Innovations in Multiple Sclerosis Care: The Impact of Artificial Intelligence via Machine Learning on Clinical Research and Decision-Making

Jacob Cartwright, MS; Kristof Kipp, PhD; and Alexander V. Ng, PhD, FACSM

CE INFORMATION

ACTIVITY AVAILABLE ONLINE: To access the article and evaluation online, go to https://www.highmarksce.com/mscare.

TARGET AUDIENCE: The target audience for this activity is physicians, advanced practice clinicians, nursing professionals, mental health professionals, social workers, and other health care providers involved in the management of individuals with multiple sclerosis (MS).

LEARNING OBJECTIVE:
1. Recognize differences between supervised and unsupervised learning to better understand and evaluate their strengths, limitations, and relevance to the diagnosis and care for individuals with MS.
2. Describe how machine learning techniques can assist with MS diagnosis, personalize treatment plans, and optimize rehabilitation strategies for improved patient outcomes in order to be able to apply this technology to patient care.

ACCREDITATION:

In support of improving patient care, this activity has been planned and implemented by the Consortium of Multiple Sclerosis Centers (CMSC) and Intellisphere, LLC. The CMSC is jointly accredited by the Accreditation Council for Continuing Medical Education (ACCME), the Accreditation Council for Pharmacy Education (ACPE), and the American Nurses Credentialing Center (ANCC), to provide continuing education for the healthcare team.

This activity was planned by and for the health care team, and learners will receive .5 Interprofessional Continuing Education (IPCE) credit for learning and change.

PHYSICIANS: The CMSC designates this journal-based activity for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

NURSES: The CMSC designates this enduring material for 1.0 contact hour of nursing continuing professional development (NCPD) (none in the area of pharmacology).

PSYCHOLOGISTS: This activity is awarded 1.0 CE credits.

SOCIAL WORKERS: As a Jointly Accredited Organization, the CMSC is approved to offer social work continuing education by the Association of Social Work Boards (ASWB) Approved Continuing Education (ACE) program. Organizations, not individual courses, are approved under this program. Regulatory boards are the final authority on courses accepted for continuing education credit. Social workers completing this course receive 1.0 general continuing education credits.

DISCLOSURES: It is the policy of the Consortium of Multiple Sclerosis Centers to mitigate all relevant financial disclosures from planners, faculty, and other persons that can affect the content of this CE activity. For this activity, all relevant disclosures have been mitigated.

Francois Bethoux, MD, editor in chief of the *International Journal of MS Care (IJMSC)*, and **Alissa Mary Willis, MD,** associate editor of *IJMSC*, have disclosed no relevant financial relationships. Authors **Jacob Cartwright, BSc; Kristof Kipp, PhD;** and **Alexander V. Ng, PhD,** have disclosed no relevant financial relationships.

The staff at *IJMSC*, CMSC, and Intellisphere, LLC who are in a position to influence content have disclosed no relevant financial relationships. **Laurie Scudder, DNP, NP,** continuing education director at CMSC, has served as a planner and reviewer for this activity. She has disclosed no relevant financial relationships.

METHOD OF PARTICIPATION:
Release Date: September 1, 2023; Valid for Credit through: September 1, 2024

To receive CE credit, participants must:
(1) Review the continuing education information, including learning objectives and author disclosures.
(2) Study the educational content.
(3) Complete the evaluation, which is available at https://www.highmarksce.com/mscare.

Statements of Credit are awarded upon successful completion of the evaluation. There is no fee to participate in this activity.

DISCLOSURE OF UNLABELED USE: This educational activity may contain discussion of published and/or investigational uses of agents that are not approved by the FDA. The CMSC and Intellisphere, LLC do not recommend the use of any agent outside of the labeled indications. The opinions expressed in the educational activity are those of the faculty and do not necessarily represent the views of the CMSC or Intellisphere, LLC.

DISCLAIMER: Participants have an implied responsibility to use the newly acquired information to enhance patient outcomes and their own professional development. The information presented in this activity is not meant to serve as a guideline for patient management. Any medications, diagnostic procedures, or treatments discussed in this publication should not be used by clinicians or other health care professionals without first evaluating their patients' conditions, considering possible contraindications or risks, reviewing any applicable manufacturer's product information, and comparing any therapeutic approach with the recommendations of other authorities.

ABSTRACT

Artificial intelligence (AI) and its specialized subcomponent machine learning are becoming increasingly popular analytic techniques. With this growth, clinicians and health care professionals should soon expect to see an increase in diagnostic, therapeutic, and rehabilitative technologies and processes that use elements of AI. The purpose of this review is twofold. First, we provide foundational knowledge that will help health care professionals understand these modern algorithmic techniques and their implementation for classification and clustering tasks. The phrases *artificial intelligence* and *machine learning* are defined and distinguished, as are the metrics by which they are assessed and delineated. Subsequently, 7 broad categories of algorithms are discussed, and their uses explained. Second, this review highlights several key studies that exemplify advances in diagnosis, treatment, and rehabilitation for individuals with multiple sclerosis using a variety of data sources—from wearable sensors to questionnaires and serology—and elements of AI. This review will help health care professionals and clinicians better understand AI-dependent diagnostic, therapeutic, and rehabilitative techniques, thereby facilitating a greater quality of care.

Artificial intelligence (AI) refers to a broad class of algorithms that, as the name suggests, result in displays of intelligence by machines.[1] Contemporary AI manifests in many forms, from self-learning chatbots to simple linear regressions. If the exact definition of AI seems nebulous and esoteric, that is because the exact definition of AI *is* nebulous and esoteric. As a writer described it, "... it's part of the history of the field of artificial intelligence that every time somebody figured out how to make a computer do something—play good checkers, solve simple but relatively informal problems—there was a chorus of critics to say, but that's not thinking."[2] It is this steadily moving set of goalposts that makes AI so hard to define.

