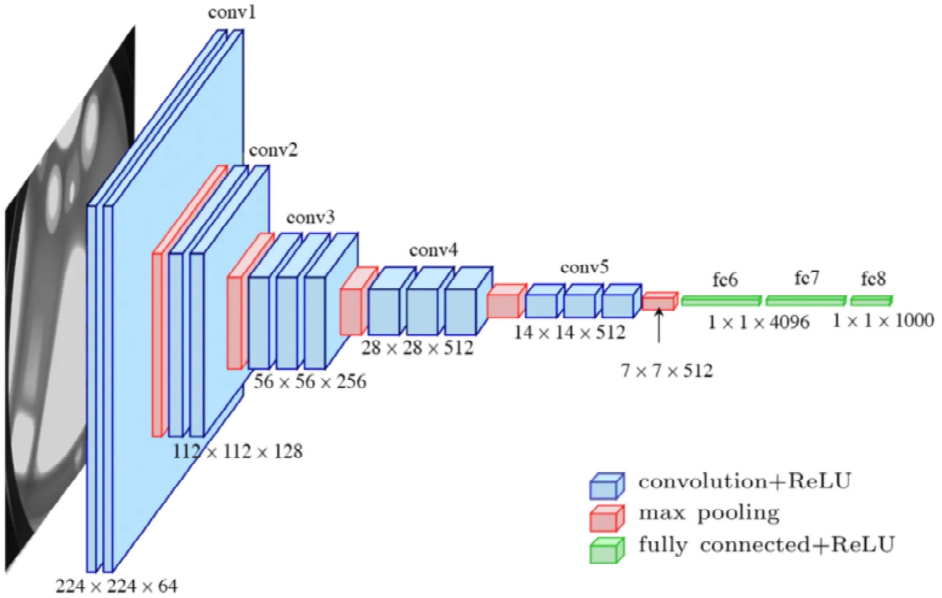
Original VGG-16

VGG-16 is a deep CNN for image classification, comprising 13 convolutional layers, 5 max-pooling layers, and 3 FCs. It learned intricate features via convolutions, pooling, and Fully Connected layers (FCs), yielding a class probability distribution using a softmax activation function. This was expressed as:

$$y = \text{softmax}\left(f_3\left(f_2\left(f_1(x)\right)\right)\right) \quad (2)$$

where x denoted the input image, f_1 , f_2 , and f_3 represented the convolutional, max pooling, and FC layers, respectively, and y signified the resulting probability distribution across classes (Simonyan and Zisserman, 2015). Figure 3 shows the original VGG-16 architecture (Rohini, 2021; Khuyen, 2021).

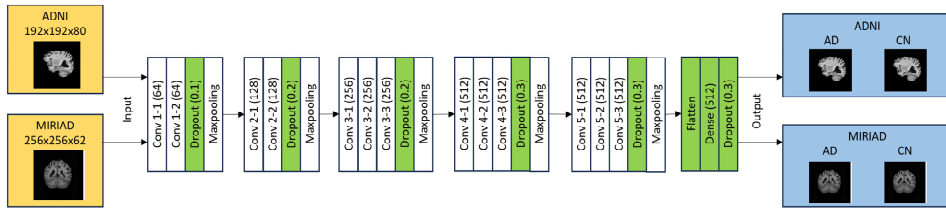
Figure 3 Original VGG-16 architecture (see online version for colours)



Improved 3D-VGG-16

Several modifications were applied to improve the VGG-16 architecture for our task. First, this architecture was designed for 3D datasets, so it is named 3D-VGG-16. To design the 3D-VGG-16, 3-dimensional convolutional filters and max-pooling were used to ensure compatibility with 3D datasets. The number of filters in each convolutional layer is indicated in brackets in Figure 4. For example, 'Conv 1-1 (64)' means the first block's first layer used 64 filters, where the first number represents the block number, the second number represents the layer number within that block, and the number in brackets represents the filters used in that layer. Second, this architecture used a dropout regularisation technique to optimise its performance. Dropout regularisation is a technique used in training neural networks to prevent overfitting. During training, dropout worked by randomly 'dropping out' or ignoring a certain percentage of neurons in the network for each iteration. This meant that each training iteration updated the model using a slightly different subset of neurons, forcing the network to become less reliant on specific neurons and to learn more robust features. By preventing the network from depending too heavily on particular neurons, dropout helped the model generalise better to new data, ultimately improving its performance. Third, the original VGG-16 included three FC layers with 40,96, 4,096, and 1,000 units after the 13th convolutional layer. However, these FC layers were removed, and new FCs were added. The last FC layer originally had 1,000 units designed to classify 1,000 classes. Since this task involved binary classification, the FC layers were modified accordingly. The improved version of the VGG-16 architecture is shown in Figure 4. This architecture was used for feature extraction; these features were further applied to ML classifiers – GPC, SVM, and RF. These classifiers are discussed in detail further.

Figure 4 Proposed improved 3D-VGG-16 architecture (see online version for colours)



The proposed network architecture includes five convolutional blocks, each block consisting of two to three convolutional layers (Conv) with ReLU activation functions, followed by max-pooling layers for spatial downsampling and dropout layers for regularisation to prevent overfitting. Dropout rates varied from 0.1 to 0.3. Different dropout rates were used for convolutional blocks: 0.1 for conv1, 0.2 for conv2 and conv3 blocks, and 0.3 for conv4 and conv5 blocks. The output layer used a 0.3 dropout rate. The convolutional filters increased in number progressively across the blocks, with 64, 128, 256, and 512 filters, respectively. After the final convolutional block, the feature maps were flattened and passed through a fully connected dense layer with 512 units, followed by a dropout layer. The network concluded with a softmax layer for binary classification,

outputting probabilities for AD and CN categories for both the ADNI and MIRIAD datasets. In this architecture, both datasets were applied simultaneously at the input, and the classification output was displayed simultaneously at the output. So, this architecture has one more benefit: it acts as a multiple input and output architecture.

3.3 *ML classifiers*

Various ML classifiers, such as SVM, RF, AdaBoost, decision tree, K-nearest neighbour (KNN), Gaussian Naive Bayes, quadratic discriminant analysis (QDA), and GPC, were available. The extracted feature set from the 3D-VGG-16 model was applied to all these classifiers to find the best three classifiers. Out of all these classifiers, GPC, SVM, and RF performed much better than other classifiers. So, these three classifiers were selected for this work.

GPC

A Gaussian process involves random variables, with an infinite number having a joint Gaussian distribution (Rasmussen and Williams, 2005). Unlike Gaussian distribution for vectors, it deals with functions defined by a mean function $\mu(x) = E[Y(x)]$ and a covariance function $k(x, x') = E[(Y(x) - \mu(x))(Y(x') - \mu(x'))]$ (Williams and Barber, 1998). Like SVM, various kernel functions can be the covariance function for Gaussian processes. A commonly used form is

$$k(x, x') = \theta_o \exp \left\{ -\frac{1}{2} \sum_{l=1}^d w_l (x_l - x'_l)^2 \right\} + \theta_1 \quad (3)$$

Here, x_l represents the l^{th} element of x , and $\theta = (\theta^o, \theta^o, w_1, \dots, w_d)$ represents the positive hyperparameters of the covariance function. The GP model's output spans the real axis, but a transfer function can be applied to map it to the (0, 1) interval, representing the probability of class 1. For binary classification, the logistic function $y = \sigma(a)$ is commonly employed as the transfer function, where $\sigma(a) = 1/(1 + e^{-a})$. Unlike regression, which uses the Gaussian distribution, classification employs the Bernoulli distribution to characterise the probability distribution over the target value $t \in \{0, 1\}$, where $t \sim \text{Bern}(\sigma(a))$. In two-class problems, calculating the predictive distribution is essential for making predictions.

$$\begin{aligned} P(t_* = 1 | t) &= \int P(t_* = 1 | a_*) P(a_* | t) da_* \\ &= \int \sigma_* P(a_* | t) da_* \end{aligned} \quad (4)$$

where $a_* = a(x_*)$, and $\sigma_* = \sigma(a_*)$ (Wang et al., 2009).

