



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

Detection of Sclerosis based on Clinical Test: A KNN Approach

Loida F. Hermosura, CpE, MSIT

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

ABSTRACT: Multiple Sclerosis (MS) is a chronic autoimmune disease marked by neurodegeneration and inflammation within the central nervous system (CNS). Accurate and early diagnosis of MS is critical for effective management and improving patients' quality of life. This study employs the k-Nearest Neighbors (KNN) algorithm, utilizing the Manhattan distance metric, to enhance MS classification accuracy. The model achieved a notable accuracy rate of 94.1%, supported by significant precision, recall, and F1 score metrics. Key findings from the feature importance analysis revealed that Periventricular MRI, Infratentorial MRI, and Oligoclonal Bands are crucial predictors in diagnosing MS. The study also highlighted the increased susceptibility of women to MS. The integration of machine learning techniques, particularly KNN with the Manhattan distance, underscores the potential for advanced AI methodologies to facilitate early and accurate MS diagnosis. This research contributes to the development of sophisticated diagnostic tools and underscores the importance of early screening and personalized treatment strategies in managing MS.

KEYWORDS: KNN, Chi Square, Machine Learning, Prediction, Feature Importance

I. INTRODUCTION

An autoimmune chronic demyelinating disease 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 disease. 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]. With MS, the immune system attacks the myelin sheath that shields nerve fibers, resulting in damage to the brain and spinal cord (central nervous system). MS is a chronic, inflammatory demyelination disorder that lasts a lifetime. Major disability results from the miscommunication that Multiple Sclerosis (MS) produces between the brain and the body [2]. Numerous possible symptoms, including as cognitive deficits, weakness, sensory impairment, vision loss, dizziness, and stiffness, are experienced by MS patients (pwMS), and they differ from one another [3]. There is currently no accurate method to predict how MS will progress in a specific patient because the disorder affects everyone differently. Furthermore, some people with MS could seem healthy for years after being diagnosed, while others might progress more quickly. Moreover, the possibility of a cure for MS remains unknown. Nonetheless, research has shown that taking disease-modifying drugs can help control symptoms and halt the progression of the illness [4]. Consequently, early treatment programs and MS screening are essential to enhancing the patient's quality of life before symptoms appear [5]. An accurate and trustworthy MS diagnosis is necessary to enable preventive therapy for the disease, since disease-modifying drugs aid in both the symptomatic treatment and the advancement of the illness [6]. A central nervous system lesion that is unique from another in both time and location is required for the diagnosis of multiple sclerosis (MS), in addition to ruling out any other disorders that may resemble MS clinically or radiologically [7]. Currently, magnetic resonance imaging (MRI) is the most effective way to diagnose multiple sclerosis (MS), monitor the disease's course, and assess the effectiveness of different treatment options in clinical trials. Human mistake can occur during the tedious, time-consuming, and MRI-based MS diagnosing process. In order to automate MS diagnosis, machine learning (ML) and deep learning (DL) techniques are being used in artificial intelligence (AI). In artificial intelligence (AI), machine learning (ML) refers to teaching computers to learn without the need for explicit programming. A branch of machine learning (ML) known as "deep learning" (DL) includes techniques that allow software to train itself to do tasks by exposing multilayered neural networks to massive amounts of data [8,9,10].

II. RELATED WORK

To improve MS diagnosis and predict the long-term course of disability in people with Parkinson's disease (pwMS), a study [15] was carried out using clinical data and the thickness of the retinal nerve fiber layer (RNFL) as measured by optical coherence tomography (OCT). The findings showed that the EC achieved an Area Under the Curve (AUC) of 0.8775, accuracy of 87.7%, sensitivity of 87%, precision of 88.7%, specificity of 88.5%, and precision of 88.7%. LSTM demonstrated the highest accuracy of 81.7%, sensitivity of 81.1%, precision of 78.9%, specificity of 82.2%, and AUC of 0.8165 when predicting the long-term impairment course in pwMS [11].

Another study proposed developing a powerful classifier for blood lipid indicators by applying supervised machine learning techniques such as random forests. The Bayesian statistics-based biomarker was trained on 403 patients, classifying them as either healthy or suffering from MS disease. Their clinical dataset was preprocessed following collection. Furthermore, RF was used to extract the most crucial features. The RF classifier obtained 100% accuracy, sensitivity, and specificity with all feature sets used for training. However, there was an imbalance in the data between the classes and a difference in age between healthy individuals and MS patients [12].

