Epidemiology & Health Data Insights
(eISSN 3080-8111)

Review Article

Machine Learning for Predictive Modeling of Climate-Sensitive Autoimmune Diseases

Ikemefula Oriaku[1], Jennifer Payin Baiden[2], Oluwafemi Johnson Bamidele[3], Olukunle O. Akanbi[4], Vivian Ukamaka Nwokedi[5]

[1]*Department of Medical Laboratory Science, University of Nigeria, Nsukka, Nigeria*
[2]*Department of Nanoscience and Nanoengineering, University of North Carolina, Greensboro, USA*
[3]*Department of Computer Science, College of Engineering, Prairie View A&M University, USA*
[4]*Department of Psychology and Behavioral Sciences, National Louis University, FL, USA*
[5]*Department of Clinical Pharmacy, Faculty of Pharmacy, University of Benin, Benin City, Nigeria*

Received: Nov 24, 2025
Accepted: Dec 23, 2025

Corresponding author's email:
viviannwokedi250@gmail.com

Abstract:
Autoimmune diseases, including Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis (RA), and Multiple Sclerosis (MS), represent a growing global health burden. These diseases disproportionately affect women and the young, and their complex aetiology involves an interplay between genetic susceptibility and environmental triggers. In light of climate change's increasing influence on health outcomes, this study explores the potential of machine learning (ML) models to predict climate-sensitive autoimmune diseases. We examine the integration of diverse data sources, such as electronic health records (EHRs), genomic data, and climate exposures, to enhance predictive accuracy. Current ML models in autoimmune disease prediction primarily rely on clinical and omics data, with limited consideration for environmental factors. We identify significant gaps, particularly in incorporating climate data such as particulate matter, UV radiation, and temperature variability. The study also highlights the challenges of data fusion, feature engineering, and causal inference in these models. Ethical concerns, including data privacy, model explainability, and equity, are also addressed. The research underscores the need for large-scale, prospective studies to validate climate-informed models and calls for policy-driven approaches to ensure equitable access and deployment. By bridging these gaps, climate-informed ML models hold promise for personalized, proactive disease prevention and public health planning.

Keywords: Machine Learning; Autoimmune Diseases; Climate Change; Predictive Modeling; Environmental Factors

Introduction

The Burden of Autoimmune Diseases in a Changing World

The 21st century has witnessed a notable and concerning increase in the incidence and prevalence of autoimmune diseases, a complex class of chronic conditions where the body's immune system mistakenly attacks its own tissues [1, 2]. Autoimmune Diseases such as Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis (RA), and Multiple Sclerosis (MS) represent a significant and growing global health burden [3]. Autoimmune Diseases disproportionately affect women and the young and are a leading cause of morbidity and disability worldwide, imposing immense personal, social, and economic costs [4, 5]. The aetiology of these diseases is multifactorial, arising from an intricate interplay of genetic predispositions and environmental triggers. While genetic factors provide the 'susceptibility', it is the environmental 'trigger' that often initiates the clinical presentation of the disease [6, 7].

The precise reasons for the rising trends remain elusive, but it is clear that our evolving environment plays a crucial role. The traditional medical paradigm has often focused on internal biological factors, but a more holistic and systems-based view is essential to unravel these complexities. The urgency for advanced predictive and preventive strategies has never been greater. Current diagnostic methods are often reactive, identifying the disease only after significant immune-mediated damage has occurred. There is a pressing need to shift from a reactive to a proactive model of care, one that can identify individuals at high risk and intervene before disease onset. This shift requires sophisticated tools capable of integrating vast and disparate data sources to understand complex causal pathways.

The Climate-Autoimmunity Link

Among the most significant and rapidly changing environmental factors in the modern era is the climate itself. Climate change is not merely an abstract future threat; it is a current driver of health outcomes, including the modulation of immune system function [8]. The connection between climate and autoimmune diseases is a burgeoning field of study, providing compelling evidence that climate-related stressors can act as potent environmental triggers. The mechanisms linking these seemingly disparate domains are multi-layered and involve a combination of direct biological effects and indirect ecological changes as shown in Figure 1 and Table 1.

One prominent mechanism involves air pollution, particularly particulate matter (PM2.5). Exposure to PM2.5 has been associated with increased systemic inflammation and oxidative stress, which are underlying mechanisms in the development and exacerbation of autoimmune conditions [9, 10]. For example, studies have shown that high exposure to traffic-related air pollution is associated with an increased risk of developing rheumatoid arthritis [9, 11]. The fine particles can penetrate deep into the lungs and even the bloodstream, triggering immune responses that become dysregulated over time.

Another well-established link involves ultraviolet (UV) radiation exposure and diseases like Systemic Lupus Erythematosus (SLE). UV light is a known environmental trigger for SLE flares and is believed to contribute to disease pathogenesis through mechanisms that induce cell death (apoptosis) and alter immune regulation [12]. The changing patterns of UV exposure due to ozone layer variations and altered outdoor activity patterns in a warmer climate add another layer of complexity to disease risk management.

Beyond specific pollutants and radiation, broader climate-related stressors such as extreme temperatures, altered seasonal patterns, and associated changes in local ecosystems are implicated. For instance, temperature variability has been linked to flare-ups in conditions like Multiple Sclerosis (MS) and Rheumatoid Arthritis (RA) [13, 14]. The proposed mechanisms are diverse, ranging from immune system alterations mediated by the gut microbiome (which is sensitive to dietary and environmental changes) to direct cellular stress responses [13, 14].

