Research Article
Journal of Intelligent Systems: Theory and Applications 9 (2026) 1708339
DOI: 10.38016/jista.1708339

Unveiling Multiple Sclerosis Predictors: A Proposed Artificial Intelligence Approach Using Associative Classification

İsmail Onay[1], Murat Kirişçi[2,*], Cemil Çolak[3]

[1,2] Department of Biostatistics and Medical Informatics, Istanbul University-Cerrahpaşa, Istanbul, Türkiye

[3] Department of Biostatistics and Medical Informatics, Inonu University, Malatya, Türkiye

onayismail@gmail.com, mkirisci@hotmail.com, cemil.colak@inonu.edu.tr

Abstract

Objective:

Although multiple sclerosis (MS) is the leading non-traumatic cause of disability among young adults, the reasons behind its increasing prevalence remain elusive. This study explores the use of artificial intelligence, specifically associative classification, to identify key predictors for the progression of multiple sclerosis (MS).

Methods:

The study utilized a publicly available dataset to forecast whether individuals have MS based on various personal traits. The relevant dataset originated from a cohort group study conducted on Mexican mestizo individuals recently identified with Clinically Isolated Syndrome (CIS). These individuals had sought care at the National Institute of Neurology and Neurosurgery (NINN) between 2006 and 2010. Associative rule mining was applied to uncover relationships between the variables and the development of Clinical Definite Multiple Sclerosis (CDMS). The effectiveness of the model was assessed using accuracy, balanced accuracy, sensitivity, specificity, positive & negative predictive values, and F1 measure with 95% confidence intervals.

Results:

Oligoclonal bands, breastfeeding history, education level, and specific MRI results were significant predictors of MS classification. Oligoclonal bands were particularly associated with a higher likelihood of CDMS. The proposed model performed with high accuracy (87.8%), sensitivity (89.6%), and specificity (86.3%), highlighting its effectiveness in predicting MS progression.

Conclusions:

Artificial intelligence, through associative classification, provides valuable insights into MS progression by identifying significant predictors. These findings can support early diagnosis and contribute to the development of personalized treatment strategies. Future research should incorporate more diverse datasets to validate these results further.

Keywords: Artificial intelligence, Associative Classification, Diagnostic Predictors, Clinically Isolated Syndrome, Clinical Definite Multiple Sclerosis.

Multipl Skleroz Öngörücülerinin Açılması: İlişkisel Sınıflandırma Kullanarak Önerilen Yapay Zekâ Yaklaşımı

Öz

Amaç:

Multipl skleroz (MS) genç yetişkinler arasında önde gelen travma olmayan sakatlık nedeni olmasına rağmen, artan prevalansının arkasındaki nedenler zordur. Bu çalışma, multipl sklerozun (MS) ilerlemesi için temel öngörücüleri tanımlamak için yapay zekanın, özellikle de ilişkilendirici sınıflandırmanın kullanımını araştırmayı amaçlamaktadır.

Yöntemler:

Çalışma, bireylerin çeşitli kişisel özelliklere dayalı MS olup olmadığını tahmin etmek için halka açık bir veri kümesi kullanılmıştır. İlgili veri kümesi, yakın zamanda klinik izole sendrom (CIS) ile tanımlanan Meksika mestizo bireyleri üzerinde yapılan bir kohort grubu çalışmasından kaynaklanmıştır. Bu bireyler 2006-2010 yılları arasında Ulusal Nöroloji ve Nöroşirürji Enstitüsü'nde (NINN) bakım aradılar. Değişkenler arasındaki ilişkileri ortaya çıkarmak ve klinik kesin multipl sklerozun (CDMS) gelişimini ortaya çıkarmak

* Corresponding Author.
E-mail: mkirisci@hotmail.com

Recieved : 28 May 2025
Revision : 18 May 2026
Accepted : 20 May 2026

için ilişkisel kural madenciliği uygulanmıştır. Modellerin etkinliği doğruluk, dengeli doğruluk, duyarlılık, özgüllük, pozitif ve negatif prediktif değerler ve F1 ölçüsü kullanılarak değerlendirildi.

Bulgular:

Oligoklonal bantlar, emzirme geçmişi, eğitim seviyesi ve spesifik MRI sonuçları MS sınıflandırmasının önemli belirleyicileriydi. Oligoklonal bantlar özellikle CDMS'nin daha yüksek bir olasılığı ile ilişkiliydi. Önerilen model, yüksek doğruluk (%87,8), duyarlılık (%89,6) ve özgüllük (%86,3) ile gerçekleştirerek MS ilerlemesini tahmin etmede etkinliğini vurgulamaktadır.

Sonuçlar:

Yapay zeka, ilişkisel sınıflandırma yoluyla, önemli öngörücüler belirleyerek MS ilerlemesine değerli bilgiler sağlar. Bu bulgular erken tanıyı destekleyebilir ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesine katkıda bulunabilir. Gelecekteki araştırmalar, bu sonuçları daha da doğrulamak için daha çeşitli veri kümeleri içermelidir.

Anahtar kelimeler: Yapay zeka, ilişkisel sınıflandırma, tanılama öngörücüler, klinik olarak izole edilmiş sendrom, klinik kesin multipl skleroz.

