

'Big-Data' in dermatological research

Feroze Kaliyadan, Kingshuk Chatterjee¹

Department of Dermatology, Sree Narayana Institute of Medical Sciences, Kochi, Kerala, ¹Department of Dermatology, Nil Ratan Sircar (NRS) Medical College Kolkata, Burdwan, India

Introduction

Big data in medicine involves the collection and analysis of large amounts of complex heterogeneous data. 'Big data' is a general term used for any collection of datasets whose size and complexity exceed the capabilities of traditional data processing applications. There is no specific definition for the volume which can be classified under 'big data', but generally anything ≥ 1 petabyte can be considered as big data.¹ The actual data involved includes a variety of 'omics' like genomics and epigenomics as well as biomedical data and electronic health records (EHR) data. Integrating the various omics with clinical data from EHRs would provide a rich database of information. This concept of integrating different 'omics' is referred to as 'multiomics'. The key issue here is the heterogeneity of the data at all levels. For example, clinical data in the electronic health record itself is stored in different forms at different centres. The conventional view of patient data is in the form of longitudinal data retrieved through patient records of which relevant structured data is used primarily for understanding or hypothesis testing. In big data, all types of data are collected and analysed, with the focus being on finding patterns and predictions. The initial data is unstructured.^{1,2}

Big Data

Big data is described in terms of six variables – volume, velocity, variety, veracity, variability, and value, of which the most important are volume (amount of data), velocity (the speed at which data are generated and processed) and variety (types of data).³

The advantages of big data in research include the relatively lower cost of collection of data, and the ability to use the data both prospectively and retrospectively. The data can also be used for both hypothesis generation as well as hypothesis testing, although compared to traditional data it is primarily

more useful for hypothesis generation. Concerns are the data storage, quality of data/data cleaning, privacy-related issues, and local health policy guidelines on the use of the data^{1,2}

As mentioned, the unstructured and heterogeneous nature of the data and the volume of the data lead to challenges in analysis. There are different techniques associated with the analysis of big data.

- Data mining – analysis of data to identify unsuspected and unexpected patterns
- Cluster analysis – focusing on grouping observations based on factors like demographics
- Machine learning (ML) and its types – supervised learning, unsupervised learning, semisupervised learning, and reinforcement learning

Machine Learning

ML is a component of the broader umbrella of artificial intelligence (AI) and it has become an integral and valuable part of the analysis of big data. ML uses training datasets to develop algorithms that can be used as predictive models. Deep learning is one such subset of ML, which has found application in dermatology image analysis and involves the use of 'neural networks' to develop prediction models. ML applied to big data analysis includes sub-types like decision tree learning, Bayesian networks, cognitive computing, and natural language processing. The primary use of AI in dermatology has been in the diagnosis and classification of skin diseases, especially in the context of skin cancer (mainly in melanoma). The use of artificial neural networks – deep and convoluted neural networks, in image analysis, has evolved significantly in the last few years. However, the applications of AI can go much beyond this – including the realm of dermatological research including research related to predicting disease outcomes. An important aspect of training

How to cite this article: Kaliyadan F, Chatterjee F. 'Big-Data' in dermatological research. *Indian J Dermatol Venereol Leprol.* 2024;90:342–4. doi: 10.25259/IJDVL_1298_2023

Corresponding author: Dr. Feroze Kaliyadan, Sree Narayana Institute of Medical Sciences, Kochi, India. ferozkal@gmail.com

Received: December, 2023 **Accepted:** January, 2024 **Epub Ahead of Print:** March, 2024 **Published:** April, 2024

DOI: 10.25259/IJDVL_1298_2023 **PMID:** 38594987

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

datasets is the size of the dataset. Larger the dataset, the better the expected performance, and in this context again the availability of big data is significant in improving the quality of predictive models. The other challenge is to ensure that these data sets cover natural human heterogeneity in terms of gender, skin colour, and race.

- One of the largest databases contributing to the study of AI in dermatology is the 'The International Skin Imaging Collaboration' (<https://www.isic-archive.com/>). The collaboration primarily aims to reduce the mortality associated with melanoma by developing diagnostic AI-based algorithms, which can help in early recognition. The collaborations host a large open-source archive of dermatological images which can be used for teaching, research, and development/testing of AI-based algorithms.¹⁻⁶

Important areas where big data analytics have been used in dermatology include – risk prediction models for skin cancers (mainly melanoma), developing clinical decision support systems, genome-wide association studies related to dermatological conditions, and tailored/precision medicine (identification of specific features in heterogeneous conditions which might predict treatment response).

