EBioMedicine 43 (2019) 107–113

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

# Skin cancer detection by deep learning and sound analysis algorithms: A prospective clinical study of an elementary dermoscope

A. Dascalu [a,*], E.O. David [b]

[a] Department of Physiology and Pharmacology, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
[b] Department of Computer Science, Bar-Ilan University, Ramat-Gan, Israel

ARTICLE INFO

*Article history:*
Received 18 March 2019
Received in revised form 16 April 2019
Accepted 29 April 2019
Available online 14 May 2019

*Keywords:*
Skin cancer
Deep learning
Dermoscopy
Sonification
Melanoma
Telemedicine
Artificial intelligence

ABSTRACT

*Background:* Skin cancer (SC), especially melanoma, is a growing public health burden. Experimental studies have indicated a potential diagnostic role for deep learning (DL) algorithms in identifying SC at varying sensitivities. Previously, it was demonstrated that diagnostics by dermoscopy are improved by applying an additional sonification (data to sound waves conversion) layer on DL algorithms. The aim of the study was to determine the impact of image quality on accuracy of diagnosis by sonification employing a rudimentary skin magnifier with polarized light (SMP).

*Methods:* Dermoscopy images acquired by SMP were processed by a first deep learning algorithm and sonified. Audio output was further analyzed by a different secondary DL. Study criteria outcomes of SMP were specificity and sensitivity, which were further processed by a F2-score, i.e. applying a twice extra weight to sensitivity over positive predictive values.

*Findings:* Patients ($n = 73$) fulfilling inclusion criteria were referred to biopsy. SMP analysis metrics resulted in a receiver operator characteristic curve AUC's of 0.814 (95% CI, 0.798–0.831). SMP achieved a F2-score sensitivity of 91.7%, specificity of 41.8% and positive predictive value of 57.3%. Diagnosing the same set of patients' lesions by an advanced dermoscope resulted in a F2-score sensitivity of 89.5%, specificity of 57.8% and a positive predictive value of 59.9% (P=NS).

*Interpretation:* DL processing of dermoscopic images followed by sonification results in an accurate diagnostic output for SMP, implying that the quality of the dermoscope is not the major factor influencing DL diagnosis of skin cancer. Present system might assist all healthcare providers as a feasible computer-assisted detection system.

*Fund:* Bostel Technologies.

Trial Registration clinicaltrials.gov Identifier: NCT03362138



## 1. Introduction

About 1 million non-melanoma skin cancers and 288,000 malignant melanoma (MM) cancers occurred globally in 2018 [1]. Due to an aging population and limited health care resources, accurate diagnosis and feasibility of detection are a requisite for a generalized skin cancer prevention policy. The impact of immunotherapies on survival and cost further strain the already overburden healthcare system and raise the question of financial sustainability [2]. Skin cancer and especially MM early detection is challenging for both dermatologists and general practitioners. Dermoscopy is considered the standard of care [3], but in objective tests dermatologists achieve a limited diagnostic sensitivity of 40% MM detection [4] due to the complexity of visual inputs embedded in a dermoscopy image [5]. General practitioners seem to benefit from use of a dermoscopy course, while a figure of 51% correctly diagnosed lesions calls for further improvements [6]. Likewise, specificity of diagnosis by dermatologists calls for a further improvement, as reflected by a spectrum of 28:1 to 9:1 number of biopsies that need to be excised in order to identify one melanoma and a 3:1 ratio for overall skin cancer [7,8].

Deep learning (DL) classifiers are a promising candidate for detection of skin cancer [9,10]. Nonetheless, laboratory studies reported a clinical sensitivity from 29%–87% [ 11,12], a discrepancy which might be attributed to the quality of the dataset input, therefore rendering technology as experimental. Recently, a first prospective clinical observational study [13] reported on a two step approach, adding a second layer of sonification (visual data turned into sounds) to a DL classifier in order to improve accuracy of detection. This dual DL utilized an advanced dermoscope, a relatively expensive device, and a technique

* Corresponding author at: 6 Matmon Cohen Street, Tel Aviv 6209406, Israel.
*E-mail address:* dasc@tauex.tau.ac.il (A. Dascalu).

https://doi.org/10.1016/j.ebiom.2019.04.055

**Research in context**

*Evidence before this study*

Diagnosis of skin cancer aided by deep learning is pursued in last years. Previously, a clinical prospective study published on Feb 2019 EBioMedicine indicated that sonification (data conversion to sound) imposed on deep learning algorithms improves diagnostic outcome. The clinical study utilized image capture by an advanced dermoscopic device. We searched in Pubmed and arXiv for prospective clinical trials using the search terms of "artificial intelligence" or "deep learning" and "melanoma" or "skin cancer" with addition of the term "prospective". Search was conducted on Dec 15, 2017 and repeated on Mar 15, 2019 without any finding beyond the before mentioned article. There were no additional clinical studies applying deep learning skin diagnostics on either "sonification" or a "low-cost" dermoscope.

