Application of machine learning in the diagnosis of axial spondyloarthritis

Jessica A. Walsh, Martin Rozycki, Esther Yi, and Yujin Park

INTRODUCTION

Machine learning, a specialization arising from statistics and computer science, operates on the basis of learning relationships from data sets collected from computing algorithms; in short, this scientific discipline focuses on how computers learn from data [1,2]. Driven by the digitalization and storage of data in healthcare and the availability of open-source tools and codes, machine learning can be used to effectively analyze large amounts of data on patient history, laboratory results, treatments, diagnoses, and outcomes [3–6]. Machine-learning models are not new in rheumatology, as numerous algorithms have been developed to detect disease and outcomes, particularly in rheumatoid arthritis (RA). A prediction model successfully prognosticated a 1-year response to certolizumab pegol among patients with RA using clinical and laboratory data collected at baseline and at 12 weeks [7]. Three models identified RA phenotypes [8], synovial subtypes [9], and disease activity [10]. Mortality due to RA was predicted based on demographic and clinical data collected in the first 2 years after diagnosis [11**]. Mass spectrometry and magnetic beads were used to screen for serum protein biomarkers from patients with various autoimmune disease, including RA and systemic lupus erythematosus (SLE); a machine-learning model successfully predicted a diagnosis of SLE based on four putative biomarkers [12]. Differential gene expression profiles for rheumatoid factor were used to train predictive models for the diagnosis of polyarticular
The ability to identify meaningful patterns in large data sets makes machine learning particularly attractive in screening for undiagnosed diseases such as axial spondyloarthritis (axSpA), a spectrum of chronic conditions characterized by inflammation in the axial skeleton that can cause pain, joint damage, and disability [17,18]. Ankylosing spondylitis is a subtype of axSpA characterized by radiographic sacroiliitis [19]. Here, we include ankylosing spondylitis in our definition of axSpA and will use ‘ankylosing spondylitis’ only for studies that specifically addressed ankylosing spondylitis and not axSpA.

Epidemiology data from 2009 to 2010 indicate that axSpA affects 1.4% or less of the adult general population in the United States [20]. With advances in imaging technology and therapies, evidence is mounting that axSpA is widely underrecognized, undertreated, and understudied [21–23]. The estimated delay between symptom onset and diagnosis of ankylosing spondylitis in the United States is approximately 13 years [24]. This delay in diagnosis contributes to the significant burden on patients, caregivers, and society [25–30]. Reasons for delayed diagnosis are intrinsic to the disease and systemic to the healthcare system. Intrinsic features of the disease such as the common presentation of low back pain [29], lack of awareness of differential disease presentations among men and women [31–34], insidious onset, slow progression, lack of specific biomarkers [21,35,36], absence of remarkable physical findings among patients with early stages of axSpA and ankylosing spondylitis [37], and lack of extra-articular manifestations [37] contribute to misdiagnosis [25] and complicate early identification and diagnosis. The lack of accessibility to rheumatologists and long waiting times are also barriers to a timely diagnosis [38,39]. Consequently, as patients continuously experience inaccurate diagnoses and unsuccessful interventions, they may believe that nothing can be done and refrain from seeking appropriate medical care [24].

### OPPORTUNITIES AND POTENTIAL BENEFITS WITH MACHINE LEARNING IN AXIAL SPONDYLOARTHRITIS DETECTION

Traditional approaches to improving early axSpA identification have had limited success. Machine learning approaches may create opportunities to transfer some of the burden of disease detection away from healthcare providers and patients and potentially decrease the time to diagnosis. In an attempt to aid in the earlier diagnosis of axSpA, we developed machine-learning models to predict a diagnosis of these diseases using administrative claims [40] and electronic medical record (EMR) [41–43] data. In the claims-based model, the positive predictive value in predicted patients (6.24%) was 5 times higher compared with that of a clinical model developed based on ankylosing spondylitis clinical features (1.29%) [40]. The EMR-based machine-learning model identified patients with axSpA at accuracies ranging from 82.6 to 91.8% [41–43]. Using these models, we hope to aid in the timely diagnosis of axSpA, thus improving the short-term and long-term outcomes for patients with axSpA [21,22].

