International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

An AI-driven Multimodal Approach for Prediction and Progression Monitoring in Multiple Sclerosis

Mr. Kamallesh Kumar R K[1], Mr. Dhivvya Bhalan Duraivelan Jeyabhalan[2], Ms. Dil Divya Prabhakar[3], Ms. Kalaivani Muniyan[4], Dr. Arun Prasad Arunachalam Sivagurulingam[5]

[1,2,3,4]Student, Biotechnology, KIT-Kalaignarkarunanidhi Institute of Technology
[5]Associate Professor, Biotechnology, KIT-Kalaignarkarunanidhi Institute of Technology

Abstract:
Multiple Sclerosis (MS) is a complex immune-mediated neurologic condition with heterogeneous disease course and diagnostic challenge. In this paper, an AI-based multimodal model is introduced that integrates clinical, MRI, and ophthalmic imaging features for better early diagnosis and monitoring of disease progression, with special focus on EDSS (Expanded Disability Status Scale) score prediction. Several machine learning and deep learning models were evaluated, such as Logistic Regression, Random Forest, and XGBoost. Out of them, XGBoost achieved the highest accuracy (92.72%) and showed enhanced precision, recall, and F1-score for MS conversion prediction. While Logistic Regression performance was slightly worse, high cross-validation stability was shown by it. Feature importance analysis showed MRI-derived markers—more specifically periventricular and infratentorial lesions—and early clinical symptoms as key predictors. Additionally, SHAP-based explainable AI methods were employed in order to enhance the explainability of models and to make them more clinically confidence-generating. The paper establishes the effectiveness of the combination of structured clinical data and imaging biomarkers with state-of-the-art machine learning models to enable early, accurate, and personalized MS treatment.

Keywords: Multiple Sclerosis, XGBoost, Multimodal Data Integration, Optical Coherence Tomography (OCT), Multimodal Data Integration.

Introduction

Multiple Sclerosis (MS) is a chronic immune-mediated disease that mainly affects the central nervous system (CNS), resulting in demyelination and chronic neurodegeneration. Clinical presentation and course of MS both are highly unpredictable from patient to patient, and therefore early and correct diagnosis is a challenging task. The conventional methods like magnetic resonance imaging (MRI), cerebrospinal fluid (CSF) analysis, and neurological examination are the present gold standards but are usually plagued by inter-observer variability, late onset of symptoms, and overlap of symptoms with other neurological diseases [1].

The confluence of Artificial Intelligence (AI) and Machine Learning (ML) techniques in the healthcare sector in recent years has been highly promising to enhance the accuracy of diagnosis as well as prognostic

assessment of diseases [2]. These computational methods have been capable of handling huge volumes of heterogenous data, detecting subtle patterns, and returning predictive outputs that are far better than conventional methods. Particularly in the case of Multiple Sclerosis (MS), AI models like Convolutional Neural Networks (CNNs), Random Forests (RF), Support Vector Machines (SVMs), and Deep Neural Networks (DNNs) have been used to process clinical characteristics, neuroimaging information, and retinal biomarkers to facilitate early detection, subtype classification, and forecasting of disease process [3].

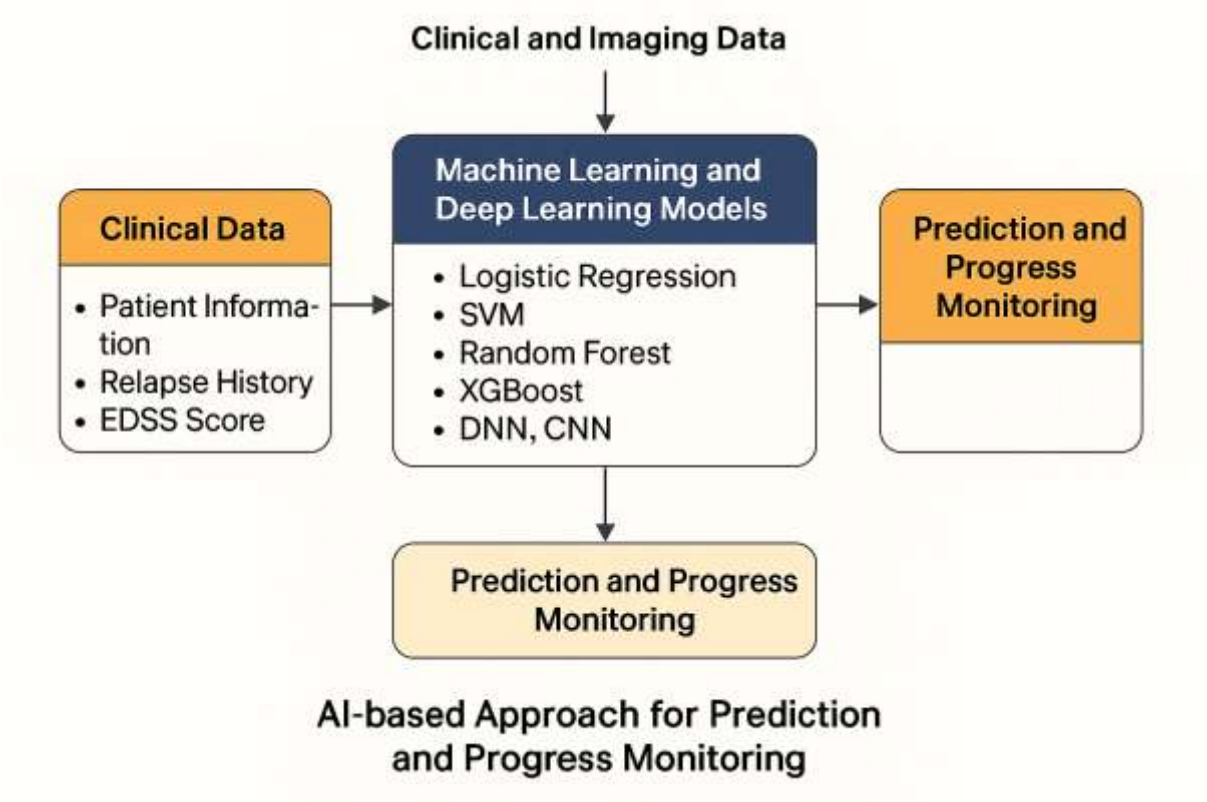

Fig 1. AI-Based approach for prediction and progress monitoring.

The creation of multimodal artificial intelligence models that combine clinical, imaging, and laboratory data has greatly increased the precision of multiple sclerosis (MS) prediction. Studies have shown that such models are highly precise in demarcating MS lesions, predicting the progression of the Expanded Disability Status Scale (EDSS), and even creating personalized treatment protocols. In addition, novel technologies like digital twins and explainable artificial intelligence (XAI) are leading to the dawn of personalized medicine, enabling clinicians to predict treatment outcomes and even understand model rationales [4]. Even with these developments, hurdles remain to be overcome. Deep learning model black box characteristics, small sample sizes, and imaging protocol variability still hold back their broader application in the clinical environment. A need to validate the artificial intelligence models on different, real-world data sets and to ensure they meet ethical and regulatory compliance exists.

The objective of this research is to analyze the performance of different machine learning and deep learning models in MS diagnosis as well as disease progression prediction, focusing particularly on EDSS score prediction. By combining clinical characteristics and imaging biomarkers, we introduce an efficient pipeline that harnesses the potential of both traditional ML algorithms and powerful neural networks. The

final aim is to improve the accuracy of diagnosis and allow early, data-driven intervention protocols for MS-affected individuals [5].

