Received: 27 April 2024 | Accepted: 19 July 2024

DOI: 10.1111/jdv.20286

JEADV JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

SYSTEMATIC REVIEW

Natural language processing in dermatology: A systematic literature review and state of the art

Alessia Paganelli[1] | **Marco Spadafora**[2,3] | **Cristian Navarrete-Dechent**[4] | **Stefania Guida**[5] | **Giovanni Pellacani**[6] | **Caterina Longo**[2,3]

[1]Dermatology Unit, Azienda Unità Sanitaria Locale—IRCCS di Reggio Emilia, Reggio Emilia, Italy

[2]University of Modena and Reggio Emilia, Modena, Italy

[3]Azienda Unità Sanitaria Locale—IRCCS di Reggio Emilia, Skin Cancer Center, Reggio Emilia, Italy

[4]Melanoma and Skin Cancer Unit, Department of Dermatology, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

[5]Dermatology Clinic, IRCCS San Raffaele Scientific Institute, Milan, Italy

[6]Department of Clinical Internal, Anesthesiological and Cardiovascular Sciences, Dermatology Clinic, Sapienza University of Rome, Rome, Italy

Correspondence
Caterina Longo, Skin Cancer Center, Azienda Unità Sanitaria Locale—IRCCS di Reggio Emilia, viale Risorgimento 80, 42123 Reggio Emilia (RE), Italy.
Email: caterina.longo@unimore.it

Abstract

Background: Natural Language Processing (NLP) is a field of both computational linguistics and artificial intelligence (AI) dedicated to analysis and interpretation of human language.

Objectives: This systematic review aims at exploring all the possible applications of NLP techniques in the dermatological setting.

Methods: Extensive search on 'natural language processing' and 'dermatology' was performed on MEDLINE and Scopus electronic databases. Only journal articles with full text electronically available and English translation were considered. The PICO (Population, Intervention or exposure, Comparison, Outcome) algorithm was applied to our study protocol.

Results: Natural Language Processing (NLP) techniques have been utilized across various dermatological domains, including atopic dermatitis, acne/rosacea, skin infections, non-melanoma skin cancers (NMSCs), melanoma and skincare. There is versatility of NLP in data extraction from diverse sources such as electronic health records (EHRs), social media platforms and online forums. We found extensive utilization of NLP techniques across diverse dermatological domains, showcasing its potential in extracting valuable insights from various sources and informing diagnosis, treatment optimization, patient preferences and unmet needs in dermatological research and clinical practice.

Conclusions: While NLP shows promise in enhancing dermatological research and clinical practice, challenges such as data quality, ambiguity, lack of standardization and privacy concerns necessitate careful consideration. Collaborative efforts between dermatologists, data scientists and ethicists are essential for addressing these challenges and maximizing the potential of NLP in dermatology.

INTRODUCTION

Natural Language Processing (NLP) is a subdomain of both computational linguistics and artificial intelligence (AI) that involves the development of computer algorithms for the analysis of human language.[1] The main objective of NLP is the generation of data to be used in various specific settings and tasks with text processing typically considered its main function. However, NLP encompasses a wide range of other tasks, including syntax and semantics analysis, translation, summarization, entity recognition, sentiment analysis and question answering.[2] All these functions give the reason for the wide range of potential applications of NLP, ranging from search engines to virtual assistants, chatbots and language translation services.

In the last few years, NLP technology has successfully been applied to the medical setting, determining a revolution not only in healthcare delivery, but also impacting

academics, scientific literature, medical education, research and health administration.[3] NLP, in fact, holds the potential to redefine the way medical professionals interact with vast amounts of textual data generated in healthcare settings.[4] With its ability to analyse human language, NLP facilitates the extraction of information from electronic health records, medical literature, clinical notes and other unstructured data sources.[5,6] The use of NLP in the medical setting is aimed at improving clinical decision-making, simplifying administrative processes and advancing medical research.

The aim of the present study is to systematically review the available literature on the use of NLPs, focusing on the dermatological setting.

Key points

Why was the study undertaken?

- The aim of our study was to systematically analyse the potential applications of natural language processing (NLP) techniques in the dermatological setting.

What does this study add?

- The present study underscores the potential of NLP in extracting clinical data for improving diagnosis and treatment, as well as for the identification of patient unmet needs.

What are the implications of this study for disease understanding and/or clinical care?

- The refinement of NLP-based technologies not only fastens diagnostic-therapeutic processes for patients with skin disorders but also provides a new valid method for data collection in the setting of dermatological research.

MATERIALS AND METHODS

This systematic review was based on the PRISMA guidelines (see Data S1). A flow chart diagram is presented in Figure 1.

Search strategy

A search was conducted in the MEDLINE and Scopus electronic databases from inception to present (up to March 2024). The detailed search strategy for MEDLINE (PubMed) used the following terms: 'natural language processing' [Title/Abstract] AND 'dermatology' [Title/Abstract] OR 'skin' [Title/Abstract]. The terms were adapted for the other databases as appropriate. All the major journals were indexed. Only journal articles were taken into consideration, while books and book chapters were excluded. Articles without full text electronically available and/or English translation were also excluded. Review articles and conference papers were not included. Only journal articles specifically focusing on the use of NLP in the dermatological setting were taken into consideration. Papers on other forms of AI, or on medical branches other than dermatology were excluded.

Study population—Selection

The following PICO (Population, Intervention or exposure, Comparison, Outcome) elements were applied as inclusion criteria for the systematic review: (i) Population: patients affected by dermatological conditions, (ii) Intervention: use of NLP to extract diagnostic/prognostic/therapeutic information (either from electronic health records or from the web), (iii) Comparator: diagnosis/classification by clinicians, (iv) Outcome: NLP performance in providing diagnostic/prognostic/therapeutic information for dermatological diseases.

