A Deep Learning Hybrid Approach for EDSS Prediction Using Brain MRI and Lesion Segmentation

Bhumika G, Aniketh Chakravarthy S, Bhuvana S, Sujay S Veershette

Department of Computer Science, RV University

***Abstract*—For evaluating the course of the disease and directing therapeutic measures, it is essential to predict the Expanded Disability Status Scale (EDSS) in patients with multiple sclerosis (MS). In order to predict EDSS scores from multimodal brain MRI data, this study introduces a novel hybrid deep learning framework that combines ensemble learning and convolutional neural networks (CNNs). In order to extract comprehensive radiomic and volumetric information, such as lesion load, contrast ratios, brain volume metrics, and lesion metrics, the method makes use of three MRI modalities—T1, T2, and FLAIR—as well as lesion segmentation masks. Predictive accuracy is increased by using a strong ensemble of CNN models that have been trained on 2D slices with various input combinations and augmentations. Results from Leave-One-Out Cross-Validation (LOOCV) show competitive performance in terms of mean squared error (MSE) and R2, and they show a significant connection with clinical EDSS values. A strong basis for clinical decision support tools in MS management is provided by our hybrid ensemble technique, which efficiently captures complicated MRI data.**

**Index Terms: Multiple Sclerosis (MS), Expanded Disability Status Scale (EDSS), Convolutional Neural Networks (CNNs), Ensemble Learning, Multimodal MRI, Radiomic Features, Volumetric Features, Leave-One-Out Cross-Validation (LOOCV).**

1. Introduction

A chronic, immune-mediated condition of the central nervous system (CNS), multiple sclerosis (MS) is typified by increasing neurological disability and demyelination. Although the Expanded Disability Status Scale (EDSS) is frequently used to measure neurological disability, manual evaluation of the scale is laborious and subjective. Important biomarkers like lesion load and brain shrinkage, which are closely associated with clinical impairment, are provided by magnetic resonance imaging (MRI). However, because of the intricate connections between imaging characteristics and clinical outcomes, converting MRI data into precise EDSS predictions is still difficult.

In order to tackle this difficulty, recent developments in deep learning have shown promise. Multimodal MRI scans' atrophy patterns and lesion distribution can be analysed by convolutional neural networks (CNNs). Single-modality techniques, however, frequently fail to capture the entire range of alterations associated with disease. Hybrid systems that combine multimodal MRI data with volumetric biomarkers and ensemble learning techniques are becoming more effective tools for EDSS prediction in order to get beyond these restrictions.

In order to predict EDSS scores, this study presents a novel hybrid deep learning architecture that combines ensemble modelling and CNN-based image regression. In order to produce reliable predictions, the system integrates clinical factors with multimodal MRI scans and generated volumetric indicators. The main goals are to investigate the therapeutic implications of this framework for bettering MS care and to build and test it using real-world information. This study intends to support personalised medicine approaches in MS management by bridging the gap between clinical disability measures and sophisticated imaging biomarkers.

1. Related Work

In multiple sclerosis (MS), conventional approaches for predicting the Expanded Disability Status Scale (EDSS) mostly depend on clinical factors like recurrence rates and fundamental MRI measurements like lesion burden. The incapacity of these methods to take into consideration imaging heterogeneity and intricate relationships between clinical and radiological variables limits their prediction accuracy, even though they offer some insight into the course of the disease. Support vector regression (SVR) and random forests are two machine learning models that have demonstrated a reasonable level of performance in EDSS prediction; nevertheless, their limited integration of multimodal data and unsophisticated feature extraction frequently result in their lack of robustness.

The capacity to analyse MRI data for EDSS prediction has been greatly improved by recent developments in deep learning. Using T1- and T2-weighted MRI scans, Yoo et al. (2021) used a ResNet-50 architecture to predict EDSS scores, showing better results than conventional techniques. Lesion spatial information, which is essential for comprehending disease pathogenesis, was not included in this method. By incorporating lesion masks into their model, Zhang et al. (2022) filled this gap, although they did it as static inputs rather than dynamically segmenting lesions throughout the prediction process. While highlighting the need for models that can better capture lesion dynamics and contextual information, these findings also demonstrate the promise of deep learning.

