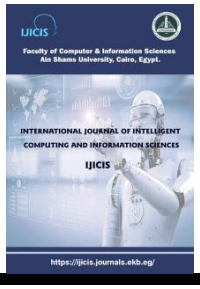
IJICIS, Vol.25, No.3, 41-54
DOI: 10.21608/ijicis.2025.402810.1410

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



CO-REGISTERED VOLUMETRIC MRI-BASED SYNTHETIC LESION TRANSPLANTATION FOR MULTIPLE SCLEROSIS CLASSIFICATION USING CUTOFF AUGMENTATION

Rezq Muhammed Thabet*

Scientific Computing Department
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
rizk.mohamed@cis.asu.edu.eg

Dina Khattab

Scientific Computing Department
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
dina.khattab@cis.asu.edu.eg

Maryam Al-Berry

Scientific Computing Department
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
maryam_nabil@cis.asu.edu.eg

Howida A. Shedeed

Scientific Computing Department
Faculty of Computer and Information Sciences, Ain Shams University,
Cairo, Egypt
dr_howida@cis.asu.edu.eg

Received 2025-07-11; Revised 2025-08-13; Accepted 2025-08-15

Abstract: *Multiple sclerosis (MS) is a chronic, immune-mediated disorder characterized by demyelinating lesions visible on MRI. Limited volumetric datasets and class imbalance hinder automated deep-learning approaches for MS classification. In this study, we extend the cutoff augmentation technique to three-dimensional (3D) MRI volumes and evaluate it using two publicly available pathological MS datasets and a large healthy cohort. Specifically, we used the 3D-MR-MS dataset (30 MS patients; modalities: FLAIR, T1-w, contrast-enhanced T1-w, and T2-w with multi-rater lesion segmentations), the long-MR-MS dataset (20 patients imaged longitudinally, two sessions per patient, with lesion-change masks), and the IXI healthy cohort (≈600 subjects; we selected T1-w volumes and used 50 T1-w scans for fold-specific pairing and background templates). After co-registering anatomical images, lesion masks, and brain masks into a shared healthy reference space, and performing skull stripping and intensity normalization, lesion voxels were transplanted into healthy volumes to generate synthetic pathological examples. A 3D DenseNet-169 trained with five-fold cross-validation demonstrated that 3D cutoff augmentation increased the mean accuracy from 61.2% to 72.7%, doubled the MS recall from 22.3% to 45.4%, and improved the F1 score from 36.3% to 61.7% while preserving precision. These results indicate that co-registered 3D cutoff augmentation effectively mitigates data scarcity and class imbalance for volumetric MS classification.*

Keywords: *Multiple Sclerosis, MRI Modalities, Cutoff Augmentation, CAD tools, Co-Registration*

***Corresponding Author**: Rezq Muhammed Thabet

Scientific Computing Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: rizk.mohamed@cis.asu.edu.eg

1. Introduction

Multiple Sclerosis (MS) is a chronic immune-mediated inflammatory disease of the central nervous system, with focal demyelinating lesions being readily visible on Magnetic Resonance Imaging (MRI) [1] [2]. Although expert radiologists can manually detect and assess these lesions, doing so is both time-consuming and subjective and it will not be possible with a large population. On the other hand, Artificial Intelligence (AI) is a robust and scalable solution that allows for automated, reproducible, and efficient analyzation of complex volumetric data. AI methodologies, particularly deep learning, are progressively used to aid in diagnosis, classification, and disease monitoring in MS.

MRI provides several imaging modalities that offer complementary views of brain tissue. T1-weighted (T1-w) scans offer detailed anatomical structure [3], T2-weighted (T2-w) sequences emphasize fluid-rich regions [4], and Fluid-Attenuated Inversion Recovery (FLAIR) nulls the cerebrospinal fluid signal to enhance lesion visibility [5]—especially near the ventricles, where MS lesions frequently occur. In addition, Proton Density–weighted (PD-w) images reflect hydrogen proton concentration and help visualize subtle tissue abnormalities [6]. Despite the diagnostic richness of multi-modal MRI, automated lesion classification remains difficult due to the small size and variable presentation of lesions, limited availability of annotated MS datasets, and pronounced class imbalance between pathological and healthy cases.

Against these issues, data augmentation has emerged as an effective approach to enhance the generalization of deep learning models. Among such techniques, cutoff augmentation—originally proposed for 2D brain tumor images [7]—involves lesion segmentation from pathological images and pasting it on healthy images to synthetically generate new lesion-bearing samples. Although effective for 2D tumor classification issues, this approach has yet to be tried in multiple sclerosis or three-dimensional (3D) MRI, where spatial alignment and anatomical realism are important.

In this study, we propose the first adaptation of cutoff augmentation for MS and 3D volumetric MRI data. Our approach addresses the challenges of volumetric lesion synthesis by co-registering three key components [8] [9]: the anatomical MS volume, its corresponding lesion mask, and the brain mask used for skull stripping [10]—all aligned to the space of a healthy brain volume. This co-registration ensures that lesion voxels extracted from pathological scans are accurately overlaid onto anatomically consistent locations in healthy volumes. Skull stripping removes non-brain tissue, improving signal-to-noise ratio and focusing the learning process on relevant regions. After preprocessing, lesion voxels are transplanted into healthy scans in 3D space, creating realistic synthetic pathological volumes for training. Also, intensity normalization is applied to enforce a uniform contrast across different subjects and folds

To demonstrate pathological cases, we use the 3D-MR-MS [11] [12] and long-MR-MS [13] [14] datasets. Additionally, the IXI dataset [15], a publicly available multi-modal MRI database of healthy participants, is the foundation for training and testing normal (healthy) volumes, as well as the reference point for the generation of synthetic lesion-augmented volumes.

