EUROPEAN UROLOGY ONCOLOGY xxx (xxxx) xxx–xxx

available at www.sciencedirect.com
journal homepage: euoncology.europeanurology.com

European Association of Urology

Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer

Ali Sabbagh[a], *Samuel L. Washington 3rd*[b,c], *Derya Tilki*[d,e], *Julian C. Hong*[a], *Jean Feng*[c], *Gilmer Valdes*[a], *Ming-Hui Chen*[f], *Jing Wu*[g], *Hartwig Huland*[d], *Markus Graefen*[d], *Thomas Wiegel*[h], *Dirk Böhmer*[i], *Janet E. Cowan*[b], *Matthew Cooperberg*[b,c], *Felix Y. Feng*[a,b], *Mack Roach 3rd*[a], *Bruce J. Trock*[j], *Alan W. Partin*[k], *Anthony V. D'Amico*[l], *Peter R. Carroll*[b], *Osama Mohamad*[a,b,*]

[a] *Department of Radiation Oncology, University of California-San Francisco, San Francisco, CA, USA;* [b] *Department of Urology, University of California-San Francisco, San Francisco, CA, USA;* [c] *Department of Epidemiology and Biostatistics, University of California-San Francisco, San Francisco, CA, USA;* [d] *Martini-Klinik Prostate Cancer Center, University Hospital Hamburg-Eppendorf, Hamburg, Germany;* [e] *Department of Urology, University Hospital Hamburg-Eppendorf, Hamburg, Germany;* [f] *Department of Statistics, University of Connecticut, Storrs, CT, USA;* [g] *Department of Computer Science and Statistics, University of Rhode Island, Kingston, RI, USA;* [h] *Department of Radio Oncology, University Hospital Ulm, Ulm, Germany;* [i] *Department of Radiation Oncology, Charité University Hospital, Berlin, Germany;* [j] *Division of Epidemiology, Brady Urological Institute, Johns Hopkins Medical Institution, Baltimore, MD, USA;* [k] *Department of Urology, Brady Urological Institute, Johns Hopkins Medical Institution, Baltimore, MD, USA;* [l] *Department of Radiation Oncology, Brigham and Women's Hospital and Dana Farber Cancer Institute, Boston, MA, USA*

Article info

Article history:
Received 13 September 2022
Received in Revised form
10 January 2023
Accepted 3 February 2023

Associate Editor:
Alberto Briganti

Statistical Editor:
Andrew Vickers

Keywords:
Prostate cancer
Lymph node metastasis
Machine leaning

Abstract

Background: Pelvic lymph node dissection (PLND) is the gold standard for diagnosis of lymph node involvement (LNI) in patients with prostate cancer. The Roach formula, Memorial Sloan Kettering Cancer Center (MSKCC) calculator, and Briganti 2012 nomogram are elegant and simple traditional tools used to estimate the risk of LNI and select patients for PLND.

Objective: To determine whether machine learning (ML) can improve patient selection and outperform currently available tools for predicting LNI using similar readily available clinicopathologic variables.

Design, setting, and participants: Retrospective data for patients treated with surgery and PLND between 1990 and 2020 in two academic institutions were used.

Outcome measurements and statistical analysis: We trained three models (two logistic regression models and one gradient-boosted trees–based model [XGBoost]) on data provided from one institution (n = 20 267) with age, prostate-specific antigen (PSA) levels, clinical T stage, percentage positive cores, and Gleason scores as inputs. We externally validated these models using data from another institution (n = 1322) and compared their performance to that of the traditional models using the area under the receiver operating characteristic curve (AUC), calibration, and decision curve analysis (DCA).

Results and limitations: LNI was present in 2563 patients (11.9%) overall, and in 119 patients (9%) in the validation data set. XGBoost had the best performance among all the models. On external validation, its AUC outperformed that of the Roach formula by 0.08 (95% confidence interval [CI] 0.042–0.12), the MSKCC nomogram by 0.05

* Corresponding author. Department of Radiation Oncology, University of California-San Francisco, 1825 4th Street, San Francisco, CA 94158, USA. Tel. +1 415 476 9065.
E-mail address: osama.mohamad@ucsf.edu (O. Mohamad).

https://doi.org/10.1016/j.euo.2023.02.006
2588-9311/© 2023 European Association of Urology. Published by Elsevier B.V. All rights reserved.

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006

(95% CI 0.016–0.070), and the Briganti nomogram by 0.03 (95% CI 0.0092–0.051; all $p < 0.05$). It also had better calibration and clinical utility in terms of net benefit on DCA across relevant clinical thresholds. The main limitation of the study is its retrospective design.

Conclusions: Taking all measures of performance together, ML using standard clinicopathologic variables outperforms traditional tools in predicting LNI.

Patient summary: Determining the risk of cancer spread to the lymph nodes in patients with prostate cancer allows surgeons to perform lymph node dissection only in patients who need it and avoid the side effects of the procedure in those who do not. In this study, we used machine learning to develop a new calculator to predict the risk of lymph node involvement that outperformed traditional tools currently used by oncologists.

© 2023 European Association of Urology. