machine learning & knowledge extraction

MDPI

Article

KhayyamNet: A Parallel Multiscale Feature Fusion Framework for Accurate Diagnosis of Multiple Sclerosis and Myelitis

Mahshid Dehghanpour [1], Mansoor Fateh [1,*], Zeynab Mohammadpoory [2] and Saideh Ferdowsi [3,*]

[1] Faculty of Computer Engineering, Shahrood University of Technology, Shahrood 3619995161, Iran; mahshid.dehghanpour@shahroodut.ac.ir
[2] Faculty of Electrical Engineering, Shahrood University of Technology, Shahrood 3619995161, Iran; z.mohammadpoory@shahroodut.ac.ir
[3] School of Mathematics, Statistics and Actuarial Science, University of Essex, Colchester CO4 3SQ, UK
* Correspondence: mansoor_fateh@shahroodut.ac.ir (M.F.); s.ferdowsi@essex.ac.uk (S.F.)

Abstract

Multiple Sclerosis (MS) and Myelitis are serious inflammatory spinal cord disorders with overlapping clinical symptoms and radiological characteristics, making accurate differentiation challenging yet clinically essential. Early and precise diagnosis is critical for guiding treatment strategies and improving patient outcomes. In this study, we propose KhayyamNet, a novel hybrid deep learning architecture designed to fuse complementary local and global representations for the accurate diagnosis of MS and Myelitis using spinal MRI. To improve robustness and generalization capability, a comprehensive preprocessing strategy including data augmentation and intensity normalization is also applied to reduce noise and address data variability. The proposed architecture combines three complementary deep learning models for feature extraction composed of Xception for high-level semantic features, Convolutional Neural Networks (CNNs) for fine-grained local patterns, and Vision Transformers (ViTs) for global contextual representations via attention mechanisms. Extracted features are then fused and refined using the Minimum Redundancy Maximum Relevance (MRMR) algorithm to eliminate redundancy and retain the most informative signals. Finally, a Random Forest (RF) classifier utilizes the optimized feature set to achieve accurate and robust differentiation between MS, Myelitis, and control spinal MRIs. Experimental results demonstrate that KhayyamNet outperforms existing methods by achieving an average classification accuracy of $98.15 \pm 0.80\%$. This framework demonstrates promising performance for the automated analysis of spinal MRIs and shows potential to assist in the differentiation of MS and Myelitis. While these findings highlight the potential of KhayyamNet for automated MRI interpretation, its evaluation is limited to a single-center dataset, and further validation on external multi-center data is required.

Keywords: multiple sclerosis; Myelitis; spinal MRI; deep learning; vision transformers; MRMR algorithm

Academic Editor: Vincenza Gianfredi

Received: 12 December 2025
Revised: 26 February 2026
Accepted: 28 February 2026
Published: 5 March 2026

1. Introduction

Multiple Sclerosis and Myelitis are two serious neurological disorders that affect the spinal cord and central nervous system. Both disorders often present with similar clinical symptoms such as motor weakness, sensory dysfunction, and inflammation [1]. Accurate and early differentiation between these two diseases is necessary for effective treatment strategies, as their medication approaches and clinical implications differ significantly [2]. Magnetic resonance imaging, especially spinal MRI, has become a vital diagnostic tool

for visualizing structural changes and lesion patterns associated with these disorders. However, the visual similarity of MS and Myelitis symptoms appeared in MRI poses a challenge even for experienced radiologists, highlighting the urgent need for intelligent, automated diagnostic systems [3]. In recent years, various machine learning and deep learning methods have been employed to detect MS and Myelitis using neuroimaging data. Traditional approaches have relied heavily on handcrafted features such as texture, edge, and intensity-based descriptors combined with conventional classifiers like Support Vector Machines (SVMs) or Random Forests (RFs) [4–6]. While these methods have demonstrated promising results, their performance heavily depends on the quality of feature engineering, and they often fail to capture the complex and hierarchical representations inherent in MRI data. More recently, deep learning models such as Convolutional Neural Networks (CNN) and ResNet-based architectures have been applied to automate feature extraction and classification [7,8]. Deep learning models significantly improve diagnostic accuracy by learning discriminative features directly from the raw image data. Additionally, some studies have explored transfer learning and ensemble learning techniques to enhance robustness and generalizability [9–11]. However, most of these methods rely on a single feature extraction strategy, which may limit the diversity and richness of the learned representations. Despite advancements, several challenges persist in this domain. Firstly, the number of publicly available spinal MRI datasets specific to MS and Myelitis is limited, and existing datasets often suffer from class imbalance and low sample diversity [12]. Secondly, current models frequently overlook the integration of both local and global features, which are essential for distinguishing subtle differences between the two disorders. Furthermore, high-dimensional features extracted by deep networks may introduce redundancy and noise, affecting the overall classification performance [12]. Overall, the existing literature reveals a clear research gap: currently, there is no unified framework that simultaneously captures complementary local, spatial, and global contextual features from spinal MRI while incorporating an effective feature selection mechanism to remove redundant information. Addressing this gap is the motivation behind our proposed hybrid deep learning framework, KhayyamNet, which utilizes the complementary strengths of Xception, CNN, and ViT architectures in a parallel configuration. Xception captures high-level hierarchical patterns, CNN focuses on detailed local structures, and ViT models long-range dependencies and global context. Then the extracted features are concatenated and are refined using the MRMR algorithm to preserve only the most relevant and non-redundant features. Feature extraction and selection followed by a RF classifier to ensure high accuracy, generalizability, and stability in distinguishing between MS and Myelitis. The main contributions of our work include the following:

- Designing a parallel feature extraction framework (KhayyamNet): parallel extraction of spatial, local, and global features using Xception, CNN, and ViT networks.
- Fusing and optimizing the extracted features: effective integration of diverse features followed by MR-based feature selection to enhance feature relevance and to reduce dimensionality.

2. Related Work

To provide a clearer structure to the literature review, we divide this section into two main sub-sections. First, we investigate state-of-the-art hybrid architectures that integrate CNNs and Transformers in medical imaging, highlighting their success and relevance. Second, we focus on approaches specifically aimed at differentiating MS and Myelitis through machine learning to contextualize the unique contribution of KhayyamNet.

2.1. Deep Learning in Medical Image Analysis Using Hybrid Architecture

Deep learning has significantly advanced medical image analysis by enabling the automatic extraction of both low-level and high-level features from complex imaging data. CNNs are highly effective at capturing fine-grained local patterns, whereas ViTs excel in modelling long-range dependencies and global contextual relationships. Recent studies have shown that hybrid models combining these architectures achieve superior performance, particularly in tasks requiring both detailed texture analysis and global contextual understanding. Ince et al. [13] proposed a novel U-Net architecture for cerebral vascular occlusion segmentation, enhanced with ConvNeXtV2 blocks and Gated Residual Networks-based Multilayer Perceptron (MLP). Evaluated on the ISLES 2022 dataset, the model achieved an IoU of 0.8015 and a Dice coefficient of 0.8894, outperforming existing U-Net variants. Pacal et al. [14] proposed a hybrid deep learning model for automated Colorectal cancer detection that combines InceptionNeXt blocks to capture local features such as nuclear morphology and glandular structures, enhanced Swin Transformer blocks to model long-range dependencies, and a Residual MLP to refine feature representation. Evaluated on two benchmark datasets, their model achieved accuracies of 99.96% and 99.06%. Pacal et al. [15] proposed using the InceptionNeXt-Transformer for breast cancer diagnosis. They combined InceptionNeXt blocks and ViT-based self-attention for extracting multi-scale features and capturing global dependencies. Their model achieved accuracies of 100% and 98.25% on seven datasets across histopathology, mammography, and ultrasound modalities, for binary and multi-class classification, respectively. Aruk et al. [16] proposed a hybrid CNN–ViT model for skin cancer classification that combines ConvNeXt blocks to extract detailed local features with Transformer blocks to capture global contextual relationships. Their model achieved accuracies of 94.30% and 92.50% on the HAM10000 and ISIC 2019 datasets, respectively.

2.2. Deep and Machine Learning Approaches for Differentiating MS and Myelitis

This section provides a comprehensive review of recent studies that are relevant to diagnosing MS and Myelitis using MRI images. To enhance clarity, we categorized the reviewed methodologies into two distinct groups, namely machine learning- and deep learning-based studies. A summary of the categories is provided in Tables 1 and 2.

Table 1. Comparison of machine learning-related works.

Ref	Method	Dataset
[17]	3D Discrete Wavelet Transform	3D MRI Dataset
[18]	Ensemble Classifier	MRI Dataset
[19]	Multi-parametric MRI with trustworthy machine learning classification	Multi-parametric MRI data from CNS demyelinating disease patients
[20]	Algorithms with Multi-modal Data Fusion	Multi-modal data involving patients with neuromyelitis optica and multiple sclerosis

Table 2. Comparison of deep learning-related works.

Ref	Method	Dataset
[21]	CNN	MRI Data
[22]	MultiResUNet and DenseNet121	MRI Data
[23]	Transformer-based Deep Learning	Neuroimaging Dataset
[24]	CNN	Noncontrast MRI Dataset

Table 2. *Cont.*

Ref	Method	Dataset
[25]	CNN	Conventional MRI images from patients with MS mimics
[26]	Layer-wise Relevance Propagation (LRP)	Conventional MRI data for diagnosing multiple sclerosis
[27]	Deep Learning Model	MRI data from patients with MS and neuromyelitis optica spectrum disorder
[28]	CNN	Multi-parametric quantitative MRI data from patients with MS and NMOSD
[29]	Exemplar MobileNetV2-based Model	MRI images from patients diagnosed with MS
[30]	Grad-CAM (Gradient-weighted Class Activation Mapping)	Clinical brain MRI data for classifying different types of multiple sclerosis
[31]	Combination of the U-Net backbone with 3D CNN	ISBI2015

2.2.1. Machine Learning Methods

Acar et al. [17] utilized 3D discrete wavelet transform (DWT) to extract features from 3D MRIs representing MS. They employed six machine learning techniques to classify the features extracted by ten different DWT wavelet families. The highest F1-score, precision, and recall of 95.0% were obtained by the SVM using the SYM4, SYM8, and Haar wavelet families. Jain et al. [18] presented an ensemble learning-based classification method to distinguish MS from healthy based on their MRIs. The approach involves extracting features from images using 18 Gray Level Co-occurrence Matrix based characteristics. Classification is performed using decision tree (DT) based ensemble learning and three boosting techniques and could obtain an accuracy of 94.91%. Huang et al. [19] focused on the diagnostic challenges of differentiating MS from Neuromyelitis Optica (NMO). They assessed the effectiveness of quantitative radiomic features extracted from brain white matter lesions in facilitating the differentiation. In this study, a Multi-parametric Multivariate RF (MM-RF) was used for classification, which demonstrated an accuracy of 87.1% and AUC of 0.902. Eshaghi et al. [20] evaluated a multi-kernel learning approach for automatic diagnosis using multi-modal data fusion. They utilized modalities included T1-weighted high-resolution scans, diffusion tensor imaging (DTI), and resting-state functional MRI (fMRI). This research proposed 18 predictors from neuroimaging, clinical, and cognitive measures and achieved an accuracy of 84% in distinguishing between MS, NMO and healthy subjects.

