Received: 16 September 2019 | Accepted: 9 November 2019
DOI: 10.1111/srt.12822

ORIGINAL ARTICLE

WILEY

Using content-based image retrieval of dermoscopic images for interpretation and education: A pilot study

Mahya Sadeghi[1] | Parmit Chilana[1] | Jordan Yap[2] | Philipp Tschandl[1,3] | M. Stella Atkins[1,4]

[1]School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
[2]MetaOptima Technology Inc., Vancouver, BC, Canada
[3]Department of Dermatology, Medical University of Vienna, Vienna, Austria
[4]School of Dermatology and Skin Science, University of British Columbia, Vancouver, BC, Canada

Correspondence
M. Stella Atkins, School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.
Email: stella@sfu.ca

Funding information
Canadian Natural Sciences and Engineering Research Council, Grant/Award Number: 311106

Abstract

Background: Dermoscopic content-based image retrieval (CBIR) systems provide a set of visually similar dermoscopic (magnified and illuminated) skin images with a pathology-confirmed diagnosis for a given dermoscopic query image of a skin lesion. Although recent advances in machine learning have spurred novel CBIR algorithms, we have few insights into how end users interact with CBIRs and to what extent CBIRs can be useful for education and image interpretation.

Materials and Methods: We developed an interactive user interface for a CBIR system with dermoscopic images as a decision support tool and investigated users' interactions and decisions with the system. We performed a pilot experiment with 14 non-medically trained users for a given set of annotated dermoscopic images.

Results: Our pilot showed that the number of correct classifications and users' confidence levels significantly increased with the CBIR interface compared with a non-CBIR interface, although the timing also increased significantly. The users found the CBIR interface of high educational value, engaging and easy to use.

Conclusion: Overall, users became more accurate, found the CBIR approach provided a useful decision aid, and had educational value for learning about skin conditions.

KEYWORDS
classification decision aids, content-based image retrieval, dermoscopy images, diagnosis accuracy, educational, user study

1 | INTRODUCTION

Recent advances in artificial intelligence (AI) and computer-aided decision support methods have produced various efficient ways to allow for learning about skin problems.[1] In particular, advances in machine learning have spurred novel retrieval algorithms and aroused interest in *content-based image retrieval* (CBIR) techniques, where computer vision methods are applied to search for similar images to a "query" image based on the content of the image and visual clues such as color, shape, and pattern, from large databases.[2] In the medical domain, CBIR is designed to assist with finding similar, labeled, medical images from a curated database. Within the dermatology context, CBIR can assist with diagnosis or education by comparing visually similar skin lesion images,[3] removing the difficulties that can arise when trying to describe images with words. Since the database and the algorithms for these systems are curated for a specific area or problem, users are less likely to encounter irrelevant images, one of the main problems with generic search engines.

Despite the proposed benefits of modern CBIR systems, most CBIR-related research to date has focused on improving the accuracy of AI systems for diagnostic decisions[4,5]: we know little about the perceived utility and usability of CBIR systems for end users from a human-computer interaction (HCI) perspective.[6] In this paper, we describe a pilot study on how an interactive dermoscopic

CBIR-based decision support tool can be used to help classify skin lesion images, and determine if the tool has educational value for users without prior dermoscopy training. Pagnanelli et al suggested that web-based training is an efficient tool for teaching dermoscopy to nonexperts,[7] so we included a brief dermoscopy web-based tutorial before the study in order provide some knowledge to participants who had no prior dermoscopy training.

Our key research questions were as follows:

RQ1: To what extent does a CBIR system help untrained dermoscopy users accurately identify a subset of four important skin conditions commonly observed in clinical practice (nevus, seborrheic keratosis, basal cell carcinoma, and malignant melanoma)?
RQ2: How does CBIR affect user confidence, trust, and timing in making a decision?
RQ3: To what extent is CBIR perceived to be educational for untrained dermoscopy users?

Our main contribution is in designing a user interface for a dermatological CBIR system and performing pilot studies that provide insights into how this system is used and perceived by untrained dermoscopy users. Our results have several implications for designing interfaces for CBIRs and other decision support tools for helping users search, explore, and learn from medical image collections.

2 | RELATED WORK

Our work builds upon existing research on information retrieval and recent CBIR systems for medicine, especially in the dermatology field. We consider current challenges that end users face in finding medical information online and summarize existing efforts on evaluating and improving CBIR-based decision support systems in other contexts.

2.1 | Challenges in Finding Relevant Visual Medical Information Online

CBIR has been proposed to be promising for medical information retrieval,[8] and the last decade has witnessed great interest in research on CBIR systems in the dermatology field.[9-11] Rise of computing power and large image datasets also has helped to pave the way for a large number of new techniques and systems for improving CBIR retrieval accuracy, and the domain of CBIR for dermatology through improved algorithms is moving forward quickly.[12,13]

2.2 | Learning about Dermatological Concepts

Dermoscopy is a useful tool for people attempting early diagnosis of melanoma.[14] Although diagnostic accuracy is directly dependent on the experience of the observer, studies have shown that dermoscopy improves diagnostic accuracy in comparison with clinical diagnosis with the naked eye for skin lesions.[15] Patient education on skin self-examination is also effective for improving the rate of early detection of melanoma.[16] Despite significant use of dermoscopy, dermoscopy training is usually time-consuming, is often limited to dermatologists, and is often not provided for general practitioners.[17,18] These issues highlight the opportunities for advanced decision support tools to be used as educational means in dermoscopy training for people with different levels of expertise.

