nature portfolio

Peer Review File

Predictive models of melanoma metastasis based on dermatoscopy in an international retrospective human reader study

Corresponding Author: Professor Aimilios Lallas

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Lallas et al. conducted an international human reader study to evaluate associations between dermatoscopic features and melanoma metastasis. Secondary analyses included evaluating predictive models of melanoma metastasis using histopathologic and dermatoscopic features.

Data source – Patient population: More information about the study sample is needed. What proportion of patients from each center were included? What proportion of patients at each center had dermatoscopic images? How representative are the selected patients relative to each center's patient population? Were patients with thicker lesions included (as suggested in the limitations section)? Were there differences between centers regarding the types of patients seen? Were any centers tertiary care centers? Regarding the patients themselves, were the included melanomas the first primary melanoma for all patients? What was the time between initial diagnosis and occurrence of metastatic melanoma for the metastatic melanoma patients? Did any patients have >1 (simultaneous) melanoma at the time of study selection? What types of melanomas did the patients have?

Methods, Participant selection: The text states that readers from each center needed to have a minimum of 5 years of experience in dermatoscopy. However, Supplementary table 12 shows the range of experience from 2 – 35 years. How many readers had experience <5 years?

Statistical analyses: Were regression analyses adjusted for potential confounders? Were any of the analyses corrected for multiple comparisons?

Table 1, Results: Almost 14% of tumors were acral and 0.8% subungual. Since acral tumors tend to be thicker than superficial spreading melanomas (SSM) and were observed with much higher frequency in the metastatic group, there is concern that these tumors might have skewed the results. The authors should conduct a sensitivity analysis excluding acral and subungual tumors from the evaluations to see what impact these specific tumor types have on the analyses. Given the differences in etiologic factors related to acral/subungual tumors versus SSM, the authors should include discussion about keeping this set of tumors in the analyses.

It is difficult to interpret the data in this manuscript without knowing the histopathologic subtypes of the melanomas evaluated. For example, are the associations reported in this paper predominantly related to superficial spreading melanomas or are they applicable across subtypes (lentigo maligna melanoma, acral melanoma, etc.)?

Results, Supplementary table 3: Most of the dermatoscopic features evaluated had only fair interrater agreement. Given this, how can these features be considered for prognostication purposes?

NCCN guidelines do not recommend sentinel lymph node (SLN) biopsy for cutaneous melanomas with a Breslow thickness <0.8 mm in the absence of ulceration. It would be clinically informative to know if the presence of a blue-white veil in thin, non-ulcerated melanomas <0.8 mm in thickness is associated with disease progression and/or worse survival (assuming there is power to conduct such an analysis).

As mentioned in the discussion and shown in supplementary table 3, the interobserver agreement among readers was low. This issue makes it problematic to use dermatoscopy (currently) as a biomarker to predict metastasis, recurrence, etc. Further discussion of this issue is needed.

Minor comments:

Abstract (line 3) – There is an extra “with” in this line.

Results, Comparative analysis of the accuracy of models: The authors should include a sentence stating that there were no significant differences between the models.

Supplementary table 1: Individualized assessment of changes in stage would be informative. This information could be added as a supplemental figure.

Supplementary table 5: Authors should add footnote listing the variables that were included in the multivariate model and include mention of the methods used to determine the variables.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3

(Remarks to the Author)
- The authors primary objective was to evaluate the role of dermoscopy in predicting melanoma prognosis by investigating whether specific dermatoscopic criteria was associated with the development of metastases. This was a multi-center study, and included a total of 524 patients with cutaneous melanoma
- This is a well-executed study that contributes valuable data to the current literature. The authors clearly stated the limitations of the study. Thus, if these findings are further validated and effectively translated into clinical practice, dermoscopy could become an essential, easy-to-access- decision-making tool for prognosis and treatment planning. In particular, it adds additional information for accurate identification of early stage but high-risk patients who may benefit from some treatments such as immunotherapy
- work and results support the conclusions

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5

(Remarks to the Author)
Major
1. Page 1, line 1. The title should specify the specific nature of the 'human reader study' (e.g., 'multicenter retrospective').
2. Page 3-4. It is recommended to supplement the comparison of results among several models in either the Abstract or Main section, such as AUC and accuracy comparisons in the Training set and Test set.
3. Page 21. Please supplement the baseline patient staging in Table 1.
4. Page 22, line 23. The wide confidence intervals for "Dermatoscopic Ulceration(>50%) " (OR=3.84, 95% CI: 1.79–8.23) and "blue-white veil" (OR=6.10, 95% CI: 3.65–10.17) in Table 2 suggest limited precision, which should be addressed in the Discussion section.
5. Page 24, Fig. 2, Fig. 3. It is recommended to supplement with risk tables when presenting survival curve visualizations from Cox model results (particularly in multivariate analyses).
6. Page 9, line 33. The Discussion section should further analyze potential reasons for the discrepancy between this study's finding ("Regression structures (>50%) = low metastatic risk") and previous conclusions .
7. Page 10, line 33. When emphasizing the preoperative advantages of dermoscopy in the Discussion, a comparative analysis with postoperative gold standards (e.g., Breslow thickness + ulceration status) is recommended.
8. The impact of initial metastatic status at diagnosis was not controlled for in the analysis.

