## REVIEW ARTICLE     OPEN

Check for updates

# A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases

I. S. Stafford [ID][1,2], M. Kellermann [ID][1], E. Mossotto[1,2], R. M. Beattie[3], B. D. MacArthur [ID][2] and S. Ennis [ID][1 ✉]

Autoimmune diseases are chronic, multifactorial conditions. Through machine learning (ML), a branch of the wider field of artificial intelligence, it is possible to extract patterns within patient data, and exploit these patterns to predict patient outcomes for improved clinical management. Here, we surveyed the use of ML methods to address clinical problems in autoimmune disease. A systematic review was conducted using MEDLINE, embase and computers and applied sciences complete databases. Relevant papers included "machine learning" or "artificial intelligence" and the autoimmune diseases search term(s) in their title, abstract or key words. Exclusion criteria: studies not written in English, no real human patient data included, publication prior to 2001, studies that were not peer reviewed, non-autoimmune disease comorbidity research and review papers. 169 (of 702) studies met the criteria for inclusion. Support vector machines and random forests were the most popular ML methods used. ML models using data on multiple sclerosis, rheumatoid arthritis and inflammatory bowel disease were most common. A small proportion of studies (7.7% or 13/169) combined different data types in the modelling process. Cross-validation, combined with a separate testing set for more robust model evaluation occurred in 8.3% of papers (14/169). The field may benefit from adopting a best practice of validation, cross-validation and independent testing of ML models. Many models achieved good predictive results in simple scenarios (e.g. classification of cases and controls). Progression to more complex predictive models may be achievable in future through integration of multiple data types.

## INTRODUCTION

### Autoimmune disease

Three elements contribute to autoimmune disease development: genetic predisposition, environmental factors and immune system dysregulation (Fig. 1). Due to the heterogeneity of onset and progression, diagnosis and prognosis for autoimmune disease is unpredictable.

A predisposition to autoimmunity is strongly linked to genetics, and caused by defects in mechanisms that result in loss of self-tolerance.[1] Autoimmune disease develops after further immune system dysregulation, in both the innate and adaptive immune system.[2] Microbial antigens, foreign antigens and cytokine dysregulation, can cause induction of self-reactive lymphocytes.[3] Moreover, hyper-activation of T and B cells may occur, along with a change in the duration and quality of their response, which further disrupts the homeostasis of the immune system.[2]

The prevalence of autoimmune disease is difficult to estimate; diseases are variably represented across different studies and no definitive list exists.[4–6] There is a reported prevalence rate of between 4.5%[5] and 9.4%,[4] across all autoimmune diseases.

### The importance of personalised medicine

Personalised care is valuable for autoimmune disease, with variability within the disorders,[7] and presence of autoimmune comorbidities for 15–29% of patients.[8–11] Arguably, patients with multiple autoimmune comorbidities would particularly benefit from personalised healthcare for the causal molecular mechanism as opposed to specialist treatment of symptoms.

### The data revolution

Standard patient care generates diverse clinical data types. Examples of such data include laboratory test results from blood or urinary samples, symptoms at diagnosis and images obtained using colonoscopies and magnetic resonance imaging (MRI). The majority of these data are reproduced longitudinally over a chronic disease course.

In addition to this wealth of clinical data, 'omic data—such as patients' genomic, transcriptomic and proteomic profiles—are now increasingly available. 'Omic data are large, as molecular measurements are made on a genome-wide scale,[12] and high throughput omics technologies have allowed fast analysis of these data. The inclusion of multiple types of 'omic data into machine learning models may give a more complete picture of autoimmune disease, leading to novel insights.

### The need for artificial intelligence and machine learning

Combined clinical and 'omic data have limited utility without methods for interpretation. Artificial intelligence and machine learning techniques have the capacity to identify clinically relevant patterns amongst an abundance of information,[13] fulfilling an unmet need. The ability to stratify patient's using these data has implications for their care, from estimation of autoimmune disease risk, diagnosis, initial and ongoing management, monitoring, treatment response and outcome.

### Defining artificial intelligence and machine learning

The terms "machine learning" (ML) and "artificial intelligence" (AI) are often conflated. Artificial intelligence is the study of methods to imitate intelligent human behaviour (for example to make

[1]Department of Human Genetics and Genomic Medicine, University of Southampton, Southampton, UK. [2]Institute for Life Sciences, University of Southampton, Southampton, UK. [3]Department of Paediatric Gastroenterology, Southampton Children's Hospital, Southampton, UK. ✉email: s.ennis@soton.ac.uk
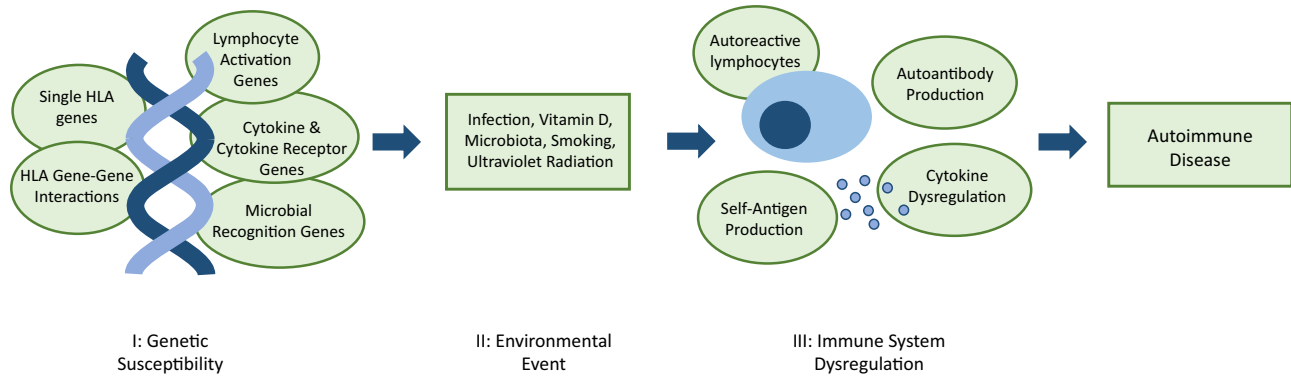
**Fig. 1 The three factors contributing to autoimmune disease development. I** genetic susceptibility is conferred by a combination of genes that may include genes encoding human leukocyte antigen (HLA) innate and adaptive immune proteins, and directly or indirectly affect the regulation of the immune system. **II** examples of potential environmental triggers for dysregulation. **III** autoantibody production alone will not always result in disease development; self-antigen production and subsequent elevated immune response is necessary.[3]

---

**Box 1** Popular supervised machine learning methods

Neural networks: outputs are learned from inputs via a series of nested nonlinear functions, encoded in a network of "neurons", which may vary in its topology.[59]
Decision trees: outputs are learned from inputs via a series of yes/no questions that successively divide the predictor space into discrete piece.[175]
Random forest: a simple ensemble method that grows a large number of decision trees, each of which see only a subset of the data, and learns output from input by combining the predictions.[79]
k nearest neighbours: learns output from input by comparing the identity of each data point to its (k) nearest neighbours.[117]
Support vector machine: a binary classification method that can be adapted to multiclass classification or regression. They seek to partition the predictor space into two, such that data points from each class are concentrated on one side of the decision boundary.[118]
Natural language processing: a set of advanced ML methods that seek to extract sentiment from text.[19]

---

decisions under conditions of uncertainty). Machine learning is a subset of AI that focuses on the study of algorithms that enable a computer to perform specific tasks (typically classification or regression) without specific instructions, but instead inferring patterns from data.[14] Both AI and ML differ from traditional statistical methods as they focus on prediction and classification from high-dimensional data, rather than inference. Successful ML requires robust data from which it can learn. These data must be sufficiently abundant to enable the model to be robust and generalisable to unseen data.

Supervised and unsupervised machine learning
Two types of ML are discussed here: supervised and unsupervised learning. During supervised learning, an algorithm is trained on a "training dataset" to recognise the patterns that are associated with specific "labels" (for example, healthy or diseased). Once predictive patterns have been learned from training data, the ML algorithm is then able to assign labels to unseen "test data". In a well-trained model, the patterns identified in the training data will generalise to the test data. Brief descriptions of some of the most common supervised ML techniques referred to in this review are summarised in Box 1.

For unsupervised learning, training data are unlabelled, and the algorithm instead attempts to find and represent patterns within the data, for example by identifying clusters based upon the similarity of the examples. Other types of ML exist, but are reviewed elsewhere.[15] Some of the more common unsupervised methods discussed in this review include hierarchical clustering and self-organising maps.

