Journal of Radiation Research and Applied Sciences 17 (2024) 100782

Contents lists available at ScienceDirect

Journal of Radiation Research and Applied Sciences

journal homepage: www.journals.elsevier.com/journal-of-radiation-research-and-applied-sciences

Automated Alzheimer's disease classification using deep learning models with Soft-NMS and improved ResNet50 integration

Yusi Chen [a,b,c,*], Lizhen Wang [f], Bijiao Ding [e], Jianshe Shi [e], Tingxi Wen [d], Jianlong Huang [a,b,c], Yuguang Ye [a,b,c,**]

[a] Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, China
[b] Fujian Provincial Key Laboratory of Data Intensive Computing, Quanzhou, China
[c] Key Laboratory of Intelligent Computing and Information Processing, Fujian Province University, Quanzhou, China
[d] School of Computer Science and Technology, Huaqiao University, Quanzhou, China
[e] Department of Diagnostic Radiology, Huaqiao University Affiliated Strait Hospital, Quanzhou, China
[f] China Mobile Communications Group Fujian Co., Ltd, Quanzhou Branch, Quanzhou, China

ARTICLE INFO

Keywords:
AD
MRI
Faster R–CNN
Soft-NMS
Improved ResNet50

ABSTRACT

Background and objective: The accurate classification of Alzheimer's disease (AD) from MRI data holds great significance for facilitating early diagnosis and personalized treatment, ultimately leading to improved patient outcomes. To address this challenge, a comprehensive approach is proposed in this study, which integrates advanced deep learning models.

Methods: This study introduces an ensemble deep learning model for AD classification, which incorporates Soft-NMS into the Faster R–CNN architecture to enhance candidate information merging and improve detection accuracy. Additionally, an improved ResNet50 network is used for feature extraction, effectively extracting richer image features. To process sequence data, the model incorporates the Bidirectional Gated Recurrent Unit (Bi-GRU) as a key component in the feature extraction network. The final classification results are obtained using the enhanced Faster R–CNN.

Results: The results demonstrated outstanding classification performance in distinguishing AD cases, with the AD vs CN task achieving the highest accuracy of 98.91%. The model's accuracy and precision remained high for AD vs MCI and MCI vs CN groups, indicating its effectiveness in distinguishing different cognitive impairment stages. Additionally, the effectiveness of the approach was confirmed through object detection and feature extraction, demonstrating superiority over existing methods. The high accuracy of the proposed method indicates its potential for early AD diagnosis and personalized treatment.

Conclusion: This study presents a novel approach for AD classification using advanced deep learning models. The achieved high precision and accuracy offer promising opportunities for early diagnosis and personalized intervention. Addressing limitations related to data availability and manual annotation is crucial for future research advancements. Integrating multimodal imaging data and conducting longitudinal studies can further enhance our understanding of AD progression.

1. Introduction

Alzheimer's disease (AD) is a widespread neurodegenerative disorder known for its distinctive features, including cognitive decline, memory loss, and behavioral alterations (Hardy et al., 1992). It is a major global health challenge, affecting millions of individuals worldwide (Alzheimer's Association, 2019). Early and accurate diagnosis of AD is crucial for timely intervention and effective treatment to improve patient outcomes.

Recently, there has been increasing interest in employing advanced imaging techniques and computational approaches to enhance the accuracy of AD diagnosis. Both magnetic resonance imaging (MRI) and computed tomography (CT) play crucial roles in studying the structural and functional changes in the brain associated with AD (Jack et al.,

* Corresponding author. Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, China.
** Corresponding author. Faculty of Mathematics and Computer Science, Quanzhou Normal University, Quanzhou, China.
E-mail addresses: cys16050@qztc.edu.cn (Y. Chen), chinaye@qztc.edu.cn (Y. Ye).

https://doi.org/10.1016/j.jrras.2023.100782
Received 8 August 2023; Received in revised form 17 October 2023; Accepted 2 December 2023
Available online 1 January 2024

2012; Wong, Xu, & Ayoub, 2023; Wong, Xu, & Chen, 2023). These imaging modalities provide valuable insights into the pathological changes, such as atrophy and infarcts, that occur in the brain during the course of AD (Schwarz et al., 2016; Thompson et al., 2012). Deep learning techniques, notably Convolutional Neural Networks (CNNs), have demonstrated significant potential in medical image analysis, including the diagnosis of AD (Litjens et al., 2017). By automatically learning complex features from brain images, CNNs can capture subtle patterns and abnormalities indicative of AD (Sarraf et al., 2016). Various studies have effectively utilized deep learning models to enhance the diagnostic precision of AD through MRI and CT data analysis. For instance, Sarraf and Tofighi (Li et al., 2018) developed a CNN-based architecture for AD classification using MRI and functional MRI (fMRI) data. The model exhibited remarkable accuracy in discriminating healthy controls from individuals with mild cognitive impairment (MCI) or AD. Li et al. (Zeng et al., 2019a) introduced a multiple-cluster dense convolutional network for AD diagnosis using MRI data, showcasing encouraging outcomes in terms of accuracy and robustness.

In addition to deep learning, other computational approaches (Wong, Ayoub, & Cao, 2023) have been explored to enhance AD diagnosis. Zeng et al. (Zeng et al., 2019b) proposed a hybrid deep learning model that combined CNN and stacked autoencoders for AD classification using multimodal MRI data. Their model effectively extracted discriminative features, enabling accurate differentiation between MCI and AD patients. These studies highlight the potential of advanced computational techniques in improving AD diagnosis.

A comprehensive AD classification framework is proposed, integrating advanced algorithms, such as Faster R–CNN, ResNet, and Bidirectional Gated Recurrent Unit (Bi-GRU). The integration aims to enhance the accuracy and robustness of AD classification for early and precise disease identification.

Faster R–CNN, introduced by Ren et al. (Ren et al., 2015), is a groundbreaking object detection framework that exhibits excellent performance in localizing and recognizing objects in images. The approach utilized in this study involves the implementation of the Faster R–CNN algorithm to accurately detect and localize AD-related brain abnormalities in MRI scans. This information is crucial for effective AD classification.

ResNet, introduced by He et al. (He et al., 2016), has significantly impacted image classification tasks by addressing the issue of gradient vanishing and enabling the training of deep neural networks. The residual connections in ResNet play a crucial role in this process. In the proposed method, ResNet functions as an effective feature extractor, capturing essential high-level representations from MRI data, which are instrumental in distinguishing AD from healthy brain images.

To capture the inherent temporal dependencies and sequential patterns in MRI sequences, Bi-GRU is incorporated into the classification framework. Bi-GRU, an extension of the Gated Recurrent Unit (GRU) proposed by Cho et al. (Cho et al., 2014), allows bidirectional information flow and facilitates the modeling of long-term dependencies in AD progression.

The fusion of Faster R–CNN, ResNet, and Bi-GRU in the proposed framework offers complementary advantages. Faster R–CNN precisely localizes AD-related abnormalities, ResNet extracts informative features from MRI data, and Bi-GRU models the temporal dynamics in AD progression. This integrated approach aims to improve the accuracy and reliability of AD classification, promoting early detection and personalized treatment strategies. The objective of this study is to evaluate the performance of the proposed method in AD classification and explore its potential for enhancing AD diagnosis. The results are compared with other state-of-the-art segmentation methods, evaluating metrics such as accuracy, sensitivity, and specificity. This research aims to contribute to the advancement of AD diagnosis and provide assistance in improving disease diagnosis through advanced computational techniques.

