

The impact of stage-related features in melanoma recurrence prediction: A machine learning approach



To the Editor: The American Joint Committee on Cancer 8th Edition (AJCC8) stages localized melanomas by tumor thickness and ulceration.¹ However, the specific role of tumor thickness and ulceration in early-stage melanoma recurrence prediction using machine learning remains understudied. We leveraged a multi-institutional cohort of early-stage melanomas to evaluate the impact of stage-related features (thickness, ulceration, anatomic level, and clinical stage) and other clinicopathologic features in recurrence prediction.

We identified a retrospective cohort of 1166 (229 recurrences vs 937 nonrecurrences) stage I/II primary melanomas diagnosed between 2000 and 2020 at Mass General Brigham (original cohort). We extracted 11 clinicopathologic features: sex, race, age at diagnosis, AJCC8 clinical stage, tumor thickness, ulceration, anatomic level, histologic type, tumor anatomic location, mitotic rate, and total surgical margins. The following histologic subtypes were included: superficial spreading, lentigo

maligna, nodular, and malignant melanoma, not otherwise specified. Based on clinical guidelines,² melanomas where sentinel lymph node biopsy was indicated but not performed were excluded from this study. Melanomas with any unknown stage-related features were also excluded. A manual chart review by 2 independent reviewers was conducted to ascertain the first recurrence. A minimum of 5-year follow-up was used to ensure sufficient time to observe a recurrence. Melanomas that were stage IV at the time of recurrence, based on AJCC8, were labeled as “distant,” and other recurrent melanomas were labeled as “regional.” We also conducted analyses where nonrecurrences were randomly selected to match the number of recurrences in each machine learning experiment (matched cohort).

We experimented with 2 classification tasks: (1) recurrence versus nonrecurrence, and (2) distant versus regional recurrence by applying 3 well-known machine learning algorithms: random forest, gradient boosting, and logistic regression.³ Model parameters were optimized by cross-validated grid search and evaluated by 50 repeated 5-fold

Table I. Best performance of machine learning models in melanoma recurrence prediction with the original cohort and the matched cohort

	Original cohort			Matched cohort		
	BACC	AUC	PPV	BACC	AUC	PPV
	Recurrence versus nonrecurrence (229 vs 937)			Recurrence versus nonrecurrence (229 vs 229)		
Stage-related features* (mean and 95% CI)	0.657 0.653-0.661	0.805 0.801-0.809	0.662 0.652-0.672	0.733 0.728-0.739	0.804 0.798-0.810	0.737 0.731-0.744
All extracted features (mean and 95% CI)	0.687 0.685-0.688	0.831 0.826-0.836	0.694 0.684-0.703	0.756 0.749-0.763	0.820 0.814-0.825	0.765 0.758-0.772
<i>P</i> value [†]	<.001	<.001	<.001	.024	<.001	<.001
	Distant versus regional (117 vs 112)			Distant versus regional (112 vs 112)		
Stage-related features (mean and 95% CI)	0.539 0.531-0.54	0.540 0.531-0.548	0.528 0.519-0.536	0.525 0.518-0.533	0.540 0.531-0.549	0.516 0.509-0.524
All extracted features (mean and 95% CI)	0.585 0.578-0.593	0.612 0.604-0.619	0.585 0.577-0.592	0.590 0.583-0.597	0.617 0.610-0.625	0.589 0.581-0.597
<i>P</i> value [†]	<.001	<.001	<.001	<.001	<.001	<.001

AUC, Area under the receiver operating characteristic curve; BACC, balanced accuracy; CI, confidence interval; PPV, positive predictive value.

*Stage-related features: thickness, ulceration, anatomic level, and clinical stage.

[†]*P* value: *t* test for comparing the results when only using the stage-related features and when using all extracted features.

