ORIGINAL ARTICLE

EAACI

Data-driven research on eczema: Systematic characterization of the field and recommendations for the future

Ariane Duverdier[1,2,3] ⦿ | Adnan Custovic[4] ⦿ | Reiko J. Tanaka[2] ⦿

[1]Department of Computing, Imperial College London, London, UK

[2]Department of Bioengineering, Imperial College London, London, UK

[3]UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, UK

[4]National Heart and Lung Institute, Imperial College London, London, UK

Correspondence
Reiko J. Tanaka, Department of Bioengineering, Imperial College London, London, UK.
Email: r.tanaka@imperial.ac.uk

Abstract

Background: The past decade has seen a substantial rise in the employment of modern data-driven methods to study atopic dermatitis (AD)/eczema. The objective of this study is to summarise the past and future of data-driven AD research, and identify areas in the field that would benefit from the application of these methods.

Methods: We retrieved the publications that applied multivariate statistics (MS), artificial intelligence (AI, including machine learning-ML), and Bayesian statistics (BS) to AD and eczema research from the SCOPUS database over the last 50 years. We conducted a bibliometric analysis to highlight the publication trends and conceptual knowledge structure of the field, and applied topic modelling to retrieve the key topics in the literature.

Results: Five key themes of data-driven research on AD and eczema were identified: (1) allergic co-morbidities, (2) image analysis and classification, (3) disaggregation, (4) quality of life and treatment response, and (5) risk factors and prevalence. ML&AI methods mapped to studies investigating quality of life, prevalence, risk factors, allergic co-morbidities and disaggregation of AD/eczema, but seldom in studies of therapies. MS was employed evenly between the topics, particularly in studies on risk factors and prevalence. BS was focused on three key topics: treatment, risk factors and allergy. The use of AD or eczema terms was not uniform, with studies applying ML&AI methods using the term eczema more often. Within MS, papers using cluster and factor analysis were often only identified with the term AD. In contrast, those using logistic regression and latent class/transition models were "eczema" papers.

Conclusions: Research areas that could benefit from the application of data-driven methods include the study of the pathogenesis of the condition and related risk factors, its disaggregation into validated subtypes, and personalised severity management and prognosis. We highlight BS as a new and promising approach in AD and eczema research.

KEYWORDS
artificial intelligence, atopic dermatitis, bibliometric analysis, statistics

1 | BACKGROUND

Atopic dermatitis (AD, also referred to as eczema or atopic eczema) is a common chronic inflammatory skin disease that affects approximately 20% of children and 10% of adults in high-income countries.[1] Recently, computational modelling[2] and data-driven analytical methods have emerged as powerful new approaches to AD research, especially to elucidate its complex pathophysiology,[3] patient-dependent response to treatment,[4,5] and endotypes or subtypes.[6–11]

Big data have revolutionized the way we study disease.[12] The increased availability of large and diverse medical datasets has favoured the adoption of modern computational methods which can integrate and interrogate large quantities of data and extract hidden patterns and associations. There are three primary analytic methodologies or disciplines for data-driven research: multivariate statistics (MS), Bayesian statistics (BS), and machine learning and other artificial intelligence methods (ML&AI). MS encompasses methods that analyse datasets with multiple independent and/or dependent variables,[13] which is a key characteristic of biomedical datasets thereby making MS a popular and powerful methodology. AI is a field concerned with building systems that can mimic human intelligence, and ML is a subfield of AI. Finally, BS allows us to combine prior knowledge and observed data,[14] contrasting the frequentist approach which bases its analysis only on the observed data,[12,15] and is a potentially promising approach to develop predictive models and utilize clinical data. Such data-driven approaches have been applied to identify biomarkers to diagnose disease and identify therapeutic targets.[12,16,17] Deep neural networks have been developed to aid in the detection and diagnosis of skin,[18] breast,[19] and prostate[20] cancer. In AD research, the Bayesian mechanistic model recently developed by Hurault et al.[21] can predict individual patients' next-day AD severity scores from their previous severity scores and treatments applied. These examples illustrate the benefit of employing a data-driven approach in medical fields with a growing quantity of data.

Within the AD community, data collection is increasing, providing a unique opportunity to leverage data-driven methods.[2] As we enter a period of further substantial growth in the employment of data-driven methods to study AD, we aimed to identify the areas in AD research where data-driven methods have been applied, their current state of development, and highlight the knowledge gaps in the field that could benefit from the application of these methods. To address our aim, we conducted a bibliometric analysis highlighting the publication trends and conceptual knowledge structure of data-driven research on AD and eczema, and applied topic modelling to retrieve the key topics present within the literature. Bibliometrics uses statistical tools to study publication trends and patterns within an area of research,[22,23] and can be used to summarise a field of research in a systematic and reproducible manner. Probabilistic topic modelling explores the knowledge structure of a field by identifying the latent thematic structure of a corpus of documents.[24] A bibliometric analysis was previously conducted to understand the knowledge structure and theme trends of AD research[25] but it considered publications with the term AD from 2015 to 2019 and did not focus on data-driven research. Also of note, the continued absence of a consensus in nomenclature has resulted in the co-existence of two main terms for the skin condition, AD and eczema, which have been shown to be linked to different findings and biased to different disciplines.[26] Our study included both AD and eczema terms and retrieved all publications available up to March 2021 without a time constraint, to provide the full picture of the field. Additionally, we included topic modelling to provide a detailed view of the key research topics in the field and methods employed.

2 | METHODS

This section summarises the analysis conducted in this paper; detailed description is presented in Appendix S1.

