



Original article

Discrimination of multiple sclerosis using scanning laser ophthalmoscopy images with autoencoder-based feature extraction

Ali Aghababaei^{a,b,1}, Roya Arian^{a,1}, Asieh Soltanipour^a, Fereshteh Ashtari^c, Hossein Rabbani^a, Raheleh Kafieh^{d,*}

^a Medical Image and Signal Processing Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

^b School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

^c Isfahan Neurosciences Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

^d Department of Engineering, Durham University, South Road, Durham, UK



ARTICLE INFO

Keywords:

Multiple Sclerosis
Optical Coherence Tomography
Scanning Laser Ophthalmoscopy
Feature Extraction
Machine Learning
Deep Learning

ABSTRACT

Objective: Optical coherence tomography (OCT) investigations have revealed that the thickness of inner retinal layers becomes decreased in multiple sclerosis (MS) patients, compared to healthy control (HC) individuals. To date, a number of studies have applied machine learning to OCT thickness measurements, aiming to enable accurate and automated diagnosis of the disease. However, there have much less emphasis on other less common retinal imaging modalities, like infrared scanning laser ophthalmoscopy (IR-SLO), for classifying MS. IR-SLO uses laser light to capture high-resolution fundus images, often performed in conjunction with OCT to lock B-scans at a fixed position.

Methods: We incorporated two independent datasets of IR-SLO images from the Isfahan and Johns Hopkins centers, consisting of 164 MS and 150 HC images. A subject-wise data splitting approach was employed to ensure that there was no leakage between training and test datasets. Several state-of-the-art convolutional neural networks (CNNs), including VGG-16, VGG-19, ResNet-50, and InceptionV3, and a CNN with a custom architecture were employed. In the next step, we designed a convolutional autoencoder (CAE) to extract semantic features subsequently given as inputs to four conventional ML classifiers, including support vector machine (SVM), k-nearest neighbor (K-NN), random forest (RF), and multi-layer perceptron (MLP).

Results: The custom CNN (85 % accuracy, 85 % sensitivity, 87 % specificity, 93 % area under the receiver operating characteristics [AUROC], and 94 % area under the precision-recall curve [AUPRC]) outperformed state-of-the-art models (84 % accuracy, 83 % sensitivity, 87 % specificity, 92 % AUROC, and 94 % AUPRC); however, utilizing a combination of the CAE and MLP yields even superior results (88 % accuracy, 86 % sensitivity, 91 % specificity, 94 % AUROC, and 95 % AUPRC).

Conclusions: We utilized IR-SLO images to differentiate between MS and HC eyes, with promising results achieved using a combination of CAE and MLP. Future multi-center studies involving more heterogeneous data are necessary to assess the feasibility of integrating IR-SLO images into routine clinical practice.

1. Introduction

Multiple Sclerosis (MS) is the most common disorder of the central nervous system (CNS) affecting young adults, characterized by demyelination, gliosis, and axonal atrophy. Diagnosis of MS is not straightforward, relying on magnetic resonance imaging (MRI) and lumbar puncture (Thompson et al., 2018). This necessitates searching for cost-effective and less invasive biomarkers to facilitate early disease

detection and prediction of the severity progression. As the retina has been known as an extension of the CNS, studying this tissue may provide a valuable window to investigate the pathological changes affecting MS patients' brain (Mehmood et al., 2021). Indeed, optical coherence tomography (OCT) investigations suggest that the inner retinal layers, including the retinal nerve fiber layer (RNFL) and the ganglion cell-inner plexiform layer (GCIPL), become thinner in MS (Petzold et al., 2017); this has been mainly attributed to a retrograde axonal degeneration

* Corresponding author.

E-mail address: raheleh.kafieh@durham.ac.uk (R. Kafieh).

¹ These authors contributed equally.

following the inflammatory processes, regardless of whether clinically-discernable episodes of optic neuritis (ON) are present (Mehmood et al., 2021).

Machine learning (ML) is a subfield of artificial intelligence that is defined as the ability of a computer-based system to perform human tasks without explicitly being programmed. There exist a variety of ML algorithms that can be utilized for both classification and regression tasks, like support vector machines (SVMs), k-nearest neighborhood (K-NN), and neural networks (NNs) (Sarker, 2021). Inspired by the structure and function of the biological NNs within the human brain, NNs consist of several inter-connected nodes that process the input data through one or more hidden layers to generate an output. A neural network with multiple hidden layers is considered a "deep" network, fabricating the term deep learning (DL) extensively utilized nowadays owing to its astonishing superiority over traditional ML approaches, esp. with abundant input data (Mehlig). Of note, ML, including its sub-set DL, has gained much attention during recent years and has become applied to a various range of medical fields, including disease classification, prognosis prediction, and drug discovery, to name a few (Sidey-Gibbons and Sidey-Gibbons, 2019). In line with these advancements, numerous studies have shown that MS can be automatically detected using thickness measurements of retinal layers provided by OCT (Montolío et al., 2022; Hernandez et al., 2023; Garcia-Martin et al., 2013; Zhang et al., 2020; Montolío et al., 2021; Ortiz et al., 2023; Garcia-Martin et al., 2015; Kenney et al., 2022; Khodabandeh et al., 2024; Khodabandeh et al., 2023; Ciftci Kavaklioglu et al., 2022; Pérez del Palomar et al., 2019; López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021). A systematic review and meta-analysis published in late 2022 reported that ML models applied to OCT thickness data can discriminate between MS and healthy control (HC) individuals with high levels of accuracy (ACC) (pooled ACC = 93 % [95 % CI: 88 %, 97%], pooled sensitivity [SEN] = 95 % CI: 88 %, 96 %, pooled specificity [SPE] = 95 % CI: 89 %, 98 %) (Nabizadeh et al., 2022). Such remarkable results have been made using spectral-domain (SD) (Montolío et al., 2022; Hernandez et al., 2023; Garcia-Martin et al., 2013; Zhang et al., 2020; Montolío et al., 2021; Ortiz et al., 2023; Garcia-Martin et al., 2015; Kenney et al., 2022; Khodabandeh et al., 2024; Khodabandeh et al., 2023; Ciftci Kavaklioglu et al., 2022) and swept-source (SS) (Pérez del Palomar et al., 2019; López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021) OCT devices that employed peripapillary (Montolío et al., 2022; Garcia-Martin et al., 2013; Montolío et al., 2021; Garcia-Martin et al., 2015; Kenney et al., 2022; Ciftci Kavaklioglu et al., 2022; Pérez del Palomar et al., 2019), macular (Montolío et al., 2022; Zhang et al., 2020; Montolío et al., 2021; Kenney et al., 2022; Khodabandeh et al., 2024; Khodabandeh et al., 2023; Ciftci Kavaklioglu et al., 2022; Pérez del Palomar et al., 2019), wide (Pérez del Palomar et al., 2019; López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021), and posterior pole (Hernandez et al., 2023; Ortiz et al., 2023) scanning protocols. However, other imaging modalities like OCT angiography (OCT-A) or fundus camera photography, have not been utilized for automated diagnosis of MS.

Infrared scanning laser ophthalmoscopy (IR-SLO) is a retinal imaging technology often performed along with OCT to lock B-scans at a fixed position. This ensures that the effect of eye motion on image quality is minimized and also allows for a more accurate assessment of disease progression during follow-up visits since the same B scan is consistently referenced (Aumann et al., 2019). IR-SLO works by illuminating the retinal tissue with laser light in a raster pattern, with the backscattered light being passed through a confocal aperture so that the unwanted signals are eliminated, creating high-resolution two-dimensional images of the retina (Fischer et al., 2019). A recent study by Wisely et al. (Wisely et al., 2022) employed multiple imaging modalities, namely OCT, OCT-A, and ultra-widefield color SLO (not IR-SLO) and fundus autofluorescence, for classifying Alzheimer's disease using DL, leading to a best area under the receiver operating characteristics (AUROC) of 0.841 (95 % CI: 0.739, 0.943); the model performance considerably dropped

when color SLO images were used as the sole input data (0.450, [95 % CI: 0.282, 0.592]) (Wisely et al., 2022).

We aimed to apply various ML and DL approaches to these images for discriminating between MS and HC states, with the hope of discovering novel retinal biomarkers that might offer diagnostic value in routine clinical practice.