Table 1: The Link between Environmental/Climate Factors and Autoimmune Diseases

Environmental/Climate Factor	Associated Autoimmune Disease(s)	Mechanism
Particulate Matter (PM2.5)	Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE)	Chronic inflammation, oxidative stress, immune dysregulation

UV-B Radiation	Systemic Lupus Erythematosus (SLE), Multiple Sclerosis (MS)	Immunosuppression, apoptosis of keratinocytes, Vitamin D synthesis modulation
Extreme Temperature Variability	Multiple Sclerosis (MS), Rheumatoid Arthritis (RA)	Physiological stress response, altered pain perception, systemic inflammation
Altered Gut Microbiome	Inflammatory Bowel Disease (IBD), Type 1 Diabetes	Dysbiosis, breakdown of gut barrier integrity, immune system activation

[AUC: Area Under the Curve, SNP: Single Nucleotide Polymorphism, RA: Rheumatoid Arthritis, SLE: Systemic Lupus Erythematosus, RF: Random Forest, SVM: Support Vector Machine, GTB: Gradient Tree Boosting, XGBoost: eXtreme Gradient Boosting, NLP: Natural Language Processing]

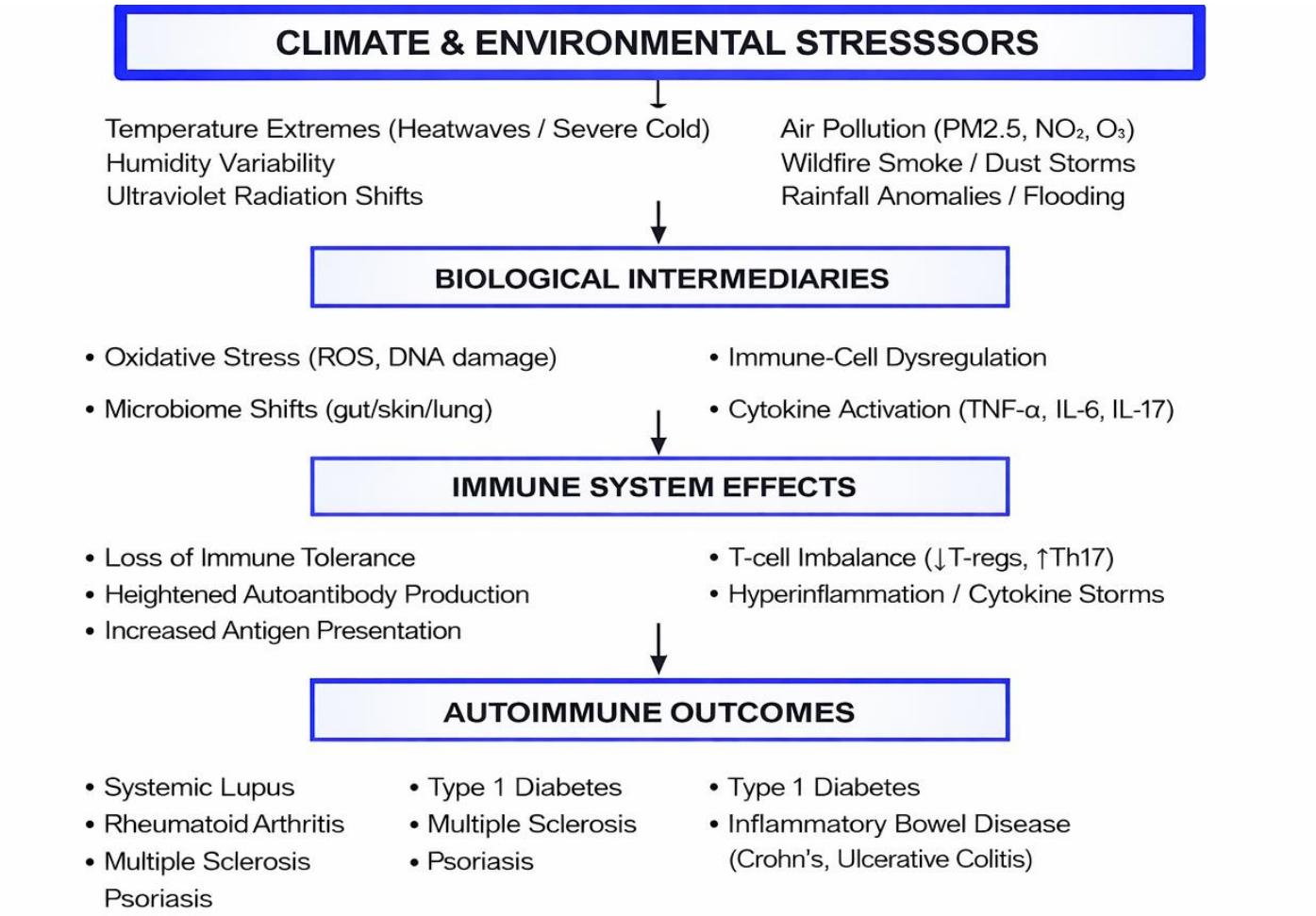

Figure 1. Conceptual Diagram of Biological Mechanisms Triggered by Climate Change

The Promise of Machine Learning in Complex Health Systems

Traditional statistical methods, while valuable, often struggle to capture the non-linear, multi-

dimensional, and time-dependent nature of the interactions between thousands of patient-specific variables and dynamic environmental factors. This is where machine learning (ML) emerges as a transformative tool for healthcare and public health [15]. ML algorithms excel at identifying subtle patterns and complex relationships within vast and heterogeneous datasets that might be invisible to human observation or traditional methods [15].