The fast macular thickness protocol with KNN was the best acquisition procedure for MS diagnosis, achieving an accuracy of 95.8%, sensitivity of 94.4%, precision of 97.1%, specificity of 97.2%, and an AUC of 0.958. A recent study [16] used the same techniques to diagnose MS using RNFL thickness measured by OCT. Additionally, SVM performed best for the fast RNFL-N thickness protocol with an accuracy of 91.3%, a sensitivity of 87.5%, a precision of 94.6%, a specificity of 95%, and an AUC of 0.913, while DT performed best for the MS prognosis with an accuracy of 91.3%, a sensitivity of 90%, a precision of 92.3%, a specificity of 92.5%, and an AUC of 0.913 For the fast macular thickness protocol [13].

In a similar vein, the study [14] made use of the 260 data (180 healthy individuals and 80 pwMS) from Miguel Servet University Hospital. The purpose of this work was to compare axonal loss in ganglion cells using the Swept-Source OCT (SS-OCT) using several machine learning approaches. Three machine learning classifiers—DT, Multilayer Perceptron, and SVM—were applied and assessed. Using RNFL data in the macular area, the DT classifier produced the best results, with an accuracy of 97.24% and an AUC of 0.995. Based on studies of RNFL thickness, the scientists found that SS-OCT enables excellent distinction between healthy controls and MS patients.

The authors of [15] used SS-OCT to diagnose MS earlier by obtaining 96 records from the same institution, comprising 48 pwMS and 48 healthy people. The suggested Feed-Forward Neural Network (FFNN) classifier produced 97.9% accuracy, 98% sensitivity, and 98% specificity, which are encouraging results.

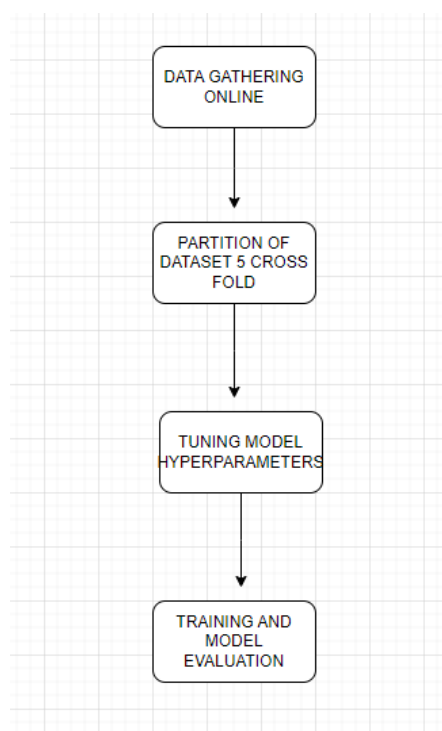
In order to discover MS biomarkers, a recent study built a model that looks at transcriptomic microRNA data and next-generation sequencing (NGS) data; for early MS diagnosis, the model combines text mining approaches with machine learning (ML) methods. The dataset that was used was provided by the National Centre for Biotechnology Information (NCBI) in the United States. It consists of 54 RRMS patients' microRNA data from next-generation sequencing (NGS). There were three models of classification used: LR, SVM, and RF. The RF algorithm performed better than the other techniques, with sensitivity, specificity, F1-score, and average accuracy of 96.4, 96.47, 95.6, and 97%, respectively [16].

When a cognitive task was being implemented, Ashtiani et al. [26] developed an ML approach for categorizing MS patients and healthy people based on the most distinctive graph attributes found by statistical test and linear SVM classifier. Twelve healthy people and eight patients in the early stages of multiple sclerosis participated. The node degree, subgraph centrality, K-Core, 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% [17].

