



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Multiple Sclerosis Detection Using Coarse Tree Classifier

Loida F. Hermosura CpE. MSIT, Isaac Angelo M. Dioses, Marlon L. Castro, Arjay Ordonio,
Jesusimo L. Dioses Jr.

Department of Information Technology, Northeastern College, Santiago City, Isabela

School of Information Technology, MAPUA University, Makati Campus, Philippines

Department of Information Technology, Northeastern College, Santiago City, Isabela

Department of Information Technology, Northeastern College, Santiago City, Isabela

College of Computing Studies, Department of Information and Communication Technology, Isabela State University,
Echague Campus, Philippines

ABSTRACT: Multiple Sclerosis (MS) is a chronic autoimmune disease that affects the central nervous system, characterized by neurodegeneration and inflammation leading to demyelinating plaques. This disease primarily affects young adults and is more prevalent in women. Accurate and early diagnosis of MS is critical for effective management and treatment, yet it remains complex due to the absence of a specific diagnostic test. Traditional diagnostic methods rely on clinical evaluation, imaging, and laboratory results based on the 2017 McDonald criteria.

Recent advancements in machine learning (ML) and deep learning (DL) have shown promise in automating and enhancing the accuracy of MS diagnosis. This paper reviews various ML and DL approaches applied to MS diagnosis, highlighting studies that utilized classifiers such as logistic regression, support vector machines (SVM), and artificial neural networks (ANN) to analyze clinical data, balance disorders, exhaled breath, blood biomarkers, and cognitive task-related graph attributes. These approaches have achieved varying degrees of accuracy, with some models reaching up to 100% accuracy in specific scenarios. In our study, we utilized a dataset from Kaggle to develop an MS classification model using MATLAB2023b. The model employed a coarse tree classifier with hyperparameter tuning, achieving an impressive accuracy of 94.5%. The model's performance was validated through a five-fold cross-validation process, and its effectiveness was demonstrated with a high prediction speed and minimal training time. Feature importance analysis using LIME highlighted that BAEP, gender, VEP, and initial EDSS were significant predictors for MS. This study underscores the potential of ML and DL in revolutionizing MS diagnosis, offering precise and efficient diagnostic tools that can significantly improve early detection and treatment outcomes for MS patients.

KEYWORDS: Machine Learning, Fine Tree, CDMS, Confusion Matrix, MATLAB

I. INTRODUCTION

An autoimmune chronic demyelinating illness that affects the central nervous system (CNS) is called multiple sclerosis (MS). Neurodegeneration and inflammation are its primary characteristics. MS plaques or lesions are the pathological manifestation of the illness. These are localized regions of demyelination that primarily impact the central nervous system's white matter. There are four forms of MS: progressive-relapsing MS (PRMS), secondary-progressive MS (SPMS), PPMS (primary progressive), and RRMS (relapsing-remitting MS) [1]. An estimated 2.8 million people worldwide are thought to have MS, with a prevalence incidence of 35.9 per 100,000. Every five minutes, a new case of MS is recorded worldwide. MS mostly affects young individuals and is more prevalent in women. Patients with MS have a wide range of symptoms. There is no known cause for multiple sclerosis (MS); however, research indicates that environmental factors may contribute to the disease's onset in genetically vulnerable individuals. Symptoms of MS include weak limbs, blurred vision, dizziness, tiredness, and tingling feelings [2,3,4]. Since disease-modifying medications can manage symptoms and stop the illness from progressing, a precise and accurate diagnosis of multiple sclerosis is essential for enabling early therapies for the condition. The existence of distinct CNS lesions in both space

and time, together with the rule out of all other illnesses that may resemble MS clinically and radiologically, are the foundations for the diagnosis of multiple sclerosis. There isn't a specific lab test available to diagnose the illness. For this reason, clinical evaluation, imaging, and laboratory results are all combined in the 2017 McDonald diagnostic criteria for MS [5,6,7,8].