*Added value of this study*

To our knowledge, this is the first prospective observational study employing an elementary skin magnifier with polarized light to successfully test and validate skin cancer diagnosis by sonification. Data processed by a deep learning algorithm was sonified and output compared between a low-cost and advanced dermoscope. It is demonstrated that sonification of data renders diagnostic accuracy of a low resolution dermoscope on par with a professional dermoscope.

*Implications of all the available evidence*

Present technology improves accuracy output and might assist all physicians to diagnose skin cancer. System advantages are cost convenience of the device and bypass of dermoscopy-related experience factors, time constraints and physical inconvenience of acquiring images. The conclusion that quality of a dermoscope is not the major factor influencing deep learning diagnostics of skin cancer implicates different pattern recognition between humans and machine learning. The study utilized teledermatology and contributes to preventive medicine, potentially supporting present budget oriented healthcare systems as a computer-assisted diagnostics.

highly dependent on physician experience [4], rendering it less suitable for widespread primary care physicians use.

Consequently, the impact of image quality on accuracy of diagnosis was further examined. It was decided to test a low priced device classified by its manufacturer as a skin magnifier with polarized light (SMP). Images quality acquired by SMP preclude in most cases a precise clinical diagnosis due to haziness and lack of fine high level dermoscopy patterns and diagnostic structures. Images were processed by DL algorithms, sonified and diagnostic metrics were validated versus the histopathology report.

## 2. Methods

### 2.1. Primary deep learning training and sonification

As previously described [13] a convolutional neural network architecture based on the Inception V2 network was utilized. Dermoscopic images validated by biopsy reports were classified into either malignant or benign and a feature representation was obtained. Publicly-available datasets, such as the International Skin Imaging Collaboration (ISIC) 2017 dataset [14] were used for training to a total of 4361 advanced dermoscope images and 800 non-dermoscopic regular photos. Data augmentation, training and fine tuning were performed as mentioned [13] and the weighted activations of all of the 1024 nodes in the penultimate layer of the DL classifier were sonified, i.e. representation of data using non-speech [15] in order to generate sounds. A K-means clustering algorithm [16] was employed to cluster the activations into groups of related observations. The K-means algorithm was initialized by randomly choosing N data points without replacement to constitute the initial cluster centers, where N is the number of clusters. In order to address the sensitivity to initialization, K-means was run 100 times, each with a different random starting point. The clustering solution with the lowest error (i.e. the one that maximizes the likelihood of the data) was chosen as the final model. Cluster centroids represented by individual pitches and malignant "alert" sounds were mapped onto loudness, timbre, and duration of a sonification, thus an audio signal for each of the centroids of data was derived, providing for an audio output that acoustically differentiated the malignant from benign lesions and conferring information about the image through a raw wave file (Fig. 1).

### 2.2. Secondary deep learning algorithm

Raw sound files were derived for each dermoscopic image referred to biopsy ($n = 73$ patients) by SMP (133 data files) alongside an advanced dermoscope (AD, 142 data files) and analyzed by a secondary learning machine. For each image a sonification audio file was produced. In order to explore the diagnosis of skin cancer based on the audio waves, a particular classifier dedicated to each acquiring device, either SMP or AD, was trained against the ground truth diagnosis in the database, using a 80% random single split of the samples (training set). The remaining 20% of the set were held back and later used for validation (test set). The classifier normalized the input (zero-mean and divide by standard deviation), and dropout was used for regularization. For the raw audio classifier each raw WAV file is single-channel (mono) audio, produced via the sonification algorithm, with sample rate of 44,100 Hz and a duration of 3 s, for a total of 132,300 data points per file. By averaging each 10 consecutive samples, the input size was reduced to 13,230 values. A 1-dimensional CNN was used, with input size 22,050, first convolutional layer with 32 filters of size $1 \times 5$; max-pooling layer with size 10; second convolutional layer with 64 filters; max-pooling layer with size 10; a fully connected layer with 128 neurons; and output softmax layer with 2 neurons. Performance of SMP and AD classifiers were quantified by the area under the curve (AUC) of the receiver operating characteristic curve (ROC).

### 2.3. Clinical study

An open, prospective, non-interventional observational study was conducted at dermatology clinics at three sites by one dermatologist (DA).The clinical trial was approved by the institutional review board of Maccabi Healthcare, Israel (protocol Aq 16842/2017), clinicaltrials.gov Identifier: NCT03362138. Enrollment occurred between 18th Dec 2017 and 31th Aug 2018. Inclusion criteria were: (i) age 18 years and older, (ii) a suspected malignant lesion identified by a dermatologist through dermoscopy resulting in clinical management of referral to biopsy and (iii) patients' consent to participate in the study. Exclusion criteria were (i) a non-intact skin, since touch dermoscopy on an ulcerated skin cancer is unadvisable as it requires physical pressure in order to assess a flat diagnostic surface, might be painful or encounter an infected tissue, (ii) >15 hairs per dermoscopic field, (iii) performance of an unsolicited biopsy by surgeon (shave), and (iv) lesion location within 1 cm of the eye or mucosae surfaces. A total of 83 consecutive biopsy reports were received, 7 patients were illegible by inclusion criteria and 3