### Cost savings

Patients often seek consultations with various types of healthcare providers in their journey toward diagnosis [25], undergoing expensive, unnecessary diagnostic workup and inappropriate medical interventions that may not directly improve patient outcomes [25,44,45]. For example, an increase in opioid use was reported in patients with ankylosing spondylitis, representing a suboptimal treatment with economic ramifications to the health system [46]. Thus, a timely diagnosis of axSpA can yield...
Data application and novel research

In recent years, there have been expanded efforts to improve knowledge on axSpA using big data and machine-learning techniques [53]. With technological advances in imaging modalities and treatment, researchers identified many individuals with nontraditional axSpA phenotypes that were previously not recognized [53,54]. Despite extensive acceptance of the broader, newly defined axSpA concepts, prominently the recognition of ankylosing spondylitis characteristics, big data axSpA research remains inhibited by antiquated axSpA definitions because international codes used for disease classification and billing (e.g., International Classification of Diseases, Ninth and Tenth Revisions) currently exist only for the traditionally recognized phenotype of ankylosing spondylitis [24,55,56]. As a result, characteristics of the overall population of patients with axSpA are not as well known as those with ankylosing spondylitis [57–59].

CHALLENGES AND STRENGTHS WITH MACHINE LEARNING

It must be noted that machine learning is still in a discovery phase and that the ubiquity of electronic health big data across many clinical domains may not mean that machine-learning techniques will prove equally valuable in each domain [60]. We must also consider that machine-learning techniques are subject to biases, including but not limited to missing data, misclassifications, and measurement errors, that may perpetuate existing healthcare inequalities. Therefore, physicians should continue to rely on clinical judgment in conjunction with applications of artificial intelligence.

Data quality, availability, and accuracy

Performance of machine-learning models is largely dependent on data quality, including data availability and accuracy. Incomplete and inaccurate data sets will limit the usefulness of predictions made by machine-learning models. Modeling of healthcare data in the United States is challenging because the relationship dynamics among stakeholders involved in patient care are complicated; variability in physician and patient behavior is somewhat expected across institutions. Furthermore, these behaviors may be dictated by reimbursement policies and coverage decisions made by the government and payers [61]. For instance, a nonuniversal healthcare system involves a multitude of commercial payers at the national level with different coverage on services and treatments compared with Medicare and Medicaid. The variability increases within Medicaid at the state level.

Data availability differs according to the data source. Most current machine-learning algorithms use claims and EMR data. In the following section, we use these data sources to illustrate our points; however, we acknowledge that other data sources, such as registries and survey data, can also be used to build machine-learning models.

Claims-based models permit the use of US healthcare data and therefore provide perspectives specific to the United States. Claims data consist of billing codes, which can be used to track the patient’s healthcare resource utilization and expenditure even if they involve multiple physician groups and practice settings [62], whereas EMR data may be limited to specific healthcare or hospital systems (e.g., the Veterans’ Health Administration) and therefore cannot be used to capture the patient’s activities outside those systems [62]. However, because claims data consist of billing codes, the clinical processes leading to a specific diagnosis, procedure, and treatment are unknown [62]. In addition, claims data may show the billing records of diagnostic tests, but not the results; if a patient pays out of pocket (e.g., for over-the-counter drugs and chiropractor visits), claims data will not be able to capture this information [62]. On the other hand, EMR data provide detailed information on physicians’ assessments and the decision-making process involved with patient care, along with information on disease severity, diagnostic results, and treatments received [62]. Although EMR data comprise smaller patient populations compared with claims...
data, the increasing interoperability of different EMR systems due to mergers of healthcare systems and the creation of common data models confer the potential of a larger patient cohort in the future [62]. Although the volume of data analyzed may be higher, claims data also offer less clinical information and do not permit validation from physicians or patients. External validation confers strength to the predictive capability of machine-learning models [8,11**].

The access to information is limited by data made available based on state-level or institution-level policy requirements or individual healthcare provider preference. For instance, variability in how healthcare providers evaluate their patients and document their case notes may affect the availability of EMR data that can be processed by machine-learning algorithms. Claims data are limited to information related to billing and reimbursement, whereas survey and registry data are bound by the research aims of investigators or study sponsors. Therefore, limited or lack of data availability may not necessarily indicate that the data set did not contain pertinent information.

Data accuracy is limited by human biases and misconceptions that are inevitably introduced into provider notes and other data sources that are used for machine-learning. For example, women and nonwhite patients with axSpA have not been well studied, recognized, or diagnosed relative to men and white patients with axSpA [20,63–68]; these limitations must be considered when designing and interpreting machine-learning outputs [69].

**Data curation and validation**

Raw data from EMR and claims databases may not be suitable for machine learning and may require restructuring in a specific manner. The resource-intensive, reiterative process of data cleaning, extraction, linking, regrouping, analyzing, and interpretation [41] is performed in the hopes of finding the underlying motif or pattern that can also be clinically relevant; this is critical to consider as we continue to build, analyze, and apply these models in the healthcare field.