2. Materials and methods

2.1. Datasets

In this study, we utilized IR-SLO images from two distinct datasets including the Isfahan and Johns Hopkins datasets, consisting of MS and HC individuals. In both datasets, OCT scans and IR-SLO images were captured using a SPECTRALIS® SD-OCT device developed by Heidelberg Engineering in Heidelberg, Germany. The Isfahan dataset was compiled during a study conducted between April 2017 and March 2019 at the Kashani Comprehensive MS Center in Isfahan, Iran, which acts as a primary referral center for MS in the region (Ashtari et al., 2021). It consists of 282 IR-SLO images sourced from 35 MS patients (146 images) and 71 HC individuals (136 images). The Johns Hopkins dataset was publicly available and comprised IR-SLO and OCT images of the right eyes from 32 individuals, including 14 HCs and 18 patients diagnosed with MS (He et al., 2019). Overall, 314 IR-SLO images from a combination of both datasets, consisting of 150 HC and 164 MS images, respectively, were utilized to train convolutional neural networks (CNNs) (Section 3.1.) and convolutional autoencoder (CAE) plus ML classifiers (Section 3.2.).

2.2. Data preprocessing

Initially, all images were adjusted to a size of $128 \times 128 \times 1$ pixels, and their pixel intensities were normalized by dividing them by 255, resulting in values ranging from 0 to 1. Additionally, images associated with the left eyes were flipped horizontally to match the orientation of the right-eye images.

2.2.1. Train and test splitting

Subsequently, we implemented random splitting to divide the data into training and test sets using k-fold cross-validation (CV), with k set to 5. K-fold CV is favored over random splitting because of its thorough and adaptable nature. With this method, the model is trained on the complete dataset, ensuring each data point is included an equal number of times in both training (k-1 times) and validation (1 time) sets. In contrast, random splitting might lead to repeated selections in the test set due to re-sampling in each iteration. Consequently, k-fold CV is preferred for its ability to maximize dataset utilization and reduce biased selection during testing. It is noteworthy that we employed a stratified K-fold sampling technique to ensure an equal distribution of classes in each fold. Additionally, to prevent any data leakage between the training and test sets, we adopted a "subject-wise" approach. This method involves grouping all images from the same participant, regardless of their left/right orientation, and assigning them exclusively to either the training or test sets. This ensures no overlap between the two sets and reduces the likelihood of overestimating model performance (Saeb et al., 2017).

2.2.2. Data augmentation

In this study, due to the limited size of the training sets in each fold, we applied numerous geometric and color space transformations to augment the IR-SLO images. Data augmentation, a widely used preprocessing technique in machine learning, aims to address overfitting by introducing slight modifications to the original input images, generating new yet similar examples. This process artificially enhances the diversity and quantity of training samples.

Our augmentation techniques in this research include vertical

flipping, adjusting height within a range of ± 5 pixels, shifting width within a range of ± 30 pixels, rotating within a range of $\pm 5^\circ$, zooming within a range of ± 0.2 , and adjusting brightness within a range of 0.2–1.5.

2.3. Classification

Of note, CNNs work by integrating the dimensionality reduction and classification processes in an end-to-end manner. Using a CNN, features are extracted from input images through the convolutional layers, with no external feature extraction method needed to be applied; the resulting features are then utilized for final classification made by the fully connected (FC) layers (Mehlig, 2021). To test whether even higher classification accuracies can be achieved, an independent feature extraction method that is suitable for image data (CAE) was applied to the dataset and the obtained features were classified with various ML algorithms.

2.3.1. Deep learning (CNNs)

To classify IR-SLO images using CNNs, two distinct approaches were pursued. First, we developed a custom CNN model tailored for full training on our dataset from scratch. Secondly, we employed several state-of-the-art CNN architectures, such as VGG-16 (Simonyan and Zisserman, 2015), VGG-19 (Simonyan and Zisserman, 2015), ResNet-50 (He et al., 2015), and InceptionV3 (Szegedy et al., 2015), utilizing both transfer learning and fine-tuning strategies. Transfer learning involves transferring knowledge from pre-trained models to a new dataset, thus avoiding the need to start learning from scratch, which can be prone to overfitting, especially with smaller datasets. These utilized state-of-the-art models have exhibited impressive classification accuracies on large-scale image datasets like ImageNet (Deng et al., 2009). Hence, by keeping their weights unaltered, these models can function as fixed feature extractors, followed by adjustments to the FC component to align with our binary classification objective.

On the other hand, fine-tuning is an alternative strategy where the initial convolutional layers, responsible for extracting coarse features, remain frozen, while the uppermost convolutional layers, tasked with

capturing domain-specific and fine features, are unfrozen, enabling their weights to be modified during training phase. Therefore, we additionally applied a fine-tuning approach to the CNN model that exhibited the best performance.

In both DL approaches discussed above, the Optuna hyperparameter optimization framework was employed to identify the optimal CNN hyperparameters, including learning rate, batch size, dropout probability, the number of hidden layers in the FC part, and the number of neurons in each hidden layer (Akiba et al., 2019).

A brief introduction of CNNs is provided in the Supplementary Material on pages 2 and 3. Moreover, Fig. 1 illustrates an overview of our deep learning methods with optimal architectures.

As shown in Fig. 1, the proposed CNN comprised five blocks, each containing 2D convolutional layer (with a kernel size of 3×3), ReLU activation function, batch normalization, and max pooling layers stacked together; the number of channels (feature maps) in each convolutional layer progressed from 32, 64, 128, 256, and finally 512 through the network. Finally, an FC part consisting of three hidden layers with 54, 194, and 1287 neurons was added to the end of the CNN, with ReLU and sigmoid as the activation functions for the first three layers and the output layer, respectively. To reduce the risk of overfitting, dropout regularization with a probability of 0.3 was applied just after flattening layer were given to the FC part and to the first two FC layers as well. A dropout rate of 0.1 was also added at the end, just before the output layer (Fig. 1). Adam and binary cross entropy loss were used as the optimizer and loss function, respectively.

2.3.2. Feature extraction and machine learning classifiers

Dimensionality reduction is a necessary preprocessing step when dealing with high-dimensional data like images in order to overcome the so-called “curse of dimensionality”, a problem leading to high computational costs and poor performance of ML models. To address this, a number of methods have been developed that project data points to a lower dimensional space which still retains the most important information of the original data, e.g., principal component analysis (PCA) (Abdi and Williams, 2010) and autoencoder neural networks (AEs) (Arian et al., 2023; Bank et al., 2021); the resulting features can then be

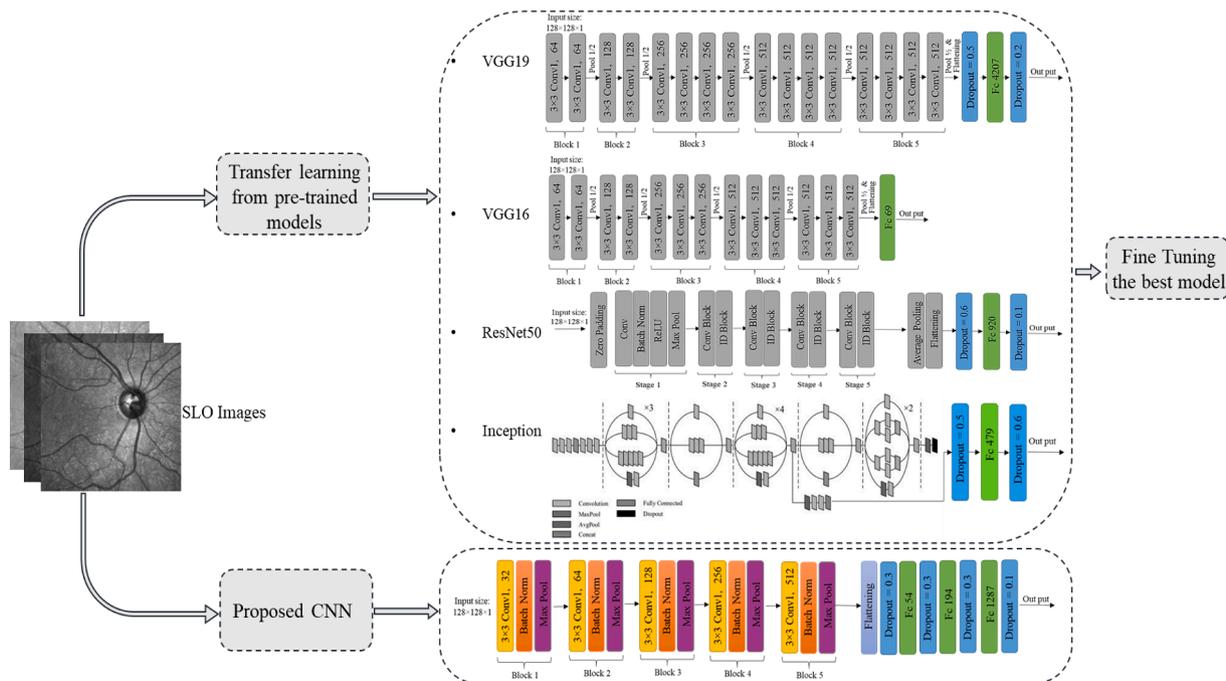


Fig. 1. Overview of MS classification based on infrared scanning laser ophthalmoscopy (IR-SLO) images using convolutional neural networks (CNNs). The gray blocks represent untrainable layers with frozen weights and the colored blocks show trainable layers.

given to ML algorithms for final classification. Indeed, when working with limited data, utilizing a combination of feature extraction and conventional machine learning algorithms is expected to outperform deep neural networks.

