Review of Machine Learning in Predicting Dermatological Outcomes

Amy X. Du[1]*, Sepideh Emam[2†] and Robert Gniadecki[1†]

[1] *Division of Dermatology, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada,* [2] *Information Services and Technology, University of Alberta, Edmonton, AB, Canada*

Artificial intelligence is a broad branch of computer science that has garnered significant interest in the field of medicine because of its problem solving, decision making and pattern recognition abilities. Machine learning, a subset of artificial intelligence, hones in on the ability of computers to receive data and learn for themselves, manipulating algorithms as they organize the information they are processing. Dermatology is at a particular advantage in the implementation of machine learning due to the availability of large clinical image databases that can be used for machine training and interpretation. While numerous studies have implemented machine learning in the diagnostic aspect of dermatology, less research has been conducted on the use of machine learning in predicting long-term outcomes in skin disease, with only a few studies published to date. Such an approach would assist physicians in selecting the best treatment methods, save patients' time, reduce treatment costs and improve the quality of treatment overall by reducing the amount of trial-and-error in the treatment process. In this review, we aim to provide a brief and relevant introduction to basic artificial intelligence processes, and to consolidate and examine the published literature on the use of machine learning in predicting clinical outcomes in dermatology.

Keywords: artificial intelligence, machine learning, dermatology, prediction, clinical outcomes

INTRODUCTION

Artificial intelligence (AI) is a broad branch of computer science that has garnered significant interest in the field of medicine because of its problem solving, decision making, and pattern recognition abilities. Machine learning (ML), a subset of AI, hones in on the ability of computers to receive data and learn for themselves, manipulating algorithms as they organize the information they are processing. Dermatology is at a particular advantage in the implementation of ML due to the availability of large clinical image databases that can be used for machine training and interpretation. In fact, studies have already demonstrated the successful use of ML in classification and diagnosis of skin diseases, such as skin cancer [1, 2], eczema [3], psoriasis [4], onychomycosis [5] at a performance level equal or superior to board-certified dermatologists.

While numerous studies have implemented ML in the diagnostic aspect of dermatology [6], less research has been conducted on the use of ML in predicting long-term outcomes in skin disease, with only a few studies published to date. In an era of personalized medicine, there is a push toward a data-driven approach allowing for accurate prediction of long-term clinical outcomes for individual patients [7–9]. Such an approach would assist physicians in selecting the best treatment methods, save patients' time, reduce treatment costs and improve the quality of treatment overall by reducing the amount of trial-and-error in the treatment process [8].

Machine learning techniques are very good at managing large amounts of high-level data from patient databases, such as electronic medical records, and are often able to detect sophisticated data patterns that traditional statistical methods are unable to delineate (7–9). These approaches have been used with increasing success to predict patient prognoses in many other areas of medicine, such as the risk of readmission after hospital discharge (10), cancer progression (11), diabetic complications (12, 13), cardiovascular mortality (14), and many others (15–17).

In this review, we aim to provide a brief and relevant introduction to basic AI processes, and to consolidate and examine the published literature on the use of ML in predicting clinical outcomes in dermatology.

BRIEF OVERVIEW OF PRINCIPLES IN ARTIFICIAL INTELLIGENCE

Artificial intelligence can be subdivided in a number of ways, but in its simplest form, it can be broken into two main categories: strong AI and weak AI (**Figure 1**).

Strong AI refers to a programmed machine that takes on human-level cognition, with the capacity for consciousness, self-awareness and ethical decision-making (18). The machine has the competence to learn, on its own, to simultaneously conduct a number of complex tasks, and the capability of learning more based on what it already knows (19). Currently, strong AI does not exist outside the realm of science fiction.

