The dermoscopic inverse approach significantly improves the accuracy of human readers for lentigo maligna diagnosis

Aimilios Lallas, MD, PhD,[a] Konstantinos Lallas, MD,[a] Philipp Tschandl, MD, PhD,[b] Harald Kittler, MD, PhD,[b] Zoe Apalla, MD, PhD,[c] Caterina Longo, MD, PhD,[d,e] and Giuseppe Argenziano, MD, PhD[f]
Thessaloniki, Greece; Vienna, Austria; and Reggio Emilia, Modena, and Naples, Italy

Background: A recently introduced dermoscopic method for the diagnosis of early lentigo maligna (LM) is based on the absence of prevalent patterns of pigmented actinic keratosis and solar lentigo/flat seborrheic keratosis. We term this the *inverse approach*.

Objective: To determine whether training on the inverse approach increases the diagnostic accuracy of readers compared to classic pattern analysis.

Methods: We used clinical and dermoscopic images of histopathologically diagnosed LMs, pigmented actinic keratoses, and solar lentigo/flat seborrheic keratoses. Participants in a dermoscopy masterclass classified the lesions at baseline and after training on pattern analysis and the inverse approach. We compared their diagnostic performance among the 3 timepoints and to that of a trained convolutional neural network.

Results: The mean sensitivity for LM without training was 51.5%; after training on pattern analysis, it increased to 56.7%; and after learning the inverse approach, it increased to 83.6%. The mean proportions of correct answers at the 3 timepoints were 62.1%, 65.5, and 78.5%. The percentages of readers outperforming the convolutional neural network were 6.4%, 15.4%, and 53.9%, respectively.

Limitations: The experimental setting and the inclusion of histopathologically diagnosed lesions only.

Conclusions: The inverse approach, added to the classic pattern analysis, significantly improves the sensitivity of human readers for early LM diagnosis. (J Am Acad Dermatol 2021;84:381-9.)

Key words: artificial intelligence; dermatoscopy; dermoscopy; diagnosis; inverse approach; maligna; melanoma; pigmented actinic keratosis; solar lentigo.

Lentigo maligna (LM) is difficult to recognize because it shares similar epidemiologic, pathogenetic, and morphologic characteristics with pigmented actinic keratosis (PAK) and solar lentigo/flat seborrheic keratosis (SL/SK).[1-4]

Dermoscopy allows the recognition of melanomas lacking macroscopic criteria.[5] However, discriminating early LM from PAK and SL/SK remains challenging even dermoscopically, although the dermoscopic morphology of LM has been extensively investigated.[3,4,6-8] The initial dermoscopic

From the First Department of Dermatology, Aristotle University, Thessaloniki[a]; Department of Dermatology, Medical University of Vienna[b]; Second Department of Dermatology, Aristotle University, Thessaloniki[c]; Centro Oncologico ad Alta Tecnologia Diagnostica, Azienda Unità Sanitaria Locale—Istituto di Ricovero e Cura a Carattere Scientifico di Reggio Emilia[d]; Department of Dermatology, University of Modena and Reggio Emilia, Modena[e]; and Dermatology Unit, University of Campania Luigi Vanvitelli, Naples.[f]
Funding sources: None.
Conflicts of interest: None disclosed.
IRB approval status: Not applicable.
Accepted for publication June 19, 2020.
Reprints not available from the authors.
Correspondence to: Aimilios Lallas, MD, PhD, First Department of Dermatology, School of Medicine, Aristotle University, 124 Delfon St, 54623, Thessaloniki, Greece. E-mail: emlallas@gmail.com.
Published online June 24, 2020.
0190-9622/$36.00

https://doi.org/10.1016/j.jaad.2020.06.085

features of LM appear on the outline of the follicular openings: they are gray and include dots, short lines, semicircles, and circles.[3,6,9-11] These features are very subtle, only focally present, and, thus, difficult to recognize. Additionally, similar structures typify PAK and SL/SK, especially when regression occurs (lichen planus–like keratosis).[3,12-14] This was highlighted in studies reporting that dermoscopic features with high sensitivity for LM had low specificity.[4,8,15-17]

CAPSULE SUMMARY

- The inverse approach dermoscopic method for the diagnosis of early lentigo maligna (LM) uses the absence of prevalent patterns of pigmented actinic keratosis and solar lentigo/flat seborrheic keratosis.
- Training in this method improved sensitivity for the diagnosis of LM from 51.5% to 83.6%, making it a useful tool.

