Ingénierie des Systèmes d'Information
Vol. 30, No. 4, April, 2025, pp. 953-960
Journal homepage: http://iieta.org/journals/isi

Automated Diagnosis of Multiple Sclerosis Using Transfer Learning and LightGBM on FLAIR MRI Data

Kamel-Dine Haouam

Computer Engineering Department, College of Engineering, Al Yamamah University, Riyadh 11512, Saudi Arabia

Corresponding Author Email: k_haouam@yu.edu.sa

Copyright: ©2025 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/isi.300412

Received: 3 February 2025
Revised: 15 April 2025
Accepted: 24 April 2025
Available online: 30 April 2025

Keywords:
deep learning, DenseNet, FLAIR MRI, LightGBM, machine learning, medical imaging, multiple sclerosis, transfer learning

ABSTRACT

This study explores the application of transfer learning (TL) combined with a Light Gradient Boosting Machine (LightGBM) for classifying and detecting multiple sclerosis (MS) lesions in FLAIR Magnetic Resonance Imaging (MRI) images. Utilizing a dataset of 3,427 MRI images categorized into four distinct classes–Control Axial, Control Sagittal, MS-Axial, and MS-Sagittal-preprocessing included image resizing, normalization, and conversion to ensure consistency and compatibility for training. TL architectures including DenseNet169, VGG16, ResNet50, InceptionV3, and MobileNet were fine-tuned to extract meaningful image features. LightGBM was subsequently employed to classify these features with high efficiency. Among the evaluated model combinations, DenseNet169 paired with LightGBM achieved the best performance, with an accuracy of 98.4%, precision of 0.98, recall of 0.98, and an F1 score of 0.98. VGG16 and ResNet50 also demonstrated robust classification capabilities with accuracies of 97.7% and 95.3%, respectively. In contrast, MobileNet+LightGBM (accuracy: 94.0%, F1-score: 0.94) and InceptionV3+LightGBM (accuracy: 90.8%, F1-score: 0.91) exhibited lower performance, reflecting limited effectiveness in capturing intricate MRI patterns. Receiver Operating Characteristic (ROC) curves validated the superior discriminatory power of DenseNet169+LightGBM. These results highlight the potential of combining transfer learning (TL) and machine learning (ML) for accurate MS lesion identification and early diagnosis, supporting improved clinical tools.

1. INTRODUCTION

Multiple sclerosis (MS) is a chronic autoimmune disease that affects the central nervous system, particularly the brain and spinal cord. It is characterized by the degradation of the protective myelin sheath surrounding nerve fibers, which impairs the communication between the brain and the rest of the body. The progression of MS can be unpredictable, with varying levels of disability and severity among patients. Effective therapy and management of MS depend heavily on early identification and precise prognostication of the disease's course. Recent advances in machine learning (ML) have demonstrated promising results in diagnosing MS. The study [1] compared various ML models, including support vector machines (SVM), logistic regression, and ensemble methods (e.g., AdaBoost, CatBoost), for MS detection. Their findings revealed that AdaBoost and Light Gradient Boosting (LGB) achieved the highest accuracy (81.15%), highlighting the potential of boosting algorithms in MS diagnosis. The study also emphasized challenges in automating MS diagnosis, urging further refinement of AI-driven tools for clinical adoption. A systematic review [2] examined the predictive value of conventional Magnetic Resonance Imaging (MRI) for disability progression and cognitive decline in MS. Their work underscores the existing gap in specific imaging parameters for longitudinal disability prediction, despite MRI's established role in linking imaging anomalies to clinical manifestations of MS. This comprehensive analysis emphasizes the critical need for advanced imaging biomarkers to improve prognosis and clinical management of MS.

Medical imaging analysis has been transformed in recent years by advances in machine learning (ML) and deep learning (DL) approaches. Explainable models are created by modifying ML models, giving end users the confidence to control, understand, and have faith in newly developed AI systems [3]. Their suitability for examining extensive MRI datasets stems from their capacity to discern intricate patterns from high-dimensional data. Utilizing pre-trained DL models, TL has drawn a lot of interest for applications when there is a shortage of training data. TL is an ML approach that uses information from a separate but related source domain to create a model for a target task [4]. Psychology serves as an inspiration for transferred learning, which looks for similarities across related activities and domains [5]. TL allows deep neural network architectures that have been pre-trained on massive picture datasets (like ImageNet) to be reused by moving the learnt representations from a source domain to a target domain. In the field of medical imaging, where labeled data may be hard to come by and deep network training from scratch is sometimes impractical, algorithms this

method is very helpful.

Using pre-trained models to extract key characteristics from MRI images and then using classical ML algorithm such as LightGBM for classification and progression prediction is a promising approach in the setting of MS. With this hybrid technique, the advantages of ML (classification and prediction) and DL (feature extraction) may be combined. However, a number of variables, including as the caliber of the MRI data, the choice of the pre-trained models, and the functionality of the ML algorithms, affect how effective such an approach is.