1. Introduction

Multiple Sclerosis (MS) is recognized as the most common non-traumatic disabling condition among young adults. Both developed and developing countries are witnessing an increase in MS incidence and prevalence, although the exact reasons for this trend remain unclear. MS is a multifaceted disease where a variety of genes marginally increase the risk, along with several environmental factors, including exposure to vitamin D or ultraviolet B light (UVB), Epstein-Barr virus (EBV) infection, obesity, and smoking. Although historically characterized as an autoimmune condition driven by organ-specific T-cells, this classification has been questioned owing to the success of therapies targeting B-cells, which contradicts the conventional T-cell autoimmune concept. MS is often considered a two-phase disease: it begins with inflammation, causing a relapsing-remitting phase, followed by a period of neurodegeneration that results in a constant progression, known as secondary and primary progressive MS (Dobson & Giovannoni, 2019). MS represents the predominant non-traumatic neurological disorder causing disability in young adults, with a global prevalence ranging from 50-300 per 100,000 population. Recent epidemiological data indicate a marked geographical variation, with higher rates observed in North America (165 per 100,000) and Europe (189 per 100,000) compared to Eastern Mediterranean regions (101.4 per 100,000). The annual incidence rates have shown a consistent upward trend, particularly in developed nations, with a 10-year cumulative increase of approximately 10%. This rising prevalence pattern presents significant implications for healthcare systems and societal resources (Walton *et al.*, 2020).

MS appears as a complicated illness, driven by a dynamic interplay between genetic differences and extrinsic circumstances that contribute to its vulnerability. The intermingling of these genetic and environmental elements constitutes a pivotal juncture in the disease's development. Importantly, numerous data from diverse research demonstrate the relevance of key environmental risk factors connected to MS. These include obesity, vitamin D deficiency, Epstein-Barr virus infection, and smoking. Given the possibility of preventive actions by altering these changeable environmental and lifestyle elements, it is crucial to establish definitive causal links between these factors and the onset of MS (Hone *et al.*, 2022). Understanding these complicated relationships intimately might change preventative methods, perhaps lowering the prevalence of MS and lessening its effect on people and healthcare systems. A collaborative, interdisciplinary approach is essential to identify the causal linkages between these variables and MS, incorporating information from genetics, immunology, epidemiology, and molecular biology. This approach not only contributes to a full knowledge of the disease's roots, but also encourages the invention of particular therapies. This lays the scene for tailored methods to minimize MS risk. At the nexus of scientific innovation and better public health, deciphering the intricate interaction of genetic and environmental variables contributing to MS susceptibility is an important undertaking (Vandebergh *et al.*, 2022).

Predicting the course and diagnosis of MS is critical for early intervention and tailored therapy. Recent improvements in artificial intelligence (AI) have shown promise in boosting the predictive accuracy of MS utilizing several data sources, including magnetic resonance imaging (MRI) examinations, gene expression characteristics, and serum cytokines. AI techniques, notably leveraging MRI data, gene expression patterns, serum cytokine levels, and so on, have shown substantial promise in predicting the diagnosis and course of MS. Combining multiple data sources and utilizing modern machine learning techniques, including ensemble approaches and deep learning, boosts forecast accuracy and gives useful insights for early intervention and individualized treatment options (Goyal *et al.*, 2019; Roca *et al.*, 2020; Ponce de Leon-Sanchez *et al.*, 2023). In the related studies, ensemble and machine learning approaches (support vector machine, decision tree, random forest, neural networks, network representation learning, etc.) applied to serum cytokines or disease-related genes

recover feature importance scores but not explicit decision rules (Goyal et al., 2019; Liu et al., 2020). Explainable-AI pipelines that retrofit post-hoc interpretation onto MRI lesion texture features supply explanations but inherit the underlying model's representational opacity (Nicolaou et al., 2023). Another paper proposes a deep learning model based on an artificial neural network for predicting the diagnosis of MS, and demonstrates the effectiveness of the suggested network (Ponce de Leon-Sanchez et al., 2023). By contrast, the associative-classification (CBA) framework adopted here generates human-readable IF–THEN rules directly from demographic, clinical, neurophysiological, and radiological variables, providing transparency at the inference stage rather than as a post-hoc overlay.

This research aims to address a critical gap in current MS prediction methodologies by leveraging artificial intelligence techniques, specifically associative classification. While recent research papers have documented the growing application of AI in MS diagnosis and prognosis, and significant advances have been made in machine learning approaches for MS clinical care (Amin *et al.*, 2024). The specific potential of associative rule mining with clinical parameters remains understudied. Current AI applications in MS have primarily focused on imaging analysis and treatment response prediction (Arrambide *et al.*, 2024), leaving a notable gap in the systematic analysis of demographic, clinical, and radiological markers through rule-based pattern recognition. Our approach uniquely addresses this gap by developing an interpretable AI model that could enhance predictive accuracy while maintaining clinical interpretability, potentially advancing personalized diagnostic strategies in MS care.