2.2.2. Deep Learning Methods

Storelli et al. [21] developed a deep learning system that predicts MS progress over a two-year follow-up using MRI of MS patients. They proposed a CNN-based architecture specifically designed for predicting illness progress. The performance of their suggested model was assessed by two expert clinicians. Krishnamoorthy et al. [22] developed a pipeline for segmenting and classifying spinal cord lesions using the 2D networks MultiResUNet and DenseNet121. The experimental outcome of this study provided an accuracy of 98% without the skull. Huang et al. [23] employed an advanced transformer-based deep learning architecture to differentiate between MS and NMOSD utilizing various imaging sequences (i.e., coronal and sagittal T2-weighted fluid-attenuated inversion recovery) and different anatomical regions (i.e., brain and spinal cord). A model that combined brain and spinal cord MRI achieved the best overall performance, with the highest accuracy of 81.4%. Narayana et al. [24] used CNN to classify growing lesions in MRI images for the diagnosis of MS. Each image segment was processed individually for classification, and participant

predictions were generated by aggregating the cutoff scores with a fully connected network. The model achieved a slice-wise sensitivity of 78 ± 4.3% and specificity of 73 ± 2.7%, while the participant-wise sensitivity and specificity were 72 ± 9.0% and 70 ± 6.3%, respectively. Rocca et al. [25] developed a method for the automated classification of MS and its mimics. They collected 268 brain MRI scans and trained a neural network using 178 of them. The results showed that the proposed model outperformed the experts, particularly in diagnosing MS, while struggling with NMOSD. Their findings suggest that deep learning could enhance diagnostic accuracy for white matter disorders. Eitel et al. [26] introduced a transparent deep learning framework using 3D CNNs and layer-wise relevance propagation (LRP) for MS diagnosis. They pre-trained a CNN on Alzheimer's MRI data before fine-tuning it to distinguish between MS patients and healthy controls. Their model achieved a balanced accuracy of 87.04% and an AUC of 96.08%. Seok et al. [27] focused on creating a deep learning model to distinguish between MS and NMOSD using brain MRI data. The model utilized a modified ResNet18 CNN, trained on 5-channel images derived from five 2D slices of 3D FLAIR images, achieving an accuracy of 76.1%, with a sensitivity of 77.3% and specificity of 74.8%. Hagiwara et al. [28] developed a CNN model to distinguish between MS and NMOSD. Their model achieved sensitivity rates of 80.0% for MS and 83.3% for NMOSD, and an overall accuracy of 81.1%. Ekmekyapar et al. [29] combined the MobileNetV2 network with exemplar-based learning, MRMR feature selection, and KNN for MS diagnosis. The proposed model achieved accuracy rates of 99.76% for axial, 99.48% for sagittal, and 98.02% for hybrid images of dataset 1. The model reached 100% accuracy on dataset 2. Zhang et al. [30] analyzed six different CNN models trained to classify brain MRI scans into three categories, including relapsing-remitting MS (RRMS), secondary progressive MS (SPMS), and healthy controls. Grad-CAM demonstrated the best localizing ability for heatmaps, while the VGG19 model with a global average pooling layer and pretrained weights achieved the highest classification performance. Notably, the 95th percentile values of Grad-CAM in SPMS were significantly higher than those in RRMS, indicating greater heterogeneity in brain regions identified by voxel-wise analysis. Dehghanpour et al. [31], proposed ZechariahNet, a hybrid deep learning model for multiple sclerosis (MS) plaque segmentation in MRI images. Their approach combines a U-Net backbone with 3D CNN components and integrates advanced modules such as Transition Down, Dense blocks, Squeeze Attention, and C-LSTM layers to better capture spatial–temporal features of MS lesions. The inclusion of C-LSTM blocks after attention modules significantly improved segmentation performance. The customized version of ZechariahNet achieved a Dice score of 84.72%, demonstrating a noticeable improvement over previously reported methods, despite the inherent difficulty of distinguishing MS plaques from surrounding normal brain tissue.

3. Proposed Method

This section presents the proposed method for the diagnosis of MS and Myelitis from spinal MRI images. The goal of the proposed approach is to leverage the strengths of deep learning models and feature selection algorithms to enhance diagnostic accuracy. Initially, the MRI images undergo a preprocessing stage that includes data augmentation to address class imbalance and normalization to standardize pixel intensity values. Then, in order to extract meaningful features, three deep networks are employed in parallel. This includes an Xception for high-level feature extraction, a CNN to capture local features, and a ViT to extract global features and long-range dependencies within the images. The features extracted by these three networks are then fused to form a comprehensive feature set. To reduce dimensionality and eliminate redundant or irrelevant features, the MRMR algorithm is applied for an optimal feature selection. Finally, the selected features are fed

into the RF classifier to perform the final classification, enabling accurate detection of MS and Myelitis. The diagram of the proposed method is shown in Figure 1.

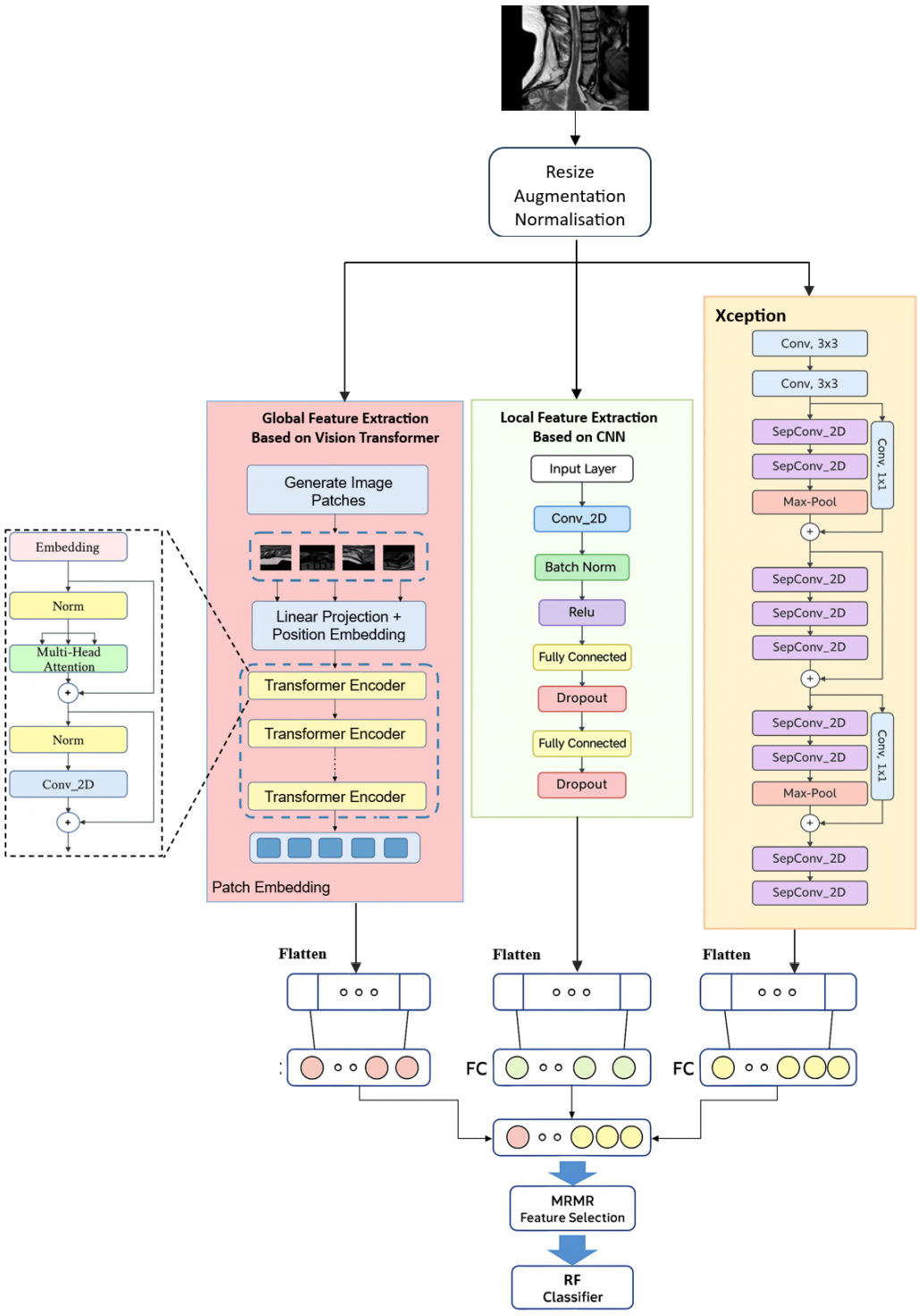


Figure 1. A block diagram of the proposed method.

3.1. Preprocessing

To increase the generalization, robustness and convergence of deep learning models, data augmentation and pixel normalization of MRI images have been applied for preprocessing. Details of the preprocessing pipeline are provided below:

Data augmentation: given that medical datasets are often imbalanced or limited in size, augmentation strategies are applied to increase the diversity of training data. Let I denotes the original image. The augmented image is obtained as:

$$I_{\text{aug}} = T(I) \tag{1}$$

$$T \in \{\text{Rotation, Flipping, Zooming, Shifting, Brightness Adjustment}\}$$

These transformations preserve semantic meaning while enriching feature space. Henceforth, all images including original and augmented images will be presented by I.