In educational settings, an ideal clinical decision support system could provide images relevant to a clinical query for special cases,[19] as these images often convey essential information and can be very valuable for interpretation and education. In,[20] authors developed a retrieval system to assist clinicians in self-learning and differential diagnosis of lung cancer to reduce the inter-observer variability using the information of similar nodules. Searching for visually similar images with different diagnoses can also be valuable for teaching. Moreover, based on,[21] their CBIR-based tool minimizes the risk of missing critical lesions. However, to improve the precision of case-based retrieval, future work is required.

2.3 | User Interaction with CBIR Systems

The need for a user test in the clinical environment to measure the impact of an efficient retrieval system on diagnosis quality is mentioned in a number of scientific publications [22,23]; however, only a few medical CBIR systems have been developed in direct collaboration with end users and evaluated in a real workflow.[7,24]

CBIR applications are inherently visual in nature, so an effective user interface plays a key role. Dermoscopic images themselves can be improved with image processing techniques such as using high dynamic range conversion [25]; user evaluation of the enhanced images was positive, but the increased accuracy of diagnostic decision was not reported. Improving the user interface is one of the biggest challenges in developing clinical CBIR decision support tools.[26] There have been some studies on user-oriented design and evaluation of medical CBIR tools. The goal of these studies was mainly to assess the system usability and identify user-centered evaluation of the CBIR tool and potential improvements in the system and interface.[27] It has been also stated that making a variety of queries, creating relationships between image features, assigning importance to features, and forming hybrid queries by intelligently combining text and image queries are some of the necessary features of a CBIR user interface.[28]

An effective decision support system must also be easy to use and should minimize the users' effort to receive and act based on the important information provided.[29] Usability evaluation based on realistic scenarios with target user groups is needed for educational and diagnostic purposes. Hence, direct cooperation between medical practitioners and medical computer scientists is necessary to determine user expectation from a decision support system.[30] It is also mentioned that one of the most important factors in image retrieval research is evaluation of the behavior of retrieval system users. Each user has different needs and preferences, so it is important to detect users' preference and provide an efficient interface to them.[31]

3 | METHOD

3.1 | Study design

We used an experimental design, where all the participants made decisions about query images under two conditions: non-CBIR and CBIR. We built our CBIR system based on a previously published CBIR image retrieval algorithm, which was shown to be also effective for classification by taking the 16 most similar images.[3]

3.2 | Study participants

Fourteen non-medically trained participants (six males and eight females, aged 25-35) successfully completed the laboratory experiment. The participants were all professionals, in software or hardware engineering (5/14), image processing/graphics (5/14), and in business (4/14), but none of them had any training in medicine. This study was advertised through an electronic mailing list and was approved by the Simon Fraser University Ethics Board.

3.3 | System and user interface description

To run this study, we designed a user interface for enabling content-based image retrieval using an existing classifier.[3] Using the existing algorithm, our decision support tool retrieved all the similar images for each query image, sorted in ascending order based on the distance of deep feature vectors from a deep neural network trained to classify images, where the first similar image was the most visually similar image to the query image. Our interface displayed thumbnails of the 15 most similar images.

3.3.1 | User interface for the non-CBIR condition

In the non-CBIR condition, each user was presented with a query image one at a time and was asked to choose the single best classification category by clicking on the appropriate button, where buttons were color coded with respective skin lesion category. Users were also requested to select a confidence level on their decision on a Likert scale score in scale of 1 (least confident) to 5 (most confident) for each query.