2. Materials and Methods

2.1. Data Collection

This research utilized a publicly available dataset to predict the status of MS by examining a range of individual characteristics. The goal was to enhance early prediction techniques and deepen our understanding of how personal traits interact with the development of MS. The original dataset was derived from a prospective cohort study, focusing on individuals of Mexican mestizo heritage who had been newly diagnosed with Clinically Isolated Syndrome (CIS), an early-stage neurological disorder. These individuals had sought care at the renowned National Institute of Neurology and Neurosurgery (NINN) in Mexico City, Mexico. The study, conducted between 2006 and 2010, collected and analyzed data to gain insights into the experiences of this particular patient group with CIS (Chavarria *et al.*, 2023). Expanding upon this result, the study's use of a robust dataset enables a detailed exploration into the nuanced relationship between various genetic, environmental, and lifestyle factors and the onset of MS in this unique population. By scrutinizing these individual traits, the study not only aims to identify potential early indicators of MS but also contributes to a broader understanding of the disease's etiology, particularly in the context of Mexican mestizo individuals. This approach could potentially lead to more personalized healthcare strategies, tailored to specific demographic groups. The research's time frame and geographical focus also allow exploring the progression and manifestation of CIS and MS within a specific cultural and genetic background, offering valuable insights that might differ from studies conducted in other regions or populations (Chavarria *et al.*, 2023). Inclusion and exclusion criteria were given as follows.

Inclusion Criteria:

- Age ≥ 15 years
- Newly diagnosed CIS according to McDonald's 2010 criteria
- Mexican mestizo ethnicity
- Complete clinical and radiological data availability

Exclusion Criteria:

- Previous neurological disorders
- Incomplete medical records
- Loss to follow-up within study period
- Other demyelinating disorders

Statistical Power and Sample Size Considerations

The sample size was calculated by power analysis (WSSPAS) (Arslan *et al.*, 2018), with $\alpha = 0.05$, power = 0.90, and an effect size of 0.3, yielding a minimum requisite sample of 236 people. The present study included a cohort of 273 individuals.

2.2. Dataset Description

This research employed a comprehensive dataset containing demographic, clinical, and radiological characteristics obtained from individuals with suspected neurological diseases. Each patient record was allocated a unique identity (ID) and included numerous attributes classed as input (group) and output variables for modeling purposes.

Input Variables

Demographic factors were age (in years), schooling (years of formal education), gender (1 = male, 2 = female), and breastfeeding history (1 = yes, 2 = no, 3 = unknown). Clinical characteristics comprised Varicella-zoster virus status (1 = positive, 2 = negative, 3 =

unknown), early symptoms (categorized into 15 unique categories), and monosymptomatic or polysymptomatic presentation (1 = monosymptomatic, 2 = polysymptomatic, 3 = unknown).

Diagnostic measures included the presence of oligoclonal bands (OCBs) in cerebrospinal fluid (0 = negative, 1 = positive, 2 = undetermined), which are crucial in identifying MS and other neurological diseases. Neurophysiological tests included Lower Limb and Upper Limb Somatosensory Evoked Potentials (LLSSEP and ULSSEP, respectively; 0 = negative, 1 = positive), Visual Evoked Potential (VEP; 0 = negative, 1 = positive), and Brainstem Auditory Evoked Potentials (BAEP; 0 = negative, 1 = positive).

Radiological factors comprised MRI results for periventricular, cortical, infratentorial, and spinal cord areas (0 = negative, 1 = positive for each), offering insights into the location and amount of neurological injury.

Outcome Variable

The key outcome variable, Group, defined patients as either Clinical Definite Multiple Sclerosis (CDMS; coded as 1) or non-CDMS (coded as 2) based on the McDonald 2010 criteria. This dataset format enables a full investigation of the correlations between numerous clinical, demographic, and radiological parameters and the development of CDMS, helping the discovery of possible prognostic indicators for disease progression (Kaggle, 2024).

Associative Classification Approach

We employed the Associative Classification (AC) methodology, specifically utilizing the Classification Based on Associations (CBA) algorithm, for rule generation and classification. The CBA algorithm was chosen for its efficiency in handling large datasets and its proven effectiveness in previous studies (Liu *et al.*, 1998). Our AC model was implemented using the open-source machine learning library, scikit-learn, with custom scripts to integrate the CBA algorithm.

Rule Generation

Rule generation involved scanning the dataset to identify frequent item sets using the a priori algorithm. The minimum support threshold was set to 0.20, and the minimum confidence threshold for rule generation was set to 0.50. This step produced a set of candidate classification rules.

Discretization and Model Building

Class-attribute interdependence Maximization (CAIM) was implemented for converting continuous data into discrete bins or intervals, making it suitable for algorithms that require categorical input. CAIM is an algorithm used in data mining, specifically in the task of discretization of continuous attributes while preserving or maximizing the relevance of these attributes to a class variable. Regularized Class Association Rules (RCARs) were applied to the related dataset. RCARs offer a powerful and efficient approach to associative classification, addressing some of the key challenges associated with traditional Class Association Rules (CARs). By incorporating regularization techniques, RCARs enhance model generalizability, interpretability, and performance, making them valuable tools in various data mining and machine learning applications. The final model consisted of rules that were both frequent and highly predictive of the target classification outcome (i.e., CDMS and non-CDMS groups).