ML models, ranging from simple regression algorithms to sophisticated deep learning neural networks, can integrate patient electronic health records (EHRs), genomic data, social determinants of health, and crucially, large-scale environmental and climate data (e.g., satellite imagery, meteorological station data, air quality monitors). This capacity for data fusion is paramount for addressing a challenge as complex as climate-sensitive autoimmune diseases [16]. For example, a model could learn that a specific combination of genetic markers, prior infection history, and exposure to a certain level of PM2.5 in a specific season leads to a significantly elevated risk of RA onset three years later [17, 18]. Such insights enable far more precise and personalised risk assessments than are currently possible.

Objective

The main objective of this narrative review is to critically synthesize the existing literature on the use of machine learning for predicting autoimmune diseases and to explore the feasibility, challenges, and necessity of integrating granular climate and environmental data into such models.

Methodologies for Narrative Review and Literature Search

In carrying out this review a narrative synthesis approach was adopted and data captured by conducting a comprehensive search across multiple electronic databases (PubMed, ScienceDirect, and Google Scholar). The search was performed on September 10, 2024 using a combination of keywords to capture relevant publications from January 2010 to the search date. Key search terms were strategically combined using Boolean operators: (Machine Learning OR AI OR Artificial Intelligence OR Predictive Model) AND (Autoimmune Disease OR Rheumatoid Arthritis OR Systemic Lupus Erythematosus OR Multiple Sclerosis) AND (Climate Change OR Environmental Factors OR Air Pollution OR UV Radiation OR Temperature OR Environmental Triggers). The search was refined iteratively based on initial results to capture the breadth of both the medical and data science literature.

The selection process was conducted in two stages. In the initial screening phase, titles and abstracts were reviewed for relevance to at least two of the three core thematic areas: machine learning, autoimmune diseases, and environmental/climate factors. Studies were included if they were primary research articles or review articles published in English. The exclusion criteria comprised editorials, commentaries, conference abstracts without full paper availability, and studies that solely focused on genetic factors without any environmental or machine learning component.

In the second stage, the full texts of potentially relevant articles were retrieved and assessed for their direct contribution to the study. We also examined the reference lists of key review articles to identify additional relevant papers not captured by the initial keyword search. Narrative synthesis was carried out using emerging themes from the included studies to discuss this study in detail.

State-of-the-Art of Machine Learning Models for Autoimmune Disease Prediction

In this section we critically review the data sources employed in existing models, the machine-learning algorithms most commonly used in the autoimmune domain, and the performance and limitations of these models — with special attention to the notable gap of climate or environmental exposure integration.

Data Sources in Existing Models

The foundation of any predictive framework is the data on which it is built. In the autoimmune domain, three major classes of data sources dominate: electronic health records (EHRs) or clinical data, genomic / omics data, and lifestyle or environmental factors (although the latter remain under-represented). Clinical or EHR data are ubiquitous: for example, models built to identify patients who warrant autoimmune disease testing leveraged EHR-derived features (lab values, ICD codes, demographic data) in a large hospital biobank [19]. Omics data—such as genome-wide SNP arrays, gene expression profiles, methylation or proteomic data—have increasingly been used, especially to refine risk prediction or subtype classification. For example, a study of Rheumatoid Arthritis (RA) and Systemic Lupus Erythematosus (SLE) used genome-wide SNP data and applied ML algorithms (random forest, SVM, gradient tree boosting) to achieve AUCs of 0.98 or higher [20]. Lifestyle and environmental factors (smoking, BMI,

diet, physical activity) have been included in some models but remain much less common and rarely integrated with multi-omics and clinical data. A recent bibliometric review noted that many autoimmune-ML studies did not include multimodal exposures beyond the clinical/omics dimension [21]. In short: the bulk of predictive modelling for autoimmune diseases has relied on clinical/EHR and genomic/omics data; inclusion of environmental, behavioural and climate exposures has been minimal so far.

Common Machine Learning Algorithms

Turning to the algorithms, a consistent pattern emerges across reviews of autoimmune-ML work. Supervised learning methods such as random forests (RF), support vector machines (SVMs) and logistic regression (LR) remain common, while neural networks and gradient boosting methods are increasingly encountered in more recent work. For instance, Stafford et al. (2020) found that in their systematic review of 169 ML studies in autoimmune diseases the most-commonly used methods across the board were SVM and random forest [22]. In a more recent study using SNP data for RA and SLE, Chung et al. (2021) compared LR (AUC ≈ 0.82) with RF (AUC ≈ 0.98), SVM (≈0.98) and gradient tree boosting/ XGBoost (≈0.99) [20]. Deep learning (DL) and hybrid approaches are now emerging: for example, a multi-task neural network designed to integrate DNA methylation data across multiple autoimmune phenotypes (RA, SLE, multiple sclerosis, type-1 diabetes) showed improved performance and interpretability [23]. One must note, however, that though deep models may give higher nominal accuracy, their transparency and reproducibility in the autoimmune space remain challenging. Reviews emphasise that the lack of standardisation of feature-engineering, validation strategy, and the heterogeneity of autoimmune cohorts complicates algorithmic comparison [24].

Performance and Limitations of Current Models

When we look at performance metrics, the results are encouraging but must be tempered with care. Many studies report high AUCs (often >0.90) and accuracies, especially where the data are well-characterised and sample size adequate. For instance, in the ML pipeline integrating clinical, laboratory and omics data the authors achieved up to 96 % accuracy in classifying autoimmune disease status [25]. A meta-analysis of AI/ML in autoimmune diseases reported a mean accuracy of ≈91.06 % (95 CI 86.38-95.73%) across a variety of conditions (though the methods and definitions varied) [26]. These results suggest that algorithmically, the field is capable of strong discriminative performance under favourable conditions.