Using volumetric cutoff augmentation together with a DenseNet-169 backbone, we demonstrate improved reliability and efficiency in automated MS classification. This improvement is achieved by co-registering lesion segmentations and anatomical scans into a common healthy reference space, then

transplanting lesion voxels to generate anatomically consistent 3D synthetic pathological volumes for training. Models trained with volumetric augmentation also showed reduced variance across folds, indicating better generalization.

The rest of the paper is organized as follows. Section 2 surveys related literature; Section 3 details the datasets, preprocessing steps, and our volumetric lesion-transplantation pipeline; Section 4 presents the experimental setup, quantitative results, and discussion; and Section 5 summarizes conclusions and outlines future work.

2. Related Work

A comprehensive review of our earlier work [16] summarizes the current state of synthetic lesion creation and multiple sclerosis categorization using artificial intelligence on MRI neuroimaging. The review summarizes a range of synthetic augmentation approaches—including encoder–decoder architectures, GAN-based image synthesis, and transplantative methods such as cutoff. It also covers AI-driven classification strategies, from basic convolutional neural networks to advanced hybrid and transfer-learning frameworks, which have driven recent progress in lesion segmentation, subtype classification, and large-scale patient studies. This assessment provides the basis for the thorough discussions that follow.

2.1 Synthetic Augmentation and Generative Models

Early work in the generation of synthetic lesions focused on improving segmentation by leveraging encoder–decoder structures. Salem et al. [17], for example, used a dual-input U-Net to generate synthetic multiple sclerosis lesions on intact images by encoding both lesion masks and MRI signals. Their results demonstrated improvements in both image realism and segmentation accuracy, particularly in the context of limited data availability. Although promising, these approaches rely on deep generative models and are primarily evaluated in segmentation contexts rather than classification tasks.

Cutoff augmentation was first introduced in 2D brain-tumor imaging by El-Assiouti et al. [7]. The method involves segmenting tumor masks and “pasting” them onto healthy scans to create synthetic examples, which improved generalization in tumor identification tasks using planar images. However, this approach has not been applied to neurological diseases like MS nor extended to 3D MRI.

Karras et al. [18] introduced StyleGAN, a sophisticated generative model intended for high-quality image generation. Its modular latent space architecture enables controlled generation, especially useful in research for enhancing medical images. Over the past few years, medical imaging research has used StyleGAN and similar models to create realistic lesions, normalize scans from various sites, and perform image translation between modalities.

Use GAN-based models to map between MRI images from different medical conditions. Zhu et al. [19] introduced CycleGAN, which is a promising approach for generating synthetic MS images from a related domain such as brain tumor MRI.

Advanced GAN architectures like StyleGAN3 [20] or diffusion models [21] can be explored to create better quality and higher resolution images of MS lesions. These can be especially useful for generating hard-to-detect small MS lesions that are underrepresented in current datasets.

2.2 AI-Driven MS Classification

Eitel et al. [22] employed a 3D CNN with Layer-Wise Relevance Propagation on T2-w MRI, achieving approximately 87% balanced accuracy in MS vs. healthy classification and revealing that the model leverages both lesion and non-lesion regions

Alijamaat et al. [23] presented a hybrid approach that combines wavelet transforms with convolutional neural networks (CNN) for automatic multiple sclerosis (MS)/healthy control (HC)/normal discrimination using brain MRI scans (T1-w/T2-w/FLAIR). The presented framework begins with the application of a two-dimensional discrete Haar Wavelet Transform (HWT), which breaks down each image slice into multi-resolution features that include both general structures and detailed textures. Next, such features extracted from wavelet analysis are fed into a streamlined CNN model. The evaluation achieved an accuracy of 99.05%, precision score of 99.43%, recall rate of 99.14%, and specificity of 98.89%.

Ye et al. [24] presented a novel approach combining Diffusion Basis Spectrum Imaging (DBSI) with artificial neural networks to classify three MS clinical courses—Relapsing-Remitting (RRMS), Primary Progressive (PPMS), and Secondary Progressive (SPMS). They further extended this framework to distinguish multiple lesion subtypes (Normal Appearing White Matter (NAWM), Acute Black Holes (ABH), Acute Gray Holes (PGH), Non-Black or gray Holes (NBH), and Persistent Black Holes (PBH)) from T1-weighted MRI scans. Their suggested methodology combined Diffusion Basis Spectrum Imaging (DBSI) [25] with Artificial Neural Networks (ANNs) to provide the best possible classification of the aforementioned categories. This approach was tested on a clinical (private) dataset, where intensity standardization was the only preprocessing step utilized. The obtained accuracy, F1-score, and Area Under the Curve (AUC) were 93.4%, 97.3%, and 99.8%, respectively.

Tatli et al. [26] developed MSNet, which utilizes a dual-stage transfer-learning technique with an iterative feature extraction approach from existing models like DenseNet201 and ResNet50. Feature selection methods like ReliefF, Chi2, and NCA were then employed along with classification algorithms K-NN and SVM. Data fusion with iterative voting resulted in the maximum classification accuracy at 97.63% for tri-class MRI datasets with examples for multiple sclerosis, myelitis, and control. The importance of feature engineering optimized was highlighted with explicit avoidance of discussions regarding synthetic lesion augmentation and volumetric segmentation.