Evaluation Metrics

The effectiveness of the models was assessed using several metrics including accuracy, balanced accuracy, sensitivity, specificity, positive & negative predictive values, and F1 measure, to pinpoint the most effective outcome. To quantify the uncertainty of these point estimates, non-parametric bootstrap resampling (B = 1,000 replicates, with stratification on the outcome variable) was used to derive bias-corrected and accelerated (BCa) 95% confidence intervals for every metric. The RCAR framework provides built-in protection against overfitting through three mechanisms: (i) a hard minimum support threshold (0.20) that eliminates spurious rules supported by few observations; (ii) a minimum confidence threshold (0.50) that filters out rules with weak class association; and (iii) a regularization term during rule-set selection that penalizes overly specific (long-antecedent) rules and therefore favors parsimonious, generalizable patterns.

Table 1. Detailed results of the association rules for the related classification task

Antecedent rule	Consequent rule	Support	Confidence	Frequency
[OLIGOCLONAL_BANDS=0, VEP=0, PERIVENTRICULAR_MRI=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.266	0.973	72
[GENDER=2, OLIGOCLONAL_BANDS=0, PERIVENTRICULAR_MRI=0]	[NON-CDMS]	0.28	0.95	76
[OLIGOCLONAL_BANDS=0, PERIVENTRICULAR_MRI=0, SPINAL_CORD_MRI=0]	[NON-CDMS]	0.251	0.944	68
[AGE=[15, 48.5), PERIVENTRICULAR_MRI=0, CORTICAL_MRI=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.244	0.93	66
[VEP=0, BAEP=0, PERIVENTRICULAR_MRI=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.306	0.922	83
[VEP=0, PERIVENTRICULAR_MRI=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.314	0.914	85
[GENDER=2, PERIVENTRICULAR_MRI=0]	[NON-CDMS]	0.314	0.904	85
[OLIGOCLONAL_BANDS=0, PERIVENTRICULAR_MRI=0]	[NON-CDMS]	0.339	0.885	92
[SCHOOLING=[0, 13), INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.218	0.868	59
[INITIAL_SYMPTOM=[1, 9.5), LLSSEP=0, ULSSEP=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.288	0.867	78
[AGE=[15, 48.5), INITIAL_SYMPTOM=[1, 9.5), LLSSEP=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.295	0.86	80
[INITIAL_SYMPTOM=[1, 9.5), OLIGOCLONAL_BANDS=0, ULSSEP=0, SPINAL_CORD_MRI=0]	[NON-CDMS]	0.258	0.814	70
[BAEP=0, INFRATENTORIAL_MRI=1]	[CDMS]	0.214	0.806	58
[SCHOOLING=[13, 25), PERIVENTRICULAR_MRI=1]	[CDMS]	0.273	0.804	74
[GENDER=2, INITIAL_SYMPTOM=[1, 9.5), INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.332	0.804	90
[SCHOOLING=[13, 25), BAEP=0, PERIVENTRICULAR_MRI=1]	[CDMS]	0.255	0.802	69
[OLIGOCLONAL_BANDS=1]	[CDMS]	0.218	0.787	59
[AGE=[15, 48.5), BAEP=0, CORTICAL_MRI=0, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.306	0.783	83
[INITIAL_SYMPTOM=[1, 9.5), MONO_OR_POLYSYMPTOMATIC=2, INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.266	0.766	72
[INITIAL_SYMPTOM=[1, 9.5), OLIGOCLONAL_BANDS=0, SPINAL_CORD_MRI=0]	[NON-CDMS]	0.314	0.766	85
[INITIAL_SYMPTOM=[1, 9.5), INFRATENTORIAL_MRI=0]	[NON-CDMS]	0.435	0.742	118
[BREASTFEEDING=3]	[NON-CDMS]	0.207	0.675	56

CDMS (Clinical Definite Multiple Sclerosis), ULSSEP (Upper-Limb Somatosensory Evoked Potentials), LLSSEP (Lower-Limb Somatosensory Evoked Potentials), VEP (Visual Evoked Potentials), BAEP (Brainstem Auditory Evoked Potentials); Polysymptomatic refers to more than one symptom; Monosymptomatic refers to a single characteristic symptom.

3. Results

The current research included a collective of 273 people. Of the total number of individuals, 61.54% (n = 168) were female and 38.46% (n = 105) were male. The average age of the participants was 34.06 years, with a standard deviation of 11.11. The participants' ages ranged from 15 to 77 years. The age interquartile range (IQR) ranged from 25 to 40 years. 25% of participants were below 25 years old, 50% were below 33 years old, and 75% were below 40 years old. The average number of years of schooling, as defined by educational attainment, was 15.18 years, with a standard deviation of 4.24 years. The participants' educational attainment varied from 0 to 25 years, with 25% of persons having fewer than 12 years of schooling, 50% with a minimum of 15 years, and 75% with 20 years or more. This study aimed to identify the association between patients' various demographic, clinical, and radiological characteristics and their classification into CDMS and non-CDMS groups. Utilizing association rule mining, we generated several rules that indicate strong relationships between these characteristics and patient group classification. Detailed results of the association rules for the related classification task are presented in Table 1.

