

International Telecommunication Union

ITU-T FG-AI4H Deliverable

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

15 September 2023

PRE-PUBLISHED VERSION

DEL10.2

**FG-AI4H Topic Description Document for the
Topic Group on AI-based dermatology (TG-
Derma)**

ITU-T

Summary

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

Keywords

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

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: harsha@flash.health
	Ivy Lee the American Academy of Dermatology	E-Mail: ivyannlee@gmail.com

Contributors: (in alphabetical order)		
	Harsha Jayakody Flash Health Sri Lanka	Tel: +947 7431 3048 E-mail: harsha@flash.health
	Ivy Lee The American Academy of Dermatology USA	Tel: +1 626 233 0275 E-Mail: ivyannlee@gmail.com

CONTENTS

	Page
1 Introduction.....	3
2 About the FG-AI4H topic group on Dermatology.....	3
2.1 Documentation.....	3
2.2 Status of this topic group	3
2.3 Topic Group participation.....	5
3 Topic description	5
3.1 Definition of the AI task	5
3.2 Current gold standard	6
3.3 Relevance and impact of an AI solution.....	6
3.4 Existing AI solutions	6
3.5 Leadership and strategies with related to AI solutions	7
4. Ethical considerations	7
5. Existing work on benchmarking	8
6. Benchmarking by the topic group.....	9
7. Overall discussion of the benchmarking.....	10
8. Regulatory considerations.....	10
References	10
Annex A: Glossary	11

List of Figures

	Page
Figure 2.2 – ELZA image processing data flow	4
Figure 5.1 – WHO-ITU eHealth Strategy toolkit.....	9

ITU-T FG-AI4H Deliverable 10.2

FG-AI4H Topic Description Document for the Topic Group on AI-based dermatology (TG-Derma)

1 Introduction

Artificial intelligence, 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 artificial intelligence have led to a new development in dermatology.

In this context, application of AI in contemporary dermatology ranges from the use of AI in melanoma¹ to atopic dermatitis². AI integration in dermatology has also progressed into public-facing tools and applications, such as Google's DermAssist³ for identifying various skin conditions.

This topic description document specifies the standardized benchmarking for TG Dermatology systems. It serves as deliverable No. 10.2 of the ITU/WHO Focus Group on AI for Health (FG-AI4H).

2 About the FG-AI4H topic group on Dermatology

The introduction highlights the potential of a standardized benchmarking of AI systems for Dermatology to help 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 the TG-Dermatology at the meeting P in Helsinki, Finland, 19-22 Sep. 2022.

FG-AI4H assigns a *topic driver* to each topic group (similar to a moderator) who coordinates the collaboration of all topic group members on the TDD. During FG-AI4H meeting TG-Dermatology at the meeting P in Helsinki, Finland, 19-22 Sep. 2022, Dr Harsha Jayakody from Flash health, Sri Lanka was appointed as topic driver for the TG-Dermatology.

2.1 Documentation

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

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

2.2 Status of this topic group

The following subsections describe the update of the collaboration within the TG-Dermatology for the official focus group meetings.

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 at <https://itu.int/go/fgai4h>.

During the meeting P in Helsinki from 19-22 September 2022, Harsha Jayakody was named as the topic group leader for TG Dermatology. Following the meeting, Eva Weicken via a zoom meeting linked Ivy Lee to the group. Harsha Jayakody together with Ivy Lee contacted the all registered participants who had shared interest to join TG Dermatology group via email. "Call for Topic Group Participation" document was updated and shared.