SVM

The SVM classifier maximises the margin around the hyperplane ($w^T x + b$) that separates samples from different classes. Here, w is the weight vector, x is the input vector, and b is the bias term (Vapnik, 2013). Each sample $i = 1, \dots, m$, consists of an

N-dimensional feature vector x_i and a class label $y_i \in \{+1, -1\}$. Maximising the margin is equivalent to minimising this expression:

$$w^*, b^*, \zeta^* = \arg \min_{w, b, \zeta} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \zeta_i \quad (5)$$

$$\text{s.t. } y_i (w^T x_i + b) \geq 1 - \zeta_i; \quad \zeta_i \geq 0; i = 1, \dots, m$$

In the soft-margin SVM equation, parameter C determines the weight of the misclassification penalty ζ_i within the margin. The final weight vector w^* represents the combined influence of all features on the classifier (Bron et al., 2015).

RF

A RF classifier comprises multiple decision tree classification models, $\{h(x, \theta_k), k = 1, 2, 3, \dots, N\}$, where each parameter $\{\theta_k\}$ is a distinct stochastic vector. Given an independent variable X , a decision tree classification model is selected through a voting process to determine the optimal classification result (Sun et al., 2021). During tree construction, features are chosen randomly, and the Gini index $G(T)$ for a specific feature is calculated using equation (6).

$$G(T) = 1 - \sum_{i=1}^c p_i^2 \quad (6)$$

The equation uses 'c' for the number of classes and ' p_i ' for the class ' i ' proportion in T . The split node is determined by selecting the feature with the lowest Gini index. This decision tree ranks features, helping the RF selection algorithm choose the most relevant ones, reducing computation time, and improving classification accuracy (Thushara et al., 2020).

3.4 Performance analysis

The effectiveness of the proposed methodology was evaluated using several key parameters, including accuracy (ACC), sensitivity (SEN), specificity (SPE), precision, recall, and F-measure (F1-score). These parameters were defined based on the counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) as specified in equations (7)–(11) (AlSaeed and Omar, 2022; Hwang et al., 2019; Beheshti et al., 2017; KP and P, 2022):

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

$$SEN / Recall = \frac{TP}{TP + FN} \quad (8)$$

$$SPE = \frac{TN}{TN + FP} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$F\text{-measure} = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (11)$$

4 Experimental results and discussion

This section presented the experimental setup and results, including software and hardware configuration details. It discussed the 3D-VGG-16 model for feature extraction and GPC, SVM, and RF classifiers for classification. To validate the proposed architecture, computed outcomes were compared with other techniques.

4.1 Experimental configuration

The experiments were conducted in a Python environment using Google Colaboratory Pro (Google Colaboratory, n.d.), which offered a hosted GPU for Python code execution. Python libraries, including TensorFlow (n.d.), Keras (Medium, n.d.), Scikit-learn (Medium, n.d.), Numpy (Medium, n.d.), and Nibabel (Neuroimaging in Python – NiBabel 5.3.0.dev5+g773e3c40 Documentation, n.d.), were employed. The study focused on brain anatomy analysis using ADNI (2003) and MIRIAD (2012) MRI datasets in the Neuroimaging Informatics Technology Initiative (NIFTI) format. The analysis of these images focused particularly on the coronal plane (x-z plane), which was known to provide advantages (Park and Moon, 2016). The dataset was randomly divided into the ratio 75:25 for training and testing. A weight initialisation technique was applied to mitigate the vanishing gradient problem. Adam optimiser was used here with default parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and it is initialised with a learning rate of 0.0001. Different dropout rates were used for different convolutional layers. 0.1 for conv1 block, 0.2 for conv2 and conv3 block, and 0.3 for conv4 and conv5 block is used. The 3D-VGG-16 model used categorical cross-entropy loss and a dropout rate of 0.3 for the output layer. During training, the learning rate was reduced by 20% if validation accuracy remained stagnant for ten consecutive epochs. Early stopping halted training if validation loss didn't decrease for 30 consecutive epochs. The hyperparameter selected for the model's training is stated in Table 2. A grid search method was used to select these specific parameters. Various combinations of parameters were trialled and tested to choose the selected parameters. With the help of these parameters, the proposed model achieved optimised results.

Table 2 Hyperparameters used for the proposed model

Regularisation technique	Dropout
Model	3D-VGG-16
Learning rate	0.0001
No. of epochs	300
Batch size	16
Loss function	Categorical cross-entropy
Optimiser	Adam
Dropout rate	0.1, 0.2, 0.3
Activation function	ReLU, Softmax

4.2 Performance analysis

Various experiments aimed to find the best 3D-VGG-16 and classifier model combination for efficient AD classification. Fusion architecture used various ML classifiers: 3D-VGG-16+GPC, 3D-VGG-16+SVM, and 3D-VGG-16+RF. To implement this approach, classifiers' parameters required tuning for optimisation through a grid search across various combinations. The GPC and SVM classifiers utilised radial basis function (RBF) and Linear kernels, necessitating tuning of '*c*' and '*gamma*' for SVM, '*n_estimators*' and '*max_depth*' for RF, and '*gamma value*' for GPC's RBF kernel. These parameter values were chosen from [*c* = 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1,000], [*gamma* = 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1,000], [*n_estimators* = 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20], [*max_depth* = 5, 6, 7, 8, 9], and [*gamma value* = 0.11, 0.22, 0.33, 0.44, 0.55, 0.6, 0.7, 0.8, 0.9, 1.0]. Figure 5 illustrates the parameter tuning process for optimising classification accuracy using GPC, SVM, and RF classifiers. Each graph displayed the classification accuracy of the ADNI and MIRIAD datasets for different parameter combinations. In the GPC graph, accuracy improved with tuning and peaked at around 0.987 for ADNI and 0.962 for MIRIAD. The SVM graph showed a steady increase in accuracy, reaching 0.975 for ADNI and 0.937 for MIRIAD. Similarly, the RF graph indicated that accuracy improved with tuning, achieving approximately 0.937 for ADNI and 0.912 for MIRIAD. In summary, tuning parameters enhanced accuracy for all classifiers, with GPC achieving the highest accuracy on both datasets.

