Scilight

Dermatology and Skin Science

https://www.sciltp.com/journals/dss

Letter

Patient-Reported Adverse Reactions and Early Detection of Unresolved Interactions in a Canadian Dermatology Chatbot

Sepideh Emam [1,*] and Robert Gniadecki [2]

[1] Institutional Data and Analytics, University of Alberta, Edmonton, AB T5J 0B2, Canada
[2] Division of Dermatology, Faculty of Medicine, University of Alberta, Edmonton, AB T6G 2R3, Canada
* Correspondence: emam@ualberta.ca

How To Cite: Emam, S.; Gniadecki, R. Patient-Reported Adverse Reactions and Early Detection of Unresolved Interactions in a Canadian Dermatology Chatbot. *Dermatology and Skin Science* **2026**, *1*(1), 4.

Received: 8 May 2026
Revised: 30 May 2026
Accepted: 2 June 2026
Published: 24 June 2026

Abstract: Background: Dermatology chatbots are increasingly used for patient engagement and triage, generating large volumes of patient-reported data. **Objective:** To examine the nature of patient-reported adverse reactions and evaluate machine learning approaches for early detection of unresolved chatbot interactions. **Methods:** A dataset of 103 anonymized chatbot conversations was analyzed. Latent intent topics were identified using non-negative matrix factorization. Interactions were labeled as resolved or unresolved using rule-based heuristics. Logistic regression models with TF-IDF features and conversation-level metadata were evaluated using 5-fold cross-validation. Early-warning models were restricted to the first one or two user messages. Kaplan–Meier analysis was used to assess time-to-unresolved outcomes. **Results:** Reported side effects were predominantly dermatologic and generally moderate in severity but lacked specificity regarding treatment, dosage, and duration. Machine learning models demonstrated moderate predictive performance (ROC-AUC ≈ 0.72), improving to ≈0.78 using the first two user messages. Unresolved interactions were most likely to occur within the first 3–5 conversational turns. **Conclusion:** Dermatology chatbots capture valuable patient-reported signals but lack structured data for clinical interpretation. Early detection of unresolved interactions using machine learning is feasible and may support real-time intervention and improved patient safety.

Keywords: dermatology chatbot; adverse reactions; machine learning; natural language processing; digital health; patient safety

1. Introduction

Artificial intelligence–driven chatbots are increasingly integrated into dermatology practice to support patient engagement, triage, and education [1–3]. These systems provide a low-barrier interface for patients to report symptoms and concerns outside traditional clinical encounters. As a result, they generate real-world patient-reported data that may offer insight into treatment experiences and system usability.

However, the clinical utility of such data depends on its specificity, structure, and interpretability. In parallel, unresolved or failed chatbot interactions represent a potential risk to patient experience and safety. Advances in natural language processing and machine learning provide opportunities to detect such failures early and enable timely intervention [2,4].

This study examines two key aspects of dermatology chatbot use: (1) the characteristics and limitations of patient-reported adverse reactions and (2) the application of machine learning models to identify unresolved interactions.

2. Methods

2.1. Sample Collection

The dataset consisted of 103 anonymized chatbot conversations collected from a Canadian dermatology clinic website. Conversations were included if they contained at least one patient-generated query and sufficient textual content for analysis. Duplicate records, test conversations, administrative messages, and conversations lacking meaningful textual content were excluded. Consecutive eligible conversations during the study period were included to minimize selection bias.

2.2. Topic Modeling and Labeling

Non-negative matrix factorization was used to identify latent intent topics within conversations. Latent intent topics represent underlying user goals or themes inferred from recurring language patterns rather than explicitly assigned labels. Examples identified in the dataset included appointment access, treatment-related questions, symptom inquiries, and medication concerns. Interactions were classified as resolved or unresolved/redirected using rule-based heuristics.

2.3. Predictive Modeling

Logistic regression models were developed using TF-IDF text features and conversation-level metadata. This approach represents a commonly used baseline method in clinical natural language processing tasks [4]. Model performance was evaluated using 5-fold cross-validation.

2.4. Early-Warning Models

To assess early detection capability, models were trained using only the first one or two user messages.

2.5. Statistical Analysis

Kaplan–Meier survival analysis was used to model time-to-unresolved outcomes as a function of conversation length. Association between intent and outcome was assessed using Cramér's V.