Right now, the best method for diagnosing multiple sclerosis (MS), tracking the progression of the condition, and evaluating the efficacy of various therapies in trials is magnetic resonance imaging (MRI). MRI-based MS diagnosis is laborious, time-consuming, and prone to human error. Thus, machine learning (ML) and deep learning (DL) approaches are being applied in artificial intelligence (AI) to automate MS diagnosis. Machine learning (ML) is a branch of artificial intelligence (AI) in which computers are allowed to learn without explicit programming. Deep learning (DL) is a subset of ML that consists of methods that will enable the software to train itself to execute tasks by subjecting multilayered neural networks to enormous volumes of data [9,10,11].

II. RELATED WORK

A machine learning algorithm was developed by Fiorini et al. to evaluate clinical data and identify the MS disease progression. Making a distinction between benign and advancing structures was the goal. Ordinary least squares (OLS) regularized least squares (RLS), logistic regression (LR), K-nearest neighbors (KNN), and linear SVM were the classifiers that were employed. First, 457 patients provided 91 attributes for collection. The median was then used to impute missing data. The dataset characteristics were then normalized using a min-max scaling technique to fit them within the [0:1] interval. 78.32% was the highest accuracy attained when the OLS algorithm with L1L2 feature selection was used. Additionally, the RLS method with L1L2 feature selection produced the greatest F1-score of 70.2% [12]. Another research sought to create a decision support system (DSS) that uses a straightforward, noninvasive technique to identify MS patients who depend on balance disorders. In the study, 20 healthy controls and 14 MS patients were included. Every participant had a marking placed between their eyebrows on their forehead. After then, subjects were videotaped for three minutes while facing a black background. We used an image processing method to study and assess the relocation of these markers. Using a "tan-sigmoid" transfer function, an ANN was employed. Finding the characteristics that were demonstrated to be substantially different between the MS patients and the healthy controls was necessary for feature extraction. The accuracy of the ANN was 92.35% [13]. The usefulness of an electronic nose (eNose) in identifying MS via exhaled breath analysis was investigated by Ettema et al. [14].

129 healthy controls and 124 MS patients with a verified diagnosis of the disease underwent this procedure, breathing into the Aeonose™ for five minutes each. The Aeonose™ diagnostic test equipment can identify volatile organic substances in exhaled breath. An ANN was trained using exhaled breath data, and Aeonose™ was tested to see if it could discriminate between healthy control participants and patients with MS. Using a subset of MS patients without prescriptions for MS drugs, a second prediction model was developed. Based on the ANN model, there was a 75% sensitivity and 60% specificity difference between MS patients and healthy controls. The model developed using MS patients off medication and healthy controls had accuracy, sensitivity, and specificity of 80%, 93%, and 74%, respectively [15].

Another research suggested utilizing supervised machine learning techniques like RF to create a sophisticated classifier for blood lipid biomarkers. 403 patients were used to train the Bayesian statistics-based biomarker, which classified them as either healthy or having MS illness. After collection, preprocessing was done on their clinical dataset. Additionally, the most essential characteristics were extracted using RF. With all of the feature sets used for training, the RF classifier achieved 100% accuracy, sensitivity, and specificity. Nevertheless, there was a discrepancy in the ages of the healthy participants and MS patients, and there was an imbalance in the data among the classes. In a previous article, researchers suggested an ML technique that used the most distinctive graph attributes identified by a statistical test and a linear SVM classifier during the execution of a cognitive activity to categorize MS patients and healthy people. Twelve healthy people and eight patients in the early stages of multiple sclerosis participated. The node degree, subgraph centrality, K-Core-ness, and PageRank centralities evaluated in the left fusiform, hippocampus, and parahippocampal gyri regions obtained an accuracy of 85% by the integration of all local metrics. The identification of the MS patient was improved by two ideal global measures, modularity and small-worldness index, and individual betweenness centrality, yielding a sensitivity of 81.25% [16]. A recent study developed a model that looks at transcriptomic microRNA data and next-generation sequencing (NGS) data to identify MS biomarkers; it also combines text mining techniques with machine learning (ML) methods for early MS identification. The National

Centre for Biotechnology Information (NCBI) in the United States provided the dataset that was utilised. It is made up of microRNA data from next-generation sequencing (NGS) for 54 RRMS patients. Three models of classification—RF, SVM, and LR—were used. with sensitivity, specificity, F1-score, and average accuracy of 96.4, 96.47, 95.6, and 97%, respectively, the RF algorithm fared better than the other methods.