AEs aim to generate a compressed and meaningful representation of input data by learning to reconstruct it in an unsupervised manner. AEs consist of an encoder and a decoder, which are NNs in most scenarios, and a bottleneck layer between the two. The encoder network tries to map input data to a latent feature space with a lower dimensionality (bottleneck layer) and the decoder will then utilize the resulting features to create output data as much similar to the original data. During training, AEs minimize the difference between the original and reconstructed data, often using mean square error as the loss function for such a regression problem. Some of the various applications of AEs include image denoising, image generation, anomaly detection, and feature extraction; the latter can be seen as a non-linear extension of PCA. Indeed, in cases where no non-linear activation functions are used in AEs, the latent feature space created by the encoder is the same as the PCA output (Bank et al., 2021).

Interestingly, combining AEs with CNNs seems to yield more informative representations for image data by preserving the two-dimensional features of images, e.g., edges and corners. In the current study, the resulting features from such CAE were then given to a SVM classifier. Subsequently, CAE hyperparameters like the number of convolutional and pooling layers, number of filters, kernel size, learning rate, batch size, and dropout probability were optimized empirically to get the highest possible classification accuracies (without considering the reconstruction error). Furthermore, the features obtained from the CAE were also used to train a number of other ML classifiers, namely, K-NN, MLP, and random forest (RF) to achieve the highest possible level of accuracy. Each of these three classifiers is briefly introduced in the Supplementary Material on page 1. In this study, the grid search algorithm (Liashchynskiy and Liashchynskiy, 2019) and the Optuna library (Akiba et al., 2019) were utilized to find the optimal hyperparameters

for each classifier. Grid search algorithm involves calculating the ACC of each combination of all specified hyperparameters and then selecting the best value for them (Liashchynskiy and Liashchynskiy, 2019) while Optuna uses Bayesian optimization to find the optimal set of hyperparameters. Fig. 2 provides an overview of this section (feature extraction plus final classification), including the architecture of the designed CAE. All Conv2D layers in the CAE employed the ReLU activation function, except for the final layer where a sigmoid function was applied. Additionally, the optimization process was carried out using the Adam optimizer, and the loss was computed using the Huber loss function. The mean squared error (MSE) is a great loss function for learning outliers while the mean absolute error (MAE) ignores them. However, the Huber loss function demonstrates lower sensitivity to the outliers compared to the MSE by balancing the MSE and the MAE together. It behaves quadratically (like MSE) for small values of the difference between the predicted value y and the actual value $f(x)$, and linearly (like MAE) for large values (Huber, 1964). The mathematical formula for Huber loss function is as follows:

$$L_{\delta}(y, f(x)) = \begin{cases} \frac{1}{2}(y - f(x))^2, & \text{for } |y - f(x)| \leq \delta \\ \delta \cdot (|y - f(x)| - \frac{1}{2}\delta), & \text{otherwise} \end{cases} \quad (1)$$

Where the hyperparameter δ introduces a threshold which was set to 1 in this study. 2.4. Experimentation environment and evaluation metrics

All the experiments in this study were implemented using Python programming language, in the Keras platform backend in Python 3.7 software environment.

In this study, evaluation metrics such as ACC, SE, SP, precision (PR), and F1-score (F1) were utilized to assess the performance of the implemented models, as formulated below. Moreover, the receiver operating characteristics (ROC) and precision-recall curves were plotted and the areas under these curves, referred to as AUROC and area under the precision-recall curve (AUPRC), were also computed as two other

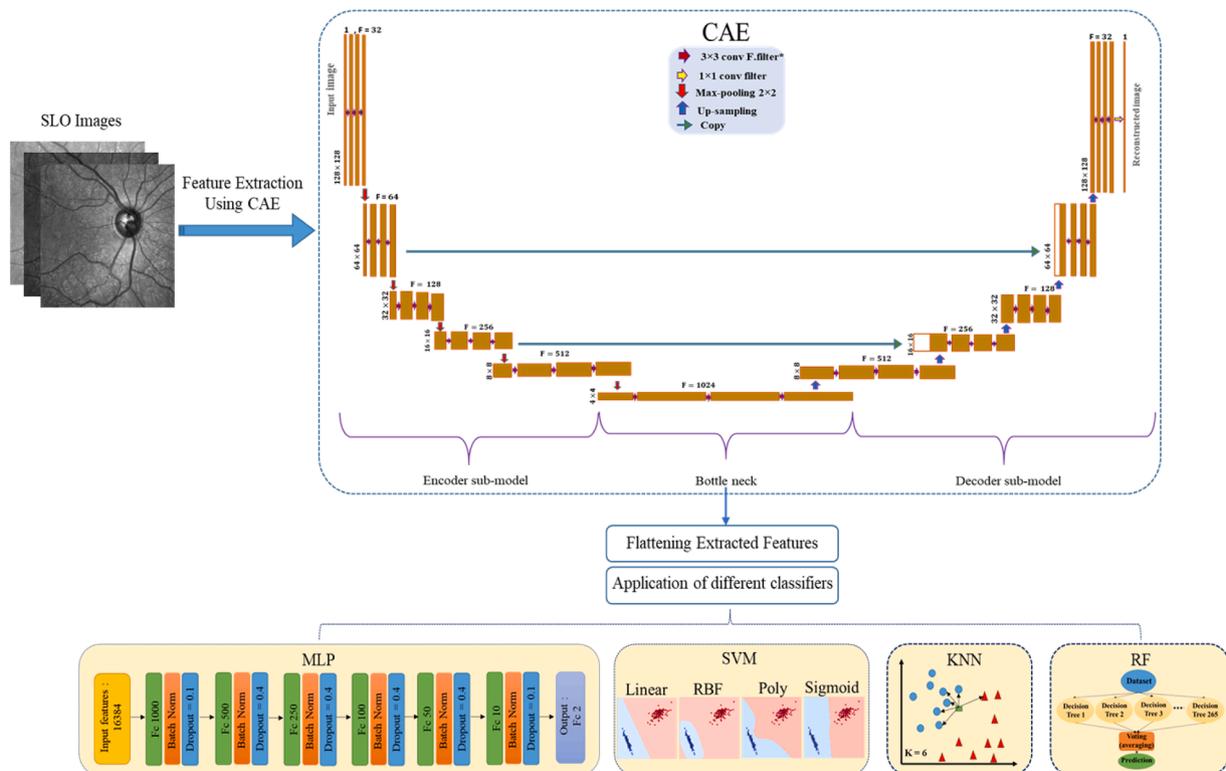


Fig. 2. Overview of MS classification based on IR-SLO images using a CAE as the feature extractor followed by conventional machine learning classifiers; MLP, multi-layer perceptron; SVM, support vector machine; K-NN, k-nearest neighbor; RF, random forest.

distinct metrics.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$SE = \frac{TP}{TP + FN} \quad (3)$$

$$SP = \frac{TN}{TN + FP} \quad (4)$$

$$PR = \frac{TP}{TP + FP} \quad (5)$$

$$F_1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (6)$$

where, TP, FN, TN, and FP represent true positives, false negatives, true negatives, and false positives, respectively.

For interpreting the predictions generated by CNN and CAE models, two types of saliency heat maps, namely the gradient-based class activation map (Grad-CAM) (Selvaraju et al., 2020) and the gradient-weighted latent activation mapping (Grad-LAM) (Bartler et al., 2021), an extension of Grad-CAM designed for unsupervised representation learning, were employed.