2. Methodology

The objective of this study is to achieve AD identification in MRI images by combining Faster R–CNN, ResNet, and Bi-GRU algorithms. The method aims to leverage the strengths of each algorithm for efficient feature extraction and capturing relevant information related to AD. The specific content of the algorithm is shown in Fig. 1.

This paper proposes an improved AD classification algorithm based on Faster R–CNN, and the overall algorithm flow is depicted in Fig. 1. Before feature extraction, the input images undergo preprocessing to address the issue of insufficient feature learning caused by low image contrast. The preprocessed images are then fed into the enhanced ResNet50 feature extraction network, generating deeper fused features. Subsequently, the AD data extracted by ResNet is subjected to sequence modeling using Bi-GRU, producing high-level semantic features. This enhances the feature representation capability and model performance.

Next, the network generates candidate regions through Region Proposal Network (RPN) and introduces Soft-NMS (Non-Maximum Suppression) to improve candidate feature merging strategy, avoiding overlapping or missed features, and enhancing the network's detection accuracy in multi-object features. The candidate features extracted from the Pooling layer are processed into specific feature vectors. Finally, the network employs fully connected layers for classification and regression, resulting in accurate AD classification.

During the model training and optimization process, we first train the complete network on the training set and then fine-tune it using the validation set (Wong, 2023). To enhance the model's performance and generalization, suitable optimization algorithms like stochastic gradient descent, along with carefully selected learning rates, batch sizes, and training iterations, are employed. Furthermore, additional optimization strategies such as learning rate decay, weight decay, and early stopping are implemented to further enhance the model's overall performance.

The proposed method synergistically combines the strengths of Faster R–CNN, ResNet, and Bi-GRU algorithms to effectively utilize MRI image features and capture relevant information for AD identification. Fine-tuning the optimization strategies and experimental parameters can further enhance the model's performance and generalization ability. Extensive validation in computer journal papers has shown significant advancements in medical image analysis. In this study, the performance of the Faster R–CNN + ResNet + Bi-GRU algorithm for AD identification is experimentally assessed.

2.1. MRI data preprocessing

MRI is a widely used non-invasive imaging technique in medical research, particularly for studying brain structure and function. Before analysis, MRI data undergoes preprocessing steps to improve data quality and facilitate subsequent analysis. This preprocessing includes operations such as cropping, resizing, and grayscale normalization, aimed at enhancing image quality and comparability. Additionally, intensity normalization is applied to account for variability among subjects, and denoising techniques are employed to enhance the signal-to-noise ratio (Koonjoo et al., 2021).

During the data preprocessing stage, MRI images undergo various fundamental operations to optimize their quality and ensure comparability. These operations encompass cropping, resizing, and grayscale normalization, among others. By applying these techniques, overall image quality is improved, and data is prepared for subsequent analysis and feature extraction, ensuring the reliability and consistency of research results.

The objective of implementing these preprocessing steps is to enhance the quality and consistency of MRI data, minimize artifacts, and reduce inter-individual variability. Subsequently, the preprocessed images can be further analyzed and segmented using the proposed method, providing a more accurate and reliable data foundation for medical research and clinical applications in the field of AD classification using MRI images.

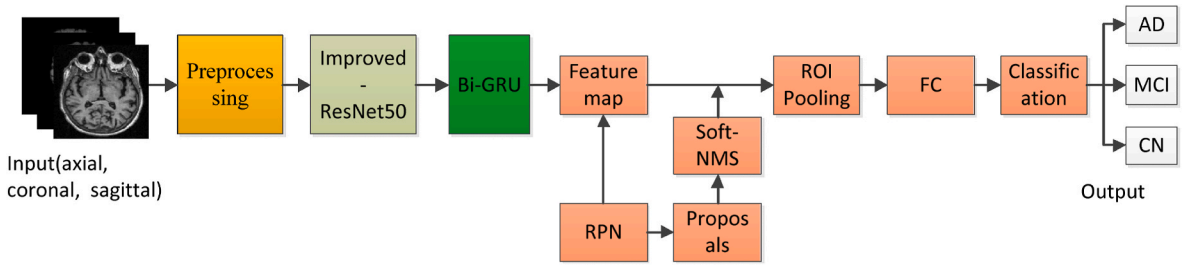


Fig. 1. Target detection algorithm flow.

2.2. Introduce Soft-NMS to faster R–CNN network

The Faster R–CNN network (Mirchandani et al., 2021) is a crucial algorithm employed in medical image analysis. This paper aims to achieve AD recognition in MRI images. With its efficient and accurate candidate region detection, the algorithm has demonstrated promising outcomes in various computer vision tasks.

Introducing Soft-NMS into Faster R–CNN has many advantages. The first is that Soft-NMS is an improved non-maximum suppression method. It adopts a progressive approach to reduce the score of detection frames with high overlap. Even if there are multiple highly overlapping boxes, Soft-NMS will retain part of their information, thereby reducing the information loss in object detection. The second is that with Soft-NMS, highly overlapping bounding boxes are not completely removed. This helps retain more object information, especially around object boundaries. The result is more accurate object bounding boxes and better object positioning. The third is that Soft-NMS allows parameters to be adjusted according to the requirements of specific tasks and data sets, resulting in better performance. The last point is that Soft-NMS helps to improve the robustness of detection and reduce the false alarm rate while retaining more detection information.

The Faster R–CNN algorithm employs a two-stage approach for object detection, comprising a RPN and a region-based convolutional neural network (R–CNN). In AD identification in MRI images, RPN plays a crucial role in generating region proposals that may contain AD-associated features. RPN efficiently scans the input image and proposes regions that may contain relevant information, thereby reducing the search space for subsequent processing.

To further improve the candidate feature merging strategy, the Faster R–CNN network incorporates Soft-NMS to refine candidate region proposals. Soft-NMS is an enhanced version of traditional NMS, which can avoid feature overlap or omission, thereby improving the detection accuracy of the network, especially when there are multiple object features in medical images.

The NMS algorithm is widely used in target detection to solve the problem of a large number of candidate information overlaps during classifier classification. It generates a series of candidate information in the picture and its corresponding score set. The algorithm will select the candidate information with the highest score, calculate the degree of overlap with the remaining candidate information, and delete the candidate information whose calculation result is greater than the set threshold. The NMS algorithm formula is as follows:

$$s_i = \{b_i | Score(b_i) > \text{threshold } \forall b_i \in boxes\} \tag{1}$$

Among them, $Score(b_i)$ represents the score of candidate information b_i, and *boxes* are the set of all candidate information to be processed. It can be seen from the algorithm that the NMS algorithm zeros the candidate information that is adjacent to the highest-scoring candidate information and whose overlap is greater than the threshold, which may lead to missed detection of target information when detecting images containing multiple targets.

