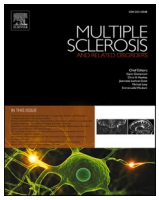
Multiple Sclerosis and Related Disorders 87 (2024) 105642

Contents lists available at ScienceDirect

Multiple Sclerosis and Related Disorders

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



Original article

Deep learning for discrimination of active and inactive lesions in multiple sclerosis using non-contrast FLAIR MRI: A multicenter study

AmirAbbas Amini [a], Azin Shayganfar [b], Zahra Amini [a], Leila Ostovar [b], Somayeh HajiAhmadi [b], Navid Chitsaz [b], Masoud Rabbani [b], Raheleh Kafieh [c,*]

[a] School of Advanced Technologies in Medicine, Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[b] Department of Radiology, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran
[c] Department of Engineering, Durham University, Durham, UK

ARTICLE INFO

Keywords:
Multiple sclerosis (MS)
Non-contrast MRI
Deep learning

ABSTRACT

Background: Within the domain of multiple sclerosis (MS), the precise discrimination between active and inactive lesions bears immense significance. Active lesions are enhanced on T1-weighted MRI images after administration of gadolinium-based contrast agents, which brings about associated complexities. This study investigates the potential of deep learning to differentiate between active and inactive lesions in MS using non-contrast FLAIR-type MRI data, presenting a non-invasive alternative to conventional gadolinium-based MRI methods.

Methods: The dataset encompasses 9097 lesion images collected from 130 MS patients across four distinct imaging centers, with post-contrast T1-weighted images as the benchmark reference. We initially identified and labeled the lesions and carefully selected corresponding regions of interest (ROIs). These ROIs were employed as inputs for a convolutional neural network (CNN) to predict lesion status. Also, transfer learning was utilized, incorporating 12 pre-trained CNN models. Subsequently, an ensemble technique was applied to 3 of best models, followed by a systematic comparison of the results.

Results: Through a 5-fold cross-validation, our custom designed network exhibited an average accuracy of 85 %, a sensitivity of 95 %, a specificity of 75 %, and an AUC value of 0.90. Among the pre-trained models, ResNet50 emerged as the most effective, achieving a specificity of 58 %, an accuracy of 75 %, a sensitivity of 91 %, and an AUC value of 0.81. Our comprehensive evaluations encompassed the receiver operating characteristic curve, precision-recall curve, and confusion matrix analyses.

Conclusion: The findings underscore the efficacy of the proposed CNN, trained on FLAIR MRI data ROIs, in accurately discerning active and inactive lesions without reliance on contrast agents. Our multicenter study of 130 patients with diverse imaging devices outperforms the other single-center studies, achieving superior sensitivity and specificity. Unlike studies using multiple modalities, our exclusive use of FLAIR images streamlines the process, and our streamlined approach, excluding conventional pre-processing, demonstrates efficiency. The external validation conducted on diverse datasets, coupled with the analysis of dilated masks, underscores the adaptability and efficacy of our custom CNN model in discerning between active and inactive lesions.

1. Introduction

Multiple sclerosis (MS) is an autoimmune disease impacting the central nervous system, affecting over 2.8 million individuals worldwide (Walton et al., 2020). A key pathological feature of MS is the presence of lesions on MRI images with fluid-attenuated inversion recovery (FLAIR), although not all lesions are active. Active lesions are discernible on T1-weighted MRI images post-injection of gadolinium-based contrast agents (GBCAs), enhancing visibility. However, concerns arise as studies (Burke et al., 2016; Hu et al., 2016; Beomonte Zobel et al., 2016; McDonald et al., 2017; Fraum et al., 2017) demonstrate GBCAs' accumulation in bones, brain, and other body parts after multiple injections, prompting regulatory agencies such as the US Food and Drug Administration (2015), the European Medicines Agency (Agency, 2017), the Consortium of Multiple Sclerosis Centers (Traboulsee et al., 2017), and the International Society of Magnetic Resonance in Medicine to advise

* Corresponding author at: Department of Engineering, Durham University, Durham, UK.
E-mail address: raheleh.kafieh@durham.ac.uk (R. Kafieh).

https://doi.org/10.1016/j.msard.2024.105642
Received 10 November 2023; Received in revised form 18 February 2024; Accepted 20 April 2024
Available online 21 April 2024

limiting GBCA use to essential cases. Moreover, GBCA injection poses risks, such as nephrogenic systemic fibrosis in renal-compromised patients (Grobner, 2006).

Hence, diagnosing active MS lesions from MRI scans sans GBCAs is a crucial research focus. Previous approaches, including logistic regression, chemical exchange saturation transfer techniques (Shinohara et al., 2012; Lin et al., 2018) and texture analysis (Ardakani et al., 2017) have limitations like single-center data, small sample sizes, and manual feature identification. Artificial intelligence, particularly Deep Learning (DL), emerges as a promising alternative.