Big data can also help improve the efficiency of clinical trials in dermatology by better identification of ideal candidates for clinical trials. Like in other specialties, it can also improve post-marketing surveillance for drug safety

AI and ML could go beyond big data in clinical research. Some of the areas in which AI could play a more significant role in the future could include

- Designing studies/protocols
- Sorting of data, statistical analysis
- Collection and sorting of references
- Even images can be generated using AI, for illustrating concepts/graphics

Recently platforms like Chat generative pre-training transformer (ChatGPT) have shown abilities of error detection in software codes and also providing possible solutions to the errors. It is possible that the same might be used to improve the quality of research protocols and also resultant research data. Platforms like ChatGPT could play a role in data analysis, data summarization, and report writing in the future.

Mobile health (m-health) data is a valuable source of big data. Health and fitness apps can record significant amounts of personal health data for analysis. The often-unstructured nature and patient-dependent data entry and factors might affect the reliability of m-health data. Social media can also serve as an input for big data in the context of medical research especially, in the context of qualitative research dealing with areas like knowledge, attitudes, practice, and quality of life.

Evidence-based medicine and big data can function in synergy, especially in some niche areas like genomics. The quality of evidence can be improved by the sheer size and computational power of big data and equally importantly big data by itself does not have much meaning when not looked at from the lens of evidence-based medicine.⁷

Limitations, Problems and Challenges

AI-based language algorithms can actually make the process of research writing simpler, but in turn, the ethical aspects and limits of using these for research are debatable and need to be discussed to avoid the possibility of research misconduct/plagiarism-related issues.

As of now we still need counter-checks to ensure the veracity of AI in dermatology – whether it be in the context of image recognition or for more advanced uses. For diagnostic purposes, we need more extensive datasets which cover a larger variety of skin conditions as well as skin types (the skewed representation of 'skin of colour' has always been a concern). Privacy and security concerns for clinical image data need to be addressed for the training datasets.

Patients already have access to some of the AI-based diagnostic software/apps. These are also accessible to non-dermatologists/alternative medicine practitioners, which opens up the possibility of wanton use of the same learning to possible misdiagnosis and wrong treatment

Finally, of course, the key question is if the developments in AI can lead to the dermatologist being less important or worse still replaceable. The answer at this point is of course a resounding no. However, the influence of AI in clinical medicine will only keep growing. It is pertinent to mention the terminology 'augmented intelligence' rather than 'artificial intelligence' to demonstrate the complementary role of AI in decision-making.⁸ What we as dermatologists need to do is to get involved in the process as important stakeholders who can use AI as a strength rather than consider it as an enemy. Our involvement is also necessary to ensure that any misuse of the technology can be prevented through patient education, advocacy, and legislation. The key is for dermatologists to get involved in the process.

To conclude, AI and big data are complementary and both are here to stay. As dermatologists, we must try to understand how we can use these newer developments to augment our decision-making skills, and research capabilities and generally improve patient outcomes. To do so we first have a basic understanding of these concepts and try to get more involved in the development and fine-tuning of the same.

Declaration of patient consent

Patient's consent not required as there are no patients in this study.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

Use of artificial intelligence (AI)-assisted technology for manuscript preparation

The authors confirm that there was no use of artificial intelligence (AI)-assisted technology for assisting in the writing or editing of the manuscript and no images were manipulated using AI.

References

1. Wehner MR, Levandoski KA, Kulldorff M, Asgari MM. Research techniques made simple: An introduction to use and analysis of big data in dermatology. *J Invest Dermatol* 2017;137:e153–8.
2. Risteovski B, Chen M. Big data analytics in medicine and healthcare. *J Integr Bioinform* 2018;15:20170030
3. Laney D. 3D Data management: Controlling data volume, variety and velocity. *Application Delivery Strategies* 2001;6:949.
4. Murphree DH, Puri P, Shamim H, Bezalel SA, Drage LA, Wang M, *et al.* Deep learning for dermatologists: Part I. Fundamental concepts. *J Am Acad Dermatol* 2022;87:1343–51.
5. Puri P, Comfere N, Drage LA, Shamim H, Bezalel SA, Pittelkow MR, *et al.* Deep learning for dermatologists: Part II. Current applications. *J Am Acad Dermatol* 2022;87:1352–60.
6. Liopyris K, Gregoriou S, Dias J, Stratigos AJ. Artificial intelligence in dermatology: Challenges and perspectives. *Dermatol Ther (Heidelb)* 2022;12:2637–51.
7. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: Challenges and opportunities. *Genome Med* 2016;8:134.
8. American Academy of Dermatology. Position statement on augmented intelligence. 2019. Available at: <https://server.aad.org/forms/Policies/Uploads/PS/PS-Augmented%20Intelligence.pdf> (Accessed January 25, 2023).