During the meeting Q in Doula from 6-9 December 2022;

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

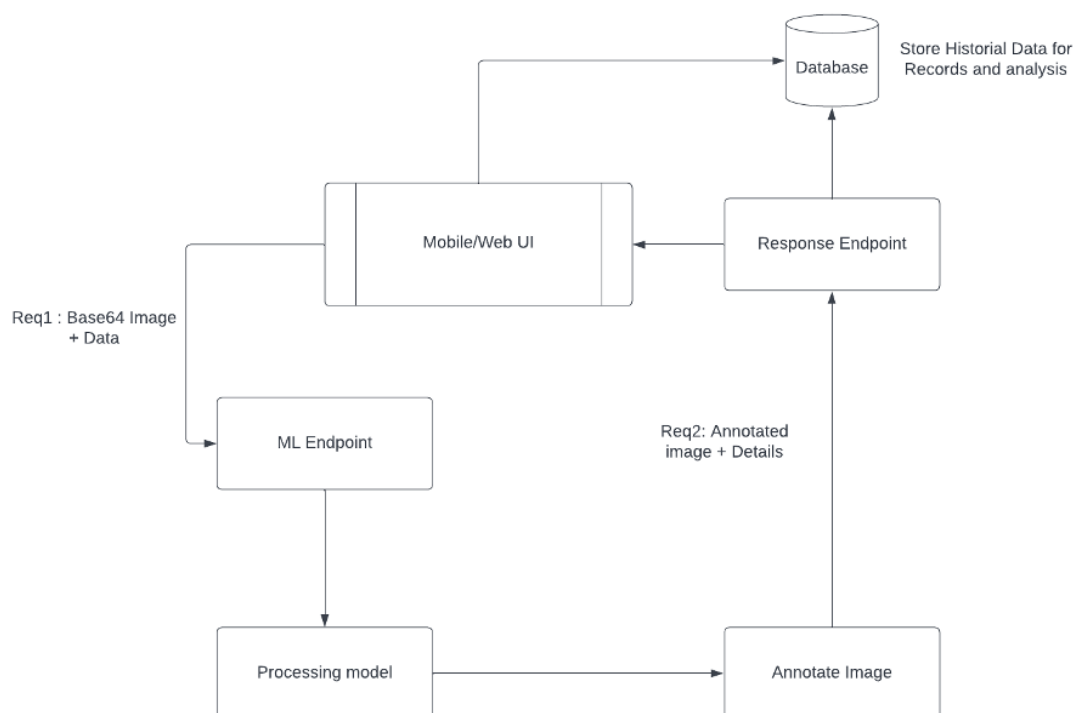


Figure 2.2 – ELZA image processing data flow

- During the discussion differentiation of different skin colour with ethnicities was discussed with the model.
- At the end of the meeting Ivy Lee from LA 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 Dermatology on different applications used in different countries.

During the meeting R in Boston from 21-24 March 2023;

- Harsha Jayakody presented the four areas where AI is used in Dermatology such as for automated diagnosis, disease classification, treatment planning & tele-dermatology
- First half of the presentation, "Skin Vision" application and "Milskin PRO" was discussed in detail.
- Ivy Lee presented the second half of the presentation on showcasing leadership and strategy in Tanzania and Great Britain on AI applications used for Dermatology.

- Furthermore, building a community to identify academic, commercial, and regulatory stakeholders was stressed in the presentation.

During the meeting S in Geneva from 3-4 July 2023:

- Ivy Lee presented the American Academy of Dermatology approach on AI for Dermatology
- Further explained that the Focus on AI's assistive role to enhance human intelligence and the physician/patient relationship rather than replace it.

2.3 Topic Group participation

The participation in both, the Focus Group on AI for Health and in a TG were generally open to anyone (with a free ITU account). For this TG, the most recent corresponding "Call for TG participation" (CfTGP) can be found here:

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

Each topic group also has a corresponding subpage on the ITU collaboration site. The subpage for this topic group can be found here:

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

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

- <https://itu.zoom.us/my/fgai4h>

All relevant administrative information about FG-AI4H—like upcoming meetings or document deadlines—were announced via the general FG-AI4H mailing list fgai4h@lists.itu.int.

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

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

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:

- <https://itu.int/go/fgai4h>

3 Topic description

This section contains a detailed description and background information of the specific health topic for the benchmarking of AI in Dermatology and how this can help to solve a relevant 'real-world' problem.