Consequently, there is no consensus concerning the distinction between AI and machine learning (ML); some individuals consider both terms synonymous, and others consider ML to lack sufficient complexity to constitute AI. A third group considers ML to be a subcategory of AI. In the latter case, the range of AI is vast: Expert systems codify the knowledge of experts,[3] artificially intelligent chatbots simulate conversation,[4] and ML finds patterns in data.[1] Although *AI* accurately describes each of these domains, for clarity, subsequent discussion uses the most precise term.

Expert systems can help clinicians diagnose lesions and white matter abnormalities resulting from multiple sclerosis (MS) via magnetic resonance imaging (MRI) by offering plausible diagnoses with confidence levels.[5] In contrast, AI chatbots are versatile and are used for many tasks, including improving language (both code and narrative paragraphs), inspiring paths for analysis, and solving mathematical equations.[4] Finally, ML, the focus of this review, consists of iterative algorithms adept at exploiting patterns in data to group observations, establish class boundaries, or predict continuous variables.

Although not all specific to MS, numerous recent reviews have highlighted the importance of ML across the health care sector.[6-8] This narrative review explores ML algorithms and their use in MS research, including disease detection from serum biomarkers and medical imaging, discovery of clusters of pathology, analysis of movement, inference of cognitive impairment, and prediction of disease progression.
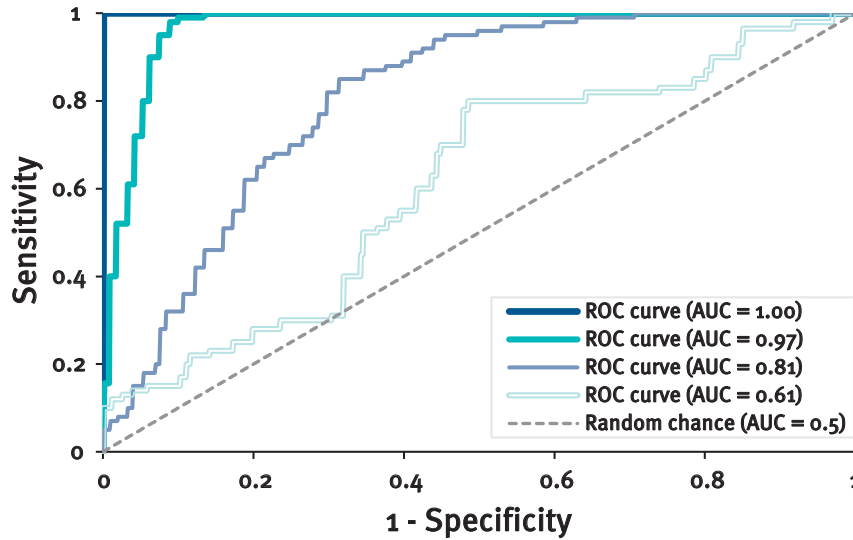
Multiple Sclerosis

A chronic neurodegenerative disease predominantly affecting women, MS is characterized by demyelination of nerves in the brain and spinal cord[9] and often follows a pattern of cyclical periods of remission and relapse, ultimately with progression.[9] Although the exact etiology is unknown,[9,10] several genes "... including those for protein tyrosine phosphatase (CD45), the IL-7 receptor, and CD24"[10] have been linked to susceptibility to developing MS, in addition to environmental factors, including Epstein-Barr virus,[11] increased latitude,[12] and decreased levels of vitamin D.[13] Common symptoms in individuals with MS include symptomatic fatigue, muscle weakness, impaired gait, spasticity, ataxia, dysarthria, and vision loss.[10]

Because MS is such a heterogeneous disease, with differently presenting preclinical stages—including a seemingly nonspecific prodrome[14]—and phenotypes (eg, relapsing-remitting, progressive),[9] advanced techniques are required to investigate certain aspects of its pathophysiology and treatment. Use of ML enables analyses that would be impractical or impossible through traditional methods. Possible applications are limitless, spanning from serology-based disease diagnosis to uncovering the prognostic value of novel subgroupings of disease pathology. Consequently, ML can enhance disease diagnosis, monitoring, prognosis, precision medicine, and the overall management of MS.

Machine Learning
Algorithm Assessment

Machine learning is a process by which an algorithm iteratively alters a predictive or descriptive mathematical model to fit a data set. However, a model bespoke to specific data may not generalize well to new data. To address this, a data set

From the Program in Exercise Science, Department of Physical Therapy, Marquette University, Milwaukee, WI, USA (JC, KK, AVN). *Correspondence*: Alexander V. Ng, PhD, FACSM, Program in Exercise Science, Department of Physical Therapy, Marquette University, Milwaukee, WI 53201-1881; email: alexander.ng@marquette.edu.

FIGURE 1. Gauge of Performance by Model

Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) for 3 classification models.
An AUC of 1.00 indicates perfect classification.

ispartitioned into 2 or 3 eponymously named representative data sets: training, validation, and test sets. Models are trained on the training set before validation on the validation set, where parameters are further fine-tuned to increase performance. The optional test set allows for a final evaluation of the refined model on fresh, unseen data.

Machine learning can be split into 2 primary categories: supervised learning, in which the algorithm is trained to detect distinct classes or predict a continuous variable using prelabeled data, and unsupervised learning, in which the algorithm looks for patterns in data without direction.[15]

Supervised learning. Classification algorithms—algorithms that recognize certain types of observations from a larger collection—are assessed by their ability to categorize different classes of observations in the validation or test set. The true-positive rate, sometimes called sensitivity or recall, is the proportion of true-positives to true-positives plus false-negatives. Similarly, the true-negative rate, sometimes called specificity, is the ratio of true-negatives to true-negatives plus false-positives. Plotting the false-positive (1–specificity) rate against sensitivity creates a receiver operating characteristic curve, and the area under the curve (AUC) is used as a gauge of success, where an AUC of 1.0 indicates a perfectly discriminating algorithm (**FIGURE 1**).[16] Finally, the proportion of correctly labeled observations among all observations is known as accuracy.