Minor
1. Page 7, line 7. Please specify the corresponding table or figure for the metastasis prediction data during follow-up.
2. Page 16, line 18. In the Disclosures section, could the affiliation of author HPS with a dermoscopy company potentially

influence the image analysis criteria?
3. The formatting of Supplementary Table 8 requires correction.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have responded to the questions raised in the initial review. It is unfortunate that data that would permit further examination of important issues is unavailable, but the authors at least mention this in the limitations.

An additional sensitivity analysis would strengthen the manuscript: please perform a sensitivity analysis adjusting for histopathologic subtype among the 291 cases in which these data are available. This analysis should be possible with all the caveats of limited sample size.

Reviewer #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5

(Remarks to the Author)
The authors have satisfactorily addressed my comments. Lallas et al. conducted an international, multicenter retrospective study to investigate the association between dermatoscopic features of primary melanoma and metastasis. The authors noted the limitations of this work. If these findings are validated in larger prospective studies in the future, the clinical applications will be something to look forward to. It is suggested that dermoscopy could potentially serve as a non-invasive and convenient tool for early prediction of metastasis risk in patients with primary melanoma, thereby providing valuable guidance for subsequent treatment decisions.

Predictive models of melanoma metastasis based on dermatoscopy. An international retrospective human ready study

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Lallas et al. conducted an international human reader study to evaluate associations between dermatoscopic features and melanoma metastasis. Secondary analyses included evaluating predictive models of melanoma metastasis using histopathologic and dermatoscopic features.

- Data source – Patient population: More information about the study sample is needed. What proportion of patients from each center were included? What proportion of patients at each center had dermatoscopic images? How representative are the selected patients relative to each center's patient population? Were patients with thicker lesions included (as suggested in the limitations section)? Were there differences between centers regarding the types of patients seen? Were any centers tertiary care centers? Regarding the patients themselves, were the included melanomas the first primary melanoma for all patients? What was the time between initial diagnosis and occurrence of metastatic melanoma for the metastatic melanoma patients? Did any patients have >1 (simultaneous) melanoma at the time of study selection? What types of melanomas did the patients have?

- Methods, Participant selection: The text states that readers from each center needed to have a minimum of 5 years of experience in dermatoscopy. However, Supplementary table 12 shows the range of experience from 2 – 35 years. How many readers had experience <5 years?

- Statistical analyses: Were regression analyses adjusted for potential confounders? Were any of the analyses corrected for multiple comparisons?

- Table 1, Results: Almost 14% of tumors were acral and 0.8% subungual. Since acral tumors tend to be thicker than superficial spreading melanomas (SSM) and were observed with much higher frequency in the metastatic group, there is concern that these tumors might have skewed the results. The authors should conduct a sensitivity analysis excluding acral and subungual tumors from the evaluations to see what impact these specific tumor types have on the analyses. Given the differences in etiologic factors related to acral/subungual tumors versus SSM, the authors should include discussion about keeping this set of tumors in the analyses.

- It is difficult to interpret the data in this manuscript without knowing the histopathologic subtypes of the melanomas evaluated. For example, are the associations reported in this paper predominantly related to superficial spreading melanomas or are they applicable across subtypes (lentigo maligna melanoma, acral melanoma, etc.)?

- Results, Supplementary table 3: Most of the dermatoscopic features evaluated had only fair interrater agreement. Given this, how can these features be considered for prognostication purposes?

- NCCN guidelines do not recommend sentinel lymph node (SLN) biopsy for cutaneous melanomas with a Breslow thickness <0.8 mm in the absence of ulceration. It would be clinically informative to know if the presence of a blue-white veil in thin, non-ulcerated melanomas <0.8 mm in thickness is associated with disease progression and/or worse survival (assuming there is power to conduct such an analysis)
- As mentioned in the discussion and shown in supplementary table 3, the interobserver agreement among readers was low. This issue makes it problematic to use dermatoscopy (currently) as a biomarker to predict metastasis, recurrence, etc. Further discussion of this issue is needed.