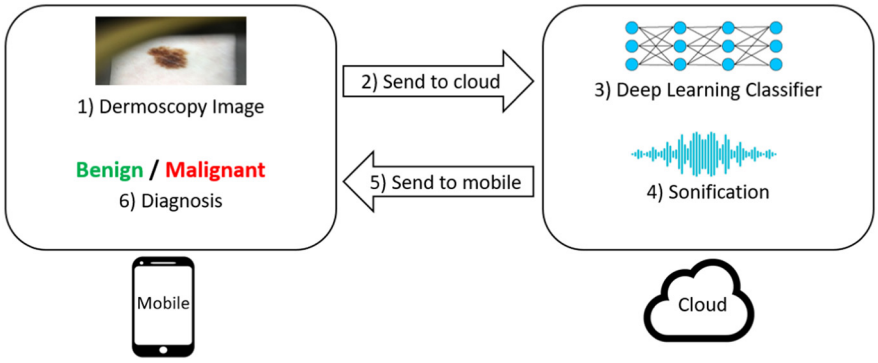


**Fig. 1.** Flowchart of image processing. A dermoscopy image is acquired by a smartphone and conveyed to cloud computing. A deep learning classifier predicts primary findings which are further processed by sonification. Final diagnosis is conferred to user as a malign or benign lesion diagnosis, i.e. excise or not indication.

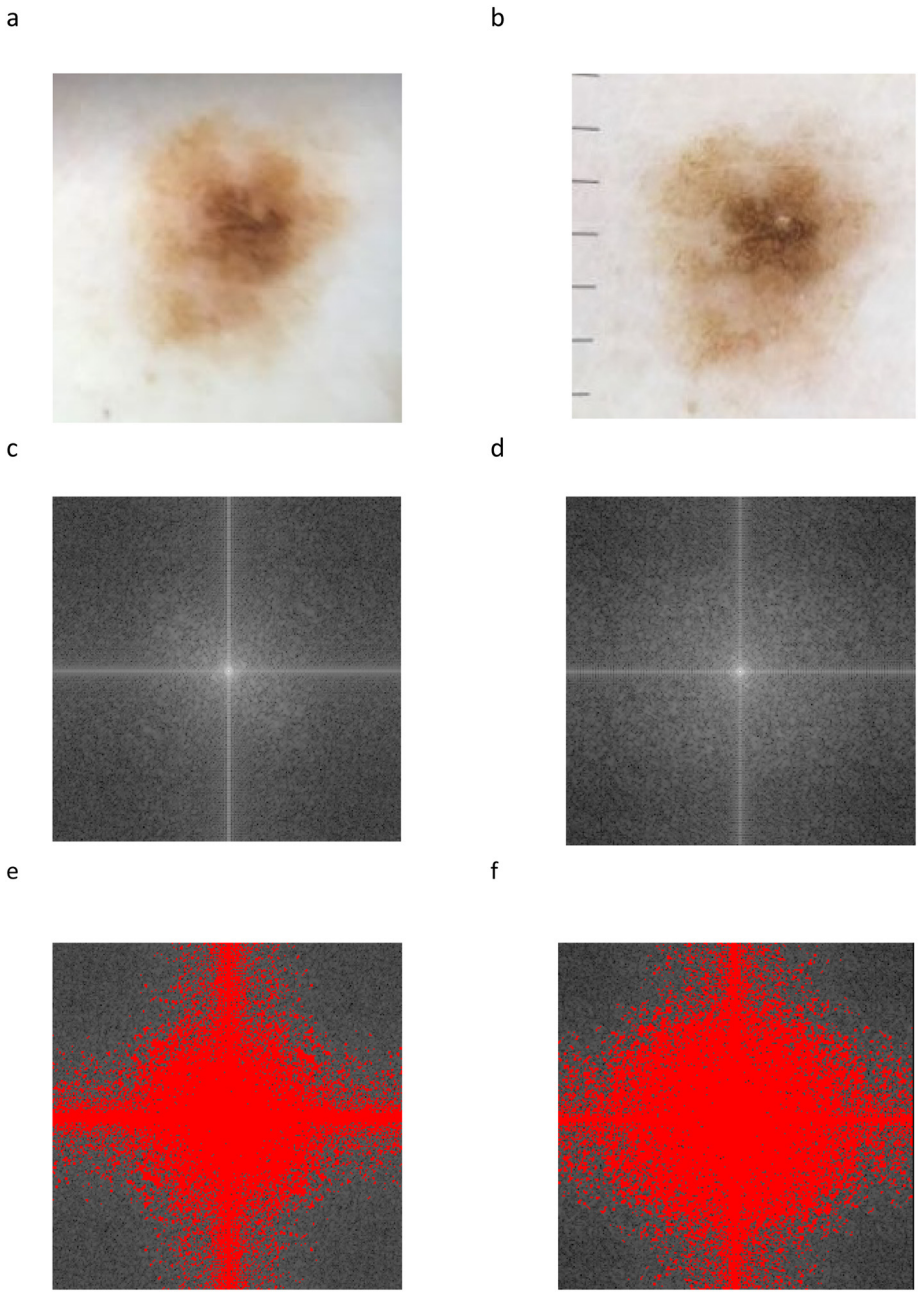


**Fig. 2.** An image from an elementary dermoscope (2a) was compared to the advanced dermoscope version (2b). A Fast Fourier Transformation was applied on images (2c, b), white areas were further thresholded (Fig. 2e, f) and area fraction was measured. The advanced dermoscope possesses higher spatial frequencies ($p > .001$) and therefore inputs a more detailed image.

lesions regressed prior to biopsy (likely keratoacanthomas) rendering 73 eligible subjects by inclusion criteria.