Table 3 Performance of classifiers on ADNI and MIRIAD

<i>Dataset</i>	<i>Model</i>	<i>Classifiers</i>	<i>Accuracy</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>Precision</i>	<i>F-measure</i>
ADNI	3D-VGG-16	GPC	98.75	99.9	97.5	99.9	98.73
	3D-VGG-16	SVM	97.5	99.9	95	97.5	97.56
	3D-VGG-16	RF	93.75	95	92.5	94.9	93.82
MIRIAD	3D-VGG-16	GPC	96.25	92.5	99.9	95.12	96.10
	3D-VGG-16	SVM	93.75	99.9	87.5	92.7	94
	3D-VGG-16	RF	91.25	85	97.5	90.24	90.66

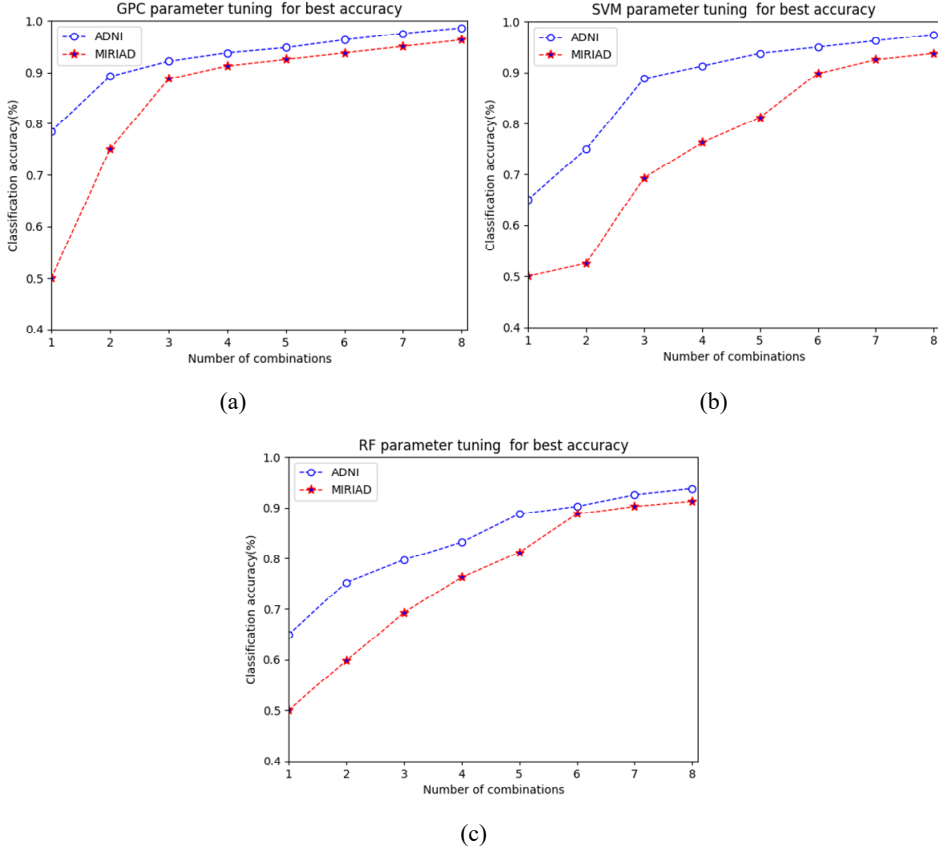
Figure 5 Parameter tuning to achieve best classification accuracy with (a) GPC (b) SVM and (c) RF classifiers (see online version for colours)

Table 3 presents the classifier's performance based on the selected parameters after fine-tuning. For the ADNI dataset, 3D-VGG-16+GPC achieved an accuracy of 98.75%, sensitivity of 99.9%, specificity of 97.5%, precision of 99.9%, and F-measure of 98.73%. In comparison, 3D-VGG-16+SVM performed well but not as effectively as 3D-VGG-16+GPC, with parameters of 97.5%, 99.9%, 95%, 97.5%, and 97.56%. 3D-VGG-16+RF exhibited slightly lower performance than both classifiers, with metrics of 93.75%, 95%, 92.5%, 94.9%, and 93.82%. Similarly, for the MIRIAD dataset, 3D-VGG-16+GPC achieved an accuracy of 96.25%, sensitivity of 92.5%, specificity of 99.9%, precision of 95.12%, and F-measure of 96.1%. The 3D-VGG-16+SVM performed well but not better than the 3D-VGG-16+GPC, with parameters of 93.75%, 99.9%, 87.5%, 92.7%, and 94%. The performance of 3D-VGG-16+RF was slightly lower than that of both classifiers, with metrics of 91.25%, 85%, 97.5%, 90.24%, and 90.66%. The results indicated that the GPC classifier performs better than SVM and RF for both datasets. The confusion matrices for all approaches are shown in Figure 6. Each matrix compared actual and predicted classifications: Figures 6(a) and 6(d) showed results for 3D-VGG-16 with GPC, Figures 6(b) and 6(e) used SVM, and 6(c) and 6(f) used RF. For the ADNI dataset, the models achieved high accuracies: Figures 6(a) and 6(b) both had much better classification, with only one and two misclassifications each, while Figure 6(c) showed

slightly more errors. For the MIRIAD dataset, Figure 6(d) maintained high accuracy with only three errors, Figure 6(e) had slightly more misclassifications, and Figure 6(f) showed the highest error rate among the models. These results implied that while all models performed well, the GPC were generally more accurate in distinguishing AD from CN subjects. Additionally, Table 4 presents class-specific performance using the confusion matrix for the 3D-VGG-16+GPC classifier on both datasets. This showed that the GPC classifier maintained consistent performance for both classes in the datasets. All classifiers illustrated consistently impressive performance, with GPC exhibiting the highest accuracy, followed by SVM and RF. These results indicated that the proposed model could perform effectively with different datasets.

Figure 6 Confusion matrices, (a) 3D-VGG-16+GPC (b) 3D-VGG-16+SVM, and (c) 3D-VGG-16+RF for ADNI dataset and (d) 3D-VGG-16+GPC (e) 3D-VGG-16+SVM, and (f) 3D-VGG-16+RF for MIRIAD dataset (see online version for colours)

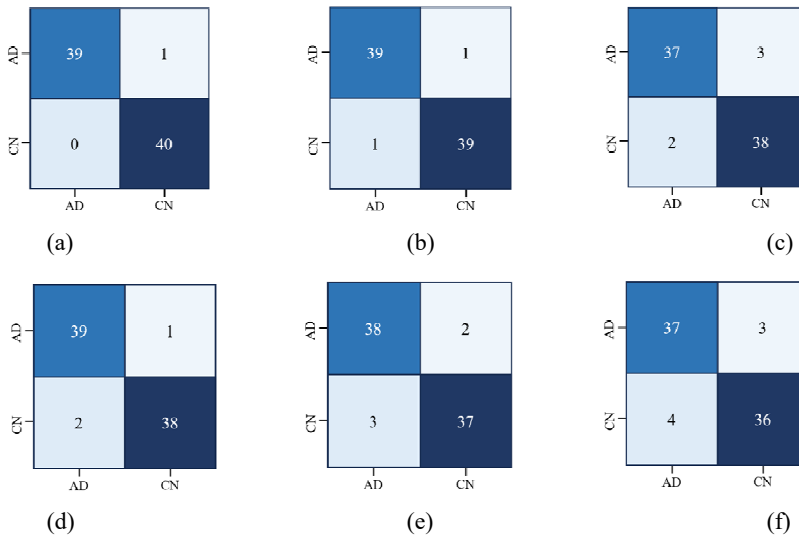


Table 4 3D-VGG-16+GPC experimental results with ADNI and MIRIAD

Dataset	Category	Precision	Recall	F1-score	Support (sample count)
ADNI	CN	97.56	99.9	98.72	40
	AD	99.9	97.5	98.69	40
	Accuracy			98.75	80
	Macro avg	98.73	98.7	98.74	80
	Weighted avg	98.73	98.7	98.74	80
MIRIAD	HC	99.9	92.5	96.06	40
	AD	93.02	99.9	96.34	40
	Accuracy			96.25	80
	Macro avg	96.51	96.25	96.24	80
	Weighted avg	96.51	96.25	96.24	80

Figure 7 Graphical analysis of classification performance based on ROC, PR, KS statistic plot for (a), (b), (c) GPC and (d), (e), (f) SVM classifier (see online version for colours)