To address this problem, we introduced a new method suggesting that the diagnosis of early LM should not be based on the presence of LM-specific features but on the absence of prevalent nonmelanoma patterns. We now term this method the *inverse approach*. The sensitivity of the method in the initial study was 88.5, and the specificity was 66.9%.[18] The method is based on the notion that PAK and SL/SK can be dermoscopically diagnosed by the detection of at least 1 of the features shown in Table I, provided that the feature is predominant, meaning that it occupies more than half of the lesion's surface. If these features are absent or present only in small areas, then this is enough to consider the lesion as suspicious for LM, even if none of the features known to typify LM can be seen (see Supplementary Material available via Mendeley at https://data.mendeley.com/datasets/htdcngcvr9/1#file-6a1ee9d5-3248-4198-8408-a2d2527dee9b).

The primary aim of the present study was to investigate whether the inverse approach increases the diagnostic accuracy of human readers as compared to the classic pattern analysis. In addition to the classic accuracy measures, we also used a convolutional neural network (CNN) as an objective reference with reasonable accuracy.

METHODS

This diagnostic study was held during a 3-day dermoscopy masterclass, using a data set of facial pigmented macules, histopathologically diagnosed as LM, PAK, or SL/SK. The participants were asked to classify the lesions at 3 different timepoints, using a voting system with manual devices. One screen was placed in front of every 3 participants, and a video wall projection was visible to all of them. The study was conducted by using appropriately anonymized data sets and, therefore, ethics committee approval was waived.

Participants

All 78 participants of the masterclass were invited and agreed to participate in the study. Of them, there were 45 board-certified dermatologists, 30 residents in dermatology, 2 general practitioners, and 1 nurse. Before the masterclass, they evaluated clinical and dermoscopic images of 66 randomly selected skin tumors (not facial), which included all common diagnoses (nevus, seborrheic keratosis, angioma, dermatofibroma, melanoma, and basal and squamous cell carcinoma). Based on the number of correct answers, participants were classified in 3 groups (higher, intermediate, and lower skills in dermoscopic diagnosis) by a simple split into thirds. We selected this method to assess the baseline skills of participants because we consider it more objective than the commonly used method of asking how many years of experience each participant reports having.

Test set

A sample size calculation was conducted to validate the number of lesions to include in the test set. The number of readers was fixed at 78, and with an aim of increasing sensitivity by 30%, with an alpha level at 0.05 and power of 80%, a number of 56 lesions was calculated as adequate. The database of the Department of Dermatology at the University of Campania was screened for eligible cases. Facial pigmented macules with a definite histopathologic diagnosis of LM, SL/SK, or PAK were eligible. The search identified 407 eligible lesions, excised or biopsied between July 2009 and July 2018, including 162 LMs, 169 SL/SKs, and 76 PAKs. Of them, 60 lesions were selected with the use of random numbers. The final test set consisted of clinical and dermoscopic images of 23 LMs, 26 SL/SKs, and 11 PAKs. All images were captured with a camera with a dermoscopic lens (polarized light at 10-fold magnification).

Procedures

Participants evaluated the 60 images at 3 timepoints. The first evaluation was performed on day 1, before teaching activities. On day 2, the

Abbreviations used:

AI: artificial intelligence
CNN: convolutional neural network
LM: lentigo maligna
PAK: pigmented actinic keratosis
SL/SK: solar lentigo/flat seborrheic keratosis

participants attended a 45-minutes lecture on dermoscopy of LMs, PAKs, and SL/SKs, based on pattern analysis and including an analytic description of global patterns and local features with several examples. At the end of day 2, approximately 30 hours after the first evaluation, the second evaluation was performed. On day 3, participants attended a 30-minute lecture describing the inverse approach, with numerous examples. At the end of day 3, approximately 30 hours after the second evaluation, the third evaluation was performed. The sequence of the 60 images in the 3 evaluations was randomly altered. The readers were informed about the correct answers only after the last evaluation, and none of the images of the test set were included in the training lectures.

Outcomes and measures

The primary outcome was the sensitivity and specificity for LM without training and after training on pattern analysis and on the inverse approach. Secondary outcomes were the number of correct answers at the 3 timepoints, the sensitivity and specificity for PAK and SL/SK, and the comparison of the performance of participants to that of a previously trained CNN.