DL a subset of machine learning, eliminates manual feature identification, making it valuable in image processing and computer vision (Al-Saffar et al., 2017). Notably, DL has been employed to identify active MS lesions with reduced contrast agent usage. Gong et al. (2018) reduced contrast agent dosage using DL methods but still required injection (Gong et al., 2018) Similarly, another study utilized a convolutional neural network (CNN) with VGG16 architecture but necessitated T1, T2, and FLAIR MRI scans (Narayana et al., 2020). Freire et al. (2020) focused on FLAIR image texture features but lacked data volume and independent neuroradiologist validation, among other drawbacks (Freire et al., 2020). Caba et al. (2023) developed a machine learning approach using non-contrast T1 and T2 images for automatic diagnosis of acute MS lesions, albeit requiring both sequence types (Caba et al., 2023). One of the disadvantages of their study is the need for 2 types of sequences T1 and T2.

Common weaknesses across these studies include multi-step preprocessing, time complexity, and potential error rate escalation. In light of the previously mentioned shortcomings in existing research, our work contributes significantly in several key aspects. Our primary objective is to evaluate the effectiveness of a meticulously designed DL network in accurately predicting active lesions within multiple sclerosis patients, while avoiding the need for contrast agents. By leveraging advancements in convolutional neural networks (CNNs) and their applications in medical image processing, we aim to demonstrate the utility of DL in this context. In addition to our custom-designed CNN, we conduct a comprehensive comparison of 12 distinct transfer learning network models to predict active lesions. This approach enables us to gain insights into the most effective model for the task at hand. We address the challenges introduced by multicenter studies, which often result in data inhomogeneities and various artifacts. Our research explores how DL techniques can effectively manage these complexities. Our study exclusively relies on FLAIR MR sequence images for lesion classification, simplifying the diagnostic process by eliminating the need for multiple image types. Unlike similar studies that involve labor-intensive preprocessing steps, such as brain extraction and noise reduction, we streamline the diagnostic process by omitting these procedures. This approach aims to reduce complexity and enhance efficiency.

It's important to note that our study focuses on classifying lesions as active or inactive, rather than quantifying their size and dimensions. This distinction positions our work as a classification problem rather than a segmentation task.

2. Materials and methods

In this section, we will commence by presenting the patient data included in the study, detailing the MRI acquisition process, and elaborating the pre-processing procedures that were employed on the images. Subsequently, we will detail the input data utilized by our network, explain the architecture of our proposed DL network, delineate the network training methodology, and clarify the criteria employed for result evaluation. A visual representation of the proposed method is illustrated in Fig. 1 for clarity.

2.1. Subjects

This study exclusively enrolled patients diagnosed with relapsing-remitting multiple sclerosis (RRMS). Data collection spanned the years 2019 to 2022, encompassing information from four MRI imaging centers situated in Isfahan City (Including Kashani, Alzahra, Askarieh, and Shafa centers). The clinical trial involved patients aged between 17 and 56 years, participating in phase III. Strict measures were undertaken to ensure patient anonymity, and all participating centers obtained institutional review board approval for patient imaging (Ethics code: IR.MUI.MED.REC.1400.474). In addition, written informed consent was diligently secured from all patients, further solidifying the ethical foundation of this study. Our institutional review board meticulously approved the MRI data analysis, aligning with the regulations set forth by the Health Insurance Portability and Accountability Act (HIPAA).

A total of 345 patients initially participated in this study, with varying numbers stemming from each imaging center (95 from the Alzahra center, 153 from the Kashani center, 30 from the Askarieh center, and 67 from the Shafa center). Data from 34 patients had to be excluded due to image quality concerns. Among the remaining 311 patients, 197 displayed no active lesions. To ensure a more balanced learning approach for the network, data from the majority of these patients were excluded, and only the data from 16 were retained for study. Consequently, the final cohort under investigation consisted of 130 patients, with 114 of them manifesting at least one active lesion and the remaining 16 displaying no active lesions. In total, a comprehensive analysis identified a grand total of 9097 lesions among these patients, comprising 966 active lesions and 8131 inactive lesions.

For a visual representation of the patient distribution and lesion categories across different imaging centers, please refer to Fig. 2. This figure delineates the total patient count, the number of patients with excluded low-quality data, and the distribution of patients with active and inactive lesions within the context of the various imaging centers.

2.2. Data acquisition

The MRI systems employed in this study were sourced from four distinct imaging centers and comprised a combination of field strengths. Specifically, three 1.5-Tesla MRI systems manufactured by Siemens Medical Systems, Germany, were utilized (in Kashani, Askarieh, and Shafa centers), alongside one 1.5-Tesla MRI system crafted by Philips

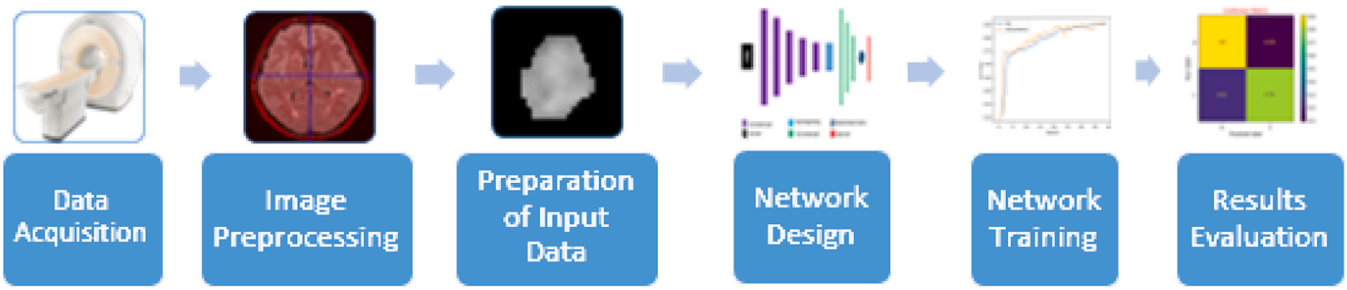


Fig. 1. Pipeline of the proposed method.