Weak AI, on the other hand, does currently exist and is the process by which we train a machine to complete a specific, designated task. The machine simply acts upon and is bound by the rules and algorithms that are set for it. It does not have the capacity, unlike strong AI, to think and act beyond those parameters (19).
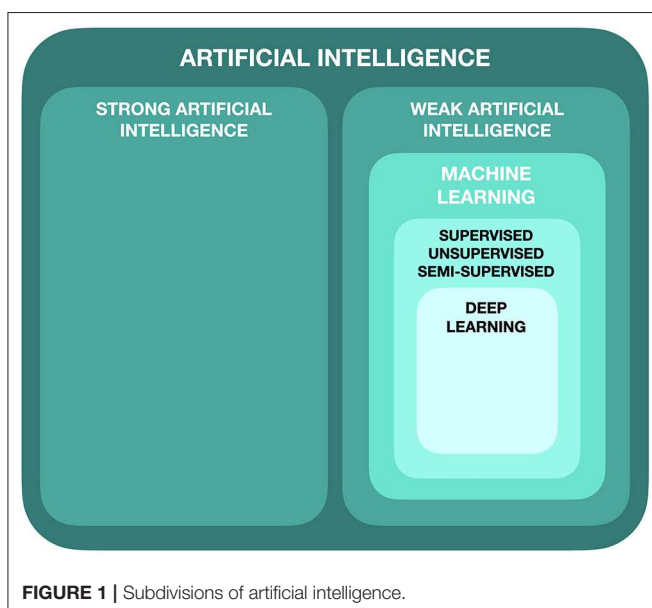

FIGURE 1 | Subdivisions of artificial intelligence.

Machine Learning

Machine learning is a subdivision of AI in which algorithmic models are trained to perform specific tasks by recognizing and learning patterns from the data it sees, rather than through explicit computer programming by a human expert. This process can be categorized as supervised, semi-supervised or unsupervised, with the most common method being supervised (20).

Supervised learning occurs when the algorithm system gains experience through training with a labeled dataset, and is then expected to categorize a new, unfamiliar data point. For example, in the case of recognizing benign vs. malignant skin lesions, the computer system would be provided with many images of skin lesions that have already been labeled as either being benign or malignant. Once training with these images is completed, the algorithm would then be tested by being presented with novel, unlabeled images to classify as either being benign or malignant (21).

When no training dataset is available for the corresponding output data, it is known as unsupervised learning. Very much like supervised learning, the goal of this type of learning is to place input data into categories. The main difference is that, in unsupervised learning, the input data are not labeled, and therefore, the model aims to categorize data based on their inherent features. This method of machine learning allows us to take a more open-ended approach to learn about the underlying distribution of data that may have been missed otherwise (22).

Finally, a hybrid method of machine learning, known as semi-supervised machine learning, combines aspects of supervised and unsupervised learning. In this approach, a large amount of unlabeled inputs are combined with a small amount of labeled inputs in an effort to lessen the challenge of data labeling (22).

Deep Learning

Deep learning is a further subdivision of ML and refers to a specific type of learning that involves the use of artificial neural networks (ANN) (23). It is often used for unsupervised learning, as it is capable of learning from data that is unstructured and unlabeled. It is able to detect patterns in datasets that it has not been previously trained on. Deep learning functions by imitating the neural connections made in the human brain and are connected in a network of nodes, forming multiple layers.

Performance Analysis of Machine Learning Algorithms

In order to statistically evaluate the performance of learning approaches, and to best determine which approach predicts with the highest accuracy, machine learning algorithms are often assessed using the area under the curve receiver operating characteristic (AUC-ROC). This test quantifies how accurately a model is able to distinguish between categories, typically in medicine, "disease" vs. "no disease." The ROC curve is plotted with the true positive rate on the y-axis against the false positive rate on the x-axis. The closer the AUC is to 1.0 for any given model, the better and more accurate the performance of that model (22).

METHODS

A literature search on Ovid MEDLINE® was conducted in January 2020 for papers published from 2000-2019, to focus on recently published literature. The database was searched with relevant keywords in combination with the Boolean operators "AND" and "OR." The search included keywords from each of the following lists: dermatology, skin disease, skin cancer, psoriasis, or atopic dermatitis, AND artificial intelligence, machine learning, deep learning, or neural network, AND prediction, predicting, or outcome.