Shrwan et al. [27] proposed a model of a two-dimensional convolutional neural network (2D-CNN) for the detection and classification of two conditions: pituitary tumors and MS. The researchers used a proprietary clinical database and a preprocessing technique that included image resizing. Their approach achieved performance metrics, namely accuracy, sensitivity, and F1-score, of 99.55%, 99.15%, and 99.15%, respectively.

Kushol et al. [28] developed an approach to classify Amyotrophic Lateral Sclerosis (ALS) cases from HC ones. Their 3D-CNN network was trained to take a T1-w MRI input image and depended on a new

idea to exploit the long-range relationships among image features; they also combined spatial and frequency domain information to enhance the network's performance. They trained their network on a clinical data and a set of preprocessing methods were applied including bias field correction, image normalization, skull stripping, spatial and frequency transformation, and patches extraction. They got an accuracy of 88%, F1-score of 90%, precision of 100%, recall of 81.3%, and a specificity of 100%.

Taken together, prior work shows strong progress in lesion synthesis and MS classification but also consistent gaps: many methods focus on 2D or segmentation tasks, rely on generative models whose anatomical fidelity is not guaranteed for classification, or report results on limited/private cohorts without addressing class imbalance or small-lesion sensitivity. These limitations motivate our anatomically consistent, volumetric lesion-transplantation pipeline and its evaluation with a volumetric DenseNet-169 backbone to better address data scarcity and improve robustness for 3D MS classification.

3. Methodology

We evaluate our 3D cutoff augmentation pipeline on two pathological MS datasets (3D-MR-MS and long-MR-MS) and one healthy cohort (IXI). Preprocessing comprises skull stripping and rigid + nonrigid co-registration of anatomical volumes, lesion masks, and brain masks into a common healthy template space. The cutoff augmentation method is then applied to extract lesion voxels from registered MS scans and transplants them into healthy volumes to generate synthetic pathological examples. We train and evaluate a DenseNet-169 classifier under two five-fold cross-validation experiments—with and without synthetic augmentation—using a consistent split of training and validation volumes. Implementation details cover network hyperparameters, optimizer settings, and evaluation metrics (accuracy, precision, recall, F1-score).

3.1 Datasets

3.1.1 Dataset Description

The current study employs three additional MRI datasets, which are: 3D-MR-MS [11] [12], long-MR-MS [13] [14], and IXI [15].

3D-MR-MS is a newly released public dataset of 30 multiple sclerosis patients, each with four raw MRI modalities—FLAIR, T1-w, T1-w post-contrast (T1-wKS), and T2-w—alongside a brain mask, consensus white-matter lesion segmentation (multi-rater), and preprocessed versions of all four modalities. This dataset provides high-quality, expert-validated lesion masks that form the foundation for our pathological training and validation splits.

long-MR-MS involves longitudinal imaging information from 20 multiple sclerosis patients, each imaged twice at different time intervals. Each imaging session has co-registered and N4 bias-field–corrected T1-w, T2-w, and FLAIR volumes, as well as unprocessed native-space images, brain masks, and masks outlining changes in white matter lesions. The dataset allows for the evaluation of model robustness to lesion appearance and contributes to our larger validation cohort.

The IXI dataset is a large, publicly available archive consisting of around 600 healthy adult brain MRIs, acquired from three London medical centers (Institute of Psychiatry, Guy's Hospital, Hammersmith Hospital) using GE and Philips scanners at 1.5 T and 3 T. The imaging protocol of each subject includes T1-w, T2-w, proton density-weighted, magnetic resonance angiography, and diffusion-weighted imaging. In this work, we use the IXI volumes as the "normal" training and validation class, and also as background templates for the synthetic lesion-augmentation process. Figure 1. presents representative slices from the three datasets used in this study. Each row shows one anatomical plane (axial, coronal, sagittal) and each column corresponds to an example subject from the 3D-MR-MS, long-MR-MS, and IXI collections. These examples were selected to illustrate typical inter-dataset differences in contrast and anatomical appearance that the preprocessing and co-registration steps must accommodate.