Yet, several key limitations undermine the translation of these models into real-world, climate-sensitive prediction frameworks. First, generalisability: many models are trained on relatively homogeneous (often single-centre) cohorts, and external validation is weak. The heterogeneity of autoimmune diseases (phenotypic variation, disease progression, comorbidities) further complicates transferability [24]. Second, data integration remains a challenge: combining EHR, genomics, imaging, lifestyle, and emerging climate- or exposome-derived data is rare. Third, climate/environmental exposures (such as temperature extremes, UV exposure, air pollutants, humidity changes) are virtually absent in the majority of published autoimmune-ML models — yet these exposures are biologically plausible triggers for immune dysregulation and autoimmunity. In other words, if one's aim is to build a predictive model for climate-sensitive autoimmune disease, this gap is acute. Moreover, interpretability and transparency pose further hurdles: many high-performing models are "black boxes," and their clinical adoption is hampered by lack of explainability. Ethical and data governance issues (privacy, biased sampling, missingness) also remain under-addressed [21]. Finally, even when exposures are captured, temporal linkages (lag effects of climate on immune-response) and fine-grained spatial exposure data are seldom included.

The Integration of Climate Data into Predictive Models: Challenges and Opportunities

Data Acquisition and Fusion

Bringing together high-resolution climate data with individual health records and other biomedical streams is no trivial task. Climate information is often captured at macro-scales (satellite imagery, gridded model outputs, weather station networks) and at temporal granularities that may not align with clinical or patient-level data. For example, a satellite-derived measure of surface temperature or UV irradiance may come in daily or hourly bins over a spatial grid of kilometres, whereas individual-level health records (e.g., from electronic health records (EHRs) or cohort databases) register exposures and events at the level of visits, diagnoses, biomarkers or even self-reported symptoms [27]. Literature in climate-health modelling emphasises that the integration of such distinct modalities — health, environment, social-determinant data — requires both careful alignment in space and

time and rigorous handling of mismatches in resolution and missingness [28].

The fusion process typically involves geocoding patient residence or service-use locations, linking them to the nearest climate grid cell or weather station, aggregating exposures over relevant windows (e.g., months or seasons) and then merging with routine clinical or demographic variables. Challenges arise when climate data are proprietary, when patient privacy limits granular location linking, or when temporal windows are arbitrarily chosen without proper justification. Further complexity emerges when climate exposures need to be harmonised with other environmental sensors (air pollution, pollen counts, humidity indices), lifestyle factors and genomic or biomarker data streams. The literature points out that many ML health models ignore this fusion step or treat climate data as an add-on rather than a core modality [29].

Feature Engineering from Climate Data

Once climate exposures are aligned to patient records or cohorts, the raw data still require transformation into features that an ML model can make sense of. For example, rather than simply including "daily maximum temperature" for a patient's location, one might compute the cumulative number of heatwave days over the past six months, the rolling average of UV index exposures, or the number of days with extreme humidity above a certain threshold. These engineered variables potentiate the ability of the model to detect patterns tied to immune dysregulation or flare-risk. In agricultural health-modelling contexts, studies have translated raw climate variables into meaningful predictors (e.g., cumulative precipitation deficit) for crop yields; analogous methods hold for autoimmune disease contexts, though they remain rare [30].

Strategic feature engineering entails domain knowledge: determining whether short-term exposures (e.g., a heatwave in the past week) or long-term exposures (e.g., seasonal shifts over the past year) matter more for autoimmune pathogenesis, selecting appropriate aggregation windows, and controlling for collinearity (for example, temperature and humidity often move together). In addition, transformations for lags, moving averages, thresholds and interaction terms (such as UV × smoking status) may yield richer predictors. Importantly, such features should be interpretable; for clinical translation, features labelled in human-readable fashion (e.g., "6-month cumulative days with UV index > 8") facilitate stakeholder engagement. As the climatic drivers of immune dysfunction become better understood (e.g., pollen loads, drought stress, wild-fire-derived particulate matter), the feature-engineering toolkit must expand accordingly [28].

Addressing the Causal Inference Problem

Integrating climate features into ML models raises fundamental questions of causality versus correlation. It is comparatively straightforward to train a model that finds a statistical association between "high UV index exposures" and "autoimmune flare within 30 days", but far harder to assert that the exposure caused the flare. The pathophysiology of autoimmune disease is complex, involving genetics, immune history, environment, and stochastic events. As the literature on immune-mediated disease emphasises, the exposome (cumulative environmental exposures) is rapidly changing due to climate change, but linking that to individual-level disease onset remains scientifically immature [28].

To strengthen causal interpretation in this context, it becomes essential to address temporal precedence, ensuring that environmental exposures not only correlate with—but clearly precede—disease events within biologically plausible windows. Recent methodological work underscores that establishing temporality is foundational to causal inference, particularly for environmental exposures whose effects may be delayed, cumulative, or episodic [31]. Autoimmune flares may arise weeks or months after relevant exposures, making it necessary for models to incorporate temporal lags and exposure windows rather than rely on static, cross-sectional measures.

From a modelling vantage point, causal inference demands more than high accuracy: one must consider bias, confounding (for example, patients living in urban heat-island zones may also have higher pollution, different access to care, or different socioeconomic status), and temporal precedence (does the exposure precede the disease event in a biologically plausible window?). Ignoring this may produce models that predict well but offer little mechanistic insight or clinical trust. One way to strengthen temporal reasoning is through longitudinal datasets, which allow researchers to track exposures and disease outcomes over time, enabling clearer separation of cause and effect [32]. Longitudinal climate-health designs also reduce the risk of reverse causation—a frequent limitation in ecological and cross-sectional studies.