Complementary to classification is regression. Whereas the target variable in classification is categorical, it is continuous in regression. In linear regression, for example, a vector is fit to data where the independent axis is predictive of the dependent axis. Although linear regression typically uses a closed-form solution (a solution calculated exactly using a finite number of operations), it may also be calculated iteratively. In contrast, many other regression models are limited to iterative learning.

Whereas classification models are assessed by how well observations are placed in discrete categories, regression models are graded on how closely the predictions match the data. Standard assessment metrics include sum of squares, mean square error, and mean absolute error.[17]

Unsupervised learning. Unlike supervised learning, unsupervised learning aims to identify data patterns without a target variable's guidance. Unsupervised learning is commonly used for dimensionality reduction and clustering (algorithmic grouping of similar data). Standard metrics include the Davies-Bouldin index, the Dunn index, and the silhouette coefficient.[18] Although these metrics may indicate the best-performing model, their cross–data set generalization ability is limited.

Algorithms

Classification algorithms seek boundaries that best separate classes. For 2-dimensional data, this boundary may be a line; for 3-dimensional data, it is a plane. Humans can easily envision 1-, 2-, or 3-dimensional space; however, computer algorithms can determine divisions in high-dimensional space. When separating data arranged in high-dimensional space, the dividing boundary is often a hyperplane—a subspace of exactly 1 dimension fewer than the space it divides.

Support vector machines are commonly used for both classification and regression. A support vector classifier calculates the hyperplane that maximizes the distance between support vectors—the closest observations to a hyperplane—to maximize class-class difference (**FIGURE 2**).[15] Similarly, a support vector regressor uses an analogous strategy to minimize the error of the support vectors, in this case, points outside the ε-tube—a predetermined space around the hyperplane. Thus, because variables in the ε-tube are not included in the error function, a support vector regression can be less sensitive to

FIGURE 2. Maximization of Class-Class Difference
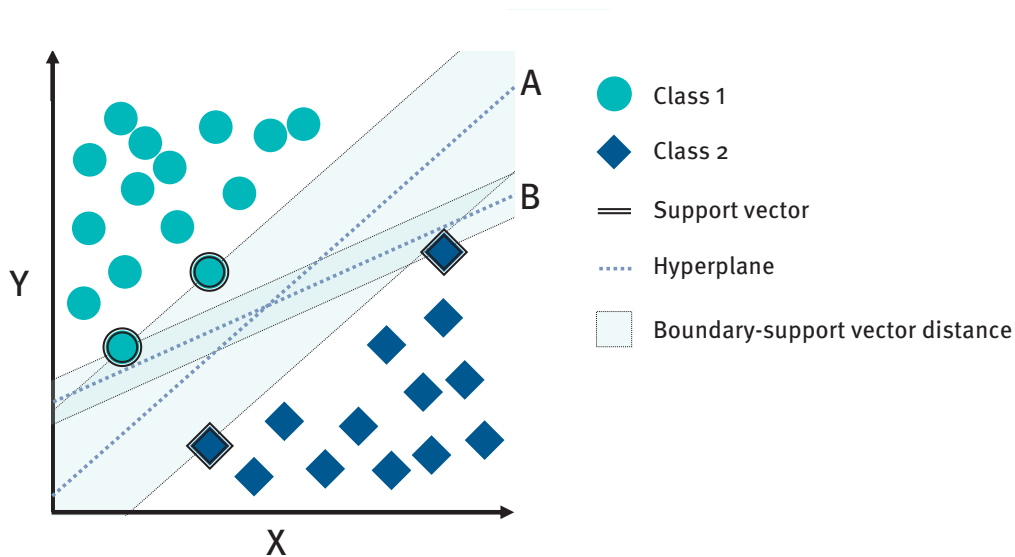
Better (line A) and worse (line B) decision boundaries of a support vector classifier as determined by the distance between the support vectors and the hyperplane.

clusters of observations relative to other algorithms, such as linear regression, that use all the data in their cost functions.[19] In addition, nonlinear hyperplanes can be used for both support vector classification[15] and regression[19] by remapping data onto a new feature space.

Neural networks consist of linked units (nodes) that create a network akin to a brain.[1,20] Each node may have multiple inputs and outputs; a weight coefficient determines the connection strength between nodes.[1,20] Whereas a living brain may upregulate or downregulate excitatory or inhibitory neural transmitters or receptors, a neural network can alter a weight coefficient of a mathematical function (**FIGURE S1**).[1] A neural network makes a prediction, and if the prediction is correct, the model remains unchanged. However, if the prediction is incorrect, coefficient weights are adjusted.[1,20] Neural networks consist of several layers: the input, the output, and 1 or more intermediary layers called hidden layers. Using multiple neurons and layers, neural networks analyze data points in relation to other data and are thus adept at pattern recognition.[20]

A decision tree classifier develops branching rules to best explain divisions between classes (**FIGURE S2**).[21] However, decision trees are particularly prone to overfitting, which reduces generalization outside the training data set.[22] Overfitting can be mitigated by building multiple weaker models and combining them into a single, stronger ensemble model; techniques include gradient boosting[23] and Bayesian ensemble learning.[24] Perhaps the most common ensemble algorithm for tree-based learning is the random forest classifier, which builds multiple decision trees, each trained on a subset of data, later combined into a single model (**FIGURE S3**)[22]; an increased number of trees diminishes overfitting, thus increasing accuracy.[21,22]

In contrast to supervised learning, unsupervised learning algorithms make conclusions about which observations are most similar without human prespecification. Common unsupervised learning applications include dimensionality reduction and clustering.