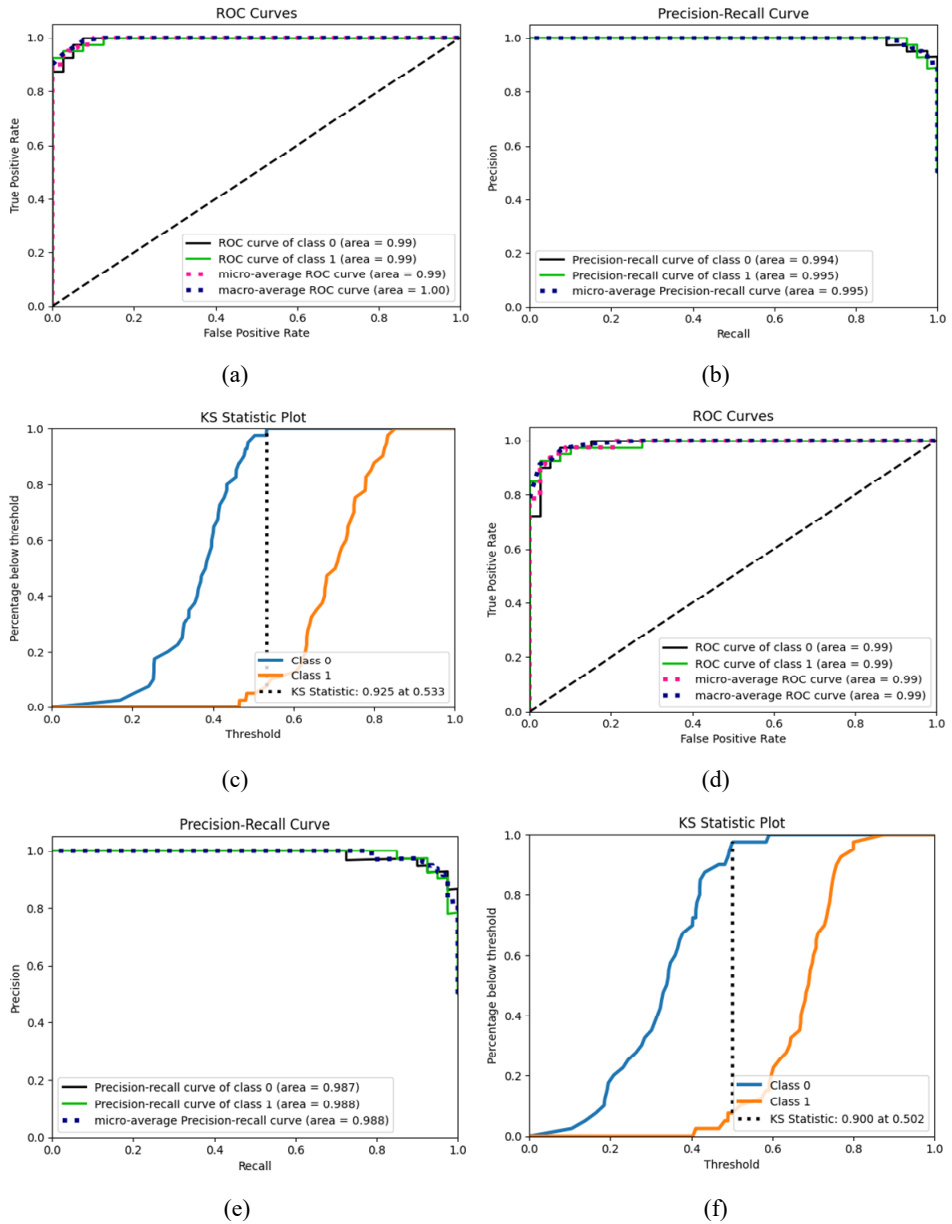
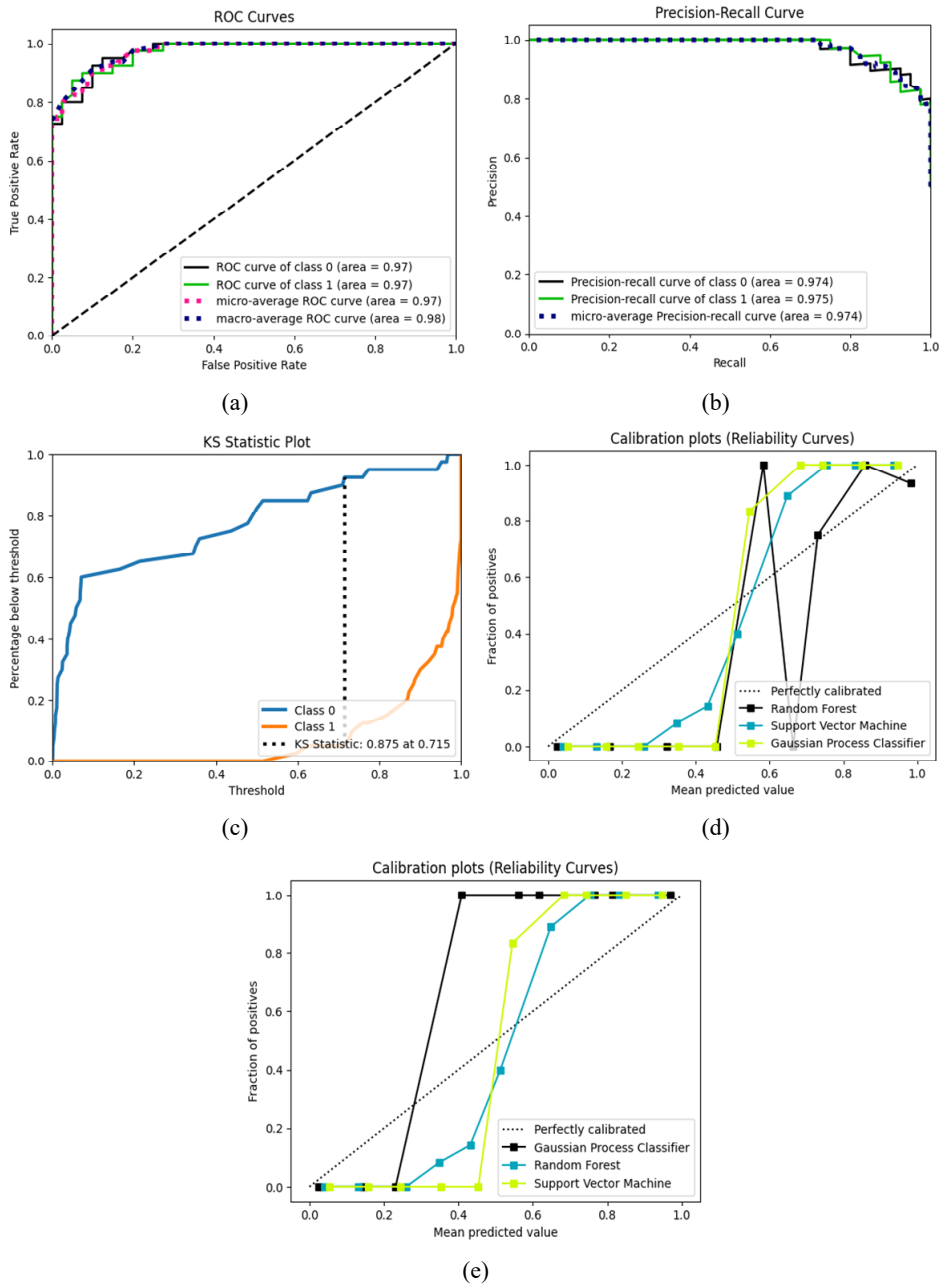


Figure 8 Graphical analysis of classification performance based on ROC, PR, KS statistic plot for (a), (b), (c) RF classifier, calibration plot for (d) MIRIAD and (e) ADNI dataset (see online version for colours)