Normalization: to ensure consistency and improve the learning stability, each image is normalized as follows:

$$I_{\text{norm}}(x, y) = \frac{I(x, y) - \mu}{\sigma} \tag{2}$$

where $I(x, y)$ is the augmented image, μ is the average pixel intensity and σ is the standard deviation.

3.2. Feature Extraction

To capture diverse and complementary information from spinal MRI images, we adopt a parallel architecture comprising three powerful deep learning models:

- Xception for learning hierarchical features with efficient convolutions.
- Custom CNN for local spatial patterns and fine texture detection.
- ViT for global context via attention mechanisms.

The selection of Xception, CNN, and ViT as parallel feature extraction backbones is motivated by the specific characteristics of spinal cord MRI lesions in MS and Myelitis. MS-related lesions often appear as small, irregular, and spatially localized demyelinated regions. Such fine-grained details can effectively be captured by CNN layers, which excel at detecting local edge-level and texture-level features. On the other hand, Myelitis lesions frequently extend across longer segments of the spinal cord and exhibit diffuse inflammation patterns. These global contextual dependencies are better modeled by ViT, which uses self-attention to capture long-range spatial relationships across the entire spinal cord structure. Complementarily, the Xception network, with its depth-wise separable convolutions, efficiently captures mid-to-high-level semantic representations of structural abnormalities while reducing computational complexity and overfitting risk. The integration of these three models ensures that localized lesion boundaries, global contextual cues, and abstract semantic patterns are simultaneously captured, which is particularly critical given the subtle yet clinically significant radiological differences between MS and Myelitis. Each network contributes distinct perspectives of the data, and their combined representations lead to more accurate diagnosis. The following explains how each network extracts features from images.

Feature extraction using Xception network: The Xception (Extreme Inception) network is a convolutional architecture that leverages depth-wise separable convolutions to improve efficiency and reduce redundancy in feature extraction [32]. In this design, a standard convolution is factorized into two operations: (i) a depth-wise convolution, which applies a spatial filter to each input channel separately, and (ii) a pointwise convolution, which combines the outputs across all channels. Formally, for an input image tensor $I \in \mathbb{R}^{H \times W \times C}$, depthwise convolution is applied to each channel independently, followed by a 1×1 pointwise convolution to aggregate the features:

$$F_{\text{out}} = \text{ReLU}(\text{BN}(\text{PWConv}(\text{DWConv}(F_{\text{in}})))) \tag{3}$$

where DWConv and PWConv denote depthwise and pointwise convolutions, respectively, and BN indicates batch normalization. In the context of spinal MRI, this architecture is advantageous because it captures multi-scale and hierarchical lesion patterns with fewer parameters, which reduces the risk of overfitting while maintaining discriminative ability. MS lesions typically appear as small, irregular plaques, while Myelitis lesions can extend longitudinally. By decomposing convolutions, Xception can effectively learn both localized and abstract structural variations relevant for differentiating these conditions. After multiple convolutional blocks, a Global Average Pooling (GAP) layer is applied to compress each feature map into a single scalar:

$$f = \frac{1}{H' \times W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} F_{\text{out}}(i, j) \tag{4}$$

where F_{out} is the final output feature map. This step reduces dimensionality while retaining essential information, producing a compact feature vector $f \in \mathbb{R}^{C'}$. Overall, the Xception branch in KhayyamNet provides an efficient high-level feature representation that complements local features from CNN and global contextual features from ViT. The overall structure of the Xception network is shown in Figure 2.

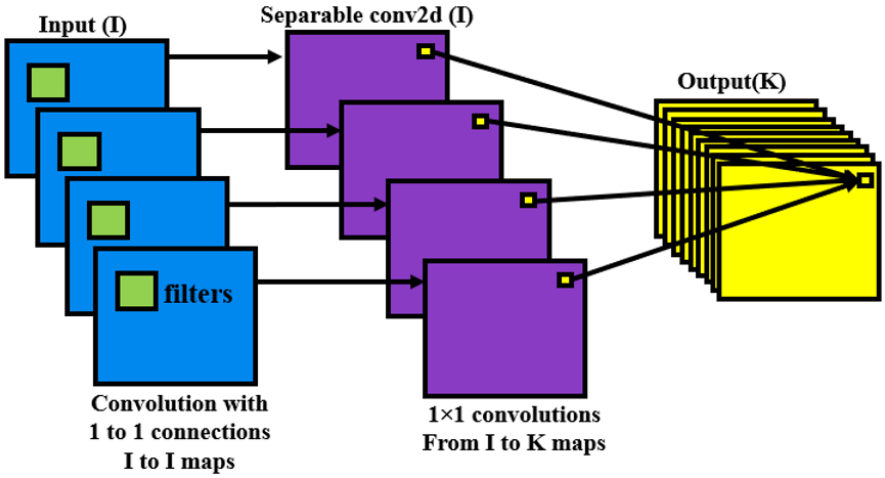


Figure 2. A schematic illustration of the Xception Architectural [33].

Feature extraction using a custom CNN: CNNs are particularly effective for capturing local spatial features in medical images, such as edges, lesion boundaries, and fine structural variations [34]. In this study, the CNN branch was designed to emphasize localized lesion features that may be less visible to high-level architectures like Xception or ViT. A standard convolution operation applies a kernel weight for the k-th filter W_k of size $K_h \times K_w$ to the input image I, generating a feature map:

$$F_k(i, j) = \sum_{m=1}^{K_h} \sum_{n=1}^{K_w} \sum_{c=1}^{C} W_k(m, n, c) \cdot I(i + m, j + n, c) + b_k \tag{5}$$

where b_k is the bias term or the k-th filter. This operation captures spatial correlations in small neighborhoods, which is essential for identifying focal MS plaques and other localized abnormalities. Following convolution, non-linear activation functions are applied to introduce sparsity and enhance representational power. In our model, the Rectified Linear Unit (ReLU) was used. This ensures efficient training and alleviates vanishing gradient problems. In addition, Batch Normalization (BN) was applied to stabilize learning and accelerate convergence by normalizing feature distributions:

$$BN(x) = \gamma \cdot \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta \tag{6}$$

where μ and σ^2 are the batch mean and variance, and γ, β are learnable parameters. To further refine the representation, pooling layers reduce spatial dimensions while retaining the most salient features. For example, max-pooling over a $p \times p$ window is expressed as

$$P(i, j) = \max_{(m,n) \in p \times p} F(i + m, j + n) \tag{7}$$

This operation improves translational invariance and reduces overfitting, which is crucial in spinal MRI where lesions may vary in size, shape, and orientation. Through successive convolution–BN–ReLU–pooling stages, the CNN branch learns fine-grained lesion characteristics that complement the high-level semantic features extracted by Xception. Finally, a Global Average Pooling (GAP) layer is applied to condense the multi-dimensional feature maps into a compact vector representation. This operation reduces dimensionality without introducing additional trainable parameters, thereby ensuring computational efficiency. The resulting CNN feature vector is then passed into the fusion stage of KhayyamNet, where it contributes detailed local information to the combined representation. Together with the abstract features from Xception and the global contextual features from ViT, the CNN-derived features enhance the model's discriminative ability for differentiating MS from Myelitis. Figure 3 shows the architecture of the proposed CNN used in this paper.

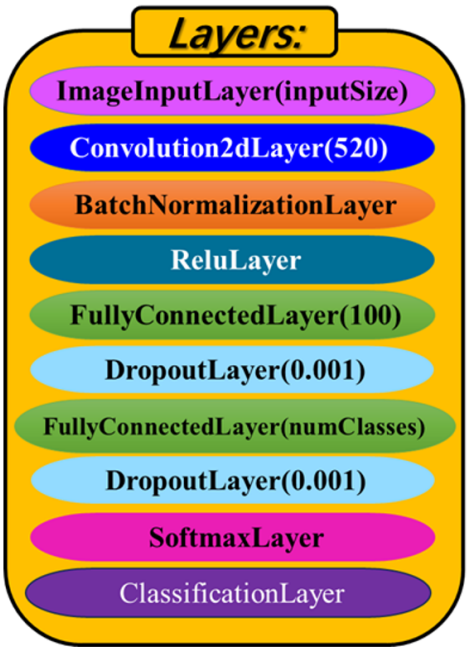


Figure 3. The architecture of the proposed CNN.

Feature extraction using ViT: The ViT is a powerful deep learning architecture inspired by the transformer models used in Natural Language Processing (NLP) [35]. Unlike CNNs, which use convolutional filters, ViT processes images by splitting them into patches, embedding them, and modeling global dependencies using self-attention mechanisms. This makes ViT ideal for capturing global contextual features from medical images [35]. Figure 4 represents the ViT's structure, and the details of this structure are illustrated as follows.

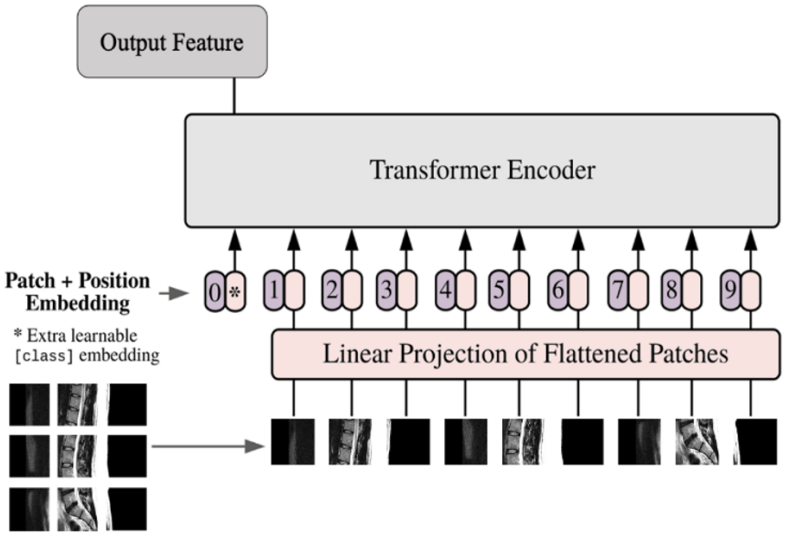


Figure 4. The structure of the ViT network inspired from [35].