OCBs and Group Association

We found a significant association between the presence of OCBs and the classification into the CDMS group. Specifically, patients with positive OCBs were 78.7% likely to be classified as CDMS (Support: 21.8%, Frequency: 59). This result underscores the importance of OCBs as a biomarker of diagnosing and progressing MS.

Breastfeeding Status and Group Association

The analysis revealed that patients whose breastfeeding status was unknown were predominantly classified into the non-CDMS group with a confidence of 67.5% (Support: 20.7%, Frequency: 56). This finding suggests the potential influence of early-life nutritional factors on the risk of developing MS-related conditions.

Educational Background and MRI Findings

Patients with lower educational levels (0-13 years of schooling) and a negative infratentorial MRI were more likely to be classified into the non-CDMS group with high confidence of 86.8% (Support: 21.8%, Frequency: 59). Conversely, patients with higher educational attainment (13-25 years of schooling) and positive periventricular MRI findings were associated with the CDMS group (Confidence: 80.4%, Support: 27.3%, Frequency: 74). These findings highlight the potential interplay between socio-economic factors, educational attainment, and the biological markers of MS.

Gender, MRI Findings, and Group Association

Female patients with negative periventricular MRI findings were significantly associated with the non-CDMS group, with a confidence level of 90.4% (Support: 31.4%, Frequency: 85). This suggests that gender and specific MRI findings could play a crucial role in MS classification and potentially in tailoring patient-specific therapeutic strategies.

Initial Symptoms and Group Association

The initial symptoms of patients, when combined with negative infratentorial MRI results, were indicative of a higher likelihood of classification into the non-CDMS group (Confidence: 74.2%, Support: 43.5%, Frequency: 118). This emphasizes the prognostic value of initial clinical presentations in conjunction with radiological findings in MS.

Age, MRI Findings, and Group Association

Patients within the age range of 15-48.5 years, exhibiting negative periventricular, cortical, and infratentorial MRI findings, were predominantly associated with the non-CDMS group (Confidence: 93%, Support: 24.4%, Frequency: 66). This result highlights the significance of age and comprehensive MRI findings in distinguishing between CDMS and non-CDMS groups.

Table 2 presents the performance metrics of the proposed CBA-based associative classifier, all reported with 95% confidence intervals. The classifier demonstrated strong overall performance, achieving an accuracy of 0.878 (95% CI: 0.835–0.913) and a balanced accuracy of 0.880 (0.836–0.917). Sensitivity was 0.896 (0.840–0.942) and specificity 0.863 (0.802–0.913). The positive predictive value reached 0.848 (0.789–0.899), while the negative predictive value was notably higher at 0.906 (0.857–0.946). The resulting F1-score was 0.872 (0.821–0.913), indicating a well-balanced trade-off between precision and recall.

Table 2. Performance of the proposed CBA-based associative classifier reported with 95% confidence intervals (CI)

Metric	Point estimate	95% CI
Accuracy	0.878	0.835 – 0.913
Balanced accuracy	0.880	0.836 – 0.917
Sensitivity	0.896	0.840 – 0.942
Specificity	0.863	0.802 – 0.913
Positive predictive value	0.848	0.789 – 0.899
Negative predictive value	0.906	0.857 – 0.946
F1-score	0.872	0.821 – 0.913

4. Discussion

The relationship between many factors and patients' CDMS and non-CDMS categories was assessed in this study using the association rule mining technique. The research findings strongly support early indicators in the prognosis of MS, socio-demographic factors, and biomarkers. One of the artificial intelligence techniques used in this study was association rule mining, which successfully linked the classification of the MS type with the significant characteristics of the patients. This approach allowed the identification of independent predictors of demographic, clinical, and radiologic character connected with CDMS and non-CDMS states. As a result, association rule mining may provide field knowledge on new prospective markers and risk factors—i.e., elements not apparent in a first statistical analysis. Association rule mining helps to highlight pertinent prediction markers and risk factors that may not be immediately apparent from conventional statistical studies by spotting these interactions. In this context, association rule mining highlights the need for AI-driven methods in medical research and provides a solid foundation for revealing latent trends in large-scale data. Including association rule mining in clinical evaluation processes might help healthcare professionals be more capable of making data-driven decisions, thereby enhancing patient outcomes. Our study was able to uncover intricate patterns and interactions among various MS predictors using AI techniques, particularly association rule mining. This approach not only validated our research question about the feasibility of AI in MS research but also highlighted potential areas for further investigation into modifiable risk factors. Various AI techniques, including rule-based methods, network representation learning, gradient-boosting trees, deep learning, and so on, show high accuracy and sensitivity in diagnosing and predicting MS (Arani *et al.*, 2018; Liu *et al.*, 2020; Briggs & Sept, 2021; Nicolaou *et al.*, 2023).