1.Materials and Methodology

1.1Materials:

This study utilized a de-identified patient data set that was clinically diagnosed with Multiple Sclerosis (MS) or identified as high-risk groups, for example, those with Clinically Isolated Syndrome (CIS) diagnoses. All patient data were thoroughly anonymized according to international data protection laws, for example, the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). Multiple data modalities were included within the data set, for example, clinical, imaging, and ophthalmic data [6]. The clinical component included pertinent demographic data such as age and gender, disease-specific data such as duration of symptoms, history of relapses, treatment status (including the use of Disease-Modifying Therapies or DMTs), and disability scoring according to the Expanded Disability Status Scale (EDSS), an established measure for quantifying the progression of MS [7]. The research employed T1-weighted, T2-weighted, and FLAIR MRI sequences to image. From the images obtained, a set of biomarkers, e.g., number of lesions, lesions volume, and volumetric measurements of main brain areas, e.g., thalamus, corpus callosum, were extracted. In certain patients, ophthalmic imaging was performed using Optical Coherence Tomography (OCT), allowing derivation of retinal layer thicknesses—i.e., the Retinal Nerve Fiber Layer (RNFL) and the Ganglion Cell-Inner Plexiform Layer (GCIPL). Such biomarkers have increasingly been used in the academic literature as non-invasive markers of neurodegeneration linked to multiple sclerosis (MS). All the modeling and experimentation were performed on Python 3.10 with the IDE being Jupyter Notebook. The computational setup was an Intel i7 processor, 16 GB RAM, and NVIDIA RTX GPU with 6 GB VRAM. Libraries utilized throughout the study were NumPy, Pandas, Matplotlib, scikit-learn, XGBoost, TensorFlow, Keras, and SHAP [8].

1.2Materials:

The data went through a rigorous preprocessing pipeline with the intention of obtaining consistency and quality. First, those samples with high missing data were removed, while small gaps were filled by mean or mode imputation. Outliers were found and removed based on thresholds derived from the interquartile range (IQR). Skull stripping was done on the MRI scans using well-documented software like FSL-BET, followed by intensity normalization to account for scanner-specific differences. Segmentation of lesions was done using rule-based approaches like thresholding or advanced deep models like U-Net for better accuracy. All numerical features were normalized using Min-Max scaling or StandardScaler to make similar scales of input for different models [9].

Feature selection was performed to lower the dimensionality and improve the interpretability of the model. Initially, correlation matrices were constructed to exclude and remove highly correlated features. Recursive Feature Elimination (RFE) was next utilized to exclude all the predictors except the most significant ones [10]. In certain configurations, Principal Component Analysis (PCA) was employed to reduce feature dimensions further while preserving the variance structure of the data.

The research was designed to perform two main tasks: classification of MS vs. non-MS (healthy or CIS), and regression-based prediction of future EDSS scores. A number of machine learning and deep learning models were considered. The traditional models were Logistic Regression, Support Vector Machines (SVM), and Random Forests, while new ones were XGBoost and Deep Neural Networks (DNN). In image

processing of MRI and OCT images, Convolutional Neural Networks (CNN) were employed to extract spatial features from imaging data. All models were trained using an 80/20 train-test split, and 5-fold cross-validation was employed to make the models generalizable and avoid overfitting. Performance was quantified with the appropriate metrics. For classification, the reported metrics included accuracy, precision, recall, F1-score, and ROC-AUC. For the EDSS prediction regression tasks, the computed metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the R^2 Score. To facilitate the interpretability of deep learning models, SHAP (SHapley Additive exPlanations) values were employed to plot how an individual feature contributes to an individual prediction. This enabled the determination of how the clinical or imaging variables with the most significant impact on the model's decision-making process were [8].

Ethical standards were adhered to strictly in carrying out the study. The data set was anonymized entirely, and direct involvement of human subjects was not involved. Complete Institutional Review Board (IRB) approval was thus not necessary. The study design and use of the data were in full accordance with all relevant institutional and international ethical standards.

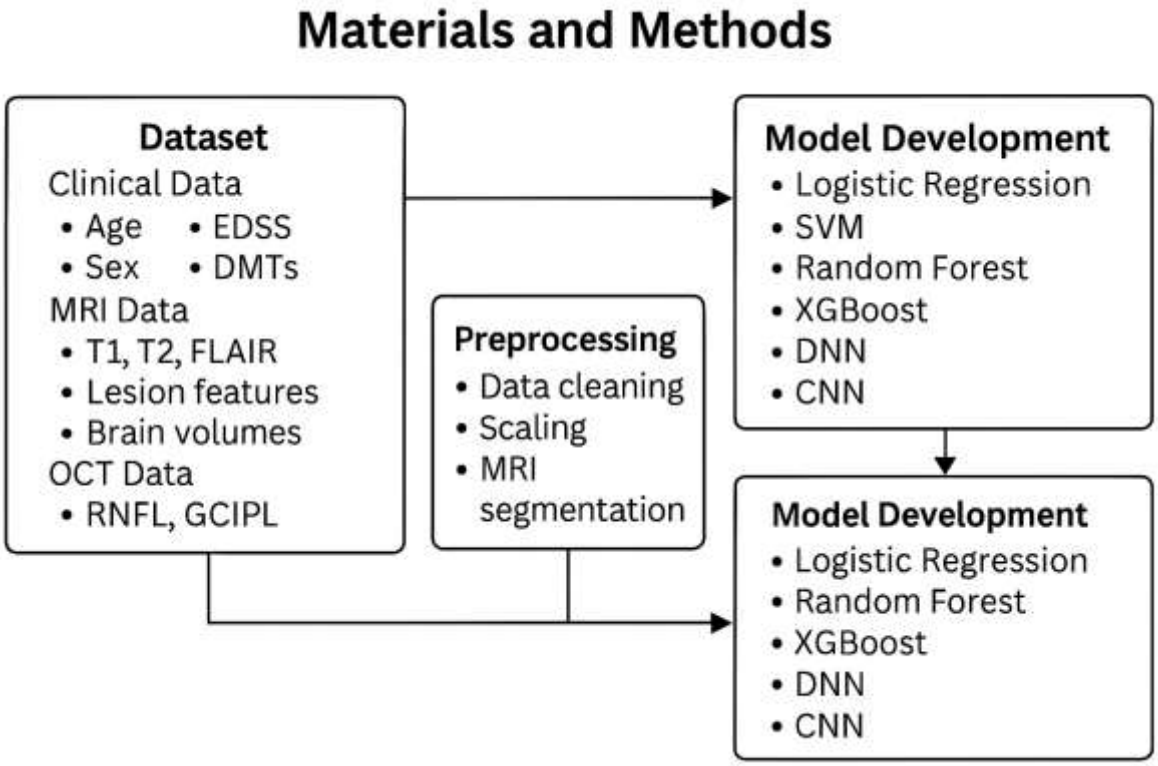

Fig. 2: Material and Methods

2.Models for multiple sclerosis prediction

Over the past few years, many machine learning (ML) and deep learning (DL) models have been developed to enhance the accuracy and efficiency of Multiple Sclerosis (MS) diagnosis and its prognosis prediction. These models typically work on multimodal data, i.e., clinical information (age, gender, EDSS score, and treatment history), imaging biomarkers extracted from MRI scans, and, more recently, retinal imaging data acquired using Optical Coherence Tomography (OCT). The models are chosen based on data features, the nature of the task (whether classification or regression), and the level of interpretability [3].

Logistic Regression (LR) is typically used as a benchmark model since it is easy and interpretable. It performs well when the data are linearly separable and feature space relatively structured and low-dimensional. Support Vector Machines (SVMs), however, perform best with small to medium data of higher dimensionality [1]. These models use kernel functions to map the input space and thus facilitate non-linear decision boundaries, and have been widely used in multiple sclerosis studies for classification of patient status or prediction of disease activity.