Principal components analysis is a dimensionality reduction technique that transforms correlated variables into a smaller set of uncorrelated variables called principal components.[25] Principal components are vectors fit to data to explain the maximum amount of variance[26]; the vector with the highest variance is the first principal component. Subsequent principal components are orthogonal to every previous component and explain the maximum amount of the remaining variance (**FIGURE S4**).[27] As each vector explains the maximum amount of remaining variance, each principal component explains less than all previous components. Principal components analysis helps show which variables most contribute to data variance and simplifies high-dimensional data, thus increasing interpretability.

K-means is one of many clustering algorithms used to group observations by similarity. K-means clustering sets a predetermined number of clusters, k, onto a feature space; the center of each is a point referred to as a *centroid* (**FIGURE S5**).[26] Two steps are subsequently repeated until further iterations cease to minimize the sum of squares error within clusters: (1) each observation is assigned to the cluster with the closest centroid[17,26] and (2) the centroids are moved to minimize the sum of squares within the cluster.

Subtype and Stage Inference (SuStaIn) is a novel and unorthodox clustering technique that clusters across disease stage, stratified by subtype.[28] This is in contrast to k-means clustering, which may group individuals with low disease progression into 1 category and individuals with high disease progression into another with little delineation between subtypes. Importantly, SuStaIn may infer disease progression from both cross-sectional and longitudinal data. In SuStaIn, each feature

is converted to a z score normalized to a control population. Disease trajectory is linearly modeled in a piecewise fashion by connecting the expected z score at arbitrary time points; each calculated trajectory is considered a subtype (**FIGURE S6**). SuStaIn minimizes error for both subtype and trajectory for each observation.

ML in Clinical Decision-Making and Research

We examine the application of ML to 5 areas of MS patient care: direct classification of disease following serology or medical imaging; typing of patient lesions, which may help clinicians streamline diagnostic processes; discovery of new clusters of pathology; patient movement analysis, which may assist targeted therapeutic care; and finally, projection of disease progression.

Disease Detection

Due to their pattern recognition ability, ML algorithms are commonly used to differentiate between classes of observations. Regarding MS, differentiation is often used for disease detection.

Serum Biomarkers

Machine learning approaches are frequently used for classification problems that require tabular data. Several studies have used ML to identify protein and lipid biomarker concentration patterns to differentiate between individuals with MS and individuals without MS.[29,30]

Goyal et al[29] applied 4 classification algorithms to detect MS from concentrations of 8 serum cytokines: interleukin (IL) 1β, IL-2, IL-4, IL-8, IL-10, IL-13, interferon gamma, and tumor necrosis factor alpha. These data were compiled from previous studies and comprised 956 individuals with MS and 199 controls without MS. Of the 4 models, the random forest performed best for all comparative categories, with accuracy of 90.91%, sensitivity of 0.756, specificity of 0.857, and an AUC of 0.957. The support vector machine performed second-best in all categories; however, its sensitivity and specificity were both lacking, with scores of 0.500 and 0.633, respectively. The performances of the decision tree and neural network were poor. The random forest distinguished between individuals with remitting and nonremitting MS with an accuracy of 70%.

Tsoukalas et al[30] applied a neural network to detect autoimmune disease via serum concentrations of 28 biomarkers, of which 23 were fatty acids. Their final model had an AUC of 0.792 and a predictive accuracy of 76.2%. The most predictive biomarker was cis-11-eicosenoic acid (C20:1n9), which accounted for more than 10% of the strength of the model. A principal components analysis discovered 7 principal components, accounting for more than 70% of the variance among those data.

These studies support hematology as a modality to detect autoimmune diseases; however, note that the range of biomarkers that these authors investigated was not exhaustive. Further investigation is needed to identify biomarkers indicative of clinically isolated syndrome, radiologic syndrome, and MS prodrome. Continuance of this research may enhance the ancillary diagnostic ability of serology and enable routine evaluation of patients who might otherwise remain undiagnosed due to early disease progression or symptom obfuscation by another condition.

Medical Imaging

Much MS research involving ML has revolved around differentiating individuals with MS from controls without MS and distinguishing disease stages. Such classification algorithms have been developed using various imaging techniques, including MRI and optical coherence tomography (OCT).

Eitel et al[31] developed a convoluted neural network—a subcategory of neural networks—to distinguish individuals with MS from controls via MRI. The authors acknowledged their small sample size (N = 147, individuals with MS = 76) and mitigated this restriction by pretraining the model on a larger data set to distinguish between controls and patients with Alzheimer disease. The final model in MS achieved an AUC of 96.08%, demonstrating a strong capacity for detection. In addition, the authors generated visual representations showing the significance of each pixel and, therefore, each brain region in the neural network's decisions. As such, appropriate neural network use may highlight clinically relevant lesions to assist in diagnostic decisions by radiologists.

Using OCT—a technique that measures retinal cell layer thickness—Cavaliere et al[32] developed a support vector classifier to distinguish individuals with MS from age-matched controls. The OCT images contained scans of 3 cell layers segmented into smaller regions around the macula and optical nerve head. After training, the 3 measurements with the greatest differential ability were (1) the whole ganglion cell layer, between the retinal pigment epithelium boundaries and the inner limiting membrane; (2) the inner nasal retina; and (3) the outer nasal retina. A combined model using these 3 regions displayed a classification ability with an AUC of 0.97. This study highlights a potential avenue for diagnosis in patients for whom typical diagnosis via MRI is not recommended.

Although ML is neither necessary nor sufficient for disease diagnosis, its implementation in diagnostic processes may create an assistive layer to support clinicians. ML may detect disease markers missed by clinicians and serve as a second opinion for medical professionals. Further investigation may work toward integrating ML-based disease detection into practice and focus on differentiating people with MS from those with other neurologic diseases.