The pros and cons of alternative machine learning models
Recommendations cannot be made on the best model to use in general, as this is always situation specific and dependent on data type, size and dimensionality. Decision trees are simple and highly interpretable, but they rarely achieve performance accuracies higher than other algorithms. Using the random forest method can improve performance, at the cost of losing some interpretability. K nearest neighbours is a non-parametric method, and copes well when complex boundaries separate classes, but this flexibility can lead to poor classification results due to over-fitting.[16] Neural networks and support vector machines have similar strengths and weakness: they achieve high accuracies, and can extract linear combinations of features, but interpretability is poor, scaling to very large data can be difficult, and they are not robust to outliers.[17]

Technical aspects regarding the operation and fitting of machine learning algorithms are outside the scope of this systematic review, but comprehensively discussed elsewhere.[16,17]

Avoiding overfitting
Machine learning models are often complicated, and can involve optimizing many free parameters. For this reason, they are prone to overfitting. Overfitting is the process by which the algorithm learns patterns that are specific to the training data but do not generalise to test data. For example, there may be some random technical error in the training data that is not of clinical relevance, yet is learned by the algorithm. Training any model accurately while avoiding overfitting is a central part of an ML pipeline. If data are abundant and/or the ML model is computationally expensive to train then the standard strategy is to remove a portion of the data for training, optimize the model on the remaining portion, and finally determine the model performance by comparison with the unseen test portion. If data are not abundant, a process known as cross-validation is typically employed (Fig. 2). There are many variations of cross-validation but they are all essentially generalisations of the training/test splitting process described above. For example, in k-fold cross validation, the data are randomly split into k subsets, with all but one subset used to train an ML model, and the remaining subset used to test the model. This process leads to the generation of k ML models. Each subset of the data is used only once as a test set, and overall model performance is determined by averaging the performance of the k models (Box 2 describes model evaluation metrics).[18]
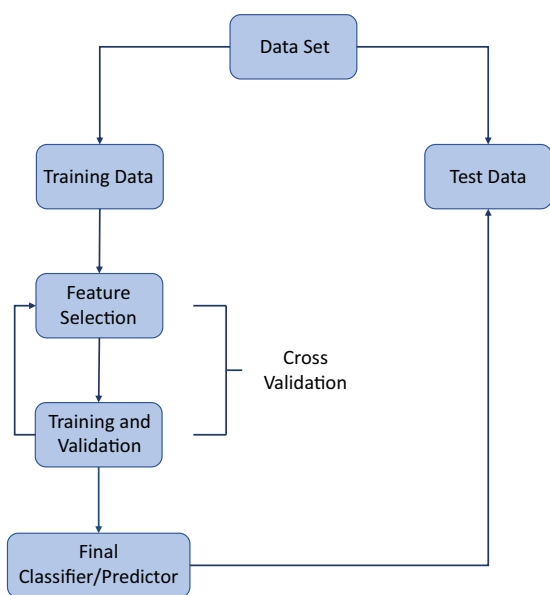
**Fig. 2 Simplified workflow for developing a machine learning model.** This includes the cycle of feature selection, training and validation that is required to avoid overfitting (cross validation).

---

**Box 2** Metrics for ML method evaluation

Accuracy: percentage of correct predictions.[198]
Area under the receiver-operator curve (AUC): appropriate for binary classification problems, this method uses a plot of sensitivity versus specificity to determine model performance.[16]
Balanced accuracy: measure of the total number of correct predictions in either class, therefore taking into account an unbalanced dataset.[198]
F-score: an accuracy measure calculated using precision and recall.[199]
Out-of-bag error: this metric applies to tree-based ensemble methods, and measures the test error by comparing predictions with true labels for samples that were not used in the construction of a particular decision tree.[16]
Precision: equivalent to positive predictive value.[16]
Recall: another term for sensitivity.[16]
$R^2$: measures the amount of variation explained by the model regression.[16]
Sensitivity: correctly identified true positives.[16]

---

Artificial intelligence, machine learning and autoimmune disease
This systematic review aims to inform on the current status of the application of artificial intelligence and machine learning methods to autoimmune disease to improve patient care. To the best knowledge of the researchers, this is the first study on this topic. The review identifies the most common methods, data and applications, the issues surrounding this exciting interdisciplinary approach, and promising future possibilities.

## RESULTS

### Summary of results
Of 702 papers identified in database searches, 169 were selected for inclusion in the analysis, 227 duplicates were removed, 273 records were excluded based on the abstract and 33 were excluded after reading the full article (Fig. 3) using the criteria described above. A summary and detailed information for qualifying studies are described in Table 1 and Supplementary Table 1, respectively. Six diseases included in the database search returned no studies that met the inclusion and exclusion criteria (Addison disease, myasthenia gravis, polymyalgia rheumatica, Sjögren syndrome, systemic vasculitis and uveitis).

Machine learning and artificial intelligence are most commonly applied to multiple sclerosis (MS), rheumatoid arthritis

(RA) and inflammatory bowel disease (IBD). MS, IBD and RA models used the most types of data, including 13 studies generating models using two data types (always including clinical data). Random forests and support vector machines were the most commonly used methods throughout diseases and applications. Clinical data were used in models for every type of autoimmune disease, and models using genetic data were created for the majority of disorders. The variety in methodological approaches, applications and data, as well as use of validation methods (Supplementary Table 1) renders meta-analysis of these methods inappropriate.

The applications for ML can be categorised into six broad topics: patient identification, risk prediction, diagnosis, disease subtype classification, disease progression and outcome and monitoring and management.

### Identification of patients
Studies utilised ML methods to identify patients with autoimmune diseases from electronic medical records,[19–25] and employed natural language processing. Gronsbell et al. worked to improve the efficiency of algorithms for this purpose.[26,27] These algorithms are intended to replace International Classification of Diseases billing codes, which have error rates of between 17.1–76.9% due to inconsistent terminology.[19] Electronic medical records also identified comorbidities associated with alopecia and vitiligo using natural language processing. This identified similar autoimmune comorbidities for both diseases.[28,29]

### Identifying and assessing autoimmune disease risk
Prediction of disease risk[30–39] and identification of novel risk factors through feature selection[40–44] was documented for IBD, type 1 diabetes (T1D), RA, systemic lupus erythematosus (SLE) and MS. Fifteen studies employed genetic data, using either sequencing arrays (GWAS) or exome data (nine studies), individual SNPs[38] within in the HLA regions[37,45] or pre-selected genes,[41] or gene expression data.[30,43] Only one study employed clinical data,[31] and two others combined clinical and genomic data.[30,45] Popular models included random forest, support vector machine and logistic regression.

### Diagnosis
Patient diagnosis was the most frequent ML application, and this approach was used for all autoimmune diseases. Distinguishing cases from healthy controls was an aim for 27 studies. Diagnostic classification models used patients with other autoimmune diseases as controls,[46–49] differentiated between diseases with overlapping or similar symptoms or phenotype, for example stratifying coeliac disease and irritable bowel syndrome,[50–56] or examined classification of multiple autoimmune diseases.[57,58] ML specifically for early diagnosis was specified by seven studies for the later onset degenerative conditions MS and RA.[48,59–64] Other diagnostic applications included distinguishing coeliac disease from an at-risk group[65,66] and differentiating those who have complications in T1D.[67,68] Random forests and support vector machine most frequently utilised.

### Classifying disease subtypes
Disease subtypes in one RA, two IBD, and six MS studies were classified by ML. Three types of unsupervised clustering were used by these studies: hierarchical clustering for identifying novel IBD subtypes;[69] consensus clustering to identify high, low and mixed levels of inflammation in RA;[70] and agglomerative hierarchical clustering to cluster MS by genetic signature.[71] Two of these studies employed support vector machine,[69,70] which is a popular supervised method in general, as well as random forest. There was wide variation in data types used. These
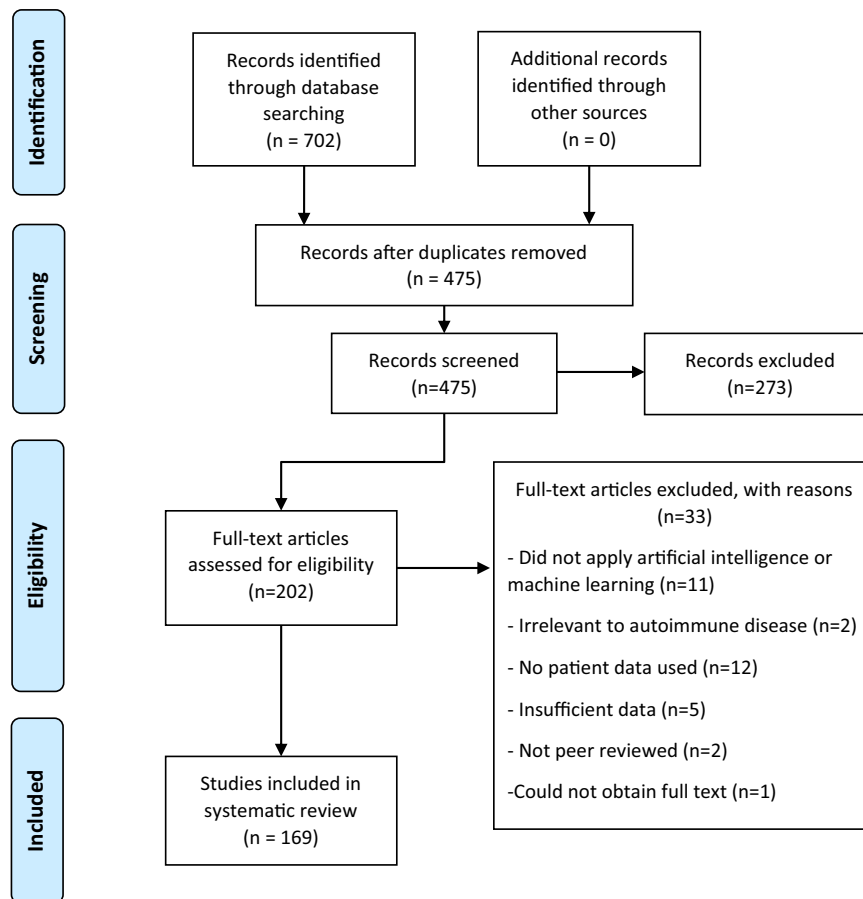
**Fig. 3  Methodological flowchart and number of papers reviewed at each stage.** The inclusion and exclusion criteria are applied to the title and abstract at the screening step and to the full article at the eligibility step. During the screening step, it was unclear from some abstracts if the article fulfilled the criteria, and therefore a full read is completed at the eligibility step to clarify the status of those records. Two reviewers completed screening independently, and where consensus could not be reached, a third reviewer assessed these articles and decided whether they were included or excluded.