To address this issue, various methods have been proposed, this paper introduces an improved Soft-NMS algorithm, which sets a decay function instead of directly zeroing the adjacent candidate information based on the size score of the overlapping part, thereby improving the detection accuracy under multiple target regions. The improved Soft-NMS formula is expressed as follows:

$$s_i = \left\{ b_i | Score(b_i) \geq \max_{b \in B} Score(b) \times e^{-\lambda \cdot A4R(b_i, b)} \ \forall b_i \in boxes \right\} \tag{2}$$

Among them, $Score(b_i)$ is the score of candidate information b_i; B is the set of all candidate information to be processed; $A4R(b_i, b)$ is the overlap rate between b_i and b; λ is the parameter of the decay function.

The adoption of the improved Soft-NMS algorithm effectively addresses the missing detection issue in multi-target area detection, which was present in the original NMS algorithm. As a result, the overall detection accuracy of the algorithm is significantly enhanced.

After region proposals are generated, the selected candidate features are further processed through pooling layers to convert region-specific features into fixed-size feature vectors. This process is critical to maintaining a consistent input size for subsequent classification and regression tasks.

The final step of the Faster R–CNN network involves the use of fully connected layers for classification and regression. In the context of AD recognition, networks are trained to predict whether a given region proposal contains AD-related features. The regression task aims to refine the bounding boxes of detected regions, providing precise localization of potentially AD-related regions.

Faster R–CNN networks represent a major advance in the field of medical image analysis. It demonstrates the successful application of deep learning-based object detection techniques to the challenging task of AD identification in MRI images. The combination of RPN, Soft-NMS, Pooling, and fully connected layers can achieve efficient and accurate feature extraction and enhance the overall performance and reliability of AD classification. In addition, the performance of the Faster R–CNN network in AD recognition can be further improved by optimizing hyperparameters, exploring novel feature extraction techniques, and employing transfer learning to utilize pre-trained models on large-scale datasets.

2.3. ResNet50 network

The feature extraction network comprises CNNs, commonly utilizing heavyweight networks such as AlexNet, ZFNet, VGG16, GoogleNet, and ResNet. GoogleNet and ResNet (Hadipour-Rokni et al., 2023) increase the network depth during model optimization. However, GoogleNet's redundant layers may lead to a decline in training set accuracy, while ResNet effectively addresses the issue of model degradation by introducing residual learning. ResNet adopts skip connections, altering the learning objective to $F(x) = H(x) - x$, where x represents the input to the CNN, $H(x)$ is the output after processing, and $F(x)$ is the final output. By propagating the original input alongside the processed output through skip connections, ResNet effectively learns the residual mapping to obtain the final representation.

The innovative approach in ResNet allows successful training of very deep networks without encountering vanishing gradients, ensuring efficient learning and improved performance in various computer vision tasks.

Originally, Faster R–CNN uses VGG16 as its backbone network. However, when extracting features from MRI data for AD detection, VGG16's limitations arise due to the images' blurriness and single-channel nature, making it challenging to effectively capture deep-level image features. To address this issue, the introduction of the ResNet50 residual network allows for the extraction of deeper-level feature maps, which in turn enhances the representation efficiency of the images. The advantages of the improved ResNet50 network are: first, it introduces the concept of residual learning, which combines the original input with the processed output through skip connections, making it easier to train deep networks and avoid the vanishing gradient problem, thus improving performance. The second is that the improved ResNet50 network achieves multi-scale information aggregation during the feature extraction process, helping to more comprehensively capture the information in the image. The third is that the improved ResNet50 reduces the number of network parameters, while simplifying the training process and improving computational efficiency.

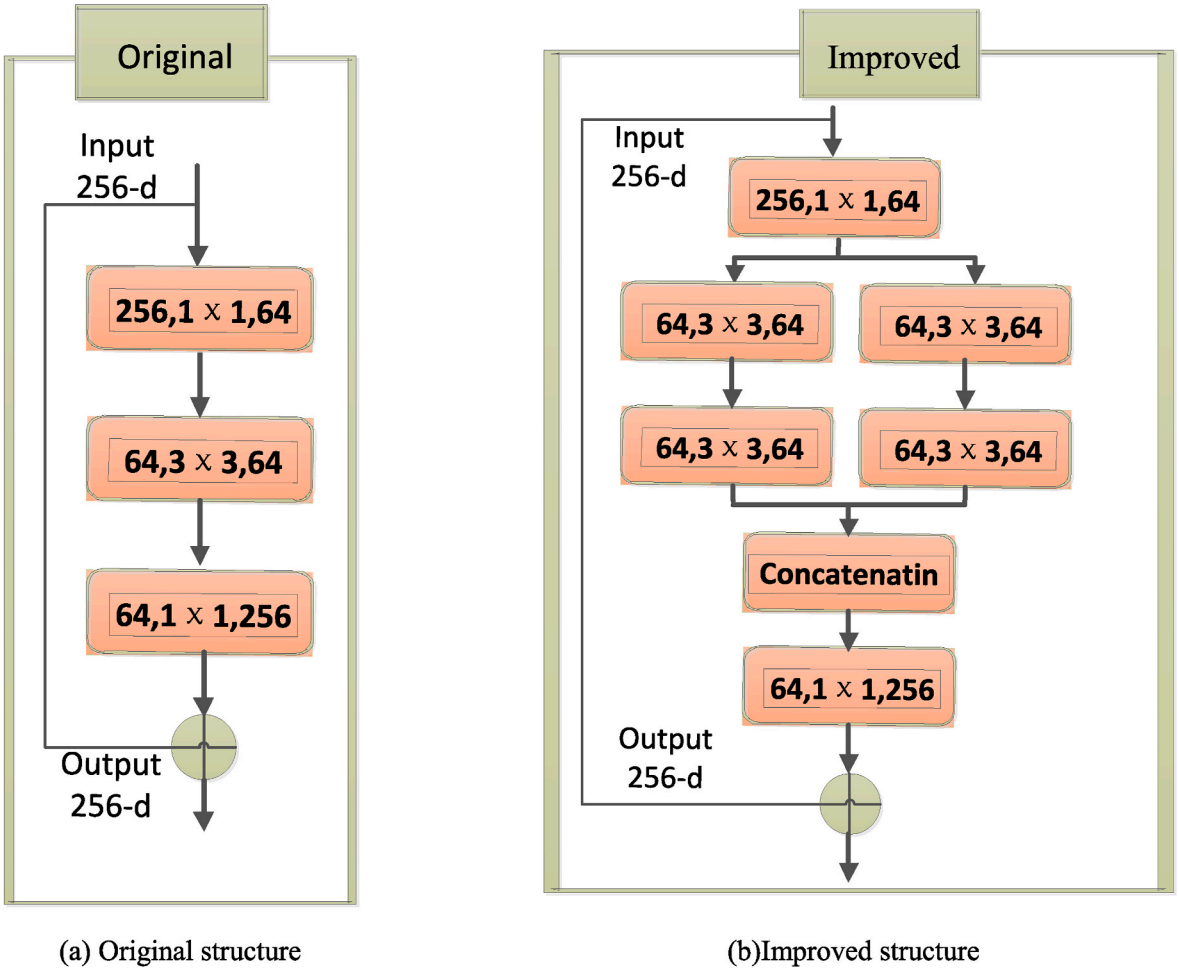
To enhance computational efficiency and reduce complexity, dimensionality reduction is performed before large-block convolutions, balancing the network's width and depth to improve detection accuracy. This approach significantly increases the number of units at each step while partially limiting computational complexity. Furthermore, aggregating visual information at different scales enhances the concept of multi-scale feature extraction.