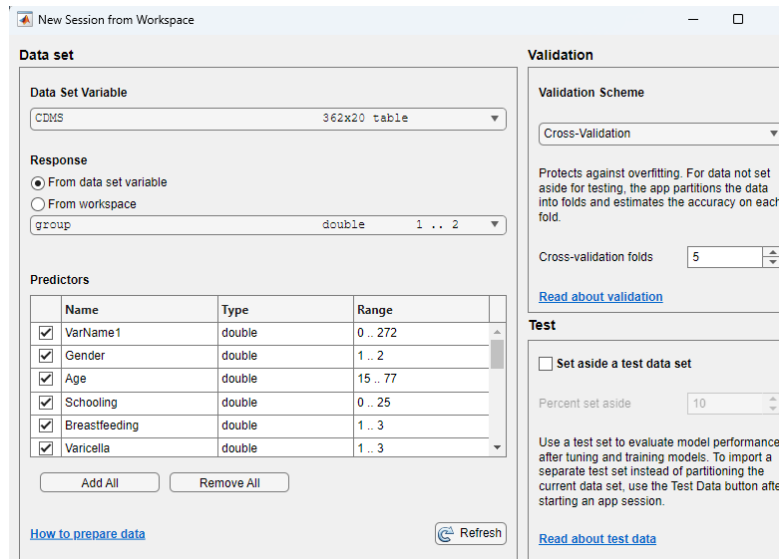
III. METHODOLOGY

The Dataset was taken from Kaggle total of 320 samples consist of 19 features or predictors, some of notable predictors are the following: breastfeeding, varicella, Initial Symptoms 1=visual, 2=sensory, 3=motor, 4=other, 5=visual and sensory, 6=visual and motor, 7=visual and others, 8=sensory and motor, 9=sensory and other, 10=motor and other, 11=Visual, sensory and motor, 12=visual, sensory and other, 13=Visual, motor and other, 14=Sensory, motor and other, 15=visual,sensory,motor and other, Mono or Polysymptomatic, Olygloconal bands, LLSSEP, ULSSEP, VEP, BAEP, Periventricular_MRI, Cortical_MRI, Infratentorial_MRI, Spinal_Cord_MRI, initial_EDSS, and final_EDSS. After the data gathering, the data was imported into MATLAB2023b. Figure (a) shows the data importation from CSV file to matlab workspace.