included clinical (in particular MRI), genetic, RNA sequencing and gene expression data.

### Disease progression and outcome

Disease progression and outcome was a focus for 27 studies. Other considered issues were disease severity[72–78] in psoriasis, RA, IBD and coeliac disease; treatment response[79–87] in IBD, RA and primary biliary cirrhosis (PBC); and survival prediction[88–90] in PBC, RA and SLE. Other models focused on improved image segmentation to aid prognoses[91–96] for IBD and MS. Disease progression and outcome was the second-most prevalent area for model development. Throughout, the most common models were support vector machines, random forest and neural networks. The majority of data used was clinical, with very few papers utilising 'omic data.[86,97–99]

### Monitoring and management

Ten different studies of type 1 diabetes (T1D) used ML for monitoring and management: four predicted blood glucose level, four identified or predicted hypoglycaemic events, and two supported decision making using case-based reasoning or decision support systems. The majority of models used clinical data. Three models were developed using activity measurements for monitoring movement in MS, and one in RA. Support vector regression was used most frequently.[100–104]

### DISCUSSION

#### Validation and independent testing

Eighteen studies only used hold-out validation, not including studies with random forest models, where cross validation is unnecessary, or neural networks, where this process can be too computationally intensive. Eleven studies did not use any validation method, and so model integrity and applicability is unconfirmed. Methods that use hold-out validation have the potential to provide useful information, but it is accepted that unless the dataset is very large, these models are not as robustly validated as those that have used k-fold or leave-one-out cross validation, or a combination of cross-validation and testing on an independent dataset.

Only 14 of 169 studies combined cross-validation with independent test data for evaluating their models. These studies did not have any model types or applications in common. Clinical and genomic data were most common inputs for these studies. Models that used cross-validation and independent test data were applied to a number of the autoimmune diseases.

The research reviewed here demonstrates that, much like the disease studied, the ML models and methods used are heterogeneous. It can be difficult then, to determine which methods should be taken forward to clinical application. Alternatively, models from existing studies could be combined. Models have utilised different types of 'omic data, including proteomic, metagenomic and exome data. More popular has been sequencing array (SNP/GWAS) data, particularly when predicting

**Table 1.** Machine learning and artificial intelligence applications to autoimmune diseases.

| Disease | Number of studies | Years | Most popular classification/prediction application(s) | Most popular machine learning method(s) | Median sample size (min, max) | Data types used |
|---|---|---|---|---|---|---|
| Multiple sclerosis | 41[30,45,50,51,60,61,71,91–93,100,101,111,117–144] | 2008–2019 | Diagnosis, Prognosis, Disease Subtype | Type of Regression, Random Forest, Support Vector Machine | 99 (12, 12566) | Clinical, Survey, Genetic, MRI, Lipid Markers, SNPs, Gait Data, Immune repertoire, Gene Expression |
| Rheumatoid arthritis | 32[20–22,26,27,31,32,40–42,46–48,52,59,62–64,70,72,80–82,88,97,145–151] | 2003–2018 | Risk, Diagnosis, Early Diagnosis, Identify Patients | Support Vector Machine, Variations of Random Forest, Neural Network and Decision Tree | 338 (22, 922199) | Medical Database, Immunoassay, Metagenomic, Microbiome, GWAS/SNP, Clinical, Movement Data, Amino acid analytes, Transcriptomic, EMRs, Ultrasound images, Proteomic, Laser images |
| Inflammatory bowel disease | 30[33–36,43,57,69,73,79,83–86,94,95,98,152–165] | 2007–2018 | Diagnosis, Response to Treatment, Disease Risk, Disease Severity | Random Forest, Support Vector Machine | 273 (50, 53279) | Clinical, Colonoscopy Images, Metagenomic, Gene Expression, GWAS, Microbiota, miRNA Expression, EMRs, Exome, MRI |
| Type 1 diabetes | 17[37–39,67,68,102–104,166–174] | 2009–2018 | Disease Management | Novel Methods/Hybrid Models, Neural Network, Support Vector Regression | 23 (10, 10579) | Clinical, Red Blood Cell Images, VOCs, GWAS/SNPs |
| Systemic lupus erythematosus | 14[19,23,44,49,89,96,175–182] | 2009–2018 | Variations of prognosis, Diagnosis | Logistic Regression, Neural Network, Random Forest Decision Tree | 318 (14, 17057) | Clinical, Electronic Health Records, Drug Treatment, SNPs, MRI, Exome, Gene Expression, Proteomic, Urine Biomarkers |
| Psoriasis | 11[53,74–77,99,112,183–186] | 2007–2018 | Diagnosis, Disease Severity | Support Vector Machine | 540 (80, 22181) | Digital Image, GWAS, Proteomic, RNA Biomarkers |
| Coeliac disease | 7[24,25,54,65,66,78,187] | 2011–2018 | Diagnosis | Random Forest, Logistic Regression, Bayesian Classifier, Support Vector Machine, Logistic Model, Natural Language Processing, Combined Fuzzy Cognitive Map and Possibilistic Fuzzy c-means clustering. | 465 (47, 1498) | VOCs, Clinical, Peptide, EMRs |
| Thyroid diseases | 6[188–193] | 2008–2018 | Diagnosis | Hybrid Models | 215 (215, 7200) | Clinical |
| Autoimmune liver diseases | 5[58,87,90,194,195] | 2009–2018 | Prognosis | Variations on Random Forest | 288 (64, 787) | Clinical, Clinical Trial, Microbiome |
| Systemic sclerosis | 4[55,113,196,197] | 2016–2018 | Diagnosis, Treatment, Prognosis | Support Vector Machine, Random Forest | 119 (37, 991) | Gene Expression, Nailfold capillaroscopy images, Peripheral Blood Mononuclear cell data (flow cytometry, DNA, mRNA) |

Information includes the number of studies per autoimmune disease, the years they occurred, popular applications and methods and data types used. Median sample size was a better representation than mean, due to large cohorts in studies using data from genome-wide association studies and electronic medical records.

*EMR* electronic medical record, *GWAS* genome-wide association study, *miRNA* micro RNA, *MRI* magnetic resonance imaging, *SNP* single nucleotide polymorphism, *VOC* volatile organic compound.

autoimmune disease risk. By far the most prevalent type of data is the use of clinical and laboratory data.

To optimise the use of these data types, accessibility is key, and EMRs allow easy extraction of these data. Some researchers have moved beyond only storing medical data in these systems. The eMERGE (electronic medical records and genomics) network combines the genomic and EMR repositories to further genomic medicine research.[105] Other studies such as SPOKE (Scalable Precision Medicine Oriented Knowledge Engine), wish to integrate these data within the storage platform, by building a knowledge network using unsupervised machine learning that informs on how data types such as GWAS, gene ontology, pathways and drug data are connected to EMRs.[106] Improving knowledge of how these data are related is a key step towards implementing precision medicine.

Many models were created for autoimmune disease diagnosis, more specifically classifying those with disease and controls. The majority achieved high classifier performance (where any combination of the following metrics are over these thresholds: accuracy > 81%, AUC > 0.95, Sensitivity > 82, Specificity > 84), and provided evidence of machine learning's utility in diagnostics.