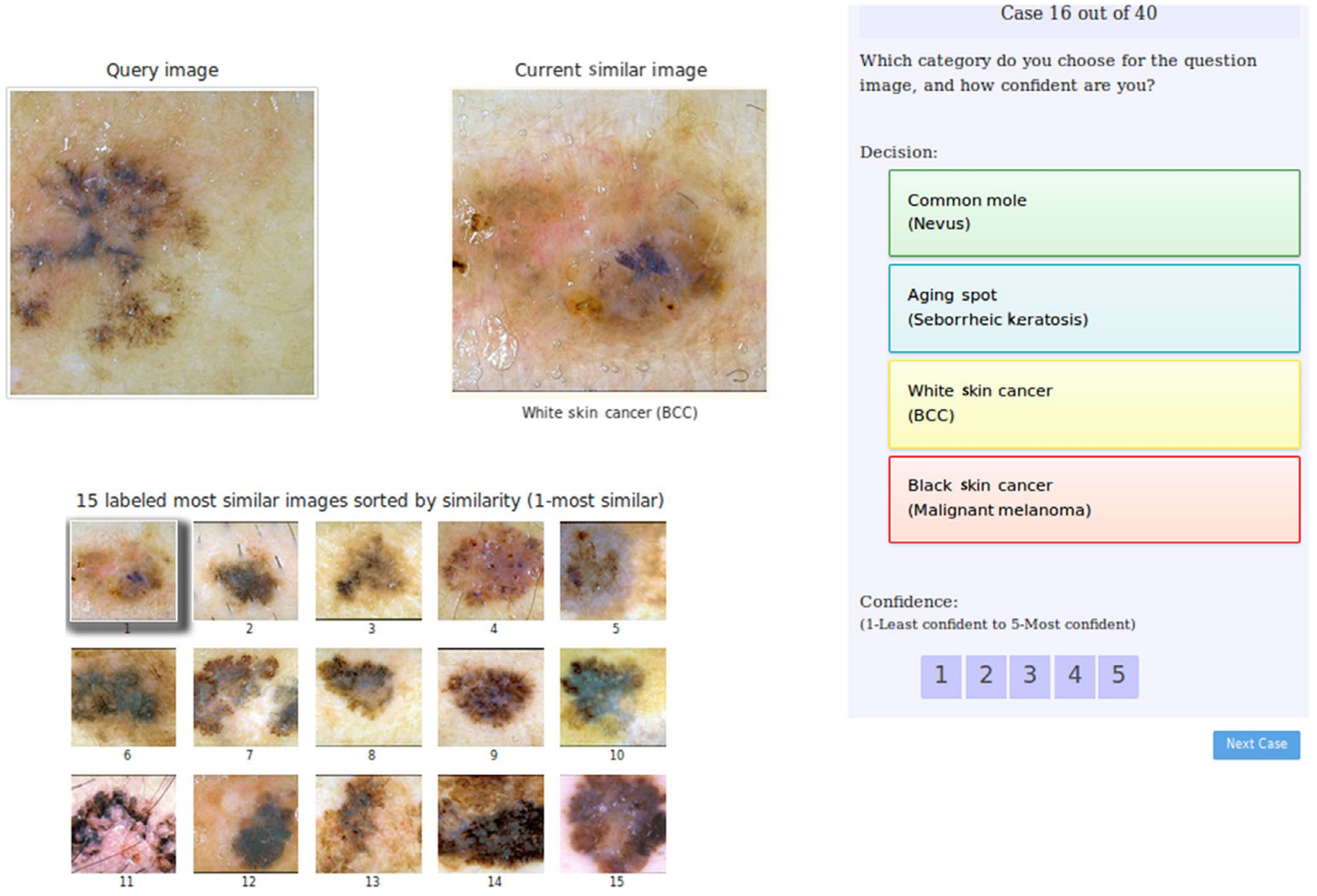

FIGURE 1 Sample screenshot of the CBIR condition

3.3.2 | User interface for the CBIR condition

In the CBIR condition, each user was presented with a query image and asked to classify it in the same manner as the non-CBIR condition. However, during the CBIR condition, the 15 most visually similar images of the collection were returned as thumbnails for each query image, sorted from top left row to the bottom right row, and participants had freedom to interpret the results and make a decision at their discretion. A sample screenshot of the CBIR condition can be found in Figure 1.

3.4 | Dataset

We selected all of the images from publicly available datasets, such as a Dermoscopy Atlas,[32] and the International Skin Imaging Collaboration (ISIC) archive.[33] We limited our study to 4 important skin lesion categories commonly observed in clinical practice, including those used in the ISIC classification challenge 2017,[34] that is, nevus, seborrheic keratosis (SK), basal cell carcinoma (BCC), and malignant melanoma (MM), resulting in 1021 usable images, with 448 nevi, 43 SK's, 42 BCC's, and 246 MMs. The remaining images included melanoma metastatis, vascular lesion, blue nevus, combined nevus, dermatofibroma, lentigo, melanosis, recurrent nevus, and spitz nevus, none of which appeared in the visual search of the query image. We also removed anything which is "acral" (fingers and toes) as these lesions have unique linear patterns. Note there were only 43 SK's in the dataset, because dermatologists do not usually capture easy SK's to diagnose. The classification accuracy of all images was confirmed by pathology, and the visual quality was approved by an expert dermatologist who had experience working with dermoscopic images. To simplify complex medical terms for general users, we used simple terminologies for each skin lesion category as shown in Figure 1.

3.5 | Experiment Protocol

3.5.1 | Pre-pilot studies

Before shaping the study, we performed two pre-pilot studies. In the first, we explored several interface design options, including showing differing numbers of similar images, and a pie chart representing proportions of the different diagnoses in the most similar images. In the second, we tested the interface with another design option, using colored borders around the retrieved images to indicate their diagnosis. We also used a different task order to discover learning effects: half of the users started with CBIR condition, and half started without CBIR.[35] From these studies, we learned that the pie chart and colored borders affect and inform the decision, so there were only a few clicks on thumbnails for full size image comparison. We also found a learning effect for users who started with CBIR, which confused the accuracy statistics. Hence, for this pilot, we used the following protocol:

3.5.2 | Pre-task Questionnaire

Participants answered their past experience in medical image search.

3.5.3 | Brief tutorial

For around 10 minutes, participants went through a brief tutorial session to learn about 4 skin categories commonly observed in clinical practice.

3.5.4 | Experimental task

Participants started the study by classifying 20 skin lesion images first in the non-CBIR (20 images) in random order, and then in the CBIR condition in random order.

3.5.5 | Post-task Questionnaire

Participants filled out a post-task questionnaire about the CBIR tool.

Users were asked to think aloud while classifying each image. We collected all these data which were mainly about images, difficulties they encounter, and suggestions for improvements.

Since our study contained complex dermoscopic images, users needed to have some knowledge about each skin lesion category (eg, features, patterns, and colors). In a realistic scenario, users will have freedom to explore the web, textbooks, and other resources to find relevant information. To simulate this scenario and provide some background knowledge, we provided users with a brief tutorial consisting of educational slides for 5-10 minutes before the experiment. All of these images, information, and simplified terminologies were approved by an expert dermatologist, who had experience working with dermoscopic images. During the study, users relied only on their knowledge from our trial or the CBIR results and did not have access to information from the educational slides.

Based on the results of our pre-pilot studies, for our experiment, we decided to use 40 randomly selected query images out of the 1021 images in our dataset: 20 without CBIR and 20 with CBIR. We selected 5 query images from each category for each condition to provide an equal disease distribution.

To reduce possible bias resulting from fatigue or learning effects, the order of "query" images was randomly selected. Once the study ended, participants were provided with total feedback on their performance. The same normal lighting condition with a large screen was provided for all users.