CNN

To estimate readers' accuracy to an objective reference, we compared it to automated image analysis via a reasonably accurate single CNN. Using the pytorch framework,[19] we fine-tuned an ImageNet[20]-pretrained ResNet34[21] architecture for dermoscopic classification of 7 pigmented skin tumors on the publicly available HAM10000 data set.[22,23] The resulting CNN takes a single dermoscopic image as an input, produces a probability value (range, 0-1) for every diagnosis, and the one with the highest value is used as the final prediction. We measured a mean recall value of 77.7% for the resulting trained CNN on the official International Skin Imaging Collaboration-2018 test set, with more correct answers than the average human rater of a recent study.[24] Because participants in this study were given 3 possible answers, probabilities for all other classes were set to 0 before the softmax layer during inference. The class with the highest probability value was taken as the final CNN prediction.

Table I. The dermoscopic inverse approach for diagnosis of lentigo maligna*

Nonmelanoma feature	Suggested diagnosis
Scales	Actinic keratosis
White and wide follicular openings	Actinic keratosis
Erythema	Actinic keratosis
Reticular or parallel brown lines	Solar lentigo/seborrheic keratosis
Sharply demarcated border	Solar lentigo/seborrheic keratosis
Milia-like cysts/ comedo-like openings	Solar lentigo/seborrheic keratosis

*The presence of 1 (or more) of these structures as a predominant feature in the lesion is highly suggestive of actinic keratosis or seborrheic keratosis, as shown. If none of these criteria can be seen as a predominant feature, then the lesion is assessed as suspicious for lentigo maligna.

Statistical analysis

A descriptive analysis was performed to calculate the sensitivity and specificity of each reader, the most voted answer, and the correct answers per day. The mean sensitivity and specificity were calculated with 95% confidence intervals. The McNemar test was used to compare the correct answers among evaluations. Considering sensitivity as a continuous variable, we used paired *t* tests to compare the sensitivity per day and per group in the subgroup analysis. Finally, we conducted scatterplots for sensitivity, specificity, and correct answers per day. All statistical tests were 2 sided, and the level of significance was a *P* value of less than .05. The analysis was made with SPSS Statistics, version 25.0 (IBM, Armonk, NY) and GraphPad Prism, version 8.0.0, (GraphPad Software, San Diego, CA).

RESULTS

The male-to-female ratio of the 78 participants was 1:3.4, and the mean age was 41.1 years, ranging from 24 to 74 years. The main results are shown in Table II.

Diagnostic accuracy of readers for LM

The mean sensitivity of readers for LM diagnosis without training was 51.5%. After training on pattern analysis, the mean sensitivity increased to 56.7%. After learning the inverse approach, it further increased to 83.6% (Fig 1, *A*). All differences in mean sensitivity among the 3 evaluations were

Table II. Results of the evaluation of 78 readers

Measure	No training (day 1)	Pattern analysis (day 2)	Difference (day 2 - day 1)	Inverse approach (day 3)	Difference (day 3 - day 2)
All readers					
Diagnostic accuracy for LM					
Mean sensitivity, % (95% CI)	51.5 (48.3 to 55.0)	56.7 (53.5 to 59.8)	5.2 95% CI: 1.6 to 8.5 $P = .005$	83.6 (80.7 to 86.4)	26.9 95% CI: 24.4 to 29.4 $P < .001$
Mean specificity, % (95% CI)	85.2 (83.4 to 87.0)	86.1 (84.5 to 87.7)	0.9 95% CI: −1.2 to 3.0 $P = .39$	85.6 (83.9 to 87.1)	−0.5 95% CI: −2.5 to 1.4 $P = .56$
Mean number of correct answers (%)	37.3 (62.1)	39.3 (65.5)	2.0 (3.4) 95% CI: 0.7 to 3.4 $P = .003$	47.1 (78.5)	7.81 (13.0) 95% CI: 6.4 to 9.2 $P < .001$
Most voted answer					
Diagnostic accuracy of the most voted answer, % (95% CI)					
Sensitivity for LM	69.6 (49.1 to 84.4)	69.6 (49.1 to 84.4)	0	100 (85.6 to 100)	30.4
Specificity for LM	83.8 (68.8 to 92.3)	89.2 (75.2 to 95.7)	5.4	89.2 (75.2 to 95.7)	0
Accuracy for LM	78.3 (66.4 to 86.9)	81.7 (70.1 to 89.4)	3.4	93.3 (84.1 to 97.4)	11.6
Sensitivity for PAK	50.0 (21.1 to 78.9)	66.7 (34.9 to 90.1)	16.7	66.7 (34.9 to 90.1)	0
Specificity for PAK	95.8 (85.8 to 99.5)	97.9 (88.9 to 100.0)	2.1	100.0 (86.3 to 100.0)	2.1
Accuracy for PAK	86.7 (75.4 to 94.1)	91.7 (81.6 to 97.2)	5.0	93.3 (83.8 to 98.2)	2.6
Sensitivity for SL/SK	92.3 (74.8 to 99.1)	92.3 (74.8 to 99.1)	0	88.5 (69.9 to 97.6)	−3.8
Specificity for SL/SK	82.4 (65.5 to 93.2)	79.4 (62.1 to 91.3)	−3.0	94.1 (80.3 to 99.3)	14.7
Accuracy for SL/SK	86.7 (75.4 to 94.1)	85.0 (73.4 to 93.0)	−1.7	93.3 (83.8 to 98.2)	8.3
Correct most voted answers, n (%)	47 (78.3)	49 (81.7)	2 (3.4) 95% CI: −4.9 to 11.5 $P = .41$	56 (93.3)	7 (11.6) 95% CI: 3.5 to 26.5 $P = .01$
Readers by group, % (95% CI)					
High skills in dermoscopy					
Mean sensitivity for LM	58.3 (52.3 to 64.3)	64.4 (59.2 to 69.2)	6.1 95% CI: −0.4 to 12.7 $P = .06$	89.1 (84.3 to 93.9)	24.7 95% CI: 20.1 to 29.2 $P < .001$
Mean specificity for LM	87.3 (84.7 to 89.8)	88.2 (85.6 to 90.8)	0.9 95% CI: −2.1 to 4.0 $P = .53$	88.6 (86.1 to 91.0)	0.4 95% CI: −2.4 to 3.2 $P = .79$
Intermediate skills in dermoscopy					
Mean sensitivity for LM	49.9 (44.9 to 54.9)	56.5 (52.6 to 60.3)	6.6 95% CI: 1.1 to 12.0 $P = .02$	82.2 (77.4 to 86.9)	25.7 95% CI: 21.3 to 29.5 $P < .001$