Minor comments:

-Abstract (line 3) – There is an extra “with” in this line.

-Results, Comparative analysis of the accuracy of models: The authors should include a sentence stating that there were no significant differences between the models.

-Supplementary table 1: Individualized assessment of changes in stage would be informative. This information could be added as a supplemental figure.

-Supplementary table 5: Authors should add footnote listing the variables that were included in the multivariate model and include mention of the methods used to determine the variables.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #3 (Remarks to the Author):

- The authors primary objective was to evaluate the role of dermoscopy in predicting melanoma prognosis by investigating whether specific dermatoscopic criteria was associated with the development of metastases. This was a multi-center study, and included a total of 524 patients with cutaneous melanoma
- This is a well-executed study that contributes valuable data to the current literature. The authors clearly stated the limitations of the study. Thus, if these

findings are further validated and effectively translated into clinical practice, dermoscopy could become an essential, easy-to-access- decision-making tool for prognosis and treatment planning. In particular, it adds additional information for accurate identification of early stage but high-risk patients who may benefit from some treatments such as immunotherapy
- work and results support the conclusions

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5 (Remarks to the Author):

Major
1. Page 1, line 1. The title should specify the specific nature of the 'human reader study' (e.g., 'multicenter retrospective').

2. Page 3-4. It is recommended to supplement the comparison of results among several models in either the Abstract or Main section, such as AUC and accuracy comparisons in the Training set and Test set.

3. Page 21. Please supplement the baseline patient staging in Table 1

4. Page 22, line 23. The wide confidence intervals for "Dermatoscopic Ulceration (>50%) " (OR=3.84, 95% CI: 1.79–8.23) and "blue-white veil" (OR=6.10, 95% CI: 3.65–10.17) in Table 2 suggest limited precision, which should be addressed in the Discussion section.

5. Page 24, Fig. 2, Fig. 3. It is recommended to supplement with risk tables when presenting survival curve visualizations from Cox model results (particularly in multivariate analyses).

6. Page 9, line 33. The Discussion section should further analyze potential reasons for the discrepancy between this study's finding ("Regression structures (>50%) = low metastatic risk") and previous conclusions.

7. Page 10, line 33. When emphasizing the preoperative advantages of dermoscopy in the Discussion, a comparative analysis with postoperative gold standards (e.g., Breslow thickness + ulceration status) is recommended.

8. The impact of initial metastatic status at diagnosis was not controlled for in the analysis.

Minor

1. Page 7, line 7. Please specify the corresponding table or figure for the metastasis prediction data during follow-up.

2. Page 16, line 18. In the Disclosures section, could the affiliation of author HPS with a dermoscopy company potentially influence the image analysis criteria?

3. The formatting of Supplementary Table 8 requires correction.

Point-by-point reply to reviewers' comments

We thank the reviewers for their insightful questions and their constructive comments. Below, we provide a point-by-point reply to reviewers' comments.

Reviewer #1 (Remarks to the Author):

Lallas et al. conducted an international human reader study to evaluate associations between dermatoscopic features and melanoma metastasis. Secondary analyses included evaluating predictive models of melanoma metastasis using histopathologic and dermatoscopic features.

- Data source – Patient population: More information about the study sample is needed. What proportion of patients from each center were included? What proportion of patients at each center had dermatoscopic images? How representative are the selected patients relative to each center's patient population? Were patients with thicker lesions included (as suggested in the limitations section)? Were there differences between centers regarding the types of patients seen? Were any centers tertiary care centers? Regarding the patients themselves, were the included melanomas the first primary melanoma for all patients? What was the time between initial diagnosis and occurrence of metastatic melanoma for the metastatic melanoma patients? Did any patients have >1 (simultaneous) melanoma at the time of study selection? What types of melanomas did the patients have?

We thank the reviewer for the important comment. We agree that the proportion of patients recruited from each center and the percentage with available dermatoscopic images would add interesting information. Based on the study inclusion criteria, participating centers were asked to provide cases that meet all of the inclusion criteria (including the availability of dermatoscopic images). Therefore, the total number of melanoma patients treated in these centers is not available to the study team. We understand that this represents a selection bias which is a limitation related to the retrospective study design. However, we have no reason to believe that the sample might not be representative in terms of dermatoscopic characteristics,

since tumors of all stages were included. In addition, aiming to enhance the representativeness of our sample, we included patients from multiple tertiary skin cancers across the globe (amended accordingly in "Methods, p15, subheading "Data source-population"), while readers from different skin cancer centers were recruited.