Fig. 2. Dataset flowchart (Total number of patients).

Table 1
Data acquisition devices and imaging protocols.

Imaging Center	Company	FLAIR images			Post-contrast T1-weighted images		
		TR (ms)	TE (ms)	Voxel dimensions (mm)	TR (ms)	TE (ms)	Voxel dimensions (mm)
Alzahra Hospital	Philips	4800	340	0.72 × 0.72 × 3	550	15	0.45 × 0.45 × 6.9
Kashani Hospital	Siemens	8000	90	0.90 × 0.90 × 6.5	400	10	0.90 × 0.90 × 7.15
Askarieh Hospital	Siemens	9000	90	1.04 × 1.04 × 5	550	15	0.73 × 0.73 × 7.8
Shafa Medical Imaging Center	Siemens	9000	90	0.45 × 0.45 × 8.45	485	12	0.45 × 0.45 × 8.45

Healthcare (in Alzahra center).

The MRI imaging protocols for FLAIR images encompassed a diversity of techniques. One center (Alzahra Hospital Center) employed both 3D and 2D imaging approaches, while the remaining centers exclusively utilized 2D imaging methods for FLAIR sequences. In the case of contrast-enhanced T1 images, all centers employed the 2D imaging modality. A comprehensive summary of the data acquisition devices and imaging protocols for each center is provided in Table 1 for reference.

2.3. Image preprocessing

The preprocessing step includes brain extraction, noise reduction, registration, intensity normalization, and anisotropic diffusion filtering, as described in (Sajja et al., 2006; Datta et al., 2006). In this study, unlike other similar studies (Caba et al., 2023; Sajja et al., 2006), pre-processing operations were not used

Instead, during the active lesion identification phase, we opted to register FLAIR scans with post-contrast T1-weighted images for individual patients, a choice driven by the need to address the uneven availability of these two types of data for some patients.

2.4. Input data

The input data to the network were only FLAIR images. The images were masked with their respective manually segmented lesion masks. Indeed, the images underwent masking with lesion masks that were dilated to various dimensions, including 3, 6, 10, and 20 voxels in size. In addition, unmasked images were used as another mode of input data to compare the output results of the network for these 5 different situations (unmasked images and those masked with different sizes of dilated masks). In each condition, a 60 × 60 square was cut around each lesion as a region of interest (ROI). Lesions with a length or width greater than 60 voxels (which constitute less than 3 % of the total data) were also shrunk to fit into a 60 × 60 square. Some examples of these ROIs are shown in Fig. 3. Then, the ROIs were augmented by rotation, reflection, and magnification operations to avoid overfitting during network training (Dietterich, 1995; Ying, 2019).

Two expert radiologists identified active and inactive MS lesions on FLAIR images with the help of post-contrast T1-weighted images. Their diagnoses were labeled as active or inactive lesions and will be utilized as the correct reference.

This study only used data from the relapsing-remitting MS phenotype, which is the most common form of MS. Because lesion increase is less observed in other MS phenotypes.

2.5. Network description and training

In this investigation, we initially devised a custom CNN model, followed by a comprehensive comparative analysis with 12 transfer learning models.

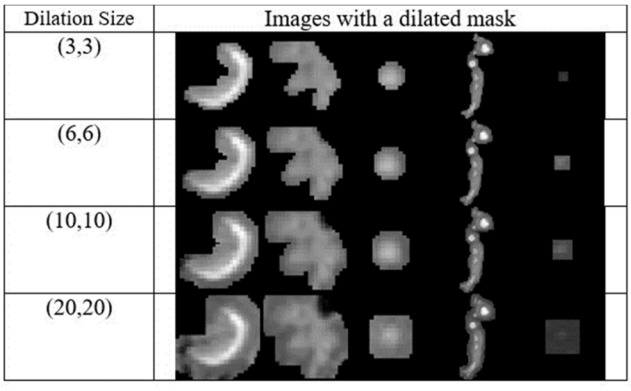


Fig. 3. Some examples of lesion ROIs with different dilation sizes.