Mean specificity for LM	83.8 (80.8 to 86.8)	84.7 (82.8 to 87.2)	0.9 95% CI: −3.0 to 4.9 *P* = .63	83.5 (81.0 to 86.0)	−1.2 95% CI: −4.5 to 2.0 *P* = .42
Low skills in dermoscopy					
Mean sensitivity for LM	45.5 (39.4 to 51.6)	48.3 (42.0 to 54.5)	2.8 95% CI: −3.9 to 9.5 *P* = .40	79.1 (73.7 to 84.5)	30.8 95% CI: 26.2 to 35.4 *P* < .001
Mean specificity for LM	84.6 (80.4 to 88.8)	85.6 (82.2 to 89.0)	1.0 95% CI: −3.5 to 5.5 *P* = .66	84.4 (81.0 to 87.9)	−1.1 95% CI: −6.0 to 3.7 *P* = .62

CI, Confidence interval; *LM*, lentigo maligna; *PAK*, pigmented actinic keratosis; *SL/SK*, solar lentigo/flat seborrheic keratosis.

statistically significant (Table II). The mean specificity was approximately 85% in all evaluations.

Correct answers of readers

The mean number of correct answers without training was 37.3 of 60 (62.1%). After training on pattern analysis, the mean number of correct answers increased to 39.3 (65.5%; $P = .003$). After training on the inverse approach, the mean number of correct answers increased to 47.1 (78.5%; $P < .001$). The correct answers of every reader per day are shown in Fig 1, *B*.

Diagnostic accuracy of the most voted answer

Based on the most voted answer per lesion, the sensitivity for LM without training was 69.6%, the specificity was 83.8%, and the accuracy was 78.3%. After training on pattern analysis, the sensitivity remained equal to 69.6%, whereas the specificity increased to 89.2% and the accuracy to 81.7%. After learning the inverse approach, the sensitivity for LM increased to 100%, the specificity remained equal to 89.2%, and the accuracy increased to 88.3%.

Correct answers based on the most voted answer

Without training, the most voted answer was correct in 47 of 60 lesions (78.3%). After training on pattern analysis, the number of correct answers increased to 49 (81.7%), but this increase was not significant. After training on the inverse approach, the most voted answer was correct in 56 lesions (93.3%), which was a statistically significant improvement (Table II).

