British Association of Dermatologists Position Statement on Artificial Intelligence (AI) Interventions

Artificial intelligence (AI) collectively describes computer algorithms that perform tasks which normally require human intelligence. There is immense potential to improve the diagnosis and precision management of skin diseases through AI Interventions (any health intervention which relies upon AI to serve its purpose). The BAD welcomes adoption of appropriately regulated and governed uses of AI interventions to enhance safe clinical practice and improve patient outcomes. We believe that AI has the potential to improve clinical care, optimise processes and allow greater use of clinical data to inform best practice and outcomes. AI technologies should be developed in areas which clearly address an unmet clinical need; by improving disease management and quality of care and by enhancing patient experience without compromising safety.

The BAD aims to support dermatology departments to ensure that safe, ethical, and effective AI interventions are adopted through a robust regulatory framework.

The evidence base for AI in dermatology

AI is a rapidly advancing field and the COVID-19 pandemic has accelerated the commercial drive to adopt digital technologies and integrate AI algorithms into clinical practice. However, currently the evidence-base for effectiveness of AI interventions in dermatology is limited [1,2,3,4] due to studies being undertaken in artificial conditions which do not adequately reflect real-life clinical settings. Several AI interventions in dermatology focus on differentiating between benign and malignant skin lesions with a particular emphasis on melanoma diagnosis[5]. Other potential applications exist, including monitoring of inflammatory skin disorders e.g., psoriasis, atopic dermatitis, acne vulgaris, leg ulcer assessment, and nail disease[5] but are earlier in their phase of development. The accuracy of AI algorithms intended to support skin cancer diagnoses may be overestimated where studies are conducted in settings which do not reflect clinical practice. For example, a study using a retrospective image database without supplementary clinical information, excluding atypical presentations (e.g., body sites such as palms/soles), using highly selected patient groups (e.g., excluding skin of colour)[6], or limiting study cases to those already selected for excision introduces significant bias, risks missing serious but rare diagnoses which can lead to patient harm and is unlikely to provide strong evidence for widespread use[2].

Furthermore, some producers of skin smartphone apps claim that they can classify (diagnose), monitor and treat a range of skin disorders. Many of these are flawed in their design due to the lack of availability of clinical images representing the full breadth and depth of skin disorders on which these algorithms are developed. This leads to many apps focusing on diagnosis of a limited number of skin cancers and neglecting potentially more serious and rare diseases. Additionally, it is not always

clear that these apps have been appropriately investigated in the clinical study setting in which the app will be deployed and a systematic review of algorithm-based smartphone apps used for skin cancer diagnosis identified serious flaws[8] which can lead to overestimating accuracy[7]. Moreover, evaluation of diagnostics is not confined to accuracy and the full scope of harms and benefits from implementing a new technology must also be assessed[9]. The cost-effectiveness of AI apps (like all medical devices) should be considered before adoption in the NHS. This generally requires a cost-effectiveness or cost-minimisation analysis to have been conducted. This analysis should compare the AI intervention to the current standard of care and measure any additional costs and benefits associated with the AI intervention. The NHS runs on a constrained budget and any new spending on medical devices needs to be considered good value for money before adoption. We encourage product developers to use NICE Scientific Advice services to help determine the type of evidence required to demonstrate the value of their product to the NHS. The NICE Scientific Advice programme operates on a cost-recovery basis and so there are fees associated with these services.

We believe it is essential for AI app developers to develop algorithms within a tightly integrated ecosystem including computer scientists, clinicians and patient organisations and to be aware of the relevant regulatory[10,11] and ethics[12] governance frameworks for the usage of their products. These principal areas are summarised under the following headings:

Regulation of medical devices apply to AI interventions

The Medicines & Healthcare products Regulatory Agency (MHRA) explains that the following types of software may be classed as medical devices[13]:

- software (including artificial intelligence) and apps (either incorporated into an existing device or supplied separately) that are used for contributing to diagnostic processes
- software and apps for helping patients to manage their health conditions
- software and apps for monitoring patients (including remotely)
- software and apps to support clinical decision making

From 1st Jan 2020 all medical devices are either CE (Conformité Européene), UKCA (UK Conformity Assessed) or UKNI (UK Northern Ireland) marked to be placed on the UK market[14]. From July 2023, Great Britain will transition to UKCA marking only, but devices on the NI market can continue to be CE or UKNI marked[15]. All devices require a clinical evaluation, this is a critical evaluation of the relevant scientific literature of equivalent devices and/or of all the clinical investigations of the product. A clinical investigation is required to verify that, under normal conditions of use, in accordance with the manufacturer's instructions for use, the product performs in the way the manufacturer intends. The clinical evaluation report should be appropriate to the device under evaluation, its specific properties, and its intended purpose. The evidence presented should adequately support its intended use, for example, the study should be undertaken in the same clinical setting (primary versus secondary care) and with the same patient population in which it is intended for deployment.

Whilst there is no requirement that manufacturers provide a copy of the clinical evaluation, we recommend that commissioners request a copy before purchase of a device. The BAD AI WPG can be contacted to assist in critical evaluation of the evidence presented.

All devices must be accompanied by the information needed to use them safely and properly, taking account of the training and knowledge of potential users. This should include (but is not limited to) a clear description of the tests and data used for validation and a clear statement of limitations emerging from validation studies. It is important to distinguish between use within the scope of an existing CE/UKCA legislation marking (1 January 2021)[16] and use in order to evaluate the product for

a purpose that is not within that scope. This requires careful consideration of the manufacturer's statements about the intended use of the product, and careful scrutiny of the instructions for use. Manufacturer's statements on labelling and instructions for use should be consistent with information provided in any promotional material. This can include adverts, information provided in the app stores, information on the product's landing page or in the manufacturer's social media channels. There is a risk of liability for example, for negligence or breach of product safety legislation, if a product is used outside the scope of its CE/UKCA marking.