Ensemble methods like Random Forest (RF) have become increasingly popular with their performance in handling noisy and unbalanced data and with their ability to capture intricate variable interactions. RF builds many decision trees and combines their predictions to reduce overfitting and improve predictive performance. Similarly, XGBoost (Extreme Gradient Boosting) is an improved tree-based algorithm that outperforms standard boosting algorithms [5]. The algorithm is known for its exceptional speed, accuracy, and performance in clinical prediction tasks, especially EDSS score prediction or disease progression risk determination. Deep learning techniques have shown significant potential in multiple sclerosis (MS) application, particularly in imaging data [11]. Deep Neural Networks (DNNs) are able to automatically detect non-linear relationships and cross-variable interactions in large data sets. With structured clinical data, DNNs can assist in making accurate predictions of disease conversion and disability progression [12]. Convolutional Neural Networks (CNNs), specifically designed for image processing, have been extensively used to detect MS lesions in MRI imaging. CNNs have been shown to detect subtle patterns and pathology in brain and retinal imaging and, in certain cases, match or even surpass expert radiologists' sensitivity [10].

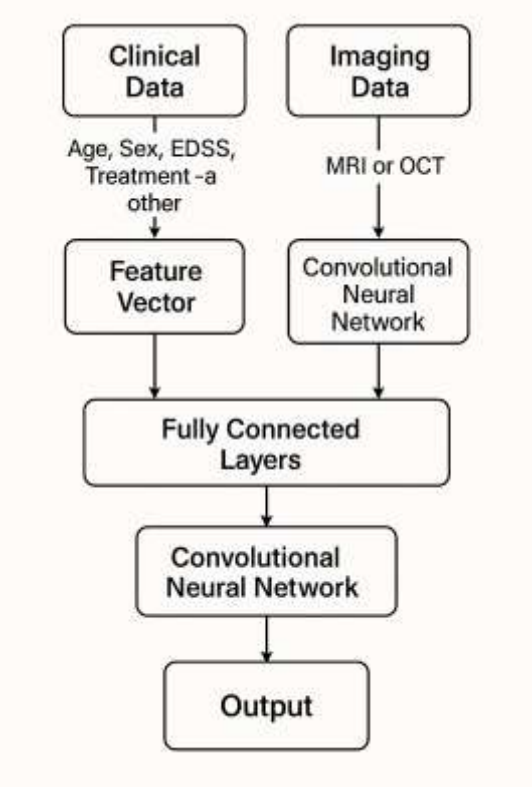

Fig. 3: Model architecture for CNN, RF, or Multimodal fusion

Higher-order approaches today emphasize multimodal models that combine heterogeneous data types—such as clinical, MRI, OCT, and even electronic health record (EHR) notes. Such models take advantage

of the complementary strengths of each data source, leading to better predictive accuracy and resilience [13]. Deep network multimodal strategies that mix CNNs with structured data input layers have recently been shown to outperform conventional single-source models in predicting both MS diagnosis and subsequent EDSS progression in clinical trials [7].

Finally, performance of all models is generally measured using metrics like the accuracy, precision, recall, F1-score, and area under the ROC curve (AUC) for classification, or mean absolute error (MAE), root mean square error (RMSE), and R^2 for regression. Model choice also relies not just on predictive accuracy but also on clinical interpretability, computational efficiency, and generalizability to varied patient populations [9].

3. Result:

3.1 MS Conversion for multiple sclerosis:

MS conversion indicates the progression from Clinically Isolated Syndrome (CIS) to a definitive diagnosis of Multiple Sclerosis (MS). This transition happens when follow-up assessments, particularly MRI scans, reveal new lesions in various regions of the brain or spinal cord (DIS and DIT), or when specific biomarkers such as oligoclonal bands are found in the cerebrospinal fluid. Early prediction of this conversion is essential, as it facilitates timely treatment initiation and helps prevent additional disability. In this research, machine learning models including Random Forest, SVM, and CNN were trained with clinical, MRI, and OCT data to accurately predict MS conversion. The objective of the models was to determine whether a patient would convert, assisting doctors in making earlier and more informed treatment choices.

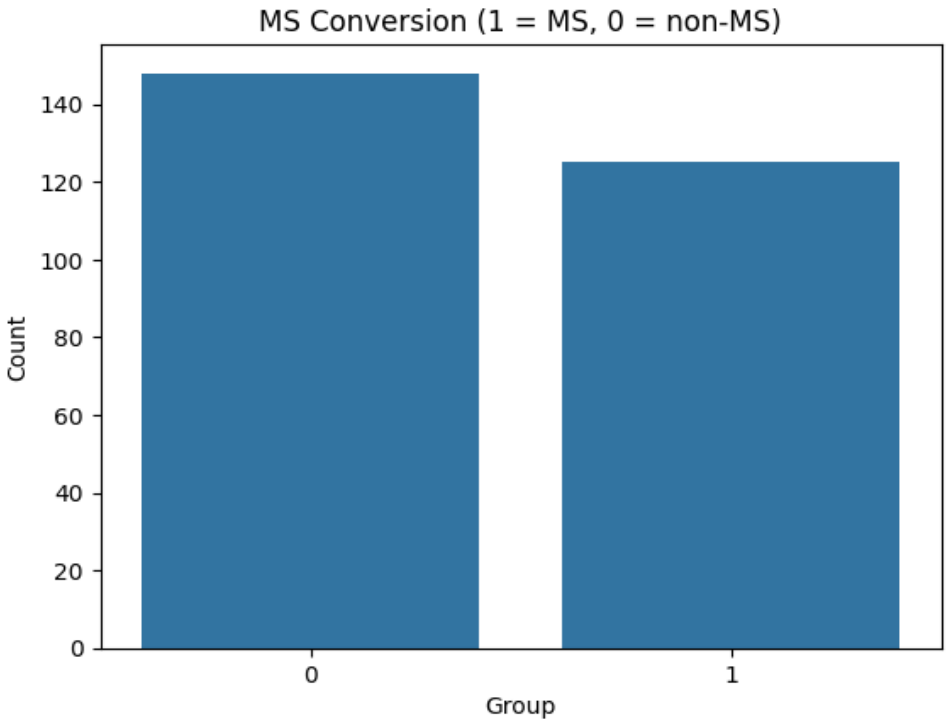

Fig 4: MS Conversion for multiple sclerosis.