Subgroup analysis by diagnostic skills

Of 78 participants, 27 scored more than 58 of 66 correct answers in the pretest on nonfacial lesions and were classified as highly skilled in dermoscopic diagnosis, 28 scored between 54 and 58 correct answers and were classified as intermediately skilled, and 23 scored fewer than 54 correct answers and were classified as less skilled. The mean sensitivity and specificity for LM diagnosis of readers of the 3 groups are shown in Table II. Highly skilled readers scored better at all timepoints, followed by the intermediately and less skilled. The sensitivity for LM slightly improved after training on pattern analysis, but the difference was significant only in the intermediate group. In contrast, the sensitivity significantly increased in all 3 groups after participants learned the inverse approach.

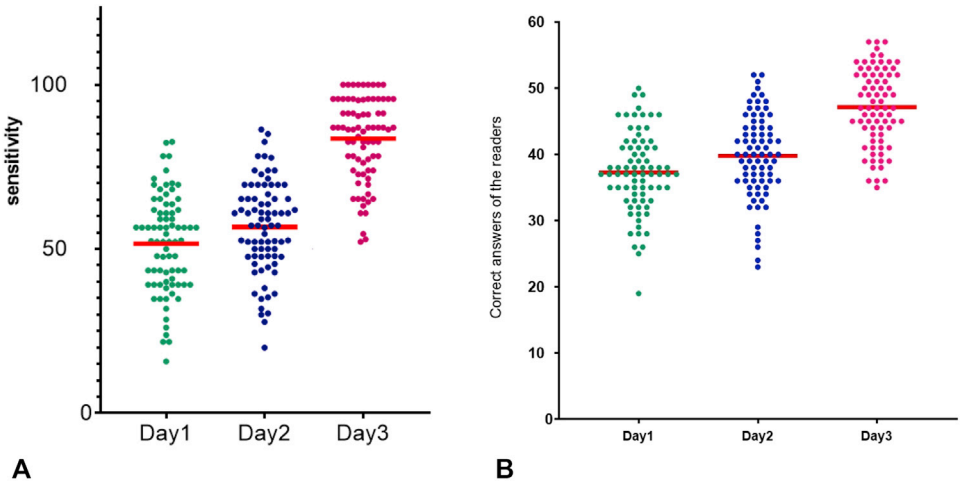

Fig 1. Diagnostic performance of 78 human readers. **A**, Scatterplot showing the sensitivity of the readers for lentigo maligna on days 1, 2, and 3. The red lines depict the mean sensitivity. **B**, Scatterplot of correct answers of the readers per day. The mean number of correct answers corresponds to the lines inside each box.

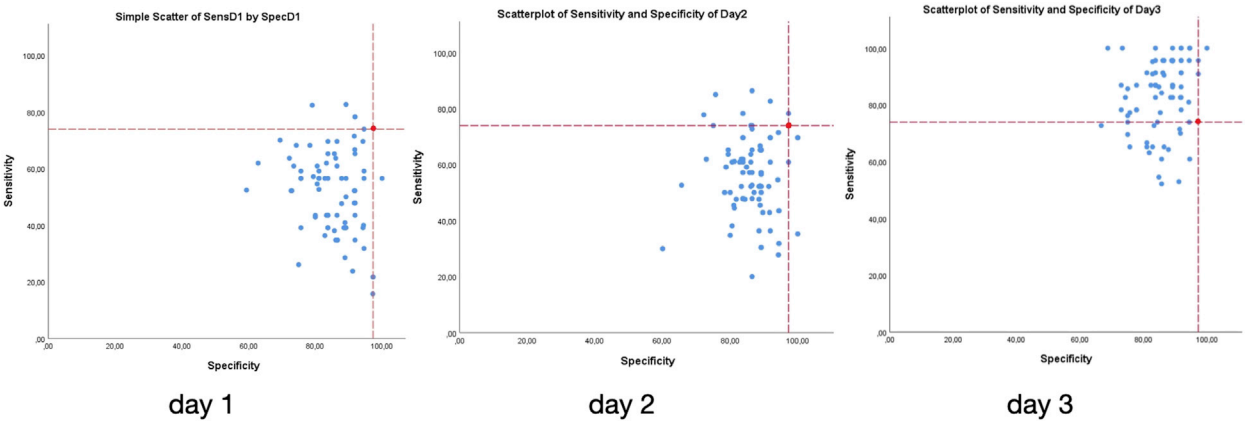

Fig 2. Diagnostic performance of 78 humans readers and a convolutional neural network. Scatterplots of sensitivity and specificity of readers for the diagnosis lentigo maligna on days 1, 2, and 3. The red dot indicates the sensitivity and specificity of the convolutional neural network. The horizontal red reference line highlights the readers who supersede the sensitivity of the convolutional neural network in each evaluation. *SensD1*, Sensitivity on day 1; *SpecD1*, specificity on day 1.