3. Results

The outcomes derived from CNNs, ML classifiers added to the CAE, and the ablation study are depicted in Sections 3.1, 3.2, and 3.3, respectively.

3.1. Deep learning (CNNs)

As mentioned in Section 2.3.1, state-of-the-art CNN models were first used as feature extractors with the weights of all layers kept frozen, followed by FC layers with a custom number of hidden layers containing varied number of neurons (Fig. 1). VGG-19 (with 14,714,688 untrainable parameters) showed the best performance, with a mean ACC of 84 % (SE = 83 %, SP = 87 %, PR = 87 %, F1-score = 84 %, AUROC = 92 %) over five consecutive executions (Table 1). Moreover, the custom CNN model that was trained from scratch showed more desirable results (ACC = 85 %, SE = 85 %, SP = 87 %) (Table 1).

In an effort to enhance the outcomes, we proceeded to unfreeze specific convolutional layers of VGG19, the selected transfer learning model. Specifically, the first, second, and fourth topmost convolutional layers were unfrozen, resulting in 2359,808, 4719,616, and 7079,424 parameters available for retraining with the IR-SLO image dataset, respectively. Despite this fine-tuning approach, it did not yield improved performance. For instance, when the first topmost layer was unfrozen, the obtained results were as follows: ACC = 84 %, SE = 83 %, SP = 85 %, PR = 85 %, AUROC = 92 %, F1-score = 83 % (as shown in Fig. A in the Supplementary Material).

Box plots for the state-of-the-art and the custom CNN models are

Table 1

Performance metrics of the state-of-the-art and the proposed convolutional neural networks (CNNs) for classification of MS using IR-SLO images. Best results are bolded, revealing that the custom CNN is the winning classifier.

Model	ACC	SP	SE	PR	F1	AUROC	AUPRC	Optimal hyper parameters	
								Batch size	Learning rate
InceptionV3	0.81	0.83	0.80	0.83	0.81	0.90	0.92	16	6e-4
ResNet-50	0.71	0.84	0.71	0.83	0.70	0.79	0.85	64	1.4e-4
VGG16	0.83	0.86	0.81	0.85	0.83	0.92	0.93	8	1.7e-4
VGG19	0.84	0.87	0.83	0.87	0.84	0.92	0.94	8	1.2e-4
Proposed CNN	0.85	0.87	0.85	0.87	0.85	0.93	0.94	16	2.5e-4

Abbreviations: ACC, accuracy; SP, specificity; SE, sensitivity; PR, precision; F1, F1-score; AUROC, area under the receiver operating characteristics curve; AUPRC, area under the precision-recall curve; MS, multiple sclerosis; HC, healthy control.

drawn in Fig. 3.

3.2. Feature extraction and machine learning classifiers

The proposed architecture for the CAE is illustrated in Fig. 2, with a brief explanation available in the Supplementary Material on page 3. The flattened representation of the bottleneck output, suggestive of the most important features of IR-SLO images, was then utilized for classifying MS and HC using SVM, MLP, RF, and K-NN algorithms. The SVM classifier achieved an ACC of 86 % (SE = 82 %, SP = 90 %) using a radial basis function (RBF) kernel as the winning kernel (Table A in the Supplementary Material); the optimal values for CAE hyperparameters, including batch size (= 16), learning rate (= 1e-5), and dropout probability (= 0.2), were then kept unchanged when other classifiers were utilized. Furthermore, to visualize which regions have the most contribution to the CAE reconstruction score, the saliency maps of several MS and HC images from the test dataset were created using Grad-LAM, illustrated in Fig. 4. Comparison of These heat maps and the heat maps obtained by the custom CNN (using the Grad-CAM algorithm) for the same MS and HC images can provide some perspective of the regions of the images with most impact for each model (Fig. 4).

Subsequently, other conventional ML algorithms, namely, K-NN, MLP, and RF, were applied to the CAE-extracted features, with MLP achieving the best results (ACC= 88 %, SE= 86 %, SP= 91 %, PR= 90 %, F1-score= 88 %; AUROC = 94 %; AUPRC = 95 %) (Table 2). Fig. 5 shows the ROC/precision-recall curves and the confusion matrices for all the four classifiers. The best hyperparameters of K-NN, MLP, and RF are depicted in the Supplementary Material Table B. The high discriminant capacity of the features extracted by the CAE is visualized in Fig. 6 using Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction technique (McInnes et al., 2018) In the same figure, the discriminant capacity of the CAE is also compared with that of PCA as a simple feature extraction technique where the number of components was set to 300.

3.3. Ablation study

An ablation study, investigating the effect of feature extraction prior to the classification, was also undertaken when SVM with an RBF kernel and MLP were utilized. The SVM classifier (with RBF kernel) separates data points by projecting them to an infinite dimensional kernel space (Cervantes et al., 2020); thus, it may not hypothetically seem reasonable to reduce the dimensionality of our dataset before applying this algorithm. Also, NNs generally exhibit superior performance when dealing with higher dimensional data, not passed through feature extraction/selection algorithms (LeCun et al., 2015). Therefore, classification performance of SVM (with RBF kernel) and MLP was evaluated with and without applying the dimensionality reduction method, with the corresponding results summarized in Table 3.

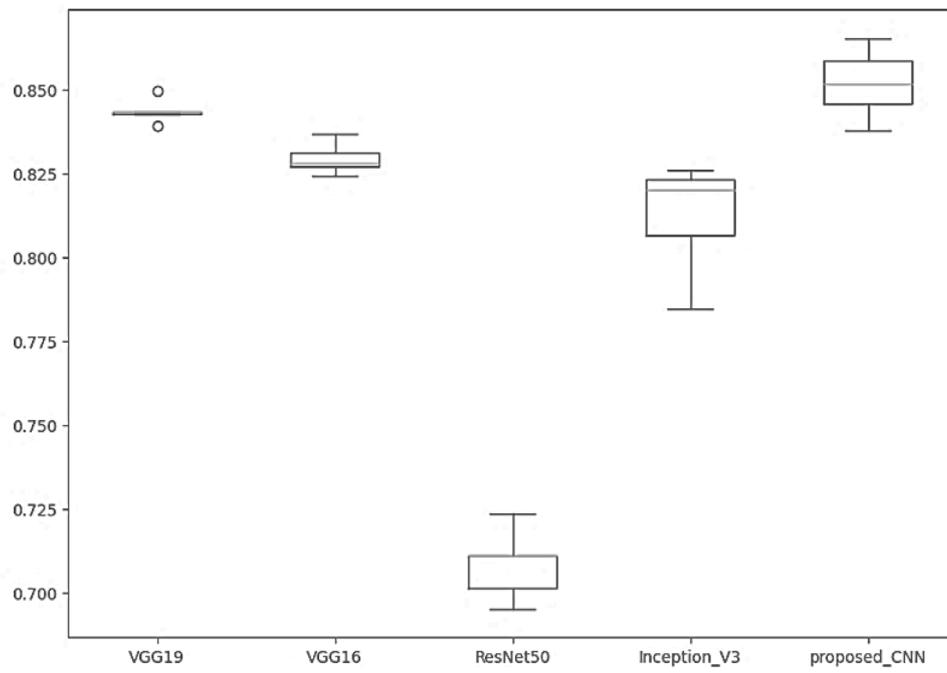


Fig. 3. Box plots for state-of-the-art convolutional neural networks (CNNs) and the proposed model, drawn following five repeated executions.