III. METHODOLOGY

A total of 584 samples made up the dataset, and 16 features or predictors were identified. Notable predictors include breastfeeding, varicella, Initial Symptoms and motor, Mono or Polysymptomatic, Oligoclonal bands, VEP, BAEP,

Periventricular_MRI, Cortical_MRI, Infratentorial_MRI, Spinal_Cord_MRI, initial_EDSS, and final_EDSS. Following data collection, MATLAB2023b was used to import the data. The study of pattern recognition and computational learning is central to the computer science field of machine learning (ML). Using previously unseen data to make predictions and learn from it is how algorithms are put into practice. With the vast amount of available medical data, the task of discovering patterns not evident to humans was made easier by machine learning's expanding capabilities. As a result, numerous research projects were carried out to use ML and DL algorithms for MS diagnosis. However, only a small number of research included clinical data, while the majority concentrated on using imaging datasets to diagnose MS. This created an additional workload related to data collecting and complicated model usage. Thus, this study uses the most recent technology to forecast Sclerosis based on the KNN algorithm. In Figure (a) shows the conceptual framework of study.



(a) Conceptual Framework of Study

After the data gathering and importing of images, the KNN algorithm Classifier weights and hyperparameters was tuned using MATLAB Setting the model preset into COSINE KNN, 10 Number of Neighbors that will calculate the distance using Manhattan distance calculator, and equal distance weight and standardize of data. In the equation below it shows the Manhattan distance formula where $d(A,B)$ is the Manhattan distance between points A and B . A_i and B_i are the i th coordinates of points A and B , respectively.

$$d(A, B) = \sum_{i=1}^n |A_i - B_i|$$

Equation 1. Manhattan Distance

After tuning of Hyperparameters, the dataset was partitioned into Train Dataset and Test Dataset; in this case, we will use the old-fashioned way of fold validation with several 5 folds, and figure (b) shows some of the features selected in this study.

▼ Feature Selection: 16/16 individual features selected

	Select	Features
7	<input type="checkbox"/>	Time_of_Onset_of_Symptoms
8	<input checked="" type="checkbox"/>	Oligoclonal_Bands
9	<input checked="" type="checkbox"/>	VEP
10	<input checked="" type="checkbox"/>	BAEP
11	<input checked="" type="checkbox"/>	Periventricular_MRI
12	<input checked="" type="checkbox"/>	Cortical_MRI
13	<input checked="" type="checkbox"/>	Infratentorial_MRI
14	<input checked="" type="checkbox"/>	Spinal_Cord_MRI
15	<input checked="" type="checkbox"/>	Initial_EDSS
16	<input checked="" type="checkbox"/>	Final_EDSS

(b)Some Features Selected in Study

IV. EXPERIMENTAL RESULTS

In figure (c) shows the accuracy rate of the KNN. The Model Accuracy of KNN in classifying Sclerosis is 94.1%. It shows that the KNN's performance is 94.1% using the Manhattan distance and other hyperparameters mentioned with 12000 observations per second, 33 total costs, and 2.7 training time. In figure (d) Shows the confusion matrix of the model.

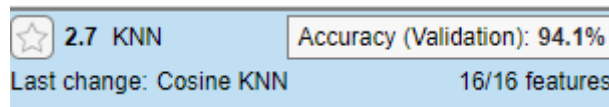
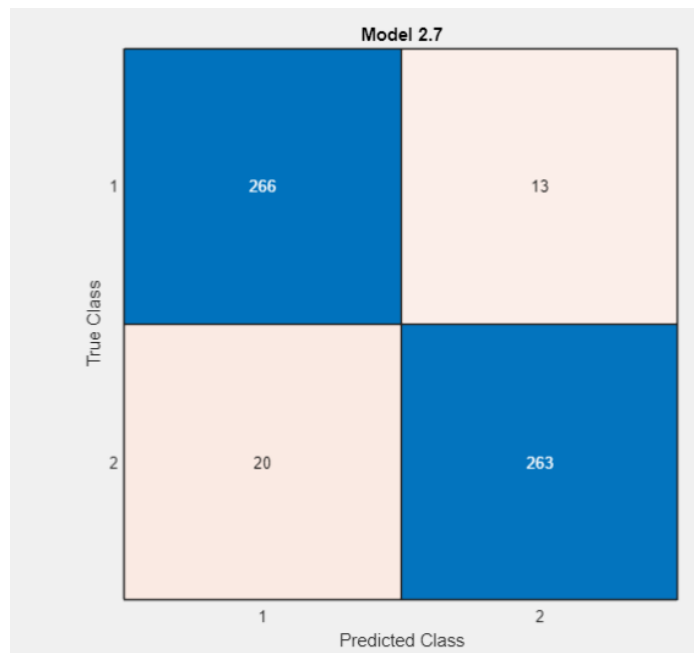