4 | RESULTS

4.1 | Accuracy

We recorded the users' decisions for each query image. Each image category has 70 entries: Five images classified by 14 users. Figure 2 shows the overall accuracy for each user, without CBIR and with CBIR.

Table 1 shows the confusion matrix between categories for classifications without CBIR and when using CBIR. Results with significant changes in accuracy where $P < .05$ between the two conditions are shown in bold, showing that accuracy for classifying nevus and MM images improved significantly with the CBIR. Diagonal elements show the number of correct decisions for each category and summing the accurate diagonal decisions total 139 (49.5%) without the CBIR condition and 196 (70%) with CBIR, a significant increase in accuracy with CBIR ($P < .05$).

4.2 | User confidence level

The confidence level of every decision was recorded automatically (see Section 3.3 and Figure 1) and summarized in Table 2. User confidence level for both the correct and incorrect classifications increased significantly in the CBIR condition.

4.3 | Timing

Computer logs recorded time spent on each condition by each participant. There was a significant difference ($P < .05$) in mean time from 451 seconds (SD = 16) in the non-CBIR condition to 815 seconds (SD = 16) in the CBIR condition. In the CBIR condition, the average number of clicks on each thumbnail to zoom the image to a higher resolution is shown in Figure 3.

4.4 | Educational Value

The educational value was measured in the post-task questionnaire as the level of usefulness on a 5-point Likert scale for learning about skin problems. Average value was 4.6/5, with 71% of users recording "Very Useful" and 21% recording "Useful."

About 5/14 users explicitly mentioned the educational value of the system: "*it is a fantastic educational tool …it is extremely important for supporting clinical decisions*," and "*Educational for both experts and non experts*."

The extent that CBIR can complement other resources such as books or Google image search was measured on a 5-point Likert scale, with a similar result; about 71% of users found it could complement other resources "a lot", and 21% recorded "Moderately", with average value 4.6/5.

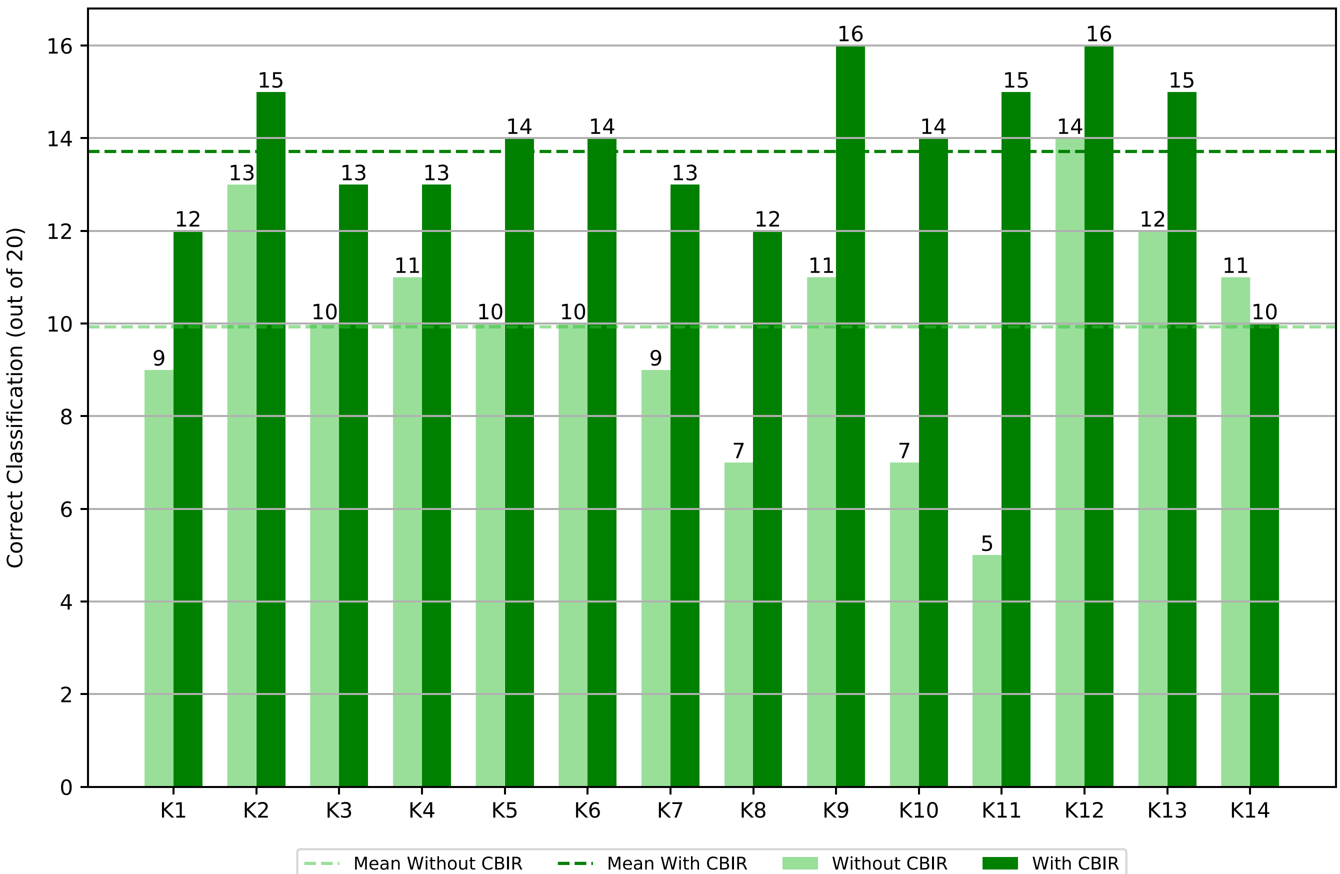

FIGURE 2 Accuracy of every user without and with CBIR

In addition, "self-education" was the first major motivation in future use of CBIR for the 13/14 users.