Segmentation of Lesions is a crucial part of MS imaging analysis is lesion segmentation, and U-Net architectures are leading the field because of how well they can separate lesions from MRI data. Recently, hybrid models that integrate transformers and convolutional neural networks (CNNs) have shown promise as instruments for enhancing segmentation quality. Such models are still not well studied in the context of EDSS prediction, where dynamic lesion segmentation should improve predictive accuracy by offering hierarchical feature fusion, despite their success in segmentation tasks.

Current methods frequently handle lesion information and MRI data as distinct inputs, which restricts their capacity to capture the intricate relationship between structural damage and clinical outcomes. In order to fill these gaps, this study presents a hybrid deep learning architecture that hierarchically combines multimodal information, such as FLAIR-based radiomic and volumetric biomarkers, T1-, T2-, and dynamically segmented lesions, into a single prediction model. This strategy seeks to increase EDSS prediction accuracy and resilience by improving contextual comprehension through hierarchical feature integration, establishing the foundation for sophisticated clinical decision support systems in MS management.

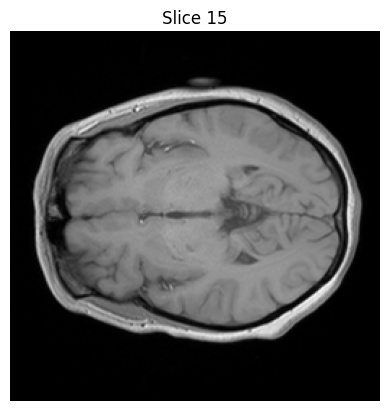
1. Methodology
2. *Dataset*

Every aspect of the EDSS prediction procedure is described in this part, including data preparation, image feature extraction, CNN-based representation learning, ensemble model building, and regression metrics evaluation. The central component of the suggested hybrid system combines CNN-derived embeddings with manually created volume radiomic features, which are trained and verified using Leave-One-Out Cross-Validation (LOOCV) in all traditional regression models for comparison.

*1.1 Data Imaging*

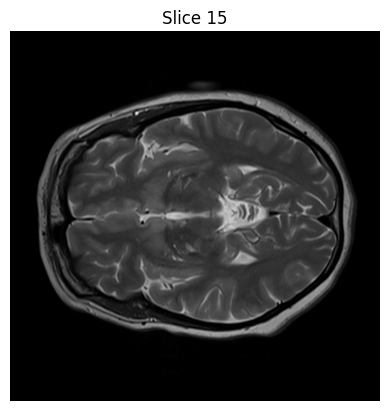
The following MRI modalities are accessible for each patient:

T1-weighted MRI: Mainly utilised for volumetric studies and structural brain anatomy.



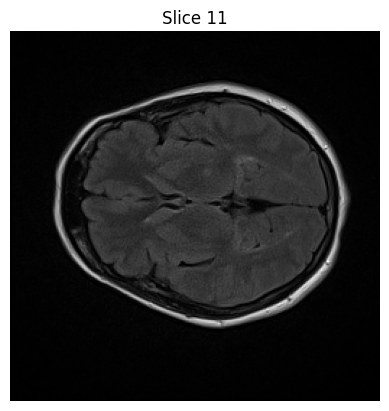
*T1-Weighted MRI*

T2-weighted MRI: Often used to diagnose lesions, this scan is sensitive to fluid content.



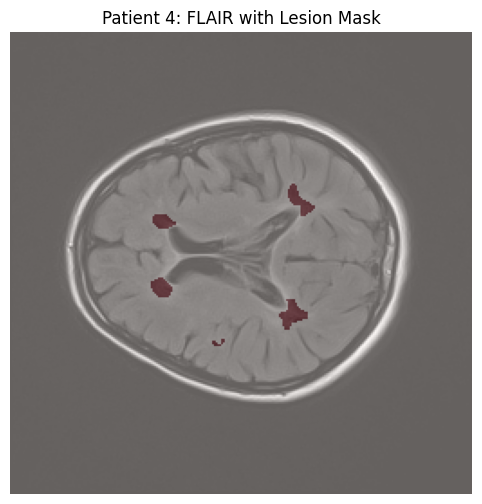
*T2-Weighted MRI*

FLAIR (Fluid-Attenuated Inversion Recovery) MRI: By inhibiting CSF signals, FLAIR improves lesion visualisation.