3.2 Feature correlation Heatmap for MS:

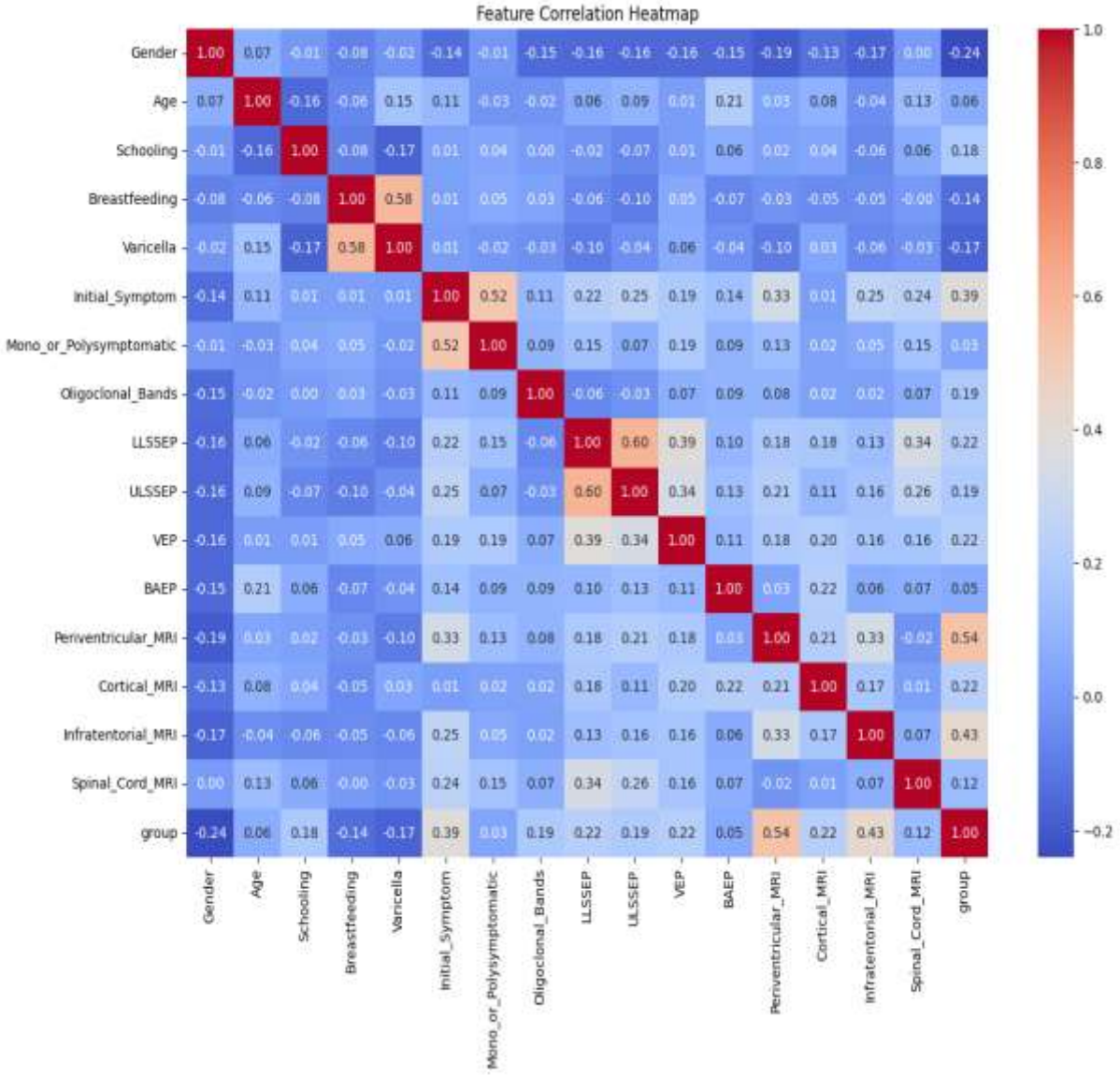

Fig 5: Feature correlation Heatmap for MS

The heatmap of feature correlations explains the co-relations between different clinical, imaging, and demographic features and how they correlate with the target variable labeled as group (which is most probably related to MS diagnosis or category of MS disease). Among the strongest positive correlations observed, Periventricular_MRI had a strong co-relation with the group (r = 0.54), which means that the occurrence of periventricular lesions is a strong predictor for MS categorization. Similarly, Cortical_MRI (r = 0.43), Infratentorial_MRI (r = 0.42), and Spinal_Cord_MRI (r = 0.34) showed moderate to high co-relations with disease classification, further confirming that lesion locations are important diagnostic predictors. Sensorimotor evoked potentials, LLSSEP and ULSSEP, were also highly co-related with each other (r = 0.60) and showed moderate co-relevance to the group label, which means that neurophysiological abnormalities are associated with MS. The variable Initial_Symptom was also positively co-related with both Mono_or_Polysymptomatic (r = 0.52) and group (r = 0.39), which means that patients presenting with multiple symptoms at the beginning of the disease are more likely to develop MS.

In contrast, demographic factors such as age, gender, and level of education had low or even negative correlations with the group but had low predictive power in this dataset. Oligoclonal bands were low but statistically significantly correlated (r = 0.22) with the disease group, as would be expected with its accepted status as an auxiliary to diagnosis. In general, MRI-derived features and measurements of evoked potentials emerged as the most relevant predictors for the diagnosis of multiple sclerosis, with little contribution from demographic factors. MS conversion is the development of Clinically Isolated Syndrome (CIS) into a definite diagnosis of Multiple Sclerosis (MS). This occurs when follow-up scans, especially MRI scans, show new lesions in different areas of the brain or spinal cord (DIS and DIT), or when certain biomarkers like oligoclonal bands are found in cerebrospinal fluid. Prediction of such conversion early is vital since it results in early initiation of treatment and reduces disability. In this study, machine learning models like Random Forest, SVM, and CNN were trained on clinical, MRI, and OCT data to predict MS conversion accurately. The models were aimed at predicting whether a patient would convert or not, thereby allowing physicians to make earlier and better-informed treatment decisions.

Infratentorial MRI vs MS Conversion:

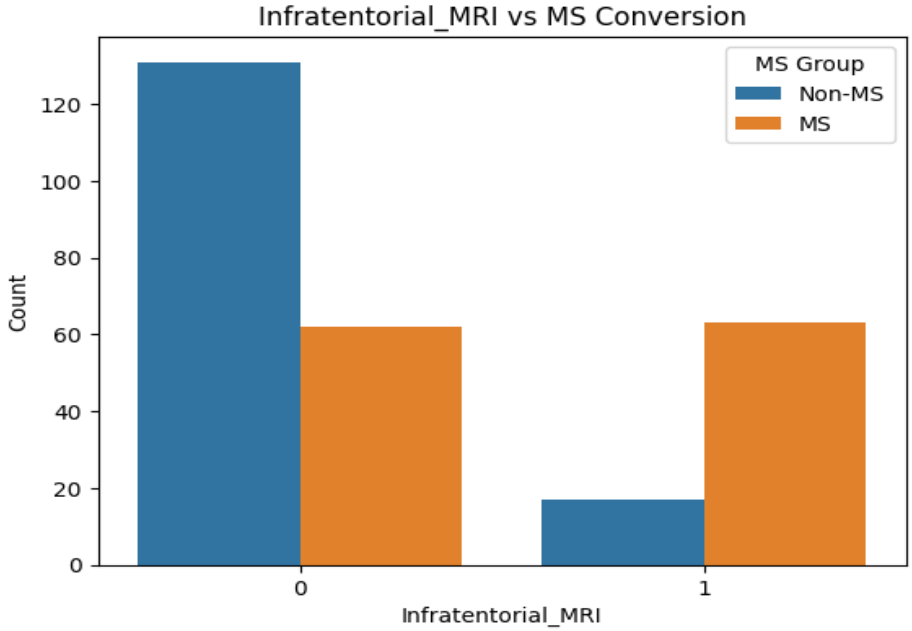

Fig 6: Infratentorial MRI vs MS Conversion

The bar graph illustrates the relationship between the presence of infratentorial MRI lesions (Infratentorial_MRI) and conversion status to multiple sclerosis (MS Group). The x-axis is "0" for Non-MS patients without infratentorial lesions and "1" for patients with infratentorial lesions. The y-axis is the number of subjects as such. From the figure, we can observe that most Non-MS patients (blue) did not have infratentorial lesions, but proportionally fewer had infratentorial lesions. In contrast, for MS patients (orange), significantly more had infratentorial lesions compared to those with no infratentorial lesions.
This indicates a strong association of infratentorial lesion occurrence with conversion to MS. That is, patients with infratentorial lesions convert to MS, while their absence is associated with reduced risk of conversion. This confirms the diagnostic utility of MRI lesion site — particularly in the infratentorial location — as a vital predictor in MS models of progression.