Inclusion criteria included: English language, original studies, and focusing on the prognostic utility of artificial intelligence/machine learning in dermatology (i.e., predicting outcomes, risk stratification, selection of best treatment). Exclusion criteria included: reviews, animal studies, case reports, systematic reviews, studies not published in English. Any studies focusing on the diagnostic utility of artificial intelligence/machine learning in dermatology, focusing on a different medical field, or not using machine learning methods were also excluded.

Our literature search yielded a total of 73 articles, among which 6 were deemed relevant to this review based on our inclusion and exclusion criteria.

APPLICATIONS OF MACHINE LEARNING IN PREDICTING DERMATOLOGICAL OUTCOMES

A total of six studies on the use of machine learning in predicting dermatological outcomes have been published to date (**Table 1**). One study focused on the risk of biologic discontinuation in psoriasis patients (24), two studies investigated the risk of developing non-melanoma skin cancer (25, 26), one study looked at response to wart treatment modalities (27), one study explored the complexity of reconstructive surgery after periocular basal cell carcinoma excision (28), and one final study examined the risk of developing chronic venous ulcers in patients with cardiovascular disease (29).

Five of the six studies used a supervised approach of machine learning in their training and validation. Wang et al. (25) used a semi-supervised approach. Generally, the results of each study were presented with varying outcomes, but AUC was reported as the primary outcome in five of the six studies. Other outcomes reported included sensitivity and specificity (25, 26), accuracy (27), and positive and negative predictive values (28). Franciscis et al. reported outcomes in the form of "level-of-risk" (29). We are unable to directly compare the outcomes of all studies as the methodology of the studies vary.

TABLE 1 | Summary of literature on the use of machine learning in predicting dermatological outcomes.

Study	Type of supervision	End point	Results	Software(s) utilized	Data used
Emam et al. (24)	Supervised	Risk of discontinuation of biologic	AUC for predicted risk of discontinuation due to: Any reason 0.95 Lack of efficacy 0.91 Adverse event 0.88 Other reasons 0.80	Generalized linear model, support vector machine, decision tree, random forest, gradient boosted trees, deep learning	n = 681 psoriasis patients 13 clinically relevant features per patient
Wang et al. (25)	Semi-supervised	Risk of developing non-melanoma skin cancer	AUC 0.89 Sensitivity 83.1% Specificity 82.3%	Convolutional neural network (deep learning)	n = 9,494, 1,829 non-melanoma skin cancer patients, 7,665 random non-cancer controls 20 clinically relevant features per patient
Roffman et al. (26)	Supervised	Risk of developing non-melanoma skin cancer	AUC 0.81 Sensitivity 86.2% Specificity 62.7%	Artificial neural network (deep learning)	n = 462,630, 2,056 non-melanoma skin cancer patients, 460,574 non-cancer patients 13 clinically relevant features per patient
Khozeimeh et al. (27)	Supervised	Response to wart treatment method	Cryotherapy: AUC 0.902, accuracy 80% Immunotherapy: AUC 0.813, accuracy 98%	Fuzzy logic and adaptive network-based fuzzy inference system (ANFIS)	n = 180, 90 patients in cryotherapy group, 90 patients in immunotherapy group 7 clinically relevant features per patient in the cryotherapy group, 8 in the immunotherapy group
Tan et al. (28)	Supervised	Complexity of reconstructive surgery after periocular basal cell carcinoma excision	Naïve Bayesian Classifier: AUC 0.854 PPV 38.1% NPV 94.1% ADTree: AUC 0.835 PPV 31% NPV 97%	Decision table, Bayesian, tree-based methods, multivariate logistic regression, nearest neighbor classifier, support vector machine	n = 156 periocular BCC patients 7 clinically relevant features per patient
de Franciscis et al. (29)	Supervised	Risk of developing chronic venous ulcers in patients with chronic venous disease	CVU group level of risk 32.38 ± 7.19% Non-CVU group level of risk 8.34 ± 3.38%	Fuzzy logic to stratify CVD patients into CVU and non-CVU groups	n = 77, 40 patients with CVU, 37 patients without CVU 27 clinically relevant features