The strong association between the presence of OCBs and the classification into the CDMS group reinforces previous research indicating the significance of OCBs in MS diagnosis and prognosis (Smith & McDonald, 1999). OCBs in cerebrospinal fluid are widely recognized as a hallmark of MS, reflecting chronic inflammation within the central nervous system. This biomarker is critical in clinical evaluations, aiding in early and accurate diagnosis, which is crucial for initiating timely treatment to alter potential disease progression (Coles, 2009; Langer-Gould *et al.*, 2013). Clinically, the presence of OCBs can guide neurologists in confirming a diagnosis of MS, especially in cases where MRI findings may be inconclusive. Incorporating OCBs testing into routine clinical evaluation enhances diagnostic accuracy and allows for early intervention strategies, which are crucial for better patient outcomes (Thompson *et al.*, 2018). Moreover, OCBs presence in cerebrospinal fluid is associated with disease activity and progression, making it a valuable marker for monitoring response to treatment and disease course over time (Link & Huang, 2006). Therefore, OCBs testing should be integrated into standard diagnostic protocols for MS, complementing other diagnostic tools such as MRI and clinical evaluations. The presence of OCBs in cerebrospinal fluid has emerged as a significant predictor of MS in our analysis, confirming our hypothesis about its diagnostic value. This finding is consistent with the literature that identifies these bands as markers of disease activity and progression in MS, thereby reinforcing their importance in clinical evaluations (Link & Huang, 2006).

Our analysis also highlighted the potential influence of unknown breastfeeding status on MS classification, which suggests a possible connection between early-life nutritional factors and MS risk. Although the exact mechanisms remain unclear, early nutritional influences have been associated with the development of the immune system and could potentially contribute to the risk of autoimmune disorders later in life (Jacobs *et al.*, 2020). This finding emphasizes the importance of gathering more detailed clinical histories in MS research and patient evaluations to better understand the impact of these early influences. Clinically, obtaining comprehensive early-life histories, including breastfeeding status, can provide valuable context for risk assessment and early detection of MS in individuals presenting with initial neurological symptoms. Understanding the role of early-life factors in MS development could help identify at-risk populations and guide the implementation of preventive strategies.

The observed relationship between educational attainment and MS classification may reflect broader socio-economic determinants of health. Higher education levels are often associated with better access to healthcare resources, including early diagnosis and treatment options for MS. Furthermore, this correlation might indicate the role of cognitive reserve in coping with neurodegenerative diseases, where individuals with higher education exhibit greater resilience to cognitive decline. These insights highlight the importance of considering socio-economic factors in clinical evaluations and treatment planning for MS patients (Stern, 2012; Marrie *et al.*, 2015). Clinicians should consider educational background as part of a comprehensive patient evaluation, recognizing that higher cognitive reserve may influence disease progression and management strategies. This relationship highlights the need for public health initiatives to promote education as a means of potentially mitigating the impact of neurodegenerative diseases.

Demographic variables such as age and gender have shown significant correlations with the classification of

MS in our study. This supports the research hypothesis that demographic factors are crucial in MS prediction, mirroring findings in other studies that suggest the age of onset and gender-specific prevalence patterns in MS (Harbo *et al.*, 2013). The gender-specific associations observed in our study, particularly the higher likelihood of females with certain MRI findings being classified into the non-CDMS group, align with existing literature documenting a higher prevalence and distinct clinical course of MS in females compared to males (Harbo *et al.*, 2013). This finding emphasizes the necessity of gender-specific research and clinical approaches in understanding and managing MS. Clinicians should consider these differences when evaluating MRI results and developing treatment plans, ensuring that gender-specific factors are accounted for in clinical assessments. Gender differences in MS presentation and progression should inform personalized treatment plans, ensuring that female patients receive tailored interventions that address their unique clinical profiles. Additionally, further research is needed to understand the biological mechanisms underlying these gender differences, which could lead to the development of more effective gender-specific therapies.

The significance of initial symptoms combined with MRI findings in classifying patients emphasizes the importance of a comprehensive clinical evaluation of MS prognosis. Early symptom recognition and appropriate imaging can lead to timely intervention, potentially altering the disease course (Brownlee *et al.*, 2017). Clinicians should prioritize thorough initial assessments, incorporating both clinical and radiological data, to ensure accurate diagnosis and effective treatment planning. Detailed documentation of initial symptoms and their progression is essential for developing personalized therapeutic strategies. Early recognition and intervention based on initial symptoms can significantly improve long-term outcomes for MS patients. Specific MRI findings, such as lesions in the periventricular and cortical regions, were notably associated with MS classifications. These results underscore the value of radiological assessments in MS diagnosis and support our hypothesis that advanced imaging techniques can significantly enhance predictive accuracy.