Identifying the molecular diagnosis to inform tailored treatment strategies has revolutionised cancer prognoses, improving patient outcomes and quality of life, along with economic benefits to the treatment provider. Targeted therapies such as monoclonal antibodies and small molecule inhibitors transformed treatment of some cancers, or improved patient survival times.[107] Key to precision treatment has been the identification of the driver mutations specific to the cancer type.[108] Machine learning has been utilised for cancer classification[109,110] and discovery of relevant pathways.[109] Across the spectrum of autoimmune diseases, there has traditionally been a one-size-fits-all approach to patient therapeutics. The expectation is that machine learning represents a necessary key tool that will use 'big' data to stratify patients and move towards personalised treatment approaches that have proven so effective in cancer. Proof of this concept has already been demonstrated through machine learning to stratify patient's inflammation status in RA,[70] and further investigate IBD subtypes.[69]

Six models from the evaluated studies returned more than one of the following measures as either 1 or 100%: AUC, accuracy, precision and recall, sensitivity and specificity.[59,67,68,111–113] This perfect performance indicates that a model may not be required, as there exists data that classifies the groups without error. An alternative explanation of apparently optimal performance may reside in poor implementation of cross-validation strategies.

Common metrics reported are accuracy, AUC, and sensitivity and specificity. However, accuracy is inferior to AUC, particularly when imbalanced datasets are used.[114] The AUC measure is unaffected by imbalanced data, but precision-recall curves may reflect model performance more accurately.[115] Dataset rebalancing methods should potentially be utilised more for a thorough review of model performance.

When creating and evaluating a model, increasing focus could be placed on which measure is more important, sensitivity or specificity. Scully et al. demonstrated this, where a lesion segmentation model could achieve high specificity (99.9%) through labelling all tissue as non-lesion.[96]

An ML model by Ahmed et al.[62] provides evidence for using an additional independent test dataset subsequent to cross validation. In their study, the AUC dropped by 0.25, indicating decreased model performance on new data.

Studies included in this systematic review have shown that artificial intelligence and machine learning models provide useful insight, despite the heterogeneity of presentation, diagnosis, disease course and patient outcome. However, the heterogeneity in data used, models and model evaluation cause difficulties in obtaining consensus. Furthermore, the number of autoimmune

diseases this literature search focussed on was restricted, and may have resulted in an incomplete picture of ML applied to autoimmune diseases.

From this analysis, it seems appropriate to advocate for standardised methods of model evaluation, by utilising a combination of cross validation and independent test data for model validation. Increased confidence in model results allows for more complex model creation, by layering data types or combining classifiers. These models could be applied to more difficult tasks that reflect the complexity of autoimmune disease. With these advances, AI and ML have the potential to bring personalised medicine closer for patients with complex and chronic disease.

## METHODS

### Autoimmune disease selection

Autoimmune diseases selected for the systematic review are based on prevalence estimates[4] and include Addison disease, alopecia, Coeliac disease, Crohn's disease, ulcerative colitis, type 1 diabetes, autoimmune liver diseases, hyper- and hypo-thyroidism, multiple sclerosis, myasthenia gravis, polymyalgia rheumatica, psoriasis, psoriatic arthritis, rheumatoid arthritis, Sjögren syndrome, systemic sclerosis, systemic lupus erythematosus, systemic vasculitis, uveitis and vitiligo.

**Table 2.** Search terms used in OvidSP and EBSCO for each autoimmune disease.

| Autoimmune disease | Disease Search Term(s) Used |
| --- | --- |
| Addison's disease | Addison* |
| Alopecia | Alopecia |
| Celiac disease | Celiac, Coeliac |
| Inflammatory bowel disease | Inflammatory bowel disease, Crohn* disease, ulcerative colitis |
| Type 1 diabetes | Type 1 Diabetes, Insulin dependent Diabetes? |
| Autoimmune hepatitis | Autoimmune hepatitis, chronic active hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis |
| Thyroid disease | Autoimmune thyroiditis, Hashimoto* thyroiditis, Hashimoto* disease, Grave* disease, hyperthyroid*, hypothyroid* |
| Multiple sclerosis | Multiple sclerosis |
| Myasthenia gravis | Myasthenia gravis |
| Polymyalgia rheumatica | Polymyalgia rheumatica |
| Psoriasis | Psoriasis |
| Psoriatic arthritis | Psoriatic arthritis |
| Rheumatoid arthritis | Rheumatoid Arthritis |
| Sjögren syndrome | Sjogren syndrome |
| Systemic sclerosis | Systemic sclerosis |
| Systemic lupus erythematosus | Lupus |
| Systemic vasculitis | Polyarteritis nodosa, microscopic polyangiitis, granulomatosis with polyangiitis, eosinophilic granulomatosis with polyangiitis. |
| Uveitis (iridocyclitis) | Uvetitis, iridocyclitis |
| Vitiligo | Vitiligo |

Asterisk (*) and question mark (?) are wildcard characters used for searching the databases OvidSP and EBSCO.

## Literature search

The literature search was performed electronically with OvidSP using MEDLINE from 1946, and EMBASE from 1974. A search was also performed on the Computers & Applied Sciences Complete database available on EBSCO. The literature search was completed in December 2018. All searches conformed to the same structure: the words "machine learning" or "artificial intelligence" combined with the chosen search term(s) for each autoimmune disease (see Table 2). Boolean operators OR and AND (for combining search terms) were used in order to streamline the procedure. In both databases, the title, abstract and subject terms/keyword headings assigned by authors were searched (last search 17/12/2018).

## Inclusion and exclusion criteria

Studies that applied ML methods to autoimmune diseases listed above, or to complications that arise from autoimmune diseases were included. Studies not written in English, published prior to 2001, that did not use real human patient data, were not peer reviewed, or were review papers were also excluded. This systematic review conforms to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards.[116]

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## DATA AVAILABILITY

The data (papers) that support the findings of this study are available publicly. Full list of records identified through database searching are available on reasonable request from the authors.