Data extraction

For studies fulfilling the inclusion criteria, two independent reviewers (A.P. and M.S.) extracted data in a standardized and predefined form. Disagreements were resolved by a third reviewer (C.L.). The following data was collected for each paper: first author, year of publication, disease(s) of interest, source data, and main outcomes (Table 1). All co-authors reviewed the final table and figures. The number of papers published per year was also calculated to get an idea of the field growth.

RESULTS

The bibliographic research performed with the aforementioned criteria identified 173 publications, with 122 papers being considered after duplicate removal. No other systematic reviews or meta-analyses on the topic were identifiable with the chosen key terms during the included years. Nine publications were excluded due to formatting requirements. No complete fulfilment of the other inclusion criteria (focus on NLP in dermatology) was found in 83 papers. A total of 30 papers met all the research criteria and were therefore considered for the present work (Figure 1). More detailed information on included articles is shown in Table 1.

The retrieved results showed great variability either in terms of the topic (e.g. immune-mediated dermatoses, skin cancer and skincare) or of sources (electronic medical records, pathology reports, social media, forums and ChatGPT). The most common use across studies was the application of NLP to electronic health records (EHRs) in order to extract clinical data. There has been a growing interest in NLPs over the years. In particular, an exponential increase in publications focusing

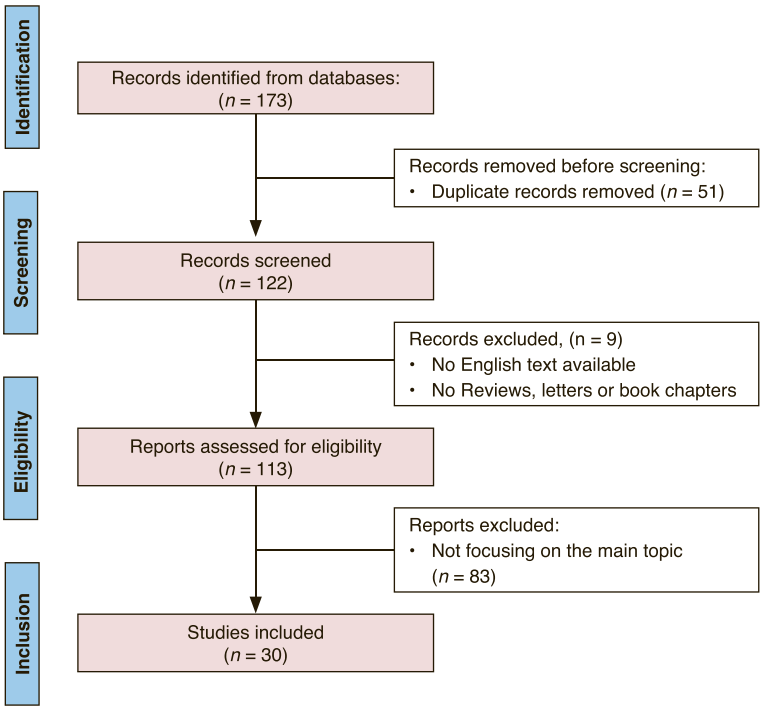

FIGURE 1 Workflow diagram describing the systematic selection of studies for inclusion in the present review (PRISMA flowchart).

on NLP has been observed in the last 5 years (Figure 2). Of note, ChatGPT was used in three papers only, while all the other publications were based on different NLP systems, such as Python 3 or Google BigQuery.[7,8] Several authors also employed other software for further data analysis, including Pharos.[9] Complete list of the software's and/or NLP tools is available in Table S1. Most of the authors developed their own specific NLP algorithms.[10–13] We describe the main findings of our search divided into subject areas.

Atopic dermatitis

In the setting of atopic dermatitis (AD), NLP has been successfully employed for many different purposes, such as diagnosis, severity assessment and identification of unmet needs. Wang et al.[10] published an automated approach for identifying AD patients through NLP of EHRs. NLP led to the identification of otherwise undiagnosed cases of AD, with potential applications for the improvement in the process of patient recruitment for AD studies. In the Spanish study 'DERMACLEAR', NLP of EHRs enabled retrospective collection of data from 54,458 patients with the aim of assessing the incidence of hidradenitis suppurativa (HS), psoriasis, ulcerative colitis (CU) and/or AD.[14] NLP proved to be an effective tool for the detection of HS, psoriasis, CU and/or AD cases as well as for providing new insights on disease burden. A study by Pierce et al.[15] aimed at assessing AD-related symptoms and unmet needs from NLP of EHRs. Healthcare-provider notes mostly focused on symptoms (especially itch) rather than on AD-related impairment in terms of quality of life, therefore highlighting a potential care gap that is worth looking into more deeply. Additionally, other publications focused on patient unmet needs. Falissard et al.[16] performed NLP patient responses to a large-scale international cross-sectional survey. Not surprisingly, the emotional burden reported by interviewed patients was even more important than the whole set of AD-related physical signs and symptoms.

Lakdawala et al.[17] assessed the reliability and preciseness of ChatGPT in offering clinical guidance for AD. The authors found that 78% of ChatGPT responses were correct but inadequate, with more than 50% of the answers not being totally exhaustive; however, no responses were fully unreliable.