Regarding lesions' thickness, yes, thicker lesions were included. We included patients with stage IB melanoma and above (8th edition, AJCC stage). The main reason for not including stage IA was the very low risk of metastasis, considering that our aim was to identify dermatoscopic criteria that predict metastasis. We set the IB AJCC stage as inclusion criterion in "Methods" section, p15 and we acknowledged also as a limitation of our study in p14 "Indeed … patient population".

Regarding the participating centers, all are tertiary skin cancer centers. No difference was found among them concerning the types of melanoma or any other patient or tumor characteristic.

Regarding types of melanomas, we included only patients with first primary melanoma, while patients with more that 1 simultaneous melanoma were excluded from the analysis a priori. We amended inclusion and exclusion criteria accordingly in "Methods", p15, subheading "Data source – Patient population". In addition, we provided in Results, p7, subheading "RFS and DMFS in early-stage melanomas" the median time to metastasis for patients with early-stage melanomas.

- Methods, Participant selection: The text states that readers from each center needed to have a minimum of 5 years of experience in dermatoscopy. However, Supplementary table 12 shows the range of experience from 2 – 35 years. How many readers had experience <5 years?

We thank the reviewer for the comment. We set as prerequisite for evaluator participation in the study at least a 5 years' experience in dermatoscopy in order to secure reliability of image evaluation and avoiding significant deviations during evaluation process. Indeed, one evaluator had 2 years' experience in dermatoscopy. We provided further clarifications in methods section, p15, subheading "participant selection- Web-based interface – Procedure for lesion evaluation".

- Statistical analyses: Were regression analyses adjusted for potential confounders? Were any of the analyses corrected for multiple comparisons?

We thank the reviewer for the constructive comment. In our primary analyses, we did not account for potential confounders such as clinical factors. According to reviewer's suggestion, we included age, gender and anatomic location in our multivariable modes and the logistic regression and multivariable Cox regression models derived after adjusting for confounders are provided in Table 2 and Supp. Table 6. Also, models' performance after adjusting for those factors are provided in Table 3, Supplementary Table 6,8,10 and Figures 2,3. Regarding multiple comparisons and AUC values

comparisons, our analysis did not find statistically significant results, so we did not conduct multiple comparisons tests. Regarding regression models, we now apply Holm-Bonferroni step down procedure for family-wise error rate control on univariate regression analyses, aiming to examine a possible effect of multiple comparisons in significance of predictors. After adjusting p-values with this approach, regression structures render marginally not statistically significant (p=0.114) for the prediction of metastasis and not significant for RFS and DMFS. Those results are now provided in Results section, subheading "Sensitivity analysis" and in Supplementary Table 12.

- Table 1, Results: Almost 14% of tumors were acral and 0.8% subungual. Since acral tumors tend to be thicker than superficial spreading melanomas (SSM) and were observed with much higher frequency in the metastatic group, there is concern that these tumors might have skewed the results. The authors should conduct a sensitivity analysis excluding acral and subungual tumors from the evaluations to see what impact these specific tumor types have on the analyses. Given the differences in etiologic factors related to acral/subungual tumors versus SSM, the authors should include discussion about keeping this set of tumors in the analyses.

We thank the reviewer for this important comment. We acknowledge the biologic aggressiveness of acral and subungual lesions, which could possibly alter the risk for metastasis and the dermatoscopic characteristics of those lesions. For that reason, and according to reviewer's suggestion we conducted further analyses, excluding those patients and the results are provided in Results section (subheading "Sensitivity analyses") and in supplementary Table 11.

- It is difficult to interpret the data in this manuscript without knowing the histopathologic subtypes of the melanomas evaluated. For example, are the associations reported in this paper predominantly related to superficial spreading melanomas or are they applicable across subtypes (lentigo maligna melanoma, acral melanoma, etc.)?

We thank the reviewer for this insightful comment. The dermatoscopic features selected for lesion evaluation are common and well described criteria in the literature and they are detected across various histologic subtypes with similar frequencies. This might support the broader applicability of our results. Nevertheless, we acknowledge that information on histologic subtype would be valuable for interpreting our results and for assessing the generalizability of our findings. Unfortunately, this information was available only for a subset of patients, as shown in the table below. Despite the distribution of subtypes among patients with available data follows the frequencies reported in the literature, it was not possible to include histopathologic subtype as a covariate in our multivariable models. We added this as a limitation of our study and we add a corresponding statement in Discussion section, p13, "in addition, … could not be possible".