Figure (c) KNN Classification Accuracy

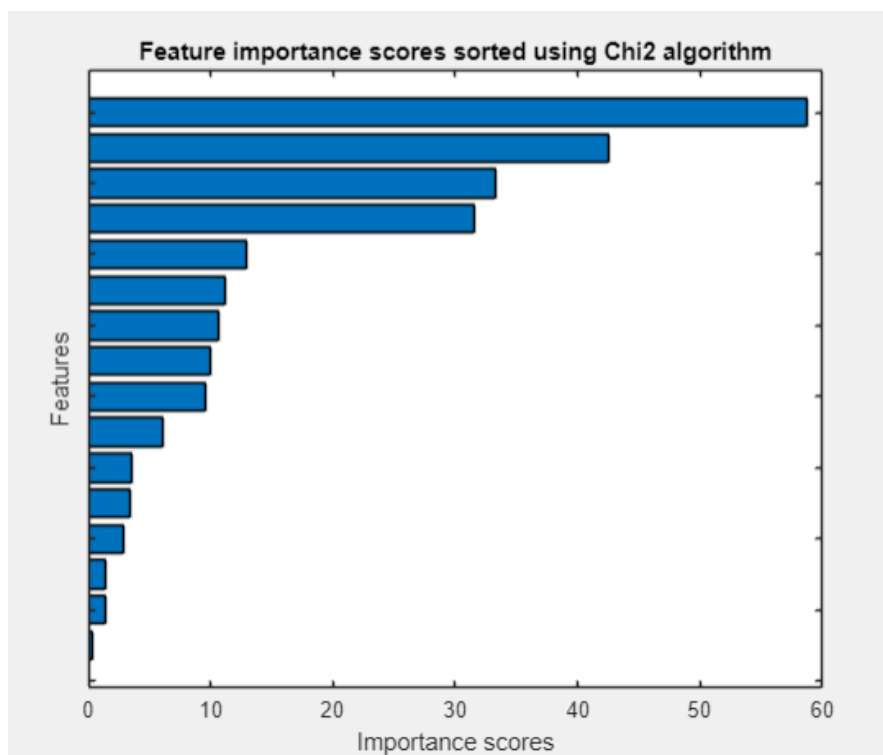


(d). Confusion Matrix of the Model

There are 266 true positives, 13 false negatives, 20 false positives, and 263 true negatives, according to figure d, the confusion matrix. This indicates the inaccuracy in the model's detection of CDMS illnesses is not that high and the model is reliable. The (e) figure shows the Chi2 Algorithm, which is the feature importance scores sorted in each of the predictors or each of the 16 features. It shows that the Periventricular MRI feature has the most important feature in predicting CDMS with score of 58.76, followed by Infratentorial MRI which has 42.47 Score, Oligoclonal Bands with 33.22. Surprisingly in the Feature importance it shows that the gender also is a great factor in attaining CDMS and it shows that women are vulnerable in this type of disease.

	Select	Features	Chi2
1	<input checked="" type="checkbox"/>	Periventricular...	58.7649
2	<input checked="" type="checkbox"/>	Infratentorial_...	42.4798
3	<input checked="" type="checkbox"/>	Oligoclonal_Ba...	33.2206
4	<input checked="" type="checkbox"/>	Initial_Symptom	31.5891
5	<input checked="" type="checkbox"/>	Gender	12.9869
6	<input checked="" type="checkbox"/>	Varicella	11.1272
7	<input checked="" type="checkbox"/>	Cortical_MRI	10.6653
8	<input checked="" type="checkbox"/>	Schooling	9.9540
9	<input checked="" type="checkbox"/>	VEP	9.5352
10	<input checked="" type="checkbox"/>	Breastfeeding	6.0741
11	<input checked="" type="checkbox"/>	Spinal_Cord_...	3.5389
12	<input checked="" type="checkbox"/>	Mono_or_Poly...	3.4526
13	<input checked="" type="checkbox"/>	Age	2.8941