3.2.1. Input Image and Patch Embedding

Let the input MRI image be

$$I \in \mathbb{R}^{H \times W \times C} \tag{8}$$

The image is divided into non-overlapping patches of size $P \times P$, resulting in

$$N = \frac{HW}{P^2} \tag{9}$$

where N is the total number of patches such that each patch is flattened into a vector $x_i \in \mathbb{R}^{P^2 \cdot C}$. These patch vectors are then linearly projected into a D-dimensional embedding space:

$$z_0^i = x_i \cdot E + b \quad \text{for } i = 1, \ldots, N \tag{10}$$

where $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$ is the learnable projection matrix, $z_0^i \in \mathbb{R}^D$is the initial patch embedding, and b is the bias term.

3.2.2. Positional Embedding

Since Transformers have no inherent understanding of spatial order, positional embeddings are added to the patch embeddings:

$$z_0^i = z_0^i + p_i \tag{11}$$

where $p_i \in \mathbb{R}^D$ is a learnable positional embedding for patch i.

3.2.3. Transformer Encoder Layers

Each transformer encoder block consists of the following:

1. Multi-Head Self-Attention (MHSA)
2. Layer Normalization (LN)
3. Feed-Forward Network (FFN)
4. Residual Connections

Let us denote the patch embeddings at layer l as $Z^{(l)} \in \mathbb{R}^{N \times D}$.

(a) Multi-Head Self-Attention (MHSA): each patch attends to all other patches via the attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{12}$$

where the following are used:

- $Q = Z^{(l)}W_Q, \quad K = Z^{(l)}W_K, \quad V = Z^{(l)}W_V$.
- $W_Q, W_K, W_V \in \mathbb{R}^{D\times d_k}$ are learnable matrices.
- d_k denotes the dimension of each head.

The multi-head version concatenates outputs from h attention heads is

$$\mathrm{MHSA}(Z^{(l)}) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W_O \tag{13}$$

(b) Layer Normalization (LN): after the attention operation, layer normalization is employed to standardize feature distributions, which stabilizes optimization and facilitates efficient training of deep transformer layers.

(c) Feed-Forward Network (FFN): after MHSA, a two-layer MLP is applied:

$$\mathrm{FFN}(x) = W_2 \cdot \sigma(W_1 x + b_1) + b_2 \tag{14}$$

where σ is typically the GELU activation function.

(d) Residual Connections and Layer Normalization: each transformer block is wrapped with residual connections:

$$Z^{(l+1)} = \mathrm{LN}\big(Z^{(l)} + \mathrm{MHSA}(Z^{(l)})\big) \tag{15}$$

$$Z^{(l+2)} = \mathrm{LN}\big(Z^{(l+1)} + \mathrm{FFN}(Z^{(l+1)})\big) \tag{16}$$

3.2.4. Global Feature Vector

In ViT, the CLS token (an abbreviation for "classification token") is a learnable embedding that is prepended to the sequence of patch tokens. This specialized token is designed to capture and consolidate global contextual information from the input image, thereby facilitating image classification. The final output of this token is

$$f_{\mathrm{ViT}} = Z^{(L)}_{[\mathrm{CLS}]} \in \mathbb{R}^D \tag{17}$$

where L is the number of Transformer layers. The vector f_{ViT} is the global feature representation of the input MRI image, capturing contextual relationships across all patches. The overall advantages of using the ViT network for feature extraction include its ability to capture spatially separated lesion patterns, learn image-level attention across the entire spinal cord, and effectively detect complex disease markers that are not confined to a single region.

Feature fusion: the feature vectors extracted by the three models are concatenated:

$$F_{\mathrm{fused}} = [\, F_X \parallel F_{\mathrm{CNN}} \parallel F_{\mathrm{ViT}} \,] \tag{18}$$

where $|.|$ denotes vector concatenation. By concatenating the features extracted by three proposed models, hierarchical, local, and global features are combined, leading to the enhancement of the discriminative power of the feature space and enabling robust classification under diversity in MRI patterns.

3.3. Feature Selection with MRMR

Regarding feature selection, the MRMR algorithm [36] was chosen because it directly balances two essential criteria: maximizing the relevance of features with respect to class labels while minimizing redundancy among selected features. This is especially important in medical imaging, where deep learning models often produce high-dimensional representations with strong inter-feature correlations. Unlike PCA, which focuses solely

on variance without accounting for class separability, or LASSO, which may eliminate clinically important yet correlated features, MRMR employs an information-theoretic criterion that preserves the most discriminative and non-redundant features. Compared to wrapper-based methods, MRMR is computationally more efficient and scalable to the large feature sets produced by deep neural networks. These properties make MRMR particularly well-suited for refining fused multi-scale features in the KhayyamNet framework. This has been done through the following steps:

Let S be the selected subset from the full feature set F_{fused}, and y be the class label.

- Relevance is defined as the mutual information between feature f_i and the class label y:

$$D = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; y) \tag{19}$$

- Redundancy is defined as the average mutual information between features:

$$R = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \tag{20}$$

The objective of the MRMR algorithm is to select features with high relevance and low redundancy. This can be formulated by the difference of relevance and redundancy:

$$\text{MRMR}(f_i) = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; y) - \frac{1}{|S|^2} \sum_{f_j \in S} I(f_i; f_j) \tag{21}$$

The features with the highest MRMR scores are iteratively selected until a predefined threshold k is reached. The final output is a reduced set of discriminative and non-redundant features:

$$F_{\text{sel}} = [\, f_{i_1}, f_{i_2}, \ldots, f_{i_k} \,] \quad \text{with } k < n \tag{22}$$

The selected feature set is then passed to the RF classifier for training and predicting labels. The overall advantages of the MRMR feature selection algorithm include reducing feature dimensionality, improving model generalization, and eliminating noisy or irrelevant features. MRMR feature selection was performed using MATLAB's fscmrmr function. For continuous-valued deep features extracted from CNN, Xception, and ViT, the function internally estimates mutual information using its default procedure, which discretizes continuous features to compute the mutual information with class labels.

3.4. Classification

In the classification stage, the RF classifier was utilized to perform the final classification based on the selected features, due to its robustness, ability to handle high-dimensional feature spaces, and reliable performance in complex medical imaging tasks.

4. Results

This section outlines the implementation details, evaluation criteria, and analysis of the experimental results based on the MRI dataset provided by Elazig Fırat University [11]. The proposed KHAYYAMNET model was developed using MATLAB 2024 on a system featuring an Intel Core *i*9 processor, 64 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU with 24 GB of memory to assess performance efficacy.

4.1. Dataset

In this study, we used a spinal MRI dataset provided by Elazig Fırat University [11]. The collection of MR images was approved by the Ethics Committee of Elazığ Fırat Uni-

versity, and informed consent was obtained from all subjects before data collection. This dataset comprises a total of 2746 MRI images from 409 individuals and is classified into three categories: Myelitis, MS, and a healthy control groups based on the size of spinal lesions observed in each subject. The dataset includes 706 MRI images associated with MS patients, 667 images related to Myelitis, and 1373 images from healthy individuals. Figure 5 illustrates two sample MRI images from each category.

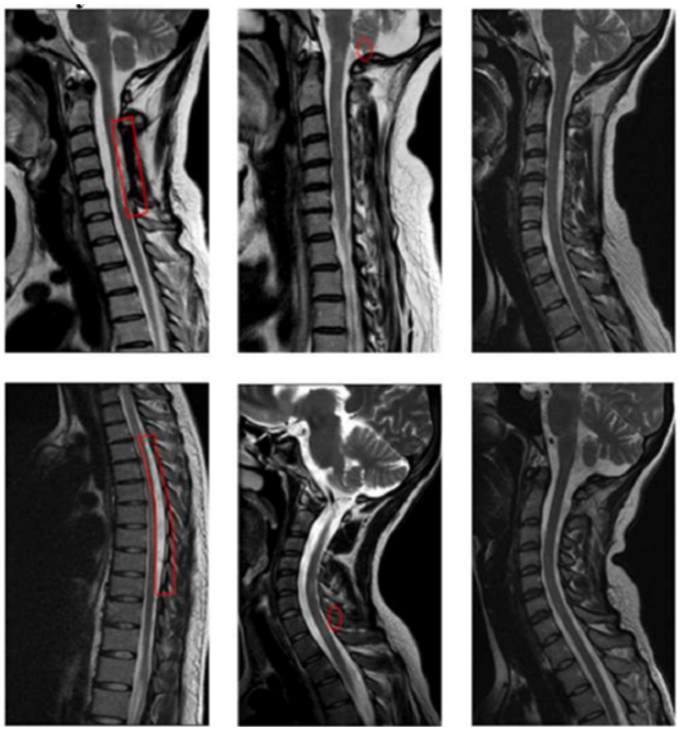


Figure 5. Sample images from the dataset provided in [11]. **Left**, **middle**, and **right** columns represent two random samples of Myelitis, MS, and Control groups, respectively. Red boxes highlight representative regions of interest within the spinal cord, emphasizing lesion-related areas for visual comparison between Myelitis, MS, and Control samples.

4.2. Evaluation Metrics

The performance evaluation metrics employed in this study, which also facilitate comparisons with other methodologies, include accuracy, precision, recall, and F1-score. The mathematical formulations for these metrics are provided below. In the Equations (23)–(26), TP (True Positive) refers to the number of MS samples that were correctly classified as MS, TN (True Negative) represents the number of non-MS samples (i.e., Myelitis or healthy) that were correctly classified as non-MS (i.e., Myelitis or healthy), FP (False Positive) denotes the number of non-MS samples (i.e., Myelitis or healthy) that were incorrectly classified as MS, and FN (False Negative) indicates the number of MS samples that were incorrectly classified as non-MS (i.e., Myelitis or healthy).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{23}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{24}$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{25}$$