Typing of Lesion Clusters

Beyond the broad clinical applicability of machine-based MS detection, ML has also demonstrated efficacy in discriminating between lesions with high and low clinical relevance. Kocsis et al[33] investigated ML's ability to differentiate MS lesion types and assess their clinical relevance using 3 MRI sequences: spin echo (SE), fast spoiled gradient echo (FSPGR), and fluid-attenuated inversion recovery. Lesions were identified, and k-means clustering (k = 2) was applied using the median intensity values of the white matter lesions from all sequences. Fluid-attenuated inversion recovery had a negligible effect

PRACTICE POINTS

Artificial intelligence, including machine learning, may assist with earlier diagnosis of multiple sclerosis by identifying previously unrecognized neurologic conditions even during unrelated health care procedures.

Artificial intelligence, including machine learning, may be used to monitor therapeutic or rehabilitation efficacy or disease progression in free-living conditions outside the constraints of a clinic. ■

on clustering. In cluster 1, 100% and 69% of lesions were easily discernable (median intensity Z ≥ 2.3) for FSPGR and SE, respectively. In contrast, for cluster 2 (high intensity), 78.7% were easily discernable for FSPGR, but this number shrank to 17.7% for SE. Cluster 2 alone revealed an association between lesions and the Expanded Disability Status Scale score.[34]

According to Kocsis et al,[33] expansion of this technique could render SE sequencing obsolete, thereby reducing the number of redundant measurements. Further investigation may explore relationships between lesion clusters and other metrics of disability and additional ways of clustering beyond median intensity values.

New Clusters of Pathology

By classifying people with comparable disease patterns, cluster analysis can be a potent tool in precision medicine. Such clustering empowers clinicians to provide patients with more individualized care, leading to better health outcomes.

Silveira et al[35] investigated symptom clusters (groupings of related and concurrent symptoms) and their relationship with quality of life (QOL) across 205 individuals with MS aged 20 to 79 years stratified into 20-year age groups. Psychological symptoms were recorded via the Pittsburgh Sleep Quality Index, the Hospital Anxiety and Depression Scale, and the Fatigue Severity Scale. In addition, the 36-Item Short Form Health Survey measured QOL.

Participants were grouped via k-means clustering (k = 3) on the psychological symptom questionnaire results, and the resultant clusters represented mild, moderate, and severe symptom experiences. Correlation and partial correlation analysis confirmed relationships between mental health measures and QOL. With 1 notable exception, the severe symptom cluster consistently displayed the highest mean symptom scores across fatigue, depression, anxiety, and sleep; the inverse was true for the mild symptom cluster. The oldest age group (60-79 years) was the lone exception, where both moderate and severe symptom clusters were associated with comparably poor sleep quality. Because 1 symptom may indicate other latent symptoms, this study reinforces the need for comprehensive mental health treatment. Furthermore, the increased prevalence of sleep-related symptoms for the eldest age group confirms the necessity for an age-related focus on mitigating sleep disorders.

Unlike the cross-sectional study by Silveira et al,[35] Eshaghi et al[36] sought to identify novel subtypes of MS while accounting for disease progression within each subtype. SuStaIn was applied to brain MRIs from 9390 individuals with MS, which identified 4 distinct subtypes, later defined as "cortex-led, normal-appearing, white matter–led, and lesion-led." Significantly, these subtypes outperformed traditional subtyping (ie, relapsing-remitting, primary progressive, and secondary progressive MS) as predictors of disability progression. Furthermore, predictive power was improved by introducing other features, including standard clinical phenotypes and the Nine-Hole Peg Test. Interestingly, persons in the lesion-led subtype were most responsive to pharmaceutical-based treatment. Such novel groupings of individuals with MS empower clinicians to better predict disease progression and provide more individualized care.

Cluster analysis can group individuals with comparable psychological or physical symptoms. However, such groupings are currently of low clinical value because it is unclear how they relate to disease progression or symptoms. Future studies will investigate new groupings and respective clinical relevance.

Analysis of Movement

Imaging is essential for physicians to assess many aspects of disease progression; however, other aspects can be measured directly at a lower cost. Several studies have implemented ML to assess movement patterns of individuals with MS compared with controls.

Movement Classification

With the rapidly increasing prevalence of so-called smartwatches, much attention has been paid to algorithmic movement classification using data gathered via their onboard sensors. Chitnis et al[37] used multiple wearable accelerometers to record dynamic movement in both laboratory and free-living conditions. A neural network was first trained on adults without MS before its application to data from adults with MS. The algorithm classified data into "run," "walk," "idle," or "other" and further classified gait segments such as "stance" for subsequent analysis. Several correlations were found between movement features gathered from machine-labeled data during free living and the Multiple Sclerosis Functional Composite,[38] including both stance time (r_s = −0.56) and leg movement rate during sleep (r_s = −0.45).

Movement classification promises to assist in automating time series segmentation and labeling, enabling research that, preceding such techniques, was impossible or insurmountably time-consuming. Wearable sensors, bolstered by movement classification, may facilitate physician assessment of free-living movement for individuals with MS and provide evaluations of MS progression without needing in-office tests. Significant research and development will be needed to create algorithms to accurately assess movement in free-living situations; however, smartwatch–based sensors offer a promising avenue to vast quantities of data.

Detection of Disease Severity

In addition to discrimination between movement types, classification via movement sensors can differentiate between individuals with MS and controls. Moreover, such models have highlighted the movement features most predictive or indicative of various stages of disability for individuals with MS.

Sun et al[39] used a random forest on force plate data to identify which postural movements best differentiate individuals with MS from controls. Balance differences were identified between individuals with MS with low fall risk and control participants, which may be difficult to measure in standard therapy. The 3 best predictors for individuals with MS with mild symptoms (by Expanded Disability Status Scale) were sway anteroposterior sample entropy, mediolateral sway range, and sway area, accounting for 15.3%, 15.1%, and 14.8% of the diagnostic performance from the 19 measured variables. Mediolateral sway range accounted for 62.3% of the performance in individuals with MS with moderate disability. Sway range in the mediolateral direction, mediolateral sway path, and mediolateral mean velocity were the best predictors for those with severe symptoms, accounting for 32.9%, 20.1%, and 18.1% of the model's performance, respectively. The final model differentiated between individuals with MS and controls with accuracy, sensitivity, and specificity all greater than 0.95.