Emam et al. considered seven different modeling techniques in evaluating a dataset of 681 psoriasis patients to determine which learner performed best in terms of accuracy, interpretability and runtime to predict risk of biologic discontinuation. Thirteen clinically relevant features per patient were analyzed. The Generalized Linear Model (GLM) outperformed the six other models that were tested. The AUC for predicted risk of discontinuation due to any reason was found to be 0.95, lack of efficacy was 0.91, adverse event was 0.88, and other reasons was 0.80 using the GLM (24).

Wang et al. and Roffman et al. used convolutional neural networks (CNN) (25) and artificial neural networks (ANN) (26), respectively, to delineate the risk of developing non-melanoma skin cancer. Both approaches are branches of deep learning and make use of the algorithm's ability to extract important classifying information at each node of a network of data. Both studies included data from non-melanoma skin cancer patients as well as an abundance of data from non-cancer patients. The system by Wang et al. analyzed data from a total of 9,494 patients, using 20 clinically relevant features per patient, and reported higher outcomes (AUC 0.89, sensitivity 83.1%, specificity 82.3%) than Roffman et al., which analyzed data from a total of 462,630 patients, using 13 clinically relevant features per patient (AUC 0.81, sensitivity 86.2%, specificity 62.7%).

Two studies used fuzzy rule-based systems to stratify patients into groups. Fuzzy logic is a flexible mathematical system that can model non-linear functions with arbitrary meaning. It is a system that very closely models human thinking and is able to handle a great degree of uncertainty (30). Khozeimeh et al. (27) aimed to predict patient responses to two wart treatment modalities: cryotherapy and immunotherapy. Important clinically relevant features were extracted from the dataset using the Apriori algorithm and converted into fuzzy rules for each group. Data from a total of 180 patients, 90 in each group, were analyzed. Seven fuzzy rules were generated for the cryotherapy group and eight fuzzy rules were generated for the immunotherapy group. The resulting AUC of the cryotherapy and immunotherapy datasets was 0.902 and 0.813, respectively. The accuracy of both datasets was 80 and 98%, respectively.

Franciscis et al. (29) also used fuzzy logic to stratify the risk factors for developing chronic leg ulcers in patients in patients living with chronic venous disease (CVD). Data from seventy-seven CVD patients, 40 patients with active ulceration, 37 without, was analyzed. Twenty-seven clinically relevant features were generated for each patient. Results of the study were reported as risk scores, with the group of CVD patients with active venous ulceration being 32.38 ± 7.19%, and the group of CVD patients without active venous ulceration being 8.34 ± 3.38%.

Finally, Tan et al. (28) considered ten machine learning algorithms to determine the most predictive model for surgical complexity post-periocular basal cell carcinoma (BCC) excision. Data from 156 periocular BCC patients was analyzed, with seven clinically relevant features per patient. The most predictive model was Naive Bayesian classifier, with an average AUC of 0.854, and positive (PPV) and negative predictive values (NPV) of 38.1 and 94.1%, respectively. The second-best model was Alternating Decision Tree, achieving an AUC of 0.835, PPV of 31% and NPV of 97%.

DISCUSSION

Our review summarizes the current literature exploring the use of machine learning in predicting various dermatological outcomes. All studies conducted on this topic thus far have demonstrated promising outcomes. During a time where precision medicine is a focus of many clinicians, ML techniques provide a method for dermatologists to more accurately predict the clinical outcomes and prognoses of their patients in a variety of skin conditions.

When compared to traditional statistical methods, which focuses on inference, ML methodology focuses more on prediction (31). This is to say that ML methods aim to anticipate future behavior, rather than just drawing associations between data. ML is also particularly useful when looking at complex and detailed datasets with a large number of input variables. In fact, a larger sample size allows ML algorithms to better make associations within the data and thus, form more accurate outputs (32). Traditional statistical methods were designed to be most accurate and successful with a small to moderate number of input variables. As the number of inputs increases, the statistical models tend to become less precise.