$$F1_{\text{score}} = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{26}$$

4.3. Simulation Setting

To evaluate the performance of the proposed hybrid deep learning framework for classifying MS and Myelitis using spinal MRI images, several simulation settings and parameter configurations were applied at each stage of the methodology. All MRI images were first normalized to have zero mean and unit variance and were then fed into three mentioned deep feature extraction networks, CNN, Xception, and ViT. The CNN architecture was designed to capture local fine-grained spatial patterns and produced a 100-dimensional feature vector, while the Xception and ViT models were employed to extract higher-level semantic and global contextual information, each yielding a 1000-dimensional feature representation. The features extracted from the three networks were concatenated to form a unified 2100-dimensional fused feature vector for each sample. The hyperparameters of the CNN, Xception, and ViT networks were optimized prior to the final model evaluation and were kept fixed throughout all subsequent experiments. Table 3 reports the final parameter values used for these three networks. To determine suitable hyperparameter configurations, a two-step tuning strategy was adopted. First, a coarse grid search was conducted over a broad range of candidate values, including filter sizes $\{64, 128, 256\}$, attention heads $\{4, 8, 12\}$, hidden dimensions $\{256, 512, 1024\}$, and pooling types $\{\text{max}, \text{average}\}$. Subsequently, a randomized search was applied within the most promising ranges identified in the first step. For each configuration, the average classification accuracy was computed using 10-fold cross-validation on training data only. The final parameter settings were selected based on the highest validation accuracy while maintaining computational efficiency. Importantly, once selected, these parameters were fixed and were not further optimized within the nested cross-validation framework, ensuring that feature extraction remained independent of the final performance evaluation. To rigorously evaluate the proposed framework and to avoid optimistic bias and overfitting, a nested cross-validation strategy was adopted. Specifically, a stratified 10-fold cross-validation scheme was employed in the outer loop, where the entire dataset was partitioned into ten folds while preserving class proportions. In each iteration, one fold was used as the test set, and the remaining nine folds were used for training. Within each outer training set, an inner 5-fold cross-validation loop was applied exclusively to the training data to perform feature selection and to determine the optimal number of features. This design strictly prevents any information leakage from the outer test folds into the feature selection process. To reduce the dimensionality of the fused 2100-dimensional feature space, the MRMR algorithm was employed. Feature subsets consisting of the top-ranked features, as determined by the MRMR relevance–redundancy scores, were evaluated in the inner cross-validation loop across a range of 50 to 1500 features. This range was selected to provide a practical balance between aggressive dimensionality reduction and retaining sufficient discriminative information, while avoiding excessive feature dimensionality that could increase computational cost and risk of overfitting. The optimal number of features was selected based on the highest average classification accuracy achieved across the inner folds, and the corresponding feature indices were then applied unchanged to the associated outer test fold. The dataset is inherently imbalanced across the Healthy, MS, and Myelitis classes. To mitigate the adverse effects of class imbalance, data augmentation was applied only to the training samples within each outer fold, while the test samples were kept completely untouched to ensure an unbiased evaluation on unseen real data. A class-specific augmentation strategy was employed, in which transformations were applied more extensively to the underrepresented MS and Myelitis classes than to the Healthy class. The level of augmentation was carefully controlled to approximately balance the number of training samples across classes without excessively inflating the dataset. For transparency, the average number of training and test samples per class before and after augmentation across

the 10-fold cross-validation procedure is reported in Table 4. During data augmentation, a series of controlled geometric and photometric transformations was applied to the training images to enhance data diversity while maintaining the anatomical integrity of spinal MRI. Random rotations were applied within a range of $\pm 15^{\circ}$. Zoom augmentation was implemented using a scaling factor uniformly sampled from the interval $[0.9, 1.1]$. Horizontal and vertical translations (i.e., shifting) were limited to a maximum of $\pm 10\%$ of the image width and height, respectively. Horizontal flipping was applied with a probability of 0.5, whereas vertical flipping was excluded to prevent anatomically implausible distortions. In addition, brightness adjustment was performed by scaling pixel intensities within the range $[0.8, 1.2]$. These augmentation parameters were empirically selected to enhance model generalization without introducing clinically implausible variations. For classification, the RF classifier based on the bagging paradigm was employed. The RF classifier was used both in the inner cross-validation loop, to evaluate the classification performances of different feature subset sizes, and in the outer loop, to train and test the final model. The model was configured with 200 trees, a minimum leaf size of 1, and Gini's diversity index ('gdi') as the split criterion. The maximum number of splits was left unlimited, and the number of predictors sampled at each split was set to $\sqrt{p}$, where p denotes the number of selected features. Out-of-Bag prediction was enabled to provide an internal estimate of model performance. These parameters were kept fixed throughout all experiments and were not optimized during cross-validation in order to ensure reproducibility and to avoid additional sources of bias. For each outer fold, the trained RF model was evaluated using standard performance metrics, including Accuracy, Precision, Recall, and F1-score. Precision, Recall, and F1-score were first computed separately for each class. To obtain an overall performance measure and properly account for class imbalance, these class-wise metrics were then combined using weighted averaging, where the weights were determined by the number of test samples in each class. Accuracy was calculated over all test samples within each fold and therefore did not require class-wise weighting. All evaluation metrics were computed independently for each of the ten folds and are finally reported as mean $\pm$ standard deviation across folds. This evaluation strategy provides a reliable and robust estimate of the model's overall performance as well as its stability across different data splits. In addition, confusion matrices were accumulated over all folds to visualize class-wise prediction behavior. Receiver Operating Characteristic (ROC) curves were generated using a one-vs.-all strategy, and class-specific Area Under the Curve (AUC) values were reported separately for the Healthy, MS, and Myelitis classes, providing a detailed assessment of discriminative performance for each class.

Table 3. The values of the parameters of the deep networks used for feature extraction.

Network	Parameter	No. of Parameters	Purpose
CNN	No. of Filters	20	To capture local spatial patterns
CNN	No. of outputs	100	Captures high-level abstract features
ViT	No. of Transformer	12	Captures global dependencies using attention mechanisms
ViT	No. of Attention Heads	8	Attention mechanism to focus on key regions of the image
ViT	Hidden Size	512	Size of the hidden representation in each layer
ViT	No. of Tokens	16 ($patchsize : 16 \times 16$)	Divides the image into patches for processing
Xception	No. of Filters	128	Learn complex hierarchical features
Xception	No. of Depth-wise Convolutions	8 layers	Learns spatially separable features
Xception	Pooling Size	7×7	Reduces dimensionality while retaining important features
Xception	No. of outputs	1000	Abstract representation of image content

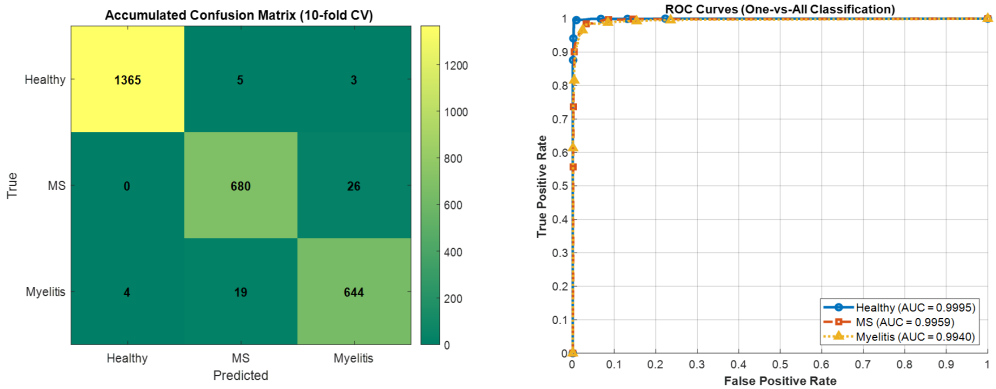
Table 4. The number of training samples per class before and after augmentation, and the number of test samples for each class.

Classes	No. of Training Samples Before Augmentation	No. of Training Samples After Augmentation	No. of Test Samples
Healthy	1236	1910	137
MS	635	1910	71
Myelitis	600	1910	67
Total	2471	5730	275

4.4. Evaluation of Results

This section presents a thorough evaluation of the proposed KHAYYAMNET model's performance in distinguishing MS, Myelitis and healthy condition using spinal MRI images. The dataset includes a diverse range of spinal lesions associated with both diseases, as well as images from a healthy control group. Figure 6a presents the accumulated confusion matrix over all folds. The confusion matrix comprises three classes corresponding to healthy samples, MS patients, and Myelitis patients. The values along the main diagonal of the confusion matrix indicate the number of correctly classified samples for each category including 1365 for healthy individuals, 680 for MS, and 644 for Myelitis. The obtained confusion matrix indicates the model's exceptional performance in diagnosing MS and myelitis from spinal MRI images. This is evidenced by the minimal misclassification of only 57 samples out of 2746 total samples, highlighting its strong capability to distinguish among MS patients, myelitis patients, and healthy individuals. Table 5 summarizes the classification performance of the proposed model evaluated using 10-fold cross-validation. For each class, Precision, Recall, and F1-score are reported as mean ± standard deviation across folds, reflecting the stability of the model under different data splits. To account for class imbalance, overall Precision, Recall, and F1-score were computed as weighted averages based on the number of samples in each class. Overall Accuracy was calculated across all test samples in each fold and then summarized as mean ± standard deviation. The model achieved an overall classification accuracy of $98.15 \pm 0.80\%$, along with weighted precision, recall, and F1-score of $98.25 \pm 0.79\%$, $98.15 \pm 0.80\%$, and $98.15 \pm 0.80\%$, respectively. At the class level, the model demonstrated excellent performance for the Healthy class (F1-score: $99.53 \pm 0.46\%$), indicating highly reliable discrimination. Strong and balanced performance was also observed for MS (F1-score: $96.48 \pm 1.66\%$) and Myelitis (F1-score: $95.97 \pm 1.81\%$), highlighting the model's robustness across clinically relevant classes. Overall, these results confirm that the proposed approach achieves high accuracy and consistent class-wise performance, while maintaining robustness against inter-fold variability. Figure 6b illustrates the ROC curves for the three classes, obtained by aggregating prediction scores across all folds of the 10-fold cross-validation procedure. A one-vs-all strategy was adopted for each class, in which the target class was treated as positive and the remaining two classes as negative, enabling class-wise AUC computation and a detailed evaluation of discriminative performance. As shown in the figure, all ROC curves lie very close to the top-left corner, indicating high sensitivity combined with low false positive rates. The curves for Healthy, MS, and Myelitis classes largely overlap, suggesting consistent and balanced performance across all classes without noticeable trade-offs. The proposed model achieves excellent discrimination, with AUC values of 99.95% for Healthy, 99.59% for MS, and 99.40% for Myelitis, resulting in an average AUC of 99.65%. These results demonstrate the robustness and reliability of the model in effectively distinguishing each class from the others over a wide range of decision thresholds. It is important to emphasize that the MRI data analyzed in this study consist of spinal MRI rather than brain MRI. Accordingly, the

regions of interest identified by the attention maps and Grad-CAM visualizations are not expected to correspond to cranial anatomy. Instead, the model predominantly attends to the spinal cord and adjacent tissues, where multiple sclerosis plaques and myelitis-associated inflammatory lesions are known to occur clinically. In certain cases, the activation maps appear spatially diffuse or only partially aligned with well-defined anatomical boundaries. This behavior can be attributed to several contributing factors, including image normalization and resizing procedures, the relatively low contrast between the spinal cord and surrounding tissues, and the global attention mechanism of the Vision Transformer, which distributes focus across broader contextual regions rather than concentrating exclusively on localized edges. Despite these effects, qualitative evaluation demonstrates that the highlighted regions consistently overlap with lesion-prone segments of the spinal cord, particularly areas exhibiting hyperintensity or longitudinal inflammatory patterns. These findings suggest that the model's decision-making process is driven by clinically meaningful features rather than irrelevant background information, thereby supporting the reliability and interpretability of the proposed KhayyamNet framework.