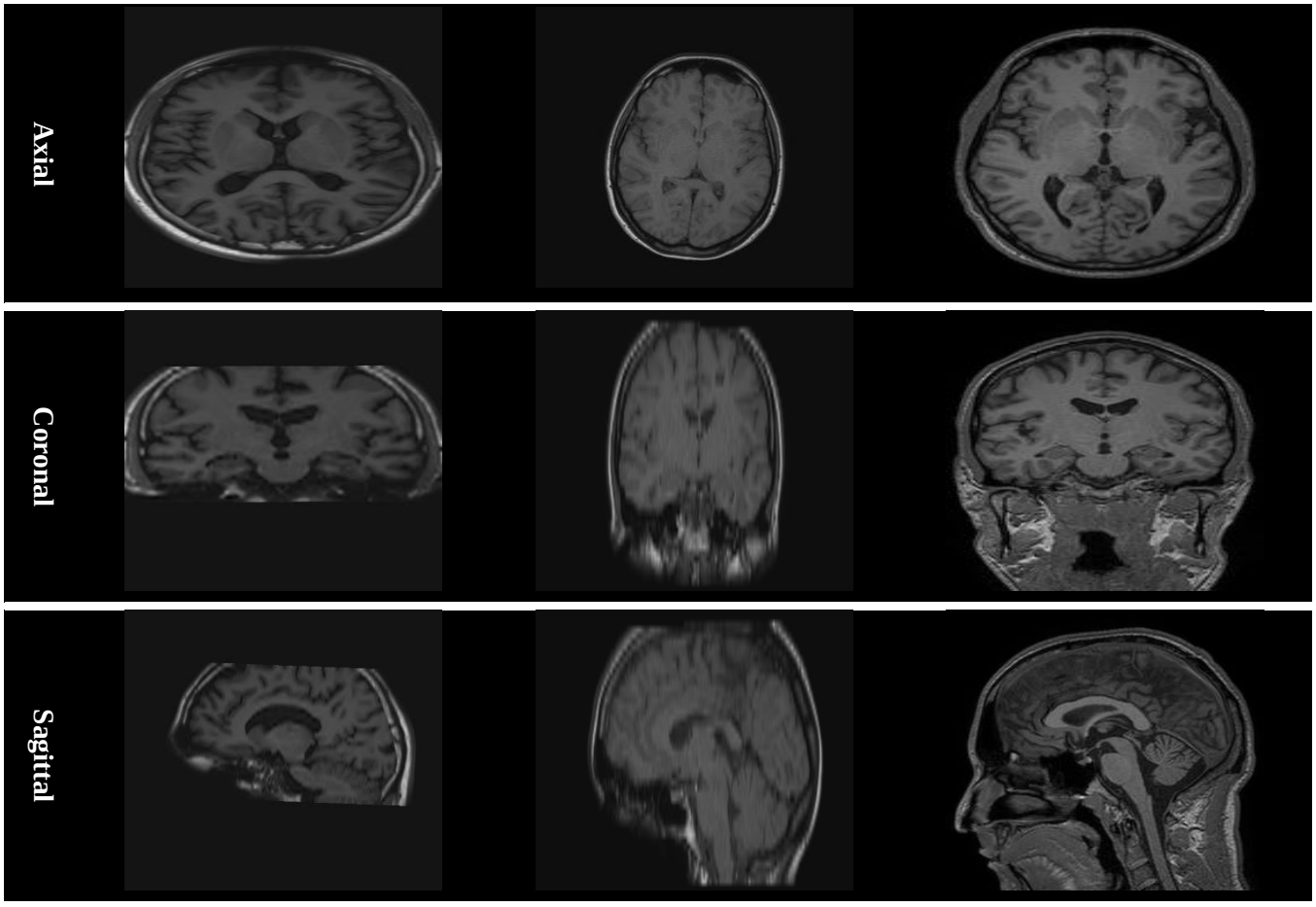


Figure 1. Representative axial, coronal, and sagittal slices from the 3D-MR-MS, long-MR-MS, and IXI datasets.

3.1.2 Anatomical–Healthy Pair Selection

In this work, the IXI cohort is utilized, which contains around 600 scans, to choose the first 50 T1-w volumes for fold-specific pairing. The first 30 volumes are exclusively matched with the 30 3D-MR-MS scans, with 24 for training and 6 for validation in each fold, and the other 20 volumes are matched with the 20 long-MR-MS validation scans. In each individual fold, just the 24 IXI volumes that belong to the

anatomical training subset are used as healthy templates for synthetic lesion transplantation in the second experiment, ensuring that augmentation is strictly matched to the training cases.

3.2 Preprocessing and Synthetic Data Generation

All analyses are performed on T1-w volumes exclusively guided by the demonstration of its known importance in lesion detection [29]. Also, the reason behind selecting T1-w was that the absence of FLAIR modality from the IXI dataset. The applied pipeline was described in the following steps:

3.2.1 Registration to Shared Space

For each anatomical–healthy pair:

- **Rigid + Nonrigid Registration:** We use the technique of ANTs' SyN [30] (rigid, then deformable) to align the anatomical T1-w volumes (original ones) to their corresponding healthy T1-w pairs (original ones).
- **Application of Transformation:** We apply the computed composite transformation to the anatomical brain mask and the lesion segmentation mask, thus registering these components as well as the anatomical image into the healthy volume's coordinate space.

3.2.2 Synthetic Data Generation

During the second experiment only, we apply the following steps to the 24 anatomical training volumes per fold:

- **Lesion Extraction:** Extract all lesion voxels (mask = 1) from each anatomically registered scan.
- **Cutoff transplantation:** Consists of incorporating these lesion voxels into correctly registered healthy T1-w images with the same spatial coordinates.
- **Synthetic Volume Generation:** Generate one synthetic volume per anatomical training case, resulting in 24 synthetic T1-w images. When combined with the 24 registered and skull-stripped original anatomical volumes, together they form 48 training samples.

3.2.3 Synthetic Data Generation Independent Healthy Splits

The last 74 volumes of the IXI scans consist of separate cohorts of healthy participants. We apply for each iteration n ($1 \leq n \leq 5$):

- **Healthy validation (26):** Select volumes ranging from $[(n–1)\times26 + 1]$ to $[n\cdot26]$.
- **Healthy Training (48):** The other 48 volumes—used in full for the second experiment training (or the first 24 for the first experiment).