Another critical consideration is the selection and structure of the study population. Heterogeneity in exposure is common in climate-health research, and failing to account for socioeconomic, geographic, and health-related differences can introduce substantial selection bias. Stratifying populations based on geography, environmental burden, baseline immune vulnerability, or socioeconomic position helps ensure

that observed associations are not artefacts of unequal exposure distributions [33, 34]. Such stratification improves validity and better reflects the uneven climate-related burdens many communities face.

Similarly, the classification of both exposure and outcome requires careful attention. Climate exposures differ in intensity, duration, and timing, and autoimmune outcomes differ in onset, severity, and chronicity. Defining exposures using cumulative, lagged, or threshold-based metrics helps align environmental measures with the biological timelines of autoimmune activation [35]. Likewise, clearly distinguishing incident autoimmune disease from flare-ups or exacerbations is crucial to avoid outcome misclassification, a frequent source of bias in observational epidemiology.

Hybrid approaches combining ML with causal-inference frameworks (e.g., directed acyclic graphs, instrumental variables) are gaining traction in climate-health research, though rarely applied in autoimmune-ML work to date [36]. Incorporating methods such as propensity score matching, inverse probability weighting, or instrumental variable estimation can reduce confounding and help approximate causal effects in the absence of randomized designs [37, 38]. Such methods are increasingly recommended for complex environmental-health settings where exposures cannot be randomized and confounders are interrelated [39].

Furthermore, interpretability remains crucial: clinician stakeholders must understand how a climate-derived predictor contributes to risk, else model outputs may not be adopted. The ethical dimension is also non-trivial: exposures may be unequally distributed (e.g., heat stress in low-income communities), and if climate-informed models do not account for those equity dimensions, they risk replicating bias or deepening disparities. This interplay between climate, health, and equity is a theme emerging in recent work [28].

Ultimately, integrating climate data into autoimmune-disease predictive modelling presents real promise: it opens a pathway to anticipate disease risks tied to global environmental change. Yet the path is strewn with methodological, technical and ethical obstacles. By carefully acquiring and fusing data, engineering meaningful features, and attending to causality and transparency, researchers can build models that are not only predictive but also clinically and socially responsible. The conceptual framework depicted in Figure 1 underscores that success will require bridging raw data streams into a refined ML architecture that delivers individualised risk scores while maintaining clear lines of explanation and governance.

Potential Applications and Clinical Implications of Climate-Informed Models

Personalised Risk Stratification

By incorporating climate exposures such as cumulative UV-index, heatwave days, humidity shifts or air-pollution peaks, a model can refine the baseline risk derived from genetics, lifestyle and clinical history. Evidence suggests that climate change is altering the exposome in ways relevant to autoimmunity — for instance, shifts in pollen burden, wildfire smoke and heat-stress have been linked to immune dysregulation [28]. With climate-informed features embedded in an ML-pipeline, clinicians could identify those whose risk profile is unexpectedly elevated because of recent environmental exposures, even when their genomic or lifestyle risk seems modest. This opens up the possibility of true anticipatory care: intervening before irreversible immune damage accumulates. From a patient-empowerment perspective, being given a personalised risk score tied to modifiable exposures can catalyse engagement, trust and preventive behaviour.

Preventive Interventions

Once a high-risk individual is identified, climate-informed modelling enables targeted preventive actions. Rather than generic advice, the clinician could issue timesensitive recommendations: for example, “Avoid outdoor activity on forecasted high-UV or high-pollution days” or “Increase sun-protection following a sustained heat-wave over the past fortnight”. The literature on climate change adaptation in public health underlines the importance of such context-specific behavioural guidance [40]. Moreover, when a model flags heightened risk because of a hot-spell or wildfire exposure, health systems could trigger early-warning alerts to those patients, prompting check-ups or pre-emptive anti-inflammatory monitoring. This proactive model offers a dynamic, rather than static, preventive paradigm—especially important in an era where climate variability is increasing the frequency of triggering exposures [41].

Resource Allocation and Public Health Policy

Beyond individual care, climate-informed autoimmune-disease models offer tremendous value for public-health planning and resource allocation. If a region is predicted to experience a cluster of exposures (e.g., heatwave followed by high ozone/pollution days) that historically drive increased autoimmune flare-ups,

then the local health system can pre-position resources: staffing for rheumatology clinics, alerting primary-care networks, and ensuring diagnostic capacity is ramped up. The broader literature on climate-sensitive health risk emphasises that detection and prediction of environmental-health hotspots is key to resilience [42]. In policy terms, aggregated outputs from many individuals' risk-scores could inform zoning, heat-mitigation programmes, and environmental health regulation — by linking climate stressors to disease-burden in actionable models, health policy can move from reactive to anticipatory.

In sum, the integration of climate data into predictive modelling for autoimmune diseases promises to shift practice in three interlinked domains: personalised risk stratification, time-sensitive preventive action and system-level planning. The challenge will be ensuring the models are robust, interpretable and ethically deployed so that they serve individuals equitably and enhance public-health resilience in a warming world.

Table 2: Framework for Integrating Climate and Health Data in Predictive Modeling for Autoimmune Disease Risk

Stage	Data Sources / Processes	Description
1. Data Acquisition	Electronic Health Records (EHRs), Climate Sensors, Satellite Imagery	These data sources provide crucial health, environmental, and geographical information.
2. Data Preprocessing & Feature Engineering	Integration and transformation of raw data into meaningful features	Raw data is processed, cleaned, and transformed into usable features, e.g., cumulative exposure.
3. Integrated Dataset	Merged dataset combining health and climate data	Data from different sources are fused into one cohesive dataset, ready for machine learning.
4. Machine Learning Model	ML algorithms (e.g., Random Forest, Support Vector Machine (SVM), etc.)	The integrated dataset is fed into a machine learning model to predict disease risk.
5. Predictive Output	Individual Risk Score (e.g., likelihood of flare)	The model generates a predictive output that assesses the individual's risk level for disease.