Frequencies of histologic subtypes for patients with melanoma and available data and distribution between non-metastatic and metastatic lesions at last follow up (n=291).			
	Total (n, %)	Non-metastatic lesions	Metastatic
SSM	165 (56.7)	91 (61.1)	74 (52.1)
NM	59 (20.2)	26 (17.4)	33 (23.2)
ALM	36 (12.4)	15 (10.1)	21 (14.8)
LMM	31 (10.6)	17 (11.4)	14 (9.9)
Total	291	149	142

SSM: superficial spreading melanoma, NM: nodular melanoma, ALM: acral lentiginous melanoma, LMM: lentigo maligna melanoma, n: number of patients

- Results, Supplementary table 3: Most of the dermatoscopic features evaluated had only fair interrater agreement. Given this, how can these features be considered for prognostication purposes?

We thank the reviewer for the constructive comment. Indeed, interobserver agreement was fair in some dermatoscopic features categories. We tried to secure a robust and reliable image evaluation by inviting readers from multiple centers and with median 19 years experience in dermatoscopy, aiming to eliminate the risk of single-center lesion evaluation or the inclusion of non-experts in dermatoscopy, which could possibly cause deviations in image evaluation process. Also, we included common dermatoscopic criteria which significantly predicted prognostic factors such as Breslow thickness and ulceration in previous published studies. Also, among the most prominent variables included in the models, such as pigmentation and ulceration, have moderate agreement, while bue/gray colors and blue-white veil demonstrate alpha values in the upper quartile of the fair agreement range. However, we acknowledge this fair agreement as a possible limitation of our analysis and present it in Discussion section (p 13).

- NCCN guidelines do not recommend sentinel lymph node (SLN) biopsy for cutaneous melanomas with a Breslow thickness <0.8 mm in the absence of ulceration. It would be clinically informative to know if the presence of a blue-white veil in thin, non-ulcerated melanomas <0.8 mm in thickness is associated with disease progression and/or worse survival (assuming there is power to conduct such an analysis).

We thank the reviewer for the thoughtful and clinically relevant comment. As explained above, our dataset did not include stage IA melanoma, therefore Is unfortunately not powered to investigate this very interesting question. However, motivated by reviewer's question, we conducted 2 additional

discrete sensitivity analyses. First, we included patients with Breslow thickness (BT) ≤1mm (n=96 patients) and examined the association of predictors from model 1 for the prediction of metastasis. Blue-white veil remained a statistically significant predictor of metastasis development in multivariable analysis, despite the small number of patients included (OR 8.39, 95%CI 1.38 – 50.86). Then, we included patients with BT≤2mm, the current threshold for selection for adjuvant immunotherapy. 295 patients were included in the analysis and blue-white veil and extensive dermatoscopic ulceration remained statistically significant predictors in multivariable analysis, both leading to increased risk for metastasis development [OR 4.10, p<0.001, 95%CI 2.03 – 8.35 and OR 9.76, p=0.030, 95%CU 1.30 – 73.14, respectively]. Contrary to, for patients with Breslow thickness >2mm (n=229), heavy pigmentation and extensive regression deemed negative predictors for metastasis development, as shown in the table provided in Supplementary File. Regarding RFS and DMFS, the number of events in patients with BT ≤1mm was very low, rendering the analysis not feasible. This was not the case for BT threshold of 2.0 mm. Similar predictors were drawn for lesions with BT ≤2.0mm, while extensive regression was associated with longer DMFS in patients with BT >2.0mm. The main results of the analyses are now available in Results section (Subheading "Sensitivity analyses", " In addition … early-stage melanomas, RFS, DMFS") and in Supplementary Table 15.

- As mentioned in the discussion and shown in supplementary table 3, the interobserver agreement among readers was low. This issue makes it problematic to use dermatoscopy (currently) as a biomarker to predict metastasis, recurrence, etc. Further discussion of this issue is needed.

We thank the reviewer for the kind comment. We have addressed this issue in a reply to previous reviewer's comment and added a relevant limitation.

Minor comments:

-Abstract (line 3) – There is an extra "with" in this line.

We amended accordingly in Abstract.

-Results, Comparative analysis of the accuracy of models: The authors should include a sentence stating that there were no significant differences between the models.

We thank the reviewer for the comment. According to the suggestion, we added the phrase "and no statistically significant different" in "Results" section, subheading "comparative analysis of the accuracy of models". Also, it is already stated for RFS and DMFS in "Results" section, subheading "RFS and DMFS in early-stage melanomas" with the sentence "There was no significant difference in accuracy among the models (DeLong's test, p > 0.05)".

-Supplementary table 1: Individualized assessment of changes in stage

would be informative. This information could be added as a supplemental figure.

We thank the reviewer for the constructive comment. We now created a plot showing transition of melanoma stage at diagnosis and last follow up and was added as figure 1f.

Supplementary table 5: Authors should add footnote listing the variables that were included in the multivariate model and include mention of the methods used to determine the variables.