## REFERENCES

1. Goodnow, C. C., Sprent, J., de St Groth, B. F. & Vinuesa, C. G. Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* **435**, 590–597 (2005).
2. Kuchroo, V. K., Ohashi, P. S., Sartor, R. B. & Vinuesa, C. G. Dysregulation of immune homeostasis in autoimmune diseases. *Nat. Med.* **18**, 42–47 (2012).
3. Male, D. K., Roitt, I. M., Roth, D. B., Roitt, I. M. *Immunology*. 8th edn. (Saunders, 2013).
4. Cooper, G. S., Bynum, M. L. & Somers, E. C. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. *J. Autoimmun.* **33**, 197–207 (2009).
5. Hayter, S. M. & Cook, M. C. Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. *Autoimmun. Rev.* **11**, 754–765 (2012).
6. Eaton, W. W., Rose, N. R., Kalaydjian, A., Pedersen, M. G. & Mortensen, P. B. Epidemiology of autoimmune diseases in Denmark. *J. Autoimmun.* **29**, 1–9 (2007).
7. Cho, J. H. & Feldman, M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nat. Med.* **21**, 730 (2015).
8. Simon, T. A. et al. Prevalence of co-existing autoimmune disease in rheumatoid arthritis: a cross-sectional study. *Adv. Ther.* **34**, 2481–2490 (2017).
9. Gilhus, N. E., Nacu, A., Andersen, J. B. & Owe, J. F. Myasthenia gravis and risks for comorbidity. *Eur. J. Neurol.* **22**, 17–23 (2015).
10. Ruggeri, R. M. et al. Autoimmune comorbidities in Hashimoto's thyroiditis: different patterns of association in adulthood and childhood/adolescence. *Eur. J. Endocrinol.* **176**, 133 (2017).
11. Gill, L. et al. Comorbid autoimmune diseases in patients with vitiligo: a cross-sectional study. *J. Am. Acad. Dermatol.* **74**, 295–302 (2016).
12. Teschendorff, A. E. Avoiding common pitfalls in machine learning omic data science. *Nat. Mater.* **18**, 422–427 (2019).
13. Jiang, F. et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. Neurol.* **2**, 230–243 (2017).
14. Kersting, K. Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. *Front. Big Data.* **1**, 6 (2018). https://doi.org/10.3389/fdata.2018.00006.
15. Fatima, M. & Pasha, M. Survey of machine learning algorithms for disease diagnostic. *J. Intell. Learn. Syst. Appl.* **09**, 1–16 (2017).
16. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning with Applications in R. 1 ed, Vol. XIV*, (426. Springer-Verlag, New York, 2013).
17. Hastie, T., Tibshirani, R. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction.* 2nd ed. ed. (Springer, New York, 2009).
18. Figueroa, R. L., Zeng-Treitler, Q., Kandula, S. & Ngo, L. H. Predicting sample size required for classification performance. *BMC Med. Inf. Decis. Mak.* **12**, 8 (2012).
19. Turner, C. A. et al. Word2Vec inversion and traditional text classifiers for phenotyping lupus. *BMC Med. Inform. Decis. Mak.* **17**, 126 (2017).
20. Zhou, S. M. 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 [Electron. Resour.].* **11**, e0154515 (2016).
21. Lin, C. et al. Automatic identification of methotrexate-induced liver toxicity in patients with rheumatoid arthritis from the electronic medical record. *J. Am. Med. Inf. Assoc.* **22**, e151–e161 (2015).
22. Chen, Y. et al. Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *J. Am. Med. Inf. Assoc.* **20**, e253–e259 (2013).
23. Murray, S. G., Avati, A., Schmajuk, G. & Yazdany, J. Automated and flexible identification of complex disease: building a model for systemic lupus erythematosus using noisy labeling. *J. Am. Med. Inf. Assoc.* **26**, 61–65 (2018).
24. Chen, W., Huang, Y., Boyle, B. & Lin, S. The utility of including pathology reports in improving the computational identification of patients. *J. Pathol. Inform.* **7**, 46 (2016).
25. Ludvigsson, J. F. et al. Use of computerized algorithm to identify individuals in need of testing for celiac disease. *J. Am. Med. Inf. Assoc.* **20**, e306–e310 (2013).
26. Gronsbell, J., Minnier, J., Yu, S., Liao, K., Cai, T. Automated feature selection of predictors in electronic medical records data. *Biometrics*. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.12987 (2018).
27. Gronsbell, J. L. & Cai, T. Semi-supervised approaches to efficient evaluation of model prediction performance. *J. R. Stat. Soc. Ser. B (Stat. Methodol.).* **80**, 579–594 (2018).
28. Huang, K. P., Mullangi, S., Guo, Y. & Qureshi, A. A. Autoimmune, atopic, and mental health comorbid conditions associated with alopecia areata in the United States.[Erratum appears in JAMA Dermatol. 2014 Jun;150(6):674]. *JAMA Dermatol.* **149**, 789–794 (2013).
29. Sheth, V. M., Guo, Y. & Qureshi, A. A. Comorbidities associated with vitiligo: a ten-year retrospective study. *Dermatology* **227**, 311–315 (2013).
30. Corvol, J. C. et al. Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. *Proc. Natl Acad. Sci. USA* **105**, 11839–11844 (2008).
31. Chin, C. Y., Hsieh, S. Y., Tseng, V. S. EDram: Effective early disease risk assessment with matrix factorization on a large-scale medical database: a case study on rheumatoid arthritis. *PLoS ONE* **13**, e0207579 (2018).
32. Liu, C., Ackerman, H. H. & Carulli, J. P. A genome-wide screen of gene-gene interactions for rheumatoid arthritis susceptibility. *Hum. Genet.* **129**, 473–485 (2011).
33. Wei, Z. et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. *Am. J. Hum. Genet.* **92**, 1008–1012 (2013).
34. Daneshjou, R. et al. Working toward precision medicine: predicting phenotypes from exomes in the Critical Assessment of Genome Interpretation (CAGI) challenges. *Hum. Mutat.* **38**, 1182–1192 (2017).
35. Giollo, M. et al. Crohn disease risk prediction-Best practices and pitfalls with exome data. *Hum. Mutat.* **38**, 1193–1200 (2017).
36. Pal, L. R., Kundu, K., Yin, Y. & Moult, J. CAGI4 Crohn's exome challenge: marker SNP versus exome variant models for assigning risk of Crohn disease. *Hum. Mutat.* **38**, 1225–1234 (2017).
37. Zhao, L. P., Bolouri, H., Zhao, M., Geraghty, D. E. & Lernmark, A. An object-oriented regression for building disease predictive models with multiallelic HLA genes. *Genet. Epidemiol.* **40**, 315–332 (2016).
38. Nguyen, C., Varney, M. D., Harrison, L. C. & Morahan, G. Definition of high-risk type 1 diabetes HLA-DR and HLA-DQ types using only three single nucleotide polymorphisms. *Diabetes* **62**, 2135–2140 (2013).
39. Wei, Z. et al. From disease association to risk assessment: an optimistic view from genome-wide association studies on type 1 diabetes. *PLoS Genet.* **5** (2009).
40. Negi, S. et al. A genome-wide association study reveals ARL15, a novel non-HLA susceptibility gene for rheumatoid arthritis in North Indians. *Arthritis Rheumatism* **65**, 3026–3035 (2013).

41. Briggs, F. B. S. et al. Supervised machine learning and logistic regression identifies novel epistatic risk factors with PTPN22 for rheumatoid arthritis. *Genes Immun.* **11**, 199–208 (2010).

42. Gonzalez-Recio, O., de Maturana, E. L., Vega, A. T., Engelman, C. D. & Broman, K. W. Detecting single-nucleotide polymorphism by single-nucleotide polymorphism interactions in rheumatoid arthritis using a two-step approach with machine learning and a Bayesian threshold least absolute shrinkage and selection operator (LASSO) model. *BMC Proc.* **3**, S63 (2009).

43. Isakov, O., Dotan, I. & Ben-Shachar, S. Machine learning-based gene prioritization identifies novel candidate risk genes for inflammatory bowel disease. *Inflamm. Bowel Dis.* **23**, 1516–1523 (2017).

44. Davis, N. A. et al. Encore: genetic association interaction network centrality pipeline and application to SLE exome data. *Genet. Epidemiol.* **37**, 614–621 (2013).

45. Mowry, E. M. et al. Incorporating machine learning approaches to assess putative environmental risk factors for multiple sclerosis. *Mult. Scler. Relat. Disord.* **24**, 135–141 (2018).

46. Niu, Q. et al. Specific serum protein biomarkers of rheumatoid arthritis detected by MALDI-TOF-MS combined with magnetic beads. *Int. Immunol.* **22**, 611–618 (2010).

47. Geurts, P. et al. Proteomic mass spectra classification using decision tree based ensemble methods. *Bioinformatics* **21**, 3138–3145 (2005).

48. De Seny, D. et al. Discovery of new rheumatoid arthritis biomarkers using the surface-enhanced laser desorption/ionization time-of-flight mass spectrometry proteinchip approach. *Arthritis Rheumatism* **52**, 3801–3812 (2005).

49. Huang, Z. 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* **48**, 626–631 (2009).

50. Alaqtash, M. et al. Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms. *Conf. Proc. IEEE Eng. Med Biol. Soc.* **2011**, 453–457 (2011).

51. Ohanian, D. et al. Identifying key symptoms differentiating myalgic encephalomyelitis and chronic fatigue syndrome from multiple sclerosis. *Neurol. (E-Cronico.).* **4**, 41–45 (2016).

52. Singh, S., Kumar, A., Panneerselvam, K. & Vennila, J. J. Diagnosis of arthritis through fuzzy inference system. *J. Med. Syst.* **36**, 1459–1468 (2012).

53. Cowen, E. W. et al. Differentiation of tumour-stage mycosis fungoides, psoriasis vulgaris and normal controls in a pilot study using serum proteomic analysis. *Br. J. Dermatol.* **157**, 946–953 (2007).

54. Arasaradnam, R. P. et al. Differentiating coeliac disease from irritable bowel syndrome by urinary volatile organic compound analysis—A pilot study. *PLoS ONE.* **9**, e107312 (2014).

55. Berks, M. et al. An automated system for detecting and measuring nailfold capillaries. Medical image computing and computer-assisted intervention: MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention. 17, 658–665 (2014).

56. Armananzas, R. et al. Microarray analysis of autoimmune diseases by machine learning procedures. IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology. *Society* **13**, 341–350 (2009).

57. Forbes, J. D. et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? *Microbiome* **6**, 221 (2018).

58. Iwasawa, K. et al. Dysbiosis of the salivary microbiota in pediatric-onset primary sclerosing cholangitis and its potential as a biomarker. *Sci. Rep.* **8**, 5480 (2018).

59. Heard, B. J. et al. A computational method to differentiate normal individuals, osteoarthritis and rheumatoid arthritis patients using serum biomarkers. *J. R. Soc. Interface* **11**, 20140428 (2014).

60. Saccà, V. et al. Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data. *Brain Imaging Behav.* https://link.springer.com/article/10.1007%2Fs11682-018-9926-9 (2018).

61. Yoo, Y. et al. Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *NeuroImage: Clin.* **17**, 169–178 (2018).

62. Ahmed, U., Anwar, A., Savage, R. S., Thornalley, P. J. & Rabbani, N. Protein oxidation, nitration and glycation biomarkers for early-stage diagnosis of osteoarthritis of the knee and typing and progression of arthritic disease. *Arthritis Res. Ther.* **18**, 250 (2016).

63. Scheel, A. K. et al. Laser imaging techniques for follow-up analysis of joint inflammation in patients with rheumatoid arthritis. *Med. Laser Appl.* **18**, 198–205 (2003).

64. Wyns, B. et al. Prediction of arthritis using a modified Kohonen mapping and case based reasoning. *Eng. Appl. Artif. Intell.* **17**, 205 (2004).

