Examine the Influence of AI Tools on the Workflow of Dermatology Practice, Focusing on Diagnostic Efficiency, Patient Management, and Clinician Decision-Making

Haifa Hassan Jannah[1], Dalia Fahad Ramadan[2], Ranad Mohammed Khashab[2], Shokran Faisal Sulaimani[2], Abeer Muala Alahmadi[3], Nada Abdulghani Sarooji[4], Hanan Fawaz Sabban[5], Saud Ghalib Almahdaly[6], Sara Waleed Hefni[2]

[1]*Consultant of Dermatology and Head of Dermatology Department in Pioneer Care Polyclinic, Saudi Arabia.*
[2]*MBBS Program, Fakeeh College for Medical Sciences, 21461, Jeddah, Saudi Arabia.*
[3]*Post Graduate Program of Family Medicine, Jeddah First Health Cluster, – Jeddah, Saudi Arabia.*
[4]*Jeddah First Health Cluster, – Jeddah, Saudi Arabia.*
[5]*King Abdullah Medical Complex - Academic Affair Department, Ministry of Health – Jeddah, Saudi Arabia.*
[6]*East Jeddah Hospital Health Cluster 1, Saudi Arabia.*

Abstract: Introduction: Skin conditions are an important health issue. On average, individuals experience 1.6 skin conditions each year, with skin-related doctor appointments making up 20% of all primary care visits, of which approximately 35% are directed to a dermatologist. Machine learning (ML) models have the potential to assist primary care professionals by analyzing and enhancing intricate datasets. Furthermore, ML models are being more commonly used in dermatology for aiding in diagnosis through image analysis, particularly for detecting and categorizing skin cancer.

Aim: This research seeks to validate a machine learning image analysis model prospectively as a diagnostic aid for diagnosing dermatological conditions.

Method: In this upcoming study, 100 successive patients seeking care for a skin issue from a participating general practitioner (GP) in central Catalonia were selected. The planned duration for data collection was set at 7 months. Anonymized images of skin conditions were captured and fed into the ML model interface (able to detect 44 different skin conditions), which provided the top 5 diagnoses with the highest probability. The identical picture was also transmitted for a teledermatology consultation in accordance with the established workflow. The GP, ML model, and dermatologist's evaluations will be compared to determine the precision, sensitivity, specificity, and accuracy of the ML model. Each type of skin disease class will have its results displayed globally and individually through a confusion matrix and the one-versus-all approach. The amount of time needed to conduct the diagnosis will also be factored in.

Results: Patient enrollment started in June 2021 and continued for a duration of 5 months. At present, all participants have been enrolled and the images have been presented to the GPs and dermatologists. The examination of the findings has already commenced.

Conclusion: This research will offer insights into the efficacy and constraints of ML models. External testing is necessary for controlling these diagnostic systems for the implementation of ML models in a primary
care environment.

1. Introduction

Health care systems in Western nations are facing a growing number of difficulties such as rising demand, older populations, long-term illnesses, multiple health conditions, and the ongoing global pandemic. These elements, along with the shortage of professionals like general practitioners (GPs), lead to the necessity of discovering innovative solutions to enhance both the quality of care and the efficiency of professionals [1].

Skin conditions are a notable health issue, and dermatological problems are a primary factor for patients seeking care from their general practitioners. Each person typically experiences 1.6 skin conditions annually [2]. Around one-fifth of GP appointments are linked to skin concerns, while around 15% of telehealth appointments are related to dermatology. [3,4] Approximately 7.6% of Catalonia's overall population seek care at a primary care center each year for skin issues, with approximately 35% being directed to a dermatologist. [5,6] Currently, in central Catalonia's healthcare sector, teledermatology appointments are frequently utilized to direct patients to a dermatologist located in a hospital. It is predicted that over 70% of PCC patients experiencing a skin issue can be efficiently evaluated through teledermatology and avoid an in-person appointment with a dermatologist [7,8].

