ITU-T Focus Group Deliverable

(09/2023)

Focus Group on Artificial Intelligence for Health

(FG-AI4H)

FG-AI4H DEL10.2

Topic description document for the Topic Group on Al-based Dermatology (TG-Derma)



ITU-T FG-AI4H Deliverable

DEL10.2 – Topic description document for the Topic Group on AI-based Dermatology (TG-Derma)

Summary

This topic description document (TDD) specifies standardized benchmarking for AI-based dermatology, and covers all scientific, technical and administrative aspects relevant for setting it up.

Keywords

Artificial intelligence, benchmarking; health, topic groups; overview, ethics; regulations, data quality, data audit, clinical relevance, topic description, dermatology.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1 of the Deliverable DEL10.2 on *FG-AI4H topic description document for the Topic Group on AI-based Dermatology (TG-Derma)* approved on 15 September 2023 via the online approval process for the ITU-T Focus Group on AI for Health (FG-AI4H).

Editors:	Harsha Jayakody Flash Health Sri Lanka	E-mail:	<u>harsha@flash.health</u>
	Ivy Lee the American Academy of Dermatology	E-Mail:	ivyannlee@gmail.com
Contributors:	(in alphabetical order)		
	Harsha Jayakody Flash Health Sri Lanka	Tel: +94 E-mail:	47 7431 3048 <u>harsha@flash.health</u>
	Ivy Lee The American Academy of Dermatology USA	Tel: +1 E-Mail:	626 233 0275 ivyannlee@gmail.com

© ITU 2025

Some rights reserved. This publication is available under the Creative Commons Attribution-Non Commercial-Share Alike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <u>https://creativecommons.org/licenses/by-nc-sa/3.0/igo</u>).

For any uses of this publication that are not included in this licence, please seek permission from ITU by contacting <u>TSBmail@itu.int</u>.

Table of Contents

Page

1	Introduction		
2	About the FG-AI4H Topic Group on AI-based Dermatology		
	2.1	Documentation	1
	2.2	Status of the Topic Group on AI-based Dermatology	1
	2.3	Topic Group participation	3
3 Topic description			3
	3.1	Definition of the artificial intelligence task	3
	3.2	Current gold standard	3
	3.3	Relevance and impact of an artificial intelligence solution	4
	3.4	Existing artificial intelligence solutions	4
	3.5	Leadership and strategies with related to artificial intelligence solutions	5
4	Ethical considerations		
5	5 Existing work on benchmarking		
6	Benchmarking by the topic group		
7	7 Overall discussion of the benchmarking		
8	Regulatory considerations		
Annex A – Glossary			9
Biblio	graphy		10

ITU-T FG-AI4H Deliverable

DEL10.2 – Topic description document for the Topic Group on AI-based Dermatology (TG-Derma)

1 Introduction

Artificial intelligence (AI), which incorporates the concept of computers and digital systems synthesizing and inferring new information, has recently been at the forefront of computer science research. Recent advances in digital healthcare and AI have led to new developments in dermatology.

In this context, application of AI in contemporary dermatology ranges from the use of AI in melanoma [1] to atopic dermatitis [1]. AI integration in dermatology has also progressed into public-facing tools and applications, such as Google's DermAssist [2] for identification of various skin conditions.

This topic description document (TDD) specifies standardized benchmarking for Topic Group on AIbased Dermatology (TG-Derma) systems. It serves as deliverable No. 10.2 of the International Telecommunication Union/World Health Organization (ITU/WHO) Focus Group on AI for Health (FG-AI4H).

2 About the FG-AI4H Topic Group on AI-based Dermatology

Clause 1 highlights the potential of standardized benchmarking of AI systems for dermatology to solve important health issues and provide decision-makers with the necessary insight to successfully address these challenges.

To develop this benchmarking framework, FG-AI4H decided to create TG-Derma at meeting P in Helsinki, Finland, 19-22 Sep. 2022.

FG-AI4H assigns a *topic driver* to each topic group (TG) (similar to a moderator) who coordinates the collaboration of all its members on the TDD. During meeting P, Dr Harsha Jayakody from Flash Health, Sri Lanka was appointed as TG-Derma topic driver.

2.1 Documentation

This document is the TDD for the TG-Derma. It introduces the health topic, including the AI task, outlines its relevance and the potential impact that benchmarking will have on health systems and patient outcomes, and provides an overview of existing AI solutions for dermatology. It describes existing approaches for assessing the quality of dermatology systems and provides details that are likely to be relevant for setting up a new standardized benchmarking. It specifies actual benchmarking methods for all subtopics at a level of detail that includes technological and operational implementation. There are individual clauses for all versions of the benchmarking. Finally, it summarizes the results of the TG-Derma initiative and runs of benchmarking. In addition, the TDD addresses ethical and regulatory aspects.