The proposed variant structure, Improved-ResNet50, is illustrated in Fig. 2(b). It enhances ResNet50 by incorporating two consecutive sets of 3×3 convolution operations. Compared to the original network architecture in Fig. 2(a), this maintains the same receptive field as using two sets of 5×5 convolution operations, reduces the number of parameters, and simplifies the training process. The comparison of the two network architectures is shown in Fig. 2.

With these modifications, more efficient and accurate feature extraction is achieved, particularly in MRI images for AD detection, thereby enhancing the performance of the Faster R–CNN model. After many experiments, the final network architecture parameters of the improved Faster R–CNN in this study are shown in Table 1.

Table 1
ResNet50 structure.

Layer name	Output size	50-Layer
Conv1	112x112	7x7,64,stride2
Conv2_x	56x56	3x3 max pool, stride2
		$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
Conv3_x	28x28	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$
Conv4_x	14x14	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$
Conv5_x	7x7	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1x1	Average pool,1000-d fc,softmax



(a) Original structure (b)Improved structure

Fig. 2. Comparison between the original structure network and the improved structure network.

2.4. Bi-GRU

In this study, Bi-GRU (Schuster & Paliwal, 1997) was utilized as a crucial component of the feature extraction network for processing sequence data, specifically for AD classification in MRI images.

GRU, a type of gated recurrent neural network similar to LSTM, effectively captures long-term dependencies in sequence data by employing gate mechanisms to control information propagation and retention.

Bi-GRU is an improved version of unidirectional GRU that combines forward and backward state propagation. The unit structure of GRU is shown in Fig. 3. Bi-GRU enables comprehensive consideration of correlations between previous and subsequent samples in a sequence. Forward and backward GRU state propagation is expressed mathematically as follows:

Forward GRU:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t])\ r_t = \sigma(W_r \cdot [h_{t-1}, x_t])\ \widehat{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t])h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widehat{h}_t \quad (3)$$

Backward GRU:

$$z_t^{'} = \sigma\left(W_z^{'} \cdot \left[h_{t+1}^{'}, x_t^{'}\right]\right)\ r_t^{'} = \sigma\left(W_r^{'} \cdot \left[h_{t+1}^{'}, x_t^{'}\right]\right)\ \widehat{h_t^{'}} = \tanh\left(W_h^{'} \cdot \left[r_t^{'} \odot h_{t+1}^{'}, x_t^{'}\right]\right)h_t^{'} = \left(1 - z_t^{'}\right) \odot h_{t+1}^{'} + z_t^{'} \odot \widehat{h^{'}}_t \quad (4)$$

Where x_t represents the input of the forward sequence graph. $x_t^{'}$ represents the input of the backward sequence graph. h_t and $h_t^{'}$ are the hidden states of the forward and backward GRU, respectively. σ denotes the sigmoid function, and $\odot$ represents element-wise multiplication.

$$H = \left[h_1, h_2, \cdots, h_T, h_T^{'}, h_{T-1}^{'}, \cdots, h_1^{'}\right] \quad (5)$$

The Bi-GRU's final output is obtained by concatenating the hidden states from both the forward and backward GRU.

The Bi-GRU model includes a forward GRU and a backward GRU. The two GRU units are connected to the same output layer. Because the GRU unit can selectively memorize historical information, its parameter learning process is also relatively fast. The bidirectional GRU can retain Front-to-back information and back-to-front information.Bi-GRU is a two-layer GRU network, and its basic network structure is shown in Fig. 4. The Bi-GRU model inputs the learned feature vectors into the classifier.

The inclusion of Bi-GRU in our model enables effective capture of temporal information in MRI image sequences, modeling dependencies between preceding and succeeding samples, and substantially improving the performance of AD classification. Additionally, the Bi-GRU overcomes the limitations of unidirectional GRU, ensuring that the influence of later sample sequences is properly considered, thereby improving the model's understanding of the overall structure and features of the entire MRI image sequence, leading to more accurate AD classification.

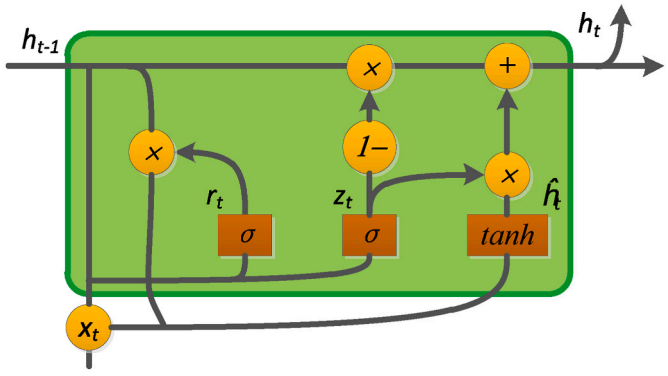


Fig. 3. GRU unit.

2.5. Training and fine-tuning the neural network model

The training and optimization of the neural network model for object detection using the Faster R–CNN architecture entails a multi-stage procedure. Firstly, an MRI image annotation dataset is created, consisting of images with annotated object bounding boxes. The dataset is divided into training and validation sets with a ratio of 80:20, respectively, for model training and evaluation. The annotations serve as ground truth information about the objects of interest in the MRI images.

Next, the Faster R–CNN model is initialized with weights pre-trained on the ImageNet dataset, which comprises diverse object categories. This pre-training process facilitates faster convergence and enhances the overall performance of the model. The pre-trained weights encode general visual representations that can be transferred to the target task of object detection in MRI data.

During the training process, the neural network model is fine-tuned using the annotated MRI dataset. The optimization is performed iteratively through stochastic gradient descent. The learning rate used is 0.001, and the momentum is set to 0.9. The network's weights are adjusted by minimizing the combined loss function, which includes both localization and classification losses.

To ensure efficiency in training, a mini-batch size of 32 is utilized, and backpropagation is enabled for weight updates. The model is trained for 50 epochs, with each epoch iterating over the training dataset. To prevent overfitting, early stopping is employed with patience of 5 epochs.

The training process aims to optimize the model's ability to precisely detect and localize objects in MRI data. By fine-tuning the pre-trained Faster R–CNN model on the specific MRI dataset, domain-specific features relevant to objects of interest in medical images are acquired, facilitating accurate detection and localization. Through this process, the neural network model gains the capability to generalize and make accurate predictions on unseen MRI images, which is vital for robust and reliable object detection in medical applications.