Computer-assisted diagnosis has been used in radiology since the 1960s in the field of medicine. The first mention of artificial intelligence (AI) in dermatopathology can be traced back to 1987 with the creation of the text-based system TEGUMENT. TEGUMENT featured a semantic tree containing 986 possible diagnoses to help dermatologists with the histopathologic differential diagnosis of skin diseases and tumors. Computer-aided melanoma diagnosis was first implemented in dermatology in the early 2000s, utilizing rule-based classifiers that use predefined features to categorize images into specific groups. [9,10].

The use of teledermatology has grown globally in recent years. It has been widely used in numerous PCC environments and has been strongly supported by extensive research as an effective method of triage, especially for skin cancer lesions [11]. Research comparing the overall precision of in-person dermatology appointments with teledermatology shows varying outcomes. Overall, in-person appointments result in greater diagnostic precision compared to teledermatology. Nonetheless, certain researches found the effectiveness of teledermatology in detecting skin cancer to be considerably high [12]. However, it is important to initially confirm that the clinicians exhibit a strong interrater reliability; without this, it is challenging to determine if the lack of agreement in diagnoses is due to the technology or inherent variations in clinical judgement. In this scenario, research has examined the consistency of diagnoses between GPs utilizing telemedicine and dermatologists. The findings from the research indicated a general diagnostic concordance of 65.52%, illustrating GPs' tendency to overdiagnose certain illnesses [13].

The teledermatology concordance achieved a rate of 94.7%. While this method proved effective in sorting quality, it exhibited poor precision in inflammatory issues [13]. Teledermatology can improve access by making referrals easier and providing convenience, shorter wait times, diagnostic help, and increased satisfaction for patients and providers alike [14-17]. In primary care, it is crucial to understand the actual requirements and create a user-friendly interface for AI adoption, in order to minimize resistance towards transitioning from traditional to touch-based interfaces in healthcare settings [18].

AI has been developed, researched, and utilized in various medical fields in recent times. Pictures are frequently used types of information for the advancement of AI, like electrocardiograms or radiologic images [19-21]. Dermatopathology is especially well suited for deep learning algorithms, as pattern recognition at scanning magnification is crucial for diagnosis [10,22-24]. Moreover, machine learning (ML) is more and more utilized in the field of dermatology, specifically in the area of skin cancer detection through image analysis with ML models incorporating deep convolutional neural networks (CNNs) [25,26]. It was not until 2012 when the ImageNet competition showcased the potential of algorithms and models with CNNs, even though they were introduced in the 1980s [23]. Afterwards, CNN has gained popularity as a machine learning method in various fields such as dermatology [27]. Additionally, there are machine learning research projects analyzing a broader range of skin conditions suitable for primary healthcare purposes [28]. ML saw a breakthrough in 2010 with the emergence of deep learning, which has transformed tasks like image classification, segmentation, and speech recognition.

Despite encountering numerous skin conditions, there have been minimal prospective studies conducted in primary care settings. Nonetheless, certain research has involved general practitioners alongside dermatologists as participants in the comparison group in order to evaluate the effectiveness of ML when compared to medical professionals [11,28,30], and has determined that AI technology could be applied in primary healthcare settings [28]. The study's primary goal is to validate a machine learning model as a diagnostic tool for dermatological conditions in a rural area of Catalonia, Spain, through prospective testing in primary care settings.

2. Method:

Study design:

This study aims to assess the performance of a machine learning model by comparing its diagnostic abilities with those of general practitioners and dermatologists. The Autoderm application programming interface (API; iDoc24 Inc) was combined with a secure, anonymous, and stand-alone web interface that works on all mobile devices.

To perform this study, the necessary steps were completed until the desired sample size was achieved: (1) a patient with a skin issue agreed to participate and signed the study agreement; (2) General Practitioners (GPs) identified the skin problem; (3) GPs captured a high-quality image of the skin issue; (4) GPs sent the photo for teledermatology consultation through the established process; (5) the image was uploaded to the Autoderm ML platform; and (6) dermatologists provided a diagnosis for the skin condition.