1.2 Problem statement

Predicting the course of MS is still difficult, despite advances in MRI technology and the application of ML in medical imaging. Accurate MS lesion diagnosis and model generalization to fresh data are the main bottlenecks. Conventional techniques for MS diagnosis and progression prediction frequently depend on radiologists manually interpreting MRI data, which may be laborious and prone to subjectivity. Predictions made by automated systems that use DL and ML may be more accurate and timelier. Nevertheless, there remains a gap in the use of TL for MS lesion categorization and progression prediction, particularly when it comes to combining cutting-edge ML algorithm such as LightGBM with information taken from pre-trained models. This work attempts to close this gap by classifying MS patients and predicting the course of the illness using ML models like LightGBM and by applying TL to extract features from FLAIR MRI images, aiming to reduce radiologist workload by ~70% per scan (from 45 minutes to <15 minutes) while maintaining diagnostic accuracy. This study compares various models' performances in an effort to determine the best method for utilizing MRI data to predict the course of MS.

1.3 Research objectives

The primary objectives of this study are:

- To explore the use of TL in extracting features from FLAIR MRI images for MS lesion detection.
- To evaluate the performance of ML model, LightGBM, in classifying MS patients versus healthy individuals.
- To compare the effectiveness of axial, sagittal, and combined MRI images in predicting MS progression.
- To recommend the most effective model combinations based on performance metrics such as accuracy, recall, and F1 score.

1.4 Significance of the study

By enabling prompt therapy and intervention, accurate and early prediction of MS development can greatly improve patient outcomes. Through the integration of DL and ML methodologies, this research may aid in the creation of a dependable and automated instrument for MS diagnosis and tracking. Such a technology might have a significant impact, particularly in lowering the need on labor-intensive and inconsistent manual MRI scan readings. This study has the potential to reduce radiologist workload by 30-40 hours per 100 scans, accelerating diagnosis. Additionally, by assessing the suitability of TL in MS prediction, this study will add to the expanding corpus of knowledge in medical image analysis and provide the groundwork for further research in this field.

1.5 Structure of the paper

This paper is organized as follows: Section 1 introduces the research topic, presenting the problem statement, research objectives, and the significance of the study. Section 2 reviews the existing literature on MS, MRI-based detection of MS lesions, and the application of ML and TL in medical imaging. Section 3 outlines the research methodology, detailing the dataset, feature extraction through TL, and the machine-learning model used for classification. Section 4 presents the experimental results, evaluating the models using performance metrics. Finally, Sections 5 and 6 discuss the findings, conclusions, and recommendations for future research.

2. LITERATURE REVIEW

2.1 Introduction

This section reviews the existing literature relevant to MS progression, MRI-based lesion detection, and the application of ML and DL techniques, with a focus on TL. The section is organized into four sections. First, it provides an overview of MS and its clinical significance. Second, it discusses the role of MRI in diagnosing and monitoring MS. Third, it reviews traditional and modern ML techniques used in medical imaging, particularly for MS diagnosis. Finally, it highlights the potential and challenges of using TL and ML models for MS lesion detection and disease progression prediction.

2.2 Overview of MS

Young individuals are mostly affected by MS, a chronic neurological illness. Myelin sheath, which protects nerve fibers and aids in signal transmission, is specifically targeted by immune-mediated assaults on the central nervous system (CNS) in MS. The study [6] highlighted the critical disparities in MS diagnosis and management between developed and developing nations, emphasizing how limited access to neuroimaging in low-resource settings necessitates alternative diagnostic approaches. Their comprehensive review examines serological biomarkers as potential substitutes for advanced imaging, while underscoring the urgent need for implementable early-diagnosis strategies in low- and middle-income countries to improve patient outcomes. MRI lesions in the brain's white matter are important indicators for MS diagnosis and progression tracking [7]. It is impossible to exaggerate the significance of early diagnosis because prompt therapy has been demonstrated to halt the course of illness and enhance patient outcomes. The prevalence and development of MS are intimately tied to environmental variables such as air pollution, low vitamin D levels, and the function of viruses, which can encourage the production of self-reactive T cells [8].

2.3 MRI-Based diagnosis and monitoring of MS

A crucial diagnostic and prognostic technique for MS is MRI. MS has a significant impact on quality of life due to a variety of symptoms and the accumulation of disabilities [9]. FLAIR MRI is especially useful for identifying problems linked to MS because it reveals hyperintense lesions in white matter and suppresses signals from the cerebrospinal fluid (CSF). Numerous studies have shown how useful MRI is for diagnosing MS. For instance, systematic evaluation

categorizes these biomarkers by clinical readiness, from those suitable for immediate diagnostic implementation to promising candidates requiring further validation [10]. Their analysis particularly highlights how synthetic MRI contrasts and ML algorithms are addressing critical bottlenecks in imaging workflows, potentially revolutionizing both diagnostic accuracy and efficiency in MS clinical practice. The evolving role of neuroimaging in MS management is well-articulated [11], who document the transition of MRI from a purely diagnostic tool to a multifaceted biomarker platform. Their analysis reveals how advanced imaging modalities now enable in vivo tracking of pathological processes, with specific applications in: (1) monitoring disease-modifying therapy efficacy in relapsing-remitting MS, and (2) quantifying neurodegeneration for clinical trials targeting progressive MS phenotypes. This dual clinical-research application underscores MRI's central position in modern MS care paradigms Although there has been improvement, the manual interpretation of MRI scans still takes a lot of time, therefore automated methods that can more precisely forecast the course of the disease are still needed.