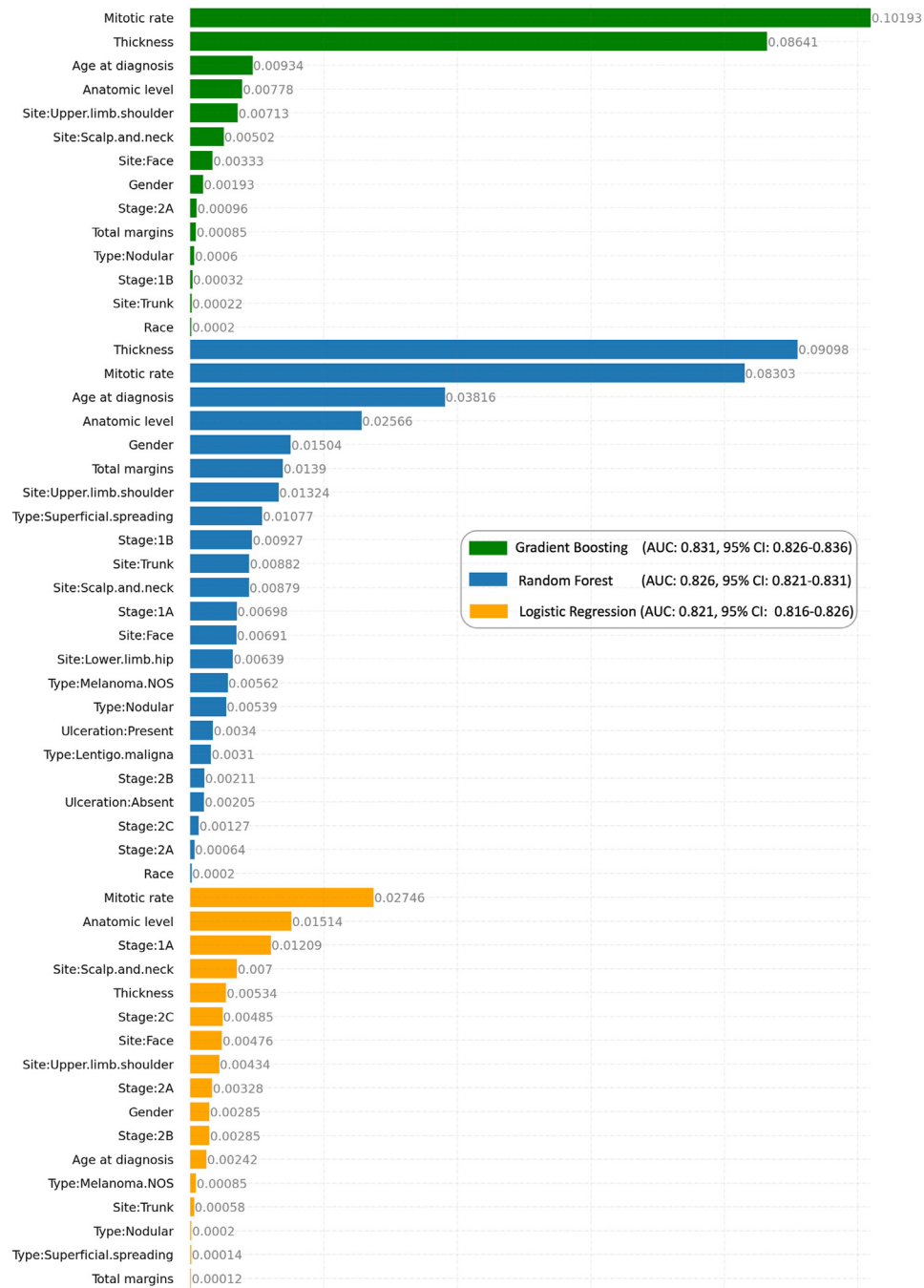


Fig 1. The ranked average feature importance in the recurrence versus nonrecurrence prediction by the 3 machine learning models. The experiments were conducted on the original cohort. Categorical features were converted by one-hot encoding. Features with zero importance were ignored. All extracted features were presented for the random forest model. *AUC*, Area under the receiver operating characteristic curve; *CI*, confidence interval.

cross-validations using 3 criteria: balanced accuracy (BACC),³ area under the receiver operating characteristic curve (AUC), and positive predictive value (PPV).

In the first task, the achieved AUC (0.805) was consistent with previous studies,⁴ but the BACC

(0.657) and PPV (0.662) were limited (Table D). The BACC and PPV on the matched cohort were better (BACC: 0.733; PPV: 0.737; $P < .001$) than those on the original cohort. In the second task, significant performance deterioration was observed in both cohorts ($P < .001$). We ranked features by conducting

permutation importance³ with 50 repeats and AUC for scoring (Fig 1). Mitotic rate, which is no longer a criterion used in the AJCC8, appeared more important than ulceration in all 3 models.