Ethical Considerations and Future Directions

As the field of predictive modelling for climate-sensitive autoimmune diseases moves forward, it is crucial to hold a mirror up to the ethical dimensions and to chart a responsible path for future research. The interplay between sensitive health data, complex machine learning frameworks and the broader pathogenic role of climate exposures demands a reflexive mindset.

Firstly, any model that integrates health records with granular climate and geolocation data must grapple with data privacy and security in a rigorous way. When clinicians or researchers merge individual-level EHRs with spatial-temporal grids of climate or pollutant exposure, the potential to inadvertently re-identify individuals or expose sensitive location histories becomes real. Literature on AI-driven healthcare emphasises that patient consent, data ownership, secure transmission and storage protocols are non-negotiable. Federated learning and differential privacy techniques offer promise, but they are neither universal nor fool-proof [43]. Beyond technical solutions, equity matters: those living in marginalized communities may be less protected by data governance frameworks, raising justice and fairness concerns. If risk-scores derived from climate-linked ML models are misused (for insurance, employment, or discrimination) the very promise of

personalized prevention could flip into a new axis of inequality.

Secondly, model explainability and trust are at the heart of clinical acceptance and ethical deployment. Building a high-accuracy algorithm is only half the battle; the clinician and ultimately the patient must understand, at least in part, why a risk assessment was made. "Black box" models, which amalgamate hundreds of features (clinical, genomic, climate) into an opaque output, increase the likelihood of scepticism or outright rejection in practice. Reviews in medical AI stress that explainable AI (XAI), domain-adaptive transfer learning and transparency are key to safe deployment [44]. In our climate-informed autoimmune context, it means that the feature "72-day heatwave exposure + UV index > 8" should map coherently to immune system activation in clinician-familiar language. If a model says "High risk" without linking back to interpretable biological or environmental logic, trust will erode. Moreover, model failures must be auditable, and governance frameworks should ensure fairness across different geographies, socio-economic strata and climate zones.

Thirdly, turning to future directions, the research agenda needs expansion and deepening. There is a clear need for large-scale, prospective studies that follow cohorts over time, capturing not only clinical and genomic data but detailed climate- and environment-exposure histories. While retrospective modelling is useful, true validation of climate-informed risk predictions requires longitudinal data that reflects changing exposures, immune system priming and eventual disease onset. The mechanistic biology also calls for deeper enquiry: as illustrated in the diagram above, climate-related exposures (for instance, particulate matter, UV radiation, humidity shifts) may trigger oxidative stress, DNA damage, microbiome dysbiosis, epigenetic reprogramming and ultimately immune activation—and these pathways remain insufficiently mapped in autoimmune disease contexts [45]. Additionally, regional equity should not be an afterthought. Climate exposures differ dramatically by geography, and models built in temperate nations may not translate to tropical or low-income settings without recalibration and local data. In parallel, there is the ethical imperative to ensure that the benefits of such predictive models (early-warning, prevention, resource allocation) do not widen health disparities.

In essence, while the promise of climate-informed predictive modelling for autoimmune diseases is substantial, the pathway is bound to ethical, technical and equity-related constraints. Data-protection architecture must be robust and context-sensitive; model interpretability must be baked into design not afterthought; and research must scale into prospective, globally representative efforts. Only then can we build models that not only predict but serve—in a way that is scientifically sound, ethically grounded and socially just.

Conclusion

Integrating climate data into predictive models for autoimmune diseases offers transformative potential for early risk stratification and personalized interventions. While the approach promises more precise prevention strategies, challenges in data fusion, ethical considerations, and model explainability remain. Future research must focus on large-scale, longitudinal studies and address regional disparities to ensure equitable, actionable outcomes. Policymakers should prioritize climate-informed health strategies to improve disease prevention and resource allocation in a changing environment.

Acknowledgments

Competing Interests: All the authors declare that they have no conflict of interest.

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: All the resources consulted in the review are provided in the reference section.

Funding: No funding of any kind was received.

Authors' Contributions: This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

References

1. Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. Curr Opin Immunol. 2023 Feb;80:102266. https://doi.org/10.1016/j.coi.2022.102266
2. Conrad N, Misra S, Verbakel JY, Verbeke G, Molenberghs G, Taylor PN, et al. Incidence, prevalence, and co-occurrence of autoimmune

disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. Lancet. 2023 Jun;401(10391):1878-90. https://doi.org/10.1016/S0140-6736(23)00457-9