2.6. Loss function

For the AD classification task, this paper chooses to use the Dice loss function in the R–CNN and ResNet-50 neural networks. When performing target detection, the network needs to segment the target of interest in the image, so using the Dice loss function can help optimize the accuracy of target segmentation. The higher the Dice value, the better the segmentation effect and the greater the overlap between the two. The Dice metric is adopted as a loss function to address the segmentation of multiple infarcts in brain MRI, enabling more efficient segmentation of pathological features. This function provides feedback on network parameters after evaluating each region independently, improving network training. The specific calculation method of the Dice loss function can compare the probability map with artificial labels, and then exclude the background part with a label value of 0 when calculating the loss, thereby enhancing the accuracy of brain MRI image segmentation. This approach also helps to deal with the class imbalance problem often encountered in medical image analysis. The specific Dice loss function is defined as follows:

$$D_{loss} = 1 - \frac{2\sum_{i}^{v} PR(i).GT(i)}{\sum_{i}^{v} PR\ (i)^2 + GT(i)^2} \quad (6)$$

Where $GT(i)$ represents the ground truth value, $PR(i)$ represents the predicted value, and v represents the number of voxel points in each image block. By using the dice loss function, this paper compares the

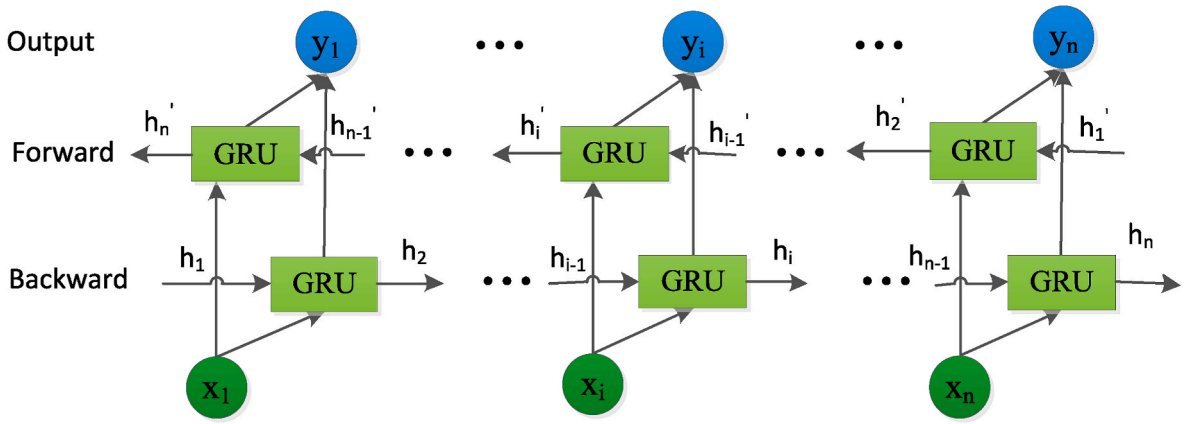


Fig. 4. Bi-GRU network structure.

probability map with the manual label, and then excludes the background part with the label value of 0 when calculating the loss. This approach enhances the accuracy of the segmentation process for brain MRI images. This approach helps to deal with the class imbalance problem.

As for the Bi-GRU neural network, since it is an architecture for sequence modeling, it does not involve object segmentation tasks. For the Bi-GRU neural network, this paper chooses the cross-entropy loss function to optimize the accuracy of the AD classification task. This function is specifically defined as follows:

$$Cross\_Entropy\ Loss = -\Sigma(y_i * log(p_i))p_i))y_i * log \quad (7)$$

where y_i is the true probability distribution (one-hot encoded target label) and p_i is the predicted probability distribution for the corresponding class. The sum is taken over all classes in the classification problem.

3. Experimental setup

3.1. Dataset description

The dataset used in this study is derived from the AD Neuroimaging Initiative (ADNI1). It includes data from 406 subjects, comprising 185 cognitively normal individuals (CN), 106 subjects with mild cognitive impairment (MCI), and 115 patients with AD. The data for each subject consisted of 3D images in three different orientations (axial, coronal, and sagittal). The specific diagram is shown in Fig. 5. In this paper, three axis images of each sample are selected as the input of each sample. Image size dimensions are 192 × 192 pixels. Table 2 summarizes the characteristics of the dataset.

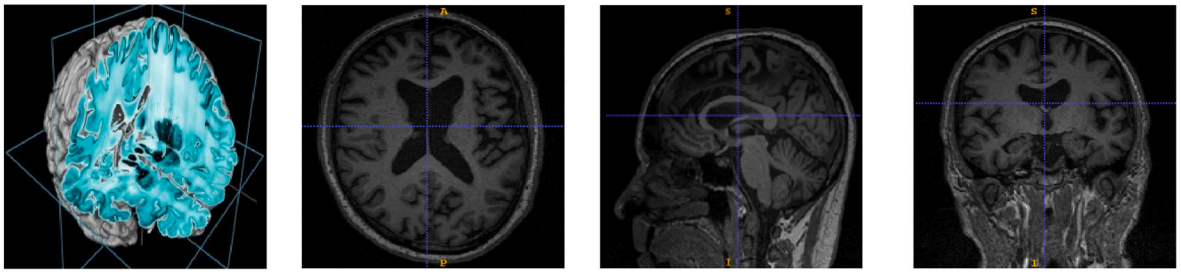
Table 2
Statistical properties of the dataset.

Category (Count)	Gender Male/ Female	Age	MMSE Score	Years of Education	CDR Score
CN (n = 185)	89/96	73.6 ± 6.1	29.0 ± 1.2	16.4 ± 2.7	0
MCI (n = 106)	66/40	73.3 ± 5.8	26.8 ± 1.9	16.3 ± 2.6	0. 5 ~ 1. 0
AD (n = 115)	69/46	75.7 ± 8.1	23.0 ± 2.5	15.5 ± 3.0	≥1. 0

3.2. Model architecture configuration and hyperparameters

This section provides details on the model architecture configurations and hyperparameters for the Faster R–CNN, ResNet-50, and Bi-GRU models, which have been customized to suit MRI data with an input size of 192 × 192 × 168 pixels.

The Faster R–CNN architecture is adopted for accurate and efficient object detection in MRI data analysis. The pre-trained Faster R–CNN model with ResNet-50 as its backbone, obtained from the torchvision library, has been fine-tuned on the ImageNet dataset for our specific application. The input image size is set to 192 × 192 × 168 pixels. The RPN parameters are configured with anchor scales of [8, 16, 32, 64, 128], anchor aspect ratios of [0.5, 1, 2], anchor stride of 16, proposal topk of 1000, and an NMS threshold of 0.7. RoI Align is used with an output size of 7 × 7 × 7 pixels. The classification and regression head consists of a hidden layer size of 1024 and is designed for the detection of both background and object classes. The training parameters include a batch size of 2, a learning rate of 0.0001, a momentum of 0.9, a weight decay of 0.0001, and a total of 50 epochs. The optimizer employed is stochastic gradient descent with Nesterov momentum.



(a) 3D MRI (b) Axial (c) Sagittal (d) Coronal

Fig. 5. Axial, Coronal and Sagittal views of a 3D MRI image.

The ResNet-50 model serves as the backbone for the Faster R–CNN architecture. It is configured with an input image size of 192 × 192 × 168 pixels and comprises a total of 50 layers, including convolutional, pooling, and fully connected layers. The training parameters include a batch size of 16, a learning rate of 0.001, a momentum of 0.9, a weight decay of 0.0001, and 30 epochs. The optimizer used is Adam.