4.5 | Trust, ease of use, and engagement

Trust in the CBIR results was measured in the post-task questionnaire as: "To what extent did you trust the similar images in the CBIR tool?" on the Likert scale score on the scale of 1 (least confident) to 5 (most confident). The users self-reported a mean of 3.18/5 trust.

The majority of the users found our CBIR-based decision support tool "very easy to use" and "engaging" (see Figure 4).

4.6 | Users' Perceptions and Feedback

Participants' comments, concerns, or suggestions about usability and usefulness of the system were noted as "think-aloud protocol" during the study and were also collected in the post-task questionnaire.

One of the key reasons users did not incorporate diagnoses of visually similar images with their decisions was their perception of similarity. It was often stated by the different users that *"Even though system suggests… I go with…"* or *"in some cases that are skin cancer are very similar to common mole, so I can't distinguish between them."* The second situation usually happened when query image or similar images were rare cases.

Sometimes, no visually similar images were enlarged, which resulted in both correct and incorrect decisions.

TABLE 1 Confusion matrix without CBIR (A) when using CBIR (B) for all 70 decisions in each category (Five images in each category for 14 subjects). Results with significant changes in accuracy where $P < .05$ between the two conditions are shown in bold

	Decision			
True class	Nevus	SK	BCC	MM
(A) Without CBIR				
Nevus	**47**	17	2	4
SK	16	26	14	14
BCC	0	9	36	25
MM	6	12	24	**28**
Total	69	64	76	71
(B) With CBIR				
Nevus	**69**	0	0	1
SK	16	33	1	20
BCC	1	2	42	25
MM	11	2	5	**52**
Total	97	37	48	98

During the think-aloud protocol, we also observed that although the same educational trial was provided for all the users, different people made their decisions based on different features in the images (eg, colors, vessels, and dots).

In many cases, users found similar images very similar to query images. It was frequently reported: *"… it's very similar to this one."* In many cases, viewing similar images was helpful; however, in some cases, the similar images made the task challenging for the users: *"… some skin cancer cases are very similar to common mole, so I can't distinguish between them."* Some users mentioned that they would like to have a numeric "similarity to the query image" score on each retrieved image, and 3/14 users stated they need an explanation of why the images are found to be visually similar.

"Maybe the retrieval system can bring up images that look more similar and also point to similar features eg, dots." These three users also stated they want a statistical chart—and on more than 15 images.

Some users stated that they need "zoom option"; 3/14 users stated that they need "Information on images" or "Image Scale" for making a more accurate and confident decision.

"Having a scale on the images would be very helpful, I wasn't able to tell how magnified an image was or tell if some of the skin features were the same size."

One user stated more information about each disease category would be useful:

"Having a tip or info section and showing some features of each type of diagnosis for people who need a refresh on the concept."

5 | DISCUSSION

In this paper, we used a multidisciplinary approach to investigate how a dermatological CBIR system is used and perceived by users. Our results have several implications for designing CBIRs and other decision support tools to help end users search, explore, and learn from medical image collections.

5.1 | Accuracy of classification

The overall accuracy of non-medical users in classifying the images increased significantly with the CBIR-based decision support tool, so our results suggest that CBIR can help users in image interpretation. However, the users' performance on SK images did not significantly increase in the CBIR condition. The reason for the poor performance of identifying the SK category likely arose from the small number of SK images (43/1021) in the dataset which resulted in fewer similar images from the SK category. The SK cases included in the database are also difficult cases which were confused with malignant melanoma and sent to the pathology laboratory and means the database may be biased toward difficult cases and may not have enough representation for easy benign cases such as obvious and easy to diagnose SK images.

These findings show the importance of creating a comprehensive database that can represent the distribution of cases that the user group will be facing in their real-life clinical practice. This can be extended to all other classes of skin disease that may not be present in the database used for image retrieval. Hence, it is very important to inform the users about the database size and the level of coverage for different disease categories.

TABLE 2 Average user confidence level without and with CBIR

Classification	without CBIR	with CBIR
Confidence on correct decisions	**3.60 (0.56)**	**4.02 (0.46)**
Confidence on incorrect decisions	**3.09 (0.37)**	**3.61 (0.56)**
Average confidence	**3.35 (0.4)**	**3.82 (0.47)**

Note: Significant results where $P < .05$ are shown in bold (SD of classifications are shown in parentheses).

5.2 | Confidence

Overall, the mean confidence on every decision, both accurate and inaccurate, increased significantly when using CBIR, suggesting that it is important to ensure that decision support tools and the algorithms behind them are safe and effective. Hence, scientific evaluation and clinical testing are necessary to measure factors related to user confidence level in the decision making process.