The ML tool's impact on the satisfaction of healthcare professionals was evaluated with 3 questions included in the survey. The questions pertain to the tool's potential in aiding diagnosis or prompting additional consideration beyond initial plans, as well as its potential to prevent the need for a dermatology referral.

Population

The research was carried out in primary care centers overseen by the Catalan Health Institute, the primary care services provider in central Catalonia, covering Anoia, Bages, Moianès, Berguedà, and Osona. Approximately 512,050 residents were part of the population reference in the study. Prospective

subjects were recruited in a consecutive manner.

Data collection:

Doctors gathered information from consecutive patients who fit the criteria after receiving signed informed consent. The data gathered were only presented in a case report format. The skin condition was diagnosed by the GP and a questionnaire was completed. The GP took a high-quality close-up photo of the skin issue for every patient using a smartphone camera. The image remains anonymous because the patients cannot be identified. Next, the general practitioner utilized the Autoderm ML platform to submit the de-identified image and completed the questionnaire with the ML model's top 5 generated diagnoses. The assessment of the Autoderm API tool serves as a validation study for aiding in the diagnosis of skin lesions in real primary care settings. Hence, despite the closed source code used by the tool, this research serves as an initial exploration to determine the potential use of similar tools in practical clinical settings. Autoderm is a dermatology search engine supported by research, marked with Conformité Européenne, and utilizing ML technology to offer quicker and more precise skin diagnoses. The existing ML model has the capability to detect 44 various types of skin diseases, such as inflammatory skin conditions, skin growths, and genital skin issues, and is available through an API. A user-friendly web interface was created for uploading images easily from either the smartphone library or the smartphone camera for this research. Using only a photo taken with a smartphone, the model creates a ranked list of the top 5 skin diseases based on probability. This ML model is projected to have a lifespan of approximately 3 months. Once this time has passed, the model will be enhanced for better precision.

In its present state, the ML model utilizes a 34-layer pretrained ResNet model from TorchVision (PyTorch) for tasks like computer vision and natural language processing. Moreover, the model underwent training through transfer learning on a private data set comprising 55,364 images for training and 13,841 images for testing. The model's average accuracy is 31.7% for the top 1 diagnosis and 68.1% for the top 5. Certain skin conditions have varying levels of accuracy due to the number of images used to train the machine learning model and the complexity of diagnosing certain diseases, especially when they occur in specific anatomical locations. Prior to implementation, the ML model underwent manual testing using a dataset sourced from different websites containing skin disease images captured using a mobile camera. The ML model was put into operation once it was considered sturdy. The 44 unique categories of skin diseases account for approximately 90% of the issues that the public worries about and seeks advice for.

Incorporating the anonymized image and an accurate description of the skin lesion into the patient's medical history, the GP sought a second opinion as part of the current teledermatology workflow. After receiving the information, the dermatologist proceeded to complete the "Assessment by teledermatology" questionnaire. It was anticipated that the response time would range from 2 to 7 days.

If a referral to a dermatologist is necessary, the GP completed the questionnaire for "Assessment by inperson dermatologist" by checking the electronic health records when they are accessible. The typical wait time for a dermatology referral varies between 30 and 90 days. The case number for the questionnaire was decided in advance prior to starting data collection, with each questionnaire having the same number, preventing patient identification.

Inclusion criteria:

Individuals over 18 years old were included in the prospective study if they visited a participating PCC for a cutaneous disease and gave written consent

Exclusion criteria

Patients were not included in the study if they had a cutaneous lesion that could not be captured with a smartphone or had conditions that could lead to poor protocol adherence. Pictures of low quality were also not included in the research.

Analysis of statistics

Determination of Sample Size

In order to evaluate the ML model's performance relative to that of GPs and dermatologists, a set of 100 images of skin diseases from eligible patients is needed for comparison. The recommended sample size is determined by the sample size calculation method utilized in comparable research studies [31-33].