Cummins et al.[7] used NLP tools to examine the evolution of patient treatment preferences overtime based on related comments in subreddits communities. Posts regarding eczema/AD and psoriasis were considered. As for AD, positive comments on antibacterial treatments, lifestyle modifications and corticosteroids decreased from 2014 to 2020, while those on phototherapy remained stable; concomitantly a significant increase in dupilumab-related comments was detected. Similarly, an increase in terms of comment volume regarding newer therapeutics for psoriasis (including new biologic drugs) was observed after FDA approval, whereas lower general interest could be appreciated for older drugs, including either traditional systemic immunosuppressants and anti-TNFa (tumour necrosis factor alpha) drugs.

As for patient risk stratification, Kwon et al.[18] recently developed an algorithm integrating text features derived from Simplified Molecular-Input Line-Entry System (SMILES) using NLP with in vitro test outcomes for predicting the risk of skin sensitization. Such an algorithm could be particularly

useful for AD patients, which notably have a higher risk of cutaneous sensitization.

Acne/rosacea

Sachar et al.[19] used NLP to collect patient reviews of over-the-counter (OTC) acne treatments and classify them based on sentiment (positive or negative). This process, also known as 'sentiment analysis' or 'opinion mining', involves the employment of NLP algorithms to ascertain the sentiment conveyed in a text with the primary objective of categorizing it as positive, negative or neutral.[20] The authors found a patient preference for salicylic acid and benzoyl peroxide.[19] Products with active ingredients were frequently reported by the patients to be more effective ($p<0.001$) and to cause potential side effects, including erythema and hypersensitivity. The authors also highlighted that over one-third of top-selling products contained solely 'natural' ingredients, which were reported to improve skin texture and to be easily applied. Other studies focused on the use of chatbots in the setting of acne. A recent study by Lakdawala et al.[17] found ChatGPT to lack accuracy in providing clinical recommendations for acne, especially regarding responses on therapeutic management. Rajalingam et al.[21] aimed at better understanding rosacea patient's unmet needs through NLP of online social media. Four major categories were established to organize and identify the main subjects of discussion: management (50.33%), clinical presentation (24.14%), emotion (21.97%) and information appraisal (3.57%).

Skin and soft tissue infections

An algorithm for NLP-based extraction of data on skin and soft tissue infection (SSTIs) from EHRs was recently developed by Rhoads et al.,[11] enabling subtype classification from clinical notes. Even more surprisingly, the group of Bucher et al.[22] efficiently developed and validated a portable NLP system capable of automated surveillance of EHRs for the detection of surgical site infections. Similarly, Woo and collaborators[12] extracted wound-infection-related information from nursing notes in order to obtain a more detailed overview of patient needs in the community setting. Blumenthal et al.[23] focused on SSTI secondary to allergen immunotherapy (AIT). The researchers performed NLP of large social media databases comparing AIT-related SSTIs to similar procedures specifically using a sterile pharmaceutical (e.g. influenza vaccination). The authors concluded that SSTIs from AIT do not carry a higher risk of SSTI.

Non-melanoma skin cancers (NMSCs)

Eide et al.[24] were the first to use NLP of electronic pathology reports (EPRs) for the detection of NMSCs cases. Such strategy was used to validate a specific search algorithm with the aim of identifying NMSCs from claims data. Ali et al.[25] also performed extensive research on NLP in the setting of basal cell carcinoma (BCC) and employed NLP techniques to extract clinical information from EHRs in this setting. A preliminary study published in 2022 demonstrated efficient clinical data extraction from pathology reports, using NLP rule-based approach for named entity recognition in BCC.[26] The authors validated a pipeline with possible utilization for enhancing cancer registry and research data. Such a NLP algorithm was used by the same group in 2023 to provide an update on the excision margins required to achieve complete BCC clearance.[13] The authors used NLP to extract margin status from EPRs in 34,955 BCCs and found that a clinical peripheral margin of 6 mm attained a histological clearance rate of 95%. Additionally, tumour thickness demonstrated an inverse correlation with deep-margin histological clearance. More recently, the same group also successfully validated the NLP-based web platform to help in evidence-based multidisciplinary-team decisions.[27]

A study published in 2016 by Strekalova et al.[28] focused on recommendations on skin cancer prevention left as a comment on social media. Most of the comments on skin cancer prevention were found to be biased by uncertain language (e.g. 'would', 'confused/confusing' and 'may').

With regards to precancerous lesions such as actinic keratoses (AKs), Lent et al.[29] assessed the accuracy of ChatGPT in providing information on pathogenesis, risk factors, diagnosis and treatment. According to an expert panel judgement, ChatGPT was found to perform better in general patient education than in providing accurate diagnostic and/or therapeutic strategies. ChatGPT was confirmed by an expert panel to be an appealing and effective tool, particularly for its quick information retrieval. However, the chatbot was criticized for being both inaccurate and overly wordy.

Melanoma

Lott et al.[30] aimed at determining the frequencies of histologically confirmed melanocytic lesions through NLP-based analysis of around 80,000 EPRs. The authors found that around 25% of skin biopsies were of melanocytic lesions, highlighting the relevance of a reliable classification system for this category. The authors were also able to subclassify the lesions according to the MPATH-Dx (Melanocytic Pathology Assessment Tool and Hierarchy for Diagnosis) classification system and to define the proportion of invasive melanomas.[31] Yang et al.[32] published a study on the prognostic significance of tumour-infiltrating lymphocytes (TILs) in a large cohort of melanoma patients established through NLP algorithms. Brisk TILs were found to represent an independent prognostic factor, confirming that NLP could represent an efficient tool to provide new insights in the dermato-oncological setting.

Skincare

Okon et al.[8] performed NLP of Reddit comments about dermatology. The authors identified tanning, acne and

TABLE 1 Complete list of included articles and main findings.