3.2.4 Skull Extraction in Aligned Spatial Coordinates

After the registration and synthetic data generation operations, we use the registered anatomical brain mask to skull-strip both the transformed anatomical T1-w and the augmented T1-w lesion volume, ensuring identical brain extraction in the shared coordinate system.

3.2.5 Intensity Normalization

All co-registered T1-w volumes (original, healthy, and synthetic) are normalized through linear scaling of voxel intensities into the range [0, 1], thus enforcing a uniform contrast across different subjects and folds.

This pipeline ensures that for Experiment 1 each fold's training comprises 24 anatomical and 24 healthy volumes. For validation it uses 26 anatomical and 26 healthy volumes. As for the second experiment, training expands to 24 anatomical, 24 synthetic and 48 healthy, with the same validation split as in the first experiment. Figure **2**. Shows the whole pre-processing process.

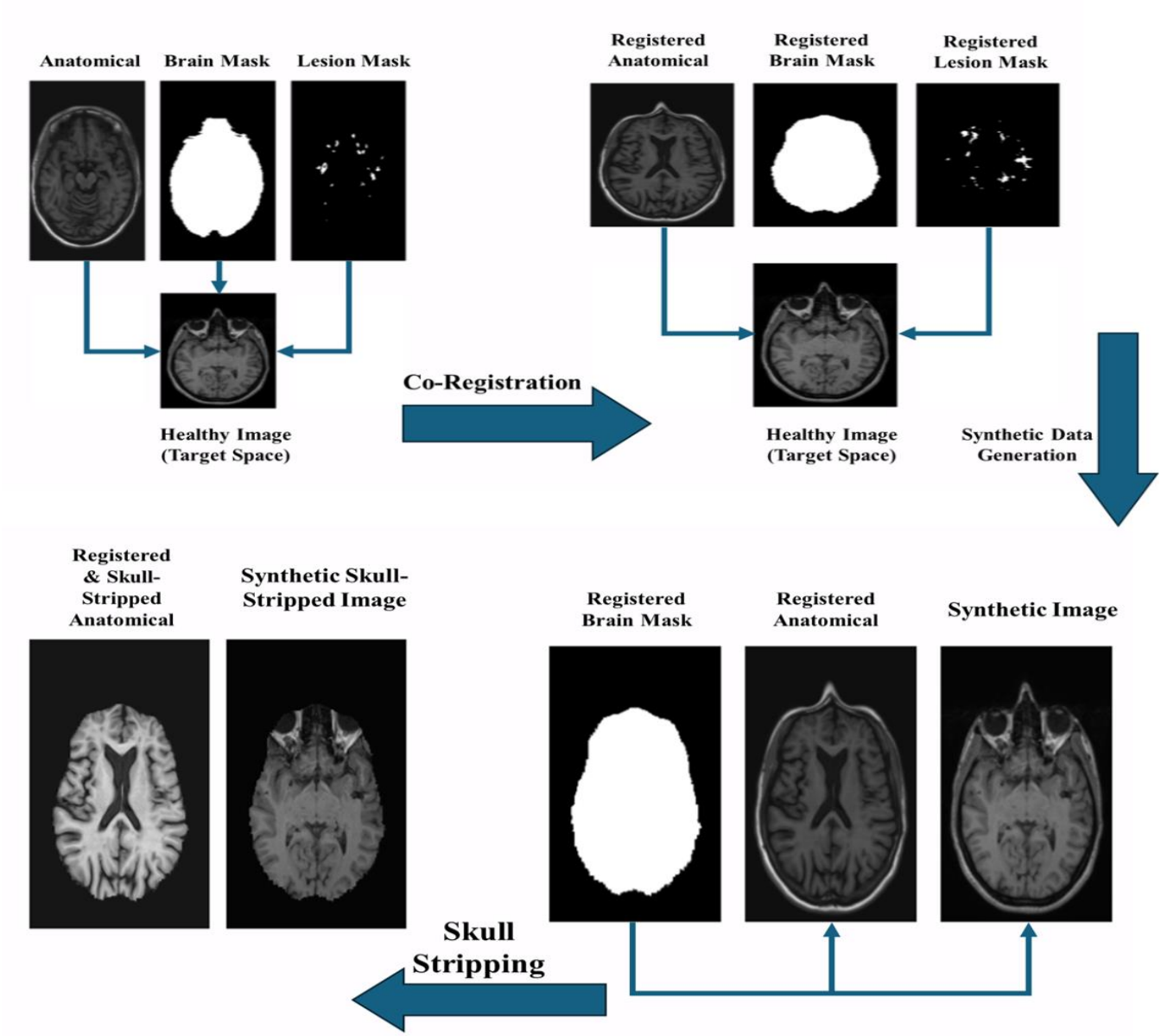


Figure 2. A detailed overview of the preprocessing pipeline, comprising co-registration, synthetic data generation, and skull stripping. The large blue arrows denote the sequential progression between the main preprocessing stages. The smaller directional arrows

3.3 Model Architecture and Loss Function

We adopt a 3D DenseNet-169 [31] as our classification backbone. The model is trained from scratch to directly learn volumetric MRI features.