While there are many benefits to the implementation of ML in a clinical dermatology setting, it is critical to discuss potential limitations to its implementation as well. One area of concern is the quantity of data required to operate ML algorithms. Some large dermatology patient registries do exist (33), however, there will need to be significantly more national and international collaboration to ensure that there is comprehensive coverage of dermatologic data.

Another limitation is the lack of ability for human operators to explain how ML algorithms make their conclusions. While there are methods to assess an algorithm's performance, there is no way to rationalize its decision. As such, they are often called "black box" technology (34). In this case, interpretation of the ML results by an experienced clinician is of utmost importance.

FUTURE DIRECTIONS

Although early studies assessing the prognostic value of machine learning in dermatology have demonstrated promising outcomes, further research is needed. To address whether the use of ML in predicting outcomes is truly a worthwhile avenue for clinicians to explore, prospective randomized clinical trials are needed.

CONCLUSION

Machine learning is a quickly advancing field in medicine and can be of great utility to clinicians in the near future, particularly in predicting the prognoses of complex dermatological conditions. As this technology advances, dermatologists will need to develop a foundational understanding of how it works and when it should be appropriately used in their clinical practice.

AUTHOR CONTRIBUTIONS

AD, SE, and RG have all provided substantial contributions to the conception and design of the work. AD drafted the work, SE and RG revised it critically for important intellectual content. All authors provide approval for publication of the content and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

REFERENCES

1. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Corrigendum: dermatologist-level classification of skin cancer with deep neural networks. *Nature.* (2017) 546:686. doi: 10.1038/nature 22985

2. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur J Cancer.* (2019) 111:148–54. doi: 10.1016/j.ejca.2019. 02.005

3. Guzman LCD, De Guzman LC, Maglaque RPC, Torres VMB, Zapido SPA, Cordel MO. Design and evaluation of a multi-model, multi-level artificial neural network for eczema skin lesion detection. In: *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)* (Kota Kinabalu) (2015). doi: 10.1109/AIMS.2015.17

4. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: A first comparative study of its kind. *Comput Methods Programs Biomed.* (2016) 126:98–109. doi: 10.1016/j.cmpb.2015.11.013

5. Han SS, Park GH, Lim W, Kim MS, Im Na J, Park I, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS ONE.* (2018) 13:e0191493. doi: 10.1371/journal.pone. 0191493

6. Ferris LK, Harkes JA, Gilbert B, Winger DG, Golubets K, Akilov O, et al. Computer-aided classification of melanocytic lesions using dermoscopic images. *J Am Acad Dermatol.* (2015) 73:28. doi: 10.1016/j.jaad.2015. 07.028

7. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* (2018) 22:1589–604. doi: 10.1109/JBHI.2017.2 767063

8. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc.* (2017) 24:198–208. doi: 10.1093/jamia/ocw042

9. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: a deep learning approach. In: *Proceedings of the 2016 SIAM International Conference on Data Mining Proceedings. (Society for Industrial and Applied Mathematics)* (Miami, FL) (2016). p. 432–40. doi: 10.1137/1.978161197 4348.49

10. Morgan DJ, Bame B, Zimand P, Dooley P, Thom KA, Harris AD, et al. Assessment of machine learning vs standard prediction rules for predicting hospital readmissions. *JAMA Netw Open.* (2019) 2:e190348. doi: 10.1001/jamanetworkopen.2019.0348

11. Kehl KL, Elmarakeby H, Nishino M, Van Allen EM, Lepisto EM, Hassett MJ, et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol.* (2019) 5:1421–9. doi: 10.1001/jamaoncol.2019.1800

12. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine Learning Methods to Predict Diabetes Complications. *J Diabetes Sci Technol.* (2018) 12:295–302. doi: 10.1177/19322968177 06375

13. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* (2016) 316:2402–10. doi: 10.1001/jama.2016.17216

14. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* (2017) 38:500–7. doi: 10.1093/eurheartj/ ehw188