Performance of the CNN and comparison to human raters

The sensitivity of the CNN for LM diagnosis was 73.9%, the specificity was 97.3%, and the number of correct specific answers was 46 of 60 (76.7%). Without training, 6.4% of raters performed better than, 7.7% equal to, and 85.9% worse than the CNN. After receiving training on pattern analysis, 15.4% of raters performed better than, 5.1% equal to, and 79.5% worse than the CNN. After learning the inverse approach, 53.9% of raters were superior, 5.1% were equal, and 41.0% were inferior to the CNN (Fig 2).

DISCUSSION

Our study shows that the inverse approach significantly improves the ability of clinicians to accurately classify flat pigmented facial lesions. The improvement is most pronounced in the sensitivity for LM, which is the most relevant diagnostic measure from an outcome perspective. Although our study was not conducted in a clinical setting, the remarkable improvement in all diagnostic measures strongly suggests that the application of the inverse approach could significantly facilitate the clinical recognition of early LM.

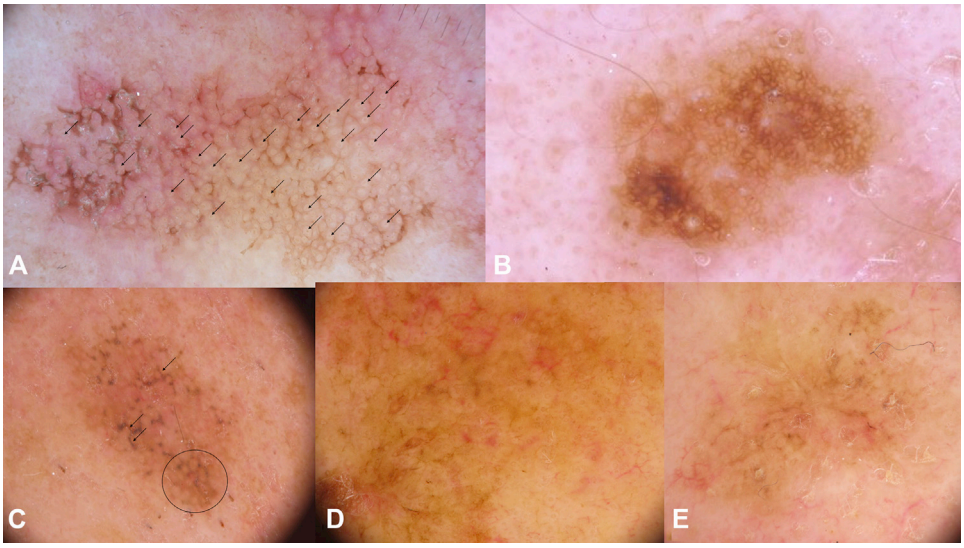

Fig 3. Examples of lesions of the data set. **A**, A pigmented actinic keratosis displaying erythema and white-to-yellow and wide follicular openings (arrows) compared to the follicles outside the lesion, which are hardly visible. **B**, A solar lentigo typified by reticular brown lines. Both lesions were correctly classified by the majority of readers in all 3 evaluations, as well as by artificial intelligence. (**C**) A lentigo maligna displaying gray angulated lines (arrows). An area with pigment network is also seen at the lower part (circle), but the feature is not predominant because it occupies a small proportion of the lesion's surface. It was correctly diagnosed by the majority of readers in all 3 evaluations, as well as by artificial intelligence. (**D**) A lentigo maligna lacking melanoma-specific criteria. It was misdiagnosed by the majority of readers on days 1 and 2 but correctly diagnosed by using the inverse approach. The lesion was also correctly classified by artificial intelligence. (**E**) A lentigo maligna lacking melanoma-specific criteria. It was misdiagnosed by the majority of readers on days 1 and 2 and misclassified by artificial intelligence. The lesion was correctly diagnosed as melanoma by the majority of the readers on day 3 using the inverse approach.

Our results confirm the common notion that the differential diagnosis of pigmented facial macules is highly challenging. In the initial test used to classify participants, which did not include facial lesions, the mean percentage of correct answers was 92.4%, much higher than in the test set of facial lesions without training (62.1%) or after training on pattern analysis (65.5%). The fact that early LM is very difficult to recognize was previously demonstrated.[3,4,8,15,16] In our study, the mean sensitivity for LM diagnosis was 51.5% without training and 56.7% after training based on pattern analysis, which means that approximately half of melanomas escaped detection.