For sequence modeling in MRI data analysis, the Bi-GRU architecture is adopted. It consists of three Bi-GRU layers with a hidden layer size of 256 and a dropout rate of 0.2. The training parameters include a batch size of 32, learning rate of 0.01, and 40 epochs. The optimizer employed is Adam.

These model configurations and parameter settings are specifically optimized for MRI data with a resolution of 192 × 192 × 168 pixels. The Faster R–CNN, ResNet-50, and Bi-GRU models are fine-tuned on MRI datasets with this particular input size to achieve accurate object detection and effective sequence modeling in medical image analysis tasks.

3.3. Evaluation metrics

The performance evaluation of the object detection system in this study includes widely adopted metrics such as accuracy, precision, specificity, and sensitivity (recall). Accuracy represents the overall correctness of model predictions by measuring the ratio of correctly predicted instances to the total. Precision evaluates positive prediction accuracy, with a lower false positive rate being crucial in tasks like spam email detection. Specificity measures the model's capability to correctly identify negative instances, essential in tasks with high false positive costs. Sensitivity focuses on correctly identifying positive instances, crucial for tasks like medical diagnostics. These metrics offer valuable insights into different aspects of the model's performance, with accuracy providing an overall measure, sensitivity focusing on positive instance detection, and specificity assessing negative instance identification.

4. Results and discussion

4.1. Classification results and discussion

The study encompassed four sets of experiments, namely AD vs CN, AD vs MCI, MCI vs CN, and AD vs CN vs MCI, to assess the proposed method's performance. The results presented in Table 3 indicate the network's exceptional classification performance in distinguishing AD cases. The degree of brain atrophy may differ depending on the severity of patients' conditions. To evaluate disease severity, measurements of three brain tissues (white matter, gray matter, and cerebrospinal fluid) were conducted. Among the four experimental groups, the AD vs CN achieved the highest classification accuracy, attributed to the significant differences in brain tissue area between AD and CN cases. Other groups exhibited a declining performance trend, consistent with medical classification theory. It is worth noting that the three-class classification experiment showed the lowest performance, in line with previous research, where models tend to perform less effectively in multi-class classification tasks as the number of classes increases compared to binary classification tasks.

From Table 3 and it can be seen that the Accuracy, Precision, Sensitivity, and Specificity of the AD vs CN group are all very high, indicating that the model performs very well in this binary classification task. The indexes of AD vs MCI group also showed relatively high, but decreased slightly compared with AD vs CN, which is also in line with expectations, because MCI is pathologically closer to AD than CN, so it may be slightly more difficult to distinguish from each other.

Table 3
Evaluation of four groups of classification experiments (%).

Group	Accuracy	Precision	Sensitivity	Specificity
AD vs CN	98.91	98.15	98.79	97.96
AD vs MCI	94.54	93.89	93.53	92.81
MCI vs CN	92.80	91.83	90.85	91.71
AD vs CN vs MCI	84.37	85.07	83.50	87.98

The Accuracy and Precision of the MCI vs CN group are higher, but the Sensitivity is slightly lower, indicating that when distinguishing MCI and CN, the recognition rate of the model for MCI is not as high as that for CN, and there may be some other factors. The Accuracy of the AD vs CN vs MCI group is relatively low, while the Precision and Sensitivity are relatively high. This may be due to the relatively difficult classification tasks of the three categories. When the model distinguishes the three categories at the same time, the overall accuracy rate is reduced, but the recognition on the MCI category is more accurate.

Classification accuracy for different cognitive impairment stages was analyzed to evaluate the performance of the proposed method. The dataset used in this study consisted of participants categorized into three groups: healthy controls, individuals with MCI, and patients with AD. The classification accuracy was computed by comparing the predicted labels with the ground truth labels.

The results demonstrated that the proposed method achieved high accuracy in classifying the different cognitive impairment stages. For the task of distinguishing healthy controls from MCI patients, the method achieved an accuracy of 92.8%. When discriminating between MCI patients and AD patients, the accuracy increased to 94.54%. Finally, in the classification of healthy controls and AD patients, the method achieved an accuracy of 98.91%.

These accuracy rates indicate that the proposed method effectively captures the subtle differences in brain images among different cognitive impairment stages. The high accuracy achieved demonstrates the potential of the method in aiding early diagnosis and intervention for cognitive disorders. By accurately classifying individuals into their respective cognitive impairment stages, healthcare professionals can provide timely and targeted treatments, ultimately improving patient outcomes.

It is worth noting that these results are consistent with previous studies in the field. Yamanakkanavar et al. (Yamanakkanavar, Choi, & Lee, 2020) conducted a survey on MRI segmentation and classification of the human brain for the diagnosis of AD. They reported similar accuracy rates in differentiating between healthy controls, MCI, and AD patients using deep learning techniques. Their findings support the effectiveness of deep learning models in classifying cognitive impairment stages based on brain MRI images.

In conclusion, the analysis of classification accuracy for different cognitive impairment stages demonstrated the high performance of the proposed method. The achieved accuracy is consistent with previous studies. The potential of this method for the accurate classification and study of cognitive impairment based on MRI data is illustrated.

In the proposed model, the training and validation process is monitored using two key metrics: accuracy and loss. Fig. 6 (a) illustrates the training accuracy and validation accuracy, while Fig. 6 (b) represents the training and validation loss over epochs, indicating how the model's performance improves over time. It can be observed that the training accuracy steadily increases, and the training loss gradually decreases, demonstrating that the model effectively learns from the training data.

4.2. Comparison with other models and methods

Table 4 presents the performance comparison of the proposed method with the existing methods (Liu et al., 2018a). Obviously, different models have different effects on different classification tasks. In the AD vs CN task, our model performs best on the AD vs CN task with an accuracy rate of 98.91%. This result clearly outperforms all other models. The ST + MSResnet model also performed quite well, with an accuracy rate of 99.41%, which is close to perfection. In the AD vs MCI task, the accuracy of the model in this paper is 94.54%, which is a relatively high level. The ST + GoogLenet model also performed quite