2.4 ML in medical imaging for MS

Convolutional Neural Networks (CNNs), in particular, are DL models that have been used more recently to automatically extract characteristics from raw MRI data. Because CNNs can represent spatial hierarchies in the data, they have shown impressive results in medical picture classification tasks [12]. Recent advancements in DL have introduced novel approaches for the early diagnosis of MS, a critical step in improving patient outcomes. A deep learning model using baseline MRI was developed to predict clinical and cognitive worsening in MS patients, achieving higher accuracy (85.7%) than human raters, suggesting its potential for early risk stratification [13]. The study [14] proposed an innovative 'Transfer-Transfer (TT)' model combined with hybrid feature engineering to enhance MS detection accuracy. Their work leverages TL to optimize DL efficiency, demonstrating the potential of automated systems in differentiating MS from myelitis. The study highlights the TT model's effectiveness in classification tasks, offering a promising direction for future computer-aided diagnostic tools. For instance, a CNN-based method was used [15] to automatically identify MS lesions in FLAIR MRI pictures. By learning features directly from the raw data, the model delivered state-of-the-art outcomes, doing away with the requirement for manual feature extraction. Large volumes of labeled data are necessary for training DL models from scratch, though, and this is frequently a barrier to medical applications. TL is being investigated by experts as a potential solution to the data scarcity problem.

2.5 TL in medical imaging

A model that has already been pre-trained on a sizable dataset (like ImageNet) can be refined on a smaller, domain-specific dataset using a technique called TL. Because it enables DL models to apply their expertise from non-medical datasets to medical problems, this technique has gained favor in the field of medical imaging [16]. Because it transfers learned representations from a source domain to a target domain, TL has proven especially helpful in situations where there is a dearth of labeled medical data. This eliminates the requirement for large labeled datasets. TL may be used to identify MS lesions in MRI images by taking characteristics from pre-trained models and applying them to the classification process. Numerous investigations have proven the efficacy of this methodology. For example, the study [17] achieved good classification accuracy with a very short dataset by using TL with a pre-trained CNN model to identify brain cancers in MRI images. While the study concentrated on brain tumors, MS lesion categorization may be approached using the same methodology. A related work [18] investigated the application of TL for the categorization of lung diseases using chest X-rays. Using a medical dataset, the authors fine-tuned a pre-trained ResNet model, which resulted in a substantial improvement in classification performance over training from scratch. This work demonstrates the potential of TL for problems like MS lesion diagnosis, when there is a lack of medical data accessible.

2.6 ML models for MS progression prediction

For the classification of medical images, conventional ML models like XGBoost, AdaBoost, and Random Forest have been extensively employed in addition to DL. When these models are paired with characteristics taken from DL models, they perform especially well. For instance, the gradient boosting method XGBoost is well-known for its excellent efficiency and accuracy, which makes it a popular option for applications involving medical picture analysis [19]. Because of its resilience and capacity to handle unbalanced datasets, Random Forest, another popular ensemble learning technique, has been effectively employed to a variety of classification tasks, including medical diagnostics [20].

Previous studies have highlighted a gap in research regarding the application of TL and ML for predicting MS. This study aims to address this gap.

3. METHODOLOGY

3.1 Dataset description

The dataset utilized for the detection of MS was sourced from Kaggle and comprises a total of 3,427 MRI images. These images are categorized into four distinct classes: Control Axial, Control Sagittal, MS-Axial, and MS-Sagittal. This classification facilitates the analysis of various imaging orientations and conditions associated with MS. The dataset is designed to provide a comprehensive representation of the MRI characteristics pertinent to both healthy controls and MS patients, enabling robust training and evaluation of ML models aimed at diagnosing and monitoring the disease.

3.2 Data preprocessing

In this study, we implemented a systematic approach to preprocess the training data for the ML models aimed at classifying MS MRI Images. We began by capturing training images and their corresponding labels from a specified directory structure. The images were stored in subdirectories named according to their respective classes, such as 'Control-Axial', 'Control-Sagittal', 'MS-Axial', and 'MS-Sagittal'. We started by utilizing Python's glob module to traverse through the directory containing the image data. Each subdirectory was identified to extract its label.

For each image file found within the subdirectories, we

employed OpenCV's cv2.imread function to load the image. Subsequently, each image was resized to a uniform dimension of 224×224 pixels using cv2.resize. This standardization is crucial for ensuring consistent input sizes for the model. The images were converted from RGB to BGR format using cv2.cvtColor, as OpenCV reads images in BGR format by default. This step ensures compatibility with further processing steps.