3. Results

3.1. Patient-Reported Adverse Reactions

Reported side effects were predominantly dermatologic and consistent with expected reactions to topical or systemic treatments. These included irritation (burning, stinging, redness), worsening symptoms (flare-ups), discomfort or pain (sensitivity), and swelling or inflammation.

Patients rarely identified specific medications, instead referring to treatments using general terms such as "the cream" or "the medication". Occasional references to duration were noted; however, reports lacked sufficient detail to link adverse effects to specific therapies, compare treatments, or assess dosage and duration.

When severity was reported, reactions were generally moderate, with no high-severity or emergent events observed. Interactions were typically brief and reassurance-seeking.

3.2. Predictors of Unresolved Interactions

Conversation-level features, particularly number of user turns and message length, were the strongest predictors of unresolved outcomes (Figure 1).

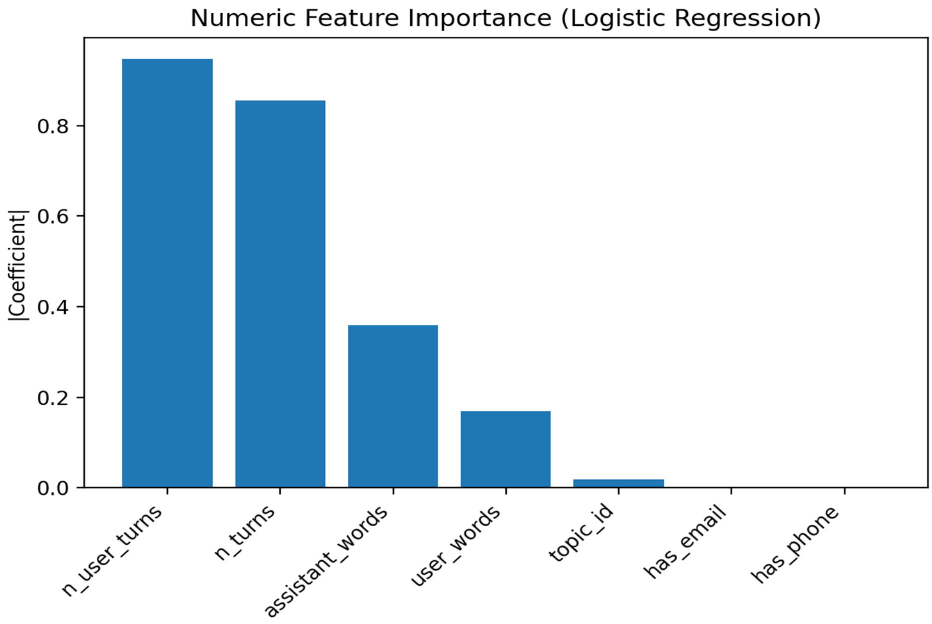

Figure 1. Feature importance derived from logistic regression, highlighting the dominant role of conversation-level variables such as number of user turns and message length.

3.3. Intent and Outcome Association

Access-related intents dominated the dataset and exhibited higher unresolved rates. A moderate association between intent category and outcome was observed (Cramér's V ≈ 0.45) (Figure 2).

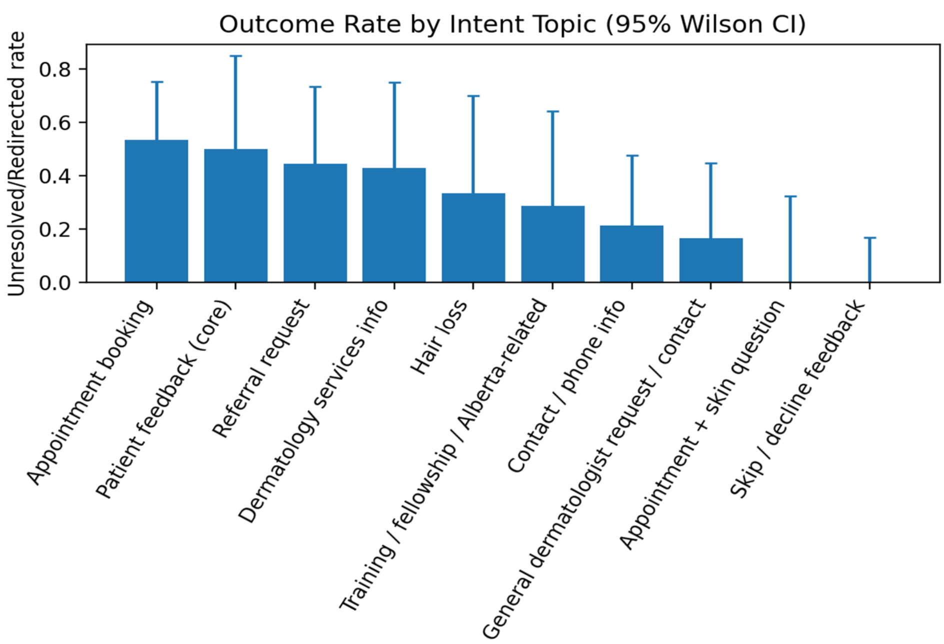

Figure 2. Unresolved outcome rates by intent category with confidence intervals.