To construct machine-learning models that are practical for use in the field, continuous input from healthcare providers will be necessary. Consider an example involving computerized clinical decision support (CDS) systems, which provide alerts and notifications to providers with regard to medication dosing. Clinicians typically individualize dosing based on a myriad of factors, such as the age and weight of the patient. CDS systems are programmed to warn providers of overdosing or underdosing based on strict clinical guidelines. However, customization of these systems is required for practical use, for example, for an adult patient who may need pediatric dosing support due to contraindications. Therefore, feedback from healthcare providers is necessary and crucial to customize and improve these systems for optimal use. In turn, their participation in the building or customization of these machine-learning algorithms will compel them to better understand and apply these models. Drawing from our experience in building our claims-based machine-learning model, we knew that hypertension was initially identified by the algorithm as a predictor for ankylosing spondylitis diagnosis. Although hypertension is frequently billed by primary care physicians, it is a common comorbidity and may not necessarily qualify as a predictor of ankylosing spondylitis. Thus, we excluded it from further training. However, we later realized that this predictor may indicate that the patients were repeatedly visiting primary care physicians and was coded along with back pain during consultations with general practitioners. Although it is not clinically intuitive, hypertension may have been an indirect marker for axSpA, and in combination with other variables, could potentially be used as a predictor. As machine-learning models are still in early phases of development, the use of these models in combination with targeted clinical evaluations may be more effective in obtaining an early axSpA diagnosis. Last, validation of machine-learning models in various data sets and preferably in a prospective manner will be important to improve its predictability and demonstrate its clinical relevance.

**Ethical considerations**

It is critical to provide transparency about the performance, clinical relevance, and limitations of machine-learning models to relevant stakeholders and to continuously evaluate their regulatory, legal, ethical, and social consequences. As an example, regulations on reuse of patient data vary based on jurisdiction [70**], and considerations will need to be made when choosing data sets. In the European Union, the European General Data Protection Regulation enacted informed-consent conditions for the use of data from European Union residents, regardless of where the data are processed [71]. On the other hand, the Health Insurance Portability and Accountability Act in the United States focuses on primary data sources of patient data, such as patient charts, but not on secondary data sources where patient data originate from noncovered entities, such as life insurance companies [72**]. Lastly, developers and institutions deploying these machine-learning models will also need to consider
legal liabilities and potential adverse events associated with its integration into CDS system.

CONCLUSION

Machine-learning algorithms may have a substantial role in medical diagnosis, especially in under-recognized diseases, such as axSpA. Machine-learning models that account for clinical significance appear to be the most promising [73]; however, it will be important to explain its limitations in addition to the opportunities for healthcare providers and patients. A timely diagnosis of axSpA may be possible with this analytic approach, but further refinements will be needed to optimize its operability and ability to correctly distinguish patients with ankylosing spondylitis/axSpA from the general population. As more applications employ these machine-learning techniques, we must not overlook the need to consider potential ethical and regulatory issues.

Acknowledgements
Support for third-party writing assistance for this article, furnished by Kheng Bekdache, PhD, of Health Interactions, Inc, was provided by Novartis Pharmaceuticals Corporation, East Hanover, NJ.

Financial support and sponsorship
The article was funded by Novartis Pharmaceuticals Corporation, East Hanover, NJ.

Conflicts of interest
J.A.W. is a consultant for Novartis Pharmaceuticals Corporation. M.R. is an employee of HVH Precision Analytics, LLC. E.Y. is a postdoctoral fellow at the University of Texas at Austin and Baylor Scott and White Health, providing services to Novartis Pharmaceuticals Corporation. Y.P. is an employee of Novartis Pharmaceuticals Corporation.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as: ■ of special interest ■ ■ of outstanding interest