The TDD was developed cooperatively by all members of the TG-Derma over time and updated TDD iterations were presented at each FG-AI4H meeting.

2.2 Status of the Topic Group on AI-based Dermatology

The following paragraphs describe the update of the collaboration within the TG-Derma for the official FG-AI4H meetings.

Regular FG-AI4H workshops and meetings proceed about every two months at various locations around the globe or remotely. More information can be found on the official FG-AI4H website [3].

During meeting P, Harsha Jayakody was named as the leader of TG-Derma. Following the meeting, Eva Weicken linked Ivy Lee to the group. Harsha Jayakody together with Ivy Lee contacted all registered participants who had shared interest in joining the TG-Derma group via email. The call for topic group participation (CfTGP) was updated and shared.

Highlights of meeting Q, Doula. 6-9 December 2022, follow.

- Harsha Jayakody presented Elza, an AI solution developed in Sri Lanka, which is an AI application (app) that detects the cosmetic concerns of an individual and gives a range for each cosmetic concern following aesthetic skin treatment options to the end-user.
- Furthermore, the module in Figure 1 was discussed at the meeting.



Figure 1 – Elza image-processing data flow

- Differentiation of skin colour with ethnicities was discussed in relation to the model.
- At the end of the meeting, Ivy Lee was named as a co-topic driver with Harsha Jayakody.
- Post meeting discussions were held between Harsha Jayakody and Ivy Lee who were the two active members of the TG-Derma on different applications used in different countries.

Highlights of meeting R in Boston, 21-24 March 2023, follow.

- Harsha Jayakody presented the four areas where AI is used in dermatology, for: automated diagnosis; disease classification; treatment planning; and tele-dermatology.
- In the first half of the presentation, the SkinVision app and Milskin PRO were discussed in detail.
- The second half of the presentation by Ivy Lee showcased leadership and strategy in Tanzania and Great Britain on AI apps used for dermatology.
- Furthermore, building a community to identify academic, commercial, and regulatory stakeholders was stressed in the presentation.

Highlights of meeting S, Geneva, 3-4 July 2023, follow.

– Ivy Lee presented the American Academy of Dermatology approach to AI for dermatology.

- Ivy Lee further explained the assistive role of FG-AI4H to enhance human intelligence and the physician/patient relationship rather than replace it.

2.3 Topic Group participation

The participation in both the FG-AI4H and in a TG were generally open to anyone (with a free ITU account). For TG-Derma, at the time of publication, the most recent corresponding CfTGP) can be found here:

https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/docs/FGAI4H-R-007-A02.docx

Each TG also has a corresponding subpage on the ITU collaboration site. The subpage for TG-Dermagroup can be found here:

https://extranet.itu.int/sites/itu-t/focusgroups/ai4h/tg/SitePages/TG-Derma.aspx

For participation in TG-Derma, interested parties can also join regular online meetings. For all TGs, the link will be the standard ITU-TG Zoom link:

– <u>https://itu.zoom.us/my/fgai4h</u>

All relevant administrative information about FG-AI4H, such as upcoming meetings or document deadlines, are announced via the general FG-AI4H mailing list <u>fgai4h@lists.itu.int</u>.

In addition to the general one for FG-AI4H, TG-Derma used a specific mailing list:

– <u>fgai4htgderma@lists.itu.int</u> (mailing list archive: <u>https://itu.int/ml/lists/arc/fgai4htgderma</u>)

Regular FG-AI4H workshops and meetings proceed about every two months at changing locations around the globe or remotely. More information can be found on the official FG-AI4H website [3].

3 Topic description

This clause contains a detailed description of and background information specifically about the health topic for the benchmarking of AI in dermatology and how this can help to solve a relevant real-world problem.

TGs summarize related benchmarking AI subjects to reduce redundancy, leverage synergies and streamline FG-AI4H meetings. However, in some cases, different subtopic groups can be established within one TG to pursue different topic-specific fields of expertise. TG- Derma has no subtopics at the time of publication.

3.1 Definition of the artificial intelligence task

The description of the AI task for TG-Derma involves clearly specifying AI tasks and objectives to be addressed within the scope of benchmarking health AI models in dermatology. This encompasses tasks such as skin disease classification, lesion segmentation, disease severity assessment, treatment recommendation, and other relevant diagnostic and prognostic tasks. Each task will be further specified to include target skin conditions, datasets and evaluation metrics. The AI task definitions will provide a clear roadmap for TG-Derma members, enabling them to focus their efforts on developing and evaluating AI models that address critical challenges in dermatological diagnosis and patient care. These definitions will be continuously refined and updated as new insights and advancements in the field emerge, ensuring that the benchmarking process remains relevant and aligned with the latest developments in AI for dermatology.