2.1 | Literature search

We retrieved all publications to March 17th, 2021, on atopic dermatitis (AD) and eczema that apply MS, ML&AI, and BS methodologies from the SCOPUS database. The keywords 'AD' and 'eczema' were used with each of MS, ML&AI, and BS methodologies.

2.2 | Bibliometric analysis

We performed a bibliometric analysis on the bibliographic information (including the authors, sources, countries, citations, and keywords) of the publications obtained from the literature search. Using the *bibliometrix* R package,[27] we obtained descriptive statistics on the collection of publications, including the most productive countries and the general publication trends. We also performed co-word analysis to produce keyword co-occurrence networks and thematic maps.

2.3 | Probabilistic topic modelling

We used the Latent Dirichlet allocation (LDA) algorithm[28] to explore the main topics present in the publications obtained by the literature search. LDA is an unsupervised ML method that estimates both the distribution of topics within each document and the distribution of words within each topic, by assuming each document consists of a mixture of topics and each topic consists of a mixture of words. Here, each document consisted of the title, keywords, and abstract. We used the *tm* R package[29] to clean the data (tokenization, lowercase conversion, removal of special characters and stop-words, standardization of words) and remove words with low frequency (words that occurred in less than 10 publications), the *topicmodels* R package[30] to run the LDA algorithm on the corpus, and generated the plots of the results using R packages such as *ggplot2*[31] and *wordcloud*.[32]

3 | RESULTS

3.1 | Publication trends of data-driven research on AD and eczema

620 unique articles, published between September 1973 and March 2021, were retrieved from the SCOPUS database on the application of MS, ML&AI and BS to AD/eczema research. Of these, 474 articles employed MS methods, 150 ML&AI, and 37 BS (Table S1, Figure S1). Publications increased over time, with most papers published in the past decade (473/620, 76.3%) (Figure 1). The earliest application of MS in AD and eczema research dates to 1973, ML&AI made its first appearance in 1996, and BS in 2001 (Figure 1A). The application of ML&AI has seen a particular emergence in the past 10 years, with a substantial increase since 2019, approaching the popularity of MS. BS is the least commonly used of the three methodologies to date. More details on publication trends are presented in Appendix S2.

Most publications are labelled as either AD or eczema, and only a small portion are annotated with both terms (100 of 620 articles). This phenomenon is similarly found within the individual methodologies (Table S1). Publication numbers for each term are similar throughout the years, showing at first glance no significant frequency preference of the field in general for one term over the other (Figure 1B). Geographical distribution of manuscripts is shown in Table S2.

3.2 | Analytical methods and the use of AD or eczema terms

Table 1 summarizes the key methods used within the collection of publications. Cluster and factor analysis are the two most common methods. Of the 37 BS papers, a manual inspection found that only eight[21,33–39] specifically study AD. Of these eight, half[33–36] used random-effects Bayesian network meta-analysis to compare treatments for AD, and one[21] uses a Bayesian mechanistic machine learning model to predict next-day AD severity for individual patients.

The use of AD and/or eczema terms is not uniform throughout the different methods. Detailed analysis is presented in the Appendix S3. Briefly, papers applying ML&AI methods use the term eczema more often. Within MS, papers using cluster and factor analysis are often only identified with the term AD. In contrast, those using logistic regression and latent class/transition models are eczema papers.

3.3 | Five central themes of data-driven AD and eczema research and their level of development

The bibliometric analysis identified five key themes within AD/eczema research employing MS, BS, and ML&AI methods, as visualized in a thematic map (Figure 2), where themes are mapped onto a two-dimensional space according to their centrality and density. The centrality is the degree of interaction of the theme with other themes and measures the significance or relevance of a theme in the development of the field at large.[40] The density measures the development of the theme.[40] Using these two measures, themes can be separated into four quadrants: emerging or declining themes (low centrality and density), niche themes (low centrality and high density), motor themes (high centrality and density), and basic themes (high centrality and low density).[40] We named the five identified themes retroactively, ordered by decreasing density:

Theme 1 *Allergic co-morbidities. This theme includes articles that study atopic/allergic diseases and the development of IgE sensitisation in childhood using analytical methods such as cluster analysis, latent class analysis, and AI. It is a motor theme, highly relevant and already developed (orange in Figure 2).*

Theme 2 *Image analysis and classification. This theme includes articles that use ML methods, including deep learning and support vector*
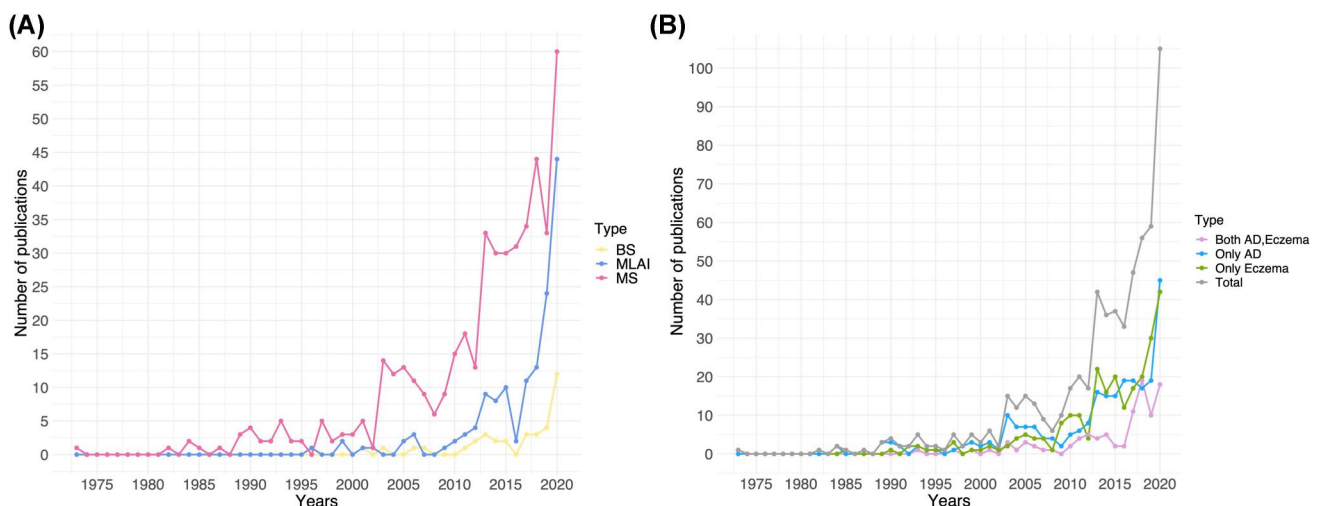