Breastfeedi...	Varicella	Initial_Sym...	Mono_or_P...	Oligoclonal...	LLSSEP	ULSSEP	VEP	BAEP	Periventric...	Cortical_MRI	Infratentoi...	Spinal_Cor...	Initial_EDSS	Final_EDSS	group
Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
Breastfeedi...	Varicella	Initial_Sym...	Mono_or_P...	Oligoclonal...	LLSSEP	ULSSEP	VEP	BAEP	Periventric...	Cortical_MRI	Infratentoi...	Spinal_Cor...	Initial_EDSS	Final_EDSS	group
1	1	2	1	0	1	1	0	0	0	1	0	1	1	1	1
3	2	10	2	1	1	0	1	0	0	0	0	1	2	2	1
3	1	3	1	1	0	0	0	0	0	1	0	0	1	1	1
1	1	7	2	1	0	1	1	0	1	1	0	0	1	1	1
2	1	6	2	0	1	0	0	0	1	0	0	0	1	1	1
1	1	6	2	0	1	0	0	0	1	0	1	0	1	1	1
1	1	14	2	0	1	0	1	0	1	1	0	1	1	1	1
1	1	14	2	0	1	1	0	0	1	1	1	1	2	2	1
1	1	8	2	0	1	1	1	0	1	0	0	0	1	1	1
1	1	8	2	0	0	0	0	0	1	0	1	0	1	1	1
3	2	15	2	0	1	0	0	0	1	0	0	1	1	1	1
1	1	5	2	0	1	0	1	0	0	1	0	1	1	1	1
1	1	11	2	0	1	1	1	0	1	1	1	0	2	2	1
1	1	13	2	1	0	0	0	0	1	0	1	0	1	1	1
1	1	1	1	0	1	1	1	0	1	1	1	1	2	2	1
1	2	5	2	0	1	1	1	0	0	1	0	0	1	1	1
1	1	15	2	0	1	1	1	0	1	0	1	1	3	3	1
1	1	8	2	1	0	0	1	0	0	1	1	0	1	1	1
1	2	10	2	0	1	1	1	0	1	0	0	1	2	3	1
1	1	8	2	1	1	0	0	0	1	1	1	0	1	1	1
3	1	11	2	1	0	0	0	0	1	0	0	0	1	1	1
3	2	15	2	1	0	0	0	0	1	0	0	1	2	2	1
1	2	13	2	1	1	1	1	0	1	1	1	1	2	2	1
1	2	11	2	1	0	1	1	0	1	1	0	0	1	1	1
3	2	5	2	0	1	0	1	0	1	1	1	1	1	1	1
2	2	13	2	1	1	1	0	0	1	0	1	0	3	3	1
2	1	8	2	1	1	1	1	0	1	1	1	1	2	2	1
2	1	1	1	0	0	0	0	0	1	0	0	0	1	1	1
2	2	11	2	1	1	1	1	0	1	0	1	1	3	3	1
2	1	8	2	1	1	1	0	0	1	1	0	1	1	1	1
2	2	8	2	0	1	1	1	0	1	1	0	1	1	1	1
2	2	5	2	1	1	1	0	1	1	0	0	1	1	1	1
1	2	11	2	0	1	0	1	0	1	1	0	0	2	2	1
2	1	3	1	0	0	1	0	1	0	0	0	1	1	1	1
2	2	13	2	0	1	1	1	0	1	0	1	0	1	2	1
1	1	11	2	0	1	0	0	0	1	0	1	0	1	1	1
3	2	11	2	0	0	0	0	0	1	0	0	0	1	1	1

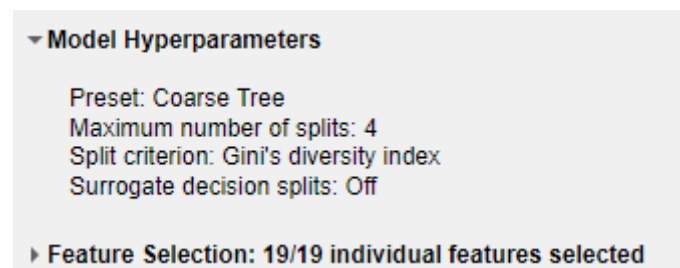
(a). Data Importation

After importing the datasets, MATLAB's classification learner was used to process the text data for machine learning. Procedure. Additionally, a five-fold cross-fold validation was used during pre-processing to improve accuracy validation, as seen in Figure (b).



(b). Data splitting

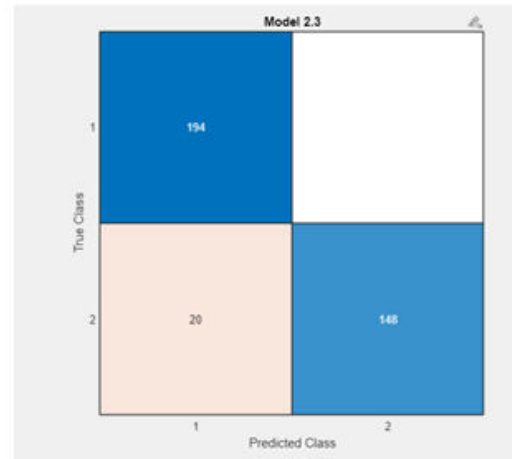
In figure (c), it shows the model's hyperparameters. Its preset was set into coarse tree, maximum number of split is set into 4, and surrogate decision splits was off, overall 19 features or predictors was selected in classifying Multiple Sclerosis.