Figures 7 and 8 present graphical analyses illustrating the effectiveness of the architecture in classification. The Receiver Operating Characteristic (ROC) curve assessed the True Positive Rate- False Positive Rate (TPR-FPR) trade-off for binary classification models. The precision-recall (PR) curve evaluated the balance between precision and recall. The Kolmogorov-Smirnov (KS) statistic plot compared data distribution differences between the two sets. Calibration plots evaluated the alignment between predicted probabilities and actual outcomes, indicating the reliability of probability estimates and potential bias. Models with a larger area under the curve (AUC) values in the ROC and PR curves were deemed to have better predictive power. The AUC values of the ROC curves for GPC, SVM, and RF classifiers were 0.99, 0.99, and 0.97, respectively. Similarly, the AUC values of the PR curves for GPC, SVM, and RF classifiers were 0.995, 0.988, and 0.974, respectively. The GPC classifier demonstrated a higher AUC value than the SVM and RF classifiers. Models with higher KS statistic values demonstrated superior discriminatory power. The KS statistic values for the GPC, SVM, and RF classifiers were 0.925, 0.90, and 0.875, respectively. The GPC classifier exhibited a greater KS statistic value than the SVM and RF classifiers. Hence, the analysis highlights that the GPC classifier outperformed the SVM and RF classifiers.

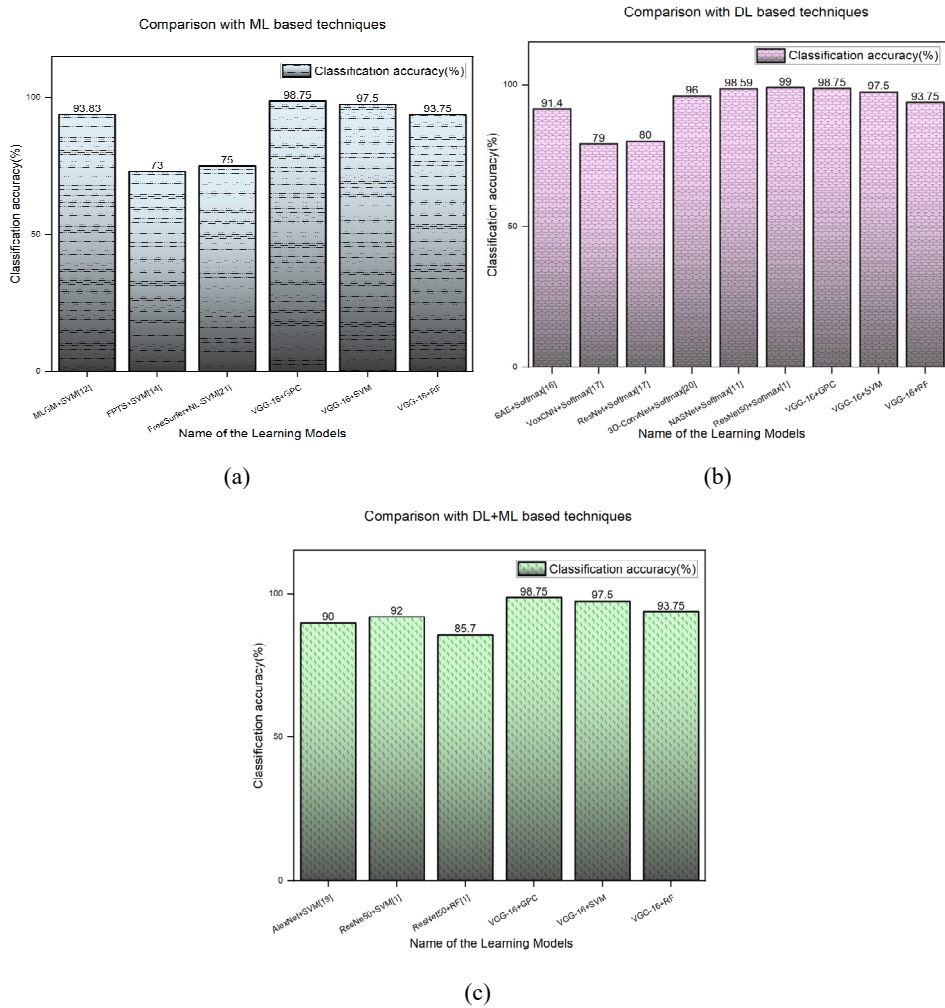
4.3 *Comparison with other studies*

The suggested approach has demonstrated promising results on both datasets, using three different classifiers for performance assessment. These outcomes underline the effectiveness of our approach, which is further compared to state-of-the-art techniques. Table 5 presents the performance comparison between our proposed method and other techniques. The 3D-VGG-16 based approach achieved better performance, particularly with 3D-VGG-16+GPC, achieving an accuracy of 98.75%, nearly on par with ResNet50+Softmax (AlSaeed and Omar, 2022). Similarly, 3D-VGG-16+SVM and 3D-VGG-16+RF also delivered impressive results compared to SVM-based (Liu et al., 2016; Lazli et al., 2018; Lin et al., 2018; Rallabandi et al., 2020; AlSaeed and Omar, 2022) and ResNet50+RF (AlSaeed and Omar, 2022) techniques. Figure 9 compared various methods based on their feature extraction and classification strategies. Our approach combined DL and ML, using a DL architecture for feature extraction and an ML classifier for classification. The comparison revealed that our approach performed better than other DL+ML and ML-based techniques. It achieved better results than specific DL-based methods (Liu et al., 2015; Korolev et al., 2017; Backstrom et al., 2018) and was nearly equivalent to others (Ji et al., 2019; AlSaeed and Omar, 2022). Thus, our approach exhibited better performance than other methods.

Table 5 Performance comparison of the proposed approach with other techniques

References	Data (scans or subjects)	Modalities	Techniques	Accuracy	Specificity	Sensitivity
Liu et. al. (2016)	ADNI (459 subjects)	MRI	Multiview learning using GM + SVM	93.83%	95.69%	92.78%
Lazli et. al. (2018)	ADNI (95 subjects)	MRI&PET	Fuzzy-Possibilistic Tissue Segmentation + SVM	73%	-	-
Liu et. al. (2015)	ADNI (758 subjects) (331 subjects)	MRI MRI&PET	Stacked auto-encoder SAE + Softmax	91.40%	90.42%	92.32%
Korolev et. al. (2017)	ADNI (231 scans)	MRI	VoxCNN + Softmax	79%	-	-
Lin et al. (2018)	ADNI (422 subjects)	MRI	ResNet + Softmax	80%	-	-
Backstrom et al. (2018)	ADNI (340 subjects)	MRI	AlexNet + SVM	90%	91%	87%
Ji et al. (2019)	ADNI (615 scans)	MRI	3D ConvNet + Softmax	98.37%	-	-
Rallabandi et al. (2020)	ADNI (1,167 scans)	MRI	ResNet50, NASNet, and MobileNet + Softmax	98.59%	-	-
AlSaeed and Omar (2022)	ADNI&MIRIAD (741 and 708 scans)	MRI	Non-linear SVM (RBF kernel)	75%	79%	75%
			ResNet50 + Softmax	99%	98%	99%
			ResNet50 + SVM	92%	91%	87%
			ResNet50 + RF	85.7%	88%	79%
			3D-VGG-16+GPC	98.75%	97.5%	99.9%
Proposed approach	ADNI&MIRIAD (239 and 206 scans)	MRI	3D-VGG-16+SVM	97.5%	95%	99.9%
			3D-VGG-16+RF	93.75%	92.5%	95%

Figure 9 Performance comparison of our approach with techniques uses, (a) ML for feature extraction and classification (b) DL for feature extraction and classification (c) DL for feature extraction and ML for classification (see online version for colours)



4.4 Discussion

The study proposed a fusion architecture by combining 3D-VGG-16 with three ML classifiers, resulting in three fused models. Among them, the 3D-VGG-16 + GPC model demonstrated better performance. The dropout regularisation in the 3D-VGG-16 and parameter tuning of classifiers through grid search significantly optimised the model's performance. Evaluations with the ADNI and MIRIAD datasets showed that the 3D-VGG-16+GPC model outperformed SVM and RF classifiers in accuracy, sensitivity, specificity, precision, and F-measure. Table 3 shows the evaluated parametric values for both datasets. For ADNI data, the GPC classifier achieved the highest accuracy (98.75%), specificity (97.5%), precision (99.9%), and F-measure (98.73%). The highest sensitivity, at 99.9%, was attained by both the GPC and SVM classifiers. For MIRIAD data, the GPC

classifier again achieved the highest accuracy (96.25%), specificity (99.9%), precision (95.12%), and F-measure (96.10%), while the SVM classifier achieved the highest sensitivity (99.9%). These results indicate that the GPC classifier outperformed SVM and RF for AD classification across both datasets.