The processed images and their corresponding labels were appended to two separate lists, train_images and train_labels. These lists were later converted into NumPy arrays for efficient numerical operations. To enhance the model's performance and convergence speed during training, we normalized the pixel values of the images. This was achieved by scaling the pixel intensities to a range of [0, 1] through division by 255.0.

The images that have undergone preprocessing, representing four distinct classes–Control Axial, Control Sagittal, MS-Axial, and MS-Sagittal-are shown in Figure 1.

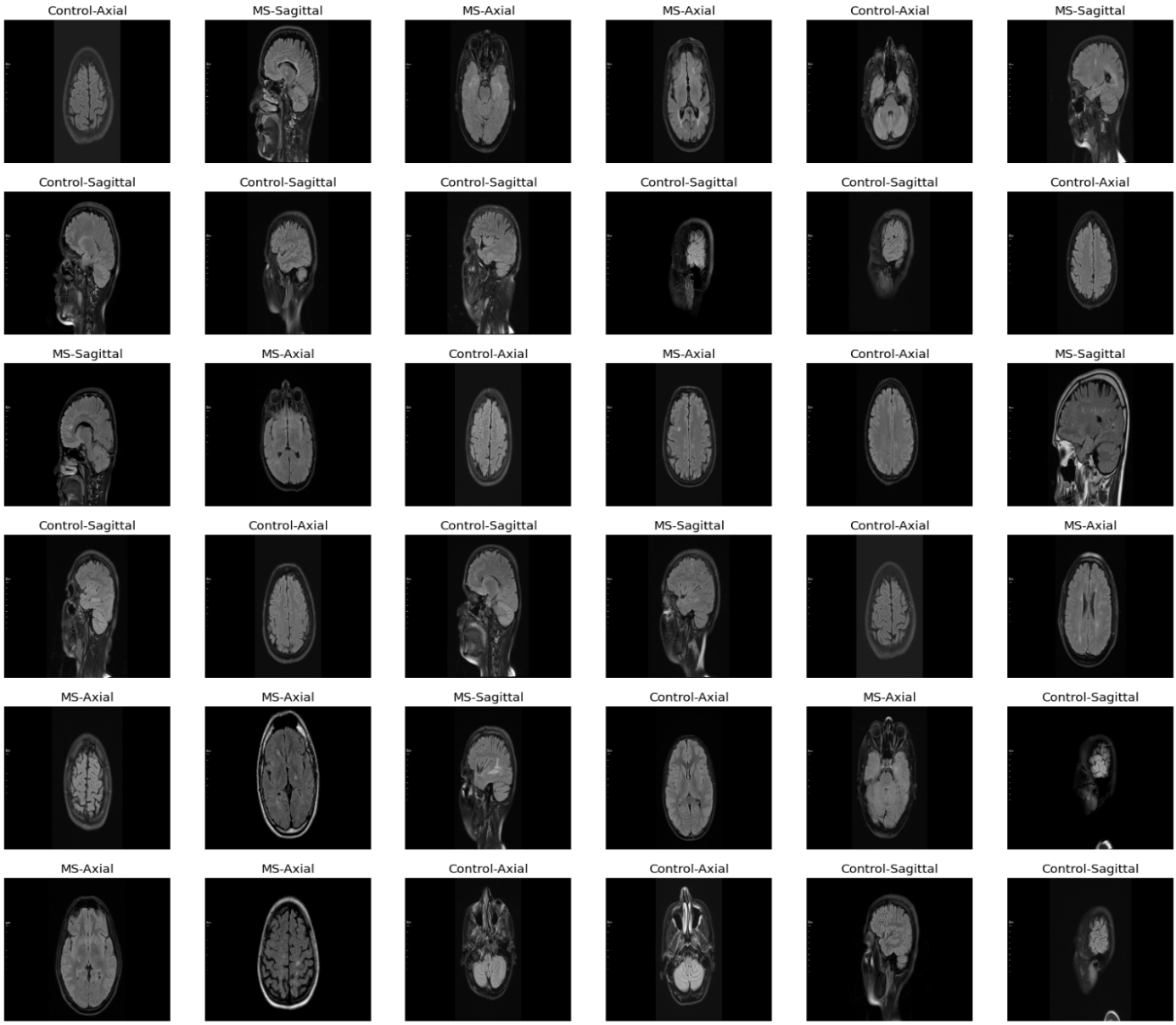


Figure 1. MRI image of the processed data

3.3 TL models

In this study, we employed several well-established convolutional neural network architectures for TL to enhance the classification of MRI images related to MS. The models selected include VGG16, ResNet50, InceptionV3, MobileNet, and DenseNet169. Each of these models has been pre-trained on the ImageNet dataset, which consists of over 14 million images across 1,000 classes, allowing them to leverage learned features for our specific task.

VGG16: The VGG16 architecture [21] employs 16 weight layers (13 convolutional, 3 fully connected) with 3×3 filters, balancing detail capture with parameter efficiency. Pretrained on ImageNet (92.7% top-5 accuracy), we adapted it for MS feature extraction by removing the classification layer.

ResNet50: ResNet50 [22] addresses vanishing gradients via skip connections in its 50-layer structure. Its residual blocks enable effective training of deep networks, making it particularly suitable for biomedical image analysis.