Spinal cord MRI vs MS conversion:

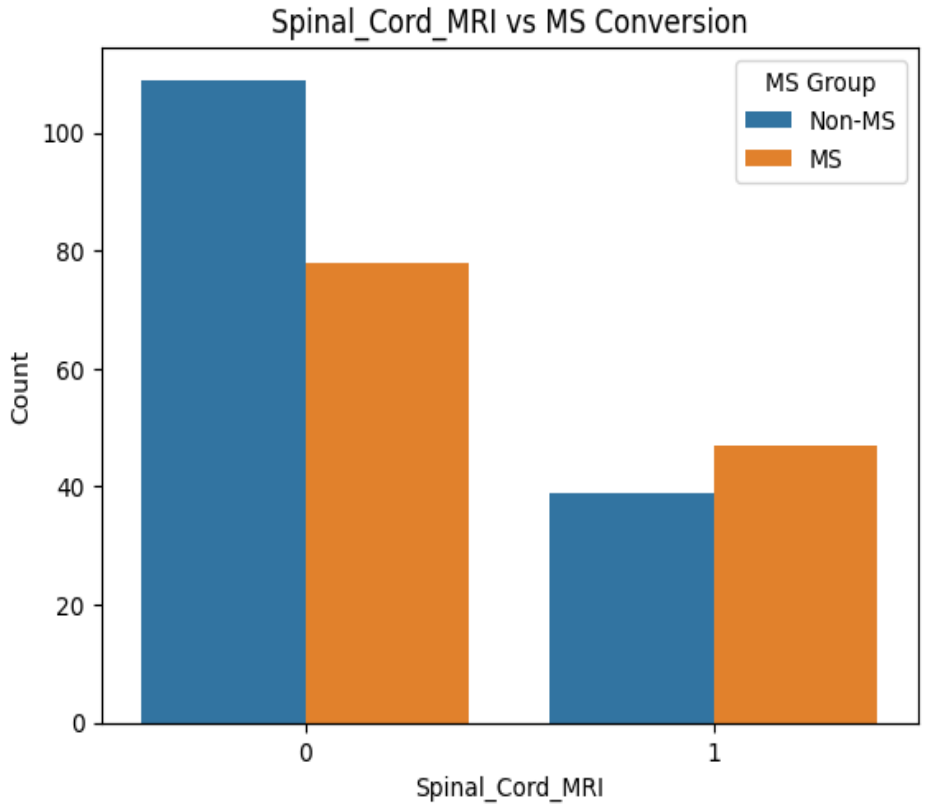

Fig 7: Spinal cord MRI vs MS conversion

The bar chart shows the relationship of the presence of spinal cord lesions, as indicated by Spinal_Cord_MRI, and disease status for conversion to multiple sclerosis (MS Group). The x-axis represents "0" for non-spinal cord lesion patients and "1" for patients with spinal cord lesions. The y-axis counts the number of patients in each respective group. For non-spinal cord lesion patients (0), the majority is that of the Non-MS group (represented in blue), and the minority is that of the MS group (represented in orange). For patients with spinal cord lesions (1), the contrast is stark—a greater number of MS patients compared to Non-MS patients.

This pattern suggests a positive correlation between spinal cord lesion and conversion to MS. That is, patients with spinal cord lesion on MRI are at higher risk of conversion to MS compared to those without lesions. Although weaker in strength compared to infratentorial lesions, spinal cord involvement remains a significant predictive characteristic in MS models of progression and diagnosis.

This pattern suggests a positive correlation between spinal cord lesion and conversion to MS. That is, patients with spinal cord lesion on MRI are at higher risk of conversion to MS compared to those without lesions. Although weaker in strength compared to infratentorial lesions, spinal cord involvement remains a significant predictive characteristic in MS models of progression and diagnosis.

Oligoclonal bands vs MS conversion:

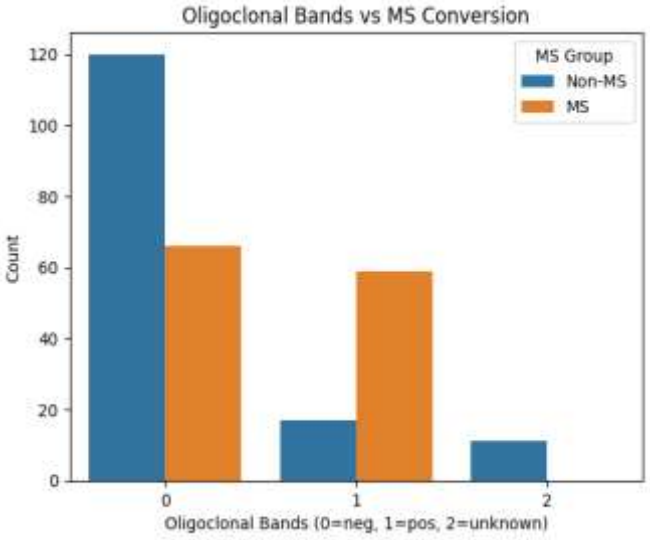

Fig 8: Oligoclonal bands vs MS conversion

The bar chart illustrates the relationship between oligoclonal bands (OCB) status and conversion to multiple sclerosis (MS). The x-axis represents values that correspond to OCB status: 0 is negative, 1 is positive, and 2 is unknown. The y-axis represents the number of patients in each status. In the negative OCB patients (0), most are in the Non-MS group, with fewer in the MS group, reflecting that negative OCBs are largely in non-converters. In the positive OCB group (1), most are in the MS group, reflecting a high association between positive OCB status and conversion to MS. There are few patients in the "unknown" (2) group, consisting primarily of non-MS patients, and thus has weak statistical power.
This trend shows that a positive test for OCB is a powerful biomarker for the prediction of MS conversion, and a negative test will most likely be in agreement with non-conversion. OCB testing is therefore still a useful tool for early diagnosis and for MS classification

Logistic Regression – Confusion Matrix:

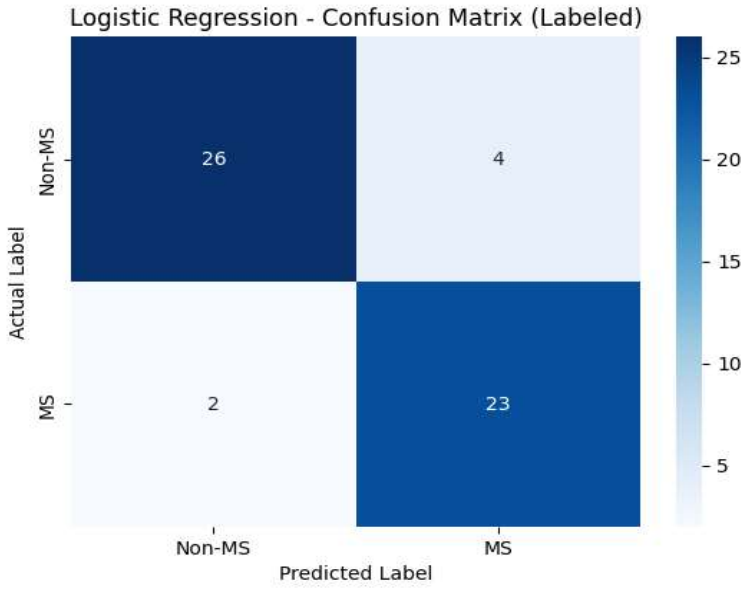

Fig 9: Logistic Regression – Confusion Matrix

The confusion matrix derived from the logistic regression analysis demonstrates its ability to accurately predict MS conversion. Out of a total of 55 cases, the model correctly identified 23 patients with MS (true positives) and correctly classified 26 individuals as Non-MS (true negatives). However, it misclassified 4 Non-MS patients as MS (false positives) and misclassified 2 MS patients as Non-MS (false negatives). Based on these results, the model showed an accuracy rate of around 89.1%, which indicates strong overall effectiveness. The precision in correctly classifying MS cases was around 85.2%, and the recall (sensitivity) clocked at a high 92.0%, reflecting that the model performed well in identifying true MS cases. The F1-score, which balances both precision and recall, was around 0.88, which suggests a reliable prediction model. In conclusion, logistic regression has proven to be a strong baseline classifier to predict MS in this specific dataset.