References	Disease(s) of interest	Source	Main outcomes	Screened records	First author	Year
10	AD	Medical notes/charts	NLP techniques enable the identification of AD patients, therefore facilitating recruitment for research studies.	*n* = 1926	Wang et al.	2024
7	AD, Pso	Social media	NLP analysis of user comments on Reddit may be a valuable tool for identifying potential adverse drug reactions (ADRs). The number of comments is influenced by the diffusion of the drugs, peaking with the FDA approval of new treatments.	*n* = 123,144 (Pso) *n* = 196,571 (eczema)	Cummins et al.	2023
15	AD	Medical notes/charts	Healthcare providers mainly concentrate on symptoms and treatments, overlooking the impact of AD on patients' QoL.	*n* = 133,025	Pierce et al.	2021
16	AD	Surveys	Information on disease burden may be collected through NLP of patients' surveys. The emotional burden reported by interviewed AD patients was even more important than the whole set of disease-related physical signs and symptoms.	*n* = 639	Falissard et al.	2020
14	HS, Pso, CU and AD	Medical notes/charts	This study tested the precision of NLP system across different tertiary hospitals. It showed high accuracy in identifying patients diagnosed with HS, PsO, CU and/or AD through both external and internal verification.	*n* = 54,458	Ortiz de Frutos et al.	2023
18	Allergic contact dermatitis	Scientific datasets	The integration of NLP of Simplified Molecular-Input Line-Entry System (SMILES) with in vitro test results can enhance the prediction of hazards of skin sensitizers	N/A	Kwon et al.	2024
17	AD and Acne	ChatGPT	ChatGPT accuracy is relatively low for questions on acne treatment compared to AD (omitted information on treatment effectiveness, no guidance on treatment expectations, missing age-appropriate and patient-specific recommendations)	N/A	Lakdawala et al.	2023
19	Acne	Online retailers	OTC products for acne treatment with active ingredients (salicylic acid and benzoyl peroxide) are more effective but lead to hypersensitivity reactions and redness. Products with 'natural' ingredients are easier to apply and enhance the skin's appearance.	*n* = 12,441 (OTC acne products)	Sachar et al.	2021
21	Rosacea	Social media	NLP of posts on social media platforms can help identify the main worries of patients affected by rosacea; with therapeutic management being the most frequently discussed topic and emotional burden being identified as one of the main unmet needs.	*n* = 207,038	Rajalingam et al.	2023
11	SSTIs	Medical notes/charts	NLP was effectively used for automatically extracting data on SSTIs and further classifying them as purulent and non-purulent; data on incision and drainage procedures could also be collected.	*n* = 6576	Rhoads et al.	2024
12	SSTIs	Medical notes/charts	NLP is effectively applied to obtain prevalence data of wound infection among patients admitted to ED or hospitalized during a homecare episode; SSTIs are more common in case of diabetes, PVD and skin ulcers.	*n* = 112,789 (care episodes)	Woo et al.	2021
23	SSTIs	Medical notes/charts	SSTIs from AIT (allergen immunotherapy) and influenza vaccination are equally uncommon occurrences.	*n* = 25,126	Blumenthal et al.	2019
22	SSTIs	Medical notes/charts	NLP of clinical notes from the operative events can help in surveillance of SSIs. This study validated a rules-based NLP system with high sensitivity and specificity	*n* = 21,784	Bucher et al.	2020

(Continues)

TABLE 1 (Continued)

References	Disease(s) of interest	Source	Main outcomes	Screened records	First author	Year
24	NMSC	Medical notes/charts	Administrative data collection was compared with NLP analysis of EPR to identify NMSCs. The results showed that while administrative data had high sensitivity and specificity rates, NLP demonstrated higher accuracy in identifying NMSC.	$n = 4883$	Eide et al.	2012
28	Skin cancer prevention	Social media surveys	NLP may be used in medical surveys to help monitor patient perceptions, focusing on cancer prevention and communication campaigns. Several indicators of uncertainty were identified by NLP as potentially determining decisional conflict related to skin cancer prevention recommendations.	$n = 762$	Strekalova et al.	2016
26	BCC	Medical notes/charts	An NLP rule-based approach was used to develop and validate a pipeline. The goal of this study was to enhance the quality of cancer registry data, aiding in service planning and improving the data collection in the research setting.	$n = 200$	Ali et al.	2022
13	BCC	Medical notes/charts	NLP can be efficiently used to extract data on surgical margin status from EPR.	$n = 34,955$	Ali et al.	2023
27	BCC	Medical notes/charts	The study validates an NLP-based platform for automating evidence-based decisions in a multiclass clinical decision support system.	$n = 1045$	Ali et al.	2024
25	BCC	Medical notes/charts	Updated RCPath reporting standards minimally impact the peripheral margin histological clearance in BCC.	$n = 34,955$	Ali et al.	2024
29	AK	ChatGPT	Chatbot performed best in patient education but struggled with diagnosis and treatment. Deficits were seen in the AK classification and provided up-to-date information on treatment.	N/A	Lent et al.	2024
30	Melanocytic lesions	Medical notes/charts	The application of NLP techniques to EPR from skin biopsies demonstrated that around one-quarter of all skin biopsies led to diagnoses of melanocytic proliferations.	$n = 80,368$	Lott et al.	2018
32	Melanoma	Medical notes/charts	Brisk TILs are an independent prognostic factor for overall survival in patients with primary cutaneous melanoma.	$n = 2624$	Yang et al.	2021
8	Skincare, Acne and Pso	Social media	NLP applied to user comments in dermatology forums helps to assess patient engagement, especially for common dermatology topics such as tanning, acne and psoriasis; most comments focused on homoeopathic treatments and retinoids.	$n = 176,000$	Okon et al.	2019
34	Skincare	Social media	NLP techniques can be applied to dermatology forums discussing 'tanning and essential oils', effectively identifying examples of misinformation.	N/A	Sager et al.	2021
33	Skincare	Social media	NLP can be helpful in analysing comments from dermatology discussion forums to track how often certain words are used and to show how language preferences change over time. Sentiments about various topics can also be gathered.	$n > 3,000,000$	Cummins et al.	2022
35	General dermatology	ChatGPT	This study determines if ChatGPT for dermatology conditions is a reliable tool. The physician-generated response was preferred over the ChatGPT response on 31 different answers to medical questions, by both physician and non-physician reviewers with higher ratings for 'readability' and 'level of empathy'.	N/A	Reynolds et al.	2024
37	EB	Medical notes/charts	Categorizing pubertal status in patients with EB using NLP software reveals that delayed puberty is a commonly under-recognized comorbidity.	$n = 186$	Wasserman et al.	2023