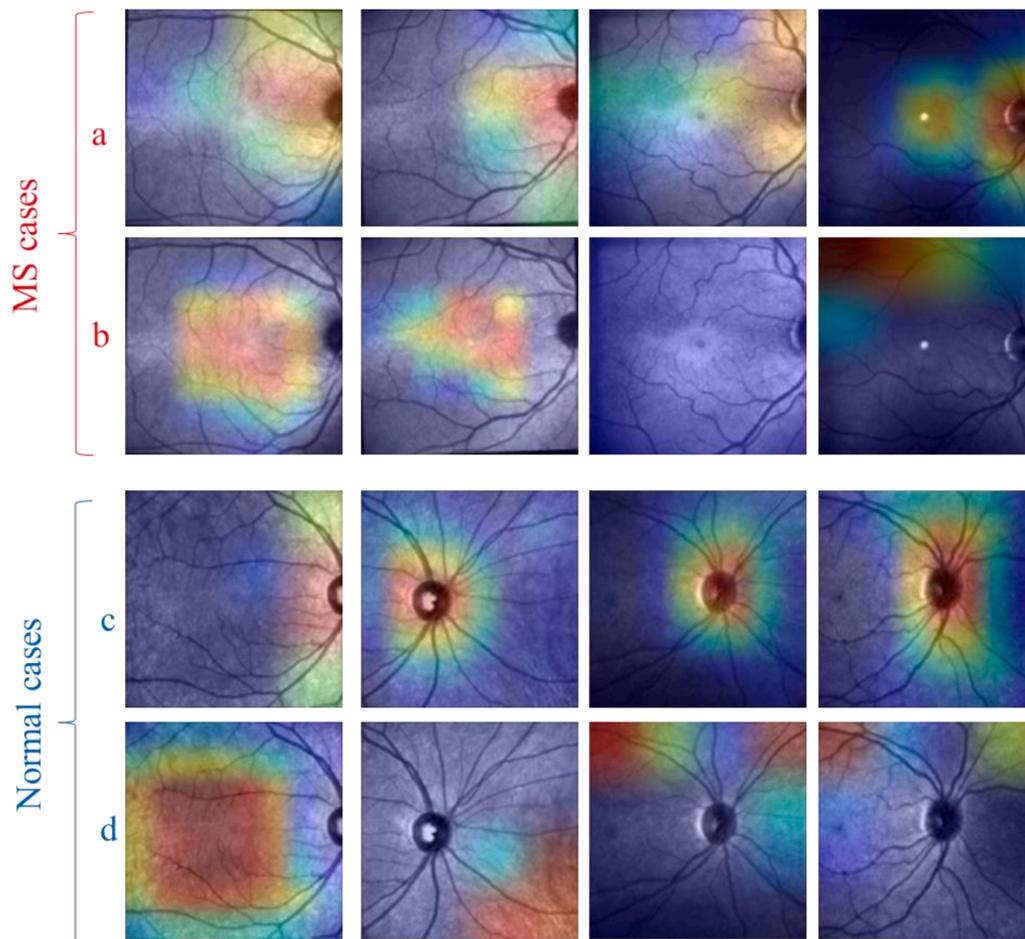


Fig. 4. Saliency maps generated for four IR-SLO images of MS and HC subjects from the test dataset when the proposed CAE (a and c) and CNN (b and d) were utilized.

Table 2

Performance metrics of the four machine learning models for classifying MS using infrared scanning laser ophthalmoscopy (IR-SLO) images. The classifiers were applied on the features extracted using the proposed convolutional autoencoder, with MLP being the winning model.

Model	ACC	SE	SP	PR	F1	AUROC	AUPRC
RF	0.83	0.76	0.89	0.92	0.82	0.9	0.92
K-NN	0.8	0.81	0.79	0.8	0.8	0.88	0.91
RBF-SVM	0.86	0.82	0.9	0.89	0.86	0.92	0.94
MLP	0.88	0.86	0.91	0.9	0.88	0.94	0.95

Abbreviations: ACC, accuracy; SP, specificity; SE, sensitivity; PR, precision; F1, F1-score; AUROC, area under the receiver operating characteristics curve; AUPRC, area under the precision-recall curve; RF, random forest; K-NN, k-nearest neighbor; RBF-SVM, support vector machine with radial basis function kernel; NN, neural network.

4. Discussion

In the current study, we showed that training ML classifiers with IR-SLO images yields encouraging results in classifying MS (ACC = 88 %, SE = 0.86, SP = 0.91); the proposed model was indeed a combination of CAE for feature extraction and a MLP for final classification.

All published studies in this field have focused solely on OCT data, i. e., thickness measurements of different retinal layers, mainly RNFL and GCIP. Although three works have been able to reach accuracies more than 90 % (Montolío et al., 2022; Hernandez et al., 2023; Khodabandeh et al., 2024), our model led to results comparable to the majority of previous studies utilizing SD-OCT devices (Garcia-Martin et al., 2013; Zhang et al., 2020; Montolío et al., 2021; Ortiz et al., 2023; Garcia-Martin et al., 2015; Kenney et al., 2022; Khodabandeh et al., 2023; Ciftci Kavaklioglu et al., 2022), with accuracies ranging from 80 % (Ciftci Kavaklioglu et al., 2022) to 88 % (Montolío et al., 2021; Khodabandeh et al., 2023). However, classifying MS based on SS-OCT data results in superior performance (Pérez del Palomar et al., 2019; López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021), even reaching 100 % (López-Dorado et al., 2021). This could be attributed, at least partially, to the greater precision of information offered by SS-OCT devices. Inherent differences between datasets concerning demographic characteristics such as age, gender, and ethnicity of participants may have also contributed to the discrepancy in performance between the model proposed in this study and SS-OCT-based models. It is worth noting that three of the four studies that employed SS-OCT used a same dataset with a small size (López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021), and therefore, it is crucial not to overemphasize the positive outcomes of these studies. Furthermore, given the limited availability of SS-OCT devices in many clinical settings, esp. in less developed regions, it is of relevance to prioritize studies utilizing the less costly SD-OCT and compare the performance of IR-SLO-based models with them.

According to the Table 1, the custom CNN model that was trained from scratch showed more desirable results (ACC = 85 %, SE = 85 %, SP = 87 %). This could partly be attributed to the fact that the architecture of our proposed CNN was not as complex compared to that of the state-of-the-art models, which is probably more suitable for the IR-SLO images containing relatively simple patterns to be recognized. For the same reasons, the fine-tuning approach did not lead to a superior performance compared to the transfer learning given an extreme increase in the number of trainable parameters.

Applying a CAE for feature extraction led to higher results compared to CNNs; therefore, we speculate that to reconstruct the input image effectively, the CAE should learn more semantically meaningful representations. The idea behind this hypothesis is similar to the paper published by Pathak et al. (2016), introducing the “context encoders” as a solution for semantic inpainting. Context encoders are indeed a type of CAEs that are aimed to generate missing parts of an image with respect to the contextual information from the surrounding regions. The latent

layer of a context encoder was shown to provide valuable features that serve as reliable indicators of the input images, resulting in appealing results across different tasks, including object detection, semantic segmentation, and classification (Pathak et al., 2016). Furthermore, the capability of the proposed CAE in detecting more informative representations could partly be attributed to the connections between two of the encoder and decoder blocks. This will propagate the information between encoder and decoder parts and prevent the precise information to be lost during the up-sampling process, since a feature map with higher resolution is constructed and then processed by the decoder convolutional layers. In addition, as illustrated in Fig. 4, MS causes pathological changes in the optic nerve head (ONH) and the area surrounding which can be detected through the proposed CAE but remain unnoticed by human physicians. In comparison, the custom CNN trained from scratch identified other regions than those within the optic disc area.

Utilizing OCT thickness data in previous ML studies is reasonable given that statistical investigations have highlighted a significant reduction in RNFL and GCIPL thickness (Petzold et al., 2017); however, there is far less evidence to support the notion that MS-related pathological changes can also be identified within en-face images like fundus camera photographs and IR-SLO images. RNFL damage can be recognized when a physician undertakes ophthalmoscopic examination, but the changes are not visible to human eye until at least half the RNFL thickness has been attenuated (Quigley et al., 1960). The prominence of the optic disc area depicted in Fig. 4 is in line with previous structural OCT studies demonstrating a substantial reduction in peripapillary RNFL thickness (Petzold et al., 2017). Furthermore, the vascular changes around the ONH may have also played a role in the model decision to give greater attention to the peripapillary regions. In this regard, OCT-A studies on MS patients have recently revealed that capillary vessel densities become decreased in peripapillary and macular regions, especially in individuals who have a prior history of ON (Feucht et al., 2019; Murphy et al., 2020; Ulusoy et al., 2020). Interestingly, Spain et al. (2018) highlighted that not only minor but also major vessels in the optic disc area could be affected. Their findings demonstrated substantial reductions in the ONH flow index, defined as the averaged OCT-A flow signal over the optic disc, among MS patients, regardless of their ON status, compared to HCs (Spain et al., 2018). The vascular damage in MS can be linked to ganglion cell atrophy leading to decreased metabolic demand and subsequent vessel attenuation via autoregulatory mechanisms (Murphy et al., 2020); conversely, impaired retinal blood perfusion resulting in a hypoxic condition may serve as the primary event, activating inflammatory mechanisms that cause demyelination and neurodegeneration (Halder and Milner, 2021). Overall, we showed that despite the absence of readily observable clinical alterations in en-face retinal images, such as IR-SLO images utilized here, employing ML approaches might offer insights into this conundrum and enable automated MS detection.