3. Zeng SQ, Zhen JH, Pu Y, Liu C, Hu JM, Chen JJ, et al. Global, regional, and national burden of autoimmune disease in older adults (≥ 60 years) from 1990 to 2021: Results from the Global Burden of Disease Study 2021. J Nutr Health Aging. 2025 Nov;29(11):100681. https://doi.org/10.1016/j.jnha.2025.100681
4. Desai MK, Brinton RD. Autoimmune disease in women: endocrine transition and risk across the lifespan. Front Endocrinol (Lausanne). 2019 Apr;10:265. https://doi.org/10.3389/fendo.2019.00265
5. Angum F, Khan T, Kaler J, Siddiqui L, Hussain A. The prevalence of autoimmune disorders in women: a narrative review. Cureus. 2020 May;12(5):e8094. https://doi.org/10.7759/cureus.8094
6. Kumar M, Yip L, Wang F, Marty SE, Fathman CG. Autoimmune disease: genetic susceptibility, environmental triggers, and immune dysregulation. Where can we develop therapies? Front Immunol. 2025 Aug;16:1626082. https://doi.org/10.3389/fimmu.2025.1626082
7. Yasmeen F, Pirzada RH, Ahmad B, Choi B, Choi S. Understanding autoimmunity: mechanisms, predisposing factors, and cytokine therapies. Int J Mol Sci. 2024 Jul;25(14):7666. https://doi.org/10.3390/ijms25147666
8. Imberti L, Tiecco G, Logiudice J, Castelli F, Quiros-Roldan E. Effects of Climate Change on the Immune System: A Narrative Review. Health Sci Rep. 2025 Apr;8(4):e70627. https://doi.org/10.1002/hsr2.70627
9. Adami G, Pontalti M, Cattani G, Rossini M, Viapiana O, Orsolini G, et al. Association between long-term exposure to air pollution and immune-mediated diseases: a population-based cohort study. RMD Open. 2022 Feb;8(1):e002055. https://doi.org/10.1136/rmdopen-2021-002055
10. Sangkham S, Phairuang W, Sherchan SP, Pansakun N, Munkong N, Sarndhong K, et al. An update on adverse health effects from exposure to PM2.5. Environ Adv. 2024 Dec;18:100603. https://doi.org/10.1016/j.envadv.2024.100603
11. Ho WC, Chou LW, Wang RY, Doan TN, Yu HL, Chou TH, et al. Association between exposure to ambient air pollution and the risk of rheumatoid arthritis in Taiwan: A population-based retrospective cohort study. Int J Environ Res Public Health. 2022 Jun;19(12):7006. https://doi.org/10.3390/ijerph19127006
12. Barbhaiya M, Hart JE, Malspeis S, Tedeschi SK, VoPham T, Sparks JA, et al. Association of ultraviolet B radiation and risk of systemic lupus erythematosus among women in the nurses' health studies. Arthritis Care Res (Hoboken). 2023 Jul;75(7):1409-15. https://doi.org/10.1002/acr.24974
13. Christogianni A, Bibb R, Davis SL, Jay O, Barnett M, Evangelou N, et al. Temperature sensitivity in multiple sclerosis: an overview of its impact on sensory and cognitive symptoms. Temperature (Austin). 2018 Jul;5(3):208-23. https://doi.org/10.1080/23328940.2018.1475831
14. EK KP, PR S, Banerjee M. Impact of seasonal cycle on Rheumatoid Arthritis based on genetic and epigenetic mechanism. Front Immunol. 2025 Sep;16:1601767. https://doi.org/10.3389/fimmu.2025.1601767
15. Rajula HS, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. Medicina (Kaunas). 2020 Sep;56(9):455. https://doi.org/10.3390/medicina56090455
16. Donovan M. The Role of Machine Learning in Enhancing Biomedical Systems for Personalized Medicine. J Biomed Syst Emerg Technol. 2024;11:207.
17. Xu Y, Wu Q. Using machine learning and single nucleotide polymorphisms for

improving rheumatoid arthritis risk Prediction in postmenopausal women. PLOS Digit Health. 2025 Apr;4(4):e0000790. https://doi.org/10.1371/journal.pdig.0000790

18. Shi Y, Zhou M, Chang C, Jiang P, Wei K, Zhao J, et al. Advancing precision rheumatology: applications of machine learning for rheumatoid arthritis management. Front Immunol. 2024 Jun;15:1409555. https://doi.org/10.3389/fimmu.2024.1409555
19. Forrest IS, Petrazzini BO, Duffy Á, Park JK, O'Neal AJ, Jordan DM, et al. A machine learning model identifies patients in need of autoimmune disease testing using electronic health records. Nat Commun. 2023 Apr;14(1):2385. https://doi.org/10.1038/s41467-023-37996-7
20. Chung CW, Hsiao TH, Huang CJ, Chen YJ, Chen HH, Lin CH, et al. Machine learning approaches for the genomic prediction of rheumatoid arthritis and systemic lupus erythematosus. BioData Min. 2021 Oct;14(1):52. https://doi.org/10.1186/s13040-021-00284-5
21. Liu S, Liu Y, Li M, Shang S, Cao Y, Shen X, et al. Artificial intelligence in autoimmune diseases: a bibliometric exploration of the past two decades. Front Immunol. 2025 Feb;16:1525462. https://doi.org/10.3389/fimmu.2025.1525462
22. Stafford IS, Kellermann M, Mossotto E, Beattie RM, MacArthur BD, Ennis S. A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. NPJ Digit Med. 2020 Mar;3(1):30. https://doi.org/10.1038/s41746-020-0229-3
23. Xu P, Cai J, Gao Y, Rong Z. MIRACLE: Multi-task learning based interpretable regulation of autoimmune diseases through common latent epigenetics. arXiv [Preprint]. 2023 Jun. Available from: https://doi.org/10.48550/arXiv.2306.13866
24. Momtazmanesh S, Nowroozi A, Rezaei N. Artificial intelligence in rheumatoid arthritis: current status and future perspectives: a state-of-the-art review. Rheumatol Ther. 2022 Oct;9(5):1249-304. https://doi.org/10.1007/s40744-022-00475-4
25. Kruta J, Carapito R, Trendelenburg M, Martin T, Rizzi M, Voll RE, et al. Machine learning for precision diagnostics of autoimmunity. Sci Rep. 2024 Nov;14(1):27848. https://doi.org/10.1038/s41598-024-76093-7
26. Abbas JK, Hatem KK, Hussein WA. Use of artificial intelligence in diagnosis of various autoimmune diseases, a comprehensive literature review about current situation & future directions. Afr J Biomed Res. 2025 Jan;28(1s):952-61. https://doi.org/10.53555/AJBR.v28i1S.6291
27. White-Newsome JL, Brines SJ, Brown DG, Dvonch JT, Gronlund CJ, Zhang K, et al. Validating Satellite-Derived Land Surface Temperature with in Situ Measurements: A Public Health Perspective. Environ Health Perspect. 2013 Aug;121(8):925-31. https://doi.org/10.1289/ehp.1206176
28. Agache I, Akdis C, Akdis M, Al-Hemoud A, Annesi-Maesano I, Balmes J, et al. Immune-mediated disease caused by climate change-associated environmental hazards: mitigation and adaptation. Front Sci. 2024 Feb;2:1279192. https://doi.org/10.3389/fsci.2024.1279192
29. Tamang S. Public Health in the Age of AI: How Synthetic Patients Are Helping Uncover Environmental Drivers of Autoimmune & Rheumatic Disease. Stanford Medicine News [Internet]. 2025. Available from: https://med.stanford.edu/medicine/news/current-news/standard-news/synthetic-patients.html
30. Ku Y, Kwon SB, Yoon JH, Mun SK, Chang M. Machine learning models for predicting the occurrence of respiratory diseases using climatic and air-pollution factors. Clin Exp Otorhinolaryngol. 2022 May;15(2):168-76. https://doi.org/10.21053/ceo.2021.01536
31. Bae S, Kim HC, Ye B, Choi WJ, Hong YS, Ha M. Causal inference in environmental epidemiology. Environ Health Toxicol. 2017 Oct;32:e2017015. https://doi.org/10.5620/eht.e2017015
32. Amnuaylojaroen T, Parasin N. A Machine Learning Perspective on the Climatic and