Planned analysis

The validation dataset will contain roughly 100 instances, which will include an image and 3 or 4 evaluations: the in-person assessment by a general practitioner, the evaluation conducted through teledermatology, the top 5 potential diagnoses from the machine learning model ranked by probability, and the assessment by the in-person dermatologist (if a referral is present). The evaluation of the machine learning model will only cover 44 different skin diseases categories. A confusion matrix is necessary for determining the precision, recall (sensitivity), specificity, and accuracy of the machine learning model. The number of true positives, true negatives, false positives, and false negatives will be computed for every skin disease. In order to assess the ML multiclass classifier, the data will be viewed as separate binary problems, one for each skin disease class. The calculation of the Area under the curve and receiver operating characteristics curve for N number of skin diseases classes will be done using the one-versus-all methodology. Macro- and micro-averaging methods will be used to demonstrate the effectiveness of rare skin disease classes (adjusted for prevalence). Precision, recall, and F-measure will be computed separately for every skin disease category, and the outcomes will be merged to derive the mean precision and F-score. The ML model will also be evaluated for accuracy on the top 3 diagnoses.

Ethical approval:

The ethics committee of Institut Universitari d'Investigació en Atenció Primària Jordi Gol i Gurina approved the trial study protocol with code 20-159P. All patients involved in the study were asked to give written consent before participating.

3. Results:

Using a confusion matrix and one-versus-all approach, the results will be displayed for each skin disease class both globally and individually. The amount of time needed to make the diagnosis will be factored in as well. The evaluation of professionals’ satisfaction with the utilization of this ML tool will be conducted.

Patient enrollment started in June of 2021 and continued for a duration of 5 months. At present, all participants have been enlisted and the images have been presented to the GPs and dermatologists. The examination of the findings has commenced. We are optimistic that enough evidence will be gathered to confirm the validity of this image analysis machine learning model. We anticipate that the findings will be applied in medical settings to streamline and enhance the care of patients with skin conditions by improving the efficiency and safety of general practitioners. This research is an initial attempt at creating more extensive ML model validation studies.

Even if the ML model does not offer a superior diagnosis to the doctor's, it is anticipated to assist the practitioner in contemplating other potential diagnoses.

4. Discussion

This research aims to validate a machine learning model in the future as a tool to support diagnostic decision-making in identifying skin conditions. It would also evaluate the ML model's diagnostic accuracy and effectiveness in a PCC environment. In this situation, this research could offer benefits for patients and primary care doctors, enhancing the system's efficiency and effectiveness, while also offering insights into the effectiveness and constraints of ML models. External testing is crucial for overseeing these diagnostic systems and implementing ML models in actual PCC environments. The primary limitation of this study is the quantity of image samples used to assess the ML model's performance. Autoderm evaluates just 44 skin diseases, with many of them representing less than 1% to 5% of the images, leading to unbalanced sample data for each class. This can result in some skin conditions not being assessed properly, causing lower confidence levels and inconclusive results for those specific conditions.

Secondly, because of the small sample size and consecutive recruitment of cases, it is unlikely that we will achieve results that are representative of less common diseases. As there may be an imbalance in the class distribution of the 100 recruited patients, we will prioritize the F-Score in our analysis to prevent an overestimation of model quality based on accuracy, sensitivity, and specificity if 90% of the skin lesions are the most common. It must be noted that this research will be conducted in actual clinical settings, and patient selection will not be possible.

A diagnosis based on a single image with the best composition may have limitations compared to diagnoses made during a clinical assessment. Our ML algorithm's result was derived from just one photo, in contrast to other ML algorithms that utilize multiple photos or even the same algorithm accessible to the public that utilizes two images.

Additionally, our data will be limited as it will not cover further testing, and only a portion of potential malignancies will be confirmed through biopsy. Instead, we base our gold standard for each case on combining the differing diagnoses of a group of dermatologists. In clinical settings, the presence of uncertainties in diagnosing conditions like rashes, which are not commonly biopsied, presents a challenge in assessing the accuracy of clinicians and deep learning systems.