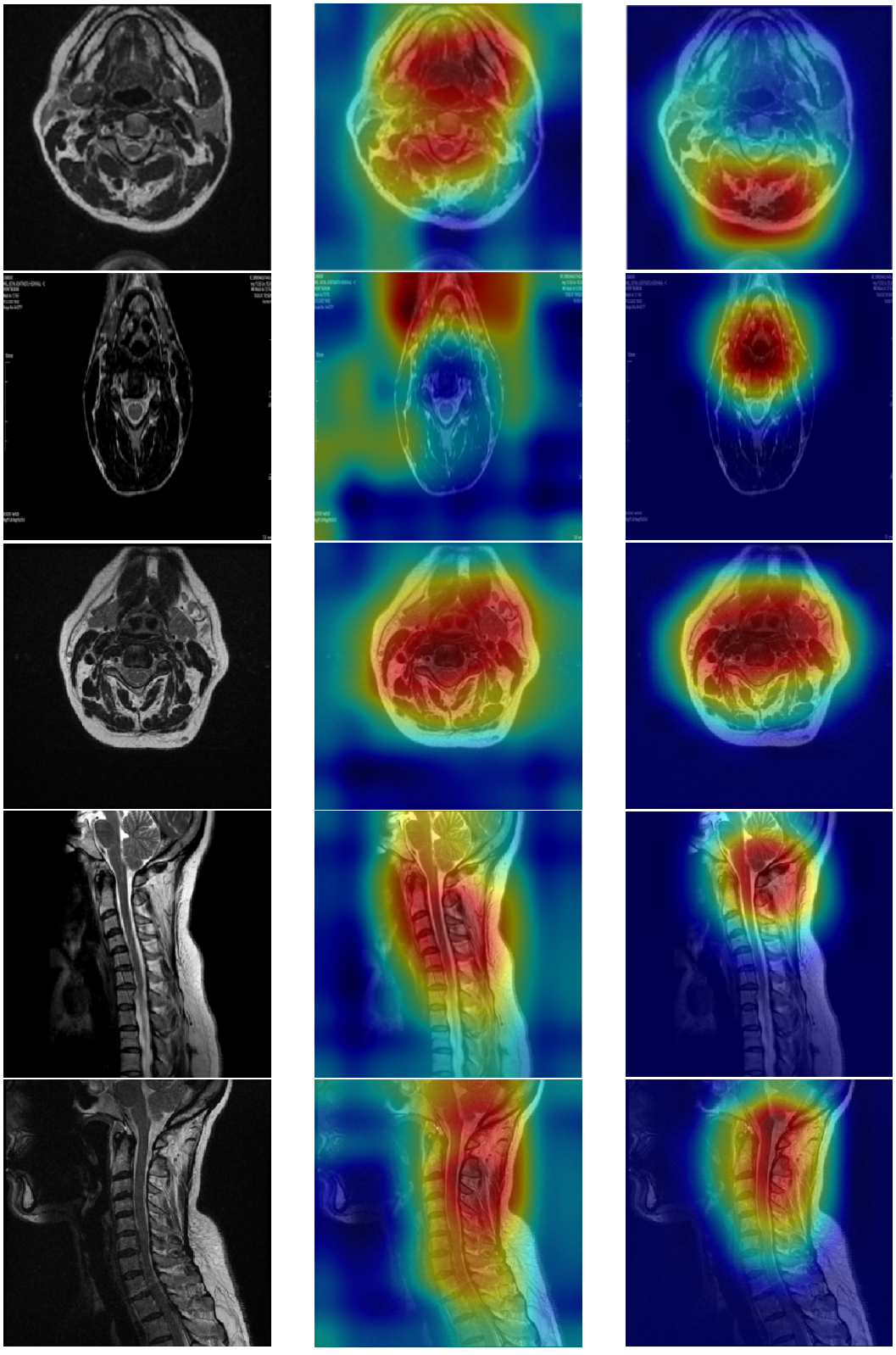
(**a**) (**b**)

Figure 6. (**a**) Confusion matrix of the proposed model for MS, myelitis and healthy condition classification. (**b**) ROC curve of the proposed model for Healthy, MS and Myelitis classification.

Table 5. Mean ± SD of Accuracy, Precision, Recall, and F1-score per class and overall across 10-fold cross-validation.

Class	Precision %	Recall %	F1-Score %	Accuracy %
Healthy	99.64 ± 0.51	99.42 ± 0.67	99.53 ± 0.46	-
MS	95.50 ± 3.47	95.60 ± 2.56	96.48 ± 1.66	-
Myelitis	95.02 ± 2.59	95.05 ± 3.75	95.97 ± 1.81	-
Overall	98.25 ± 0.79	98.15 ± 0.80	98.15 ± 0.80	98.15 ± 0.80

Figure 7 illustrates the interpretability results of the proposed KhayyamNet framework using both attention maps (Figure 7b) and Grad-CAM visualization (Figure 7c) for five the spinal MRI test images. The attention map highlights the regions automatically emphasized by the Vision Transformer branch, while Grad-CAM provides a convolution-based saliency representation of the CNN and Xception branches. As can be observed, both visualization techniques consistently focus on the lesion-prone areas of the spinal cord, particularly the hyperintense regions that indicate demyelination in MS or longitudinal inflammation in Myelitis. Importantly, the highlighted zones correspond closely with radiologically relevant patterns, suggesting that the model's decision-making process is guided by clinically meaningful features rather than irrelevant background regions. These results improve the transparency of KhayyamNet and support its potential for clinical adoption by providing visual evidence of model interpretability.



(**a**) (**b**) (**c**)

Figure 7. Axial and sagittal views of (**a**) original test image, (**b**) their corresponding attention map, and (**c**) Grad-Cam samples of five different observations. The color scale indicates the relative contribution of image regions to the model's prediction, where warm colors (red–yellow) denote highly discriminative lesion-related spinal cord areas, while cool colors (blue) represent regions with minimal influence on the classification decision.

4.4.1. Computational Cost and Efficiency Analysis

To practically validate the computational efficiency of the proposed method, Table 6 presents a comparative analysis of its performance against existing approaches. The evaluation metrics include training time and testing (inference) time. The hardware specifications were already described in Section 4. We clarified that all computational cost experiments were performed under the same hardware configuration previously reported. From Table 6, it is evident that lightweight networks such as ShuffleNet and MobileNetV2 achieve the shortest training and inference times due to their compact architecture, though often at the cost of reduced accuracy. In contrast, deeper networks like VGG16 and DarkNet19 demand significantly higher computational resources, resulting in longer training times and slower inference. The proposed KhayyamNet framework, despite employing three parallel feature extraction branches, demonstrates a balanced computational profile. Specifically, its training time (100 min) is comparable to that of other deep networks, while the integration of MRMR-based feature selection substantially reduces inference time to only 66 ms per sample. This highlights KhayyamNet's suitability for real-time and near real-time clinical applications, where rapid diagnostic feedback is essential. It is important to note that training a model typically involves learning from a large dataset, so the training time is naturally longer than the testing time. However, since training is performed only once and the trained model can be reused indefinitely, the inference time becomes the most critical factor for clinical deployment.

Table 6. Comparing the computational complexity.

Method	Training Time (min)	Testing Time (ms)
ResNet18	85	98
DarkNet19	110	105
MobileNetV2	30	58
ShuffleNet	15	55
GoogLeNet	95	102
AlexNet	60	85
Vgg16	130	125
Proposed Method (KhayyamNet)	100	66

4.4.2. Ablation Study

In order to evaluate the effectiveness of each component of the proposed KhayyamNet framework, an ablation study was conducted. Table 7 presents the mean $\pm$ SD of accuracy, precision, recall, and F1-score per class and overall across 10-fold cross-validation. Various configurations, where each feature extraction backbone was individually combined with the RF classifier, were used. The results demonstrate that using only the Xception network in combination with RF yields an accuracy of $87.13 \pm 1.25\%$. While employing the CNN network alone, the performance slightly improves to an accuracy of $88.77 \pm 1.62\%$, indicating that CNN captures more illustrative features compared to the Xception. In contrast, utilizing the ViT alongside RF Learning significantly boosts performance, achieving an accuracy of $92.91 \pm 0.92\%$, highlighting the importance of capturing global dependencies within the spinal MRI images. To further investigate the contribution of each feature branch, we evaluated the three pairwise combinations of the networks (CNN + Xception, CNN + ViT, and Xception + ViT) in addition to the single-branch and all-branch configurations. As shown in Table 7, none of the two-branch configurations outperformed the full three-branch fusion. In order to evaluate the specific contribution of the MRMR feature selection step, we conducted an additional experiment by fusing the features extracted from CNN, Xception, and ViT, followed by classification with RF, but without applying

MRMR. As shown in Table 7, this configuration achieved an accuracy of 96.93 ± 0.86%, which is notably lower than the performance of the complete KhayyamNet framework. The gap of more than 1% in accuracy demonstrates the effectiveness of MRMR in removing redundant and noisy features while preserving the most informative ones. This result highlights that MRMR is not only complementary to the fused deep feature representations but also crucial for maximizing diagnostic reliability in differentiating MS and Myelitis. Also, various deep neural networks such as ResNet18, DarkNet19, MobileNetV2, ShuffleNet, GoogLeNet, AlexNet, and VGG16 combined with MRMR and RF were employed for feature extraction, and these methodologies were compared with our proposed method. Some classical models were also used for classification instead of RF, such as SVM, Linear Discriminant Analysis (LDA), and a Bagged Ensemble Classifier (BEC) employing decision trees as base learners [37]. All models were implemented in MATLAB 2024a using the default settings. The results of comparison are shown in Table 7. Based on this table, our proposed method clearly demonstrates a significant advantage over other methods. Notably, it surpasses well-known models like MobileNetV2 and DarkNet19, which have accuracy of 94.23 ± 1.82% and 93.32 ± 1.06%, respectively, highlighting the robustness and efficiency of our approach. Finally, the full proposed method (KhayyamNet), which fuses the features extracted by CNN, Xception, and ViT in a parallel manner and applies MRMR feature selection before final classification, achieves the best performance, with an accuracy of 98.15 ± 0.80%. These results clearly demonstrate the effectiveness of the proposed feature fusion strategy and the importance of combining both local and global information to enhance the diagnosis of MS and Myelitis.