Socioeconomic Determinants of Mental Health in Southeast Asia. World. 2025 Apr;6(2):48.
33. Casey JA, Daouda M, Babadi RS, Do V, Flores NM, Berzansky I, et al. Methods in public health environmental justice research: a scoping review from 2018 to 2021. Curr Environ Health Rep. 2023 Sep;10(3):312-36. https://doi.org/10.1007/s40572-023-00406-7
34. Cardona OD, Van Aalst MK, Birkmann J, Fordham M, Mc Gregor G, Rosa P, et al. Determinants of risk: exposure and vulnerability. In: Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, editors. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press; 2012. p. 65-108. https://doi.org/10.1017/CBO9781139177245.005
35. Miller FW, Alfredsson L, Costenbader KH, Kamen DL, Nelson LM, Norris JM, et al. Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop. J Autoimmun. 2012 Dec;39(4):259-71. https://doi.org/10.1016/j.jaut.2012.05.002
36. Abdullahi T, Nitschke G. Predicting disease outbreaks with climate data. ClimaHealth [Internet]. 2021. Available from: https://climahealth.info/resource-library/predicting-disease-outbreaks-with-climate-data/
37. Rosenbaum PR, Rubin DB. Propensity scores in the design of observational studies for causal effects. Biometrika. 2023 Mar;110(1):1-3. https://doi.org/10.1093/biomet/asac054
38. Zhao Y. The application of propensity score methods in observational studies. Front Appl Math Stat. 2024 Apr;10:1384217. https://doi.org/10.3389/fams.2024.1384217
39. VanderWeele TJ. Principles of confounder selection. Eur J Epidemiol. 2019 Mar;34(3):211-9. https://doi.org/10.1007/s10654-019-00494-6
40. Walter TG, Bricknell LK, Preston RG, Crawford EG. Climate change adaptation methods for public health prevention in Australia: an integrative review. Curr Environ Health Rep. 2024 Mar;11(1):71-87. https://doi.org/10.1007/s40572-023-00422-7
41. Bianco G, Espinoza-Chávez RM, Ashigbie PG, Junio H, Borhani C, Miles-Richardson S, et al. Projected impact of climate change on human health in low-and middle-income countries: a systematic review. BMJ Glob Health. 2024 Mar;8(Suppl 3):e015550. https://doi.org/10.1136/bmjgh-2024-015550
42. Yenew C, Bayeh GM, Gebeyehu AA, Enawgaw AS, Asmare ZA, Ejigu AG, et al. Scoping review on assessing climate-sensitive health risks. BMC Public Health. 2025 Mar;25(1):914. https://doi.org/10.1186/s12889-025-22148-x
43. Yadav N, Pandey S, Gupta A, Dudani P, Gupta S, Rangarajan K. Data Privacy in Healthcare: In the Era of Artificial Intelligence. Indian Dermatol Online J. 2023 Nov-Dec;14(6):788-92. https://doi.org/10.4103/idoj.idoj_543_23
44. Chaddad A, Lu Q, Li J, Katib Y, Kateb R, Tanougast C, et al. Explainable, domain-adaptive, and federated artificial intelligence in medicine. IEEE/CAA J Autom Sin. 2023 Apr;10(4):859-76. https://doi.org/10.48550/arXiv.2211.09317
45. Skevaki C, Nadeau KC, Rothenberg ME, Alahmad B, Mmbaga BT, Masenga GG, et al. Impact of climate change on immune responses and barrier defense. J Allergy Clin Immunol. 2024 May;153(5):1194-205. https://doi.org/10.1016/j.jaci.2024.01.016