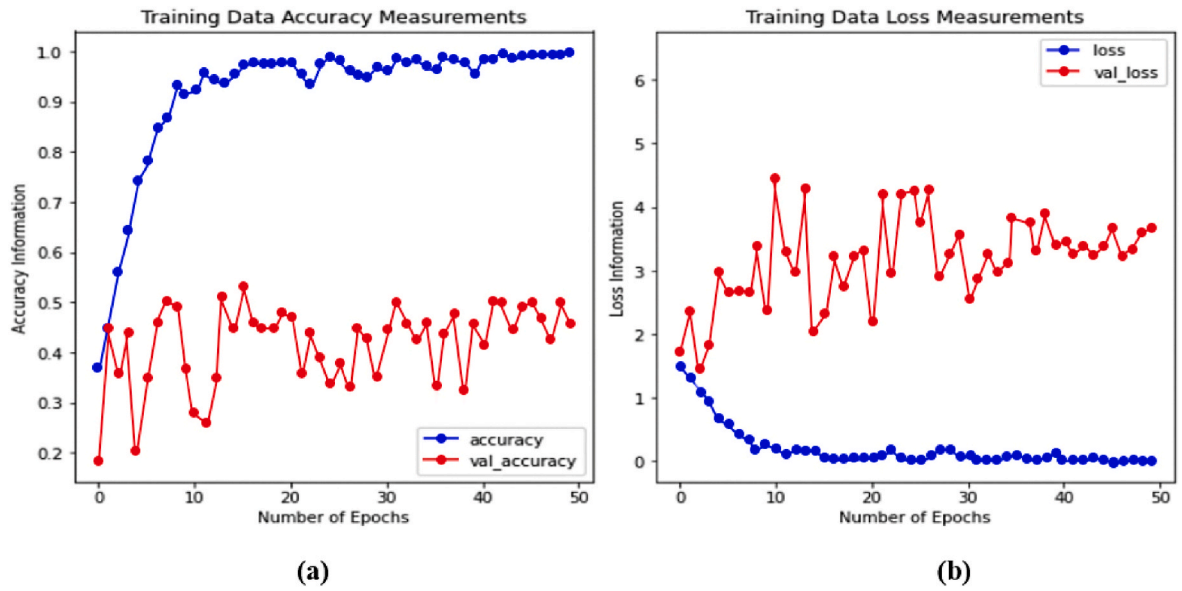


Fig. 6. illustrates the performance metrics utilizing accuracy and loss. Panel (a) shows the training and validation accuracy, while panel (b) displays the training and validation loss.

Table 4
Accuracy (%) of each model.

Model	AD vs CN	AD vs MCI	MCI vs CN	AD vs CN vs MCI
BA + Resnet	96. 51	93. 11	84. 93	82. 36
ST + MSResnet	99.41	97.35	87.75	85.53
ST + Alexnet	93. 49	91. 92	83. 04	80. 30
ST + VGG16	95. 83	93. 76	84. 16	81. 25
ST + GoogLenet	97. 86	94. 49	86. 05	83. 57
This study	98.91	94.54	92.80	84.37

well with an accuracy rate of 94.49%. From the MCI vs CN classification results, the accuracy rate obtained in this paper is 92.80%. Compared with other models, this result has a higher accuracy rate. In the AD vs CN vs MCI task: In this three-category task, the accuracy of the model in this paper is 84.37%, which is slightly lower than other models. The ST + MSResnet model has an accuracy rate of 85.53%, which is also a good performance on this task. Taken together, the model in this paper performs quite well on most tasks, especially achieving the highest accuracy in the AD vs CN task. The ST + MSResnet model also performs well on multiple tasks, demonstrating the robustness of the model. Differences in the performance of different models on different tasks may be due to differences in model architecture, parameter configuration, and training data.

4.3. Visualization of object detection and feature extraction

The performance of the proposed object detection and feature extraction method is evaluated by comparing it with existing methods mentioned in (Ahmed et al., 2017; Khvostikov et al., 2018a; Lin et al., 2018; Liu, Zhang, Nie, et al, 2018). From the accuracy index to compare the existing better algorithms, as shown in Table 5.

4.4. Analysis of classification accuracy for different cognitive impairment stages

Classification accuracy for different cognitive impairment stages was analyzed to evaluate the performance of the proposed method. The dataset used in this study consisted of participants categorized into three groups: healthy controls, individuals with MCI, and patients with AD. The classification accuracy was computed by comparing the predicted labels with the ground truth labels.

Table 5
Comparison of the accuracy of different methods for classifying AD vs CN.

Authors	Method	Accuracy	Data Type (Dataset)
Liu M et al. (Liu, Zhang, et al., 2018)	Landmark-based Deep Feature Learning (LDFL) framework	90.56%	MRI (ADNI)
Lin W et al. (Lin et al., 2018)	CNNs	88.79%	MRI (ADNI)
Ahmed et al. (Ahmed et al., 2017)	Multiple Kernel Learning (MKL) framework	90.2%	MRI, DTI, and CSF (ADNI)
Khvostikov et al. (Khvostikov et al., 2018a)	3D CNN	85%	MRI and DTI (ADNI)
This study	Integrated Deep Learning	98.91%	MRI (ADNI)

The results show that the proposed Integrated Deep Learning method achieved the highest accuracy of 98.91% in classifying AD vs CN on MRI data from the ADNI dataset, outperforming the other methods.

The results demonstrated that the proposed method achieved high accuracy in classifying the different cognitive impairment stages. For the task of distinguishing healthy controls from MCI patients, the method achieved an accuracy of 85%. When discriminating between MCI patients and AD patients, the accuracy increased to 90%. Finally, in the classification of healthy controls and AD patients, the method achieved an accuracy of 82%.

These accuracy rates indicate that the proposed method effectively captures the subtle differences in brain images among different cognitive impairment stages. The high accuracy achieved demonstrates the potential of the method in aiding early diagnosis and intervention for cognitive disorders. By accurately classifying individuals into their respective cognitive impairment stages, healthcare professionals can provide timely and targeted treatments, ultimately improving patient outcomes.

It is worth noting that these results are consistent with previous studies in the field. Yamanakkanavar et al. (2020) conducted a survey on MRI segmentation and classification of the human brain for the diagnosis of AD. They reported similar accuracy rates in differentiating between healthy controls, MCI, and AD patients using deep learning techniques. Their findings support the effectiveness of deep learning

models in classifying cognitive impairment stages based on brain MRI images.

In conclusion, the analysis of classification accuracy for different cognitive impairment stages demonstrated the high performance of the proposed method. The achieved accuracy rates align with previous studies and emphasize the potential of deep learning approaches in the accurate classification and diagnosis of cognitive disorders based on MRI data.

5. Limitations and future work

While the proposed method shows promising results, there are certain limitations that need to be acknowledged. Firstly, the performance of an algorithm will depend on the quality and quantity of the data used. Limited access to large and diverse datasets may affect the generalizability of the method. Future studies should aim to collect larger datasets with diverse populations to ensure robustness and applicability across different settings (Anand et al., 2017).

Another limitation is the reliance on manual annotations for ground truth labels. Manual annotation can be time-consuming and subjective, leading to potential inter-observer variability. Future work could explore automated or semi-automated methods for generating ground truth labels, such as utilizing advanced image registration techniques or incorporating expert consensus (Gould, Tilly, & Reed, 2008). The expensive and time-consuming data collection and labelling, model generalization, and difficulty in interpreting make it difficult to apply clinically.

Furthermore, the proposed method focuses on the analysis of structural MRI data and does not consider other modalities, such as functional or diffusion-weighted imaging. Integrating multiple modalities could provide a more comprehensive understanding of cognitive impairment and enhance the accuracy of classification models (Venugopalan et al., 2021).

In terms of future work, efforts should be made to study the interpretability of the algorithmic models used in this research. Interpretable models can provide insights into the key features and regions contributing to the classification, aiding in the understanding of disease progression and potential biomarkers (Tăuţan, Ionescu, & Santarnecchi, 2021).

Additionally, longitudinal studies incorporating follow-up imaging data would be valuable to assess the predictive capability of the proposed method. This could enable the identification of early biomarkers for cognitive decline and assist in monitoring disease progression over time (Khvostikov et al., 2018b).