FIGURE 1 Papers published per year separated by (A) methodology and (B) term

TABLE 1 Distribution of the main methodologies used in data-driven eczema and atopic dermatitis (AD) publications

		Number of publications			
Discipline	Methodology	Total (of 620)	Only eczema (of 255)	Only AD (of 265)	Both (of 100)
Multivariate statistics (MS)	Cluster analysis	206	84	**90**	32
	Factor analysis	89	26	**46**	17
	Logistic regression	56	**24**	17	15
	Latent class/Transition models	55	**31**	12	12
	Principal component analysis	46	17	**20**	9
	Discriminant analysis	28	8	**16**	4
	Markov model	28	8	**12**	8
	Structural equation modelling	11	**5**	4	2
	Mixture model	7	3	**4**	0
	Correspondence analysis	3	**2**	1	0
	Latent variable model	3	**2**	1	0
	Canonical correlation	1	0	**1**	0
Machine learning and artificial intelligence (ML&AI)	Artificial neural networks (including CNNs)	67	**44**	21	2
	Machine learning	48	**25**	17	6
	Support vector machine	36	**24**	11	1
	Artificial intelligence	17	**8**	7	2
	Decision trees	17	**7**	6	4
	Deep learning	13	**9**	3	1
	Natural language processing	12	**7**	5	0
	Random forests	9	**6**	2	1
	Supervised learning	2	**2**	0	0
	Unsupervised learning	1	**1**	0	0
Bayesian statistics (BS)	Bayesian framework	14	**8**	5	1
	Bayesian network	5	**4**	1	0
	Random-effects Bayesian network meta-analysis	4	0	1	3
	Bayesian machine learning model	3	**2**	0	1
	Bayesian spatial and temporal models	3	1	**2**	0
	Naïve Bayesian classifier	2	**2**	0	0
	Bayesian meta-regression	2	0	1	1
	Bayesian model averaging	2	**2**	0	0
	Bayesian latent class analysis	1	**1**	0	0
	Random-effects Bayesian hierarchical model	1	0	**1**	0

Note: Search strings representing the methodologies were searched for in the title, abstract, and keywords of the multivariate statistics (MS) and machine learning and artificial intelligence (ML&AI) publications. Methods for Bayesian statistics (BS) were determined manually. Number of publications are given for the total collection and additionally separated according to the term used, only eczema, only AD, or both. For each method, the highest number of publications between only eczema and only AD is bolded.
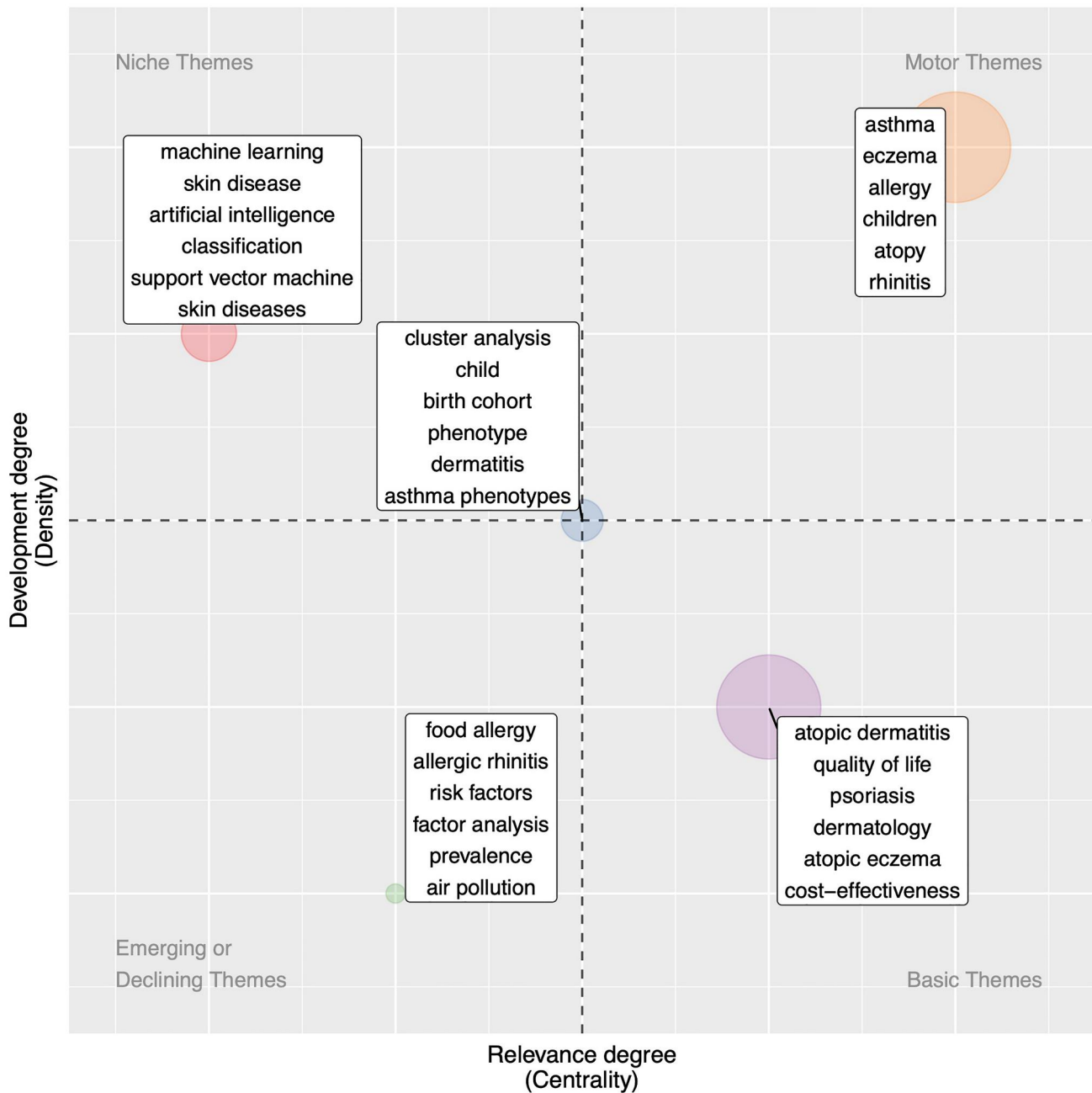
FIGURE 2 Thematic map. Themes were generated using the top 100 authors' keywords and separated according to centrality (the degree of interaction of the theme with other themes) and density (the strength of internal connections among keywords in the theme). Up to six of the most frequent keywords in the associated theme are shown on the map