Patients were enrolled serially at each of the study sites on the basis of casual visits and need to biopsy. Subsequent to a clinical decision to biopsy, patient was first referred to a surgeon and then asked to participate in the study by signing the consent form. A rudimentary SMP (DermLite HUD, 3 Gen, TX, US) and an advanced dermoscope, AD (DL4, 3 Gen, TX, US) attached to a smartphone (iPhone 6) were used used through a purpose-built application (HopLabs, Atlanta, GA, US) for acquiring the same dermoscopic image of a suspected lesion by both devices, which were securely transmitted to a server (HopLabs, Atlanta, GA, USA) via a mobile network. Participant ID was transferred as consecutive numbers, without other patient details. Images were processed on the server by the DL algorithm and sonified. A clinical diagnosis by the first DL classifier, benign or malignant, accompanied by the raw sound appeared on the smartphone screen within 6–8 s from acquiring the dermoscopic image.

Differences of image quality between both study dermoscopes (Fig. 2a,b) can be quantified by applying a Fast Fourier Transformation (FTT) to an image (Fig. 2c, d) which provides a view of its spatial frequency components. The more distant the white areas are from the origin of an FFT, the higher the spatial frequency. On the contrary, the low frequency components are found near the central FFT image point. High spatial frequencies are responsible for conveying the content of complex real-world scenes in the scene-selective visual cortex [17]. These high spatial frequencies correspond to features such as sharp edges and fine details as opposed to the coarse estimate provided by the low spatial frequencies which are less important in object recognition, an analogy relevant to our image pattern recognition discrepancy in image quality. Therefore, an increased white area corresponds to more higher spatial frequencies and can be quantified as a geometric area.

Paired images of the same nevi ($n = 12$) were captured at an identical pixel content by both SMP and AD. Images were contrast enhanced and a FFT (NIH ImageJ, v 1.51j8) was applied. FFT images of both devices were further identically tresholded (Fig. 2e, f) and white areas were measured. All FFT images were of the same pixel size and therefore areas are expressed as area fractions. Comparison between dermoscopes demonstrate that AD displays an area fraction of 43.8% $\pm$ 4.1 (median 43.8%) as compared to SMP 35.2% $\pm$ 5.9 (median 35.7%), $p < .001$, student's *t*-test. It is concluded that spatial frequencies of devices diverge significantly, AD displaying higher spatial frequencies and therefore rendering a more detailed image.

### 2.4. Outcomes

Primary outcome of the sonification system was defined as detection of malignancies at a sensitivity of at least 85%, as validated by biopsy (Sensitivity is the percentage of correctly diagnosed malignancies, i.e., true positive/positive diagnoses). Sensitivity was chosen based on data from an objective laboratory study [4], which represents and simulates clinician accuracy in detecting melanoma. This study dermatologists diagnosed "easy to recognize" melanoma correctly in dermoscopies at a 72% $\pm$ 11 sensitivity, with sensitivity dropping to 34% for "intermediate" difficulties. Thus, we aligned our sonification sensitivity criteria to the more severe criteria of 75% + 1 SD endpoint. A second primary outcome was a specificity of at least 33% for sonification, as compared to biopsy (specificity is the percentage of correctly identified normal nevi, i.e., true negative/negative diagnoses). Specificity metric was derived from a clinical field test study results of an electrical impedance device at 34% [18].

### 2.5. Statistical analysis

Baseline and demographic characteristics were summarized by standard descriptive summaries. All statistical tests used in this study (SigmaPlot v10.0, Systat Software, SanJose, CA) were 2-sided and a $p$ value <.05 was considered significant. ROC curves were used to compare the DL results to ground truth biopsies. Sensitivity, the true positive rate, was plotted on the y-axis versus [1-Specificity], the false positive rate, on the x-axis of ROC curves. AUC for such a plot has a maximum value of 1.0, and is a standard performance metric in the machine learning literature. Accuracy of the ROC is defined as the fraction of correct predictions, i.e. true positives and true negatives divided by all true and false positives and negatives.