2.5.1. CNN model

The DL model used in this study was a convolutional neural network (CNN). The reason for using CNN is its high accuracy in classifying and identifying images. This network is shown in Fig. 4 and was designed to include 5 convolutional blocks and a Global Average pooling block and 3 blocks of fully connected layers, followed by a sigmoid activation function at the output. In each convolution block, a certain number of convolution layers were allocated and each one has a modified Rectified Linear Unit (ReLU) activation function. A certain number of filters are assigned to the convolution layers in each block, starting with 8 layers for the convolution layers in the first convolution block, and there are 16, 32, 64, and 128 filters in the next blocks, respectively, Kernels were selected in each layer of size 3. In this network, instead of the normal mode of using max pooling layers, a Global Average pooling layer was used (Al-Sabaawi et al., 2021). In the fully connected layers section, 3 layers were placed, including an input layer with 128 nodes, a hidden layer consisting of 4 nodes, and an output layer with 1 node with a sigmoid activation function (Ajit et al., 2020).

In the training phase, the Adam optimizer function was used due to its fast convergence and weight-dependent learning rate. Binary cross-entropy was also used as the loss function of the model along with the ReLU activation function for all except the last layer, which had sigmoid activation. In the second step, the last convolution block and dense layers were trained on the image features. In this training stage, the operational optimizer of stochastic descent along the small learning coefficient and movement amount was used. This configuration limited major changes in the weights to preserve previously learned features (Kingma and Adam, 2014). The activation and loss functions remained the same as in the first step. In each run, the network was trained for 50 epochs (one epoch = one complete iteration of the training set). In each epoch, the network weights were updated using error backpropagation between the network output and the true value. An initial learning rate of 0.0001 was used. Dropout was also used to prevent overfitting (Ying, 2019).

2.5.2. Transfer learning methods

The transfer learning technique uses the knowledge of the model that has been previously trained to perform the primary task to solve another task. With the transfer learning technique, the weights that a network has learned for the primary task are used to perform another task. For example, one could imagine using a classification model trained on ImageNet to start learning to recognize MS lesions. Using this technique, the model parameters start with approximately appropriate initial values and are directed to the new task with some small changes. These methods are called pre-trained methods. In this study, 12 pre-trained networks including ResNet50, ResNet50V2, ResNet101, VGG16, VGG19, DenseNet121, EfficientNetB0, EfficientNetV2M, MobileNet, NASNet, InceptionV3, and Xception are used for data classification. In each of the pre-trained methods, only the last layer of the reciprocal

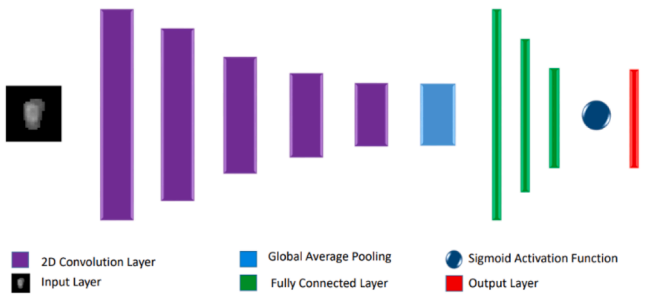


Fig. 4. CNN architecture for MS classification.

Table 2
The hyperparameters set in each pre-trained model.

Model	Number of fully connected	Optimizer function	Learning Rate	Activation function	Loss function	Dropout rate	Epochs
DenseNet121	5	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	65
ResNet101	5	Adam	0.0001	Sigmoid	Binary cross-entropy	0.3	33
ResNet50	5	Adam	0.0001	Sigmoid	Binary cross-entropy	0.3	30
ResNet50V2	6	Adam	0.0001	Sigmoid	Binary cross-entropy	0.3	32
VGG16	6	Adam	0.0001	Sigmoid	Binary cross-entropy	0.3	50
VGG19	6	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	75
Efficient NetB0	6	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	65
Efficient NetV2M	6	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	80
Inception-v3	6	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	30
Xception	6	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	85
MobileNet	5	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	42
NASNet	5	Adam	0.0001	sigmoid	Binary cross-entropy	0.3	40

neural networks is carefully adjusted. In fact, in this method, after using each of the models, a classifier was used. The hyperparameters set in each model are listed in Table 2.

2.6. Evaluation criteria

A fivefold cross-validation procedure was implemented to evaluate the stability of the DL model and its results. The data was divided into two main sets: 80 % of the data set was used for training and 20 % for validation and testing, in each iteration.

To evaluate the classification of active and inactive lesions, parameters of accuracy, sensitivity, specificity, and precision were calculated according to the formula mentioned in (1)–(4). Additionally, the area under the receiver operating characteristic curve (AUC) is computed.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Examples}} \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positive}} \quad (2)$$

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (3)$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positive}} \quad (4)$$

The implementation was done using the Keras Python library (version 2.11.0) and TensorFlow (version 2.11.0) (Abadi et al., 2016).

2.7. Ensemble technique

Ensemble learning involves training multiple models on the same dataset and aggregating their results to enhance accuracy and minimize model variance. One common ensemble technique is average ensemble, wherein the average probability score derived from each basic model is computed for each test data point. Subsequently, statistical metrics are computed based on these probability scores.

3. Results

A summary of the characteristics of the patients examined in this study is shown in Table 3.

Table 4 shows the results of accuracy for the dilation of masks in different sizes by using the CNN method.

Table 3
Summary of the characteristics of the patients.

Parameter	Value
Age (y)	17–56
F/M	98/32

Table 4
The results of accuracy for the dilation of masks in different sizes.