1. Curtis JR, Hamold LR, Asgari MM, et al. Diagnostic prevalence of ankylosing spondylitis using computerized healthcare data, to 2009; underrecognition in a US healthcare setting. Perm J 2016; 20:4–10.
2. O’Mahony C, Jchi F, Pavliou M, et al. A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). Eur Heart J 2014; 35:2010–2020.
3. Stump M, Peng L. Google AI blog: assisting pathologists in detecting cancer with deep learning. 2017. Available at: https://research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html. [Accessed 21 March 2019]
4. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016; 316:2402–2410.
5. Ferran B, Channanah AM, Behbehani K, et al. Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait – a cohort study. BMJ Open 2013; 3:e002457.
6. Waljee AK, Lipson R, Wittal WL, et al. Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning: a retrospective study. Gastrointest Endosc 2017; 84:45–53.
7. Curtis JR, Lutfens K, Kavaughn A. Predicting future response to certolizumab pegol in rheumatoid arthritis patients: features at 12 weeks associated with low disease activity at 1 year. Arthritis Care Res (Hoboken) 2012; 64:658–667.
8. Zhou SM, Fernandez-Gutierrez F, Kennedy J, et al. Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis. PLoS One 2016; 11:e0154515.
9. Orange DE, Agius P, DiCarlo EF, et al. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. Arthritis Rheumatol 2018; 70:690–701.
10. Lin C, Karlson EW, Cawtho H, et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. PLoS One 2013; 8:e69932.
11. Lezcano-Valverde JM, Salazar F, Leon L, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. Sci Rep 2017; 7:10189.
12. Authors developed a mortality prediction model applied to patients with rheumatoid arthritis (RA) and based on data collected in the first 2 years after patients received their RA diagnosis. Their model was applied to an independent RA cohort for external validation, resulting in a small increase in prediction error.
13. Huang Z, Shi Y, Cai B, et al. MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of systemic lupus erythematosus. Rheumatology (Oxford) 2009; 48:628–631.
14. Jiang K, Sawle AD, Frank MB, et al. Whole blood gene expression profiling predicts therapeutic response at six months in patients with polyarticular juvenile idiopathic arthritis. Arthritis Rheumatol 2014; 66:1363–1371.
15. Li Y, Sun X, Zhang X, et al. Establishment of a novel diagnostic model for Sjogren’s syndrome by proteomic fingerprinting. Clin Rheumatol 2014; 33:1745–1760.
16. Windrup KL, Strand V, van der Heijde D, et al. The unmet need in rheumatology: reports from the targeted therapies meeting. Clin Immunol 2018; 186:87–93.
17. Giacomelli R, Afeltra A, Alouno A, et al. International consensus: What else can we do to improve diagnosis and therapeutic strategies in patients affected by autoimmune rheumatic diseases (rheumatoid arthritis, spondyloarthritides, systemic sclerosis, systemic lupus erythematosus, antiphospholipid syndrome and Sjogren’s syndrome)? The unmet needs and the clinical grey zone in autoimmune disease management. Autoimmun Rev 2017; 16:911–924.
18. Rudwaleit M, van der Heijde D, Landewe R, et al. The assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general. Ann Rheum Dis 2011; 70:225–31.
19. Rudwaleit M, van der Heijde D, Landewe R, et al. The development of Assessment of SpondyloArthritis International Society classification criteria for axial spondyloarthritis (part II); validation and final selection. Ann Rheum Dis 2009; 68:777–783.
20. Braun J, Sieper J. Ankylosing spondylitis. Lancet 2007; 369:1379–1390.
21. Revelle JD, Witter JP, Weissman MH. Prevalence of axial spondyloarthritis in the United States: estimates from a cross-sectional survey. Arthritis Care Res (Hoboken) 2012; 64:905–910.
22. Rudwaleit M, Khan MA, Sieper J. The challenge of diagnosis and classification in early ankylosing spondylitis: do we need new criteria? Arthritis Rheum 2005; 52:1000–1008.
23. Seo MR, Baik HL, Yoon HH, et al. Delayed diagnosis is linked to worse outcomes and unfavourable treatment responses in patients with axial spondyloarthritis. Clin Rheumatol 2015; 34:1397–1405.
24. Haroon N, Inman RD, Leach TJ, et al. The impact of tumor necrosis factor alpha inhibitors on radiographic progression in ankylosing spondylitis. Arthritis Rheum 2013; 65:2645–2654.
25. Deodhar A, Mittal M, Reilly P, et al. Ankylosing spondylitis diagnosis in US patients with back pain: identifying providers involved and factors associated with rheumatology referral delay. Clin Rheumatol 2015; 35:1769–1776.
26. Ogdie A, Novell WB, Reynolds R, et al. Real-world patient experience on the path to diagnosis of ankylosing spondylitis. Rheumatol Ther 2016. (In press).
27. O’Cearc C, Taskaynatan MA, Dincer U, et al. Whole blood gene expression profiling predicts therapeutic response at six months in patients with polyarticular juvenile idiopathic arthritis. Arthritis Rheumatol 2014; 66:1363–1371.
28. Zhou SM, Fernandez-Gutierrez F, Kennedy J, et al. Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis. PLoS One 2016; 11:e0154515.
29. Orange DE, Agius P, DiCarlo EF, et al. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. Arthritis Rheumatol 2018; 70:690–701.
30. Lin C, Karlson EW, Cawtho H, et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. PLoS One 2013; 8:e69932.
31. Lezcano-Valverde JM, Salazar F, Leon L, et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. Sci Rep 2017; 7:10189.
32. Authors developed a mortality prediction model applied to patients with rheumatoid arthritis (RA) and based on data collected in the first 2 years after patients received their RA diagnosis. Their model was applied to an independent RA cohort for external validation, resulting in a small increase in prediction error.
Use of machine learning in diagnosing axial spondyloarthritis Walsh et al.