3.2 Current gold standard

In previous meetings of the TG-Derma, discussion of current gold standards used in AI for dermatology applications worldwide revealed two prominent approaches.

The first involves utilizing expert dermatologist consensus or a panel of experienced clinicians to establish a reference standard for skin disease diagnosis and classification. These experts carefully

3

review and analyse clinical data, histopathological findings and other diagnostic tests to reach a consensus on the accurate diagnosis for each case. This consensus-based gold standard serves as a benchmark against which AI model performance can be evaluated, providing a reliable measure of diagnostic accuracy and effectiveness.

The second is the utilization of standardized and well-curated dermatological image datasets that have been extensively reviewed and annotated by dermatologists. These datasets often contain a diverse range of skin conditions, capturing variations in ethnicity, age and disease severity. They are meticulously labelled with ground-truth annotations, ensuring that each image is associated with the correct diagnosis or segmentation mask. By leveraging such standardized datasets as gold standards, researchers and developers can compare the performance of their AI models against a uniform and widely accepted reference, allowing for more consistent and meaningful evaluations across different research groups and institutions. Emphasizing the use of standardized datasets encourages reproducibility, enables better collaboration and fosters advancements in AI for dermatology on a global scale.

3.3 Relevance and impact of an artificial intelligence solution

During previous meetings, the relevance and impact of AI solutions in the field of dermatology were extensively discussed. AI has shown great promise in dermatological applications, providing valuable support to healthcare professionals in diagnosing skin conditions, predicting disease progression, and recommending personalized treatment plans. The relevance of AI in dermatology lies in its ability to process and analyse vast amounts of dermatological data, enabling faster and more accurate diagnoses, which can significantly improve patient outcomes and reduce the burden on healthcare systems.

The potential impact of AI in dermatology was highlighted through various use cases, such as early detection of skin cancer, automating lesion segmentation, and assisting in the identification of rare skin diseases. By harnessing the power of AI, dermatologists can enhance their decision-making process, leading to more precise and tailored treatment strategies for patients. Moreover, AI solutions can facilitate telemedicine initiatives, extending dermatological expertise to remote and underserved regions, thereby promoting equitable access to quality healthcare services. However, challenges were also acknowledged, including the need for robust data privacy measures, ethical considerations, and ensuring AI solutions are validated and integrated into clinical workflows effectively. The discussions aimed to underscore the potential benefits while promoting responsible development and deployment of AI solutions in dermatology, ensuring they align with the highest standards of patient care and safety.

3.4 Existing artificial intelligence solutions

During the previous meetings, the AI solutions Elza app, SkinVision app and Milskin PRO app were discussed, and their inputs and outputs relevant for benchmarking examined. These AI solutions focus on dermatological applications, particularly skin disease detection and diagnosis. The inputs for benchmarking include dermatological images captured by users through smartphone cameras or uploaded from other sources. These images represent various skin conditions, ranging from common skin disorders to potentially malignant lesions. The AI models process these images and provide outputs in the form of diagnostic predictions, lesion segmentation, disease severity scores and treatment recommendations. The benchmarking process will need to consider the diversity and complexity of the skin conditions these AI solutions aim to address, as this will influence the choice of evaluation datasets and performance metrics.

The scope of these AI systems also impacts the specification of categories for benchmarking reports, metrics and scores. Each AI solution might excel in different aspects, such as accuracy in diagnosing specific skin conditions, segmentation precision for particular lesion types or efficiency in providing real-time responses. Therefore, the benchmarking criteria should encompass multiple performance

dimensions, allowing for a comprehensive evaluation of the strengths and weaknesses of each system. Moreover, factors like user-friendliness, integration with existing healthcare workflows, and adherence to ethical and privacy standards are essential aspects that need to be taken into account during benchmarking. By considering the scope and capabilities of these AI solutions, the TG can design benchmarking protocols that holistically assess their performance and ultimately contribute to advancing the field of AI in dermatology for improved patient care and outcomes.

3.5 Leadership and strategies with related to artificial intelligence solutions

This clause is dedicated to showcasing leadership and strategy in AI applications used for dermatology in Tanzania and Great Britain that were discussed during the meetings. This clause highlights the initiatives, policies, and achievements in both countries concerning the implementation of AI solutions in the field of dermatology.