5.3 | Timing

Almost twice the time was needed to make a decision in the CBIR condition, which would be very important in clinical practice. The workflow must be carefully designed to deliver value to the end users. Some of the extra time involved clicking on thumbnails to

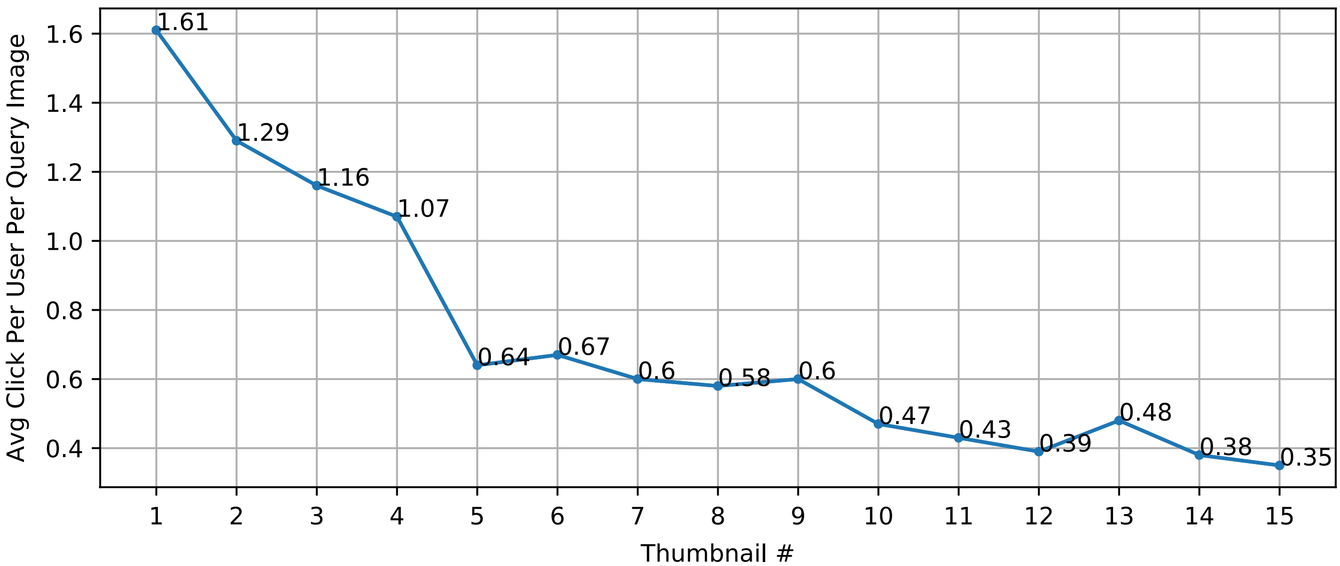

FIGURE 3 Average clicks to zoom a thumbnail to a higher resolution

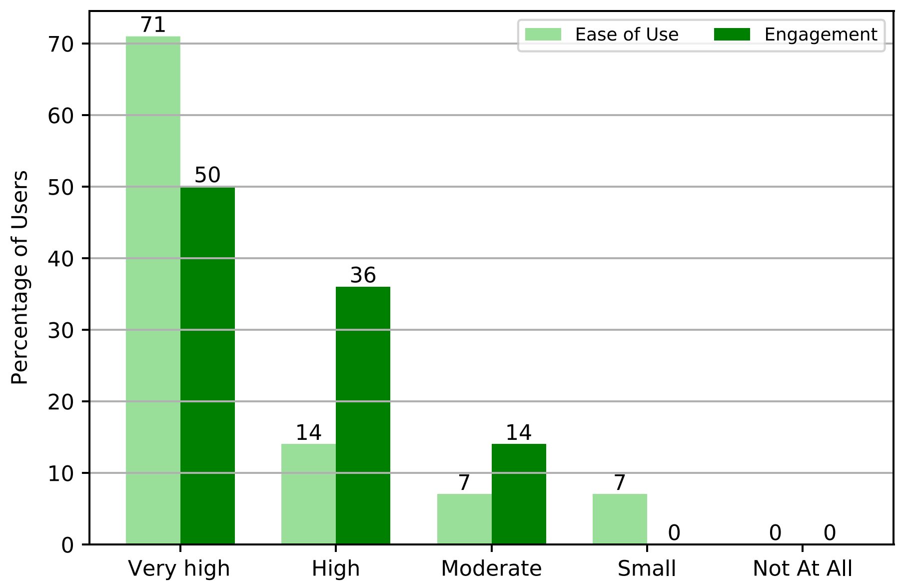

FIGURE 4 Level of ease of use and engagement (data represent percentage of the users)

zoom the images for visual comparison. On average, users magnified the first four similar images (the closest ones) almost twice as often as the remaining images (Figure 3). Future design considerations may include providing fewer visually similar image thumbnails in the search or providing one or two closest images already zoomed up.

5.4 | Educational value

According to our pre-task and post-task questionnaires, we learned that more than two thirds of the users in our study found our CBIR system “very useful” for learning about skin problems, and the CBIR system can complement other resources (eg, books, Google image search, etc). These findings demonstrate that CBIR has the high potential to be used as an educational tool for the common cases for both general and expert users.

5.5 | Trust, ease of use, and engagement

Trust is a challenging factor for medical decision support tools, and some studies propose that a second opinion by a decision support system is not always welcomed by clinicians when it does not match their own initial diagnosis.[6]

Since retrieving similar cases can provide a diagnostic support environment rather than a single second diagnosis, we investigated “trust in CBIR results” as another critical factor in the post-task questionnaire.

All of the users self-reported a medium level of trust (mean trust of 3.28/5), respectively, in the retrieved data. The main reason is likely that many of the retrieved images were not “similar enough” to the query image; this situation would be remedied with a much larger curated database of images for each kind of classification.

All of the users reported a high level of ease of use and engagement. More than 75% of the users found our CBIR system “easy to use.” More than 85% of the users ranked our CBIR system “engaging.” Although “ease of use” and “engagement” levels are not critical factors in developing medical decision support tools, they are essential for acceptance and establishing an effective user interaction.