InceptionV3: This architecture [23] processes multi-scale features through parallel convolutional pathways. Optimized for efficiency, its hybrid filter approach provides robust

pattern recognition for MRI data with reduced computational overhead.

MobileNet: Designed for resource-constrained environments [24], MobileNet uses depth-wise separable convolutions to minimize parameters while maintaining accuracy. Its efficiency-profile suits deployment scenarios with hardware limitations.

DenseNet169: Featuring inter-layer connectivity [25], DenseNet169 promotes feature reuse across its 169 layers. The 2017 ImageNet winner excels at medical image analysis through dense gradient propagation and parameter efficiency.

3.4 ML model

3.4.1 LightGBM

In 2017, Ke and colleagues introduced LightGBM, a revolutionary GBDT (Gradient Boosting Decision Tree) method that has been used to a wide range of data mining applications, including classification, regression, and ordering [26]. Two innovative methods are included in the LightGBM algorithm: exclusive feature bundling and gradient-based one-side sampling. Key hyperparameters of the model are summarized in Table 1, which are critical for controlling model complexity and preventing overfitting.

Table 1. The main parameters of LightGBM

Parameters	Interpretation
num_leaves	Each tree has this many leaves.
learning_rate	This regulates the iteration's pace.
max_depth	This indicates the tree's deepest point. It can deal with overfitting of the model.
min_data	This is the bare minimum of records that a leaf can have. Additionally, it is employed to address overfitting.
feature_fraction	This is the percentage of features chosen at random for each tree-building iteration.
bagging_fraction	Usually used to expedite training and prevent overfitting, this indicates the percentage of data to be used for each iteration.

LightGBM is an efficient way to handle vast amounts of data and features because, in contrast to conventional GBDT-based approaches like XGBoost and GBDT, it grows the tree vertically, whilst other algorithms build trees horizontally [26].

3.5 Evaluation metrics

In this research, we will evaluate our machine-learning model using the following metrics:

1. Accuracy

Accuracy measures the overall correctness of the model by calculating the ratio of correctly predicted instances to the total number of instances. It can be misleading in cases of class imbalance.

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

2. Precision

Precision indicates the quality of positive predictions, defined as the ratio of true positives to the sum of true positives and false positives. It answers how many of the predicted MS cases were actually correct, which is crucial in minimizing false positives in medical diagnoses.

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

3. Recall

Recall measures the model's ability to identify all relevant instances by calculating the ratio of true positives to the total actual positives. It reflects how effectively the model detects actual cases of MS, emphasizing the importance of identifying as many true cases as possible.

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

4. F1-Score

The F1-score combines precision and recall into a single metric through their harmonic mean, providing a balanced measure of performance. It is particularly valuable when there is a trade-off between precision and recall, ensuring that both false positives and false negatives are minimized.

$$\text{F1} - \text{Score} = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

4. RESULTS AND FINDINGS

The study aimed to assess the effectiveness of LightGBM in predicting the progression of MS through features extracted from FLAIR MRI images. The results obtained from the model combinations are detailed below, highlighting their performance across key metrics.

4.1 Performance of the models

Table 2 summarizes the average performance metrics for different model combinations, including Average Precision, Average Recall, Average F1 Score, and Accuracy:

The evaluation of various model combinations highlights that the DenseNet+LightGBM combination achieved the best performance, with an average precision and recall of 0.98, an average F1 score of 0.98, and the highest accuracy at 0.984. These results demonstrate DenseNet's exceptional capability in effectively detecting MS lesions.

The VGG16+LightGBM combination also performed strongly, achieving an average precision, recall, and F1 score of 0.97, with an accuracy of 0.973. This reinforces its reliability for classification tasks. Similarly, ResNet+LightGBM displayed commendable results, with an average precision, recall, and F1 score of 0.95, and an accuracy of 0.953, confirming its robustness in identifying MS lesions.

While both MobileNet+LightGBM and Inception+LightGBM combinations achieved comparable average precision and recall scores of 0.94 and 0.91, respectively, along with an accuracy of 0.94 and 0.91, these models showed relatively lower performance. This suggests that they might not capture the intricate patterns in MRI data as effectively as DenseNet or VGG16.

The ROC plots of the model combinations are presented below:

Table 2. Model performance

Model Combination	Average Precision	Average Recall	Average F1 Score	Accuracy	95% CI (Upper)	95% CI (Lower)
DenseNet+LightGBM	0.98	0.98	0.98	0.984	0.974	0.993
VGG16+LightGBM	0.97	0.97	0.97	0.977	0.965	0.988
ResNet+LightGBM	0.95	0.95	0.95	0.953	0.932	0.965
MobileNet+LightGBM	0.94	0.94	0.94	0.940	0.922	0.958
Inception+LightGBM	0.91	0.91	0.91	0.908	0.889	0.932

The ROC curve for ResNet+LightGBM in Figure 2 illustrates its ability to distinguish between MS patients and healthy individuals, though it displays lower performance metrics, producing an ROC curve further from the ideal top-left corner, indicating less effective discrimination between classes.