Similar models may support therapist intervention by assessing fall risk in individuals with MS. Via an increased understanding of patient movement patterns, therapy may more readily target specific patterns of ataxia. Future studies might investigate whether fall risk can be detected separately from other disability markers. Doing so may help highlight individuals with elevated risk of falling relative to their disability status.

Movement Components

When recording many variables, as is common in gait studies, several can be deeply related. The resulting collinearity may be addressed by transforming variables into their principal components. Monaghan et al[40] applied principal components analysis to 21 recorded gait characteristics gathered using sensors on the feet, wrists, chest, and lower back. The characteristics—including stride length, swing time, and range of motion knee variability—were reduced to 6 factors that explained 79.15% of the variability. Their relative contributions to gait variance were as follows: pace (24.81%), rhythm (16.57%), variability (13.02%), asymmetry (9.27%), anteroposterior dynamic stability (8.01%),

and mediolateral dynamic stability (7.47%). Relative to controls, individuals with MS displayed significantly reduced pace, variability, and mediolateral trunk motion performance. Individuals with MS who had fallen in the past 6 months exhibited reduced pace and increased asymmetry compared with those with MS who had not fallen. Finally, pace positively correlated with scores on the Stroop task and Berg Balance Scale and negatively correlated with fear of falling.

With only 6 principal components accounting for approximately 80% of the variance across 21 measured characteristics, principal components analysis highlights collinearity in gait. Indeed, principal components analysis can help eliminate superfluous measurements, and by reducing the dimensionality of data, can simplify their interpretation.

Inference of Cognitive Impairment

Monaghan et al[40] demonstrated that physical movement patterns can predict psychological symptoms. Clinicians can infer many aspects of disease progression, including cognitive impairment, without requiring specific measurements by identifying correlations between various symptoms.

Brummer et al[41] investigated how serum neurofilament light chain was correlated with MRI lesion markers, the Expanded Disability Status Scale, and the Symbol Digit Modalities Test used to measure cognitive impairment. The study of 152 individuals with MS found that serum neurofilament light chain, via a support vector regression, could predict scores on the Symbol Digit Modalities Test with $P = .004$, standard error of 0.192, and accuracy (regression coefficient) of 0.561, and when incorporating lesion and grey matter volume into the model, accuracy validated against a new cohort increased to 90.8%.

For patients facing barriers to access, eg, those with severe vision loss, similar models may allow clinicians to predict clinically relevant metrics while eschewing the testing procedures themselves. Currently, research in this area is sparse. Future studies are needed to investigate other biomarkers and their ability to predict metrics of disability.

Prediction of Disease Progression

No single factor determines the best course of treatment in MS; instead, numerous factors, including cost, patient preferences, coexisting health issues, and disease severity, are considered. To facilitate the highest level of care, clinicians must identify patients most suited to specific trials or therapies.

In a double-blind, placebo-controlled study of dirucotide (a myelin-damaging, autoimmune response–mitigating drug), Law et al[42] compared 3 decision tree variations, 2 logistic regression variations, and 2 support vector machine variations to predict secondary progressive MS progression in a cohort of 485 individuals with MS over 6 months as measured by the Multiple Sclerosis Functional Composite. The best-performing models by AUC were a standard decision tree (0.618), a random forest (0.607), and a variant decision tree (0.602).

Foreknowledge of patient condition and disease progression can improve treatment, allowing clinicians to customize care to mitigate expected symptoms. By excluding patients with

favorable short-term disability progressions and patients likely to respond well to standard care, experimental therapies can focus on high-risk patients, maximizing potential reward and mitigating potential risk. Advanced knowledge of disease progression may comfort individuals with MS living with uncertainty. Unfortunately, despite the modest success of Law et al,[42] their best AUC was not vastly better than chance. Future studies must endeavor to create more accurate predictive models.

DISCUSSION

During the past decade, AI, particularly ML, has gained prominence in MS research and is anticipated to aid future discoveries in earlier detection, slowing disease progression, regaining function, and developing a cure. Its classification, regression, feature reduction, and clustering capability have made ML a powerful analytical tool. Further advancements may create systems that alert clinicians during ancillary testing when test results are consistent with those of individuals with MS. Moreover, early disease identification may lead to earlier treatment and improved outcomes.

Unfortunately, there are several major limitations of ML in the study and treatment of MS. These limitations include the esoteric nature of ML, current diagnostic ability, the black box nature of many algorithms, the potential for harm via algorithmic misuse, the inability of a machine to consider consequences, and both the introduction and perpetuation of bias.

Due to the relatively recent expansion of ML in MS research, its use and understanding is restricted to specialized communities. As such, some clinicians may have reservations about implementing these techniques or equipment that uses them in their practices. However, the surge in studies using ML techniques in MS research demonstrates developing mainstream adoption. This limitation will gradually relax as ML is increasingly adopted in standard practice.

Similarly, use of supervised learning in diagnosis is limited by our diagnostic ability, which relies on human-labeled data; this limitation will naturally be lessened as MS diagnostic methods advance, thus strengthening future algorithms.

Even ML itself is not limitation free. Complex ML models such as neural networks may be challenging for humans to interpret due to the high number of parameters and predictors, making them black box algorithms, which contrast with more readily interpreted models such as simple decision trees. Due to a lack of explanatory power between inputs and outputs, black box models' interpretability limits their application.[43] Therefore, researchers may opt for simpler models when interpretability is required.