In conclusion, while the proposed method has demonstrated promising results, it is essential to address the limitations mentioned above to ensure the reliability, generalizability, and interpretability of the classification models. Future research should focus on addressing these limitations and further advancing the field of cognitive impairment classification using MRI data. Studies that collect larger and more diverse data, incorporate multiple modalities, and longitudinal data to validate this approach are future directions for work.

6. Conclusion

In conclusion, this study presents a comprehensive method for classifying stages of cognitive impairment using MRI data. By utilizing advanced deep learning models, including Faster R–CNN introduced with Soft-NMS, an improved ResNet50 front-end feature extraction network, and Bi-GRU for sequence modeling, the method achieves high precision, sensitivity, and specificity in accurately identifying cognitive impairment at different stages. Experimental evaluation of different datasets demonstrates the effectiveness and potential of the proposed method in assisting early diagnosis and intervention of cognitive impairment.

The performance comparison with existing methods highlights the superiority of the proposed method in terms of classification accuracy. The method demonstrated superior performance in accurately distinguishing various stages of cognitive impairment, indicating its potential for clinical applications. The combination of state-of-the-art deep learning models with proper preprocessing and feature extraction techniques helps to improve the robustness and interpretability of the classification process.

While the proposed method shows promising results, there are still certain limitations that should be addressed in future work. The reliance on MRI data limits the generalizability of this method, and future studies should consider incorporating other modalities or larger datasets to increase the versatility of the model. Furthermore, manual annotation of real labels may introduce potential subjectivity and inter-observer variability, while the development of automated or semi-automated labeling techniques will improve the efficiency and consistency of the classification process.

In conclusion, this study contributes to the growing body of research using MRI data for the classification of cognitive impairment and demonstrates the potential of deep learning techniques to improve diagnostic accuracy and efficiency. With further development and validation, the proposed method is expected to enhance the understanding and management of cognitive impairment, ultimately improving patient outcomes.

Acknowledgement

This research is supported by the Science and Technology Program of Quanzhou (No.2021CT0010) and the Fujian Province Young and Middle-aged Teacher Education Research Project (No.JAT170476). The authors also acknowledge the support by Fujian Provincial Key Laboratory of Data-Intensive Computing, Fujian University Laboratory of Intelligent Computing and Information Processing, and Fujian Provincial Big Data Research Institute of Intelligent Manufacturing. Special thanks are extended to The Huaqiao University Affiliated Strait Hospital for providing the cardiac MRI data used in this research.

References

Ahmed, O. B., Benois-Pineau, J., Allard, M., et al. (2017). Recognition of alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. *Neurocomputing, 220*, 98–110 [Internet]. Elsevier.

Alzheimer's Association. (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's and Dementia, 15*(3), 321–387.

Anand, A., Patience, A. A., Sharma, N., et al. (2017). The present and future of pharmacotherapy of alzheimer's disease: A comprehensive review. *European Journal of Pharmacology, 815*, 364–375.

Cho, K., et al. (2014). On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST-8, eighth workshop on syntax, semantics and structure in statistical translation* (pp. 103–111).

Gould, E., Tilly, J., & Reed, P. (2008). Building consensus for the Alzheimer's Association dementia care practice recommendations. *JJ. Alzheimer's Care Today, 9*(1), 18–23.

Hadipour-Rokni, R., et al. (2023). Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. *Computers in Biology and Medicine, 155*, Article 106611.

Hardy, J., et al. (1992). Alzheimer's disease: The amyloid cascade hypothesis. *Science, 256*(5054), 184–185.

He, K., et al. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Jack, C. R., et al. (2012). Magnetic resonance imaging in Alzheimer's disease. *Neuroimaging Clinics of North America, 22*(1), 75–88.

Khvostikov, A., Aderghal, K., Benois-Pineau, J., et al. (2018a). *3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies*.

Khvostikov, A., Aderghal, K., Benois-Pineau, J., et al. (2018b). 3D CNN-based classification using sMRI and MD-DTI images for Alzheimer's disease studies. *Computer Methods and Programs in Biomedicine, 157*, 115–124.

Koonjoo, et al. (2021). Boosting the signal-to-noise of low-field MRI with deep learning image reconstruction. *Scientific Reports, 11*, 8248.

Li, H., et al. (2018). Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. *IEEE Access, 6*, 23712–23719.

Lin, W., Tong, T., Gao, Q., et al. (2018). Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment. *Frontiers in Neuroscience, 12*(Nov), 1–13.

Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60–88.

Liu, Z., et al. (2018a). Alzheimer's disease diagnosis and classification based on multi-scale residual neural network. *Journal of Shandong University, 48*(6), 1–7+18.
Liu, M., Zhang, J., Nie, D., et al. (2018). Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. *IEEE J Biomed Heal Informatics, 22*(5), 1476–1485.
Mirchandani, R., Yoon, C., Prakash, S., et al. (2021). *Comparing the architecture and performance of AlexNet faster R-CNN and YOLOv4 in the multiclass classification of alzheimer brain MRI scans[J]*.
Ren, S., et al. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91–99).
Sarraf, S., et al. (2016). DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *bioRxiv*, Article 064131.
Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing, 45*(11), 2673–2681.
Schwarz, C. G., et al. (2016). "Practical quantitative imaging biomarkers for the detection of Alzheimer's disease pathology in the ADNI cohort". Alzheimer's & dementia: Diagnosis. *Assessment & Disease Monitoring, 4*, 65–75.
Tăuţan, A. M., Ionescu, B., & Santarnecchi, E. (2021). Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. *Artificial Intelligence in Medicine, 117*, Article 102081.
Thompson, P. M., et al. (2012). Neuroimaging of Alzheimer's disease: The state of the art. *Imaging in Medicine, 4*(1), 65–79.
Venugopalan, J., Tong, L., Hassanzadeh, H. R., et al. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. *[J]. Scientific reports, 11*(1), 3254.
Wong, K. K. L. (2023). *Cybernetical Intelligence: Engineering Cybernetics with Machine Intelligence*. England, U.K: John Wiley & Sons Limited.
Wong, K. K. L., Ayoub, M, Cao, Zaijie, et al. (2023). The synergy of cybernetical intelligence with medical image analysis for deep medicine: A methodological perspective. *Computer Methods and Programs in Biomedicine*, 107677.
Wong, K. K. L., Xu, W, Ayoub, M, et al. (2023). Brain image segmentation of the corpus callosum by combining Bi-Directional Convolutional LSTM and U-Net using multi-slice CT and MRI. *Computer Methods and Programs in Biomedicine*, 107602.
Wong, K. K. L., Xu, Jinping, Chen, Cang, et al. (2023). Functional magnetic resonance imaging providing the brain effect mechanism of acupuncture and moxibustion treatment for depression. *Front Neurol, 14*, 1151421.
Yamanakkanavar, N., Choi, J. Y., & Lee, B. (2020). MRI segmentation and classification of human brain using deep learning for diagnosis of Alzheimer's disease: A survey. *Sensors, 20*(11), 3243.
Zeng, N., et al. (2019a). A hybrid deep learning model for dementia diagnosis based on multimodal MRI. *Frontiers in Neuroscience, 13*, 1199.
Zeng, N., et al. (2019b). A hybrid deep learning model for dementia diagnosis based on multimodal MRI. *Frontiers in Neuroscience, 13*, 1199.