machines, to process, segment, and classify images of AD/eczema and other skin diseases. It is a niche theme with a high development degree but is less central than other themes in the field (red in Figure 2).

Theme 3 *Disaggregation of the condition*. This theme tackles the issue of disentangling the complex pathophysiology of AD/eczema and includes studies that consider biomarkers to investigate endotypes and methods such as cluster analysis to disaggregate phenotypes.

The theme also contains articles on disaggregating asthma phenotypes. Compared to other themes, it has middle relevance and development (blue in Figure 2).

Theme 4 *Quality of life and treatment response*. This theme includes studies investigating the quality of life and the cost-effectiveness of treatments, not specific to only AD/eczema but also for other skin conditions such as psoriasis. It is a basic theme with relatively low development but high relevance (purple in Figure 2).
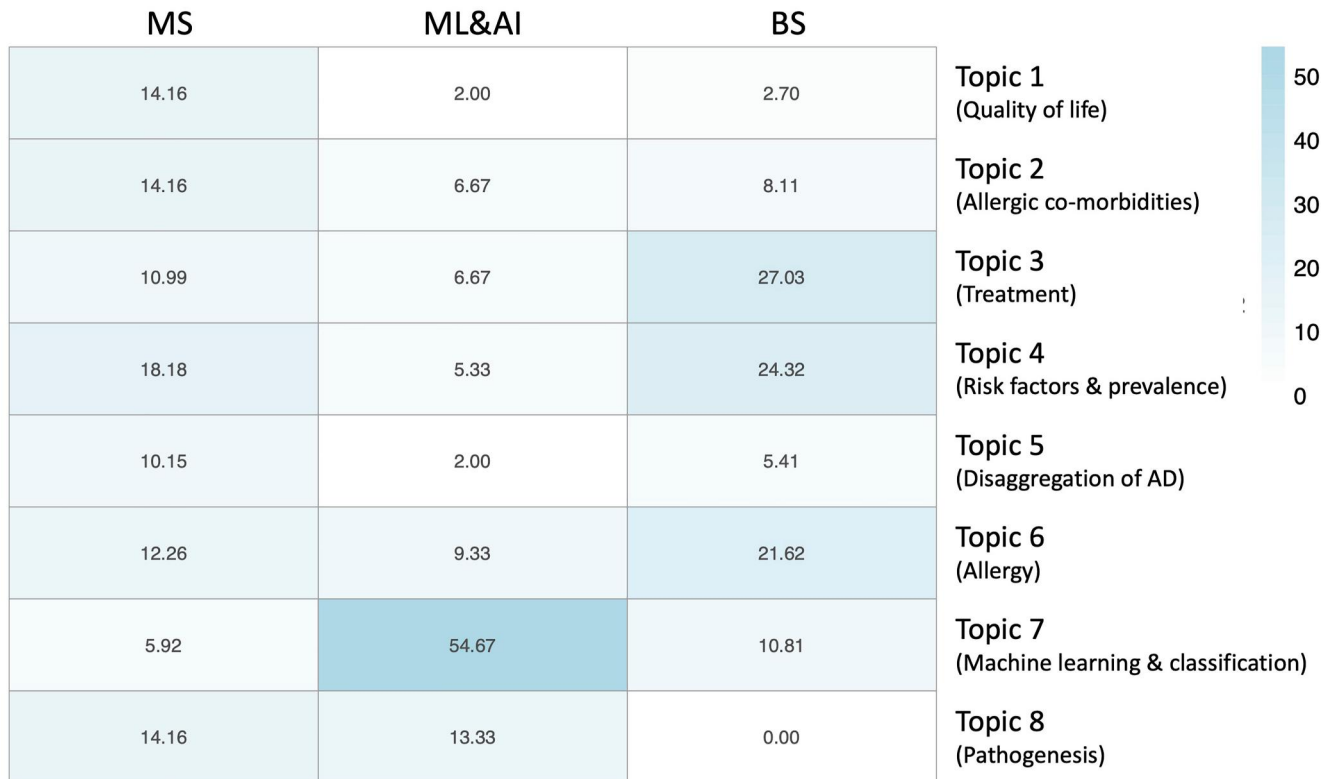
Theme 5 *Risk factors and prevalence. This theme looks at potential factors that increase the likelihood or prevalence of AD/eczema, including air pollution. It also studies the relationship with co-morbidities, including allergic rhinitis and food allergy. Methods used in this theme include factor analysis. It is an emerging theme with low development and relevance (green in Figure 2).*