Additionally, we showed that the feature extraction step seems to be an essential preprocessing step for the data utilized in this study. As shown in Table 3, results of the ablation study revealed that by removing the feature extraction step, the classification ACC dropped from 86 % to 81 % and 88 % to 79 % for SVM and MLP classifiers, respectively.

Recently, Khodabandeh et al. (2024) (not peer-reviewed yet) utilized a dataset identical to one used in this work, consisting of the Isfahan and Johns Hopkins datasets; however, instead of IR-SLO images, OCT thickness maps were employed. Similar to the current study, the authors examined the efficacy of three strategies for classifying MS, including AE for feature extraction coupled with MLP for classification, a customized CNN with its weights trained from scratch, and fine-tuning of a state-of-the-art CNN (ResNet152V2 (He et al., 2015)). For the input data, thickness/boundary maps of RNFL, GCIPL, and INL were incorporated in a channel-wise manner, creating three-channel images with different sizes, 60 * 256 * 3 and 224 * 224 * 3; also, the maps were concatenated horizontally to form a mosaic image with dimensions of

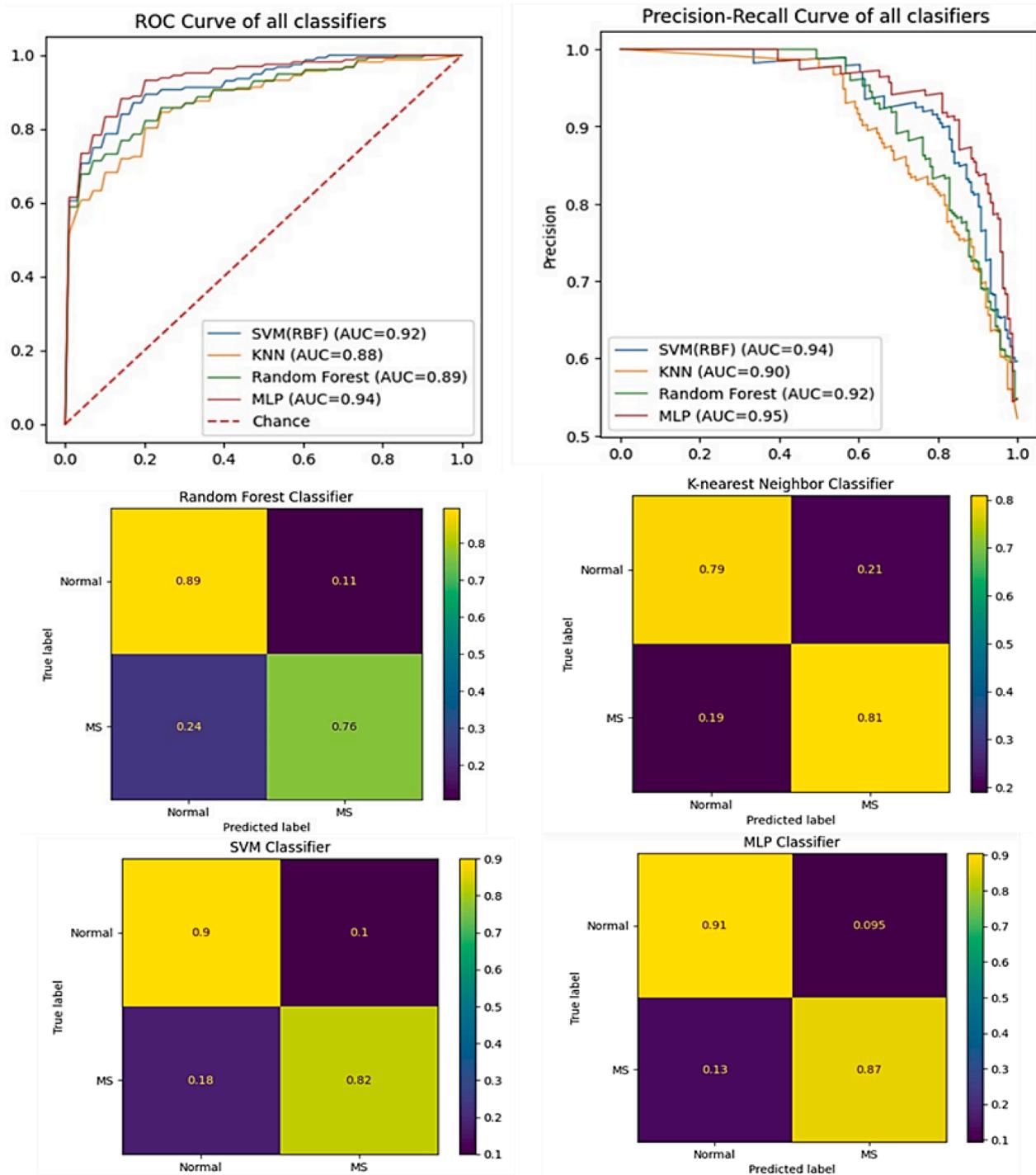


Fig. 5. Receiver operating characteristics (ROC) curve (top left), precision-recall curve (top right), and the confusion matrices (bottom) for the four machine learning classifiers. MLP, multi-layer perceptron; SVM, support vector machine; K-NN, k-nearest neighbor; RF, random forest.

672 * 224 * 1. They finally showed that a balanced ACC of 97.3 % can be achieved when thickness maps of the size 60 * 256 * 3 are integrated through the channel-wise approach using the custom CNN. Interestingly, to check for generalization ability of their models, the authors also utilized each of the two datasets exclusively for training or testing phases; for instance, a balanced ACC of 80.5 % and 84.5 % were achieved when the Isfahan dataset served as the training and test dataset, respectively. Unlike Khodabandeh et al. that employed simple AE, we took advantage of a CAE that may hypothetically have a superior performance on image data containing two-dimensional features like edges and corners. Also, we assessed the performance of different traditional

ML models, not solely the MLP, applied to the CAE-extracted features. It is worth noting that most previous studies that utilized OCT thickness measurements for classifying MS (Montolío et al., 2022; Garcia-Martin et al., 2013; Zhang et al., 2020; Montolío et al., 2021; Garcia-Martin et al., 2015; Khodabandeh et al., 2024; Ciftci Kavaklioglu et al., 2022; Pérez del Palomar et al., 2019; López-Dorado et al., 2021; Cavaliere et al., 2019; Garcia-Martin et al., 2021), like Khodabandeh et al.'s (Khodabandeh et al., 2024), employed a record-wise data splitting approach that may have led to a data leakage between train and test datasets, causing overestimation of the model performance (Saeb et al., 2017); nonetheless, we followed a less biased method in which data

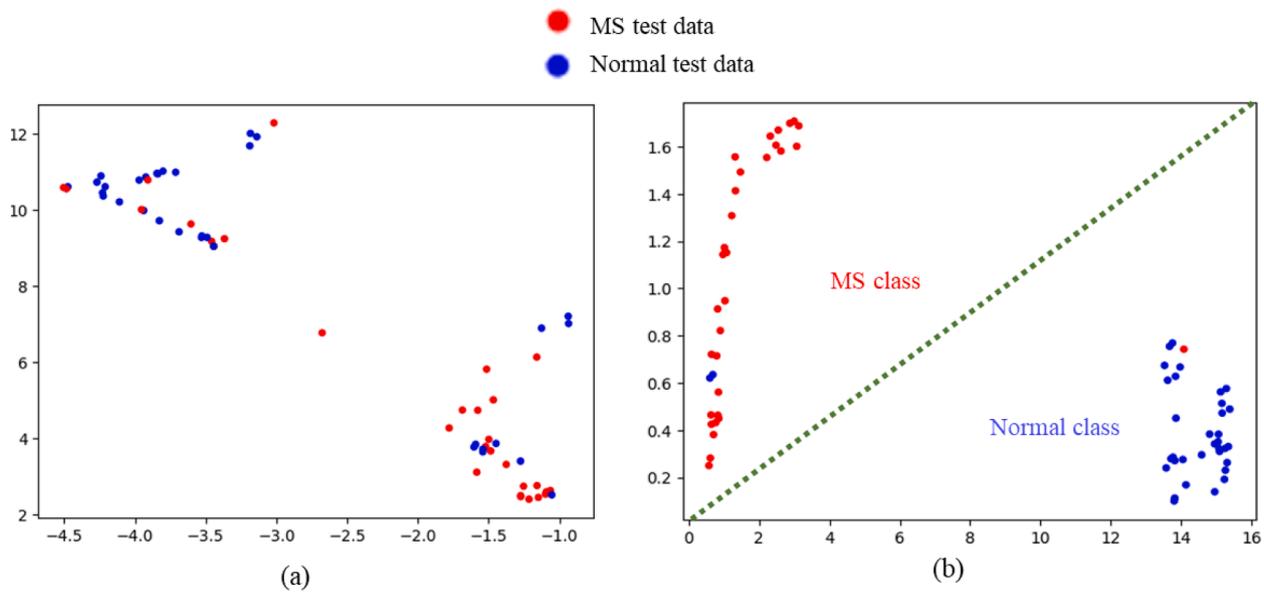


Fig. 6. Visualization of the features obtained from the convolutional autoencoder neural network using Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction technique (McInnes et al., 2018).