3.4. Subgroup Analysis by Intent Category

Intent-specific analysis revealed notable differences in unresolved interaction rates. Access-related and appointment-navigation intents demonstrated the highest proportion of unresolved outcomes, whereas symptom-related and treatment-related conversations were more likely to reach successful resolution. Side-effects–related interactions were generally shorter and reassurance-seeking in nature, with lower rates of escalation. These findings suggest that operational and access-related requests may require additional chatbot workflow refinement compared with clinically focused inquiries.

3.5. Time-to-Unresolved Outcomes

Survival analysis demonstrated that the probability of unresolved interactions increased rapidly within the first 3–5 conversational turns before stabilizing (Figure 3).

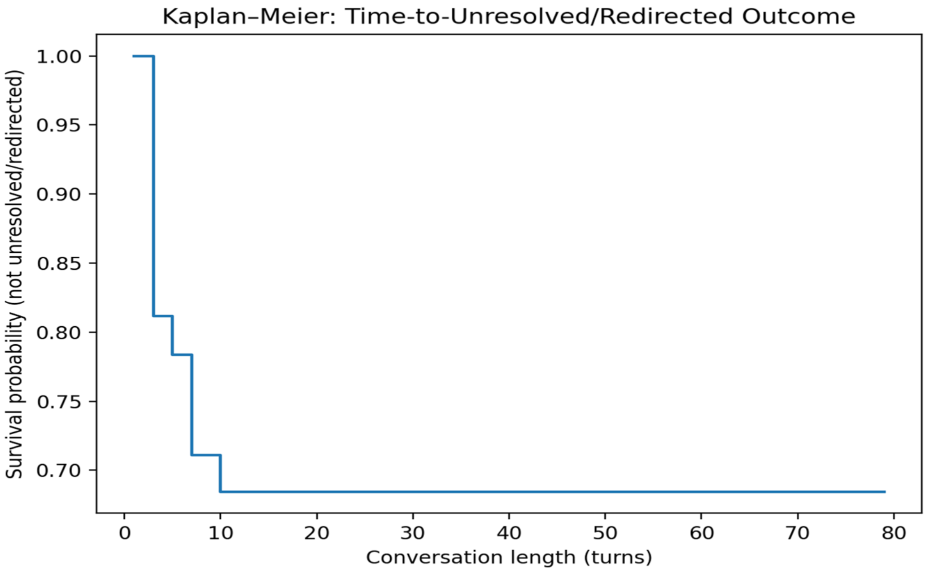

Figure 3. Kaplan–Meier curve illustrating time to unresolved outcome as a function of conversation length.

3.6. Model Performance

The baseline predictive model achieved a ROC-AUC of approximately 0.72. Early-warning models improved performance to approximately 0.74 using the first user message and 0.78 using the first two user messages (Figure 4).