Thematic maps were also generated for the three methodologies and the term used (eczema or AD), Figure 2.

3.4 | Eight key topics and identified gaps in employing modern computational methodologies

The LDA algorithm revealed eight key topics of data-driven AD and eczema research (Figure 3, Table S3), by breaking down the five themes obtained in the bibliometric analysis into their main components. It identified, in greater detail, the key areas of interest explored in the literature to date (Table S4) and their growth over time (Figure S3).

Theme 1 is represented by Topics 2 and 6, which respectively study allergic co-morbidities and the development of allergy and sensitisation. Theme 2 is represented in Topic 7 on the application of ML methods for the classification and diagnosis of skin diseases. Theme 3 is broken down into two key topics, Topics 5 and 8. Topic 5

includes studies on the disaggregation of AD and the role of the skin microbiome, especially the presence of *Staphylococcus aureus*. Topic 8 regards the pathogenesis of the condition, looking at genetic, inflammatory and immune biomarkers, and mechanisms underlying development and progression. Theme 4 encompasses Topics 1 and 3, which study AD/eczema symptoms and their management. Topic 1 includes publications on disease severity and the effect on quality of life. Topic 3 encompasses therapeutic studies, including efficacy and cost-effectiveness analyses. Finally, Theme 5 is reflected in Topic 4, which considers the prevalence of eczema and associated risk factors, including environmental exposure and parental history of atopy.

Most publications that employ ML&AI are found in Topic 7 (Figure 4). In contrast, topics including quality of life, disaggregation of AD, risk factors and prevalence, allergic co-morbidities, and treatment response see a low application of ML&AI methods.

MS is employed fairly evenly between the topics. Its most popular topic is Topic four on risk factors and prevalence.

BS focuses on three key topics: treatment, risk factors and prevalence, and allergy.

Topics on quality of life, treatment, and disaggregation of the condition contain 'atopic' and 'dermatitis' as two of the most relevant keywords, while 'eczema' is found as the top keyword in studies on risk factors and prevalence (Figure 4 and Figure S4). This reflects the trend seen in the bibliometric analysis that the term eczema tends to be used in publications that also study other allergic diseases

FIGURE 3 Word clouds for the eight topics obtained by Latent Dirichlet allocation (LDA). The topic names were retroactively chosen to best summarize the content of topics. The 40 most probable words in each topic are plotted with the size of the words proportional to their probability

MS	ML&AI	BS	
14.16	2.00	2.70	**Topic 1** (Quality of life)
14.16	6.67	8.11	**Topic 2** (Allergic co-morbidities)
10.99	6.67	27.03	**Topic 3** (Treatment)
18.18	5.33	24.32	**Topic 4** (Risk factors & prevalence)
10.15	2.00	5.41	**Topic 5** (Disaggregation of AD)
12.26	9.33	21.62	**Topic 6** (Allergy)
5.92	54.67	10.81	**Topic 7** (Machine learning & classification)
14.16	13.33	0.00	**Topic 8** (Pathogenesis)

FIGURE 4 Distribution of the application of multivariate statistics (MS), machine learning and artificial intelligence (ML&AI), and Bayesian statistics (BS) methodologies in the eight Latent Dirichlet allocation (LDA) topics. Each publication is assigned to its most probable topic

4 | DISCUSSION

The first application of data-driven methods to AD and eczema research occurred in September 1973. Since then, 620 articles have been published, with over three-fourths of the publications in the last decade. The growth in scientific production over time shows an increased interest in applying data-driven methodologies to the study of AD and eczema, similar to asthma research.[12,15]

Five central themes currently characterize the field: (1) allergic co-morbidities, (2) image analysis and classification, (3) disaggregation of the condition, (4) quality of life and treatment response, and (5) risk factors and prevalence. In 2020, Theme 2 was the most popular topic in AD and eczema research today. Theme 3 has the third-highest number of publications, indicating the continued need to delineate developmental trajectories and disease mechanisms and the subsequent characterization and validation of endotypes before these can be used and implemented in other areas of research. Theme 4 is central to the field, but not highly researched to date, suggesting the continued need to apply data-driven methods to help build personalised severity prediction tools and treatment strategies. Theme 5 has the lowest number of publications; as more data is collected on potential risk factors, data-driven tools could be

(Figure S5). In contrast, AD term is used often in publications that are more specific to the condition.

leveraged to evaluate their relevance. Mu et al.[25] conducted a bibliometric analysis on AD literature from 2015 to 2019 and similarly found that themes on quality of life, prevention and control, and pathology were undeveloped. This reinforces the need for further research, particularly employing data-driven methodologies, in these key areas.