Table 3

Results of the ablation study undertaken multi-layer perceptron (MLP) and the support vector machine (SVM) with radial basis function kernel (RBF-SVM) with and without feature extraction prior to classification. Best results for each classifier are bolded. Using convolutional autoencoder (CAE) in both methods yields higher results.

Model	CAE	ACC	SE	SP	PR	F1	AUROC	AUPRC
MLP	Used	0.88	0.86	0.91	0.9	0.88	0.94	0.95
	Removed	0.79	0.78	0.82	0.82	0.79	0.88	0.9
RBF-SVM	Used	0.86	0.82	0.9	0.89	0.86	0.92	0.94
	Removed	0.81	0.82	0.81	0.81	0.81	0.88	0.9

Abbreviations: ACC, accuracy; SP, specificity; SE, sensitivity; PR, precision; F1, F1-score; AUROC, area under the receiver operating characteristics curve; AUPRC, area under the precision-recall curve.

concerning each subject appear exclusively in either training or test datasets. Additionally, Khodabandeh et al. (2024) did not report ACC results but concentrated on the balanced ACC, which is calculated as the arithmetic mean of SE and SP. To ensure a more accurate comparison between the model proposed in this study and the Khodabandeh et al. (2024), we also employed a record-wise data splitting approach (not included in the Results section); the CAE concatenated with SVM classifier achieved a SE of 97 % and a SP of 83 %. The superior outcome observed in Khodabandeh et al. (2024) study may be attributed partly to the CNN architecture they designed, although it is important not to ignore the fact that OCT thickness data may harbor more discriminating features compared to IR-SLO images.

This study has several limitations needed to be addressed. First, although we incorporated two independent datasets of IR-SLO images ($n = 314$), the sample size is still limited, necessitating future studies to include datasets from multiple centers with diverse demographic and clinical characteristics. We tried to minimize the risk of model overfitting by artificially increasing the number of images using various data augmentation techniques. A second limitation of this study is the lack of differentiation between images with and without a history of ON. Ganglion cell atrophy and RNFL damage are more pronounced in eyes with prior episodes of ON (Petzold et al., 2017); therefore, conducting additional experiments to train ML models exclusively with images of either ON-positive or ON-negative eyes could lead to more reliable and robust outcomes. Third, we solely utilized input data from a single imaging modality. Indeed, providing the models with more heterogeneous input data can potentially improve performance and reduce the risk of overfitting to the training data. Therefore, we suggest that future studies

combine IR-SLO images with other retinal imaging modalities, like OCT and OCT-A, and even go further, integrating retinal imaging data with MRI and serum/CSF biomarkers.

5. Conclusion

Overall, we applied various ML and DL-based models to IR-SLO images, and suggested that these two-dimensional en-face retinal images may provide useful information for discriminating between MS and HC individuals. Of note, the results obtained in this study are comparable to those from many previous ML works that utilized OCT thickness measurements for classifying MS. IR-SLO and other retinal imaging modalities that are less commonly utilized compared to OCT hold the potential to be included into routine clinical practice, enabling timely diagnosis of the disease and effective therapeutic interventions. However, studies including more diverse datasets from various independent centers are needed to further evaluate the validity of this assumption.

Data and code availability

The Johns Hopkins dataset is available at: <http://iacl.jhu.edu/Resources>

Code and models are available at: <https://doi.org/10.5281/zenodo.8217281>.

The Optuna code to find optimal hyperparameters is available at: <https://doi.org/10.5281/zenodo.8218403>.

Funding

None

CRedit authorship contribution statement

Ali Aghababaei: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Roya Arian:** Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. **Asieh Soltanipour:** Methodology, Software, Visualization, Writing – original draft. **Fereshteh Ashtari:** Data curation, Resources, Writing – review & editing. **Hossein Rabbani:** Conceptualization, Writing – review & editing. **Raheleh Kafieh:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

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

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.msard.2024.105743.