Ribeiro et al[44] demonstrated the risk of black box models with a neural network seemingly trained to distinguish between photos of huskies and wolves that was instead trained to distinguish between snowy and grassy backgrounds. The same risk is true for MS research; a black box model may seem to perform well but instead relies on unintentionally powerful predictors, eg, the presence of a drug exclusively used to treat MS, to predict the presence of MS. Despite this drawback, these models have shown much promise in MS research, and this risk may be mitigated with careful data curation.

Ethical standards for ML research and clinical care must prioritize minimizing potential harm to people, just like traditional research. In some instances, the risks are comparable; the risk of an iterative linear regression, for example, is comparable with its closed-form variant. Unfortunately, ML has the potential to cause harm beyond typical research. In an infamous scenario, researchers at Target deduced that a high school student was pregnant before she told her family and unwittingly alerted her father via pregnancy-related flyers.[45] Similarly, when conducting health research in MS, researchers must be mindful of sensitive patient information and strive to protect privacy.

Like humans, AI is subject to bias, and a common way bias is introduced is through a biased training set. For example, a model trained primarily on data from a cohort of only women might not generalize well to a mixed-sex cohort. Moreover, the separation of results by sex is frequently overlooked in MS research,[46] despite significant sex-based differences in disease effects.[47] This problem could be particularly severe.

Similarly, external prejudice and practice variation can contribute to bias. For instance, because women may be less likely to be prescribed analgesics than similarly situated men,[48] a model trained on such skewed data may perpetuate this trend. Likewise, in MS, despite similar clinical circumstances, 1 sex may be prescribed certain therapies less frequently than the other.

Moreover, even unbiased algorithms may introduce automation bias: the effect of being swayed by computerized predictions that differ from one's thoughts.[43] Note that although AI might contribute to bias, it can also work to lessen it. Rather than causing more prescription disparities between men and women, a well-designed algorithm may aid in reducing them.

Finally, a machine's inability to consider the human cost of incorrect decisions may lead to catastrophic effects. Unlike humans, who may exercise caution around diagnostic uncertainty and recommend further testing, machines make no such compromise without specific instruction.[43] Indeed, because machines lack the capacity, researchers must be especially cognizant of potential harm. Currently, ML must be used only as an assistive tool, not as a decision-making authority.

Although AI, including ML, is seemingly perpetually on the frontier of medical advances, clinicians and researchers must remain conscious of the potential pitfalls of its implementation. Despite these challenges, ML is continuing to create new opportunities for research and care.

CONCLUSIONS

This literature review discusses ML's effectiveness in locating lesions, diagnosing MS, gauging level of disability, and assisting with rehabilitation. ML has created models that can identify MS from medical imaging, blood samples, and biosensor data. The capability and scope of research, diagnostics, and

rehabilitation will be expanded by further integrating ML into research and clinical equipment, ultimately enhancing the standard of care and quality of life for individuals with MS. ∎