Medical device stand-alone software including apps (including in vitro diagnostic medical devices [IVDMDs]) v1.07 specifically states (Page 26 Rule 10): any device intended to allow for direct diagnosis should be classified as Class IIA. This applies even if the word 'diagnosis' is not used but words to that effect or claims that give that impression or have a demonstrable function is sufficient. 'Indicative diagnosis' in the context of lay usage can be sufficient that the device should be Class IIA classification. 'Allow for direct diagnosis' also applies to devices that 'provide decisive information for making a diagnosis. It is incumbent upon clinicians to ensuring that AI interventions used for skin disease diagnosis or triage are appropriately classified and have the necessary supporting evidence base prior to adoption.

Guidance for Reporting Design Changes to AI

According to Article 5 MDR manufacturers are not allowed to make 'significant' or 'substantial' changes in design or significant changes to the intended purpose of their AI product without reporting this to the MHRA before implementation.

Changes to the design specifications may be substantial if they affect the indications for use, the performance of the device or if they raise new safety and performance issues. Many changes to device software will require an evaluation and acceptance by the MHRA.

Where the AI app algorithm is being used in operation in a service, as a service evaluation, but has changed sensitivity and specificity, we would advise to re-examine whether those changes mean that, in effect, the nature of the activity has now become research.

A robust regulatory framework is needed

The NHS is currently under pressure to deploy innovative technologies, and there is an understandable impetus towards adoption of experimental approaches. However, these must be appropriately verified and clearly address an unmet need in the local population. Where the evidence for safety and efficacy is lacking or the use of the product falls outside the scope of intended use for which it was CE/UKCA marked, the Health Research Authority (HRA) system for the approval of research can enable AI innovations to be trialled with appropriate safeguards. The HRA can help both developers and clinicians to recognise, understand and comply with the legal and ethics governance processes that apply to AI/data-driven innovations.

It is the responsibility of all clinical users to independently evaluate the regulatory case and intended use for current AI skin cancer diagnostic products which are being offered to improve their local /regional patient pathways. There are indications that current pre-market regulatory requirements are not always robust enough and clinicians/managers/ commissioners may not be aware of the right assessment questions to ask and/or the apparent risk if they adopt insufficiently validated AI tools. The BAD can provide guidance in this regard and can be contacted to assist in these local evaluations (please see accompanying flow chart for further guidance).

Conclusion

- We are concerned that there may be products on the market which make unsubstantiated or misleading claims about the power of AI in its use for skin cancer triage and implied diagnostics.
- The BAD aim to support our patients and professionals in engaging with AI in its current study phase for dermatology by ensuring that new AI technologies are developed in an ethically acceptable way that promotes engagement, involvement, and transparency.
- We are working closely with the MHRA, NHSX, the Department of Health and Social Care, NICE, and NHS national bodies to ensure that AI is safely deployed for its intended populations to provide maximum clinical benefit for both patients and clinicians.
- This includes highlighting the risk to local service providers where AI applications are being commissioned outside a study setting or their "approved use" for a specific skin disease population.
- We are also working with national stakeholders to develop a centralised UK clinical image library for skin disease. This will provide a large-scale image database with linked clinical data for AI algorithm development and validation.

All enquiries for AI applications for use in dermatology should be sent to our dedicated team using the email address: ai.dermatology@bad.org.uk

References

1. Liu X FL, Kale AU, et al. A comparison of deep learning performance against healthcare professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. The Lancet Digital Health 2019;1(6): E271-E97.
2. Ferrante di Ruffano L, Takwoingi Y, Dinnes J, et al. Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults. Cochrane Database Syst Rev 2018;12:CD013186. doi: 10.1002/14651858.CD013186 [published Online First: 2018/12/07]
3. Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. BMJ 2020;368:m689. doi: https://doi.org/10.1136/bmj.m689
4. Artificial Intelligence in Dermatology: A Primer.Young AT, Xiong M, Pfau J, Keiser MJ, Wei ML.J Invest Dermatol. 2020 Aug;140(8):1504-1512. DOI: 10.1016/j.jid.2020.02.026
5. Du-Harpur X, Watt FM, Luscombe NM, et al. What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol 2020 doi: 10.1111/bjd.18880 [published Online First: 2020/01/22]
6. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. J Invest Dermatol. 2018 Jul;138(7):1529-1538. doi: 10.1016/j.jid.2018.01.028. Epub 2018 Feb 8. PMID: 29428356.
7. Deeks JJ, Dinnes J, Williams HC. Sensitivity and specificity of SkinVision are likely to have been overestimated. J Eur Acad Dermatol Venereol. 2020;34(10): e582-e3.
8. Freeman K, Dinnes J, Chuchu N, et al. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. BMJ 2020;368:m127. doi: 10.1136/bmj.m127 [published Online First: 2020/02/12]
9. NICE (2019) Evidence standards framework for digital health technologies https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standardsframework-for-digital-health-technologies
10. NHSX A Buyers guide to AI in Health and Care https://www.nhsx.nhs.uk/ailab/explore-all-resources/adopt-ai/a-buyers-guide-to-ai-in-health-and-care/
11. MEDDEV 2.1/6 July 2016 - Qualification and Classification of stand-alone software
12. Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529
13. MHRA: Regulating medical devices in the UK: https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk
14. MHRA: Regulation of medical devices in Northern Ireland: https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk#NI
15. MHRA Guidance: Medical device stand-alone software including apps (including IVDMDs) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmentdata/file/957090/Software_flow_chart_Ed_1-07b-UKCA__002__FINAL.pdf
16. Guidance CE Marking updated 31 December 2021 https://www.gov.uk/guidance/ce-marking#history