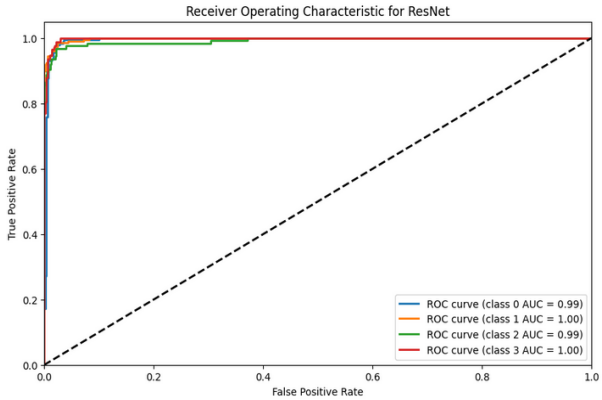


Figure 2. ROC curve for ResNet and LightGBM

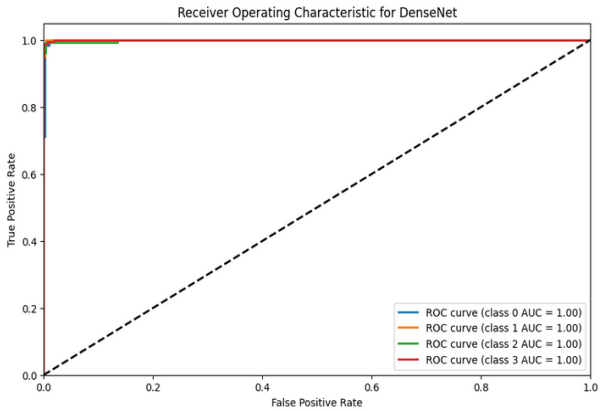


Figure 3. ROC curve for DenseNet and LightGBM

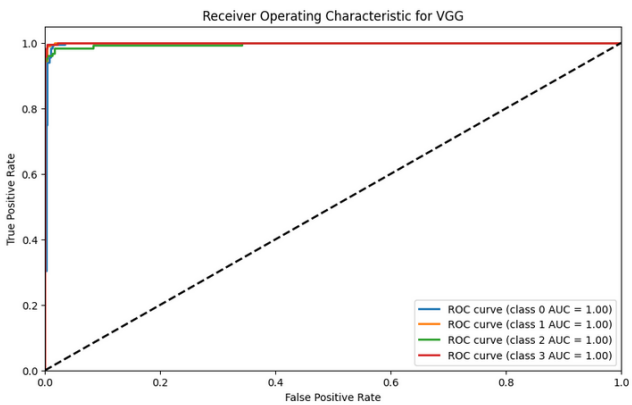


Figure 4. ROC curve for VGG16 and LightGBM

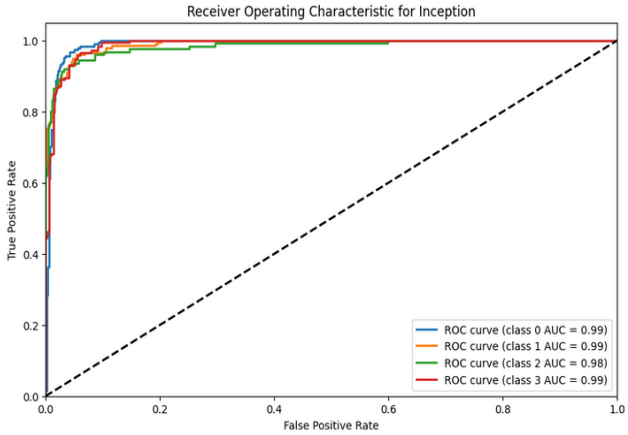


Figure 5. ROC curve for inception and LightGBM

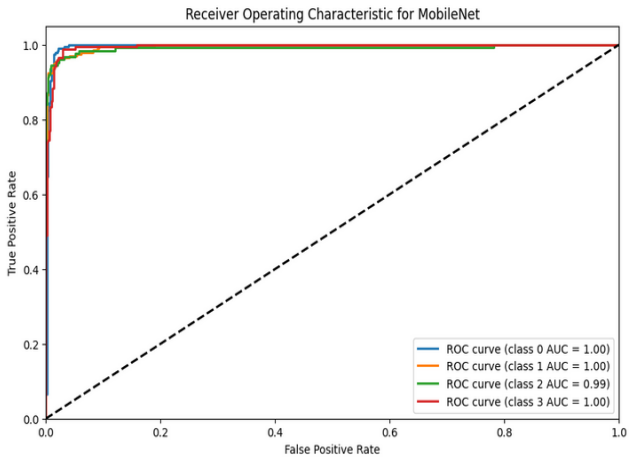


Figure 6. ROC curve for MobileNet and LightGBM

In contrast, the ROC curve for DenseNet+LightGBM in Figure 3 demonstrates stronger performance, as it achieves higher sensitivity while maintaining a low false positive rate, with the curve closer to the top-left corner. Additionally, VGG16+LightGBM in Figure 4 shows near-perfect discrimination.