After measuring the sensitivity and specificity of the trained deep learning model in 21 different settings, we selected the best operating configuration using F-score, a statistical analysis measure for binary classification. F measure, a weighted harmonic mean of recall & precision was used in order to assign a differential clinical significance to sensitivity and positive predictive scores. F-score relies on recall (i.e., sensitivity) and precision (positive predictive value) to calculate a unified score. It is calculated as follows: $F(\beta) = (1+\beta^2)\cdot(\text{Precision}\cdot\text{Recall}/(\beta^2\text{Precision} + \text{Recall}))$, where $\beta$ is a factor denoting how important recall is in comparison to precision. Since in a clinical setting recall is more important than precision (i.e., we allocate more importance to not missing a malignant classification, at the cost of higher false positives), we selected a $\beta$ of 2 (typical values are 1/4, 1/2, 1, 2, 4) indicating that twice the weight is given to recall as opposed to precision, roughly paralleling our definition of primary outcomes.

A minimal clinical sample size of 22 patients for estimating sensitivity is required assuming a 0.40 proportion for clinician group (null hypothesis), a DL sensitivity of 0.85, a statistical power of 0.80 and alpha of 0.05 (Sigmaplot for Windows, V 10.0, Systat Software, San Jose, Ca, USA). Idem, assuming a 0.10 proportion for clinician group, a DL sensitivity of 0.33, a statistical power of 0.80 and alpha of 0.05 a sample size of 58 patients is required for specificity measurement.

## 3. Results

Patients suspected of skin cancer, either carcinoma or melanoma, performed 73 consecutive biopsies (Table 1). The classifier was trained to diagnose by DL and output was further sonified as either malign or benign, i.e. an excise or do not excise decision. Atypical nevi (a clinical diagnosis) mimic small melanomas and therefore dysplastic nevi (a histopathopathology diagnosis) were part of the biopsied lesions and excise criteria as MM simulators. Patients' biopsy reports indicated 2 melanomas, one adult atypical Spitz nevus and a majority of dysplastic nevi alongside 12 carcinomas.

**Table 1**
Clinical study data.

| Epidemiologic data and characteristics of lesions | |
|---|---|
| Characteristics | No. 73 |
| Study population | |
| Patients | 73 |
| Lesions | 73 |
| Total images | |
| Age, mean (range) | 50.3 $\pm$ 14.5 (18–87) |
| Sex | |
| Male | 38 |
| Female | 35 |
| Race | |
| Caucasian | 100% |
| Anatomic site | |
| Face | 12 |
| Trunk | 47 |
| Extremities | 14 |
| Histopathology diagnosis | |
| Benign nevus | 42 |
| Skin cancer | 31 |
| Dysplastic nevus | 16 |
| Atypical spitz nevus | 1 |
| Melanoma | 2 |
| Basal cell carcinoma | 6 |
| Squamous cell carcinoma | 6 |

The images of the SMP are blurred at a close up examination as compared to AD (Fig. 3a, b). Visualization of fine dermoscopic features, such as dots and intricate networks, is inadequate which renders a clinical diagnosis difficult.

Visual inspection of the raw sound files derived from SMP does not distinguish between benign, dysplastic nevus and MM (Fig. 3 c). Consequently, a secondary machine learning was applied to the raw sound files in order to diagnose malignancy. The SMP ROC curve AUC (Fig. 4) was 0.814 (95% CI, 0.798–0.831). Concomitant diagnosis of the same lesions by AD resulted in an AUC of 0.822 (95% CI, 0.849–0.794) (p = NS). It is concluded that diagnostic output through a sonification algorithm is not influenced by the quality of the dermoscope.

From a clinical perspective, the weights of recall (sensitivity) and precision (positive predictive value) of a skin cancer are unequal, since false negative diagnoses are more compulsory to be avoided. Therefore, in line with the primary goals ratio, a 2:1 ratio weight was assigned to recall versus precision. F2-score for both SMP and AD metrics were similar and non significant (Table 2). Specificity criteria, the second primary outcome, was achieved by both SMP and AD. Specificity of the advanced dermoscope seemed to be improved, but did not reach