Topic Groups 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 topic group to pursue different topic-specific fields of expertise. The TG-Dermatology has no subtopics currently.

3.1 Definition of the AI task

The Definition of the AI task for the AI in Dermatology Topic Group involves clearly defining the specific 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 the target skin conditions, datasets, and evaluation metrics. The AI task definitions will provide a clear roadmap for the topic group 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 the previous meetings of the AI in Dermatology Topic Group, the discussion on current gold standards used in AI for dermatology applications worldwide revealed two prominent approaches. The first approach involves utilizing expert dermatologists' consensus or a panel of experienced clinicians to establish a reference standard for skin disease diagnosis and classification. These experts carefully 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 models' performance can be evaluated, providing a reliable measure of diagnostic accuracy and effectiveness.

The second approach discussed 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 AI 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 AI solutions

During the previous meetings, the AI solutions "ELZA" application, "Skin Vision" application, and "Milskin PRO" application were discussed, and their inputs and outputs relevant for benchmarking were 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 definition 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 each system's strengths and weaknesses. 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 topic group 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 AI solutions

In this document on AI for Dermatology, this section is dedicated to showcasing leadership and strategy in AI applications used for Dermatology in Tanzania and Great Britain which was discussed during the meetings. This section 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 the 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 the future the topic drivers are planning 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 be highlighting the regulatory guidelines and ethical considerations in place to govern the development and deployment of AI solutions in Dermatology in Great Britain during the future meetings.

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

4. Ethical considerations

Ethical considerations hold 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 the future engagements, topic group drivers are hoping to complete the document in the future 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 the AI systems should be conducted to identify and rectify any issues that may arise.

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 the future meetings these ethical considerations will be discussed further.

5. Existing work on benchmarking

As topic drivers, we have made significant progress in exploring existing work on benchmarking AI for Dermatology applications, which has been a central focus of our previous meetings. We have conducted a thorough review 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, we have gained valuable insights into the current state-of-the-art in AI for Dermatology and identified the strengths and limitations 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 our future meetings, we will further delve into the development of standardized datasets and protocols 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. We will also explore the use of real-world clinical data while maintaining data privacy and ethical considerations.

Additionally, our 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 different AI solutions' capabilities and allow for fair comparisons across different algorithms and applications. Moreover, we will emphasize the importance of transparency and reproducibility in benchmarking, encouraging AI developers to share their code and methodologies openly.

We recognize the need to address challenges related to generalizability and scalability of AI models in dermatology. As we move forward, our discussions will involve exploring ways to validate AI solutions across different healthcare settings and geographic regions to ensure their applicability in diverse clinical scenarios. We have identified the WHO-ITU eHealth strategy toolkit as a model to discuss the ethical considerations in each of these AI driven applications in Dermatology in the future.