3.3.1 Model Architecture

The original architecture of DenseNet-169 is modified in order to include 3D components in place of its 2D equivalents, and each skull-stripped, intensity-normalized T1-w scan is framed as one single-channel 3D input. In replacement of the primary 2D convolutional stem, 3D convolution with 3×3×3 dimensions (stride = 1, padding = 1) is used, immediately followed by batch normalization and ReLU activation functions. Then the architecture follows the DenseNet-169 bottleneck layers' structural format using 32 as the value for the growth rate, thus converting all convolution and transition stages into three dimensions; the transition stages include 1×1×1 convolution stages and 2×2×2 average pooling stages that consistently reduce spatial resolution. Lastly, the 3D global average pooling stage is used, which reduces the volumetric feature map into a 1D vector that will be passed through the fully connected layer that will predict two logits—a one that represents the MS class and another that indicates the Healthy class. Figure 3. Shows this 3D DenseNet-169 adapted framework.

3.3.2 Loss Function

We optimize the network using the binary cross-entropy loss (equivalent to multi-class cross-entropy for two classes), defined for each sample x with ground-truth label y ∈ {0 (Healthy), 1 (MS)} as:

$$\mathcal{L}(x, y) = -\left[y \log_{p1}(x) + (1 - y) \log_{p0}(x)\right],$$

Where $P_c(x)$ denotes the predicted probability for class c over the two output logits. This setup enables DenseNet-169 to be fully trained on volumetric MRI data, directly optimizing cross-entropy to distinguish between multiple sclerosis and normal scans.

3.4 Experimental Setup and Implementation Details

We do 5-fold cross-validation on both the first experiment (on original data) and the second experiment (with synthetic augmentation) on the 30-case 3D-MR-MS database. We provide for each fold, 24 anatomy MS volumes and 24 healthy IXI volumes as training set. The same volumes are allowed for the second experiment in addition to their synthetic counterparts. Validation is done on the 6 held-out anatomy volumes and on the 20 long-MR-MS exams with 26 healthy IXI volumes. This permitted each volume to contribute to validation only once and to training four times, allowing for stable patient-level performance estimates.

DenseNet-169 is trained from scratch in PyTorch on an NVIDIA P100 GPU with AdamW [32] (weight_decay = 1e-5) for 25 epochs with an initial learning rate of 5e-5. The batch size was a single volume, and the binary cross-entropy loss is trained to learn discrimination between MS and Healthy scans. The final performance is measured as the mean accuracy, precision, recall, and F1-score of the five validation folds.

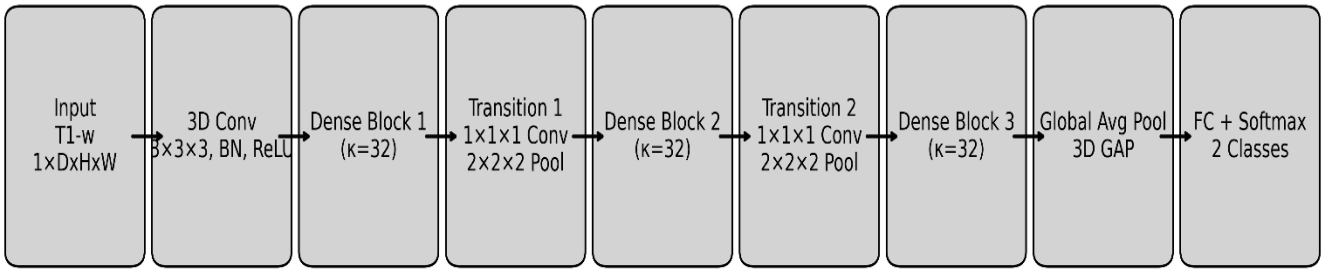


Figure 3. 3D DenseNet-169 for MS classification.

4. Experimental Results

The effectiveness of classification with our DenseNet-169 model is compared and determined using first the original database only, and then after including 3D cutoff augmentation. Performance measurements for both experiments of all folds of the validation sets are shown in Tables 1 and 2, respectively. The final row displays the average values over all folds.

4.1 Evaluation Metrics

To be able to evaluate the performance of the selected model as well as the proposed strategy, a set of well-known evaluation metrics were used for the purpose of classification evaluation which are described as follows:

Accuracy: The proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances. While intuitive, it can be misleading in imbalanced datasets. It can be calculated using equation 1.

$$\boldsymbol{Accuracy = \frac{TP + TN}{TP + TN + FP + FN}} \quad (\mathbf{1})$$

Precision: The proportion of true positive predictions among all positive predictions. It measures the exactness of the model. It can be computed using equation 2.

$$\boldsymbol{Precision = \frac{TP}{TP + FP}} \quad (\mathbf{2})$$

Recall: The proportion of true positive predictions among all actual positive instances. It measures the model's ability to find all positive samples. It can be measured using equation 3.

$$\boldsymbol{Recall = \frac{TP}{TP + FN}} \quad (\mathbf{3})$$

F1-Score: The harmonic mean of precision and recall. It provides a single metric that balances both precision and recall, particularly useful for imbalanced datasets. It can be obtained using equation 4.

$$\boldsymbol{F1 - Score = \frac{2TP}{2TP + FN + FP}} \quad (\mathbf{4})$$

Where TP represents the number of MS cases that have been properly classified as MS cases, FP represents the number non-MS being misclassified as MS cases, FN represents the number of MS cases being misclassified as non-MS, and finally TN represents the number of non-MS cases that have been properly classified as non-MS cases.