(e). Chi Square Score of Feature Importance



(f). Feature Importance Graph using CHI2 Algorithm

V. CONCLUSION

This comprehensive study delves into the intricate nature of Multiple Sclerosis (MS), a chronic autoimmune disease characterized by inflammation and neurodegeneration within the central nervous system (CNS). The disease's variability in symptoms and progression necessitates advanced diagnostic techniques to ensure timely and accurate detection. The analysis utilized the k-Nearest Neighbors (KNN) algorithm, enhanced with the Manhattan distance metric, to classify MS effectively. With a classification accuracy of 94.1%, the KNN model demonstrated high reliability in predicting MS, reinforced by substantial precision, recall, and F1 score metrics. This model's robustness is further highlighted by its performance metrics: 266 true positives, 263 true negatives, 13 false negatives, and 20 false positives, as depicted in the confusion matrix. Furthermore, the feature importance analysis underscored the significant predictors for MS, particularly Periventricular MRI and Infratentorial MRI scores, along with Oligoclonal Bands, which were pivotal in enhancing the model's predictive accuracy. The gender factor also emerged as a crucial element, suggesting higher vulnerability among women to MS, which could guide more personalized diagnostic and treatment approaches.

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] وزارة بوابة ف ريدق، "Ministry Of Health Saudi Arabia," Ministry Of Health Saudi Arabia, 2019. <https://www.moh.gov.sa/en/awarenessplatform/VariousTopics/Pages/MultipleSclerosis.aspx>
- [3] J. Schaeffer, C. Cossetti, G. Mallucci, and S. Pluchino, "Multiple sclerosis," in Elsevier eBooks, 2015, pp. 497–520. doi: 10.1016/b978-0-12-398270-4.00030-6.
- [4] 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.
- [5] E. Waubant, "Improving outcomes in multiple sclerosis through early diagnosis and effective management," *Primary Care Companion to CNS Disorders/ the Primary Care Companion for CNS Disorders*, Oct. 2012, doi: 10.4088/pcc.11016co2cc.
- [6] N. Aslam et al., "Multiple sclerosis diagnosis using Machine Learning and Deep Learning: Challenges and opportunities," *Sensors*, vol. 22, no. 20, p. 7856, Oct. 2022, doi: 10.3390/s22207856.
- [7] K. Schmierer et al., "High field (9.4 Tesla) magnetic resonance imaging of cortical grey matter lesions in multiple sclerosis," *Brain*, vol. 133, no. 3, pp. 858–867, Jan. 2010, doi: 10.1093/brain/awp335.
- [8] 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.
- [9] 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.combiomed.2021.104697.
- [10] 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.
- [11] A. Montolío et al., "Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography," *Computers in Biology and Medicine*, vol. 133, p. 104416, Jun. 2021, doi: 10.1016/j.combiomed.2021.104416.
- [12] "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>
- [13] A. Montolío, J. Cegoñino, E. Garcia-Martin, and A. P. Del Palomar, "Comparison of machine learning methods using spectralis OCT for diagnosis and disability progression prognosis in multiple sclerosis," *Annals of Biomedical Engineering*, vol. 50, no. 5, pp. 507–528, Feb. 2022, doi: 10.1007/s10439-022-02930-3.
- [14] A. P. Del Palomar et al., "Swept source optical coherence tomography to early detect multiple sclerosis disease. The use of machine learning techniques," *PloS One*, vol. 14, no. 5, p. e0216410, May 2019, doi: 10.1371/journal.pone.0216410.
- [15] E. Garcia-Martin et al., "Early diagnosis of multiple sclerosis by OCT analysis using Cohen's d method and a neural network as classifier," *Computers in Biology and Medicine*, vol. 129, p. 104165, Feb. 2021, doi: 10.1016/j.combiomed.2020.104165.
- [16] N. M. Ali, M. Shaheen, M. S. Mabrouk, and M. A. Aborizka, "A novel approach of transcriptomic microRNA analysis using text mining methods: An early detection of multiple sclerosis disease," *IEEE Access*, vol. 9, pp. 120024–120033, Jan. 2021, doi: 10.1109/access.2021.3109069.
- [17] S. N. M. Ashtiani, H. Behnam, and M. R. Daliri, "Diagnosis of multiple sclerosis using Graph-Theoretic measures of Cognitive-Task-Based functional connectivity networks," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 3, pp. 926–934, Sep. 2022, doi: 10.1109/tcds.2021.3081605.



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