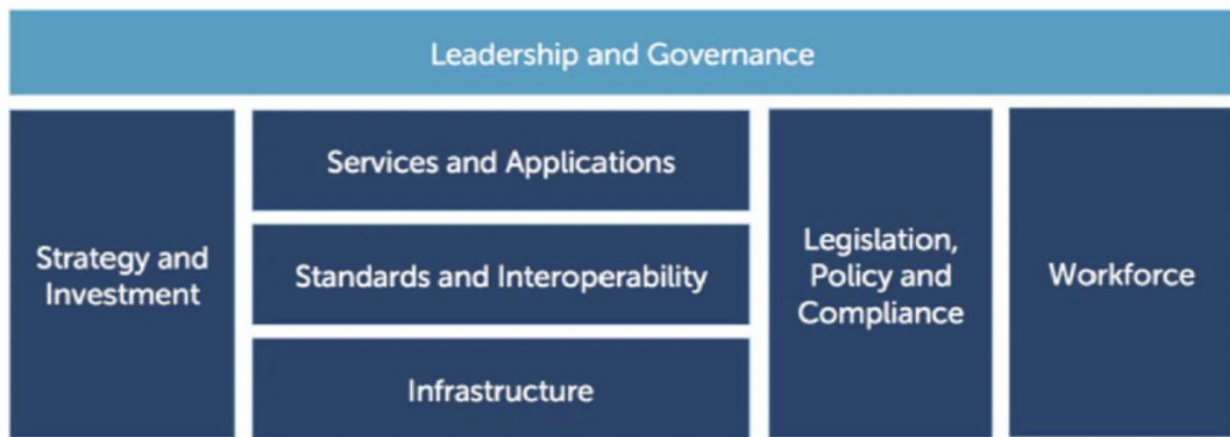


Figure 5.1 – WHO-ITU eHealth Strategy toolkit

During future meetings, we will present and document our progress on these important aspects of benchmarking AI for Dermatology. By collaboratively sharing knowledge, experiences, and insights, we will shape a comprehensive and robust benchmarking framework 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 section describes all technical and operational details regarding the benchmarking process for the TG Derma AI task including subsections for each version of the benchmarking that is iteratively improved over time.

It reflects the considerations of various deliverables: [DEL5](#) “Data specification” (introduction to deliverables 5.1-5.6), [DEL5.1](#) “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”, [DEL5.4](#) “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), [DEL5.6](#) “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), [DEL6](#) “AI training best practices specification” (which reviews best practices for proper AI model training and guidelines for model reporting), [DEL7](#) “AI for health evaluation considerations” (which discusses the validation and evaluation of AI for health models, and considers requirements for a benchmarking platform), [DEL7.1](#) “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*), [DEL7.3](#) “Data and artificial intelligence assessment methods (DAISAM)” (which provides the reference collection of WG-DAISAM on assessment methods of data and AI quality evaluation), [DEL7.4](#) “Clinical Evaluation of AI for health” (which outlines the current best practices and outstanding issues related to clinical evaluation of AI models for health), [DEL7.5](#) “FG-AI4H assessment platform” (which explores assessment platform options that can be used to evaluate AI for health for the different topic groups), [DEL9](#) “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 based AI applications,” and [DEL9 2](#) “Cloud-based AI applications”

(which describe specific requirements for the development, testing and benchmarking of mobile- and cloud-based AI applications).

This section will be completed later.

7. Overall discussion of the benchmarking

This section discusses the overall insights gained from benchmarking work in this topic group. 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 “*Regulatory considerations on AI for health*” (WG-RC) compiled the requirements that consider these challenges.

This section will be completed later.

References

- [1] Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using convolutional neural network. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:1373–6.
- [2] Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using convolutional neural network. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:1373–6.
- [3] Identify skin conditions with DermAssist [Internet]. Google Health. [cited 2022Dec4]. Available from: <https://health.google/consumers/dermassist/>
- [4] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & 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., & 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., ... & 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), 1836-1842.

Annex A: Glossary

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

Acronym/Term	Expansion	Comment
AI	Artificial intelligence	
AI4H	Artificial intelligence for health	
AI-MD	AI based medical device	
API	Application programming interface	
CfTGP	Call for topic group participation	
DEL	Deliverable	
FDA	Food and Drug administration	
FGAI4H	Focus Group on AI for Health	
GDP	Gross domestic product	
GDPR	General Data Protection Regulation	
IMDRF	International Medical Device Regulators Forum	
IP	Intellectual property	
ISO	International Standardization Organization	
ITU	International Telecommunication Union	
LMIC	Low-and middle-income countries	
MDR	Medical Device Regulation	
PII	Personal identifiable information	
SaMD	Software as a medical device	
TDD	Topic Description Document	Document specifying the standardized benchmarking for a topic on which the FG AI4H Topic Group works. This document is the TDD for the Topic Group TG Derma
TG	Topic Group	
WG	Working Group	
WHO	World Health Organization	