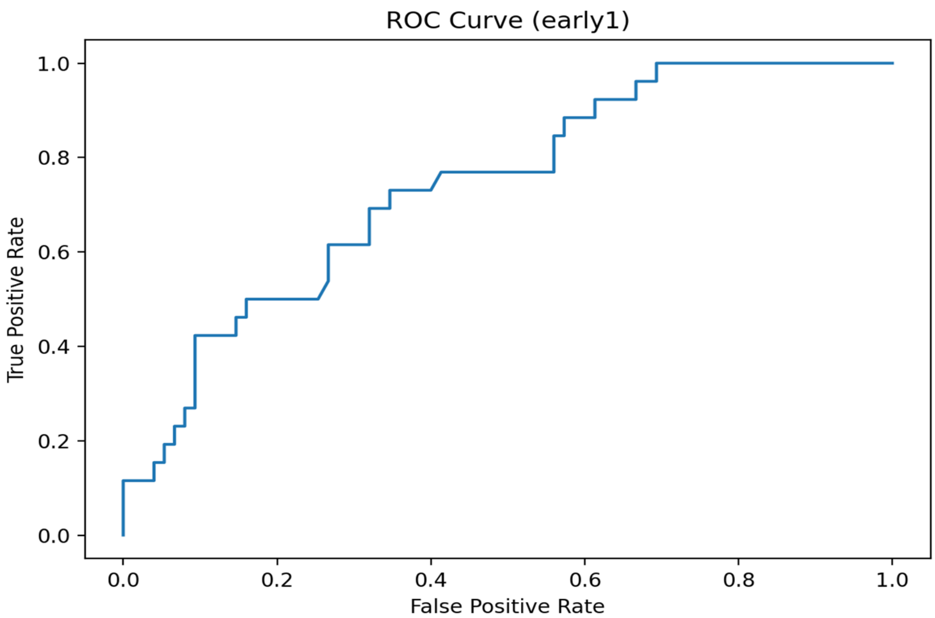

Figure 4. Receiver operating characteristic (ROC) curves comparing baseline and early-warning predictive models.

4. Discussion

This study highlights key limitations and opportunities in dermatology chatbot systems. While patient-reported adverse reactions are frequently captured, the lack of structured and specific information limits their clinical usefulness. Similar challenges in unstructured patient-reported data have been identified in broader digital health and AI applications [4,5].

Machine learning models demonstrated strong potential for early detection of unresolved interactions. The ability to predict outcomes using only the first user messages aligns with prior findings on the role of AI in improving healthcare delivery and decision support [5,6].

Additionally, side-effect–related interactions exhibited a distinct behavioral pattern—brief and reassurance-focused—suggesting that they may require specialized handling strategies compared with operational or access-related queries.

Moreover, it is worth noting that our findings are consistent with previous studies demonstrating that patient-generated chatbot data frequently lack the structured clinical details required for robust pharmacovigilance and adverse-event assessment. Laranjo et al. reported similar limitations in healthcare conversational agents, noting challenges in extracting clinically actionable information from free-text interactions. Likewise, recent digital health studies have shown that symptom and treatment descriptions provided through conversational interfaces are often incomplete with respect to medication identity, dosage, and treatment duration.

The predictive performance observed in our study (ROC-AUC approximately 0.72–0.78) is comparable to results reported in healthcare natural language processing applications where early conversational signals can be used to identify high-risk or unsuccessful interactions. Our observation that unresolved outcomes emerge within the first few conversational turns also aligns with prior work suggesting that early dialogue characteristics strongly influence user satisfaction and interaction success.

The subgroup analysis further extends previous findings by demonstrating that unresolved interactions are concentrated among access-related and operational queries rather than symptom-focused discussions. This suggests that future chatbot improvements may benefit from targeted optimization of administrative workflows in addition to clinical content delivery.

5. Limitations

This study analyzed conversations from a single Canadian dermatology clinic chatbot. Data from other chatbot platforms or healthcare organizations were not available for analysis. As chatbot design, patient populations, and operational workflows may vary across settings, the findings should be interpreted cautiously and may not be directly generalizable to other healthcare chatbots. Future multi-center studies involving multiple chatbot platforms would help validate and extend these findings.

6. Conclusions

Dermatology chatbots provide a valuable channel for capturing patient concerns but currently lack the structure needed for clinical interpretation of adverse reactions. Machine learning approaches enable early identification of unresolved interactions and offer a pathway to improve patient safety and system performance. Future work should focus on structured data capture and integration of clinical context.

Author Contributions

S.E.: conceptualization, methodology, data curation, formal analysis, investigation, visualization, software, writing—original draft preparation, and writing—review and editing. R.G.: supervision, conceptual guidance, project administration, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The research data is available from the corresponding author upon reasonable request.

Conflicts of Interest

Given the role as Editor-in-Chief, R.G. had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

Use of AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used Gemini Pro V3.1 to improve grammar, spelling, and overall language accuracy. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Esteva, A.; Kuprel, B.; Novoa, R.A.; et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118.
2. Laranjo, L.; Dunn, A.G.; Tong, H.L.; et al. Conversational agents in healthcare: a systematic review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258.
3. Miner, A.S.; Laranjo, L.; Kocaballi, A.B. Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit. Med.* **2020**, *3*, 65.
4. Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347–1358.
5. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56.
6. Liu, X.; Faes, L.; Kale, A.U.; et al. A comparison of deep learning performance against healthcare professionals. *Lancet Digit. Health* **2019**, *1*, e271–e297.