The most important result of our study is the increase of sensitivity for LM with the inverse approach. The mean sensitivity of readers increased to 83.6%, and the most voted answer correctly classified all melanomas. Notably, this increase of sensitivity occurred without any cost in specificity, which remained unaltered. This is particularly noteworthy because previous diagnostic models for LM achieved a sensitivity for melanoma higher than 90% only if the specificity decreased to 50%.[8,16]

The diagnosis of malignant neoplasms based on the exclusion of benign neoplasms is not new in dermoscopy. The 2-step algorithm is based, to some extent, on the exclusion of common benign tumors but also evaluates the presence of melanoma-specific criteria.[25] In contrast, the inverse approach for LM diagnosis is exclusively based on the presence or absence of features that typify PAK and SL/SK (Fig 3).

The application of the inverse approach also improved the diagnostic accuracy for PAK and SL/SK and the total number of correct answers, highlighting that the method also enhances the discrimination between PAK and SL/SK. Although this is less relevant from a clinical perspective, it still has some value, because PAK is considered as a premalignant lesion or as an in situ squamous cell carcinoma that requires treatment, whereas SL/SK is a benign proliferation.[26]

Our subgroup analysis showed that the effect of teaching methods was similar for all readers, irrespective of their baseline diagnostic skills on dermoscopy (Table II). This finding indicates that the inverse approach is beneficial both for inexperienced and well-trained clinicians.

Artificial intelligence (AI) was shown to perform at least equally to humans in artificial environments.[24,27-30] A recent study found a very high accuracy for LM, but the control group did not include PAK, which represents the most challenging differential diagnosis.[31] On the other hand, AI also showed excellent performance on PAK when compared to human raters.[24] In this study, we showed that AI outperforms the majority of human readers without training and after training on pattern analysis. In contrast, using the inverse approach, more than half of human readers superseded the CNN in diagnostic accuracy.

We consider our results clinically relevant, because they indicate that the inverse approach significantly improves the capacity of clinicians to recognize inconspicuous LM. Most of the problems associated with LM management result from the fact that it is often diagnosed when it is large, which limits surgical treatment, especially on the cosmetically and functionally sensitive area of the face. Diagnosing LM at an early stage could obviously simplify the surgical management. A potential criticism is that the biologic course of LM is not fully elucidated in terms of growth rate and potential to invade the dermis and metastasize. Therefore, it is not clear whether diagnosing and treating early LM is always beneficial or represents an example of overdiagnosing and, subsequently, overtreating a lesion that would never become life threatening.[32] Although this controversy is intriguing, it lies beyond the aims and the power of this study. According to current practice, LM should be treated, and, therefore, the earlier the better.[33,34]

Our study has some limitations. First, it was conducted in an experimental setting, and any conclusion on the usefulness of the method in the real clinical practice is only indicative. Second, all included PAKs and SL/SKs were biopsied, indicating that they were assessed as diagnostically equivocal. This induces a selection bias, considering that most PAKs and SL/SKs are diagnosed clinically. Therefore, the diagnostic specificity of clinicians is possibly superior to that calculated here. Finally, we used a single AI algorithm, trained only on public data. Therefore, our findings cannot be generalized to other AI algorithms.

In conclusion, our study indicates that the inverse approach, added to the classic pattern analysis, significantly improves the diagnostic performance of clinicians evaluating facial pigmented macules, especially in terms of sensitivity for LM.