Dilation Size	Accuracy (%)
(3,3)	77
(6,6)	85
(10,10)	84
(20,20)	78
Unmasked images	76

Based on the findings of Table 4, the highest accuracy is obtained by using masked images with dilation sizes of 6 and 10, so we will continue our investigations on dilated data with size 6.

The results of fivefold cross-validation for the test data are summarized in Table 5 such as average ± standard deviation. The average ± standard deviation accuracy in predicting active and inactive lesions using the CNN model is 85 % ± 0.44 %.

The results of the average ensemble technique implemented on 3 of best models (CNN, ResNet50, VGG19) for the test data are reported in Table 6.

Table 5
Comparison of accuracy, sensitivity, specificity, and AUC of all 13 deep neural network.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
CNN	85 ± 0.44	95 ± 2.79	75 ± 2.49	0.90 ± 0.01
DenseNet121	74 ± 0.58	82 ± 2.34	65 ± 2.67	0.81 ± 0.01
ResNet101	74 ± 0.74	85 ± 2.82	63 ± 2.13	0.79 ± 0.02
ResNet50	75 ± 0.65	91 ± 2.91	58 ± 2.79	0.81 ± 0.01
ResNet50V2	62 ± 0.31	68 ± 3.13	56 ± 3.07	0.64 ± 0.03
VGG16	73 ± 0.66	87 ± 2.81	59 ± 2.69	0.68 ± 0.01
VGG19	75 ± 0.37	88 ± 2.82	62 ± 2.74	0.81 ± 0.01
Efficient NetB0	74 ± 0.53	94 ± 1.88	54 ± 1.96	0.81 ± 0.03
Efficient NetV2M	72 ± 0.34	98 ± 1.19	45 ± 1.89	0.78 ± 0.03
Inception-v3	64 ± 1.65	70 ± 2.32	58 ± 2.54	0.67 ± 0.01
Xception	66 ± 0.48	74 ± 2.80	57 ± 2.43	0.71 ± 0.02
MobileNet	65 ± 0.42	73 ± 2.67	57 ± 2.81	0.67 ± 0.03
NASNet	62 ± 0.58	72 ± 2.78	51 ± 2.75	0.66 ± 0.01

Table 6
The result of average ensemble technique.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Average Ensemble	83	94	71	0.89

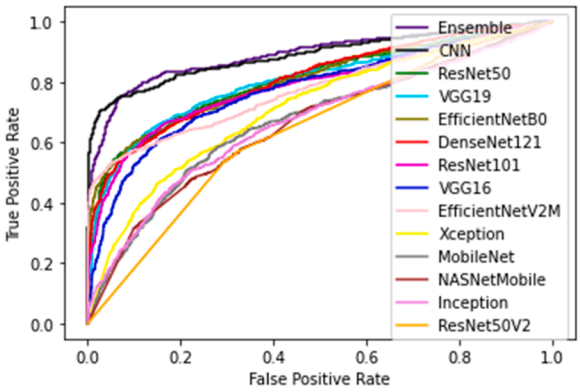


Fig. 5. The ROC curve of all 13 deep neural networks and ensemble technique.

By utilizing the receiver operating characteristic (ROC) curve, the true positive rate (also known as sensitivity) is illustrated against the false positive rate. Fig. 5 displays the ROC curves of the 13 individual models alongside the ROC curve of the ensemble method. Another approach is through the precision-recall curve, which demonstrates the precision rate relative to the recall rate. Recall is synonymous with sensitivity, and measures are defined in formulas 2 and 3. Fig. 6 exhibits the precision-recall curves of all 13 models, as well as the precision-recall curve derived from their ensemble technique.

Fig. 7 displays the confusion matrices of the three top-performing models and the ensemble method to visualize the detection rate of samples classified as active and inactive lesions. (Note: In the matrices, zero represents inactive lesions, while one represents active lesions.)

For external validation, we incorporated all data from Alzahra Imaging Center, totaling 38 patients, as the test set. These data were excluded from network training and the training data comprised samples from other imaging centers. This division allowed us to effectively train and test our custom CNN. The results of this validation are shown in Table 7.

4. Discussion

Post-contrast T1-weighted MR images are pivotal in detecting active MS lesions, but their reliance on gadolinium-based contrast agents (GBCAs) poses potential patient safety concerns and financial burdens. This study explores an AI-driven alternative for active lesion detection, employing a custom CNN model and comparing results with 12 pre-trained CNN models.

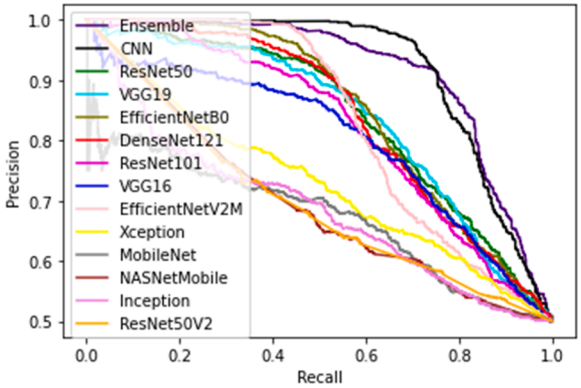


Fig. 6. The precision-recall curve of all 13 deep neural networks and ensemble technique.