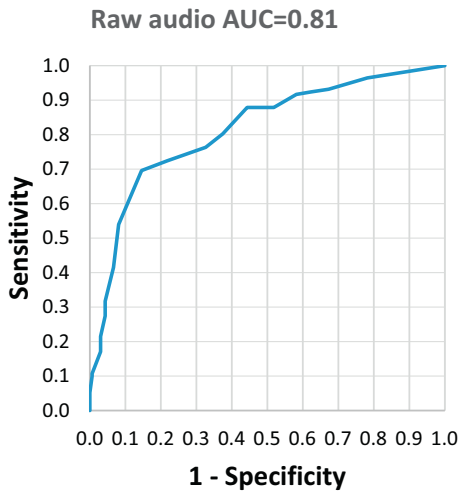


**Fig. 4.** ROC curve of raw sonified audio files derived from an elementary dermoscope.

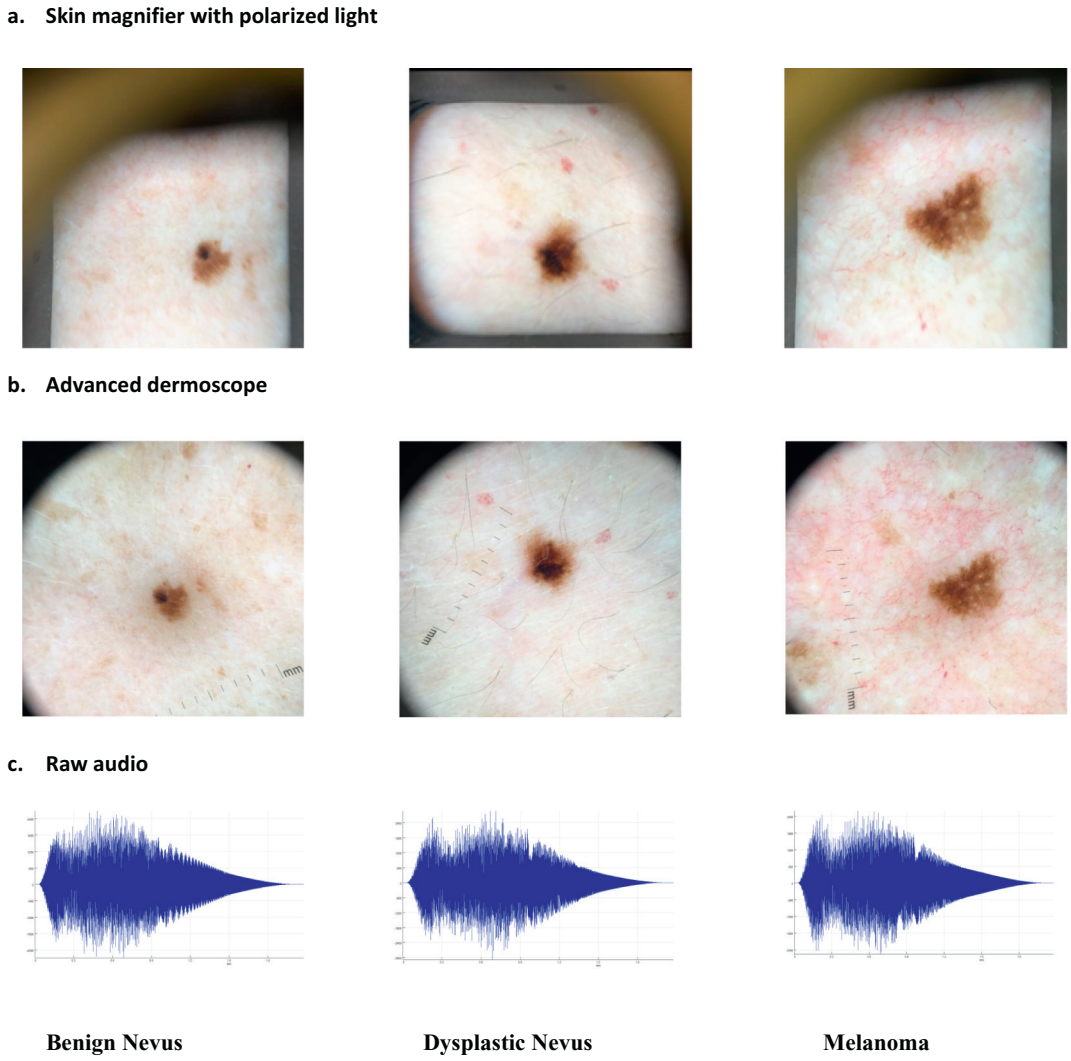


**Fig. 3.** Comparison of benign nevus (left), dysplastic nevus (middle) and melanoma (right). A. Skin magnifier with polarized light acquires hazy out of focus images, without a clear-cut pattern recognition of fine details required for diagnosis. B. Same lesions images acquired by an advanced dermoscope exhibit clear fine features and edges enabling diagnosis. C. Raw audio waves of a benign nevus, dysplastic nevus and melanoma, imperceptible to the human eye.

**Table 2**
Metrics of diagnostic analysis of elementary and advanced dermoscopes.

| Metrics | Skin magnifier with polarized light (95% CI), % | Advanced dermoscope (95% CI), % | p |
|---|---|---|---|
| Sensitivity (recall), TP/(TP + FN) | 91.7 (85.7–97.6) | 89.5 (82.5–96.4) | NS |
| Specificity | 41.8 (27.5–56.1) | 57.8 (44.0–71.6) | NS |
| Precision, TP/(TP + FP), positive predictive value | 57.3 (49.7–64.9) | 59.9 (51.4–68.4) | NS |
| F2-score (twice extra weight to recall over precision) | 81.8 (78.4–85.3) | 81.4 (77.2–85.7) | NS |

statistical significance ($p = .13$). Therefore, applying a F2-score clinical criteria render dermoscope quality without effect on diagnostic metrics.

## 4. Discussion

We report on skin cancer detection by an elementary dermoscopy device which diagnosis malignancy utilizing DL and sonification algorithms. Clinical outcome, as validated by histopathology reports, results in a sound ROC AUC of 0.81. Applying a twice extra weight to sensitivity upon positive predictive value derives a 92% sensitivity and a 42% specificity. Remarkably, diagnostic accuracy of the elementary dermoscope is comparable to a professional dermoscopy device. Bearing in mind that (i) The primary DL which outputs sonification was trained mostly on highly detailed dermoscopy images, unlike this study elementary dermoscope and that (ii) our elementary dermoscope quality output does not allow a reliable diagnosis by a clinician, the results evidentiate the difference of perception of visual images between humans and convoluted neural networks.