We amended accordingly and, in all tables, referring to multivariable models.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for the effort reviewing our manuscript and for the constructive comments.

Reviewer #3 (Remarks to the Author):

- The authors primary objective was to evaluate the role of dermoscopy in predicting melanoma prognosis by investigating whether specific dermatoscopic criteria was associated with the development of metastases. This was a multi-center study, and included a total of 524 patients with cutaneous melanoma
- This is a well-executed study that contributes valuable data to the current literature. The authors clearly stated the limitations of the study. Thus, if these findings are further validated and effectively translated into clinical practice, dermoscopy could become an essential, easy-to-access- decision-making tool for prognosis and treatment planning. In particular, it adds additional information for accurate identification of early stage but high-risk patients who may benefit from some treatments such as immunotherapy
- work and results support the conclusions

We thank the reviewer for the encouraging comments. In limitations section, we tried to report possible limitation of our work. In addition, we reported that a prospective, external validation of our results seems important for future models use.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to

facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for the effort reviewing our manuscript and for the constructive comments.

Reviewer #5 (Remarks to the Author):

Major
1. Page 1, line 1. The title should specify the specific nature of the 'human reader study' (e.g., 'multicenter retrospective').

We thank the reviewer. We amended the title accordingly.

2. Page 3-4. It is recommended to supplement the comparison of results among several models in either the Abstract or Main section, such as AUC and accuracy comparisons in the Training set and Test set.

We thank the reviewer for the kind comment. We amended accordingly, and we added the AUC values for metastasis prediction in the test set (Results, Section, Subheading "Comparative analysis of the accuracy of models"). Also, we added the AUC values for 5-fold cross validation in training set for RFS prediction (Results section, Subheading "RFS and DMFS in early-stage melanomas").

3. Page 21. Please supplement the baseline patient staging in Table 1

We thank the reviewer for the comment. We amended accordingly and now in Table 1, the baseline and at last follow up stage of patients with melanoma included in the analysis is provided.

4. Page 22, line 23. The wide confidence intervals for "Dermatoscopic Ulceration (>50%) " (OR=3.84, 95% CI: 1.79–8.23) and "blue-white veil" (OR=6.10, 95% CI: 3.65–10.17) in Table 2 suggest limited precision, which should be addressed in the Discussion section.

We thank the reviewer for the kind comment. We speculate that wide confidence intervals in some variables could derive from few observations in those categories, for example extensive ulceration in non-metastatic lesions. We now add this as a limitation (Discussion p13-14, limitations, "Another consideration… non-metastatic lesions") of our study which corroborates to the need of further validation in larger datasets.

5. Page 24, Fig. 2, Fig. 3. It is recommended to supplement with risk tables when presenting survival curve visualizations from Cox model results (particularly in multivariate analyses).

We thank the reviewer for this suggestion. As recommended, we included risk tables below the survival curves in Figure 3c, d and 4. These risk tables provide the number of patients at risk at each time point for both predicted

groups. Also, we provided survival probabilities and event status for those figures in Source Data.

6. Page 9, line 33. The Discussion section should further analyze potential reasons for the discrepancy between this study's finding ("Regression structures (>50%) = low metastatic risk") and previous conclusions.

We thank the reviewer for the thoughtful comment. The prognostic significance of histopathologic regression remains an area of great debate in the literature. Possible reasons for discrepancies in results from observational studies could be the variability in definition of histopathologic regression. For that reason, we provided results from recently published, large-scale, observational studies supporting the protective role of regression in metastasis development which were further verified in recent meta-analyses (provided in References). Now, according to reviewer's suggestion, we provided further insights on the prognostic role of dermatoscopic regression structures. Observational studies linked those structures indirectly with tumor prognosis, highlighting their detection more frequently in melanoma in situ or thin invasive lesions compared to thick melanomas. Moreover, primary tumors in patients with negative sentinel lymph node status exhibited regression structures more often compared to metastatic ones. Those findings, along with the pathophysiologic rationale from basic research and in vitro studies that we provided in Discussion section, could further support the protective role of dermatoscopic regression in metastasis development. However, a variability in dermatoscopic regression detection could not be overlooked. This was also evident from our interrater agreement analysis, where regression structures had fair agreement. We amended Discussion section accordingly.

7. Page 10, line 33. When emphasizing the preoperative advantages of dermoscopy in the Discussion, a comparative analysis with postoperative gold standards (e.g., Breslow thickness + ulceration status) is recommended.

We thank the reviewer for the constructive comment. We now added a sentence in Discussion section stating that a non-statistically significant difference between preoperative dermatoscopic models and postoperative gold standards could further support the possibility of preoperative risk assessment using digital biomarkers.