Our custom CNN model achieved an average accuracy of 85 %, with 95 % sensitivity and 75 % specificity. Among the compared models, ResNet50, and VGG19 performed well in terms of accuracy, averaging 75 %, while the Efficient NetV2M exhibited the highest sensitivity (98 %). For specificity, our custom CNN led the way at 75 %. In terms of AUC values, our custom CNN averaged 0.90, and the ResNet50, VGG19, Efficient NetB0, and DenseNet121 models followed closely at 0.81, demonstrating comparable performance to previous studies. This research underlines the potential of AI-based methods as a safer and cost-effective means of identifying active MS lesions without the need for GBCAs. Compared to Ardakani et al. (2017), which used a single-center, small sample of 16 subjects and achieved 70 % sensitivity and specificity, our multicenter study with 130 patients and various imaging devices significantly improved both sensitivity and specificity. Compared to conventional methods, DL efficiently learns lesion features from data without manual extraction and handles data variations and artifacts in multicenter studies. This robustness makes DL a perfect fit for our diverse dataset from different MRI devices. In other DL studies, Freire et al. (2020) achieved active lesion detection but still relied on contrast agents, and their study had limited patient numbers and a single-center dataset, unlike ours. Narayana et al. used DL with various MRI modalities but had lower accuracy, sensitivity, and specificity (75 %, 78 %, and 73 %). Our method outperforms both of these studies, suggesting that pre-trained networks may not surpass 75 % accuracy.

This study aimed to streamline the process by minimizing the use of multiple imaging modalities, as incorporating them leads to increased data volume and more processing operations. Registering these modalities is time-intensive and can introduce errors. Unlike (Caba et al., 2023), which employed three modalities (T1-weighted, T2-weighted, and FLAIR images), this study exclusively utilized FLAIR images, resulting in superior outcomes.

Unlike comparable studies, we abstained from conventional pre-processing techniques, streamlining the procedure while minimizing complexity and error risk. Traditional steps like noise reduction may inadvertently erase small lesions or mishandle brain extraction, leading to inaccuracies. Additionally, they are time-consuming and add complexity. In our study, we employed a singular pre-processing step, registration, aligning FLAIR images with post-contrast T1-weighted images to enhance expert lesion detection. Crucially, these registered images remained external to our DL network. Instead, by discerningly selecting ROIs and leveraging DL, we circumvented the need for data pre-processing.

Conversely, by choosing appropriate ROIs, we managed to streamline the CNN network, significantly reducing the data requirement. This was substantiated by our training and evaluation results, which demonstrated the network's effective capacity and its aptitude for successful training. Consequently, the ROI selection strategy enabled us to achieve quicker diagnoses and superior outcomes, as evidenced by the favorable statistical parameters, all accomplished with a lightweight CNN and a modest dataset.

In external validation, we diversified the test dataset by including data from Alzahra Hospital, acquired from a Philips MRI system with 3D scans. Our custom CNN model achieved an 81 % accuracy rate on this test data, highlighting its adaptability to different MRI systems. Furthermore, this study explored the pixel density around lesions by analyzing dilated masks of varying sizes. The best accuracy, notably, was achieved with a dilated size of 6. This suggests that information around and at the lesion borders plays a significant role in distinguishing active and inactive lesions.