65. Hujoel, I. A. et al. Machine learning in detection of undiagnosed celiac disease. *Clin. Gastroenterol. Hepatol.* **16**, 1354–1355.e1351 (2018).

66. Tenorio, J. M. et al. Artificial intelligence techniques applied to the development of a decision-support system for diagnosing celiac disease. *Int. J. Med. Inform.* **80**, 793–802 (2011).

67. Maulucci, G. et al. Phase separation of the plasma membrane in human red blood cells as a potential tool for diagnosis and progression monitoring of type 1 diabetes mellitus. *PLoS ONE* **12**, (2017).

68. Cordelli, E. et al. A decision support system for type 1 diabetes mellitus diagnostics based on dual channel analysis of red blood cell membrane fluidity. *Comput. Methods Prog. Biomed.* **162**, 263–271 (2018).

69. Mossotto, E. et al. Classification of paediatric inflammatory bowel disease using machine learning. *Sci. Rep.* **7**, 2427 (2017).

70. Orange, D. E. et al. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatol.* **70**, 690–701 (2018).

71. Lopez, C., Tucker, S., Salameh, T. & Tucker, C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. *J. Biomed. Inform.* **85**, 30–39 (2018).

72. Lin, C. et al. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. *PLoS ONE* **8**, (2013).

73. Niehaus, K. E., Uhlig H. H., Clifton D. A. Phenotypic characterisation of Crohn's disease severity. Conference proceedings:. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference. 2015, 7023–7026 (2015).

74. George, Y., Aldeen, M. & Garnavi, R. Psoriasis image representation using patch-based dictionary learning for erythema severity scoring. *Comput. Med. Imaging Graph.* **66**, 44–55 (2018).

75. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. A novel and robust Bayesian approach for segmentation of psoriasis lesions and its risk stratification. *Comput. Methods Prog. Biomed.* **150**, 9–22 (2017).

76. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. A novel approach to multiclass psoriasis disease risk stratification: Machine learning paradigm. *Biomed. Signal Process. Control.* **28**, 27–40 (2016).

77. Raina, A. et al. Objective measurement of erythema in psoriasis using digital color photography with color calibration. *Ski. Res. Technol.* **22**, 375–380 (2016).

78. Amirkhani, A., Mosavi, M. R., Mohammadi, K. & Papageorgiou, E. I. A novel hybrid method based on fuzzy cognitive maps and fuzzy clustering algorithms for grading celiac disease. *Neural Comput. Appl.* **30**, 1573–1588 (2018).

79. Waljee, A. K. et al. Predicting corticosteroid-free endoscopic remission with vedolizumab in ulcerative colitis. *Alimentary Pharmacol. Thera.* **47**, 763–772 (2018).

80. Miyoshi, F. et al. A novel method predicting clinical response using only background clinical data in RA patients before treatment with infliximab. *Mod. Rheumatol.* **26**, 813–816 (2016).

81. Nair, S. S., French, R. M., Laroche, D. & Thomas, E. The application of machine learning algorithms to the analysis of electromyographic patterns from arthritic patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* **18**, 174–184 (2010).

82. Van Looy, S. et al. Prediction of dose escalation for rheumatoid arthritis patients under infliximab treatment. *Eng. Appl. Artif. Intell.* **19**, 819–828 (2006).

83. Waljee, A. K. et al. Machine learning algorithms for objective remission and clinical outcomes with thiopurines. *J. Crohn's Colitis* **11**, 801–810 (2017).

84. Kang, T., Ding, W., Zhang, L., Ziemek, D. & Zarringhalam, K. A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data. *BMC Bioinforma.* **18**, 565 (2017).

85. Waljee, A. K. et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. *Clin. Gastroenterol. Hepatol.* **8**, 143–150 (2010).

86. Doherty, M. K. et al. Fecal microbiota signatures are associated with response to ustekinumab therapy among Crohn's disease patients. *mBio* **9**, e02120-02117 (2018).

87. Weiss, J., Kuusisto, F., Boyd, K., Liu, J. & Page, D. Machine learning for treatment assignment: improving individualized risk attribution. *AMIA Annu. Symp. Proc. AMIA Symp.* **2015**, 1306–1315 (2015).

88. Lezcano-Valverde, J. M. et al. Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach. *Sci. Rep.* **7**, 10189 (2017).

89. Tang, H. et al. Predicting three-year kidney graft survival in recipients with systemic lupus erythematosus. *ASAIO J. (Am. Soc. Artif. Intern. Organs: 1992)* **57**, 300–309 (2011).

90. Tsujitani, M. & Sakon, M. Analysis of survival data having time-dependent covariates. *IEEE Trans. Neural Netw.* **20**, 389–394 (2009).

91. Sweeney, E. M. et al. A comparison of supervised machine learning algorithms and feature vectors for MS lesion segmentation using multimodal structural MRI. *PLOS ONE.* **9**, e95753 (2014).

92. Commowick, O. et al. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci. Rep.* **8**, 13650 (2018).

93. Cabezas, M. et al. BOOST: a supervised approach for multiple sclerosis lesion segmentation. *J. Neurosci. Methods* **237**, 108–117 (2014).

94. Mahapatra, D., Vos, F. M. & Buhmann, J. M. Active learning based segmentation of Crohns disease from abdominal MRI. *Comput Methods Prog. Biomed.* **128**, 75–85 (2016).

95. Mahapatra, D. Combining multiple expert annotations using semi-supervised learning and graph cuts for medical image segmentation. *Computer Vis. Image Underst.* **151**, 114–123 (2016).

96. Scully, M. et al. An Automated Method for Segmenting White Matter Lesions through Multi-Level Morphometric Feature Classification with Application to Lupus. *Front Hum. Neurosci.* **4**, 27 (2010).

97. Joo, Y. B. et al. Biological function integrated prediction of severe radiographic progression in rheumatoid arthritis: A nested case control study. Arthritis and Rheumatology Conference: American College of Rheumatology/Association of Rheumatology Health Professionals Annual Scientific Meeting, ACR/ARHP. 19, 244 (2017).

98. Douglas, G. M. et al. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. *Microbiome.* **6**, 13 (2018).

99. Patrick, M. T. et al. Genetic signature to provide robust risk assessment of psoriatic arthritis development in psoriasis patients. *Nat. Commun.* **9**, 4178 (2018). https://doi.org/10.1038/s41467-018-06672-6.

100. Supratak, A. et al. Remote monitoring in the home validates clinical gait measures for multiple sclerosis. *Front. Neurol.* **9**, 561 (2018). https://doi.org/10.3389/fneur.2018.00561.

101. McGinnis, R. S. et al. A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis. *PLoS ONE* **12**, e0178366 (2017).

102. Georga, E. I., Protopappas, V. C., Ardigo, D., Polyzos, D. & Fotiadis, D. I. A Glucose Model Based on Support Vector Regression for the Prediction of Hypoglycemic Events Under Free-Living Conditions. *Diabetes Technol. Ther.* **15**, 634–643 (2013).

103. Marling, C. R., Struble, N. W., Bunescu, R. C., Shubrook, J. H. & Schwartz, F. L. A consensus-perceived glycemic variability metric. *J. Diabetes Sci. Technol.* **7**, 871–879 (2013).

104. Georga, E. I., Protopappas, V. C., Polyzos, D. & Fotiadis, D. I. Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models. *Med Biol. Eng. Comput.* **53**, 1305–1318 (2015).

105. Gottesman, O. et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet. Med.* **15**, 761–771 (2013).

106. Nelson, C. A., Butte, A. J. & Baranzini, S. E. Integrating biomedical research and electronic health records to create knowledge-based biologically meaningful machine-readable embeddings. *Nat. Commun.* **10**, 3045 (2019).

107. Gerber, D. E. Targeted therapies: a new generation of cancer treatments. *Am. Fam. Physician* **77**, 311–319 (2008).

108. Raphael, B. J., Dobson, J. R., Oesper, L. & Vandin, F. Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. *Genome Med.* **6**, 5 (2014).

109. Zeng, Z. et al. Cancer classification and pathway discovery using non-negative matrix factorization. *J. Biomed. Inform.* **96**, 103247 (2019).

110. Zhang, X., Guan, N., Jia, Z., Qiu, X. & Luo, Z. Semi-supervised projective non-negative matrix factorization for cancer classification. *PLoS ONE* **10**, e0138814 (2015).

111. Lotsch, J. et al. Machine-learning based lipid mediator serum concentration patterns allow identification of multiple sclerosis patients with high accuracy. *Sci. Rep.* **8**, 14884 (2018).

112. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind. *Comput Methods Prog. Biomed.* **126**, 98–109 (2016).

113. Zhu, H. et al. Integration of Genome-Wide DNA Methylation and Transcription Uncovered Aberrant Methylation-Regulated Genes and Pathways in the Peripheral Blood Mononuclear Cells of Systemic Sclerosis. *Int. J. Rheumatol.* **2018**, https://doi.org/10.1155/2018/7342472 (2018).