8. The impact of initial metastatic status at diagnosis was not controlled for in the analysis.

We thank the reviewer for the kind comment. In our dataset, a subset of patients presented with regional or distant metastatic disease at diagnosis. To assess whether initial metastatic status could confound our findings, we performed additional sensitivity analyses. First, we cross-tabulated dermatoscopic features according to regional or distant metastatic disease at

diagnosis and we found no statistically significant association, as shown in the table below. Moreover, we repeated our multivariable logistic regression models after excluding patients with initial metastatic disease at diagnosis. The results of those analyses remained consistent with our primary findings, except for extensive regression structures, which render marginally statistically non-significant (OR 0.35, p=0.06, 95%CI 0.12 – 1.07) (Supp. Table 14). We reported the results of those analyses in "Results" section, subheading sensitivity analyses with a statement "Furthermore … model 1" and provided a supplementary Table (Supplementary Table 13) with the results of sensitivity analyses.

Supplementary Table 13: Frequencies of dermatoscopic features according to initial metastatic status (n=143)

	Total (N=143)	**Regional metastatic disease at diagnosis (n= 95)**	**Distant metastatic disease at diagnosis (n=48)**	**X^2-test (p-value)**
Pigmentation				0.405
Absent	17 (11.9)	9 (9.5)	8 (16.7)	
25	49 (34.3)	35 (36.8)	14 (29.2)	
50	35 (24.5)	26 (27.4)	9 (18.8)	
75	19 (13.3)	12 (12.6)	7 (14.6)	
100	23 (16.1)	13 (13.7)	10 (20.8)	
Colors				
Brown	107 (74.8)	75 (78.9)	32 (66.7)	0.110
Blue/gray	100 (69.9)	68 (71.6)	32 (66.7)	0.545
Red	107 (74.8)	70 (73.7)	37 (77.1)	0.658
White	99 (69.2)	64 (67.4)	35 (72.9)	0.497
Black	63 (44.1)	39 (41.1)	24 (50)	0.309
Ulceration				0.607
Absent	71 (49.7)	47 (49.5)	24 (50)	
1-49	36 (25.2)	26 (27.4)	10 (20.8)	
>50	36 (25.2)	22 (23.2)	14 (29.2)	
Regression structures				0.773
Absent	104 (72.7)	70 (73.7)	34 (70.8)	
1-49	30 (21)	20 (21.1)	10 (20.8)	
>50	9 (6.3)	5 (5.3)	4 (8.3)	
White shiny streaks	89 (62.2)	60 (63.2)	29 (60.4)	0.749
Blue-white veil	90 (62.9)	60 (63.2)	30 (62.5)	0.939
Eccentric blotch	61 (42.7)	40 (42.1)	21 (43.8)	0.851
Angulated lines	5 (3.5)	4 (4.2)	1 (2.1)	0.513
Parallel ridge	20 (3.8)	8 (2.6)	12 (5.4)	0.104

Vessels				0.02
Absent	42 (29.4)	26 (27.4)	16 (33.3)	
1-49	46 (32.2)	25 (26.3)	21 (43.8)	
>50	55 (38.5)	44 (46.3)	11 (22.9)	

n: number of patients

Supplementary Table 14: Sensitivity analysis and results from multivariable logistic regression for metastasis prediction based on Model 1 and after exclusion of patients with metastasis at diagnosis

	OR	**p-value**	**95%CI**
Pigmentation			
Absent	Ref.		
25	0.57	0.346	0.17 – 1.83
50	0.38	0.112	0.11 – 1.25
75	0.13	<0.001	0.04 – 0.44
100	0.06	<0.001	0.02 – 0.21
Dermatoscopic Ulceration			
Absent	Ref.		
1-49	1.20	0.61	0.58 – 2.47
>50	2.85	0.04	1.05 – 7.75
Regression structures			
Absent	Ref.		
1-49	0.68	0.323	0.32 – 1.44
>50	0.35	0.06	0.12 – 1.07
Blue – white veil	6.34	<0.001	3.05 – 13.76

N: number of patients, OR: Odds ratio, CI: confidence interval, Ref.: Reference level
Model was adjusted for age, gender and anatomic location

Minor

1. Page 7, line 7. Please specify the corresponding table or figure for the metastasis prediction data during follow-up.

We thank the reviewer for the comment. We amend accordingly.

2. Page 16, line 18. In the Disclosures section, could the affiliation of author HPS with a dermoscopy company potentially influence the image analysis criteria?

We have no reason to believe so, since the raters were unaware of the device used to capture each separate image.