TABLE 1 (Continued)

References	Disease(s) of interest	Source	Main outcomes	Screened records	First author	Year
9	SLE	Social media	NLP of posts on social media platforms enabled the identification of pain and fatigue as the most reported symptoms influencing QoL in patients with LES. While pharmacological treatments poorly manage these symptoms, nonpharmacological interventions are associated with QoL improvement.	n = 45,554	Spies et al.	2024
36	Calciphylaxis	Social media	Applying the NLP algorithm to clinical data underscores a significant increase in the incidence of otherwise underreported diseases such as calciphylaxis.	n = 11,451	Nigwekar et al.	2014
38	Vesicular-pustular rash	Medical notes/charts	Afebrile neonates presenting with pustules alone do not need hospitalization and/or full bacterial infection examination.	n = 64	Manice et al.	2018

Abbreviations: AD, atopic dermatitis; AK, actinic keratosis; BCC, basal cell carcinoma; CU, chronic urticaria; EB, epidermolysis bullosa; ED, emergency department; EPR, electronic pathology reports; N/A, not applicable or available; NMSC, non-melanoma skin cancer; OTC, over-the-counter; Pso, psoriasis, HS, hidradenitis suppurativa; PVD, peripheral vascular disease; QoL, quality of life; SLE, systemic lupus erythematosus; SSIs, surgical site infections; SSTIs, skin and soft tissue infections; TIL, tumour-infiltrating lymphocytes.

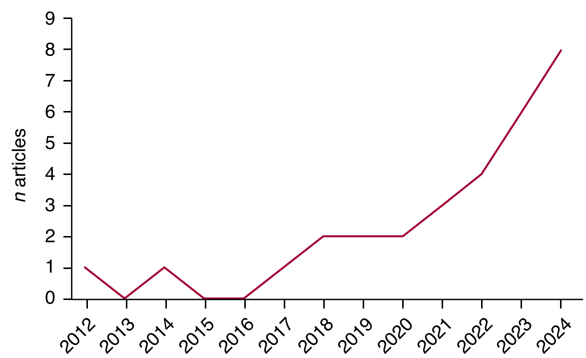

FIGURE 2 Graphical representation of the published papers on the use of NLP in the dermatological setting overtime according to our study inclusion criteria.

psoriasis as topics of interest, with potential implications for patient engagement in the dermatological research setting. Cummins et al.[33] efficiently applied NLP-based sentiment analysis and word search techniques to Reddit comments in order to gain new insights into patient perceptions in the setting of skincare. Reddit was also used as a source data by Sager et al.,[34] who performed a study with the aim of evaluating the presence of health misinformation within dermatology forums with regards to tanning and essential oils. The authors found NLP to potentially allow interception of misinformation. Recently, Reynolds et al.[35] assessed the ability of ChatGPT to generate high-quality responses to questions from patients. Not surprisingly, answers provided by physicians were largely preferred to those from ChatGPT according to both expert and non-expert reviewers.

Others

Other publications focus on potential applications of NLP in the setting of rarer dermatological conditions. Nigwekar et al.[36] developed a NLP-based method for the automatic detection of cases of calciphylaxis from EHRs, in order to assess its incidence and mortality. Interestingly, calciphylaxis incidence was observed to be higher than expected, with mortality rates 2.5–3 times higher than for other chronic haemodialysis patients.

Spies et al.[9] recently performed NLP on social media platforms to better identify unmet needs of patients living with lupus erythematosus, either systemic (SLE) or cutaneous (CLE). In total, 1925 and 106 patients were identified as having SLE and CLE, respectively. The most frequently reported symptoms included pain, fatigue and rashes, with a great impact on patient quality of life (QOL). While pharmacological interventions were found to poorly manage the most burdensome symptoms, nonpharmacological treatments, such as exercise and meditation, were associated with QOL improvement.

Wassermann et al.[37] used NLP of EHRs to assess the impact of delayed puberty and low bone mineral density (BMD) in patients affected by epidermolysis bullosa (EB). Interestingly, EB patients were found to have low BMD even

prior to the age of 10. Of those reaching adolescence, delayed puberty and further decline in BMD were not uncommon findings. Through NLP of EHRs, Manice et al.[38] assessed the management of paediatric patients presenting to the emergency department, focusing on afebrile neonates with cutaneous vesicles and pustules. Despite around one-third of such neonates receiving parenteral therapy, no cerebrospinal fluid or blood infections were found, as well as any evidence of herpes simplex virus infection. The results of the study suggested that a great proportion of afebrile neonates presenting with vesicles and/or pustules may not require hospitalization in the absence of additional risk factors.