114. Huang, J. & Ling, C. X. Using AUC and accuracy in evaluating learning algorithms. *Ieee T Knowl. Data En.* **17**, 299–310 (2005).

115. Jeni, L. A., Cohn, J. F., Torre, FDL, (eds) Facing Imbalanced Data–Recommendations for the Use of Performance Metrics. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; 2013 2-5 Sept. 2013.

116. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. The PG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Med.* **6**, e1000097 (2009).

117. Zhang, Y. et al. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine. *Simulation* **92**, 861–871 (2016).

118. Zhao, Y. et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLOS ONE.* **12**, e0174866 (2017).

119. Briggs, F. B. S. et al. Multiple sclerosis risk factors contribute to onset heterogeneity. *Mult. Scler. Relat. Disord.* **28**, 11–16 (2019).

120. Ahmadi, A., Davoudi, S. & Daliri, M. R. Computer aided diagnosis system for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention. *Comput Methods Prog. Biomed.* **169**, 9–18 (2019).

121. Zhang, H. et al. Predicting conversion from clinically isolated syndrome to multiple sclerosis–An imaging-based machine learning approach. *NeuroImage: Clin.* **21**, 101593 (2019).

122. Zurita, M. et al. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *NeuroImage: Clin.* **20**, 724–730 (2018).

123. Wang, S.-H. et al. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Front. Neuroscience.* **12**, 818 (2018). https://doi.org/10.3389/fnins.2018.00818.

124. Neeb, H. & Schenk, J. Multivariate prediction of multiple sclerosis using robust quantitative MR-based image metrics. *Zeitschrift für Medizinische Physik.* https://www.sciencedirect.com/science/article/pii/S0939388918300680?via%3Dihub (2018).

125. Tacchella, A. et al. Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: a proof-of-principle study. *F1000Research.* **6**, 2172 (2017).

126. Kiiski, H. et al. Machine learning EEG to predict cognitive functioning and processing speed over a 2-year Period in multiple sclerosis patients and controls. *Brain Topogr.* **31**, 346–363 (2018).

127. Fiorini, S. et al. A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes. *Conf. Proc. IEEE Eng. Med Biol. Soc.* **2015**, 4443–4446 (2015).

128. Zhong, J. et al. Combined structural and functional patterns discriminating upper limb motor disability in multiple sclerosis using multivariate approaches. *Brain Imaging Behav.* **11**, 754–768 (2017).

129. Lötsch, J. et al. Machine-learned data structures of lipid marker serum concentrations in multiple sclerosis patients differ from those in healthy subjects. *Int. J. Mol. Sci.* **18**, 1217 (2017).

130. Karaca, Y., Zhang, Y. D., Cattani, C. & Ayan, U. The differential diagnosis of multiple sclerosis using convex combination of infinite kernels. *CNS Neurol. Disord. Drug Targets* **16**, 36–43 (2017).

131. Ostmeyer, J. et al. Statistical classifiers for diagnosing disease from immune repertoires: a case study using multiple sclerosis. *BMC Bioinforma.* **18**, 401 (2017).

132. Ion-Mărgineanu, A. et al. Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features. *Front. Neurosci.* **11**, 398 (2017). https://doi.org/10.3389/fnins.2017.00398.

133. Kocevar, G. et al. Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Front. Neurosci.* **10**, 478 (2016). https://doi.org/10.3389/fnins.2016.00478.

134. Kosa, P. et al. Development of a sensitive outcome for economical drug screening for progressive multiple sclerosis treatment. *Front. Neurol.* **7**, 131 (2016). https://doi.org/10.3389/fneur.2016.00131.

135. Baranzini, S. E. et al. Prognostic biomarkers of IFNb therapy in multiple sclerosis patients. *Mult. Scler. (Houndmills, Basingstoke, Engl.).* **21**, 894–904 (2015).

136. Wottschel, V. et al. Predicting outcome in clinically isolated syndrome using machine learning. *Neuroimage Clin.* **7**, 281–287 (2015).

137. Crimi, A. et al. Predictive value of imaging markers at multiple sclerosis disease onset based on gadolinium- and USPIO-enhanced MRI and machine learning. *PLoS ONE* **9**, e93024 (2014).

138. Taschler, B. et al. editors. Spatial modeling of multiple sclerosis for disease subtype prediction. (Springer International Publishing, 2014).

139. Goldstein, B. A., Hubbard, A. E., Cutler, A. & Barcellos, L. F. An application of Random Forests to a genome-wide association dataset: methodological considerations & new findings. *BMC Genet.* **11**, 49 (2010).

140. Briggs, F. B. et al. Evidence for CRHR1 in multiple sclerosis using supervised machine learning and meta-analysis in 12,566 individuals. *Hum. Mol. Genet.* **19**, 4286–4295 (2010).

141. Salem, M. et al. A supervised framework with intensity subtraction and deformation field features for the detection of new T2-w lesions in multiple sclerosis. *NeuroImage: Clin.* **17**, 607–615 (2018).

142. Birenbaum, A. & Greenspan, H. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. *Eng. Appl. Artif. Intell.* **65**, 111–118 (2017).

143. Morrison, C. et al. Assessing multiple sclerosis with kinect: designing computer vision systems for real-world use. *Hum.-Computer Interact.* **31**, 191–226 (2016).

144. Liu, J., Brodley, C. E., Healy, B. C. & Chitnis, T. Removing confounding factors via constraint-based clustering: An application to finding homogeneous groups of multiple sclerosis patients. *Artif. Intell. Med.* **65**, 79–88 (2015).

145. Chocholova, E. et al. Glycomics meets artificial intelligence - Potential of glycan analysis for identification of seropositive and seronegative rheumatoid arthritis patients revealed. *Clin. Chim. Acta* **481**, 49–55 (2018).

146. Wu, H. et al. Metagenomics biomarkers selected for prediction of three different diseases in Chinese population. *BioMed Res. Int.* (2018) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5820663/.

147. Andreu-Perez, J. et al. Developing fine-grained actigraphies for rheumatoid arthritis patients from a single accelerometer using machine learning. *Sensors* **17**, 2113 (2017).

148. Yeo, L. et al. Expression of chemokines CXCL4 and CXCL7 by synovial macrophages defines an early stage of rheumatoid arthritis. *Ann. Rheum. Dis.* **75**, 763–771 (2016).

149. Pratt, A. G. et al. A CD4 T cell gene signature for early rheumatoid arthritis implicates interleukin 6-mediated STAT3 signalling, particularly in anti-citrullinated peptide antibody-negative disease. *Ann. Rheum. Dis.* **71**, 1374–1381 (2012).

150. Kruppa, J., Ziegler, A. & Konig, I. R. Risk estimation and risk prediction using machine-learning methods. *Hum. Genet.* **131**, 1639–1654 (2012).

151. Gossec, L. et al. Detection of flares by decrease in physical activity, collected using wearable activity trackers, in rheumatoid arthritis or axial spondyloarthritis: an application of Machine-Learning analyses in rheumatology. *Arthritis Care Res (Hoboken).* **22**, 22 (2018).

152. Waljee, A. K. et al. Predicting hospitalization and outpatient corticosteroid use in inflammatory bowel disease patients using machine learning. *Inflamm. Bowel Dis.* **24**, 45–53 (2018).

153. Maeda, Y. et al. Fully automated diagnostic system with artificial intelligence using endocytoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with video). *Gastrointest. Endosc.* **89**, 408–415 (2018).

154. Jain, S. et al. Predictors of long-term outcomes in patients with acute severe colitis: a northern Indian cohort study. *J. Gastroenterol. Hepatol. (Aust.).* **33**, 615–622 (2018).

155. Eck, A. et al. Interpretation of microbiota-based diagnostics by explaining individual classifier decisions. *BMC Bioinforma.* **18**, 441 (2017).

156. Menti, E. et al. Bayesian Machine Learning Techniques for revealing complex interactions among genetic and clinical factors in association with extra-intestinal Manifestations in IBD patients. *AMIA Annu. Symp. Proc. AMIA Symp.* **2016**, 884–893 (2016).

157. Hubenthal, M. et al. Sparse modeling reveals miRNA signatures for diagnostics of inflammatory bowel disease. *PLoS ONE.* **10**, e0140155 (2015). https://doi.org/10.1371/journal.pone.0140155.

158. Cui, H. & Zhang, X. Alignment-free supervised classification of metagenomes by recursive SVM. *BMC Genomics.* **14**, 641 (2013). https://doi.org/10.1186/1471-2164-14-641.

159. Firouzi, F. et al. A decision tree-based approach for determining low bone mineral density in inflammatory bowel disease using WEKA software. *Eur. J. Gastroenterol. Hepatol.* **19**, 1075–1081 (2007).