We identified a substantial increase in ML&AI publications over the past five years in AD and eczema research. Most studies that apply ML in dermatology address classifying skin lesions and primarily rely on convolutional neural networks for image recognition and classification.[41] Over half of the AD and eczema publications employing ML&AI methods are found in Theme 2 on image analysis and classification. The low application of ML&AI in the rest of the themes, including the study of the pathogenesis of the condition, disaggregation, risk factors, quality of life, treatment, and the role of allergic co-morbidities, highlights potential areas that could benefit from the application of ML&AI methods.

Bayesian approaches have been used to study asthma[12,15] and the relationship between allergic diseases.[42,43] However, only eight publications to date apply BS to study AD and eczema specifically. One of the eight developed a Bayesian mechanistic model that can predict next-day AD severity from patients' previous severity scores and treatments applied.[21] The Bayesian approach allows consistent quantification of uncertainty within parameter estimates and predictions and the incorporation of prior knowledge or data from previous studies. This highlights BS as a new and promising approach

in AD and eczema research, particularly to develop predictive models and use clinical data.

The analysis performed in this study corroborates the discrepancy in the use of AD and eczema terms within the literature that has been highlighted in previous studies.[26,44] Our results point towards a bias in term use depending on the computational method employed; this alludes to the previously articulated notion that AD and eczema terms may be associated to different research communities that have differing views on nomenclature (Appendix S3).

The main limitation of our analysis is that it is heavily dependent on the terminology used by the authors. The authors' keywords associated to each publication were used to discover the key themes of the field of research; they were also used, in part, to retrieve publications of interest. This points to the importance of keyword choice when publishing a paper and the impact of using eczema or AD terms. A second limitation is that the LDA algorithm was applied on each publication's title, keywords, and abstract, but not the full text as they were not available. Additionally, the publications were retrieved from the SCOPUS database. Although similar in content to that found on PubMed, future systematic reviews could aggregate the publications from multiple databases to ensure completeness of the collection of articles analysed.

Three key areas that could benefit from the application of data-driven approaches are the study of the disaggregation of the condition, quality of life and treatment response, and risk factors and prevalence. One of the greatest challenges for research in these areas regards data curation, particularly its collection and sharing. The study of the course of the condition, including its onset, persistence, and flare-ups, and the design of personalised treatment strategies would be greatly aided by additional longitudinal data. Previous studies have showcased the benefit and need of such data[21,42,43] and new smartphone apps could facilitate the collection of data outside of a clinical visit. The sharing of data is also crucial, as AD is a complex disease that cannot be fully characterized in a single study. It would be greatly aided by a collaborative system to share and manage data from different studies across the community.

Our study particularly underlines the need for standardized data collection, including a clear and detailed record of the criteria for diagnostics and patient selection in clinical studies to allow for proper comparison between studies. A recent study highlighted the impact of different definitions of AD in prevalence estimates, risk factors, and the performance of predictive models.[47] Further work demonstrated that development of standardized composite definitions of AD derived from multiple sources (healthcare records and validated questionnaires) may help to define AD cases with more precision.[48] A data collection tool or unified database would be particularly useful for data sharing and to ensure high quality and quantity of data needed for the proper employment of statistical methods. We may take example on similar fields of study, including asthma research,[49] to guide the next steps.

As the development and employment of machine learning and other data-driven approaches gain popularity in healthcare, experts and government agencies have increasingly collaborated to develop guidelines to facilitate the growth of the field and delineate principles of best practice.[45,46] We further underline the need and benefit of cross-disciplinary collaborations for the future of data-driven research on AD and eczema.[12]

5 | CONCLUSIONS

A bibliometric analysis of AD research published until March 2021 provides a guide for the AD community on the past and future of data-driven AD/eczema research. Data-driven methods in AD/eczema research have mainly been applied to the study of (1) allergic co-morbidities, (2) image analysis and classification, (3) disaggregation of the condition, (4) quality of life and treatment response, and (5) risk factors and prevalence. The last three of these could benefit from the further application of modern analytical methods. We additionally highlight BS as a new and promising approach in AD and eczema research.

We emphasise the importance in forming collaborations between statisticians/mathematicians and clinicians to address clinical needs and to collect/provide crucially important data. With the emergence of data-driven approaches, it is very important for clinicians to stay informed with the general concepts of new methods and their impact on the clinical studies. It is additionally crucial to keep up to date on literature using both eczema and AD keywords not to miss important findings. A computational tool based on the decision tree to disambiguate AD/eczema-related queries for both researchers and clinicians has recently been published as an open-source code, and we highlight its availability.[26]

AUTHOR CONTRIBUTIONS

Ariane Duverdier: Conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing-original draft. **Adnan Custovic and Reiko J. Tanaka:** Conceptualization, funding acquisition, project administration, resources, supervision, validation, writing-original draft, writing-review & editing.

CONFLICT OF INTEREST

Nothing to declare.

ORCID

Ariane Duverdier https://orcid.org/0000-0003-0132-3525
Adnan Custovic https://orcid.org/0000-0001-5218-7071
Reiko J. Tanaka https://orcid.org/0000-0002-0769-9382

REFERENCES

1. Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *Lancet*. 2020 (10247):360-396. https://doi.org/10.1016/S0140-6736(20)31286-1
2. Eyerich K, Brown SJ, Perez White BE, et al. Human and computational models of atopic dermatitis: a review and perspectives by an expert panel of the International Eczema Council.