In Tanzania, the leadership in AI for dermatology is demonstrated through the establishment of partnerships between the government, academic institutions and healthcare organizations. Collaborative efforts have been made to develop AI-driven tools and applications that aim to improve dermatological care in underserved regions. The strategy in Tanzania involves leveraging mobile technology to reach remote areas, providing access to dermatological expertise through AI-powered mobile applications. These initiatives focus on early detection of skin conditions, patient education and improving the efficiency of healthcare delivery. During future meetings, the co-drivers will provide a detailed overview of the specific AI applications implemented in Tanzania, the datasets used, the evaluation metrics employed and the impact achieved.

In Great Britain, the leadership in AI for dermatology is characterized by a strong emphasis on research, innovation and regulatory frameworks. The country's strategy involves fostering collaborations between research institutions, industry stakeholders and regulatory bodies to ensure the safe and effective integration of AI solutions in clinical practice. In future, the topic drivers plan to showcase successful case studies of AI applications used in dermatology within the British healthcare system. This includes AI models for skin cancer diagnosis, image analysis for lesion detection and segmentation, and the use of AI for patient risk stratification and treatment recommendations. Additionally, the document will highlight the regulatory guidelines and ethical considerations in place to govern the development and deployment of AI solutions in dermatology in Great Britain to be described during future meetings.

Furthermore, the document aims to inspire other countries and organizations to adopt best practice, foster collaborations and promote responsible AI deployment in the field of dermatology for the benefit of patients worldwide.

4 Ethical considerations

Ethical considerations are of paramount importance in the application of AI in the field of dermatology, as these technologies directly impact patient care, privacy and medical decision-making. During future engagements, TG drivers hope to complete the document, addressing the key ethical principles to ensure responsibility and trustworthy deployment.

First and foremost, data privacy and patient confidentiality must be rigorously safeguarded during the collection, storage and processing of sensitive dermatological information. Clear consent protocols should be established to inform patients about the use of their data for AI model training and validation. Transparency in AI model development is vital, with an emphasis on understanding how the algorithms reach their conclusions to enable clinicians to validate and explain the outputs to patients. Additionally, the potential biases in AI models should be thoroughly investigated and mitigated to prevent disparities in healthcare outcomes. Regular audits and assessments of AI systems should be conducted to identify and rectify any issues that may arise.

5

Ensuring equitable access to AI-driven dermatological solutions is also crucial, with a focus on addressing disparities in healthcare delivery across different patient populations. By adhering to stringent ethical guidelines, the AI in dermatology community can build trust, foster collaboration, and realize the full potential of AI technologies in improving patient care and enhancing the field of dermatology. During future meetings, these ethical considerations will be discussed further.

5 Existing work on benchmarking

The topic drivers have made substantial progress in exploring existing work on benchmarking AI for dermatology applications, which has been a central focus of previous meetings. A thorough review has been undertaken of relevant literature and research papers that outline various benchmarking approaches and methodologies used to evaluate AI models in dermatological diagnosis, lesion segmentation, disease classification and treatment recommendation. By studying these existing benchmarks, valuable insight has been gained into the current state-of-the-art in AI for dermatology and the strengths and limitations identified of different evaluation frameworks.

However, despite the progress made, there is still much more work to be done in the field of benchmarking AI for dermatology. In future meetings, further development of standardized datasets and protocols will be explored for benchmarking AI models. This includes curating diverse and representative datasets that encompass a wide range of skin conditions and variations to ensure comprehensive evaluations. The use of real-world clinical data while maintaining data privacy and ethical considerations will also be examined.

Additionally, future discussions will focus on refining and expanding the selection of evaluation metrics to capture various aspects of AI model performance. This will enable a more nuanced assessment of the different capabilities of AI solutions and allow for fair comparisons across different algorithms and applications. Moreover, the importance of transparency and reproducibility in benchmarking will be emphasized, encouraging AI developers to share their code and methodologies openly.

The need is recognized to address challenges related to generalizability and scalability of AI models in dermatology. As progress is made, discussions will involve exploring ways to validate AI solutions across different healthcare settings and geographic regions to ensure their applicability in diverse clinical scenarios. The WHO-ITU eHealth strategy toolkit has been identified as a model to discuss the ethical considerations in each of these AI driven applications in dermatology in the future. See Figure 2.



Figure 2 – WHO-ITU eHealth Strategy toolkit

During future meetings, progress on these important aspects of benchmarking AI for dermatology will be presented and documented. By collaboratively sharing knowledge, experience and insights, a comprehensive and robust benchmarking framework will be shaped that serves as a guiding resource for the wider AI in dermatology community. This ongoing work will be crucial in fostering

advancements in AI applications for dermatology, ultimately leading to improved patient care, early detection and optimized treatment strategies.