160. Ozawa, T. et al. Novel computer-assisted diagnosis system for endoscopic disease activity in patients with ulcerative colitis. *Gastrointestinal Endoscopy.* (2018). https://www.sciencedirect.com/science/article/pii/S0016510718331936?via%3Dihub.

161. Reddy, B. K., Delen, D., Agrawal, R. K. Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data. *Health Inform J.* 1460458217751015 (2018).

162. Han, L. et al. A probabilistic pathway score (PROPS) for classification with applications to inflammatory bowel disease. *Bioinformatics* **34**, 985–993 (2018).

163. Yu, S. et al. Surrogate-assisted feature extraction for high-throughput phenotyping. *J. Am. Med. Inf. Assoc.* **24**, e143–e149 (2017).

164. Wisittipanit, N. et al. Classification methods for the analysis of LH-PCR data associated with inflammatory bowel disease patients. *Int J. Bioinform Res. Appl.* **11**, 111–129 (2015).

165. Ahmed, S. et al. Effect of fuzzy partitioning in Crohn's disease classification: a neuro-fuzzy-based approach. *Med. Biol. Eng. Comput.* **55**, 101–115 (2017).

166. Stawiski, K., Pietrzak, I., Mlynarski, W., Fendler, W. & Szadkowska, A. NIRCa: An artificial neural network-based insulin resistance calculator. *Pediatr. Diabetes* **19**, 231–235 (2018).

167. Ben Ali, J. et al. Continuous blood glucose level prediction of Type 1 Diabetes based on Artificial Neural Network. *Biocybern. Biomed. Eng.* **38**, 828–840 (2018).

168. Siegel, A. P. et al. Analyzing breath samples of hypoglycemic events in type 1 diabetes patients: towards developing an alternative to diabetes alert dogs. *J. Breath Res.* **11**, 026007 (2017). https://doi.org/10.1088/1752-7163/aa6ac6.

169. Georga, E. I., Protopappas, V. C., Polyzos, D., Fotiadis, D. I. Online prediction of glucose concentration in type 1 diabetes using extreme learning machines. Conference proceedings:. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference. 2015, 3262–3265 (2015).

170. Jensen, M. H. et al. Evaluation of an algorithm for retrospective hypoglycemia detection using professional continuous glucose monitoring data. *J. Diabetes Sci. Technol.* **8**, 117–122 (2014).

171. Schwartz, F. L., Shubrook, J. H. & Marling, C. R. Use of case-based reasoning to enhance intensive management of patients on insulin pump therapy. *J. Diabetes Sci. Technol.* **2**, 603–611 (2008).

172. Sampath, S., Tkachenko, P., Renard, E. & Pereverzev, S. V. Glycemic control indices and their aggregation in the prediction of nocturnal hypoglycemia from intermittent blood glucose measurements. *J. Diabetes Sci. Technol.* **10**, 1245–1250 (2016).

173. Ling, S. H., San, P. P. & Nguyen, H. T. Non-invasive hypoglycemia monitoring system using extreme learning machine for Type 1 diabetes. *ISA Trans.* **64**, 440–446 (2016).

174. Perez-Gandia, C. et al. Decision support in diabetes care: the challenge of supporting patients in their daily living using a mobile glucose predictor. *J. Diabetes Sci. Technol.* **12**, 243–250 (2018).

175. Guy, R. T., Santago, P. & Langefeld, C. D. Bootstrap aggregating of alternating decision trees to detect sets of SNPs that associate with disease. *Genet. Epidemiol.* **36**, 99–106 (2012).

176. Ceccarelli, F. et al. Biomarkers of erosive arthritis in systemic lupus erythematosus: application of machine learning models. *PLoS ONE* **13**, e0207926 (2018). https://doi.org/10.1371/journal.pone.0207926.

177. Ceccarelli, F. et al. Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models. *PLoS ONE* **12**, e0174200 (2017). https://doi.org/10.1371/journal.pone.0174200.

178. Kan, H. et al. Longitudinal treatment patterns and associated outcomes in patients with newly diagnosed systemic lupus erythematosus. *Clin. Ther.* **38**, 610–624 (2016).

179. Wolf, B. J. et al. Development of biomarker models to predict outcomes in lupus nephritis. *Arthritis Rheumatol.* **68**, 1955–1963 (2016).

180. Armañanzas, R. et al. Microarray analysis of autoimmune diseases by machine learning procedures. *IEEE Trans. Inf. Technol. Biomedicine.* **13**, 341–350 (2009).

181. Reddy, B. K. & Delen, D. Predicting hospital readmission for lupus patients: an RNN-LSTM-based deep-learning methodology. *Comput Biol. Med.* **101**, 199–209 (2018).

182. Tang, Y. et al. Lupus nephritis pathology prediction with clinical indices. *Sci. Rep.* **8**, 10231 (2018).

183. Wang, Y. et al. Random Bits Forest: a Strong Classifier/Regressor for Big Data. *Sci. Rep.* **6**, 30086 (2016).

184. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. Exploring the color feature power for psoriasis risk stratification and classification: a data mining paradigm. *Comput. Biol. Med.* **65**, 54–68 (2015).

185. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. Reliable and accurate psoriasis disease classification in dermatology images using comprehensive feature space in machine learning paradigm. *Expert Syst. Appl.* **42**, 6184–6195 (2015).

186. Shrivastava, V. K., Londhe, N. D., Sonawane, R. S. & Suri, J. S. Reliability analysis of psoriasis decision support system in principal component analysis framework. *Data Knowl. Eng.* **106**, 1–17 (2016).

187. Choung, R. S. et al. Synthetic neoepitopes of the transglutaminase-deamidated gliadin complex as biomarkers for diagnosing and monitoring celiac disease. *Gastroenterology* **156**, 582–591.e581 (2019).

188. Ahmad, W., Ahmad, A., Lu, C., Khoso, B. A. & Huang, L. A novel hybrid decision support system for thyroid disease forecasting. Soft Computing - A Fusion of Foundations. *Methodologies Appl.* **22**, 5377–5383 (2018).

189. Baccour, L. Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets. *Expert Syst. Appl.* **99**, 115–125 (2018).

190. Morejón, R., Viana, M. & Lucena, C. An approach to generate software agents for health data mining. *Int. J. Softw. Eng. Knowl. Eng.* **27**, 1579–1589 (2017).

191. Temurtas, F. A comparative study on thyroid disease diagnosis using neural networks. *Expert Syst. Appl.* **36**, 944–949 (2009).

192. Polat, K., Şahan, S. & Güneş, S. A novel hybrid method based on artificial immune recognition system (AIRS) with fuzzy weighted pre-processing for thyroid disease diagnosis. *Expert Syst. Appl.* **32**, 1141–1147 (2007).

193. Keleş, A. & Keleş, A. ESTDD: Expert system for thyroid diseases diagnosis. *Expert Syst. Appl.* **34**, 242–246 (2008).

194. Singh, A. & Pandey, B. A KLD-LSSVM based computational method applied for feature ranking and classification of primary biliary cirrhosis stages. *Int. J. Comput. Biol. Drug Des.* **10**, 24–38 (2017).

195. Eaton, J. E. et al. Primary sclerosing cholangitis risk estimate tool (PREsTo) predicts outcomes in PSC: a derivation & validation study using machine learning. *Hepatology* (2018). https://www.ncbi.nlm.nih.gov/pubmed/29742811.

196. Taroni, J. N., Martyanov, V., Mahoney, J. M. & Whitfield, M. L. A functional genomic meta-analysis of clinical trials in systemic sclerosis: toward precision medicine and combination therapy. *J. Invest. Dermatol.* **137**, 1033–1041 (2017).

197. Huang, H. et al. A methodology for exploring biomarker–phenotype associations: application to flow cytometry data and systemic sclerosis clinical manifestations. *BMC Bioinforma.* **16**, 293 (2015).

198. Brodersen, K. H., Ong, C. S., Stephan, K. E., Buhmann, J. M., (eds) The Balanced Accuracy and Its Posterior Distribution. 2010 20th International Conference on Pattern Recognition; 2010 23–26 Aug. 2010.

199. Goutte, C., Gaussier, E., (eds) *A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation*. (Springer Berlin Heidelberg, 2005).

## ACKNOWLEDGEMENTS

## AUTHOR CONTRIBUTIONS

I.S.S., S.E. and E.M. formed the search strategy and inclusion and exclusion criteria. I.S.S. performed database searches and reviewed all records for eligibility (title and abstract, full record if necessary). M.K. reviewed all records (title and abstract). E.M. reviewed records where agreement could not be reached between I.S.S. and M.K. regarding study inclusion. I.S.S. read and evaluated all included studies, with analysis and interpretation assistance from E.M., B.D.M. (M.L. perspective) S.E. and R.M.B. (clinical perspective). I.S.S. wrote the manuscript with input from M.K., E.M., B.D.M., R.M.B. and S.E.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41746-020-0229-3.

**Correspondence** and requests for materials should be addressed to S.E.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.