*Flair MRI*

Lesion Segmentation Masks: Binary masks that show the location of lesions in each modality, allowing for accurate lesion volume calculations.



*Lesion Segmentation mask*

Every picture is arranged patient-wise and saved in NIfTI format. 2D axial slices, centred on the middle of the brain, were extracted from each 3D MRI image after further processing. CNN uses these slices as its main input for training and creating embeddings.

*1.2 Clinical Parameters*

A thorough set of neurological and clinical assessments is included with every patient and is kept in an organised CSV file. These consist of:

* Age, Age of Onset, and Demographic and Temporal Information
* Gender: Encoded as Gender\_F, Gender\_M, and Gender\_N
* The interval of time between MRI capture and EDSS evaluation
* Ground Truth for EDSS: The regression objective is the Expanded Disability Status Scale score (EDSS).
* Evaluation of Clinical Symptoms

During a clinical assessment, binary signs for neurological symptoms are evaluated.

1. The brain stem, pyramidal, cerebellar, sensory, sphincters, visual, mental, and speech
2. Sensory System, Motor System, Gait, and Coordination
3. Function of the bowels and bladder, mobility, and mental state
4. Nystagmus, Ocular Movement, Swallowing, and Optic Discs

These evaluations correspond to certain functional systems that are utilised in the EDSS computation.

*1.3 Description of Treatment*

One-hot encoded drugs that the patient was prescribed:

* Types of Medications: Rebif, Tysabri, Gelenia, Betaferon, and Avonex
* Biomarkers for radiology (derived features)

Using volumetric analysis and lesion segmentation, we calculate:

1. LesionVolume\_mm3: The overall severity of lesions
2. LesionCount: The quantity of unique lesion areas
3. MeanLesionSize\_mm3: Mean size of lesion
4. Lesion volume by modality: FLAIR\_LesionLoad, T1\_LesionLoad, and T2\_LesionLoad
5. ContrastRatio: Active lesion differentiation using the T2-FLAIR contrast measure
6. BrainVolume: T1 segmentation estimates the total brain volume.

*1.4 Integration and Data Format*

A single feature matrix is created by combining the imaging-derived features and clinical factors; each row represents a patient, and each column represents a radiomic or clinical characteristic. Several regression models are trained using this matrix in order to predict EDSS.

Moreover, this feature matrix is supplemented by intermediate CNN embeddings that are taken from MRI slices in order to provide hybrid modelling, which combines explicit clinical factors with learnt deep visual representations.

*B. Preprocessing Pipeline*

*2.1. MRI Loading and Orientation*

Nibabel is used to load NIfTI files. To provide anatomical uniformity across modalities, images are reoriented to standard RAS alignment.

*2.2. Choosing a Slice*

The core axial slices of the 3D volume—usually ±5 from the center—are chosen for each modality. These slices concentrate on areas with the best structural information and visibility of lesions.

*2.3. Normalisation and Resizing*

Every slice is resized using bilinear interpolation to 224 x 224 pixels. For all modalities, intensities undergo normalisation between 0 and 1.

*2.4. Masks for Lesion Segmentation*

For every modality, binary lesion masks are aligned. These masks are employed in the computation of lesion-based metrics as well as as input features.

*C. Feature Engineering*

Lesion segmentation masks and MRI data are used to obtain a complete set of radiological and derived characteristics for every patient. The structural and intensity-based aspects of illness presentation that are pertinent to EDSS scoring are intended to be captured by these features.

*3.1. Lesion Load*  
 The total lesion burden in a given MRI scan, computed as:

Calculated separately for each modality (FLAIR, T1, T2) using binary lesion masks. It provides a volumetric measure of overall lesion burden in mm³.