DISCUSSION

In this systematic review, including 30 articles on the use of NLPs in dermatology, we found different categories of studies, classifiable either based on study design or main subject. Major topics of interest included: atopic dermatitis, acne/rosacea, skin infections, NMSCs, melanoma, skincare and drug reactions.

Studies have demonstrated NLP effectiveness in identifying undiagnosed cases, assessing disease severity and uncovering patient unmet needs. The ability of NLP to extract valuable insights from diverse sources such as EHRs, social media platforms and online forums underscores its versatility in gathering real-world data for research and clinical purposes. However, the review also identifies limitations in certain applications, such as the reliability of NLP chatbots in providing exhaustive clinical guidance for specific dermatological conditions, such as AD or acne, indicating areas for further improvement and refinement of NLP tools. Despite ChatGPT is probably the most widely used chatbot worldwide, only little evidence is available for its use in the dermatological setting, with current literature suggesting relatively low accuracy and reliability for providing clinical guidance to patients affected by skin conditions.[17,35] However, we acknowledge that this chatbot only became available at the end of 2022: this short time horizon gives reason of the little evidence supporting its use in the dermatological setting. More studies on dermatology are needed to evaluate its true role. Our results highlight the versatility and potential of NLP in improving various aspects of dermatological care, including diagnosis, treatment optimization, patient education and understanding patient needs.

While NLP offers numerous opportunities for advancing dermatological research, there are several pitfalls and challenges that researchers may encounter in its application. Firstly, the need for high-quality input data: if the training data predominantly consists of certain demographic groups, it may not generalize well to diverse populations, leading to biased outcomes. Secondly, a need for standardized terminology: terminological differences between medical and patient vocabulary can be particularly challenging and do not allow precise matching of clinical notes and patient-reported data.[39–41] Moreover, the variability in terms of terminology and coding practices across different healthcare systems, EHRs and clinical notes may affect data extraction and analysis.[42] Another tricky point for NLP algorithms is represented by the adherence to stringent data privacy to ensure patient confidentiality and compliance with regulatory requirements, with mishandling of sensitive information potentially leading to legal and ethical ramifications.[43] Lastly, NLP models trained on specific datasets may suffer from overfitting, where they memorize patterns from the training data but fail to generalize well to unseen data.[44]

The advent of NLP within the realm of AI is revolutionizing medical practice, particularly in conjunction with image-analysis applications.[45] With regards to this, existing research in image-based dermatology underscores the enhanced diagnostic precision facilitated by AI tools for skin cancer diagnosis and beyond.[46–48] While concerns persist regarding the potential displacement of human expertise by AI, we foresee a synergistic partnership between clinicians and AI, with NLP serving as a fundamental component of this collaboration.

The main limitation of our study resides in the heterogeneity of the included publications, which did not allow us to perform further meta-analysis. Moreover, we did not perform bias assessment, with potential disparities in terms of ethnicity, sex and age of the different study populations possibly affecting the interpretation of the results.

CONCLUSION

The findings of this systematic review demonstrate the wide-ranging applications and potential of NLP in dermatology. While significant progress has been made in leveraging NLP for various purposes, further research is warranted to address existing limitations and explore novel avenues for enhancing its utility in clinical practice. Collaboration between clinicians, data scientists and technologists will be crucial in harnessing the full capabilities of NLP to improve patient care and advance dermatological research.

AUTHOR CONTRIBUTIONS

Conceptualization AP and CL; data curation AP and MS; formal analysis SG, CN and AP; supervision CL and GP; investigation AP and MS; writing—original draft AP, MS and CN; writing—review and editing CL, SG and GP.

ACKNOWLEDGEMENTS

The authors thank Dr. Marica Mirra and Margherita Raucci for their technical assistance. Open access publishing facilitated by Universita degli Studi di Modena e Reggio Emilia, as part of the Wiley - CRUI-CARE agreement.

FUNDING INFORMATION

None.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY STATEMENT

Data sharing not applicable - no new data generated.

ETHICAL APPROVAL

EC approval was not required since human subjects were not directly involved in the present study.

STATEMENTS

The present systematic review was submitted for registration on PROSPERO (ID 526135). A written protocol study was not prepared prior to systematic literature revision.

ORCID

Alessia Paganelli https://orcid.org/0000-0002-0916-7769
Marco Spadafora https://orcid.org/0000-0003-3509-4222
Cristian Navarrete-Dechent https://orcid.org/0000-0003-4040-3640
Stefania Guida https://orcid.org/0000-0002-8221-6694
Giovanni Pellacani https://orcid.org/0000-0002-7222-2951
Caterina Longo https://orcid.org/0000-0002-8218-3896