The association between age, specific MRI findings, and non-CDMS classification may provide valuable insights into the underlying pathophysiology of the disease. These findings suggest that patterns of neural damage may vary or become more detectable at different stages of life, emphasizing the importance of age in understanding MS progression (Rocca *et al.*, 2015). By recognizing these age-related variations, clinicians can better tailor diagnostic and therapeutic approaches to align with the disease's progression in individual patients. Age-specific clinical guidelines should be developed to optimize MS management across different life stages. Understanding how age-related changes in the nervous system influence MS pathophysiology can help tailor treatments to better address the needs of patients at various ages. This approach can lead to more effective management of MS, reducing the burden of the disease and improving the quality of life for patients. The integration of detailed clinical, demographic, and radiological data in our AI models has paved the way for personalized medicine approaches in MS management. This venture aligns with our initial inferences that personalized diagnostics and treatment strategies based on AI could significantly improve patient outcomes and disease management. A similar study on this subject aimed to predict MS conditions using an optimized deep-learning (DL) algorithm based on demographic and clinical characteristics of the same public dataset. The relevant study's results demonstrated that the DL model could effectively and consistently predict MS status, offering valuable insights for early diagnosis and clinical decision-making (Ucuzal *et al.*, 2024). The current research's difference and advantage lie in its ability to identify specific predictive factors for MS, such as oligoclonal bands and breastfeeding history, providing more interpretable and clinically relevant insights compared to the black-box nature of the DL approach used in the mentioned study (Ucuzal et al., 2024).

The generated rules may admit a coherent clinical reading aligned with current MS diagnostic practice. The CDMS-predictive rules anchored on positive oligoclonal bands (OCBs = 1 → CDMS; confidence 0.787) and on periventricular or infratentorial lesion involvement could directly recapitulate the cerebrospinal-fluid and dissemination-in-space pillars of the 2017 revised McDonald criteria, lending the rule set face validity from a neurological standpoint. The mirror-image NON-CDMS rules, in which the conjunction of negative OCBs with absent periventricular, infratentorial, or spinal-cord lesions could reach high confidence values, formalize the rule-out reasoning routinely applied during clinically isolated syndrome counselling and provide a quantitative anchor for reassurance when both the canonical fluid biomarker and the canonical imaging substrate are unremarkable. Two rule families warrant cautious interpretation: associations involving schooling intervals most plausibly act as proxies for healthcare access and diagnostic delay rather than as direct causal determinants, while rules dominated by negative evoked-potential patterns are clinically informative only in conjunction with structural and biochemical findings, consistent with the diminished ancillary role of evoked potentials in the post-2017 framework. Overall, the rule set offers a transparent, quantitatively annotated restatement of the diagnostic logic that experienced neurologists already apply — a feature we regard as a strength for clinical acceptability — while any deployment as decision support must be preceded by

external validation and structured adjudication by board-certified MS specialists (Thompson et al., 2018).

Several limitations of the present work warrant explicit acknowledgement. The proposed model was developed and evaluated on a single cohort of Mexican mestizo participants, and no external dataset from an independent center, country, or ethnic background was available for out-of-sample validation. Consequently, the reported performance figures must be interpreted as evidence of internal validity only, and the generalizability of both the rule set and the classification metrics to populations with different genetic backgrounds, environmental exposures, or healthcare-access patterns remains formally untested; the ethnically homogeneous composition of the cohort further constrains extrapolation to other demographic groups. Although the sample size (n = 273) exceeds the minimum required for adequate statistical power and our findings are internally robust, the sample size may nevertheless limit the statistical power to detect rules of low support and to characterize rare antecedent combinations. Taken together, these considerations imply that the present rule set should be regarded as a hypothesis-generating output rather than a definitive diagnostic instrument, and that larger, prospective, ethnically and geographically diverse multicenter investigations are required to confirm, refute, or refine these patterns and to verify the applicability of the proposed predictive framework across heterogeneous patient populations.

5. Conclusions

Our analysis demonstrates the utility of association rule mining in uncovering significant relationships between patient characteristics and the classification into CDMS and non-CDMS groups. The identified associations, particularly those involving OCBs, breastfeeding status, educational background, gender, initial symptoms, age, and MRI findings, offer valuable insights into the risk factors and diagnostic markers of MS. These findings contribute to the growing body of evidence supporting personalized approaches in the diagnosis, monitoring, and management of MS. Incorporating detailed clinical evaluations, including comprehensive histories and biomarker assessments, is essential for improving diagnostic accuracy and patient outcomes. The conclusions of this study should, however, be read in light of two structural caveats: the model has been developed only internally on a single ethnically homogeneous cohort. Accordingly, before any clinical-decision-support use, the rule set must be prospectively re-evaluated on independent, ethnically and geographically diverse multicenter cohorts; until that external validation has been performed, the present results are best framed as a hypothesis-generating output.

References

Amin M, Martínez-Heras E, Ontaneda D & Prados Carrasco F (2024). Artificial Intelligence and Multiple Sclerosis. *Current Neurology and Neuroscience Reports* **24,** 233-243.

Arani LA, Hosseini A, Asadi F, Masoud SA & Nazemi E (2018). Intelligent computer systems for multiple sclerosis diagnosis: a systematic review of reasoning techniques and methods. *Acta Informatica Medica* **26,** 258-264.

Arrambide G, Comabella M & Tur C (2024). Big data and artificial intelligence applied to blood and CSF fluid biomarkers in multiple sclerosis. *Frontiers in Immunology*.

Arslan AK, Yaşar Ş, Çolak C & Yoloğlu S (2018). WSSPAS: An interactive web application for sample size and power analysis with R using shiny. *Türkiye Klinikleri Biyoistatistik* **10,** 224-246.

Briggs FB & Sept C (2021). Mining complex genetic patterns conferring multiple sclerosis risk. *International journal of environmental research and public health* **18,** 2518.