Prospective clinical studies of skin cancer detection by DL are scarce. Publications up to date are laboratory initiated and do not account for diagnosis of dysplastic nevi, an obligatory biopsy MM mimicker, therefore selectively concentrating on coarse MM features. Dysplastic nevi, a histopathology diagnosis, are part of the atypical nevi spectrum which they mimic and are diagnosed a posteriori. It is recommended to excise only severe dysplastic nevi, and by some moderate dysplastic nevi as well. However, (i) there is no a priori technology which can identify whether a suspicious lesion is mild, moderate or severely dysplastic and (ii) even pathologists are at dispute whether a nevus belongs to the spectrum of moderate, severe or melanoma in situ, requesting a 3 panel experts for a final decision of about 15% of biopsies [19]. Furthermore, there is a 2% yield of melanoma of incompletely excised moderate dysplastic nevi at 5 years of follow up [20], an inappropriate burden for any healthcare system. Accordingly, our system does not seek to diagnose which kind of nevus the atypical suspect belongs to, but to recommend whether to diagnose it as possible malign and recommend to excise it or not. These study criteria of excision are more sensitive than studies which diagnose exclusively MM, rendering a comparison as impracticable. It is assumed that since sonification was able to detect a dysplastic nevus, remarkable for a small size and high level dermoscopic features, a melanoma diagnosis should be at least as feasible with its more prominent morphological characteristics. Indeed, histopathology reports identified 13 out of 16 pigmented lesions excised as dysplastic nevi.

A biopsy validated study comparing between extreme poles of dermoscopic devices, i.e. a SMP and AD, is unavailable. Evaluation of our different visualization technologies might roughly parallel the comparison of a polarized noncontact dermoscope to a nonpolarized contact dermoscope which results in a complementary output, i.e. a strong agreement on pattern recognition, a mixed agreement, from weak to strong, on structural elements and a weak agreement on colors [21]. Contrarily, our study demonstrates a similar diagnostic output for different visual clues provided by SMP and AD devices. It might be hypothesized that DL diagnostics are more dependent on patterns and color than on structural elements, alike blotches and blue-white veils. Recently, DL was compared with humans for object recognition by applying a weaker visual signal such as a reduced contrast and additive noise [22], a representation parallel to the downgrade in optical signal of SMP as compared to AD. The authors concluded that the human visual system is more robust to image manipulations than DL by applying weaker signals. On the contrary, SMP of present study, a relatively weak signal for human diagnostics, performed alike DL, a crisp visual signal. It might be speculated that at a higher level and unlike human or even DL processing, sonification and kmeans algorithms diagnose less by identifying similarities to previous images and structural elements and more by other means such as color recognition or even avoiding false patterns. Sensitivity of DL is notable bearing in mind a limited 5000 images dataset our classifier was trained on.

Limitations to the study might include the assertion that SMP image quality is on a diagnostic par with an advanced dermoscope. The likelihood that SMP does procure a high quality image, which renders it close to the AD, is not supported by the raw quality of the images which are blurred and present fuzzy edges on high magnification, rendering diagnosis difficult. The study may encounter a random error due to a relatively limited MM number of patients studied, but as referred, missing an MM is unlikely due to a high sensitivity of the system which includes fine features dysplastic nevi. A selection bias, due to involvement of a single dermatologist in clinical assessments might exist. Initial planning which includes consecutive recruitment of patients, diagnosis at three different sites and patients' inclusion into study exclusively after a signed decision to biopsy lessen the impact of, to be further validated by multicenter studies. The study is exposed to a confounding bias due to a 15% disagreement between pathologists, and future studies will validate its impact by applying a decision expert panel. Since all missing data are limited at 10/83 (12%) of patients, it does not seem to impact on the study results. As a first report of a computer-assisted diagnostics system which improves diagnostic accuracy of a low quality dermoscope, our clinical datasets are limited and ensuing studies with a higher patient number should handle realistic scenarios of technology deployment.

In conclusion, a computer-assisted diagnostic rudimentary dermoscope was tested whilst superimposing a sonification technique on a convoluted neural network diagnostics. Technology improves accuracy of skin cancer diagnosis and might assist physicians to diagnose skin cancer and bypass dermoscopy-related experience factors, time constraints, physical inconvenience of acquiring images and overcome or, on the contrary, assist to a physician “God complex”. The study utilized teledermatology and contributes to preventive medicine, potentially supporting present budget oriented healthcare systems. Future studies will include more substantial patient numbers and compare human to sonification algorithms performance in order to verify and upscale conclusions of this study.