15. Hogarty DT, Mackey DA, Hewitt AW. Current state and future prospects of artificial intelligence in ophthalmology: a review. *Clin Exp Ophthalmol.* (2019) 47:128–39. doi: 10.1111/ceo.13381

16. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Eng J Med.* (2019) 380:1347–58. doi: 10.1056/NEJMra18 14259

17. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med.* (2017) 376:2507–9. doi: 10.1056/NEJMp1702071

18. Hildt E. Artificial intelligence: does consciousness matter? *Front Psychol.* (2019) 10:1535. doi: 10.3389/fpsyg.2019.01535

19. Flowers JC. Strong and weak ai: deweyan considerations. In: *AAAI Spring Symposium: Towards Conscious AI Systems.* Available online at: http://ceur-ws.org/Vol-2287/paper34.pdf

20. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *Radiographics.* (2017) 37:505–15. doi: 10.1148/rg.20171 60130

21. Moore MM, Slonimsky E, Long AD, Sze RW, Iyer RS. Machine learning concepts, concerns and opportunities for a pediatric radiologist. *Pediatr Radiol.* (2019) 49:509–16. doi: 10.1007/s00247-018-4277-7

22. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. *Radiographics.* (2017) 37:2113–31. doi: 10.1148/rg.2017170077

23. Hogarty DT, Su JC, Phan K, Attia M, Hossny M, Nahavandi S, et al. Artificial intelligence in dermatology-where we are and the way to the future: a review. *Am J Clin Dermatol.* (2020) 21:41–7. doi: 10.1007/s40257-019-00462-6

24. Emam S, Du AX, Surmanowicz P, Thomsen SF, Greiner R, Gniadecki R. Predicting the Long-term Outcomes of Biologics in Psoriasis Patients using Machine Learning. *Br J Dermatol.* (2020) 182:1305–7. doi: 10.1111/bjd. 18741

25. Wang H-H, Wang Y-H, Liang C-W, Li Y-C. Assessment of deep learning using nonimaging information and sequential medical records to develop a prediction model for nonmelanoma skin cancer. *JAMA Dermatol.* (2019) 155:1277–83. doi: 10.1001/jamadermatol.2019.2335

26. Roffman D, Hart G, Girardi M, Ko CJ, Deng J. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. *Sci Rep.* (2018) 8:1701. doi: 10.1038/s41598-018-19907-9

27. Khozeimeh F, Alizadehsani R, Roshanzamir M, Khosravi A, Layegh P, Nahavandi S. An expert system for selecting wart treatment method. *Comput Biol Med.* (2017) 81:167–5. doi: 10.1016/j.compbiomed.2017.01.001

28. Tan E, Lin F, Sheck L, Salmon P, Ng S. A practical decision-tree model to predict complexity of reconstructive surgery after periocular basal cell carcinoma excision. *J Eur Acad Dermatol Venereol.* (2017) 31:14012. doi: 10.1111/jdv.14012

29. de Franciscis S, Fregola S, Gallo A, Argirò G, Barbetta A, Buffone G, et al. PredyCLU: a prediction system for chronic leg ulcers based on fuzzy logic; part I - exploring the venous side. *Int Wound J.* (2016) 13:12529. doi: 10.1111/iwj.12529

30. Phuong NH, Kreinovich V. Fuzzy logic and its applications in medicine. *Int J Med Inform.* (2001) 62:5 doi: 10.1016/S1386-5056(01)00160-5

31. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* (2018) 15:233–4. doi: 10.1038/nmeth.4642

32. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep.* (2016) 6:26094. doi: 10.1038/srep26094

33. DiMarco G, Hill D, Feldman SR. Review of patient registries in dermatology. *J Am Acad Dermatol.* (2016) 75:824–9. doi: 10.1016/j.jaad.2016.03.020

34. Wang F, Kaushal R, Khullar D. Should health care demand interpretable artificial intelligence or accept "black box" medicine? *Ann Intern Med.* (2020) 172:59. doi: 10.7326/M19-2548