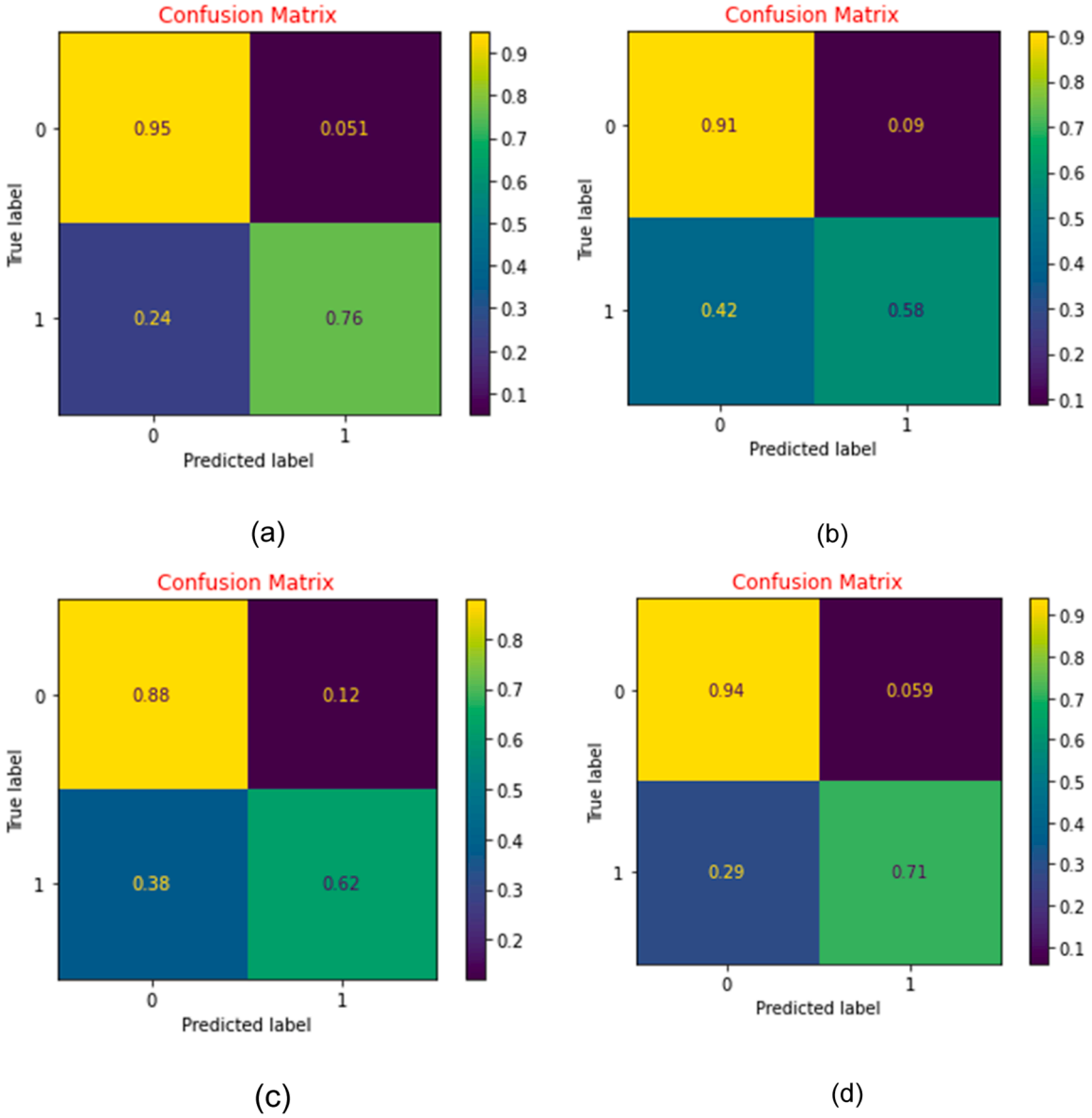


Fig. 7. Confusion Matrix of 3 top-performing models and ensemble method: (a) Custom CNN model; (b) ResNet50 model; (c) VGG19 model; (d) ensemble method.

Table 7
The result of external validation.

Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
81	93	69	0.83

5. Conclusions

In conclusion, the research to predict lesion enhancement in MR images without the need for contrast agents proved to be a challenging yet fruitful exploration of the potential of DL. This study yielded encouraging results, with our custom CNN model achieving an accuracy rate of 85 %, and pre-trained models, particularly ResNet50 and VGG19, demonstrating promise with an accuracy rate of 75 %. These outcomes signal an appealing alternative to contrast agent administration.

This study had a number of limitations that should be considered when interpreting the results:

- The process of identifying active and inactive lesions was not entirely automated. Instead, the network relied on manually selected Regions of Interest (ROIs), where lesion status had been previously diagnosed by radiologists and categorized accordingly.
- The use of 2D images might compromise image classification accuracy, suggesting the need for further exploration into incorporating 3D images to refine the methodology.
- This study exclusively utilized standard MRI images, highlighting a potential avenue for future investigation involving the inclusion of diffusion-weighted images for lesion diagnosis.
- While this study was conducted with a relatively large sample size, more data is required for robust conclusions in deep learning networks. This is already addressed through augmentation techniques.
- In the external evaluation of the model, although the test data differed in MRI system and image dimensions compared to the training data, confirming the generalizability of our custom CNN model to a large extent, further testing on a heterogeneous dataset is recommended for thorough validation.

Funding

This work was supported in part by the Vice-Chancellery for Research and Technology, Isfahan University of Medical Sciences.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Ethics Committee of Isfahan University of Medical Sciences (protocol code IR.MUI.MED.REC.1400.474)

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

CRediT authorship contribution statement

AmirAbbas Amini: Conceptualization, Data curation, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Azin Shayganfar:** Conceptualization, Data curation, Investigation, Resources, Writing – review & editing. **Zahra Amini:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **Leila Ostovar:** Investigation, Validation. **Somayeh HajiAhmadi:** Supervision, Visualization. **Navid Chitsaz:** Resources. **Masoud Rabbani:** Resources. **Raheleh Kafieh:** Conceptualization, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (Eds.), 2016. Tensorflow: a System for Large-Scale Machine Learning. Osdi, Savannah, GA, USA.

Agency EM. EMA's final opinion confirms restrictions on use of linear gadolinium agents in body scans. 2017.

A review of convolutional neural networks. In: Ajit, A., Acharya, K., Samanta, A. (Eds.), 2020. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE.

Al-Sabaawi A., Ibrahim H.M., Arkah Z.M., Al-Amidie M., Alzubaidi L., (Eds.), Amended convolutional neural network with global average pooling for image classification. Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020) held December 12–15, 2020; 2021: Springer.