4.2 Experiment 1: Baseline Classification Performance

Table 1 illustrates the achieved accuracy of 61.2% for the DenseNet-169 model, when applied to– The Precision of 100%, indicates that no false positive instances. Still, recall for MS is only 22.3% which signifies that the model only identifies fewer than one-fourth of the true MS cases. This translates to a corresponding mean F1-score of 36.3%. This trend of high precision and very poor sensitivity reflects a sharp tendency to predict healthy cases while overlooking most pathological volumes on all test folds.

4.3 Experiment 2: Enhanced Performance with 3D Cutoff Augmentation

The introduction of 3D cutoff-based synthetic lesion augmentation results in a significant performance boost (see Table 2): mean accuracy increases to 72.7%, and the recall for the MS class raises to 45.4% (more than its double). The precision has been preserved at 100%. As a consequence, the average F1-score increases to 61.7%, reflecting a much better balance between detection of MS cases and avoidance of false positives. The most significant fold-wise improvement occurs for Fold 4 (accuracy boost of 17.3 points, recall boost of 34.6 points), highlighting the important role of augmentation in enabling the model to identify a larger percentage of true MS cases without relaxing its high precision in MS predictions.

Table 1. Baseline Classification Performance on Original 3D-MR-MS, long-MR-MS, and IXI Data.

Validation set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	61.54	100	23.08	37.5
Fold 2	59.62	100	19.23	32.26
Fold 3	59.62	100	19.23	32.26
Fold 4	65.38	100	30.77	47.06
Fold 5	59.62	100	19.23	32.26
Mean	61.2	100	22.3	36.3

Table 2. Enhanced Classification Performance with 3D Cutoff-Augmented Training Data.

Validation set	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	73.08	100	46.15	63.16
Fold 2	71.15	100	42.31	59.46
Fold 3	67.31	100	34.62	51.43
Fold 4	82.69	100	65.38	79.07
Fold 5	69.23	100	38.46	55.56
Mean	72.7	100	45.4	61.7

5. Conclusion and Future Work

In this study, we introduce a new three-dimensional cutoff augmentation pipeline for MS lesion classification using volumetric MRI. By precisely registering anatomical scans, lesion masks, and brain masks to a common healthy reference space, we enable anatomical consistent transplantation of lesion

voxels. Our synthetic augmentation strategy, when applied to a small pathological training dataset, boosts model sensitivity considerably without compromising specificity. Through 5-fold cross-validation, this augmentation achieved more than an 11% improvement in overall accuracy and a twofold improvement in lesion recall, resulting in a remarkable improvement in the balanced F1-score by 25%. Training a DenseNet-169 model from scratch on T1-w volumes demonstrates that the 3D lesion transplantation strategy can successfully address the challenges of data paucity and class imbalance in MS classification. Future work will involve the inclusion of multi-modal inputs (e.g., FLAIR and T2-w sequences), sophisticated blending strategies, and evaluation on larger, multi-center datasets to enhance clinical robustness and generalizability.

References

[1] B. Channel, "Multiple sclerosis (MS)-common symptoms," 2021. [Online]. Available: https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/multiple-sclerosis-ms-common-symptoms. [Accessed 7 July 2025].

[2] C. Clinic, "MRI (Magnetic Resonance Imaging)," 2022. [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/4876-magnetic-resonance-imaging-mri. [Accessed 7 July 2025].

[3] Y. Chen, S. J. Almarzouqi, M. L. Morgan and A. G. Lee, "T1-Weighted Image," in Encyclopedia of Ophthalmology, U. Schmidt-Erfurth and T. Kohnen, Eds., Berlin, Heidelberg, Springer Berlin Heidelberg, 2018, pp. 1747-1750.

[4] Y. Chen, S. J. Almarzouqi, M. L. Morgan and A. G. Lee, "T2-Weighted Image," in Encyclopedia of Ophthalmology, U. Schmidt-Erfurth and T. Kohnen, Eds., Berlin, Heidelberg, Springer Berlin Heidelberg, 2018, pp. 1750-1752.

[5] W. M. Bailey, "Fast Fluid Attenuated Inversion Recovery (FLAIR) imaging and associated artefacts in Magnetic Resonance Imaging (MRI)," Radiography, vol. 13, no. 4, pp. 283-290, November 2007.

[6] M. Master, "Proton Density (PD) MRI," 2023. [Online]. Available: https://mrimaster.com/characterise-image-pd/. [Accessed 7 July 2025].

[7] O. S. El-Assiouti, H. El-Saadawy, H. M. Ebied, D. Khattab and G. Hamed, "RegionInpaint, Cutoff and RegionMix: Introducing Novel Augmentation Techniques for Enhancing the Generalization of Brain Tumor Identification," IEEE Access, vol. 11, pp. 83232-83250, 2023.

[8] X. Zhang, Y. Feng, W. Chen, X. Li, A. V. Faria, Q. Feng and S. Mori, "Linear Registration of Brain MRI Using Knowledge-Based Multiple Intermediator Libraries," Frontiers in Neuroscience, vol. 13, p. 909, 11 September 2019.

[9] Z. A. Dagdeviren, K. Oguz and M. Cinsdikici, "Automatic registration of structural brain MR images to MNI image space," in 2015 23th Signal Processing and Communications Applications Conference (SIU), Malatya, Turkey, 2015.

[10] P. Kalavathi and V. B. S. Prasath, "Methods on Skull Stripping of MRI Head Scan Images—a Review," Journal of Digital Imaging, vol. 29, no. 3, pp. 365-379, June 2016.