(c). Coarse Tree Hyperparameter

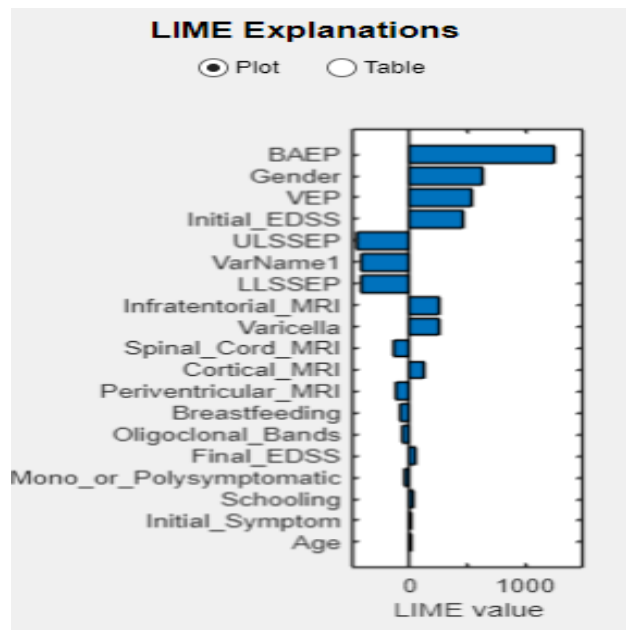
IV. EXPERIMENTAL RESULTS

After pre-processing the data, it is trained using MATLAB2023b. figure (d) shows the training results. 94.5% accuracy was achieved by this model with 14000 observations per second prediction speed, with 3.15 seconds of training time, which is promising and surprising because of its precise accuracy. In figure (e) shows the confusion matrix of the model.



(d). Accuracy and Training Results

(e). Confusion Matrix



(f). LIME Explanations

The figure e, which is the confusion matrix, shows that there are 154 true positives, no false negatives, 20 false positives, and 148 true negatives. This means that the model has no error in detecting CDMS diseases. On the other hand there are 20 mistakes in predicting NON-CDMS. In Figure f shows the lime explanation and the feature importance. It shows that ULSSEP and LLSSEP has the highest negative value means it has nothing to do with predicting CDMS disease, on the otherhand BAEP, Gender, VEP, Initial EDSS plays a big role in predicting CDMS disease.

V. CONCLUSION

Multiple Sclerosis (MS) is a chronic autoimmune disease that affects the central nervous system, characterized by neurodegeneration and inflammation leading to the formation of demyelinating plaques. This disease significantly impacts young adults and is more prevalent in women. Diagnosing MS is complex and involves clinical evaluation, imaging, and laboratory results as outlined in the 2017 McDonald criteria. Recent advances in machine learning (ML) and deep learning (DL) have been applied to automate the diagnosis of MS, aiming to improve accuracy and reduce human error. Several studies have demonstrated the efficacy of ML and DL models in diagnosing and predicting the progression of MS. This study used a dataset from Kaggle to develop an MS classification model using MATLAB2023b. The model achieved an impressive 94.5% accuracy with a prediction speed of 14,000 observations per second and a training time of 3.15 seconds. The confusion matrix showed high accuracy in detecting MS with minimal errors in non-MS predictions. LIME explanations indicated that BAEP, gender, VEP, and initial EDSS were significant predictors of MS.