Comparing results with other state-of-the-art techniques highlighted the advancements brought by the proposed model. Liu et al. (2016) reported 93.83% accuracy using multiview learning with GM + SVM on the ADNI dataset, which is substantially lower than the 98.75% achieved by the GPC classifier. Lazli et al. (2018) achieved 73% accuracy using Fuzzy-Possibilistic Segmentation + SVM, indicating a significant performance gap. Liu et al. (2015) achieved 91.40% accuracy with SAE + Softmax, and Korolev et al. (2017) reported 79% and 80% accuracy using VoxCNN + Softmax and ResNet + Softmax, respectively. Lin et al. (2018) achieved 90% accuracy with AlexNet + SVM, while Ji et al. (2019) reported 98.59% accuracy using ResNet50, NASNet, and MobileNet + Softmax. Backstrom et al. (2018) achieved 98.37% accuracy using 3D ConvNet + Softmax, and AlSaeed and Omar (2022) reported 92% and 85.7% accuracy using ResNet50 + SVM and ResNet50 + RF, respectively. The proposed approach 3D-VGG-16 + GPC model performed better than all these methods, except for a comparable accuracy of 99% reported by AlSaeed and Omar (2022) using ResNet50 + Softmax. These results highlight the effectiveness and potential of the 3D-VGG-16 + GPC model for clinical application.

The graphical analysis of different performance metrics provided further validation. The ROC curves for each classifier showed that the 3D-VGG-16+GPC model consistently achieved a higher AUC across both datasets, indicating better performance in distinguishing between AD and non-AD cases. The PR curves also showed higher PR values for the GPC classifier compared to others, validating its reliability and accuracy. Integrating DL with traditional ML classifiers significantly enhanced AD classification. The high sensitivity and specificity of the GPC classifier implied a substantial reduction in misdiagnoses, which was crucial for improving patient outcomes. Accurate and early detection of AD is critical for effective disease management and treatment planning. The proposed model's consistent high accuracy and reliability across different datasets suggested robust potential for clinical application. In summary, the 3D-VGG-16+GPC model excelled in classifying AD, surpassing other models and recent studies. Results from various metrics and graphical analyses highlight the model's robustness and reliability. This approach highlights the benefits of combining DL feature extraction with advanced ML classifiers. The key findings of the work are summarised below.

- The best fusion architecture is 3D-VGG-16+GPC, followed by SVM and RF.
- Dropout regularisation enhanced generalisability, resulting in consistent performance across both datasets.
- ROC, PR, and KS statistic plots validated the proposed approach. The AUC values were 0.99 for ROC, 0.995 for PR, and 0.925 for the KS statistic. These high AUC and KS statistic values indicated the reliability and robustness of the 3D-VGG-16 model for AD classification.
- Compared to state-of-the-art methods, the proposed architecture outperformed all except for ResNet50+Softmax (AlSaeed and Omar, 2022), which showed comparable performance.

The key differences of the proposed architecture compared to existing ones are as follows.

- Dropout was used in every layer of the architecture with varying rates.
- Classifier parameters were tuned using the grid search method.
- The proposed 3D-VGG-16 architecture was designed for 3D data, unlike the original VGG16, which was for 2D data.
- It was trained and tested with ADNI and MIRIAD datasets.
- Both datasets were applied, and their outcomes were evaluated simultaneously.

On the basis of the detailed investigation of the proposed model's performances, it is inferred that 3D-VGG-16+GPC achieved better values for the different metrics as compared to the others. The final conclusions, including various limitations of this study and future directions, are presented in the next section.

5 Conclusions

This work introduces an improved 3D-VGG-16 architecture for early-stage AD classification using structural MRI images. 3D-VGG-16 is used for feature engineering tasks that involve crucial input feature extraction. The extracted features are utilised for classification using GPC, SVM, and RF classifiers. The three approaches, 3D-VGG-16+GPC, 3D-VGG-16+SVM, and 3D-VGG-16+RF, are trained and tested on MRI images from the ADNI and MIRIAD datasets. The effectiveness of the approaches is examined using various performance metrics such as ACC, SEN, SPE, F-measure, and precision. The accuracies attained for 3D-VGG-16+GPC, 3D-VGG-16+SVM, and 3D-VGG-16+RF are 98.75%, 97.5%, and 93.75% using the ADNI dataset and 96.25%, 93.75%, and 91.25% using MIRIAD dataset, respectively. The models' performances are analysed using metrics such as ROC, PR, and KS-statistic curves. The 3D-VGG-16+GPC model shows higher AUC values for ROC (0.99), PR (0.995) curves, and KS statistic value (0.925) as compared to the other two models. The investigation of results highlights that the 3D-VGG-16+GPC classifier performs better than the 3D-VGG-16+SVM and 3D-VGG-16+RF. The class-specific performance of 3D-VGG-16+GPC, as per the confusion matrices for both datasets, shows consistent performance for both classes. All three classifiers illustrate impressive performance, where GPC exhibits the highest accuracy, followed by SVM and RF. These results indicate that the proposed fusion architecture can perform effectively with multiple datasets. While the proposed model demonstrated significant improvements, certain limitations still need to be addressed. Despite achieving a 98.75% accuracy, further enhancement is possible. Selecting varying dropout rates for each layer is time-consuming which requires numerous experiments. The proposed model only performs binary classification and is not designed for survival analysis. This opens the door for future research on techniques like transformers to refine AD classification performance. Another promising direction is to explore an adaptive or guided dropout technique to automatically select the optimal rate for different layers. In future, this research work will be expanded to include multi-class with various modalities and to perform survival analysis for disease progression with time.