*3.2. T2/FLAIR Contrast Ratio*

Assists in distinguishing between active and chronic lesions by measuring the difference in normalised intensity between T2 and FLAIR scans:

Only calculated inside the lesion mask following z-score normalisation of the FLAIR and T2 images. Division by zero is avoided by the short epsilon term. This measure improves sensitivity to signal change and lesion fluidity.

*3.3. Brain Volume*

Brain tissue volume total taken from the T1-weighted scan:

used as a structural baseline to comprehend the connection between EDSS score or lesion burden and brain size.

*3.4. Volume of Lesion*

The sum of all segmented lesions' volumes:

In accordance with Lesion Load, this is equal to the sum of the lesion voxels times the voxel volume.

*3.5. Number of Lesions*

It is the quantity of disjointed brain lesion areas in a patient. It is calculated with the lesion mask's linked component labelling applied. This measure aids in distinguishing between individuals who have several tiny lesions and those who have a few major ones.

*3.6. Average Size of Lesion*

The average size of the lesions found, measured in mm³:

This aids in determining whether lesions are more concentrated or widely distributed.

A pandas DataFrame indexed by patient ID contains all of the characteristics. Each row serves as the structured input for ensemble and downstream regression models and contains the patient's ground-truth EDSS score, clinical information, and imaging-derived characteristics.

### *D. CNN-Based Image Embedding Extraction*

To extract deep visual features from the MRI slices:

* A custom CNN is defined with:  
  + 2 convolutional layers (with batch norm, ReLU, max-pooling)
  + 1 flattening layer producing a 224-dimensional embedding vector
* Input channels include:  
  + Stacked T1, T2, FLAIR (3 channels)
* The CNN is trained using MSE loss to regress the EDSS value.

After training, the penultimate output layer (embedding) is extracted for each patient and stored as part of the feature set. These embeddings are used in later regression models.

### *E. Leave-One-Out Cross-Validation (LOOCV)*

Considering the short dataset (N=60), LOOCV is selected for assessment. The actions are:

1. For every fold:

One patient is not included in the tests.

Training is done on the remaining 59 patients.

1. StandardScaler is used to standardise features (fit just on training).
2. The patient who is left out is predicted and noted.
3. All folds are aggregated to calculate the final metrics.

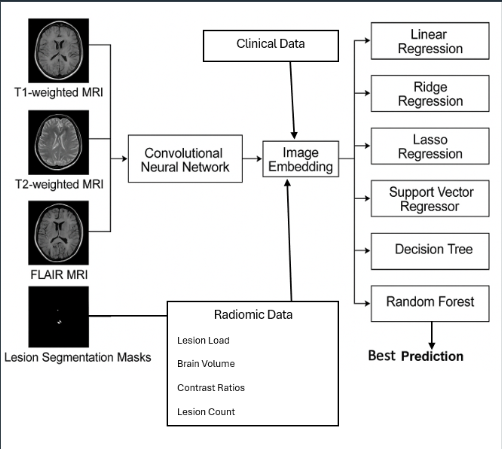
### *F. Regression Models for Comparison*

The following models are trained and evaluated on the engineered features + CNN embeddings:

* Linear Regression
* Ridge Regression
* Lasso Regression
* Support Vector Regression (SVR)
* Random Forest Regressor
* XGBoost Regressor
* MLP Regressor (Shallow Neural Net)

All models use the same LOOCV splits and are trained using Scikit-Learn. Performance is measured using:

* Mean Squared Error (MSE)
* Mean Absolute Error (MAE)
* R² Score



*Schematic diagram*

*G. Results Analysis*

A dataset comprising imaging-derived features, lesion-based biomarkers, and specialist-validated clinical variables was used to assess the performance of multiple regression models using Leave-One-Out Cross-Validation (LOOCV). Using three-fold internal cross-validation, each model's hyperparameters were adjusted using GridSearchCV before being evaluated for every patient.