3. The formatting of Supplementary Table 8 requires correction.

We thank the reviewer for the comment. Supplementary Table 8 (previously reporting frequencies of clinical, histopathological and dermatoscopic features according to training and test set) was amended according to reviewer's suggestion and is presented as supplementary Table 7 in the revised supplementary file.

Predictive models of melanoma metastasis based on dermatoscopy. An international retrospective human ready study

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have responded to the questions raised in the initial review. It is unfortunate that data that would permit further examination of important issues is unavailable, but the authors at least mention this in the limitations.

An additional sensitivity analysis would strengthen the manuscript: please perform a sensitivity analysis adjusting for histopathologic subtype among the 291 cases in which these data are available. This analysis should be possible with all the caveats of limited sample size.

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #5 (Remarks to the Author):

The authors have satisfactorily addressed my comments. Lallas et al. conducted an international, multicenter retrospective study to investigate the association between dermatoscopic features of primary melanoma and metastasis. The authors noted the limitations of this work. If these findings are validated in larger prospective studies in the future, the clinical applications will be something to look forward to. It is suggested that dermoscopy could potentially serve as a non-invasive and convenient tool for early prediction of metastasis risk in patients with primary melanoma, thereby providing valuable guidance for subsequent treatment decisions.

Point-by-point reply to reviewers

We thank the reviewers for their thoughtful comments and constructive suggestions, which have materially strengthened the manuscript. We are grateful for the positive assessment and recognition of the potential for non-invasive prognostic evaluation of melanoma by dermatoscopy, as well the importance of validating this approach in prospective cohorts. We have revised the manuscript and below is provided a point-by-point response to any reviewers’ comments.

Reviewer #1 (Remarks to the Author):

The authors have responded to the questions raised in the initial review. It is unfortunate that data that would permit further examination of important issues is unavailable, but the authors at least mention this in the limitations.

An additional sensitivity analysis would strengthen the manuscript: please perform a sensitivity analysis adjusting for histopathologic subtype among the 291 cases in which these data are available. This analysis should be possible with all the caveats of limited sample size.

We thank the reviewer and we are pleased that the amendments made following initial revision have adequately addressed the comments. We acknowledge the limitation of lack of histopathologic subtype of the tumor, and we add it as a limitation of our study. Following reviewer's suggestion, we conducted an additional sensitivity analysis, where model 1 was further adjusted for histopathologic subtype for the prediction of metastasis. The results of the analysis are presented as Supplementary Table 17 and below. Also, we added a statement in the Sensitivity analysis of Results section, arguing that the addition of histopathologic subtype did not alter the primary results. ("Also, despite…primary results").

	OR	p-value	95%CI
Pigmentation			
Absent	Ref.		
25	1.01	0.984	0.34 – 3.58
50	0.43	0.100	0.17 – 2.16
75	0.14	<0.001	0.04 – 0.30
100	0.07	<0.001	0.03 – 0.19
Dermatoscopic Ulceration			
Absent	Ref.		
1-49	1.38	0.168	0.84 – 2.79
>50	3.22	0.016	1.34 – 8.38
Regression structures			
Absent	Ref.		
1-49	0.81	0.193	0.42 – 1.67
>50	0.44	0.025	0.14 – 0.81
Blue – white veil	5.90	<0.001	3.06 – 12.68

OR: Odds ratio, CI: confidence interval, Ref.: Reference level. Pigmentation, dermatoscopic ulceration and regression structures are expressed as percentages of feature extent in the lesion. Results from last step of multivariable analysis after backward elimination were provided and Model 1 was adjusted for age, sex, anatomic location and histopathologic subtype of the primary tumor. 291 patients had available data about histopathologic subtype. P-values were two-sided and obtained from Wald

tests. Exact p-values are provided for pigmentation 75% ($p=1.30 \times 10^{-4}$), pigmentation 100% ($p=1.62 \times 10^{-8}$) and blue-white veil ($p=9.86 \times 10^{-7}$).

Reviewer #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for the effort reviewing our manuscript and for the thoughtful and constructive feedback.

Reviewer #5 (Remarks to the Author):

The authors have satisfactorily addressed my comments. Lallas et al. conducted an international, multicenter retrospective study to investigate the association between dermatoscopic features of primary melanoma and metastasis. The authors noted the limitations of this work. If these findings are validated in larger prospective studies in the future, the clinical applications will be something to look forward to. It is suggested that dermoscopy could potentially serve as a non-invasive and convenient tool for early prediction of metastasis risk in patients with primary melanoma, thereby providing valuable guidance for subsequent treatment decisions.

We appreciate reviewer's supportive assessment and constructive feedback that strengthened the manuscript. We agree that prospective multicenter validation is the next step toward clinical translation, and we highlighted this in the Discussion section, while clearly acknowledging the limitations of our work.