References

- Alp, S., Akan, T., Bhuiyan, M.S., Disbrow, E.A., Conrad, S.A., Vanchiere, J.A., Kevil, C.G. and Bhuiyan, M.a.N. (2024) 'Joint transformer architecture in brain 3D MRI classification: its application in Alzheimer's disease classification', *Scientific Reports*, Vol. 14, <https://doi.org/10.1038/s41598-024-59578-3>.
- AlSaeed, D. and Omar, S.F. (2022) 'Brain MRI analysis for alzheimer's disease diagnosis using CNN-based feature extraction and machine learning', *Sensors*, Vol. 22, No. 8, p.2911, <https://doi.org/10.3390/s22082911>.
- Backstrom, K., Nazari, M., Gu, I.Y.H. and Jakola, A.S. (2018) 'An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images', *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, <https://doi.org/10.1109/isbi.2018.8363543>.
- Beheshti, I., Demirel, H. and Matsuda, H. (2017) 'Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm', *Computers in Biology and Medicine*, Vol. 83, pp.109–119, <https://doi.org/10.1016/j.compbiomed.2017.02.011>.
- Bron, E.E., Smits, M., Niessen, W.J. and Klein, S. (2015) 'Feature selection based on the SVM weight vector for classification of dementia', *IEEE Journal of Biomedical and Health Informatics*, Vol. 19, No. 5, pp.1617–1626, <https://doi.org/10.1109/jbhi.2015.2432832>.
- Castellazzi, G., Cuzzoni, M.G., Ramusino, M.C., Martinelli, D., Denaro, F., Ricciardi, A., Vitali, P., Anzalone, N., Bernini, S., Palesi, F., Sinforiani, E., Costa, A., Micieli, G., D'Angelo, E., Magenes, G. and Wheeler-Kingshott, C.A.M.G. (2020) 'A machine learning approach for the differential diagnosis of Alzheimer and vascular dementia fed by MRI selected features', *Frontiers in Neuroinformatics*, Vol. 14, <https://doi.org/10.3389/fninf.2020.00025>.
- Chen, Q., Fu, Q., Bai, H. and Hong, Y. (2024) 'LongFormer: longitudinal transformer for Alzheimer's disease classification with structural MRIs', *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, pp.3563–3572, DOI: 10.1109/WACV57701.2024.00354.
- Christina, P. (2018) *World Alzheimer Report 2018 – The State of the Art of Dementia Research: New Frontiers*, Vol. 48, New Frontiers; Alzheimer's Disease International, London, UK [online] <https://www.alzint.org/u/WorldAlzheimerReport2018.pdf> (accessed 12 January 2024).
- Fan, Z., Li, J., Zhang, L., Zhu, G., Li, P., Lu, X., Shen, P., Shah, S.A.A., Bennamoun, M., Hua, T. and Wei, W. (2021) 'U-net based analysis of MRI for Alzheimer's disease diagnosis', *Neural Computing and Applications*, Vol. 33, No. 20, pp.13587–13599, <https://doi.org/10.1007/s00521-021-05983-y>.
- Google Colaboratory (n.d.) [online] <https://colab.research.google.com/notebooks/pro.ipynb> (accessed 29 November 2023).
- Gunawardena, K.A.N.N.P., Rajapakse, R.N. and Kodikara, N.D. (2017) 'Applying convolutional neural networks for pre-detection of alzheimer's disease from structural MRI data', *2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, <https://doi.org/10.1109/m2vip.2017.8211486>.
- Huang, J., VanZijl, P.C.M., Han, X., Dong, C.M., Cheng, G.W.Y., Tse, K.H., Knutsson, L., Chen, L., Lai, J.H.C., Wu, E.X., Xu, J. and Chan, K.W.Y. (2020) 'Altered d -glucose in brain parenchyma and cerebrospinal fluid of early Alzheimer's disease detected by dynamic glucose-enhanced MRI', *Science Advances*, Vol. 6, No. 20, <https://doi.org/10.1126/sciadv.aba3884>.
- Hwang, H., Rehman, H.Z.U. and Lee, S. (2019) '3D U-Net for skull stripping in brain MRI', *Applied Sciences*, Vol. 9, No. 3, p.569, <https://doi.org/10.3390/app9030569>.
- Ji, H., Liu, Z., Yan, W., Q. and Klette, R. (2019) 'Early diagnosis of Alzheimer's disease using deep learning', *Proceedings of the 2nd International Conference on Control and Computer Vision*, pp.87–91, <https://doi.org/10.1145/3341016.3341024>.

- Khuyen, L. (2021) *An Overview of VGG16 and NiN Models* [online] <https://lekhuyen.medium.com/an-overview-of-vgg16-and-nin-models-96e4bf398484> (accessed 6 June 2024).
- Korolev, S., Safiullin, A., Belyaev, M. and Dodonova, Y. (2017) ‘Residual and plain convolutional neural networks for 3D brain MRI classification’, *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, <https://doi.org/10.1109/isbi.2017.7950647>.
- KP, M.N. and P, T. (2022) ‘Feature selection using efficient fusion of Fisher Score and greedy searching for Alzheimer’s classification’, *Journal of King Saud University – Computer and Information Sciences*, Vol. 34, No. 8, pp.4993–5006, <https://doi.org/10.1016/j.jksuci.2020.12.009>.
- Krashenyi, I., Popov, A., Ramirez, J. and Gorriz, J.M. (2016) ‘Fuzzy computer-aided diagnosis of Alzheimer’s disease using MRI and PET statistical features’, *2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO)*, <https://doi.org/10.1109/elnano.2016.7493045>.
- Kurlowicz, L. and Wallace, M. (1999) ‘The mini-mental state examination (MMSE)’, *Journal of Gerontological Nursing*, Vol. 25, No. 5, pp.8–9, <https://doi.org/10.3928/0098-9134-19990501-08>.
- Laha, M., Tripathi, P.C. and Bag, S. (2018) ‘A skull stripping from brain MRI using adaptive iterative thresholding and mathematical morphology’, *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, <https://doi.org/10.1109/rait.2018.8389028>.
- Lazli, L., Boukadoum, M. and Mohamed, O.A. (2018) ‘Computer-aided diagnosis system for Alzheimer’s disease using fuzzy-possibilistic tissue segmentation and SVM classification’, *2018 IEEE Life Sciences Conference (LSC)*, <https://doi.org/10.1109/lsc.2018.8572122>.
- Lin, L., Zhang, B. and Wu, S. (2018) ‘Hybrid CNN-SVM for Alzheimer’s disease classification from structural MRI and the Alzheimer’s disease neuroimaging initiative (ADNI)’, *2018 International Conference on Biomedical Engineering, Machinery and Earth Science (BEMES 2018)*, <https://doi.org/10.25236/bemes.2018.041>.
- Liu, L., Liu, S., Zhang, L., To, X.V., Nasrallah, F. and Chandra, S.S. (2023) ‘Cascaded multi-modal mixing transformers for Alzheimer’s disease classification with incomplete data’, *NeuroImage*, Vol. 277, p.120267, <https://doi.org/10.1016/j.neuroimage.2023.120267>.
- Liu, M., Zhang, D., Adeli, E. and Shen, D. (2016) ‘Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer’s disease diagnosis’, *IEEE Transactions on Biomedical Engineering*, Vol. 63, No. 7, pp.1473–1482, <https://doi.org/10.1109/tbme.2015.2496233>.
- Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J. and ADNI (2015) ‘Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer’s disease’, *IEEE Transactions on Biomedical Engineering*, Vol. 62, No. 4, pp.1132–1140, <https://doi.org/10.1109/tbme.2014.2372011>.
- Medium (n.d.) *Medium* [online] <https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c> (accessed 30 January 2024).
- Neuroimaging in Python – NiBabel 5.3.0.dev5+g773e3c40 Documentation (n.d.) [online] <https://nipy.org/nibabel/> (accessed 30 November 2023).
- Park, M. and Moon, W.J. (2016) ‘Structural MR imaging in the diagnosis of Alzheimer’s disease and other neurodegenerative dementia: current imaging approach and future perspectives’, *Korean Journal of Radiology*, Vol. 17, No. 6, p.827, <https://doi.org/10.3348/kjr.2016.17.6.827>.
- Rallabandi, V.P.S., Tulpule, K. and Gattu, M. (2020) ‘Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer’s disease using structural MRI analysis’, *Informatics in Medicine Unlocked*, Vol. 18, p.100305, <https://doi.org/10.1016/j.imu.2020.100305>.