Brownlee WJ, Hardy TA, Fazekas F & Miller DH (2017). Diagnosis of multiple sclerosis: progress and challenges. *The lancet* **389,** 1336-1346.

Chavarria V, Espinosa-Ramírez G, Sotelo J, Flores-Rivera J, Anguiano O, Hernández AC, Guzmán-Ríos ED, Salazar A, Ordoñez G & Pineda B (2023). Conversion Predictors of Clinically Isolated Syndrome to Multiple Sclerosis in Mexican Patients: A Prospective Study. *Archives of Medical Research***,** 102843.

Coles A (2009). Multiple sclerosis: the bare essentials. *Practical neurology* **9,** 118-126.

Dobson R & Giovannoni G (2019). Multiple sclerosis–a review. *European Journal of Neurology* **26,** 27-40.

Goyal M, Khanna D, Rana PS, Khaibullin T, Martynova E, Rizvanov AA, Khaiboullina SF & Baranwal M (2019). Computational intelligence technique for prediction of multiple sclerosis based on serum cytokines. *Frontiers in neurology* **10,** 781.

Harbo HF, Gold R & Tintoré M (2013). Sex and gender issues in multiple sclerosis. *Therapeutic advances in neurological disorders* **6,** 237-248.

Hone L, Jacobs BM, Marshall C, Giovannoni G, Noyce A & Dobson R (2022). Age-specific effects of childhood body mass index on multiple sclerosis risk. *Journal of Neurology* **269,** 5052-5060.

Jacobs BM, Giovannoni G, Cuzick J & Dobson R (2020). Systematic review and meta-analysis of the association between Epstein–Barr virus, multiple sclerosis and other risk factors. *Multiple sclerosis journal* **26,** 1281-1297.

Kaggle (2024). Predictors of Multiple Sclerosis Disease, vol. 2024, pp. Conversion predictors of Clinically Isolated Syndrome to Multiple Sclerosis in Mexican patients, https://www.kaggle.com/. [Accessed on June 5, 2025]

Langer-Gould A, Brara SM, Beaber BE & Koebnick C (2013). Childhood obesity and risk of pediatric multiple sclerosis and clinically isolated syndrome. *Neurology* **80,** 548-552.

Link H & Huang Y-M (2006). Oligoclonal bands in multiple sclerosis cerebrospinal fluid: an update on methodology and clinical usefulness. *Journal of neuroimmunology* **180,** 17-28.

Liu B, Hsu W & Ma Y (1998). Integrating classification and association rule mining. In *Proceedings of the fourth international conference on knowledge discovery and data mining*, pp. 80-86.

Liu H, Guan J, Li H, Bao Z, Wang Q, Luo X & Xue H (2020). Predicting the disease genes of multiple sclerosis based on network representation learning. *Frontiers in Genetics* **11,** 328.

Marrie RA, Elliott L, Marriott J, Cossoy M, Blanchard J, Leung S & Yu N (2015). Effect of comorbidity on mortality in multiple sclerosis. *Neurology* **85,** 240-247.

Nicolaou A, Pantzaris M, Loizou CP, Kakas A & Pattichis CS (2023). An Explainable AI model in the assessment of Multiple Sclerosis using clinical data and Brain MRI lesion texture features. In *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 1-4. IEEE.

Ponce de Leon-Sanchez ER, Dominguez-Ramirez OA, Herrera-Navarro AM, Rodriguez-Resendiz J, Paredes-Orta C & Mendiola-Santibañez JD (2023). A deep learning approach for predicting multiple sclerosis. *Micromachines* **14,** 749.

Roca P, Attye A, Colas L, Tucholka A, Rubini P, Cackowski S, Ding J, Budzik J-F, Renard F & Doyle S (2020). Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. *Diagnostic and Interventional Imaging* **101,** 795-802.

Rocca MA, Amato MP, De Stefano N, Enzinger C, Geurts JJ, Penner I-K, Rovira A, Sumowski JF, Valsasina P & Filippi M (2015). Clinical and imaging assessment of cognitive dysfunction in multiple sclerosis. *The Lancet Neurology* **14,** 302-317.

Smith KJ & McDonald W (1999). The pathophysiology of multiple sclerosis: the mechanisms underlying the production of symptoms and the natural history of the disease. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **354,** 1649-1673.

Stern Y (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology* **11,** 1006-1012.

Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M & Freedman MS (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology* **17,** 162-173.

Ucuzal H, Colak C & Turtay MG (2024). An optimized deep learning approach for predicting multiple sclerosis conditions from an artificial intelligence perspective In *WACEM 2023 World Academic Congress of Emergency Medicine*. ed. Galwangar S, pp. 434-438. WACEM23, Antalya/Türkiye.

Vandebergh M, Degryse N, Dubois B & Goris A (2022). Environmental risk factors in multiple sclerosis: Bridging Mendelian randomization and observational studies. *Journal of Neurology* **269,** 4565-4574.

Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, Robertson N, La Rocca N, Uitdehaag B, van der Mei I, Wallin M, Helme A, Angood Napier C, Rijke N & Baneke P (2020). Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult Scler* **26,** 1816-1821.