5.6 | Visual elements in CBIR and implications for design

5.6.1 | Number of retrieved cases

In a previous study comparing CBIR accuracy with diagnostic accuracy,[3] the diagnostic accuracy obtained using CBIR to retrieve 16 visually similar images was similar to predictions made by a neural network. However, our findings demonstrate that users do not always use all 16 images; users magnified the first four similar cases far more than the rest (see Figure 3). In addition, there is a trade-off between number of thumbnails and size of thumbnails that you can fit in the view. Therefore, showing a larger number of retrieved cases results in smaller thumbnails where users may not be able to see the details.

5.6.2 | Design choices to show the diagnosis

In a CBIR decision support tool, it is also important how you present the diagnosis to the user. To show the diagnosis, we considered four different design choices as follows:

5.6.3 | Colored statistical charts

According to data from our pre-pilot study, the first impression of the colored statistical chart introduces bias to user decision. For example, one of the pre-pilot users (P2) was trying not to look at the images and chose her decision mainly based on the majority of the cases in the pie chart. The statistical chart is highly dependent on the diagnosis of similar cases and ignores the visual similarity which is an important factor in designing CBIR systems.

5.6.4 | Colored borders

In general, the colored borders used in the pre-pilot study[35] highly affected the user's decisions, interaction, and behavior. Providing users with the colored borders may encourage or discourage them from clicking on specific cases. First, impression of the colored borders introduces bias to user decision, most probably through creation of a “mental frequency estimation” of diagnoses. Using colored borders may work for a limited number of classes, for example to differentiate malignant from benign cases. However, in a real-world setting, it is impossible to use the colored borders that can be representative for a multitude of categories of skin lesions. Finally, colored borders also affected the visual appearance of similar images and how they looked to the users.

5.6.5 | No diagnosis

Users have no diagnosis for the thumbnails, so it takes significantly more time to make a decision since they need to magnify the similar thumbnails in order to see the associated diagnosis. Furthermore, users need to remember the diagnosis of cases they found similar, which can be frustrating and less helpful for the user.

5.6.6 | Mnemonic text

In a more realistic yet effective design, we can attach diagnosis to thumbnails, so users can see the diagnosis of each retrieved case underneath as a text. This method also removes the difficulties of finding appropriate colors representing different diagnosis, when the number of lesion categories is large. We infer from results that different user interface choices should be designed and tested in future studies for every use case.

5.6.7 | Visual elements and features

Zooming option, similarity scores for retrieved cases, and patient information such as age and gender are other potential choices and may help toward more accurate decisions. Showing extracted features by the AI algorithms in the tool is another interesting challenge.

5.7 | Limitations

First, our pilot study works on limited and public datasets available at the time of study commencement, which included only 1021 images. As the dataset size grows larger, the algorithm is likely to return more similar images for the query images. The study would be more comprehensive with more available newly published datasets like HAM10000,[36] which includes 10 000 skin lesion images. Size and quality of the training data is another important factor for improving retrieval accuracy.

This study was only based on images and visual characteristics of skin lesions. As such, the study may not truly reflect the real decision making process, and we need to address complete factors for a comprehensive study.

Another major limitation is level of user expertise, and our participants were limited to the non-medically trained population. Dermatologists and medical professionals are hard to recruit because of time constraints, so we tested this tool with a participant sample from the general population.

While this study presents the findings that apply to the non-medically trained users, future work can address these limitations by extending the investigations to more expert groups, medical students, and patients.

6 | CONCLUSION AND FUTURE WORK

In our user-centered design approach, the problem of skin lesion classification is considered, where our experimental results indicate that CBIR is indeed effective, showing that the number of correct classifications and user confidence level is increased with showing pre-diagnosed similar images to the user. CBIR was also perceived to be very easy to use, engaging, and useful for learning about skin problems.

For a complete evaluation, it is important to explain the technology, with its benefits, problems, and limitations to the users so that they will have a practical idea of what can be achieved or expected. Having a clear proof of retrieval quality based on standardized datasets can also be useful for medical professionals to build trust. Clinicians need to give their opinion on the usability and usefulness of the technology, and trust needs to be gained before CBIR-based decision support tools can be used in clinical settings. Such collaboration with expert and non-expert users can also improve the interface significantly, when good feedback and comments are delivered.

In future work, we need to consider whether the results transfer to other user groups. Although initial feedback from general physicians shows their knowledge of dermoscopy for skin lesions is almost as limited as the users we tested, future studies should perform user studies with healthcare professionals to confirm these findings. We also need to investigate how they can adopt CBIR in clinical settings safely for better patient care outcome and more efficient workflow.