Table 7. Contribution of each block in the Khayyamnet model.

Method	Accuracy
CNN + MRMR + RF	88.77 ± 1.62%
Xception + MRMR + RF	87.49 ± 1.25%
ViT + MRMR + RF	92.91 ± 0.92%
CNN + Xception + MRMR + RF	97.30 ± 0.86%
CNN + ViT + MRMR + RF	97.09 ± 0.73%
Xception + ViT + MRMR + RF	97.23 ± 1.04%
CNN + Xception + ViT + RF	96.93 ± 0.86%
CNN + Xception + ViT + MRMR+ SVM	96.52 ± 1.95%
CNN + Xception + ViT + MRMR+ LDA	95.02 ± 1.65%
CNN + Xception + ViT + MRMR+ BEC	96.92 ± 1.21%
ResNet18 + MRMR+ RF	92.24 ± 0.89%
DarkNet19 + MRMR+ RF	93.32 ± 1.06%
MobileNetV2 + MRMR+ RF	94.23 ± 1.82%
ShuffleNet + MRMR+ RF	94.52 ± 1.21%
GoogLeNet + MRMR+ RF	92.36 ± 1.64%
AlexNet + MRMR+ RF	93.85 ± 0.98%
Vgg16 + MRMR+ RF	92.31 ± 1.16%
Proposed Method (KhayyamNet)	**98.15 ± 0.80%**

5. Discussion

The findings of this study demonstrate the effectiveness of the proposed KhayyamNet in diagnosing MS and Myelitis using spinal MRI images. With an achieved accuracy of 98.15 ± 0.80%, the model has shown a remarkable ability to distinguish between MS, Myelitis, and Healthy cases. Unlike many existing studies that primarily focus on brain MRI for MS diagnosis, this study employs spinal MRI, which offers several clinically validated advantages. In certain cases, MS lesions first manifest within the spinal cord before becoming evident on brain MRI, making spinal imaging particularly valuable for early diagnosis [3].

Multiple prospective and retrospective studies have quantified this advantage. Furthermore, in suspected MS cases, earlier reports indicated that 87.5% exhibited spinal cord abnormalities, and the inclusion of spinal imaging increased diagnostic sensitivity to 100%, whereas brain MRI alone yielded inconclusive results [38]. These findings underscore that spinal MRI can detect lesions earlier or in cases where brain MRI is normal or ambiguous, thereby enhancing diagnostic confidence and improving differentiation between MS and Myelitis. This is particularly relevant for Myelitis, which predominantly affects the spinal cord and may present without brain abnormalities [1]. By leveraging spinal MRI as the primary imaging modality, the present study builds upon this robust clinical evidence to address a well-recognized challenge in neurological imaging and improve diagnostic accuracy. The strength of the proposed KhayyamNet model lies in its multiscale feature extraction capability. While traditional CNN-based models demonstrate strong performance in capturing fine-grained local features, they often lack the ability to model long-range dependencies, an essential factor for understanding the complex anatomical structures of the spinal cord. In contrast, ViT effectively captures global contextual information but may overlook subtle local variations in lesion patterns. The Xception model, although lightweight and computationally efficient, is less effective when used in isolation. The proposed KhayyamNet overcomes these limitations by integrating the strengths of all three architectures through parallel feature extraction and an optimized fusion strategy. Additionally, the application of the MRMR algorithm significantly enhances the model's discriminative power by removing redundant or irrelevant features. This comprehensive fusion-based approach substantially improves the model's generalization and robustness, particularly in distinguishing between MS and Myelitis, which often exhibit overlapping radiological characteristics. Beyond quantitative evaluation, we conducted a qualitative review of the 57 misclassified cases to better understand the model's limitations. Interestingly, most of the MS cases predicted as Myelitis involved longitudinally extensive lesions spanning more than two vertebral segments, an imaging characteristic more typical of Myelitis. Conversely, several Myelitis cases misclassified as MS presented with focal, short-segment lesions resembling MS plaques. A smaller subset of errors was associated with images affected by low signal-to-noise ratio or motion artifacts, which obscured lesion boundaries. These observations suggest that the majority of errors arise in clinically ambiguous cases where radiological features overlap. Importantly, such cases are also challenging for human experts, underscoring the intrinsic diagnostic complexity of these conditions. Table 8 presents a comparative analysis of the performance achieved by our proposed method against state-of-the-art techniques across the same dataset utilized in this study. The results indicate that the proposed method outperformed existing approaches in terms of accuracy for both datasets.

Table 8. A comparison of KhayyamNet against state-of-the art techniques applied on the same dataset.

Study	Method	Year	Accuracy
Tatli et al. [11]	MSNet, DenseNet201, ResNe50, NCA, RF, Ch2, SVM, KNN, IMV	2024	97.63%
NourEldeen et al. [39]	CNN+MobileNet	2024	90.22%
Alzahrani et al. [40]	MS-Trust: Causality Attention + Global Attention + Squeeze-Excitation + Convolutional Tokenizer + CutMix/ MixUp Regularization	2025	92%
Proposed Model (KhayyamNet)	CNN + Xception + ViT + MRMR+ RF	2026	$98.15 \pm 0.80\%$

Although the proposed model performs classification at the slice level, its outputs can be naturally extended to patient-level diagnosis. In a practical clinical setting, slice-

level predictions may be aggregated using a simple majority voting strategy, where the final patient-level label is determined based on the dominant class across all slices. This approach can reduce the impact of isolated misclassified slices and provide a more stable and clinically meaningful decision. Investigation of patient-level aggregation strategies will be explored in future work.

6. Conclusions

In this study, we proposed a novel hybrid deep learning framework for the accurate classification of MS and Myelitis based on spinal MRI images. The proposed method includes comprehensive preprocessing and parallel feature extraction using three powerful architectures, namely Xception, CNN, and ViT, followed by feature fusion, feature selection using the MRMR algorithm, and final classification through the RF classifier. The experimental results demonstrate that combining local, spatial, and global feature representations significantly improves classification performance and effectively addresses key challenges such as data imbalance, noise, and the strong structural and radiological similarity between MS and Myelitis. A major strength of the proposed framework lies in its ability to learn multi-level representations, reduce feature dimensionality in an optimized manner, and achieve stable and reliable classification performance compared to traditional and single-model approaches. However, an important limitation of this study is the fact that the experimental evaluation is restricted to a single-centre public dataset, and the proposed framework has not been validated on external or multicentre data. Therefore, while the results indicate strong methodological potential, the current findings should be interpreted within the scope of single-centre, slice-level evaluation, and no claims regarding direct clinical deployment or readiness are made. Future work will focus on extending the evaluation to multicentre MRI datasets with subject-level metadata, enabling patient-level analysis and external validation, as well as incorporating multimodal information such as clinical, laboratory, and demographic data. These extensions will be essential to assess the robustness, generalizability, and real-world applicability of the proposed framework across diverse clinical settings.

Author Contributions: M.D.: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing—original draft, Visualization. M.F.: Conceptualization, Validation, Writing—review and editing, Project administration, Supervision. Z.M.: Conceptualization, Validation, Writing—review and editing, Project administration, Supervision, Visualization. S.F.: Conceptualization, Writing—review and editing, Validation, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: The Ethics Committee of Elazığ Fırat University approved the collection of the MR images used in this study. All data processing procedures were conducted in accordance with relevant ethical guidelines and regulations.