REFERENCES

1. Guida G, Mauri G. Evaluation of natural language processing systems: issues and approaches. Proc IEEE. 1986;74:1026–35.
2. Goldberg Y. A primer on neural network models for natural language processing. Jair. 2016;57:345–420.
3. Masoumi S, Amirkhani H, Sadeghian N, Shahraz S. Natural language processing (NLP) to facilitate abstract review in medical research: the application of BioBERT to exploring the 20-year use of NLP in medical research. Syst Rev. 2024;13:107.
4. Bhatia A, Titus R, Porto JG, Katz J, Lopategui DM, Marcovich R, et al. Application of natural language processing in electronic health record data extraction for navigating prostate cancer care: a narrative review. J Endourol. 2024; 38:852–64.
5. Kariampuzha WZ, Alyea G, Qu S, Sanjak J, Mathé E, Sid E, et al. Precision information extraction for rare disease epidemiology at scale. J Transl Med. 2023;21:157.
6. AlShehri Y, Sidhu A, Lakshmanan LVS, Lefaivre KA. Applications of natural language processing for automated clinical data analysis in orthopaedics. J Am Acad Orthop Surg. 2024;32:439–46.
7. Cummins JA, Zhou G, Nambudiri VE. Natural language processing for large-scale analysis of eczema and psoriasis social media comments. JID Innov. 2023;3:100210.
8. Okon E, Rachakonda V, Hong HJ, Callison-Burch C, Lipoff JB. Natural language processing of reddit data to evaluate dermatology patient experiences and therapeutics. J Am Acad Dermatol. 2020;83:803–8.
9. Spies E, Andreu T, Hartung M, Park J, Kamudoni P. Exploring the perspectives of patients living with lupus: retrospective social listening study. JMIR Form Res. 2024;8:e52768.
10. Wang A, Fulton R, Hwang S, Margolis DJ, Mowery D. Patient phenotyping for atopic dermatitis with transformers and machine learning: algorithm development and validation study. JMIR Form Res. 2024;8:e52200.
11. Rhoads JLW, Christensen L, Westerdahl S, Stevens V, Chapman WW, Conway M. Automatic extraction of skin and soft tissue infection status from clinical notes. Stud Health Technol Inform. 2024;310:579–83.
12. Woo K, Song J, Adams V, Block LJ, Currie ML, Shang J, et al. Exploring prevalence of wound infections and related patient characteristics in homecare using natural language processing. Int Wound J. 2022;19:211–21.
13. Ali SR, Dobbs TD, Jovic M, Strafford H, Fonferko-Shadrach B, Lacey AS, et al. Validating a novel natural language processing pathway for automated quality assurance in surgical oncology: incomplete excision rates of 34 955 basal cell carcinomas. Br J Surg. 2023;110:1072–5.
14. Ortiz De Frutos FJ, Giménez-Arnau AM, Puig L, Silvestre JF, Serra E, Salgado-Boquete L, et al. The DERMACLEAR study: verification results of a natural language processing system in dermatology. JEADV Clin Pract. 2023;2:775–85.
15. Pierce EJ, Boytsov NN, Vasey JJ, Sudaria TC, Liu X, Lavelle KW, et al. A qualitative analysis of provider notes of atopic dermatitis-related visits using natural language processing methods. Dermatol Ther (Heidelb). 2021;11:1305–18.
16. Falissard B, Simpson EL, Guttman-Yassky E, Papp KA, Barbarot S, Gadkari A, et al. Qualitative assessment of adult patients' perception of atopic dermatitis using natural language processing analysis in a cross-sectional study. Dermatol Ther (Heidelb). 2020;10:297–305.
17. Lakdawala N, Channa L, Gronbeck C, Lakdawala N, Weston G, Sloan B, et al. Assessing the accuracy and comprehensiveness of ChatGPT in offering clinical guidance for atopic dermatitis and acne vulgaris. JMIR Dermatol. 2023;6:e50409.
18. Kwon J-H, Kim J, Lim K-M, Kim MG. Integration of the natural language processing of structural information simplified molecular-input line-entry system can improve the in vitro prediction of human skin sensitizers. Toxics. 2024;12:153.
19. Sachar M, Xiong M, Lee KC. Consumer preferences of top-rated over-the-counter acne treatment products: a cohort study. Arch Dermatol Res. 2022;314:815–21.
20. Mokhosi R, Shivachi C-S, Sethobane M. A Sesotho news headlines dataset for sentiment analysis. Data Brief. 2024;54:110371.
21. Rajalingam K, Levin N, Marques O, Grichnik J, Lin A, Chen W-S. Treatment options and emotional well-being in patients with rosacea: an unsupervised machine learning analysis of over 200,000 posts. JAAD Int. 2023;13:172–8.
22. Bucher BT, Shi J, Ferraro JP, Skarda DE, Samore MH, Hurdle JF, et al. Portable automated surveillance of surgical site infections using natural language processing: development and validation. Ann Surg. 2020;272:629–36.
23. Blumenthal KG, Topaz M, Zhou L, Harkness T, Sa'adon R, Bar-Bachar O, et al. Mining social media data to assess the risk of skin and soft tissue infections from allergen immunotherapy. J Allergy Clin Immunol. 2019;144:129–34.
24. Eide MJ, Tuthill JM, Krajenta RJ, Jacobsen GR, Levine M, Johnson CC. Validation of claims data algorithms to identify nonmelanoma skin cancer. J Invest Dermatol. 2012;132:2005–9.
25. Ali SR, Dobbs TD, Jovic M, Strafford H, Lacey AS, Williams N, et al. Revisiting basal cell carcinoma clinical margins: leveraging natural language processing and multivariate analysis with updated Royal College of pathologists histological reporting standards. J Plast Reconstr Aesthet Surg. 2024;88:443–51.
26. Ali SR, Strafford H, Dobbs TD, Fonferko-Shadrach B, Lacey AS, Pickrell WO, et al. Development and validation of an automated basal cell carcinoma histopathology information extraction system using natural language processing. Front Surg. 2022;9:870494.
27. Ali SR, Dobbs TD, Tarafdar A, Strafford H, Fonferko-Shadrach B, Lacey AS, et al. Natural language processing to automate a web-based model of care and modernize skin cancer multidisciplinary team meetings. Br J Surg. 2024;111:znad347.
28. Strekalova YA, James VS. Language of uncertainty: the expression of decisional conflict related to skin cancer prevention recommendations. J Cancer Educ. 2017;32:532–6.
29. Lent HC, Ortner VK, Karmisholt KE, Wiegell SR, Nissen CV, Omland SH, et al. A chat about actinic keratosis: examining capabilities and user experience of ChatGPT as a digital health technology in dermato-oncology. JEADV Clin Pract. 2024;3:258–65.
30. Lott JP, Boudreau DM, Barnhill RL, Weinstock MA, Knopp E, Piepkorn MW, et al. Population-based analysis of histologically confirmed melanocytic proliferations using natural language processing. JAMA Dermatol. 2018;154:24–9.
31. Barnhill RL, Elder DE, Piepkorn MW, Knezevich SR, Reisch LM, Eguchi MM, et al. Revision of the melanocytic pathology