REFERENCES

1. F. D. Lublin et al., "Defining the clinical course of multiple sclerosis," *Neurology*, vol. 83, no. 3, pp. 278–286, Jul. 2014, doi: 10.1212/wnl.0000000000000560.
2. C. Walton et al., "Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition," *Multiple Sclerosis*, vol. 26, no. 14, pp. 1816–1821, Nov. 2020, doi: 10.1177/1352458520970841.
3. R. Dobson and G. Giovannoni, "Multiple sclerosis – a review," *European Journal of Neurology*, vol. 26, no. 1, pp. 27–40, Nov. 2018, doi: 10.1111/ene.13819.
4. V. Pantazou, M. Schlupe, and R. Du Pasquier, "Environmental factors in multiple sclerosis," *La Presse Médicale*, vol. 44, no. 4, pp. e113–e120, Apr. 2015, doi: 10.1016/j.lpm.2015.01.001.
5. T. J. Murray, "Diagnosis and treatment of multiple sclerosis," *BMJ. British Medical Journal*, vol. 332, no. 7540, pp. 525–527, Mar. 2006, doi: 10.1136/bmj.332.7540.525.
6. D. Miller et al., "Differential diagnosis of suspected multiple sclerosis: a consensus approach," *Multiple Sclerosis*, vol. 14, no. 9, pp. 1157–1174, Sep. 2008, doi: 10.1177/1352458508096878.
7. T. Fangerau et al., "Diagnosis of multiple sclerosis: comparison of the Poser criteria and the new McDonald criteria," *Acta Neurologica Scandinavica*, vol. 109, no. 6, pp. 385–389, Jun. 2004, doi: 10.1111/j.1600-0404.2004.00246.x.
8. A. J. Thompson et al., "Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria," *Lancet Neurology*, vol. 17, no. 2, pp. 162–173, Feb. 2018, doi: 10.1016/s1474-4422(17)30470-2.
9. X. Lladó et al., "Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches," *Information Sciences*, vol. 186, no. 1, pp. 164–185, Mar. 2012, doi: 10.1016/j.ins.2011.10.011.
10. A. Shoeibi et al., "Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review," *Computers in Biology and Medicine*, vol. 136, p. 104697, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104697.
11. R. Bonacchi, M. Filippi, and M. A. Rocca, "Role of artificial intelligence in MS clinical practice," *NeuroImage. Clinical*, vol. 35, p. 103065, Jan. 2022, doi: 10.1016/j.nicl.2022.103065.
12. "A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes," *IEEE Conference Publication | IEEE Xplore*, Aug. 01, 2015. <https://ieeexplore.ieee.org/document/7319381/>
13. Y. Sarbaz, H. Pourakbari, M. H. Vojudi, and A. Ghanbari, "INTRODUCING a DECISION SUPPORT SYSTEM FOR MULTIPLE SCLEROSIS BASED ON POSTURAL TREMOR: a HOPE FOR SEPARATION OF PEOPLE WHO MIGHT BE AFFECTED BY MULTIPLE SCLEROSIS IN THE FUTURE," *Yixué Gōngchéng. Applications, Basis, Communications/Biomedical Engineering: Applications, Basis and Communications*, vol. 29, no. 06, p. 1750046, Dec. 2017, doi: 10.4015/s1016237217500466.
14. A. R. Ettema, M. W. P. M. Lenders, J. Vliegen, A. Slettenaar, M. C. Tjepkema-Cloostermans, and C. C. De Vos, "Detecting multiple sclerosis via breath analysis using an eNose, a pilot study," *Journal of Breath Research*, vol. 15, no. 2, p. 027101, Jan. 2021, doi: 10.1088/1752-7163/abd080.
15. J. Löscher et al., "Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy," *Scientific Reports*, vol. 8, no. 1, Oct. 2018, doi: 10.1038/s41598-018-33077-8.
16. "Diagnosis of multiple sclerosis using Graph-Theoretic measures of Cognitive-Task-Based functional connectivity networks," *IEEE Journals & Magazine | IEEE Xplore*, Sep. 01, 2022. <https://ieeexplore.ieee.org/document/9434426/>



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details