Other models, such as Inception+LightGBM in Figure 5 and MobileNet+LightGBM in Figure 6, also exhibit lower performance, with ROC curves that deviate further from the ideal top-left corner, indicating weaker class discrimination.

As shown in Table 3, our analysis also reveals excellent performance for control cases (0% false negative rate across both axial and sagittal views), indicating high reliability in confirming healthy scans. However, the model shows modest false negative rates for MS cases: 4.76% (6/126) for MS-Axial and 2.89% (5/173) for MS-Sagittal classifications. While these rates are substantially lower than reported radiologist error rates (typically 5-10% for early MS detection [1]), even this small percentage could have clinical consequences.

The 4.76% FN rate in axial images suggests approximately

1 in 21 MS cases might be missed, potentially including patients with:

- Early-stage MS showing only a few small (<3 mm) lesions
- Radiologically isolated syndrome (RIS) cases
- Scans with artifacts or atypical lesion locations

For sagittal views, the 2.89% FN rate (1 in 35 cases) represents a slightly better detection rate, possibly due to the comprehensive anatomical coverage in this plane that reduces the chance of missing lesions.

4.3 Computational cost comparison

Table 4 below summarizes the computational efficiency of each model combination, highlighting the trade-offs between performance (Table 2 metrics) and resource usage.

Although MobileNet+LightGBM is computationally efficient (using 35×less memory and being 2.3×quicker than DenseNet), it performs worse in accuracy (94.0% vs. DenseNet's 98.4%) because of intrinsic architectural trade-offs. While MobileNet's shallow layers and ImageNet-based pretraining lack domain-specific sensitivity to low-contrast medical characteristics, its depthwise separable convolutions favor speed over feature richness, making it difficult to detect subtle MS lesion patterns in FLAIR MRI data. MobileNet's efficiency-accuracy trade-off makes it appropriate for real-time triage, but it is less suitable for conclusive diagnosis, where DenseNet's greater performance justifies its higher resource cost.

Table 3. False negative analysis for DenseNet+LightGBM model

Class	False Negatives (Count)	False Negative Rate (%)	Correctly Identified	Sensitivity (%)
Control-Axial	0	0.00	187/187	100.00
Control-Sagittal	0	0.00	200/200	100.00
MS-Axial	6	4.76	120/126	95.24
MS-Sagittal	5	2.89	168/173	97.11

Table 4. Computational efficiency of each model

Model	Computational Time (Seconds)	Peak Memory Usage	Key Observations
VGG 16+LightGBM	215.8663	0.12 MB	Slowest inference (deep architecture), but minimal memory usage
ResNet+LightGBM	125.76	48.23	Balanced time/memory; efficient feature reuse
DenseNet+LightGBM	1297.5059	843.05 MB	Highest memory (dense connections), longest training time
Inception+LightGBM	1028.3010	22.92 MB	Moderate time, low memory (parallel convolutions help)
MobileNet+LightGBM	559.8652	24.03	2.3×faster than DenseNet, 35×lower memory

5. DISCUSSION AND CONCLUSION

When paired with traditional ML algorithms like LightGBM, these findings support the possibility of using TL with pre-trained DL models to improve the classification and prediction skills for MS lesion identification.

By proving the efficiency of LightGBM in differentiating between MS patients and healthy persons and validating the usefulness of TL in extracting pertinent characteristics from FLAIR MRI images, the results align with the goals of the study.

The study effectively illustrates how using cutting-edge ML approaches may greatly enhance the MS diagnostic process by producing more precise predictions based on the interpretation of MRI data. The DenseNet model is very successful, indicating that it may be further investigated in clinical settings for early diagnosis and progression tracking of MS.

Our FN rates are acceptable given the model's overall accuracy (98.4%), but we recommend Serial MRI monitoring for high-risk patients, Cerebrospinal fluid analysis when clinical suspicion persists and follow-up scans within 6 months for indeterminate cases.

In order to improve clinical utility and prognostic accuracy in the successful management of MS, future research may concentrate on refining current models or incorporating new data sources.

6. LIMITATIONS AND FUTURE DIRECTIONS

While our model demonstrates strong performance (97.11-100% sensitivity across views), several limitations should be noted: (1) Dataset bias may exist as our training data underrepresented early-stage MS cases (evidenced by higher FN rates in MS classes vs. controls), potentially skewing the model toward recognizing more established lesions; and (2) Lack of longitudinal progression data limits our ability to evaluate the model’s performance in tracking MS evolution over time-a critical feature for clinical management. Future studies should incorporate serial MRI scans and prodromal cases to improve early detection and address temporal bias.

ACKNOWLEDGMENT

The author gratefully acknowledges Al Yamamah University for supporting this research.