ORCID

M.Stella Atkins https://orcid.org/0000-0003-1507-4953

REFERENCES

1. Rigel DS, Carucci JA. Malignant melanoma: prevention, early detection, and treatment in the 21st century. *Cancer J Clin*. 2000;50(4):215-236.
2. Estrela VV, Herrmann AE. Content-based image retrieval (CBIR) in remote clinical diagnosis and healthcare. *Encyclopedia of E-Health and Telemedicine*. *IGI Global*. 2016;495-520.
3. Tschandl P, Argenziano G, Razmara M, Yap J. Diagnostic accuracy of content based dermatoscopic image retrieval with deep classification features. *Br J Dermatol*. 2019;181(1):155-165.
4. Esteva A, Kuprel B, Novoa R, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115.
5. Celebi ME, Codella N, Halpern A. Dermoscopy image analysis: overview and future directions. *IEEE J Biomed Health Inform*. 2019;23(2):474-478.
6. Cai CJ, Reif E, Hegde N, Hipp J, et al.Human-centered tools for coping with imperfect algorithms during medical decision-making. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, https://doi.org/10.1145/3290605.3300234, ACM.
7. Pagnanelli G, Soyer HP, Argenziano G, et al. Diagnosis of pigmented skin lesions by dermoscopy: web-based training improves diagnostic performance of non-experts. *Br J Dermatol*. 2003;148(4):698-702.
8. Faria AV, Oishi K, Yoshida S, Hillis A, Miller MI, Mori S. Content-based image retrieval for brain MRI: An image-searching engine and population-based analysis to utilize past clinical data for future diagnosis. *Neuroimage Clin*. 2015;7:367-376.
9. Baldi A, Murace R, Dragonetti E, et al. Definition of an automated content-based image retrieval (CBIR) system for the comparison of dermoscopic images of pigmented skin lesions. *Biomed Eng Online*. 2009;8(1):18.
10. Ballerini L, Li X, Fisher RB, Rees J.A query-by-example content-based image retrieval system of non-melanoma skin lesions. MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support, volume LNCS 5853: 31–38. Springer, 2009.
11. Nemat E, Ghasemian H. A hierarchical content-based image retrieval approach to assisting decision support in clinical dermatology. Iranian Journal of Electrical and Computer Engineering (IJECE). 2010.
12. Abbas Q. Content-based image retrieval system for clinical diagnosis of pigmented skin lesions. *IJCSNS*. 2017;17(5):238.
13. Benam A, Drew MS, Atkins MS.A CBIR system for locating and retrieving pigment network in dermoscopy images using dermoscopy

interest point detection. Biomedical Imaging (ISBI 2017), IEEE 14th International Symposium on: 122–125. IEEE, 2017.
14. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions–a valuable tool for early diagnosis of melanoma. *Lancet Oncol*. 2001;2(7):443-449.
15. Herschorn A. Dermoscopy for melanoma detection in family practice. *Can Fam Physician*. 2012;58(7):740-745.
16. Goulart JM, Malvehy J, Puig S, Martin G, Marghoob AA. Dermoscopy in skin self-examination: a useful tool for select patients. *Arch Dermatol*. 2011;147(1):53-58.
17. Chen YA, Rill J, Seiverling EV. Analysis of dermoscopy teaching modalities in United States dermatology residency programs. *Dermatol Pract Concept*. 2017;7(3):38.
18. Patel P, Khanna S, McLellan B, Krishnamurthy K. The need for improved dermoscopy training in residency: a survey of US dermatology residents and program directors. *Dermatol Pract & Concept*. 2017;7(2):17.
19. Rachana P, Ranjitha S, Suresh HN. Image retrieval by using histogram equalization and CBIR Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on: 3352–3356. IEEE, 2016.
20. Dhara AK, Mukhopadhyay S, Dutta A, Garg M, Khandelwal N. Content-based image retrieval system for pulmonary nodules: Assisting radiologists in self-learning and diagnosis of lung cancer. *J Digit Imaging*. 2017;30(1):63-77.
21. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph*. 2007;31(4-5):198-211.
22. Markonis D, Holzer M, Baroz F, et al. User-oriented evaluation of a medical image retrieval system for radiologists. *Int J Med Informatics*. 2015;84(10):774-783.
23. Agarwal M, Mostafa J. Image retrieval for Alzheimer's Disease detection. In MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support, pp 49–60. Springer, 2009.
24. Jiménez-del Toro OA, Hanbury A, Langs G, Foncubierta-Rodríguez A, Müller H. Overview of the Visceral Retrieval Benchmark 2015. In: Müller H, Jimenez del Toro OA, Hanbury A, Langs G, Rodriguez AF, eds. *Multimodal Retrieval in the Medical Domain*. Berlin, Germany: Springer; 2015:115-123.
25. Kaliyadan F. High dynamic range conversion of dermoscopy images in general dermatology conditions—A pilot study. *Skin Research and Technology*. 2019;25(6):867-870.
26. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform*. 2008;41(2):387-392.
27. McDonald S, Lai T, Tait J. Evaluating a content based image retrieval system. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval: 232–240. ACM, 2001.
28. Antani SK, Long LR, Thoma GR. Content-based image retrieval for large biomedical image archives. *Medinfo*. 2004;107(Pt 2):829-833.
29. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765.
30. Müller H, Rosset A, Garcia A, Vallée J, Geissbuhler A. Benefits of content-based visual data access in radiology. *Radiographics*. 2005;25(3):849-858.
31. Cha HJ, Kim YS, Park SH, Yoon TB, Jung YM, Lee J. *Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system*. Berlin, Germany: Springer; 2006:513-524. International Conference on Intelligent Tutoring Systems
32. Argenziano G, Soyer HP, De Giorgi V, Piccolo D, Carli P, Delfino M. Interactive atlas of dermoscopy (book and cd-rom). 2000.
33. ISIC Archive. https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery, 2018. Accessed July 31, 2019.
34. Codella NCF, Gutman D, Celebi ME, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on: 168–172. IEEE, 2018.
35. Sadeghi M, Chilana PK, Atkins MS.How users perceive content-based image retrieval for identifying skin images. Medical Image Computing and Computer Assisted Interventions, MICCAI Workshop 2018, volume LNCS 11038: 141–148. Springer, 2018.
36. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. 2018;5(1). https://doi.org/10.1038/sdata.2018.161

How to cite this article: Sadeghi M, Chilana P, Yap J, Tschandl P, Atkins MS. Using content-based image retrieval of dermoscopic images for interpretation and education: A pilot study. *Skin Res Technol*. 2019;00:1–10. https://doi.org/10.1111/srt.12822