assessment tool and hierarchy for diagnosis classification schema for melanocytic lesions: a consensus Statement. JAMA Netw Open. 2023;6:e2250613.
32. Yang J, Lian JW, Chin Y-P (Harvey), Wang L, Lian A, Murphy GF *et al*. Assessing the prognostic significance of tumor-infiltrating lymphocytes in patients with melanoma using pathologic features identified by natural language processing. JAMA Netw Open 2021;4:e2126337.
33. Cummins JA, Nambudiri VE. Natural language processing: a window to understanding skincare trends. Int J Med Inform. 2022;160: 104705.
34. Sager MA, Kashyap AM, Tamminga M, Ravoori S, Callison-Burch C, Lipoff JB. Identifying and responding to health misinformation on reddit dermatology forums with artificially intelligent bots using natural language processing: design and evaluation study. JMIR Dermatol. 2021;4:e20975.
35. Reynolds K, Nadelman D, Durgin J, Ansah-Addo S, Cole D, Fayne R, et al. Comparing the quality of ChatGPT- and physician-generated responses to Patients' dermatologic questions in the electronic medical record. Clin Exp Dermatol. 2024;llad456 49:715–18.
36. Nigwekar SU, Solid CA, Ankers E, Malhotra R, Eggert W, Turchin A, et al. Quantifying a rare disease in administrative data: the example of calciphylaxis. J Gen Intern Med. 2014;29(Suppl 3):S724–S731.
37. Wasserman H, Dumenigo A, Hornung L, Augsburger B, Marathe K, Lucky AW. Prevalence of delayed puberty and low bone density in patients with epidermolysis bullosa: insight from a large single center's experience. Pediatr Dermatol. 2023;40:100–6.
38. Manice CS, Planet PJ, Chase HS, Lauren CT. Management of afebrile neonates with pustules and vesicles in a pediatric emergency department. Pediatr Dermatol. 2018;35:660–5.
39. Donoso F, Hidalgo L, Cowen EA, Villagran S, Villablanca P, Del Puerto C, et al. Accuracy in anatomical location on dermatological surgery: a multi-centre retrospective study. J Eur Acad Dermatol Venereol. 2023;37:e93–e95.
40. Zeng Q, Kogan S, Ash N, Greenes RA. Patient and clinician vocabulary: how different are they? Stud Health Technol Inform. 2001;84(Pt 1):399–403.
41. Navarrete-Dechent C, Liopyris K, Molenda MA, Braun R, Curiel-Lewandroski C, Dusza SW, et al. Human surface anatomy terminology for dermatology: a Delphi consensus from the international skin imaging collaboration. J Eur Acad Dermatol Venereol. 2020;34:2659–63.
42. Wang X, Luo J, Wang J, Yin Z, Cui S, Zhong Y, et al. Hierarchical pretraining on multimodal electronic health records. Proc Conf Empir Methods Nat Lang Process. 2023;2023:2839–52.
43. Petit-Jean T, Gérardin C, Berthelot E, Chatellier G, Frank M, Tannier X, et al. Collaborative and privacy-enhancing workflows on a clinical data warehouse: an example developing natural language processing pipelines to detect medical conditions. J Am Med Inform Assoc. 2024;31:1280–90.
44. Hua H, Li X, Dou D, Xu C-Z, Luo J. Improving pretrained language model fine-tuning with noise stability regularization. IEEE Trans Neural Netw Learn Syst. 2023; Epub ahead of print.
45. Jena B, Saxena S, Nayak GK, Saba L, Sharma N, Suri JS. Artificial intelligence-based hybrid deep learning models for image classification: the first narrative review. Comput Biol Med. 2021;137:104803.
46. Han SS, Kim YJ, Moon IJ, Jung JM, Lee MY, Lee WJ, et al. Evaluation of artificial intelligence-assisted diagnosis of skin neoplasms: a single-center, paralleled, unmasked, randomized controlled trial. J Invest Dermatol. 2022;142:2353–62.e2.
47. Krakowski I, Kim J, Cai ZR, Daneshou R, Lapins J, Eriksson H, et al. Human-AI interaction in skin cancer diagnosis: a systematic review and meta-analysis. NPJ Digit Med. 2024;7:78.
48. Muñoz-López C, Ramírez-Cornejo C, Marchetti MA, Han SS, Del Barrio-Dìaz P, Jaque A, et al. Performance of a deep neural network in teledermatology: a single-centre prospective diagnostic study. J Eur Acad Dermatol Venereol. 2021;35:546–53.

SUPPORTING INFORMATION

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

How to cite this article: Paganelli A, Spadafora M, Navarrete-Dechent C, Guida S, Pellacani G, Longo C. Natural language processing in dermatology: A systematic literature review and state of the art. J Eur Acad Dermatol Venereol. 2024;00:1–10. https://doi.org/10.1111/jdv.20286