[11] Ž. Lesjak, A. Galimzianova, A. Koren, M. Lukin, F. Pernuš, B. Likar and Ž. Špiclin, "A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus," Neuroinformatics, vol. 16, no. 1, pp. 51-63, January 2018.

[12] L. o. I. Technologies, "3D-MR-MS," 2023. [Online]. Available: https://lit.fe.uni-lj.si/en/research/resources/3D-MR-MS/. [Accessed 7 July 2025].

[13] Ž. Lesjak, F. Pernuš, B. Likar and Ž. Špiclin, "Validation of White-Matter Lesion Change Detection Methods on a Novel Publicly Available MRI Image Database," Neuroinformatics, vol. 14, no. 4, pp. 403-420, October 2016.

[14] L. o. I. Technologies, "long-MR-MS," 2023. [Online]. Available: https://lit.fe.uni-lj.si/en/research/resources/long-MR-MS/. [Accessed 7 July 2025].

[15] Brain-development, "IXI Dataset," 2016. [Online]. Available: https://brain-development.org/ixi-dataset/. [Accessed 7 July 2025].

[16] R. M. Thabet, H. A. Shedeed, M. Al-Berry and D. Khattab, "Multiple sclerosis classification and segmentation in neuroimaging MRI using different machine and deep learning techniques: a review," Artificial Intelligence Review, vol. 58, no. 8, p. 255, 28 May 2025.

[17] M. Salem, S. Valverde, M. Cabezas, D. Pareto, A. Oliver, J. Salvi, A. Rovira and X. Llado, "Multiple Sclerosis Lesion Synthesis in MRI Using an Encoder-Decoder U-NET," IEEE Access, vol. 7, pp. 25171-25184, 2019.

[18] T. Karras, S. Laine and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019.

[19] J.-Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017.

[20] T. Zhu, J. Chen, R. Zhu and G. Gupta, "StyleGAN3: Generative Networks for Improving the Equivariance of Translation and Rotation," arXiv, 2023.

[21] J. Wu, W. Ji, H. Fu, M. Xu, Y. Jin and Y. Xu, "MedSegDiff-V2: Diffusion-Based Medical Image Segmentation with Transformer," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 6, pp. 6030-6038, 24 March 2024.

[22] F. Eitel, E. Soehler, J. Bellmann-Strobl, A. U. Brandt, K. Ruprecht, R. M. Giess, J. Kuchling, S. Asseyer, M. Weygandt, J.-D. Haynes, M. Scheel, F. Paul and K. Ritter, "Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation," NeuroImage: Clinical, vol. 24, p. 102003, 2019.

[23] A. Alijamaat, A. NikravanShalmani and P. Bayat, "Multiple sclerosis identification in brain MRI images using wavelet convolutional neural networks," International Journal of Imaging Systems and Technology, vol. 31, no. 2, pp. 778-785, June 2021.

[24] Z. Ye, A. George, A. T. Wu, X. Niu, J. Lin, G. Adusumilli, R. T. Naismith, A. H. Cross, P. Sun and S. Song, "Deep learning with diffusion basis spectrum imaging for classification of multiple sclerosis lesions," Annals of Clinical and Translational Neurology, vol. 7, no. 5, p. 2020, May 695-706.

[25] P. Sun, A. George, D. C. Perantie, K. Trinkaus, Z. Ye, R. T. Naismith, S.-K. Song and A. H. Cross, "Diffusion basis spectrum imaging provides insights into MS pathology," Neurology Neuroimmunology & Neuroinflammation, vol. 7, no. 2, p. e655, March 2020.

[26] S. Tatli, G. Macin, I. Tasci, B. Tasci, P. D. Barua, M. Baygin, T. Tuncer, S. Dogan, E. J. Ciaccio and U. R. Acharya, "Transfer-transfer model with MSNet: An automated accurate multiple sclerosis and myelitis detection system," Expert Systems with Applications, vol. 236, p. 121314,

2024.

[27] R. Shrwan and A. Gupta, "Classification of Pituitary Tumor and Multiple Sclerosis Brain Lesions through Convolutional Neural Networks," IOP Conference Series: Materials Science and Engineering, vol. 1049, no. 1, p. 012014, 1 January 2021.

[28] R. Kushol, C. C. Luk, A. Dey, M. Benatar, H. Briemberg, A. Dionne, N. Dupré, R. Frayne, A. Genge, S. Gibson, S. J. Graham, L. Korngut, P. Seres, R. C. Welsh, A. H. Wilman, L. Zinman, S. Kalra and Y.-H. Yang, "SF2Former: Amyotrophic lateral sclerosis identification from multi-center MRI data using spatial and frequency fusion transformer," Computerized Medical Imaging and Graphics, vol. 108, p. 102279, September 2023.

[29] R. M. Thabet, D. Khattab, H. A. Shedeed and M. Al-Berry, "The Effects of Diverse Brain MRI Modalities on Multiple Sclerosis Lesions Segmentation: Comprehensive Analysis," in 2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES), Giza, Egypt, 2024.

[30] GitHub, "ANTsPy," 2024. [Online]. Available: https://github.com/ANTsX/ANTsPy. [Accessed 7 July 2025].

[31] G. Huang, Z. Liu, L. v. d. Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," arXiv, 28 January 2018.

[32] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," arXiv, 4 January 2019.