In summary, we collected clinicopathologic features currently used to inform clinical decisions in surveillance and therapeutic planning for early-stage melanomas and compared their roles in predicting melanoma recurrence. Despite near-universal dependence of clinical management on the stage-related features, our results demonstrate that the prediction performance of these features has limitations. Although stage-related features play an important role in recurrence risk stratification, entirely relying on these features for therapeutic planning will lead to many missed recurrent cases and delayed treatments. Further studies with potentially predictive data, such as genomics and digital histopathology,⁵ are needed to improve recurrence risk stratification of early-stage melanomas.

Guibong Wan, PhD,^{a,b,c} Bonnie Leung, BS,^a Nga Nguyen, MD, MPH,^a Mia S. DeSimone, MD, MPH,^a Feng Liu, PhD,^e Min Seok Choi, MS,^a Diane Ho, RHIA, CTR,^a Valerie Laucks, BS,^a Stacey Duey, MCPH,^a Ryan J. Sullivan, MD,^f Genevieve M. Boland, MD, PhD,^g Nicole R. LeBoeuf, MD, MPH,^b David Liu, MD, MPH,ⁱ Alexander Gusev, PhD,ⁱ Shawn G. Kwatra, MD,^j Peter K. Sorger, PhD,^c Kun-Hsing Yu, MD, PhD,^{b,e} and Yevgeniy R. Semenov, MD, MA^{a,c}

From the Department of Dermatology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts^a; Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts^b; Department of Systems Biology, Harvard Medical School, Boston, Massachusetts^c; Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts^d; School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, New Jersey^e; Department of Oncology, Massachusetts General Hospital, Boston, Massachusetts^f; Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts^g; Department of Dermatology, Brigham and Women's Hospital, Boston, Massachusetts^b; Department of Medicine, Dana-Farber Cancer Institute, Boston, Massachusettsⁱ; and Department of Dermatology, Johns Hopkins University, Baltimore, Maryland.^j

Drs Wan and Leung are co-first authors.

Drs Yu and Semenov are co-senior authors.

Funding sources: K-H.Y. is supported in part by R35GM142879 from the National Institute of General Medical Sciences, NIH. Y.R.S. is supported in part by the Department of Defense under award number W81XWH2110819 and by the Dermatology Foundation under the Medical Dermatology Career Development Award. The other authors received no funding for this research.

IRB approval status: Reviewed and approved by Mass General Brigham Institutional Review Boards (Protocol # 2020P002179).

Key words: early-stage melanoma; machine learning; mitotic rate; recurrence prediction; stage; tumor thickness; ulceration.

Correspondence to: Yevgeniy R. Semenov, MD, MA, Department of Dermatology, Massachusetts General Hospital, Harvard Medical School, 40 Blossom Street, Bartlett Hall 6R, Room 626, Boston, MA 02114.

E-mail: ysemenov@mgh.harvard.edu

Conflicts of interest

Y.R.S. is an advisory board member/consultant and has received honoraria from Incyte Corporation, Castle Biosciences, Galderma, and Sanofi outside of the submitted work.

REFERENCES

1. Gershenwald JE, Scolyer RA. Melanoma staging: American Joint Committee on Cancer (AJCC) and beyond. *Ann Surg Oncol*. 2018;25(8):2105-2110.
2. Wong SL, Faries MB, Kennedy EB, et al. Sentinel lymph node biopsy and management of regional lymph nodes in melanoma: American Society of Clinical Oncology and Society of Surgical Oncology clinical practice guideline update. *Ann Surg Oncol*. 2018;25(2):356-377.
3. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J machine Learn Res*. 2011;12:2825-2830.
4. El Sharouni MA, Ahmed T, Varey AH, et al. Development and validation of nomograms to predict local, regional, and distant recurrence in patients with thin (T1) melanomas. *J Clin Oncol*. 2021;39(11):1243-1252.
5. Wan G, DeSimone M, Liu F, et al. 649 CNN-based histopathology image analysis for early-stage melanoma recurrence. *J Invest Dermatol*. 2022;142(8):S112.

<https://doi.org/10.1016/j.jdin.2022.08.014>