Data Availability Statement: The dataset is publicly available at https://www.kaggle.com/datasets/turkertuncer/ms-myelitis (accessed on 10 December 2025). The generated code can be obtained from the corresponding authors upon request without any additional restrictions.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-5.2 to improve grammar and readability. After that, all authors reviewed the contents, and they take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Portaccio, E.; Magyari, M.; Havrdova, E.K.; Ruet, A.; Brochet, B.; Scalfari, A.; Di Filippo, M.; Tur, C.; Montalban, X.; Amato, M.P. Multiple sclerosis: Emerging epidemiological trends and redefining the clinical course. *Lancet Reg. Health-Eur.* **2024**, *44*, 100977. [CrossRef] [PubMed]
2. Jakimovski, D.; Awan, S.; Eckert, S.; Farooq, O.; Weinstock-Guttman, B. Multiple sclerosis in children: Differential diagnosis, prognosis, and disease-modifying treatment. *CNS Drugs* **2022**, *36*, 45–59. [CrossRef]
3. Rocca, M.A.; Preziosa, P.; Barkhof, F.; Brownlee, W.; Calabrese, M.; de Stefano, N.; Granziera, C.; Ropele, S.; Toosy, A.T.; Vidal-Jordana, A.; et al. Current and future role of MRI in the diagnosis and prognosis of multiple sclerosis. *Lancet Reg. Health-Eur.* **2024**, *44*, 100978. [CrossRef]
4. Zhang, Y.; Lu, S.; Zhou, X.; Yang, M.; Wu, L.; Liu, B.; Phillips, P.; Wang, S. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: Decision tree, k-nearest neighbors, and support vector machine. *Simulation* **2016**, *92*, 861–871. [CrossRef]
5. Cavaliere, C.; Vilades, E.; Alonso-Rodríguez, M.C.; Rodrigo, M.J.; Pablo, L.E.; Miguel, J.M.; López-Guillén, E.; Sánchez Morla, E.M.; Boquete, L.; García-Martín, E. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors* **2019**, *19*, 5323. [CrossRef]
6. Saccà, V.; Sarica, A.; Novellino, F.; Barone, S.; Tallarico, T.; Filippelli, E.; Granata, A.; Chiriaco, C.; Bossio, R.B.; Valentino, P.; et al. Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state fMRI connectivity data. *Brain Imaging Behav.* **2019**, *13*, 1103–1114. [CrossRef]
7. Kaur, R.; Levy, J.; Motl, R.W.; Sowers, R.; Hernandez, M.E. Deep learning for multiple sclerosis differentiation using multi-stride dynamics in gait. *IEEE Trans. Biomed. Eng.* **2023**, *70*, 2181–2192. [CrossRef] [PubMed]
8. Kara, A.C.; Hardalaç, F. Detection and Classification of Knee Injuries from MR Images Using the MRNet Dataset with Progressively Operating Deep Learning Methods. *Mach. Learn. Knowl. Extr.* **2021**, *3*, 1009–1029. [CrossRef]
9. Ismail, K.A.; Dutta, A.K.; Sait, A.R. Ensemble Learning-Based Multiple Sclerosis Detection Technique Using Magnetic Resonance Imaging. *J. Disabil. Res.* **2024**, *3*, 20240078. [CrossRef]
10. Zhang, Y.; Liu, T.; Lanfranchi, V.; Yang, P. Explainable Tensor Multi-Task Ensemble Learning Based on Brain Structure Variation for Alzheimer's Disease Dynamic Prediction. *IEEE J. Transl. Eng. Health Med.* **2022**, *11*, 1–12. [CrossRef]
11. Tatli, S.; Macin, G.; Tasci, I.; Tasci, B.; Barua, P.D.; Baygin, M.; Tuncer, T.; Dogan, S.; Ciaccio, E.J.; Acharya, U.R. Transfer-transfer model with MSNet: An automated accurate multiple sclerosis and myelitis detection system. *Expert Syst. Appl.* **2024**, *236*, 121314. [CrossRef]
12. Mohammed Aarif, K.O.; Alam, A.; Pakruddin; Rahman, J.R. Exploring Challenges and Opportunities for the Early Detection of Multiple Sclerosis Using Deep Learning. In *Artificial Intelligence and Autoimmune Diseases: Applications in the Diagnosis, Prognosis, and Therapeutics*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 151–178. [CrossRef]
13. Ince, S.; Kunduracioglu, I.; Algarni, A.; Bayram, B.; Pacal, I. Deep Learning for Cerebral Vascular Occlusion Segmentation: A Novel ConvNeXtV2 and GRN-Integrated U-Net Framework for Diffusion-Weighted Imaging. *Neuroscience* **2025**, *574*, 42–53. [CrossRef]
14. Pacal, I.; Attallah, O. Hybrid Deep Learning Model for Automated Colorectal Cancer Detection Using Local and Global Feature Extraction. *Knowl.-Based Syst.* **2025**, *319*, 113625. [CrossRef]
15. Pacal, I.; Attallah, O. InceptionNeXt-Transformer: A Novel Multi-Scale Deep Feature Learning Architecture for Multimodal Breast Cancer Diagnosis. *Biomed. Signal Process. Control* **2025**, *110*, 108116. [CrossRef]
16. Aruk, I.; Pacal, I.; Toprak, A.N. A Novel Hybrid ConvNeXt-Based Approach for Enhanced Skin Lesion Classification. *Expert Syst. Appl.* **2025**, *283*, 127721. [CrossRef]
17. Acar, Z.Y.; Başçiftçi, F.; Ekmekci, A.H. Future Activity Prediction of Multiple Sclerosis with 3D MRI Using 3D Discrete Wavelet Transform. *Biomed. Signal Process. Control* **2022**, *78*, 103940. [CrossRef]
18. Jain, S.; Rajpal, N.; Yadav, J. Multiple Sclerosis Identification Based on Ensemble Machine Learning Technique. In Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020), Tamil Nadu, India, 29–30 November 2020. [CrossRef]
19. Huang, J.; Xin, B.; Wang, X.; Qi, Z.; Dong, H.; Li, K.; Zhou, Y.; Lu, J. Multi-Parametric MRI Phenotype with Trustworthy Machine Learning for Differentiating CNS Demyelinating Diseases. *J. Transl. Med.* **2021**, *19*, 377. [CrossRef] [PubMed]
20. Eshaghi, A.; Riyahi-Alam, S.; Saeedi, R.; Roostaei, T.; Nazeri, A.; Aghsaei, A.; Doosti, R.; Ganjgahi, H.; Bodini, B.; Shakourirad, A.; et al. Classification Algorithms with Multi-Modal Data Fusion Could Accurately Distinguish Neuromyelitis Optica from Multiple Sclerosis. *Neuroimage Clin.* **2015**, *7*, 306–314. [CrossRef] [PubMed]
21. Storelli, L.; Azzimonti, M.; Gueye, M.; Vizzino, C.M.; Preziosa, P.M.; Tedeschi, G.; De Stefano, N.M.; Pantano, P.M.; Filippi, M.; Rocca, M.A. A Deep Learning Approach to Predicting Disease Progression in Multiple Sclerosis Using Magnetic Resonance Imaging. *Investig. Radiol.* **2022**, *57*, 423–432. [CrossRef]

22. Krishnamoorthy, S.; Zhang, Y.; Kadry, S.; Khan, M.A.; Alhaisoni, M.; Mustafa, N.; Yu, W.; Alqahtani, A. Automatic Intelligent System Using Medical of Things for Multiple Sclerosis Detection. *Comput. Intell. Neurosci.* **2023**, *2023*, 4776770. [CrossRef] [PubMed]
23. Huang, C.; Chen, W.; Liu, B.; Yu, R.; Chen, X.; Tang, F.; Liu, J.; Lu, W. Transformer-Based Deep-Learning Algorithm for Discriminating Demyelinating Diseases of the Central Nervous System with Neuroimaging. *Front. Immunol.* **2022**, *13*, 897959. [CrossRef]
24. Narayana, P.A.; Coronado, I.; Sujit, S.J.; Wolinsky, J.S.; Lublin, F.D.; Gabr, R.E. Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI. *Radiology* **2020**, *294*, 398–404. [CrossRef]
25. Rocca, M.A.; Anzalone, N.; Storelli, L.; Del Poggio, A.; Cacciaguerra, L.; Manfredi, A.A.; Meani, A.; Filippi, M. Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics. *Investig. Radiol.* **2021**, *56*, 252–260. [CrossRef]
26. Eitel, F.; Soehler, E.; Bellmann-Strobl, J.; Brandt, A.U.; Ruprecht, K.; Giess, R.M.; Kuchling, J.; Asseyer, S.; Weygandt, M.; Haynes, J.-D.; et al. Uncovering Convolutional Neural Network Decisions for Diagnosing Multiple Sclerosis on Conventional MRI Using Layer-Wise Relevance Propagation. *Neuroimage Clin.* **2019**, *24*, 102003. [CrossRef]
27. Seok, J.M.; Cho, W.; Chung, Y.H.; Ju, H.; Kim, S.T.; Seong, J.-K.; Min, J.-H. Differentiation Between Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorder Using a Deep Learning Model. *Sci. Rep.* **2023**, *13*, 11625. [CrossRef] [PubMed]
28. Hagiwara, A.; Otsuka, Y.; Andica, C.; Kato, S.; Yokoyama, K.; Hori, M.; Fujita, S.; Kamagata, K.; Hattori, N.; Aoki, S. Differentiation Between Multiple Sclerosis and Neuromyelitis Optica Spectrum Disorders by Multiparametric Quantitative MRI Using Convolutional Neural Network. *J. Clin. Neurosci.* **2021**, *87*, 55–58. [CrossRef]
29. Ekmekyapar, T.; Taşcı, B. Exemplar MobileNetV2-Based Artificial Intelligence for Robust and Accurate Diagnosis of Multiple Sclerosis. *Diagnostics* **2023**, *13*, 3030. [CrossRef]
30. Zhang, Y.; Hong, D.; McClement, D.; Oladosu, O.; Pridham, G.; Slaney, G. Grad-CAM Helps Interpret the Deep Learning Models Trained to Classify Multiple Sclerosis Types Using Clinical Brain Magnetic Resonance Imaging. *J. Neurosci. Methods* **2021**, *353*, 109098. [CrossRef]
31. Dehghanpour, M.; Fateh, M.; Mohammadpoory, Z.; Ferdowsi, S. ZechariahNet: A Novel Method of MS Lesion Diagnosis Through MRI Images by the Combination of C-LSTM and 3D CNN Algorithms. *Algorithms* **2026**, *19*, 72. [CrossRef]
32. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258. [CrossRef]
33. Lu, E.H.C.; Gozdzikiewicz, M.; Chang, K.H.; Ciou, J.M. A Hierarchical Approach for Traffic Sign Recognition Based on Shape Detection and Image Classification. *Sensors* **2022**, *22*, 4768. [CrossRef] [PubMed]
34. Lei, X.; Pan, H.; Huang, X. A Dilated CNN Model for Image Classification. *IEEE Access* **2019**, *7*, 124087–124095. [CrossRef]
35. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 87–110. [CrossRef] [PubMed]
36. Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1226–1238. [CrossRef] [PubMed]
37. Plaia, A.; Buscemi, S.; Fürnkranz, J.; Mencía, E.L. Comparing Boosting and Bagging for Decision Trees of Rankings. *J. Classif.* **2022**, *39*, 78–99. [CrossRef]
38. Lycklama à Nijeholt, G.J.; Uitdehaag, B.M.; Bergers, E.; Castelijns, J.A.; Polman, C.H.; Barkhof, F. Spinal Cord Magnetic Resonance Imaging in Suspected Multiple Sclerosis. *Eur. Radiol.* **2000**, *10*, 368–376. [CrossRef] [PubMed]
39. NourEldeen, R.M.; Elshshtawy, R.A.; Elgohary, O.A.; Wahba, F.A.A.; Attia, M.K.; Elazm, Z.B.A.; Khattap, M.G.; Hassan, H.G.E.M.A.; Mostafa, M.; Ibrahim, N.; et al. Revolutionizing Neurological Diagnostics: Integrating 6G Technology with Deep Learning for Enhanced Detection of Multiple Sclerosis and Myelitis. In Proceedings of the 2024 International Telecommunications Conference (ITC-Egypt), Cairo, Egypt, 22–25 July 2024; pp. 602–608. [CrossRef]
40. Alzahrani, S.M. MS-Trust: A Transformer Model with Causal-Global Dual Attention for Enhanced MRI-Based Multiple Sclerosis and Myelitis Detection. *Complex Intell. Syst.* **2025**, *11*, 331. [CrossRef]