6 Benchmarking by the topic group

This clause describes all technical and operational details regarding the benchmarking process for the TG-Derma AI task including clauses for each version of the benchmarking that is iteratively improved over time.

It reflects the considerations of the various following deliverables.

<u>DEL5</u>* *Data specification* (introduction to deliverables 5.1-5.6).

<u>DEL5.1</u> *Data requirements* (which lists acceptance criteria for data submitted to FG-AI4H and states the governing principles and rules).

DEL5.2* Data acquisition.

DEL5.3 Data annotation specification.

<u>DEL5.4</u> *Training and test data specification* (which provides a systematic way of preparing technical requirement specifications for datasets used in training and testing of AI models).

DEL5.5 Data handling (which outlines how data will be handled once they are accepted).

<u>DEL5.6</u>* *Data sharing practices* (which provides an overview of the existing best practices for sharing health-related data based on distributed and federated environments, including the requirement to enable secure data sharing and addressing issues of data governance).

<u>DEL6</u> AI training best practices specification (which reviews best practices for proper AI model training and guidelines for model reporting).

<u>DEL7</u>* *AI for health evaluation considerations* (which discusses the validation and evaluation of AI for health models, and considers requirements for a benchmarking platform).

<u>DEL7.1</u>* *AI4H evaluation process description* (which provides an overview of the state of the art of AI evaluation principles and methods and serves as an initiator for the evaluation process of AI for health).

DEL7.2 AI technical test specification (which specifies how an AI can and should be tested in silico).

<u>DEL7.3</u>* *ML4H trial audits–Iteration 2.0 Playbook (Version 3.0).*

<u>DEL7.4</u> *Clinical evaluation of AI for health* (which outlines the current best practice and outstanding issues related to clinical evaluation of AI models for health).

<u>DEL7.5</u>* Assessment platform (which explores assessment platform options that can be used to evaluate AI for health for the different topic groups).

<u>DEL9</u>* *AI for health applications and platforms* (which introduces specific considerations of the benchmarking of mobile- and cloud-based AI applications in health)-

DEL9.1* Mobile applications.

<u>DEL9.2</u>* *Cloud-based AI applications* (which describe specific requirements for the development, testing and benchmarking of mobile- and cloud-based AI applications).

Editorial Note: Documents marked with an asterisk () were initially planned, but where not approved for publication.

7 Overall discussion of the benchmarking

This clause discusses the overall insights gained from benchmarking work in TG-Derma. This should not be confused with the discussion of the results of a concrete benchmarking run. This will be completed later.

8 Regulatory considerations

For AI-based technologies in healthcare, regulation is not only crucial to ensure the safety of patients and users, but also to accomplish market acceptance of these devices. This is challenging because there is a lack of universally accepted regulatory policies and guidelines for AI-based medical devices. To ensure that the benchmarking procedures and validation principles of FG-AI4H are secure and relevant for regulators and other stakeholders, the working group on <u>Regulatory considerations</u> <u>on AI for health (WG-RC)</u> compiled the requirements that consider these challenges.

This clause will be completed later.

Annex A

Glossary

This annex lists all the relevant abbreviations, acronyms and uncommon terms used in the document.

Acronym/Term	Expansion
AI	Artificial Intelligence
AI4H	Artificial Intelligence for Health
CfTGP	Call for Topic Group Participation
TDD	Topic Description Document
TG	Topic Group
WG	Working Group

Bibliography

- [1] Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SM, Jafari MH, Ward K, Najarian K (2016), *Melanoma detection by analysis of clinical images using convolutional neural network*, Annu Int Conf Proc IEEE Eng Med Biol Soc. 2016:1373–6. doi: 10.1109/EMBC.2016.7590963
- [2] Google Health. *DermAssist: A whole new way to help identify your skin conditions* [Internet]. Mountain View, CA: Google. Available [viewed 2023-12-27] at: <u>https://health.google/consumers/dermassist/</u>
- [3] Focus Group on Artificial Intelligence for Health. *FG-AI4H* (2023). Geneva: International Telecommunication Union. Available [viewed 2023-12-27] at: <u>https://itu.int/go/fgai4h</u>
- [4] Rajpurkar, P., Irvin, J., Ball, R.L., Zhu, K., Yang, B., Mehta, H., and Ng, A. Y. (2017), Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists, PLoS medicine, 15(11), e1002686.
- [5] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017), *Dermatologist-level classification of skin cancer with deep neural networks*, Nature, 542(7639), 115-118.
- [6] Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., and Hofmann-Wellenhof, R. (2018), *Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists*, Annals of Oncology, 29(8), 836-1842.