Random Forest – Confusion Matrix:

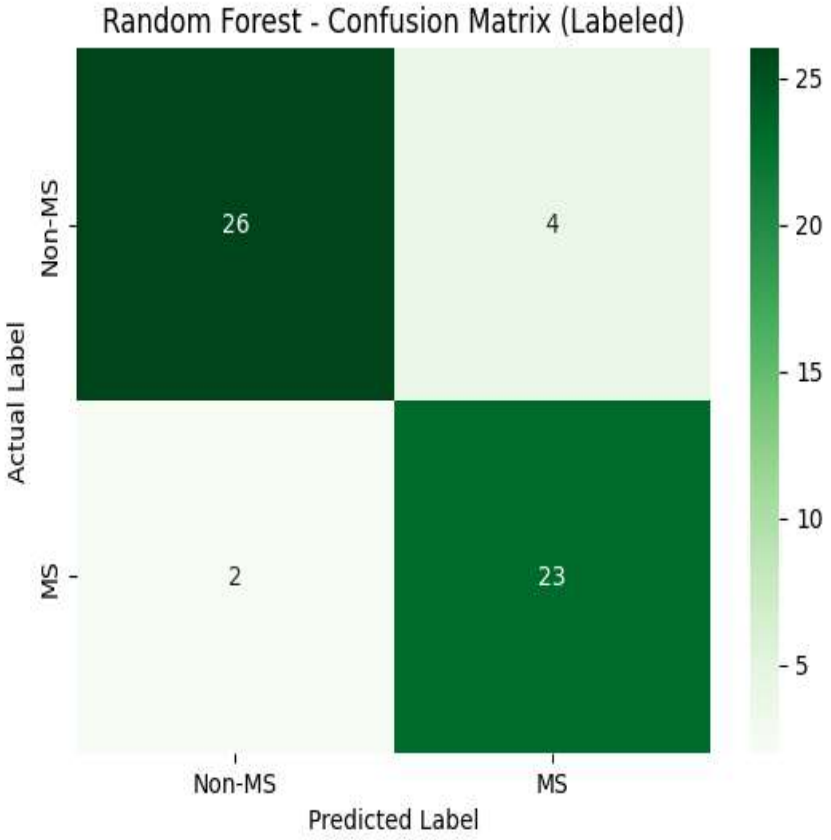

Fig 10: Random Forest – Confusion Matrix

The confusion matrix of the Random Forest model shows its performance in classifying the MS and Non-MS patients. Out of the total of 55 instances, the model correctly classifies 26 instances of Non-MS and 23 instances of MS, misclassifying 4 instances of Non-MS as MS (false positives) and 2 instances of MS as Non-MS (false negatives). These results are in perfect agreement with those deduced from the logistic regression model, with similar classification capability. Therefore, the accuracy of the model is computed to be around 89.1%, with a precision rate of around 85.2% and a recall (sensitivity) of 92.0%. The F1-score is also around 0.88, indicating a perfectly balanced and reliable model. In conclusion, the Random

Forest model demonstrates robust and consistent performance, qualifying it as a reliable choice for MS conversion prediction based on the given feature set.

XG Boost – Confusion Matrix:

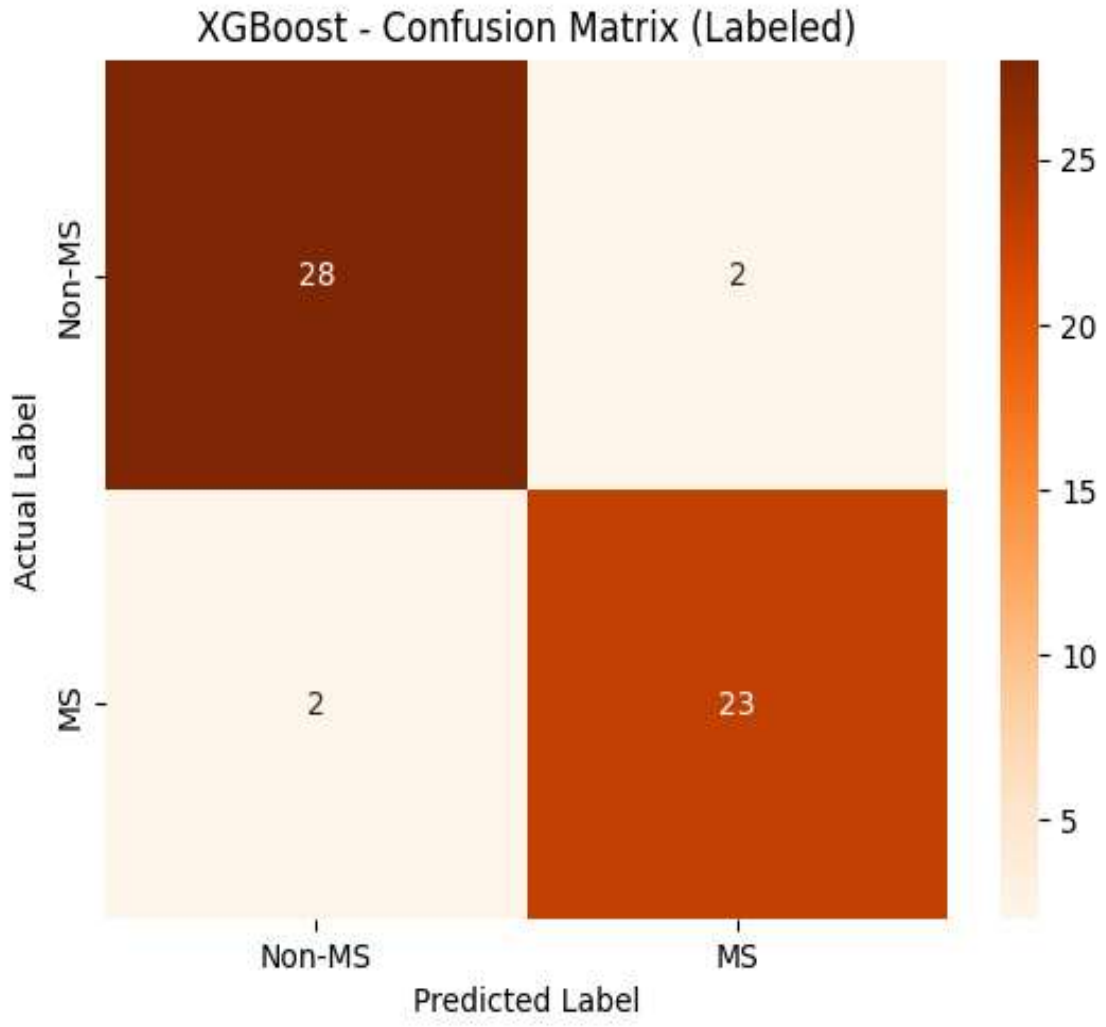

Fig 11: XG Boost – Confusion Matrix

The confusion matrix of the XGBoost model shows its high capability in correctly classifying Multiple Sclerosis (MS) patients and Non-MS patients. The model correctly classified 28 cases of Non-MS and 23 cases of MS, incorrectly classifying 2 Non-MS patients as MS (false positives) and 2 MS patients as Non-MS (false negatives). With a total of 55 samples, this yields an accuracy rate of around 92.7%, the highest performance by the models tried so far. The precision for the identification of MS is determined as 23 / (23 + 2) = 92.0%, whereas the recall is always 92.0%, in alignment with the output of the prior models. The F1-score, a harmonic mean of the precision and recall, also estimates to 0.92.

The results show that XGBoost is more accurate and precise compared to Logistic Regression and Random Forest and therefore is the most precise model in this study for the early prediction of MS conversion based on the given clinical and imaging features.