## Contributors

ED and AD conceived and designed the study. All authors take responsibility for the integrity of the data and the accuracy of the data

analysis. ED developed the algorithms. AD and ED were responsible for study supervision. AD and ED obtained and contributed to study interpretation and statistical analysis. All authors subsequently critically edited and revised the report. All authors read and approved the final report. The corresponding author had full access to all the data and final responsibility to submit for publication.

## Declaration of interests

AD is an inventor of a patent for the system used in this study; ED reported holding patents on deep Learning, unrelated to the deep learning system in this paper. AD is a shareholder at Bostel LLC. No other disclosures were reported.

## Acknowledgments

The study was performed within the framework of Maccabi Healthcare Services, Il. Authors thank Professor Bruce N Walker, Professor James M Rehg, Ankur Kalra, Ph.D., Michael R Winters, Ph.D. and Paul Drews, Ph.D. for their input to infrastructure development.

## Role of the funding source

The funding sources had no involvement in the study design; collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

The prospective observational study was approved by the institutional review board of Maccabi Healthcare, Israel (protocol Aq 16,842/2017), clinicaltrials.gov Identifier: NCT03362138

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.04.055.

## References

[1] Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer 2019 Apr 15;144(8):1941–53.
[2] Kandel M, Allayous C, Dalle S, et al. Update of survival and cost of metastatic melanoma with new drugs: Estimations from the MelBase cohort. Eur J Cancer 2018 Dec;105:33–40 Internet.
[3] Dinnes J, Deeks JJ, Chuchu N, et al. Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults. Cochrane Database Syst Rev 2018 Dec 4;12:CD011902.
[4] Tschandl P, Hofmann L, Fink C, Kittler H, Haenssle HA. Melanomas vs. nevi in high-risk patients under long-term monitoring with digital dermatoscopy: do melanomas and nevi already differ at baseline? J Eur Acad Dermatology Venereol 2017 Jun 1;31(6):972–7.
[5] Annessi G, Bono R, Sampogna F, Faraggiana T, Abeni D. Sensitivity, specificity, and diagnostic accuracy of three dermoscopic algorithmic methods in the diagnosis of doubtful melanocytic lesions: the importance of light brown structureless areas in differentiating atypical melanocytic nevi from thin melanomas. J Am Acad Dermatol 2007 May 1;56(5):759–67.
[6] Koelink CJL, Vermeulen KM, Kollen BJ, et al. Diagnostic accuracy and cost-effectiveness of dermoscopy in primary care: a cluster randomized clinical trial. J Eur Acad Dermatology Venereol 2014 Nov 1;28(11):1442–9.
[7] Matsumoto M, Secrest A, Anderson A, et al. Estimating the cost of skin cancer detection by dermatology providers in a large health care system. J Am Acad Dermatol 2018 Apr 1;78(4):701–709.e1.
[8] Argenziano G, Cerroni L, Zalaudek I, et al. Accuracy in melanoma detection: a 10-year multicenter survey. J Am Acad Dermatol 2012 Jul;67(1):54–9.
[9] Celebi ME, Codella N, Halpern A. Dermoscopy image analysis: overview and future directions. IEEE J Biomed Heal Informatics 2019 Mar;23(2):474–8.
[10] Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. Cham: Springer; 2015; 118–26.
[11] Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA. Automated dermatological diagnosis: hype or reality? J Invest Dermatol 2018 Oct 1;138(10):2277–9.
[12] Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018;29(8):1836–42.
[13] Walker BN, Rehg JM, Kalra A, et al. Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: laboratory and prospective observational studies. EBioMedicine Feb 2019;40:176–83.
[14] 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). http://arxiv.org/abs/1710.05006; 2017 Oct 13.
[15] Walker BN, Nees MA. Theory of sonification. In: Hermann T, Hunt A, Neuhoff J, editors. The sonification handbook. Berlin, Germany: Logos Publishing House; 2011. p. 9–39 [ISBN 978-3-8325-2819-5].
[16] Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 2013 Jan;40(1):200–10.
[17] Berman D, Golomb JD, Walther DB. Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. PLoS One 2017 Dec 22;12(12):e0189828.
[18] Malvehy J, Hauschild A, Curiel-Lewandrowski C, et al. Clinical performance of the Nevisense system in cutaneous melanoma detection: an international, multicentre, prospective and blinded clinical trial on efficacy and safety. Br J Dermatol 2014 Nov;171(5):1099–107.
[19] Elmore JG, Barnhill RL, Elder DE, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study. BMJ 2017;357:j2813 Jun 28.
[20] Fleming NH, Egbert BM, Kim J, Swetter SM. Reexamining the threshold for reexcision of histologically transected dysplastic nevi. JAMA Dermatol 2016;152(12):1327–34.
[21] Benvenuto-Andrade C, Dusza SW, Agero ALC, et al. Differences between polarized light dermoscopy and immersion contact dermoscopy for the evaluation of skin lesions. Arch Dermatol 2007 Mar 1;143(3):329–38.
[22] Geirhos R, Janssen DHJ, Schütt HH, Rauber J, Bethge M, Wichmann FA. Comparing deep neural networks against humans: object recognition when the signal gets weaker*. arXiv 11 Dec 2018;1706 06969v2 [cs.CV].