Additionally, our ML algorithm did not incorporate extra clinical metadata such as past medical history, symptoms, appearance, and texture, which could be a potential issue when comparing ML to physicians' diagnostic accuracy. Finally, the clinicians were asked to give only the top 3 diagnoses, even if they had additional possibilities.

5. Conclusion

This research will offer insights into the efficacy and constraints of ML models. External testing is necessary for controlling these diagnostic systems for the implementation of ML models in a primary care environment.

References

1. Sánchez-Sagrado T. Are there too many or too few physicians in Spain? migration: the eternal resource. Rev Clin Esp (English Ed) 2013 Oct;213(7):347-353. [CrossRef]
2. Lim HW, Collins SAB, Resneck JS, Bolognia JL, Hodge JA, Rohrer TA, et al. The burden of skin disease in the United States. J Am Acad Dermatol 2017 May;76(5):958-972.e2. [CrossRef] [Medline]
3. Schofield JK, Fleming D, Grindlay D, Williams H. Skin conditions are the commonest new reason people present to general practitioners in England and Wales. Br J Dermatol 2011 Nov;165(5):1044-1050. [CrossRef] [Medline]
4. Tensen E, van der Heijden JP, Jaspers MWM, Witkamp L. Two decades of teledermatology: current status and integration in national healthcare systems. Curr Dermatol Rep 2016 Mar 28;5:96-104 [FREE Full text] [CrossRef] [Medline]
5. Servei Català de la Salut. Activitat assistencial de la xarxa sanitària de Catalunya, any 2012: registre del conjunt mínim bàsic de dades (CMBD). Barcelona: Departament de Salut. 2013 Apr. URL: http://hdl.handle.net/11351/1025 [accessed 2022-08-11]
6. Lowell BA, Froelich CW, Federman DG, Kirsner RS. Dermatology in primary care: prevalence and patient disposition. J Am Acad Dermatol 2001 Aug;45(2):250-255. [CrossRef] [Medline]
7. Porta N, San Juan J, Grasa M, Simal E, Ara M, Querol I. Diagnostic agreement between primary care physicians and dermatologists in the health area of a referral hospital. Actas

Dermosifiliogr (English Ed) 2008 Apr;99(3):207-212 [FREE Full text] [CrossRef] [Medline]
8. López Seguí F, Franch Parella J, Gironès García X, Mendioroz Peña J, García Cuyàs F, Adroher Mas C, et al. A cost-minimization analysis of a medical record-based, store and forward and provider-to-provider telemedicine compared to usual care in Catalonia: more agile and efficient, especially for users. Int J Environ Res Public Health 2020 Mar 18;17(6):2008 [FREE Full text] [CrossRef] [Medline]
9. Potter B, Ronan SG. Computerized dermatopathologic diagnosis. J Am Acad Dermatol 1987 Jul;17(1):119-131. [CrossRef] [Medline]
10. Talebi-Liasi F, Markowitz O. Is artificial intelligence going to replace dermatologists? Cutis 2020 Jan;105(1):28-31. [Medline]
11. Börve A, Dahlén Gyllencreutz J, Terstappen K, Johansson Backman E, Aldenbratt A, Danielsson M, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190 [FREE Full text] [CrossRef] [Medline]
12. Finnane A, Dallest K, Janda M, Soyer HP. Teledermatology for the diagnosis and management of skin cancer: a systematic review. JAMA Dermatol 2017 Mar 01;153(3):319-327. [CrossRef] [Medline]
13. Ferrer RT, Bezares AP, Mañes AL, Mas AV, Gutiérrez IT, Lladó CN, et al. Diagnostic reliability of an asynchronous teledermatology consultation. Article in Spanish. Aten Primaria 2009 Oct;41(10):552-557 [FREE Full text] [CrossRef] [Medline]
14. Mounessa JS, Chapman S, Braunberger T, Qin R, Lipoff JB, Dellavalle RP, et al. A systematic review of satisfaction with teledermatology. J Telemed Telecare 2018 May;24(4):263-270. [CrossRef] [Medline]
15. Vidal-Alaball J, Álamo-Junquera D, López-Aguilá S, García-Altés A. Evaluation of the impact of teledermatology in decreasing the waiting list in the Bages region (2009-2012). Article in Spanish. Aten Primaria 2015 May;47(5):320-321 [FREE Full text] [CrossRef] [Medline]
16. Vidal-Alaball J, López Seguí F, Garcia Domingo JL, Flores Mateo G, Sauch Valmaña G, RuizComellas A, et al. Primary care professionals' acceptance of medical record-based, store and forward provider-to-provider telemedicine in Catalonia: results of a web-based survey. Int J Environ Res Public Health 2020 Jun 08;17(11):4092 [FREE Full text] [CrossRef] [Medline]
17. Lee MK, Rich K. Who is included in human perceptions of AI?: trust and perceived fairness around healthcare AI and cultural mistrust. 2021 May Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-14. [CrossRef]
18. Calisto FM, Ferreira A, Nascimento JC, Gonçalves D. Towards touch-based medical image diagnosis annotation. 2017 Oct 17 Presented at: ISS '17: Interactive Surfaces and Spaces; October 17-20, 2017; Brighton, United Kingdom p. 390-395. [CrossRef]
19. Calisto FM, Santiago C, Nunes N, Nascimento JC. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. Int J Hum Comput Stud 2021 Jun;150:102607. [CrossRef]
20. Calisto FM, Nunes N, Nascimento JC. BreastScreening: on the use of multi-modality in medical imaging diagnosis. 2020 Sep Presented at: AVI '20: International Conference on Advanced Visual Interfaces; September 28 to October 2, 2020; Salerno, Italy p. 1-5. [CrossRef]
21. Attia ZI, Harmon DM, Behr ER, Friedman PA. Application of artificial intelligence to the electrocardiogram. Eur Heart J 2021 Dec 07;42(46):4717-4730 [FREE Full text] [CrossRef] [Medline]