References

- Abdi, H., Williams, L.J., 2010. Principal component analysis. *WIREs Comput. Stat.* 2 (4), 433–459.
- Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: A Next-Generation Hyperparameter Optimization Framework [Internet]. arXiv; 2019 [cited 2023 Jul 25]. Available from: <http://arxiv.org/abs/1907.10902>.
- Arian, R., Mahmoudi, T., Riazi-Esfahani, H., Faghihi, H., Mirshahi, A., Ghassemi, F., Khodabandeh, A., Kafieh, R., Khalili Pour, E., 2023. Automatic choroid vascularity index calculation in optical coherence tomography images with low-contrast sclerohoroidal junction using deep learning. *Photonics* 10, 234.
- Ashtari, F., Ataei, A., Kafieh, R., Khodabandeh, Z., Barzegar, M., Raei, M., et al., 2021. Optical coherence tomography in neuromyelitis optica spectrum disorder and multiple sclerosis: a population-based study. *Mult. Scler. Relat. Disord.* 47, 102625. Jan.
- Aumann, S., Donner, S., Fischer, J., Müller, F., 2019. Optical coherence tomography (OCT): principle and technical realization. *High Resolution Imaging in Microscopy and Ophthalmology*. Springer Cham, pp. 59–85.
- Bank D., Koenigstein N., Giryres R. Autoencoders [Internet]. arXiv; 2021 [cited 2023 Jul 4]. Available from: <http://arxiv.org/abs/2003.05991>.
- Bartler, A., Hinderer, D., Grad-LAM, Yang B., 2021. Visualization of Deep Neural Networks for Unsupervised Learning. In: In: Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO). Amsterdam, Netherlands. IEEE, pp. 1407–1411 [Internet][cited 2023 Jul 26] Available from: <https://ieeexplore.ieee.org/document/9287730/>.
- Cavaliere, C., Vilades, E., Alonso-Rodríguez, M., Rodrigo, M., Pablo, L., Miguel, J., et al., 2019. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors* 19 (23), 5323. Dec 3.
- Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: applications, challenges and trends. *Neurocomputing* 408, 189–215. Sep.
- Ciftci Kavaklioglu, B., Erdman, L., Goldenberg, A., Kavaklioglu, C., Alexander, C., Oppermann, H.M., et al., 2022. Machine learning classification of multiple sclerosis in children using optical coherence tomography. *Mult. Scler. J.* 28 (14), 2253–2262. Dec.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., ImageNet, Fei-Fei L., 2009. A large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255 p.
- Feucht, N., Maier, M., Lepennetier, G., Pettenkofer, M., Wetzlmair, C., Daltrozzo, T., et al., 2019. Optical coherence tomography angiography indicates associations of the retinal vascular network and disease activity in multiple sclerosis. *Mult Scler* 25 (2), 224–234. Houndmills Basingstoke Engl.Feb.
- Fischer, J., Otto, T., Delori, F., Pace, L., Staurengli, G., 2019. Scanning laser ophthalmoscopy (SLO). In: *High Resolution Imaging in Microscopy and Ophthalmology*. Springer Cham, pp. 35–57.
- García-Martin, E., Pablo, L.E., Herrero, R., Ara, J.R., Martín, J., Larrosa, J.M., et al., 2013. Neural networks to identify multiple sclerosis with optical coherence tomography. *Acta Ophthalmol.* 91 (8), e628–e634 (Copenh)Dec.
- García-Martin, E., Herrero, R., Bambo, M.P., Ara, J.R., Martín, J., Polo, V., et al., 2015. Artificial neural network techniques to improve the ability of optical coherence tomography to detect optic neuritis. *Semin. Ophthalmol.* 30 (1), 11–19. Jan 2.
- García-Martin, E., Ortiz, M., Boquete, L., Sánchez-Morla, E.M., Barea, R., Cavaliere, C., et al., 2021. Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier. *Comput. Biol. Med.* 129, 104165. Feb.
- Halder, S.K., Milner, R., 2021. Hypoxia in multiple sclerosis; is it the chicken or the egg? *Brain* 144 (2), 402–410. Mar 3.
- He K., Zhang X., Ren S., Sun J. Deep Residual Learning For Image Recognition [Internet]. arXiv; 2015 [cited 2023 May 10]. Available from: <http://arxiv.org/abs/1512.03385>.
- He, Y., Carass, A., Solomon, S.D., Saidha, S., Calabresi, P.A., Prince, J.L., 2019. Retinal layer parcellation of optical coherence tomography images: data resource for multiple sclerosis and healthy controls. *Data Brief* 22, 601–604. Feb.
- Hernandez, M., Ramon-Julvez, U., Vilades, E., Cordon, B., Mayordomo, E., Garcia-Martin, E., Raafat, K.A., 2023. Explainable artificial intelligence toward usable and trustworthy computer-aided diagnosis of multiple sclerosis from optical coherence tomography. editor *PLOS One* 18 (8), e0289495. Aug 7.
- Huber, P.J., 1964. Robust estimation of a location parameter. *Ann. Math. Stat.* 35 (1), 73–101.
- Kenney, R.C., Liu, M., Hasanaj, L., Joseph, B., Abu Al-Hassan, A., Balk, L.J., et al., 2022. The role of optical coherence tomography criteria and machine learning in multiple sclerosis diagnosis. *Neurology* 99 (11), e1100–e1112. Sep 13.
- Khodabandeh, Z., Rabbani, H., Ashtari, F., Zimmermann, H.G., Motamedi, S., Brandt, A. U., et al., 2023. Discrimination of multiple sclerosis using multicenter OCT images. *Mult. Scler. Relat. Disord.* [Internet][Jun 23 [cited 2023 Jun 26];0(0). Available from: [https://www.msard-journal.com/article/S2211-0348\(23\)00348-6/fulltext](https://www.msard-journal.com/article/S2211-0348(23)00348-6/fulltext).
- Khodabandeh, Z., Rabbani, H., Bidabadi, N.S., Bonyani, M., Kafieh, R., 2024. Comprehensive Evaluation of Artificial Intelligence Models for Diagnosis of Multiple Sclerosis Using Information from Retinal Layers Multicenter OCT Images [Internet]. *Neurology*. Mar [cited 2024 Mar 11]. Available from: <http://medrxiv.org/lookup/doi/10.1101/2024.03.05.24303789>.
- López-Dorado, A., Ortiz, M., Satue, M., Rodrigo, M.J., Barea, R., Sánchez-Morla, E.M., et al., 2021. Early diagnosis of multiple sclerosis using swept-source optical coherence tomography and convolutional neural networks trained with data augmentation. *Sensors* 22 (1), 167. Dec 27.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444. May 28.
- Liashchynskiy P., Liashchynskiy P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS [Internet]. arXiv; 2019 [cited 2023 Jul 25]. Available from: <http://arxiv.org/abs/1912.06059>.
- McInnes, L., Healy, J., Saul, N., 2018. Großberger L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* 3 (29), 861.
- Mehlig B. Machine Learning with Neural Networks: An Introduction for Scientists and Engineers [Internet]. 1st ed. Cambridge University Press; 2021 [cited 2024 Jun 22]. Available from: <https://www.cambridge.org/core/product/identifier/9781108860604/type/book>.
- Mehmood, A., Ali, W., Song, S., Din, Z.U., Guo, R., Shah, W., et al., 2021. Optical coherence tomography monitoring and diagnosing retinal changes in multiple sclerosis. *Brain Behav.* 11 (10). Oct.
- Montolio, A., Martín-Gallego, A., Cegoñino, J., Orduna, E., Vilades, E., García-Martin, E., et al., 2021. Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. *Comput. Biol. Med.* 133, 104416. Jun.
- Montolio, A., Cegoñino, J., García-Martin, E., Pérez del Palomar, A., 2022. Comparison of machine learning methods using spectralis OCT for diagnosis and disability progression prognosis in multiple sclerosis. *Ann. Biomed. Eng.* 50 (5), 507–528. May.
- Murphy, O.C., Kwakyi, O., Iftikhar, M., Zafar, S., Lambe, J., Pellegrini, N., et al., 2020. Alterations in the retinal vasculature occur in multiple sclerosis and exhibit novel correlations with disability and visual function measures. *Mult. Scler. J.* 26 (7), 815–828. Jun.
- Nabizadeh, F., Ramezannezhad, E., Kargar, A., Sharafi, A.M., Ghaderi, A., 2022. Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta-analysis. *Neurol. Sci. Off J Ital Neurol Soc Ital Soc Clin Neurophysiol* Oct 27.
- Ortiz, M., Mallen, V., Boquete, L., Sánchez-Morla, E.M., Cordon, B., Vilades, E., et al., 2023. Diagnosis of multiple sclerosis using optical coherence tomography supported by artificial intelligence. *Mult. Scler. Relat. Disord.* 74, 104725. Jun.
- Pérez del Palomar, A., Cegoñino, J., Montolio, A., Orduna, E., Vilades, E., Sebastián, B., Bhattacharya, S., et al., 2019. Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques. editor *PLoS One* 14 (5), e0216410. May 6.
- Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: feature learning by inpainting. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Las Vegas, NV, USA. IEEE, pp. 2536–2544 [cited 2023 Jul 25] Available from: <http://ieeexplore.ieee.org/document/7780647/>.
- Petzold, A., Balcer, L.J., Calabresi, P.A., Costello, F., Frohman, T.C., Frohman, E.M., et al., 2017. Retinal layer segmentation in multiple sclerosis: a systematic review and meta-analysis. *Lancet Neurol.* 16 (10), 797–812. Oct.
- Quigley, H.A., Addicks, E.M., Green, W.R., 1960. Optic nerve damage in human glaucoma. III. Quantitative correlation of nerve fiber loss and visual field defect in glaucoma, ischemic neuropathy, papilledema, and toxic neuropathy. *Arch. Ophthalmol.* 100 (1), 135–146. Chic Ill1982 Jan.
- Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., Kording, K.P., 2017. The Need to Approximate the Use-Case in Clinical Machine Learning, 6. *GigaScience* [Internet].

- May 1 [cited 2023 Jul 3] Available from: <https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/gix019/3071704>.
- Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* 2 (3), 160.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* 128 (2), 336–359. Feb.
- Sidey-Gibbons, J.A.M., Sidey-Gibbons, C.J., 2019. Machine learning in medicine: a practical introduction. *BMC Med. Res. Methodol.* 19 (1), 64. Dec.
- Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition [Internet]. arXiv; 2015 [cited 2023 May 16]. Available from: <http://arxiv.org/abs/1409.1556>.
- Spain, R.I., Liu, L., Zhang, X., Jia, Y., Tan, O., Bourdette, D., et al., 2018. Optical coherence tomography angiography enhances the detection of optic nerve damage in multiple sclerosis. *Br. J. Ophthalmol.* 102 (4), 520–524. Apr.
- Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the Inception Architecture For Computer Vision [Internet]. arXiv; 2015 [cited 2023 May 10]. Available from: <http://arxiv.org/abs/1512.00567>.
- Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., et al., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 17 (2), 162–173. Feb.
- Ulusoy, M.O., Horasanlı, B., Işık-Ulusoy, S., 2020. Optical coherence tomography angiography findings of multiple sclerosis with or without optic neuritis. *Neurol. Res.* 42 (4), 319–326. Apr 2.
- Wisely, C.E., Wang, D., Henao, R., Grewal, D.S., Thompson, A.C., Robbins, C.B., et al., 2022. Convolutional neural network to identify symptomatic Alzheimer's disease using multimodal retinal imaging. *Br. J. Ophthalmol.* 106 (3), 388–395. Mar.
- Zhang, J., Wang, J., Wang, C., Delgado, S., Hernandez, J., Hu, H., et al., 2020. Wavelet features of the thickness map of retinal ganglion cell-inner plexiform layer best discriminate prior optic neuritis in patients with multiple sclerosis. *IEEE Access* 8, 221590–221598.