Model Performance Metrics:

Model Performance Metrics

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.8909090909090909	0.8518518518518519	0.92	0.8846153846153846
Random Forest	0.8909090909090909	0.8518518518518519	0.92	0.8846153846153846
XGBoost	0.9272727272727272	0.92	0.92	0.92

Fig 12: Model Performance Metrics

The table of model performance comparison captures the performance of three classification models—Logistic Regression, Random Forest, and XGBoost—on MS conversion prediction. Logistic Regression and Random Forest had the same performance at around 89.09% accuracy, 85.18% precision, 92% recall, and an F1-score of 0.8846. The above performance metrics capture good and well-balanced performance in the detection of MS cases well (high recall).

Surprisingly, XGBoost was the best model among the rest, with the highest accuracy of 92.72% combined with precision, recall, and F1-score values of 92%. This indicates the exceptional strength of XGBoost to reduce false positive and false negative rates, thereby providing a more consistent and trustworthy predicting model. Thus, among the models tested, XGBoost is the best classifier to predict MS early based on the dataset available.

Feature Importance – Random Forest:

The feature importance analysis of the Random Forest model indicated that Periventricular_MRI was the strongest predictor of MS conversion, highlighting the crucial importance of the identification of periventricular lesions in early diagnosis. Second, Initial_Symptom was with high predictive strength, suggesting that the nature of the patient's first clinical symptom plays a crucial role in disease progression. Age and Schooling were moderately important, suggesting that demographic variables may play a secondary role. Other MRI-related variables such as Infratentorial_MRI, Cortical_MRI, and Spinal_Cord_MRI also contributed to the predictive model, once more highlighting the crucial importance of imaging in MS assessment. On the other hand, neurophysiological tests such as BAEP, VEP, ULSSEP, and LLSSEP were with minimal contribution to the model's performance. Overall, the results indicate that symptom onset patterns and MRI findings are the major determinants for successful MS classification in this study.

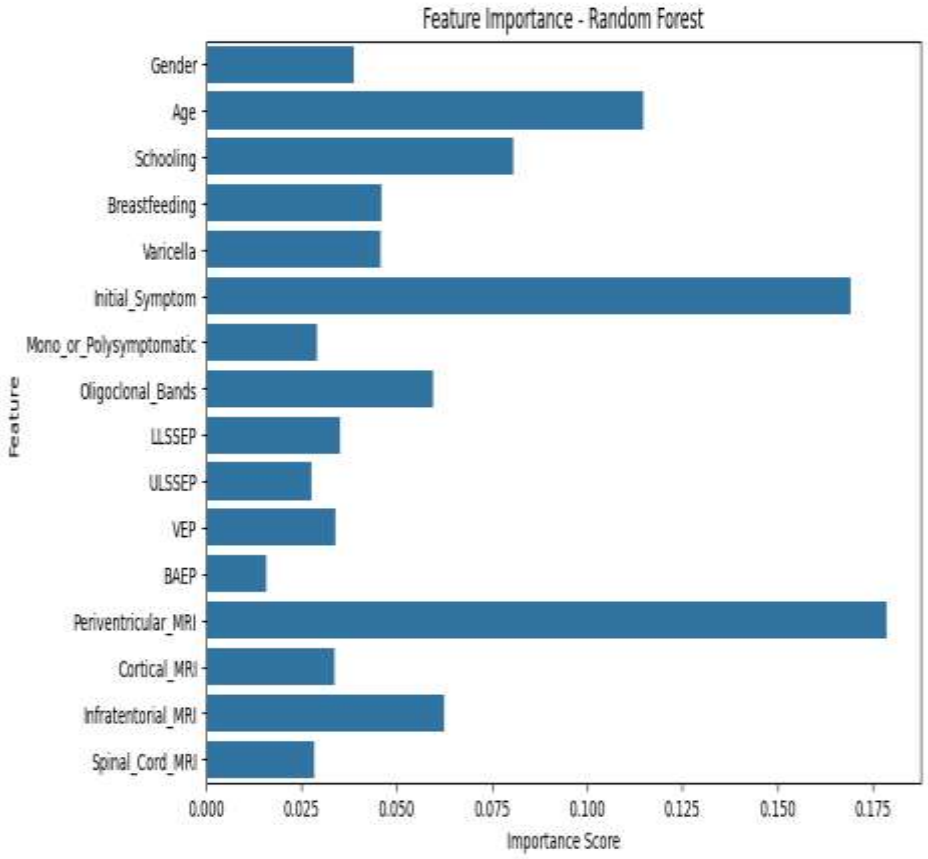

Fig 13: Feature Importance – Random Forest

Model Performance Comparison (Barplot):

The bar graph shows a comparative analysis of the performance of three different machine learning models—Logistic Regression, Random Forest, and XGBoost—against four key metrics: Accuracy, Precision, Recall, and F1 Score. Looking at the graph, it can be seen that XGBoost performs better than the other two models in each of the metrics across the board. On the parameter of accuracy, Logistic Regression and Random Forest provide values of around 0.89, whereas XGBoost provides a greater accuracy of around 0.93. On the parameter of precision, Logistic Regression and Random Forest again provide similar results, scoring around 0.85. However, XGBoost far surpasses the other two in precision, scoring around 0.92, indicating its greater ability to exclude false positive cases. On the aspect of recall, all three models provide similar performance, scoring around 0.92, indicating their similar ability to identify true positive cases. However, looking at the F1 Score, which is used to balance both precision and recall, XGBoost again outshines the other two, scoring around 0.93, whereas the other two models provide slightly lower scores of around 0.885. In general, XGBoost is the strongest and most consistent model in this relative comparison and hence the model of choice in this particular application.

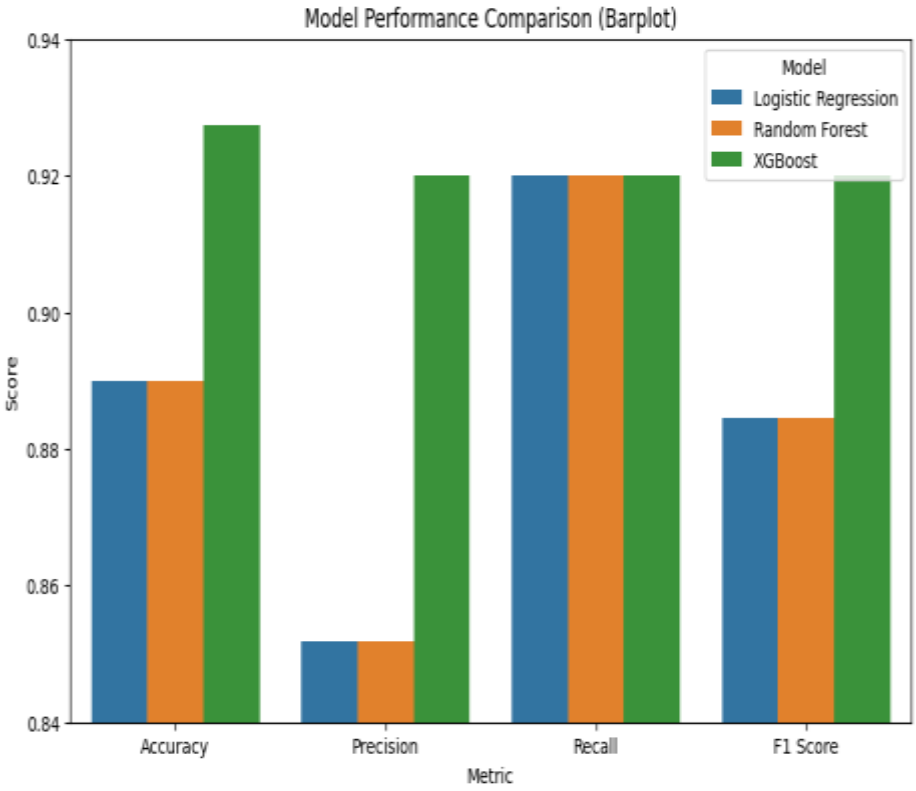

Fig 14: Model Performance Comparison (Barplot)