REFERENCES

1. Cohen LM. Lentigo maligna and lentigo maligna melanoma. *J Am Acad Dermatol*. 1995;33(6):923-940.
2. Uhlenhake EE, Sangueza OP, Lee AD, Jorizzo JL. Spreading pigmented actinic keratosis: a review. *J Am Acad Dermatol*. 2010;63(3):499-506.
3. Lallas A, Argenziano G, Moscarella E, Longo C, Simonetti V, Zalaudek I. Diagnosis and management of facial pigmented macules. *Clin Dermatol*. 2014;32(1):94-100.
4. Akay BN, Kocyigit P, Heper AO, Erdem C. Dermatoscopy of flat pigmented facial lesions: diagnostic challenge between pigmented actinic keratosis and lentigo maligna. *Br J Dermatol*. 2010;163(6):1212-1217.
5. Lallas A, Zalaudek I, Apalla Z, et al. Management rules to detect melanoma. *Dermatology*. 2013;226(1):52-60.
6. Schiffner R, Schiffner-Rohe J, Vogt T, et al. Improvement of early recognition of lentigo maligna using dermatoscopy. *J Am Acad Dermatol*. 2000;42(1 Pt 1):25-32.
7. Hongyan Z. Dermatoscopy of pigmented actinic keratosis—a striking similarity to lentigo maligna. *Int J Dermatol*. 2007;46(4):414-416.
8. Tschandl P, Rosendahl C, Kittler H. Dermatoscopy of flat pigmented facial lesions. *J Eur Acad Dermatol Venereol*. 2015;29(1):120-127.
9. Stolz W, Schiffner R, Burgdorf WHC. Dermatoscopy for facial pigmented skin lesions. *Clin Dermatol*. 2002;20(3):276-278.
10. Stante M, Giorgi V, Stanganelli I, Alfaioli B, Carli P. Dermoscopy for early detection of facial lentigo maligna. *Br J Dermatol*. 2005;152(2):361-364.
11. Slutsky JB, Marghoob AA. The zig-zag pattern of lentigo maligna. *Arch Dermatol*. 2010;146(12):1444.
12. Nascimento MM, Shitara D, Enokihara MMSS, Yamada S, Pellacani G, Rezze GG. Inner gray halo, a novel dermoscopic feature for the diagnosis of pigmented actinic keratosis: clues for the differential diagnosis with lentigo maligna. *J Am Acad Dermatol*. 2014;71(4):708-715.
13. Zalaudek I, Piana S, Moscarella E, et al. Morphologic grading and treatment of facial actinic keratosis. *Clin Dermatol*. 2014;32(1):80-87.
14. Giacomel J, Lallas A, Argenziano G, Bombonato C, Zalaudek I. Dermoscopic "signature" pattern of pigmented and non-pigmented facial actinic keratoses. *J Am Acad Dermatol*. 2015;72(2):e57-e59.
15. Zalaudek I, Ferrara G, Leinweber B, Mercogliano A, D'Ambrosio A, Argenziano G. Pitfalls in the clinical and dermoscopic diagnosis of pigmented actinic keratosis. *J Am Acad Dermatol*. 2005;53(6):1071-1074.
16. Lallas A, Tschandl P, Kyrgidis A, et al. Dermoscopic clues to differentiate facial lentigo maligna from pigmented actinic keratosis. *Br J Dermatol*. 2016;174(5):1079-1085.
17. Moscarella E, Rabinovitz H, Zalaudek I, et al. Dermoscopy and reflectance confocal microscopy of pigmented actinic keratoses: a morphological study. *J Eur Acad Dermatol Venereol*. 2014;29(2):307-314.
18. Tschandl P, Gambardella A, Boespflug A, et al. Seven non-melanoma features of facial lesions: a dermatoscopy algorithm to rule out melanoma. *Acta Derm Venereol*. 2017;97(10):1219-1224.
19. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada; December 8-14, 2019. Available at: https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf. Accessed September 3, 2020.
20. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211-252.
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Presented at: 2016 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), Las Vegas, Nevada; June 27-30, 2016. Available at: https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf. Accessed September 3, 2020.

22. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016;770-778 Available at: https://arxiv.org/abs/1512.03385. Accessed December 10, 2015.
23. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;5:180161.
24. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*. 2019;20(7):938-947.
25. Marghoob AA, Braun R. Proposal for a revised 2-step algorithm for the classification of lesions of the skin using dermoscopy. *Arch Dermatol*. 2010;146(4):426-428.
26. Werner RN, Stockfleth E, Connolly SM, et al. Evidence- and consensus-based (S3) guidelines for the treatment of actinic keratosis—International League of Dermatological Societies in cooperation with the European Dermatology Forum—short version. *J Eur Acad Dermatol Venereol*. 2015;29:2069-2079.
27. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118.
28. 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):1836-1842.
29. Haenssle HA, Fink C, Toberer F, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol*. 2020;31(1):137-143.
30. Tschandl P, Rosendahl C, Akay BN, et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. *JAMA Dermatol*. 2018;155(1):58-65.
31. Winkler JK, Sies K, Fink C, et al. Melanoma recognition by a deep learning convolutional neural network-Performance in different melanoma subtypes and localisations. *Eur J Cancer*. 2020;127:21-29.
32. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst*. 2010;102(9):605-613.
33. Garbe C, Amaral T, Peris K, et al. European consensus-based interdisciplinary guideline for melanoma. Part 1: diagnostics—update 2019. *Eur J Cancer*. 2020;126:141-158.
34. Garbe C, Amaral T, Peris K, et al. European consensus-based interdisciplinary guideline for melanoma. Part 2: treatment—update 2019. *Eur J Cancer*. 2020;126:59-177.