REFERENCES

1. Artificial intelligence. In: *Encyclopedia of Management*. 8th ed. Gale; 2019:18-22.
2. McCorduck P. *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. 2nd ed. AK Peters Ltd; 2004.
3. Expert systems. In: *Encyclopedia of Management*. 6th ed. Gale; 2009:291-293.
4. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: early experiments with GPT-4. arXiv. 2023:12712. doi:10.48550/arXiv.2023.12712
5. Esposito M, De Pietro G. An ontology-based fuzzy decision support system for multiple sclerosis. *Eng Appl Artif Intell*. 2011;24(8):1340-1354. doi:10.1016/j.engappai.2011.02.002
6. Moon S, Ahmadnezhad P, Song H-J, et al. Artificial neural networks in neurorehabilitation: a scoping review. *NeuroRehabilitation*. 2020;46:259-269. doi:10.3233/NRE-192996
7. Swetlik C, Bove R, McGinley M. Clinical and research applications of the electronic medical record in multiple sclerosis: a narrative review of current uses and future applications. *Int J MS Care*. 2022;24(6):287-294. doi:10.7224/1537-2073.2022-066
8. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. *AMIA Jt Summits Transl Sci Proc*. 2020;2020:191-200.
9. Compston A, Coles A. Multiple sclerosis. *Lancet*. 2008;372(9648):1502-1517. doi:10.1016/S0140-6736(08)61620-7
10. Multiple sclerosis. In: Lackie J, Nation B, eds. *Dictionary of Biomedicine*. 2nd ed. Oxford University Press; 2019. doi:10.1093/acref/9780191829116.001.0001
11. Bjornevik K, Cortese M, Healy BC, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*. 2022;375(6578):296-301. doi:10.1126/science.abj8222
12. Simpson S Jr, Wang W, Otahal P, Blizzard L, van der Mei IAF, Taylor BV. Latitude continues to be significantly associated with the prevalence of multiple sclerosis: an updated meta-analysis. *J Neurol Neurosurg Psychiatry*. 2019;90(11):1193-1200. doi:10.1136/jnnp-2018-320189
13. Mokry LE, Ross S, Ahmad OS, et al. Vitamin D and risk of multiple sclerosis: a mendelian randomization study *PLoS Med*. 2015;12(8):e1001866. doi:10.1371/journal.pmed.1001866
14. Marrie RA, Allegretta M, Barcellos LF, et al. From the prodromal stage of multiple sclerosis to disease prevention. *Nat Rev Neurol*. 2022;18(9):559-572. doi:10.1038/s41582-022-00686-x
15. Kim SB, Sukchotrat T. Data mining. In: Salkind NJ, ed. *Encyclopedia of Research Design*. SAGE Reference; 2010:328-332.
16. Colman AM. Receiver operating characteristic. In: *A Dictionary of Psychology*. 4th ed. Oxford University Press; 2015.
17. Tiwari AK. Introduction to machine learning. In: *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications*. IGI Global; 2020:41-51.
18. Marquez FPG. Silhouette coefficient. In: Marquez FPG, ed. *Handbook of Research on Big Data Clustering and Machine Learning*. IGI Global; 2020:90.
19. Awad M, Khanna R. Support vector regression. In: Awad M, Khanna R, eds. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress; 2015:67-80. doi:10.1007/978-1-4302-5990-9
20. Lerner KL, Lerner BW. Neural networks. In: Lerner KL, Lerner BW, eds. *Computer Sciences*. 2nd ed. Macmillan Reference USA; 2013:177-181.
21. Rokach L, Maimon O. *Data Mining With Decision Trees: Theory and Applications*. 2nd ed. World Scientific; 2008. *Machine Perception and Artificial Intelligence*; vol 81.
22. Ho TK. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol 1. Institute of Electronics and Electrical Engineers; 1995:278-282. doi:10.1109/ICDAR.1995.598994
23. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. *Artif Intell Rev*. 2021;54(3):1937-1967. doi:10.1007/s10462-020-09896-5
24. Du J, Linero AR. Interaction detection with Bayesian decision tree ensembles. *PMLR*. 2019;89:108-117.
25. Xu C, Li J, Zou C, Ni S. Application of hierarchical clustering based on principal component analysis to railway station classification. In: Liu X, Peng Q, Wang KCP, eds. *Proceedings of the Sixth International Conference on Transportation Engineering*. American Society of Civil Engineers; 2020:162-171.
26. Gurbuz AC, Ahmad F. General workflow of an ML algorithm. In: Gurbuz SZ, ed. *Deep Neural Network Design for Radar Applications*. Institution of Engineering and Technology; 2021:53-54.
27. Gurbuz SC, Amin MG. Classification with data-driven learning. In: Fioranelli F, Griffiths H, Ritchie M, Balleri A, eds. *Micro-Doppler Radar and Its Applications*. Institution of Engineering and Technology; 2020:110-115.
28. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun*. 2018;9(1):4273. doi:10.1038/s41467-018-05892-0
29. Goyal M, Khanna D, Rana PS, et al. Computational intelligence technique for prediction of multiple sclerosis based on serum cytokines. *Front Neurol*. 2019;10:781. doi:10.3389/fneur.2019.00781
30. Tsoukalas D, Fragoulakis V, Sarandi E, et al. Targeted metabolomic analysis of serum fatty acids for the prediction of autoimmune diseases. *Front Mol Biosci*. 2019;6:120. doi:10.3389/fmolb.2019.00120
31. Eitel F, Soehler E, Bellmann-Strobl J, et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *Neuroimage Clin*. 2019;24:102003. doi:10.1016/j.nicl.2019.102003
32. Cavaliere C, Vilades E, Alonso-Rodríguez MC, et al. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors (Basel)*. 2019;19(23):5323. doi:10.3390/s19235323
33. Kocsis K, Szabó N, Tóth E, et al. Two classes of T1 hypointense lesions in multiple sclerosis with different clinical relevance. *Front Neurol*. 2021;12:619135. doi:10.3389/fneur.2021.619135
34. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444-1452. doi:10.1212/wnl.33.11.1444
35. Silveira SL, Cederberg KLJ, Jeng B, et al. Symptom clusters and quality of life in persons with multiple sclerosis across the lifespan. *Qual Life Res*. 2021;30(4):1061-1071. doi:10.1007/s11136-020-0268
36. Eshaghi A, Young AL, Wijeratne PA, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun*. 2021;12(1):2078. doi:10.1038/s41467-021-22265-29-x
37. Chitnis T, Glanz BI, Gonzalez C, et al. Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis. *NPJ Digit Med*. 2019;2:123. doi:10.1038/s41746-019-0197-7
38. Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. *Mult Scler*. 1999;5(4):244-250. doi:10.1177/135245859900050040939
39. Sun R, Hsieh KL, Sosnoff JJ. Fall risk prediction in multiple sclerosis using postural sway measures: a machine learning approach. *Sci Rep*. 2019;9(1):16154. doi:10.1038/s41598-019-52697-2
40. Monaghan AS, Huisinga JM, Peterson DS. The application of principal component analysis to characterize gait and its association with falls in multiple sclerosis. *Sci Rep*. 2021;11(1):12811. doi:10.1038/s41598-021-92353-2
41. Brummer T, Muthuraman M, Steffen F, et al. Improved prediction of early cognitive impairment in multiple sclerosis combining blood and imaging biomarkers. *Brain Commun*. 2022;4(4):fcac153. doi:10.1093/braincomms/fcac153
42. Law MT, Traboulsee AL, Li DK, et al. Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. *Mult Scler J Exp Transl Clin*. 2019;5(4):2055217319885983. doi:10.1177/2055217319885983
43. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*. 2019;28(3):231-237. doi:10.1136/bmjqs-2018-008370
44. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2016:1135-1144. doi:10.1145/2939672.2939778
45. Duhigg C. How companies learn your secrets. *The New York Times Magazine*. February 16, 2012. www.nytimes.com/2012/02/19/magazine/shopping-habits.html
46. Alonso-Moreno M, Ladrón-Guevara M, Ciudad-Gutiérrez P. Systematic review of gender bias in clinical trials of monoclonal antibodies for the treatment of multiple sclerosis. *Neurologia (Engl Ed)*. 2021;S0213-4853(21)00008-6. doi:10.1016/j.nrl.2021.01.003
47. Voskuhl RR. The effect of sex on multiple sclerosis risk and disease progression. *Mult Scler*. 2020;26(5):554-560. doi:10.1177/1352458519892491
48. Chen EH, Shofer FS, Dean AJ, et al. Gender disparity in analgesic treatment of emergency department patients with acute abdominal pain. *Acad Emerg Med*. 2008;15(5):414-418. doi:10.1111/j.1553-2712.2008.00100.x