- Rasmussen, C.E. and Williams, C.K.I. (2005) *Gaussian Processes for Machine Learning*, MIT Press [online] http://books.google.ie/books?id=Tr34DwAAQBAJ&printsec=frontcover&dq=Gaussian+Processes+for+Machine+Learning&hl=&cd=1&source=gb_api (accessed 24 Dec 2023).
- Rehman, H.Z.U., Hwang, H. and Lee, S. (2020) 'Conventional and deep learning methods for skull stripping in brain MRI', *Applied Sciences*, Vol. 10, No. 5, p.1773, <https://doi.org/10.3390/app10051773>.
- Rohini, G. (2021) *Everything You Need to Know About VGG16* [online] <https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918> (accessed 24 February 2024).
- Roy, S. and Maji, P. (2015) 'A simple skull stripping algorithm for brain MRI', *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, <https://doi.org/10.1109/icapr.2015.7050671>.
- Salmon, D.P. and Bondi, M.W. (2009) 'Neuropsychological assessment of dementia', *Annual Review of Psychology*, Vol. 60, No. 1, pp.257–282, <https://doi.org/10.1146/annurev.psych.57.102904.190024>.
- Sarraf, S., Sarraf, A., DeSouza, D.D., Anderson, J.a.E. and Kabia, M. (2023) 'OViTAD: optimized vision transformer to predict various stages of alzheimer's disease using resting-state fMRI and structural MRI data', *Brain Sciences*, Vol. 13, p.260, <https://doi.org/10.3390/brainsci13020260>.
- Shattuck, D.W. and Leahy, R.M. (2002) 'BrainSuite: an automated cortical surface identification tool', *Medical Image Analysis*, Vol. 6, No. 2, pp.129–142, [https://doi.org/10.1016/s1361-8415\(02\)00054-3](https://doi.org/10.1016/s1361-8415(02)00054-3).
- Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A. and Leahy, R.M. (2001) 'Magnetic resonance image tissue classification using a partial volume model', *NeuroImage*, Vol. 13, No. 5, pp.856–876, <https://doi.org/10.1006/nimg.2000.0730>.
- Simonyan, K. and Zisserman, A. (2015) *Very Deep Convolutional Networks for Large-Scale Image Recognition*, <https://doi.org/10.48550/arXiv.1409.1556> (accessed 28 March 2023).
- Sun, H., Wang, A., Feng, Y. and Liu, C. (2021) 'An optimized random forest classification method for processing imbalanced data sets of Alzheimer's disease', *2021 33rd Chinese Control and Decision Conference (CCDC)*, <https://doi.org/10.1109/ccdc52312.2021.9602177>.
- TensorFlow (n.d.) *TensorFlow* [online] <https://www.tensorflow.org/> (accessed 30 November 2023).
- The Alzheimer's Disease Neuroimaging Initiative (ADNI) (2003) [online] <https://ida.loni.usc.edu/login.jsp?project=ADNI> (accessed 12 March 2023).
- Thulasi, N.P.K. and Varghese, D. (2018) 'A novel approach for diagnosing Alzheimer's disease using SVM', *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, <https://doi.org/10.1109/icoei.2018.8553789>.
- Thushara, A., Amma, U.D.C., John, A. and Saju, R. (2020) 'Multimodal MRI based classification and prediction of Alzheimer's disease using random forest ensemble', *2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, <https://doi.org/10.1109/accthp49271.2020.9213211>.
- UCL Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) (2012) [online] <https://www.ucl.ac.uk/drc/research/research-methods/minimal-interval-resonance-imaging-alzheimers-disease-miriad> (accessed 7 March 2023).
- Vapnik, V.N. (2013) *The Nature of Statistical Learning Theory*, Springer Science & Business Media [online] http://books.google.ie/books?id=EoDSBwAAQBAJ&pg=PR5&dq=978-1-4757-2440-0&hl=&cd=1&source=gb_api (accessed 25 September 2023).
- Wang, B., Wan, F., Mak, P.U., Mak, P.I. and Vai, M.I. (2009) 'EEG signals classification for brain computer interfaces based on Gaussian process classifier', *2009 7th International Conference on Information, Communications and Signal Processing (ICICS)*, <https://doi.org/10.1109/icics.2009.5397570>.

- Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Cedarbaum, J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Luthman, J., Morris, J.C., Petersen, R.C., Saykin, A.J., Shaw, L., Shen, L., Schwarz, A., Toga, A.W. and Trojanowski, J.Q. (2015) '2014 update of the Alzheimer's disease neuroimaging initiative: a review of papers published since its inception', *Alzheimer's & Dementia*, Vol. 11, No. 6, <https://doi.org/10.1016/j.jalz.2014.11.001>.
- Williams, C. and Barber, D. (1998) 'Bayesian classification with Gaussian processes', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, pp.1342–1351, <https://doi.org/10.1109/34.735807>.
- Xin, J., Wang, A., Guo, R., Liu, W. and Tang, X. (2023) 'CNN and swin-transformer based efficient model for Alzheimer's disease diagnosis with sMRI', *Biomedical Signal Processing and Control*, Vol. 86, p.105189, <https://doi.org/10.1016/j.bspc.2023.105189>.
- Zhang, D. and Shen, D. (2011) 'Multi-modal multi-task learning for joint prediction of clinical scores in Alzheimer's disease', *Multimodal Brain Image Analysis*, pp.60–67, https://doi.org/10.1007/978-3-642-24446-9_8.
- Zhang, D. and Shen, D. (2012) 'Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease', *NeuroImage*, Vol. 59, No. 2, pp.895–907, <https://doi.org/10.1016/j.neuroimage.2011.09.069>.
- Zimmermann, J., Perry, A., Breakspear, M., Schirner, M., Sachdev, P., Wen, W., Kochan, N.A., Mapstone, M., Ritter, P., McIntosh, A.R. and Solodkin, A. (2018) 'Differentiation of Alzheimer's disease based on local and global parameters in personalized virtual brain models', *NeuroImage: Clinical*, Vol. 19, pp.240–251, <https://doi.org/10.1016/j.nicl.2018.04.017>.

Appendix

Table A1 List of abbreviations

<i>Abbreviation</i>	<i>Full Form</i>	<i>Abbreviation</i>	<i>Full form</i>
MRI	Magnetic resonance imaging	AUC	Area under the curve
AD	Alzheimer's disease	MIRIAD	Minimal interval resonance imaging in Alzheimer's disease
ML	Machine learning	2D	Two-dimensional
DL	Deep learning	CN	Cognitively normal
GPC	Gaussian process classifier	HC	Healthy controls
SVM	Support vector machine	3D	Three-dimensional
RF	Random forest	IR-FSPGR	Inversion recovery prepared fast spoiled gradient recalled
ADNI	Alzheimer's disease neuroimaging initiative	TR	Repetition time
VGG	Visual geometry group	TE	Echo time
CSF	Cerebrospinal fluids	TI	Inversion time
fMRI	Functional MRI	BSE	Brain Surface Extractor
PET	Positron emission tomography	FCs	Fully connected layers
MMSE	Mini-mental state examinations	ACC	Accuracy
ADAS-Cog	Alzheimer's disease assessment scale cognitive	SEN	Sensitivity
CDR	Clinical dementia rating	SPE	Specificity
ROIs	Region of interests	TP	True positive
sMRI	Structural mri	TN	True negative
GM	Grey matter	FP	False positive
SURF	Speeded up robust features	FN	False negative
SAE	Stacked auto encoder	GPU	Graphics processing unit
ConvNet/CNN	Convolutional neural network	NIFTI	Neuroimaging informatics technology initiative
VoxCNN	Volumetric cnn	RBF	Radial basis function
ResNet	Residual neural network	GLCM	Grey level co-occurrence matrix
ROC	Receiver operating characteristic	TPR	True positive rate
KS	Kolmogorov-Smirnov	FPR	False positive rate