J Allergy Clin Immunol. 2019;143(1):36-45. https://doi.org/10.1016/j.jaci.2018.10.033

3. Dominguez-Huttinger E, Christodoulides P, Miyauchi K, et al. Mathematical modeling of atopic dermatitis reveals "double-switch" mechanisms underlying 4 common disease phenotypes. *J Allergy Clin Immunol.* 2017;139(6):1861-1872. https://doi.org/10.1016/j.jaci.2016.10.026

4. Miyano T, Irvine AD, Tanaka RJ. A mathematical model to identify optimal combinations of drug targets for dupilumab poor responders in atopic dermatitis. *Allergy.* 2021;77(2):582-594. https://doi.org/10.1111/all.14870

5. Tanaka G, Domínguez-Hüttinger E, Christodoulides P, Aihara K, Tanaka RJ. Bifurcation analysis of a mathematical model of atopic dermatitis to determine patient-specific effects of treatments on dynamic phenotypes. *J Theor Biol.* 2018;448:66-79. https://doi.org/10.1016/j.jtbi.2018.04.002

6. Martel BC, Litman T, Hald A, et al. Distinct molecular signatures of mild extrinsic and intrinsic atopic dermatitis. *Exp Dermatol.* 2016;25(6):453-459. https://doi.org/10.1111/exd.12967

7. Thijs JL, Strickland I, Bruijnzeel-Koomen CAFM, et al. Moving toward endotypes in atopic dermatitis: identification of patient clusters based on serum biomarker analysis. *J Allergy Clin Immunol.* 2017;140(3):730-737. https://doi.org/10.1016/j.jaci.2017.03.023

8. Leonard A, Wang J, Yu L, et al. Atopic dermatitis endotypes based on allergen sensitization, reactivity to Staphylococcus aureus antigens, and underlying systemic inflammation. *J Allergy Clin Immunol Pract.* 2019;8(1):236-247.e3. https://doi.org/10.1016/j.jaip.2019.08.013

9. Sanyal RD, Pavel AB, Glickman J, et al. Atopic dermatitis in African American patients is TH2/TH22-skewed with TH1/TH17 attenuation. *Ann Allergy Asthma Immunol.* 2019;122(1):99-110.e6. https://doi.org/10.1016/j.anai.2018.08.024

10. Leung DYM, Calatroni A, Zaramela LS, et al. The nonlesional skin surface distinguishes atopic dermatitis with food allergy as a unique endotype. *Sci Transl Med.* 2019;11(480). https://doi.org/10.1126/scitranslmed.aav2685

11. Noda S, Suárez-Fariñas M, Ungar B, et al. The Asian atopic dermatitis phenotype combines features of atopic dermatitis and psoriasis with increased TH17 polarization. *J Allergy Clin Immunol.* 2015;136(5):1254-1264. https://doi.org/10.1016/j.jaci.2015.08.015

12. Belgrave D, Henderson J, Simpson A, Buchan I, Bishop C, Custovic A. Disaggregating asthma: big investigation versus big data. *J Allergy Clin Immunol.* 2017;139(2):400-407. https://doi.org/10.1016/j.jaci.2016.11.003

13. Tabachnick BG, Fidell LS. In: Mosher J, Campanella C, Frail S, eds. *Using Multivariate Statistics.* 6th ed. Pearson Education; 2012.

14. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis.* 3rd ed.; 2021. https://users.aalto.fi/ave/BDA3.pdf

15. Fontanella S, Cucco A, Custovic A. Machine learning in asthma research: moving toward a more integrated approach. *Expet Rev Respir Med.* 2021;15(5):609-621. https://doi.org/10.1080/17476348.2021.1894133

16. Dimitrakopoulos C, Hindupur SK, Colombi M, et al. Multi-omics data integration reveals novel drug targets in hepatocellular carcinoma. *BMC Genomics.* 2021;22(1):592. https://doi.org/10.1186/s12864-021-07876-9

17. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinf Biol Insights.* 2020;14. https://doi.org/10.1177/1177932219899051

18. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118. https://doi.org/10.1038/nature21056

19. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. *Nature.* 2020;577(7788):89-94. https://doi.org/10.1038/s41586-019-1799-6

20. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digital Health.* 2020;2(8):e407-16. https://doi.org/10.1016/S2589-7500(20)30159-X

21. Hurault G, Dominguez-Huttinger E, Langan SM, Williams HC, Tanaka RJ. Personalized prediction of daily eczema severity scores using a mechanistic machine learning model. *Clin Exp Allergy.* 2020;50(11):1258-1266. https://doi.org/10.1111/cea.13717

22. McBurney MK, Novak PL. What is bibliometrics and why should you care. In *IEEE International Professional Communication Conference*; 2002:108-114. https://doi.org/10.1109/IPCC.2002.1049094

23. Cebrino J, Portero de la Cruz S. A worldwide bibliometric analysis of published literature on workplace violence in healthcare personnel. *PLOS ONE.* 2020;15(11):e0242781. https://doi.org/10.1371/journal.pone.0242781

24. Blei DM. Probabilistic topic models. *Commun ACM.* 2012;55(4):77-84. https://doi.org/10.1145/2133806.2133826. http://www.cs.columbia.edu/blei/papers/Blei2012.pdf

25. Mu Z, Zhang Y, Li L, Han X. Mapping knowledge structures and theme trends of atopic dermatitis: a co-word biclustering and quantitative analysis of the publication between 2015 and 2019. *J Dermatol Treat.* 2021:1-10. https://doi.org/10.1080/09546634.2021.1905769