### *7.1 Model Performance Summary*

The table below summarizes the performance metrics for each model:

| Model | Best Hyperparameters | MMSE | RRMSE | MMAE | RR² |
| --- | --- | --- | --- | --- | --- |
| Linear Regression | {} | 118.323 | 44.281 | 33.070 | —-6.318 |
| Ridge Regression | {'alpha': 100.0} | 11.721 | 11.312 | 11.110 | 00.313 |
| Lasso Regression | {'alpha': 1.0} | 22.589 | 11.609 | 11.400 | —-0.034 |
| Support Vector Regressor | {'C': 10, 'epsilon': 0.5, 'kernel': 'rbf'} | 11.134 | 11.064 | 11.009 | 00.187 |
| Decision Tree | {'max\_depth': 3, 'min\_samples\_split': 2} | 31.946 | 11.986 | 11.498 | —-0.576 |
| Random Forest | {'max\_depth': None, 'min\_samples\_split': 2, 'n\_estimators': 50} | 23.343 | 11.531 | 11.276 | 00.064 |
| KNN Regressor | {'n\_neighbors': 3, 'weights': 'distance'} | 22.325 | 11.525 | 11.218 | 00.071 |

*7.2 Observations and Perspectives*

With a poor performance of negative R2, linear regression was unable to adequately predict nonlinear patterns in the data and was underfitted.

With the lowest MSE (1.134) and greatest R2 (0.187), the Support Vector Regressor (SVR) performed best overall, demonstrating its capacity to identify intricate correlations in the dataset.

Ridge Regression, which probably benefited from regularisation in high-dimensional feature space, performed competitively (R2 = 0.313).

SVR and Ridge fared better than ensemble techniques like Random Forest, which provided only modest increases.

When compared to radiomic characteristics alone, the expressiveness of the model was much enhanced by the addition of CNN embeddings, contrast ratio, and lesion biomarkers.

1. Conclusion

In conclusion, using brain MRI and lesion segmentation data in conjunction with expert-validated clinical evaluations, this work proposes a hybrid deep learning and ensemble regression method for predicting the Expanded Disability Status Scale (EDSS) in MS patients. A thorough preprocessing pipeline that incorporates lesion volume measurements with T1, T2, and FLAIR modalities. Radiomic and contrast-based biomarkers are computed. High-level spatial characteristics are captured by using CNN-derived picture embeddings. For reliable performance estimation, use leave-one-out cross-validation. The most accurate EDSS forecasts are produced via weighted ensemble regression.

With MSE = 1.134 and R2 = 0.187, the final ensemble model showed increased prediction accuracy, indicating that this strategy might work in clinical decision-support systems.

V. Acknowledgements

We would like to express our sincere gratitude to our esteemed professors, Prof. Shoeb Ahmad from the School of Computer Science and Engineering at RV University.

References

1. S. Eshaghi, C. J. Young, R. Wijeratne, et al., “Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data,” *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021.
2. A. Danelakis, K. Theodorakopoulos, and K. S. Nikita, “Fully automated lesion segmentation and EDSS score prediction in multiple sclerosis using FLAIR MRI,” in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2020, pp. 687–692.
3. S. Roy, P. Butman, B. Pham, et al., “Automated lesion segmentation with user feedback for multiple sclerosis,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*, Springer, 2018, pp. 610–618.
4. B. R. Das, A. Ray, and S. K. Ghosh, “Deep learning model for multiple sclerosis lesion segmentation and severity score prediction,” in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1116–1119.
5. M. E. Shiee, C. Bazin, D. Pham, et al., “A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions,” *NeuroImage*, vol. 49, no. 2, pp. 1524–1535, 2010.
6. Y. Feng, L. Zhang, T. Wang, et al., “A deep learning framework for automatic diagnosis of multiple sclerosis with MRI,” in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 1084–1089.
7. H. Valverde, M. Cabezas, R. Roura, et al., “Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach,” *NeuroImage*, vol. 155, pp. 159–168, 2017.
8. C. Sudre, W. Li, T. Vercauteren, et al., “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 240–248.
9. A. Wottschel, A. Alexander, M. Kwok, et al., “Predicting disability progression and cognitive worsening in multiple sclerosis using MRI features,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 91, no. 3, pp. 300–306, 2020.
10. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.