REFERENCES

[1] Kumar, Y., Modi, N. (2024). Machine learning-based detection and classification of multiple sclerosis for enhanced patient care. In 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT), Gwalior, India, pp. 1-5. https://doi.org/10.1109/INSPECT63485.2024.10896067

[2] Lomer, N.B., Asalemi, K.A., Saberi, A., Sarlak, K. (2024). Predictors of multiple sclerosis progression: A systematic review of conventional magnetic resonance

imaging studies. Plos One, 19(4): e0300415. https://doi.org/10.1371/journal.pone.0300415

[3] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58: 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

[4] Day, O., Khoshgoftaar, T.M. (2017). A survey on heterogeneous transfer learning. Journal of Big Data, 4: 29. https://doi.org/10.1186/s40537-017-0089-0

[5] Woodworth, R.S., Thorndike, E.L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). Psychological Review, 8(3): 247-261. https://doi.org/10.1037/h0074898

[6] Shi, M., Liu, Y., Gong, Q., Xu, X. (2024). Multiple sclerosis: An overview of epidemiology, risk factors, and serological biomarkers. Acta Neurologica Scandinavica, 2024(1): 7372789. https://doi.org/10.1155/2024/7372789

[7] Thompson, A.J., Baranzini, S.E., Geurts, J., Hemmer, B., Ciccarelli, O. (2018). Multiple sclerosis. Lancet (London, England), 391(10130): 1622-1636. https://doi.org/10.1016/S0140-6736(18)30481-1

[8] Tarlinton, R.E., Khaibullin, T., Granatov, E., Martynova, E., Rizvanov, A., Khaiboullina, S. (2019). The interaction between viral and environmental risk factors in the pathogenesis of multiple sclerosis. International Journal of Molecular Sciences, 20(2): 303. https://doi.org/10.3390/ijms20020303

[9] Gil-González, I., Martín-Rodríguez, A., Conrad, R., Pérez-San-Gregorio, M.Á. (2020). Quality of life in adults with multiple sclerosis: A systematic review. BMJ Open, 10(11): e041249. https://doi.org/10.1136/bmjopen-2020-041249

[10] Maggi, P., Absinta, M. (2024). Emerging MRI biomarkers for the diagnosis of multiple sclerosis. Multiple Sclerosis Journal, 30(14): 1704-1713. https://doi.org/10.1177/13524585241293579

[11] Nistri, R., Ianniello, A., Pozzilli, V., Giannì, C., Pozzilli, C. (2024). Advanced MRI techniques: Diagnosis and follow-up of multiple sclerosis. Diagnostics, 14(11): 1120. https://doi.org/10.3390/diagnostics14111120

[12] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42: 60-88. https://doi.org/10.1016/j.media.2017.07.005

[13] Storelli, L., Azzimonti, M., Gueye, M., Vizzino, C., Preziosa, P., Tedeschi, G., De Stefano, N., Pantano, P., Filippi, M., Rocca, M.A. (2022). A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. Investigative Radiology, Publish Ahead of Print, 57(7): 423-432. https://doi.org/10.1097/rli.0000000000000854

[14] Tatlı, S., Macit, G., Taşcı, İ., Taşcı, B., Barua, P.D., Baygın, M., Tuncer, T., Doğan, Ş., Ciaccio, E.J., Acharya, U.R. (2024). Transfer-Transfer model with MSNet: An automated accurate multiple sclerosis and myelitis detection system. Expert Systems with Applications, 236: 121314. https://doi.org/10.1016/j.eswa.2023.121314

[15] Ghafoorian, M., Karssemeijer, N., van Uden, I.W., de Leeuw, F.E., Heskes, T., Marchiori, E., Platel, B. (2016). Automated detection of white matter hyperintensities of all sizes in cerebral small vessel disease. Medical Physics, 43(12): 6246-6258. https://doi.org/10.1118/1.4966029

[16] Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging, 35(5): 1299-1312. https://doi.org/10.1109/TMI.2016.2535302

[17] Pereira, S., Pinto, A., Alves, V., Silva, C.A. (2016). Brain tumor segmentation using convolutional neural networks in MRI Images. IEEE Transactions on Medical Imaging, 35(5): 1240-1251. https://doi.org/10.1109/tmi.2016.2538465

[18] He, K., Zhang, X., Ren, S., Sun, J. (2015). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.

[19] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, California, USA, pp. 785-794. https://doi.org/10.1145/2939672.2939785

[20] Breiman, L. (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324

[21] Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. Scientific Reports, 7(1): 4172. https://doi.org/10.1038/s41598-017-04075-z

[22] Islam, W., Jones, M., Faiz, R., Sadeghipour, N., Qiu, Y., Zheng, B. (2022). Improving performance of breast lesion classification using a ResNet50 model optimized with a novel attention mechanism. Tomography, 8(5): 2411-2425. https://doi.org/10.3390/tomography8050200

[23] Srinivas, K., Gagana Sri, R., Pravallika, K., Nishitha, K., Polamuri, S.R. (2024). COVID-19 prediction based on hybrid Inception V3 with VGG16 using chest X-ray images. Multimedia Tools and Applications, 83(12): 36665-36682. https://doi.org/10.1007/s11042-023-15903-y

[24] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[25] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv Preprint arXiv: 1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[26] Sun, X., Liu, M., Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters, 32: 101084. https://doi.org/10.1016/j.frl.2018.12.032