26. Frainay C, Pitarch Y, Filippi S, Evangelou M, Custovic A. Atopic dermatitis or Eczema? Consequences of ambiguity in disease name for biomedical literature mining. *Clin Exp Allergy.* 2021;51(9):1185-1194. https://doi.org/10.1111/cea.13981

27. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. *J Inf.* 2017;11(4):959-975. https://doi.org/10.1016/j.joi.2017.08.007

28. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res.* 2003;3:993-1022. https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

29. Feinerer I, Hornik Ktm. *Text Mining Package*; 2020. https://CRAN.R-project.org/package=tm

30. Grün B, Hornik K. Topicmodels: an R package for fitting topic models. *J Stat Software.* 2011;40(13):1-30. https://doi.org/10.18637/jss.v040.i13

31. Wickham H. *ggplot2: Elegant Graphics for Data Analysis [Internet].* Springer-Verlag; 2016. https://ggplot2.tidyverse.org

32. Fellows I. *Wordcloud: Word Clouds*; 2018. https://CRAN.R-project.org/package=wordcloud

33. Gómez-Escobar LG, Mora-Ochoa H, Vargas Villanueva A, et al. Effectiveness and adverse events of topical and allergen immunotherapy for atopic dermatitis: a systematic review and network meta-analysis protocol. *Syst Rev.* 2020;9(1):222. https://doi.org/10.1186/s13643-020-01472-w

34. Drucker AM, Ellis AG, Bohdanowicz M, et al. Systemic immunomodulatory treatments for patients with atopic dermatitis. *JAMA Dermatol.* 2020;156(6):659. https://doi.org/10.1001/jamadermatol.2020.0796

35. Drucker AM, Ellis A, Jabbar-Lopez Z, et al. Systemic immunomodulatory treatments for atopic dermatitis: protocol for a systematic review with network meta-analysis. *BMJ Open.* 2018;8:e023061. https://doi.org/10.1136/bmjopen-2018-023061

36. Liu L, Luo Y, Zhou M, et al. Tripterygium agents for the treatment of atopic eczema: a Bayesian analysis of randomized controlled trials. *Phytomedicine.* 2019;59:152914. https://doi.org/10.1016/j.phymed.2019.152914

37. Nisar H, Ch'ng YK, Ho YK. Automatic segmentation and classification of eczema skin lesions using supervised learning *IEEE Conference on Open Systems (ICOS).* IEEE; 2020:25-30. https://doi.org/10.1109/ICOS50156.2020.9293657

38. Wu Y-Y, Tang J-P, Liu Q, et al. Scanning indels in the 5q22.1 region and identification of the TMEM232 susceptibility gene

DUVERDIER ET AL.

that is associated with atopic dermatitis in the Chinese Han population. *Gene.* 2017;617:17-23. https://doi.org/10.1016/j.gene.2017.03.034

39. Yang S-I, Lee S-H, Lee S-Y, et al. Prenatal PM2.5 exposure and vitamin D–associated early persistent atopic dermatitis via placental methylation. *Ann Allergy Asthma Immunol.* 2020;125(6):665-673e1. https://doi.org/10.1016/j.anai.2020.09.008

40. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. An approach for detecting, quantifying, and visualizing the evolution of a research field: a practical application to the Fuzzy Sets Theory field. *J Inf.* 2011;5(1):146-166. https://doi.org/10.1016/j.joi.2010.10.002

41. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine learning in dermatology: current applications, opportunities, and limitations. *Dermatol Ther.* 2020;10(3):365-386. https://doi.org/10.1007/s13555-020-00372-0

42. Belgrave DCM, Granell R, Simpson A, et al. Developmental profiles of eczema, wheeze, and rhinitis: two population-based birth cohort studies. *PLoS Med.* 2014;11(10):e1001748. https://doi.org/10.1371/journal.pmed.1001748

43. Custovic A, Custovic D, Kljaić Bukvić B, Fontanella S, Haider S. Atopic phenotypes and their implication in the atopic march. *Expet Rev Clin Immunol.* 2020;16(9):873-881. https://doi.org/10.1080/1744666X.2020.1816825

44. Kantor R, Thyssen JP, Paller AS, Silverberg JI. Atopic dermatitis, atopic eczema, or eczema? A systematic review, meta-analysis, and recommendation for uniform use of "atopic dermatitis. *Allergy.* 2016;71(10):1480-1485. https://doi.org/10.1111/all.12982

45. Nakamura T, Haider S, Colicino S, et al. Different definitions of atopic dermatitis: impact on prevalence estimates and associated risk factors. *Br J Dermatol.* 2019;181(6):1272-1279. https://doi.org/10.1111/bjd.17853

46. Nakamura T, Haider S, Fontanella S, Murray CS, Simpson A, Custovic A. Modelling trajectories of parentally reported and physician-confirmed atopic dermatitis in a birth cohort study. *Br J Dermatol.* 2022;186(2):274-284. https://doi.org/10.1111/bjd.20767

47. Bulathsinhala L, Eleangovan N, Heaney LG, et al. Development of the international severe asthma registry (ISAR): a modified delphi study. *J Allergy Clin Immunol Pract.* 2019;7(2):578-588. https://doi.org/10.1016/j.jaip.2018.08.016

48. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med.* 2019;25(9):1337-1340. https://doi.org/10.1038/s41591-019-0548-6

49. Medicines and Healthcare products Regulatory Agency. *Good Machine Learning Practice for Medical Device Development: Guiding Principles*; 2021. [cited 2022 Jan 11]. https://www.gov.uk/government/publications/good-machine-learning-practice-for-medical-device-development-guiding-principles

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.