22. Wells A, Patel S, Lee JB, Motaparthi K. Artificial intelligence in dermatopathology: diagnosis, education, and research. J Cutan Pathol 2021 Aug 26;48(8):1061-1068. [CrossRef] [Medline]
23. Thomsen K, Iversen L, Titlestad TL, Winther O. Systematic review of machine learning for diagnosis and prognosis in dermatology. J Dermatolog Treat 2020 Aug;31(5):496-510. [CrossRef] [Medline]
24. Du AX, Emam S, Gniadecki R. Review of machine learning in predicting dermatological outcomes. Front Med (Lausanne) 2020 Jun 12;7:266 [FREE Full text] [CrossRef] [Medline]
25. Gomolin A, Netchiporouk E, Gniadecki R, Litvinov IV. Artificial intelligence applications in dermatology: where do we stand? Front Med (Lausanne) 2020 Mar 31;7:100 [FREE Full text] [CrossRef] [Medline]
26. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Feb 02;542(7639):115-118 [FREE Full text] [CrossRef] [Medline]
27. Morid MA, Borjali A, Del Fiol G. A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 2021 Jan;128:104115. [CrossRef] [Medline]
28. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. Nat Med 2020 Jun 18;26(6):900-908. [CrossRef] [Medline]
29. Servei Català de la Salut. Activitat assistencial de la xarxa sanitària de Catalunya, any 2012. Departament de Salut. URL: http://hdl.handle.net/11351/1025 [accessed 2022-08-11]
30. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. Lancet Oncol 2019 Jul;20(7):938-947 [FREE Full text] [CrossRef] [Medline]
31. Kamulegeya LH, Okello M, Bwanika JM, Musinguzi D, Lubega W, Rusoke D, et al. Using artificial intelligence on dermatology conditions in Uganda: a case for diversity in training data sets for machine learning. bioRxiv. Preprint posted online October 31, 2019 [FREE Full text] [CrossRef]
32. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer 2019 Sep;119:11-17 [FREE Full text] [CrossRef] [Medline]
33. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader study level-I and level-II Groups, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842 [FREE Full text] [CrossRef] [Medline]