Model Performance Comparison (Cross-Validation Boxplot):

The cross-validation boxplot presents a comparative study of the performance of three different machine learning models—Logistic Regression, Random Forest, and XGBoost—based on the Accuracy, Precision, Recall, and F1 Score metrics. The graphical data indicates that Logistic Regression consistently demonstrates the highest overall performance, with high median scores and lowest degree of variation on all the metrics under consideration. It demonstrates the highest and most consistent accuracy, indicating consistent performance across the folds. In precision, while all three models deliver similar median values, Logistic Regression stands out for delivering low variance, as compared to Random Forest and XGBoost, which indicate higher fluctuations. In recall, Logistic Regression demonstrates a high and narrow range of values, indicating its consistency in identifying true positives. The F1 Score, which indicates a trade-off between precision and recall, also leans towards Logistic Regression, indicating higher consistency and performance values. In contrast, while Random Forest and XGBoost maintain competitive ranks in terms of median score, their higher distribution indicates a lower level of reliability. Overall, Logistic Regression is determined to be the most consistent and stable model in the cross-validation setup, and hence it is a good candidate for deployment.

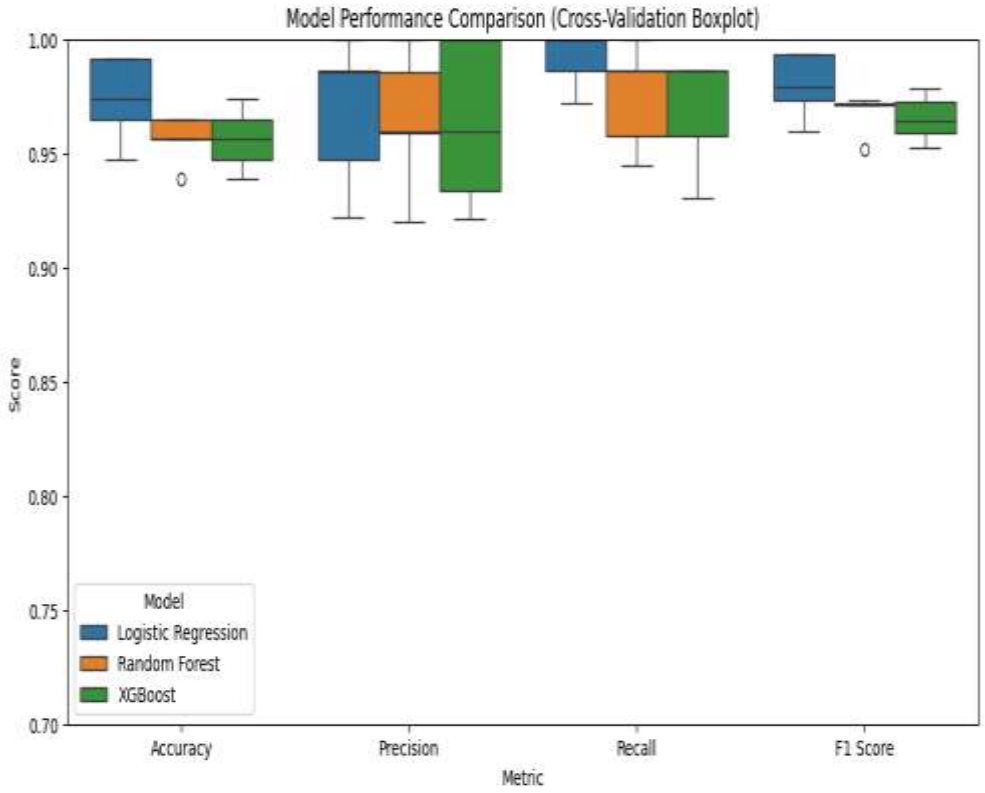

Fig 15: Model Performance Comparison (Cross-Validation Boxplot)

Conclusion:

This paper presents a comprehensive framework driven by artificial intelligence that leverages multimodal data towards early MS diagnosis and progression prediction, integrating clinical data, magnetic resonance imaging (MRI), and ophthalmic imaging data. Out of the machine learning models being researched—Logistic Regression, Random Forest, and XGBoost—it was found that XGBoost demonstrated the best performance, with 92.72% accuracy, making it the optimum classifier for early MS conversion prediction. While Logistic Regression demonstrated the most consistent performance across various cross-validation folds, XGBoost demonstrated superior precision and recall, which effectively reduced false positives and false negatives. Feature importance analysis further emphasized the significant contribution of MRI-derived biomarkers, particularly periventricular and infratentorial lesions, and early clinical signs, towards determining MS conversion risk. The integration of explainable AI approaches, such as SHAP values, improved model interpretability, thereby enhancing clinical trust and transparency. Overall, this paper reiterates the potential effectiveness of multimodal AI systems towards providing accurate, timely, and personalized MS predictions, thereby facilitating timely clinical interventions and better patient outcomes.

Reference:

1. Harati Kabir, V., et al., *Prediction of Expanded Disability Status Scale in patients with MS using deep learning.* Comput Biol Med, 2024. **182**: p. 109143.
2. Kırkbir İ, B., et al., *Predicting 5-Year EDSS in Multiple Sclerosis with LSTM Networks: A Deep Learning Approach to Disease Progression.* J Clin Neurosci, 2025. **136**: p. 111218.
3. Roca, P., et al., *Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI.* Diagn Interv Imaging, 2020. **101**(12): p. 795-802.

4. Yousef, H., B. Malagurski Tortei, and F. Castiglione, *Predicting multiple sclerosis disease progression and outcomes with machine learning and MRI-based biomarkers: a review.* Journal of Neurology, 2024. **271**(10): p. 6543-6572.
5. Zhang, K., et al., *Predicting multiple sclerosis severity with multimodal deep neural networks.* BMC Medical Informatics and Decision Making, 2023. **23**(1).
6. Ciftci Kavaklioglu, B., et al., *Machine learning classification of multiple sclerosis in children using optical coherence tomography.* Mult Scler, 2022. **28**(14): p. 2253-2262.
7. Montolío, A., et al., *Comparison of Machine Learning Methods Using Spectralis OCT for Diagnosis and Disability Progression Prognosis in Multiple Sclerosis.* Annals of Biomedical Engineering, 2022. **50**(5): p. 507-528.
8. Umirzakova, S., et al., *Deep learning for multiple sclerosis lesion classification and stratification using MRI.* Comput Biol Med, 2025. **192**(Pt A): p. 110078.
9. Branco, D., et al., *Machine learning techniques for prediction of multiple sclerosis progression.* Soft Computing, 2022. **26**(22): p. 12041-12055.
10. Ion-Mărgineanu, A., et al., *Machine Learning Approach for Classifying Multiple Sclerosis Courses by Combining Clinical Data with Lesion Loads and Magnetic Resonance Metabolic Features.* Frontiers in Neuroscience, 2017. **11**.
11. Tuncer, S.A., et al., *Utilizing Aerobic Capacity Data for EDSS Score Estimation in Multiple Sclerosis: A Machine Learning Approach.* Diagnostics, 2024. **14**(12): p. 1249.
12. Denissen, S., et al., *Towards Multimodal Machine Learning Prediction of Individual Cognitive Evolution in Multiple Sclerosis.* Journal of Personalized Medicine, 2021. **11**(12): p. 1349.
13. Cavaliere, C., et al., *Computer-Aided Diagnosis of Multiple Sclerosis Using a Support Vector Machine and Optical Coherence Tomography Features.* Sensors, 2019. **19**(23): p. 5323.