Al-Saffar A.A.M., Tao H., Talab M.A., (Eds.), Review of deep convolution neural network in image classification. 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET); 2017: IEEE.

Ardakani, A., Nabavi, S., Farzan, A., Najafabad, B., 2017. Quantitative MRI texture analysis in differentiating enhancing and non-enhancing T1-hypointense lesions without application of contrast agent in multiple sclerosis. Czech Slovak Neurol. Neurosurg. 113 (6), 700–707.

Beomonte Zobel, B., Quattrocchi, C.C., Errante, Y., Grasso, R.F., 2016. Gadolinium-based contrast agents: did we miss something in the last 25 years? Radiol. Med. 121 (6), 478–481.

Burke, L.M., Ramalho, M., AlObaidy, M., Chang, E., Jay, M., Semelka, R.C, 2016. Self-reported gadolinium toxicity: a survey of patients with chronic symptoms. Magn. Reson. Imaging 34 (8), 1078–1080.

Caba, B., Cafaro, A., Lombard, A., Arnold, D.L., Elliott, C., Liu, D., et al., 2023. Single-timepoint low-dimensional characterization and classification of acute versus chronic multiple sclerosis lesions using machine learning. Neuroimage 265, 119787.

Datta, S., Sajja, B.R., He, R., Wolinsky, J.S., Gupta, R.K., Narayana, P.A., 2006. Segmentation and quantification of black holes in multiple sclerosis. Neuroimage 29 (2), 467–474.

Dietterich, T., 1995. Overfitting and undercomputing in machine learning. ACM Comput. Surv. (CSUR) 27 (3), 326–327.

Fraum, T.J., Ludwig, D.R., Bashir, M.R., Fowler, K.J., 2017. Gadolinium-based contrast agents: a comprehensive risk assessment. J. Magn. Reson. Imaging 46 (2), 338–353.

Freire P.G., Idagawa M.H., de Oliveira E.M.L., Abdala N., Carrete H., Ferrari R.J., (Eds.),. Classification of active multiple sclerosis lesions in MRI without the aid of gadolinium-based contrast using textural and enhanced features from FLAIR images. International Conference on Computational Science and Its Applications; 2020: Springer.

Gong, E., Pauly, J.M., Wintermark, M., Zaharchuk, G, 2018. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J. Mag. Reson. Imaging 48 (2), 330–340.

Grobner, T., 2006. Gadolinium—a specific trigger for the development of nephrogenic fibrosing dermopathy and nephrogenic systemic fibrosis? Nephrol. Dial. Transplant. 21 (4), 1104–1108.

Hu, H.H., Pokorney, A., Towbin, R.B., Miller, J.H., 2016. Increased signal intensities in the dentate nucleus and globus pallidus on unenhanced T1-weighted images: evidence in children undergoing multiple gadolinium MRI exams. Pediatr. Radiol. 46 (11), 1590–1598.

Kingma D.P., Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv :14126980. 2014.

Lin, Z., Li, Y., Su, P., Mao, D., Wei, Z., Pillai, J.J., et al., 2018. Non-contrast MR imaging of blood-brain barrier permeability to water. Magn. Reson. Med. 80 (4), 1507–1520.

McDonald, R.J., McDonald, J.S., Dai, D., Schroeder, D., Jentoft, M.E., Murray, D.L., et al., 2017. Comparison of gadolinium concentrations within multiple rat organs after intravenous administration of linear versus macrocyclic gadolinium chelates. Radiology 285 (2), 536–545.

Narayana, P.A., Coronado, I., Sujit, S.J., Wolinsky, J.S., Lublin, F.D., Gabr, R.E., 2020. Deep learning for predicting enhancing lesions in multiple sclerosis from noncontrast MRI. Radiology 294 (2), 398–404.

Sajja, B.R., Datta, S., He, R., Mehta, M., Gupta, R.K., Wolinsky, J.S., et al., 2006. Unified approach for multiple sclerosis lesion segmentation on brain MRI. Ann. Biomed. Eng. 34, 142–151.

Shinohara, R.T., Goldsmith, J., Mateen, F., Crainiceanu, C., Reich, D., 2012. Predicting breakdown of the blood-brain barrier in multiple sclerosis without contrast agents. Am. J. Neuroradiol. 33 (8), 1586–1590.

Traboulsee, A., Oh, J., Barlow, L., Chan, J., Cohen, B., Costello, K., et al., 2017. Consensus statement on the use of gadolinium for magnetic resonance imaging (MRI) used in the diagnosis and follow-up of patients with multiple sclerosis (MS). J. Neurol. Sci. 381, 957.

US Food and Drug Administration, 2015. FDA Drug Safety Communication: FDA Evaluating the Risk of Brain Deposits with Repeated Use of Gadolinium-Based Contrast Agents for Magnetic Resonance Imaging (MRI). US Food and Drug Administration, Silver Spring, MD, USA.

Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R.A., et al., 2020. Rising prevalence of multiple sclerosis worldwide: insights from the Atlas of MS. Mult. Scler. J. 26 (14), 1816–1821.

Ying X., (Ed.), An overview of overfitting and its solutions. Journal of physics: Conference series; 2019: IOP Publishing.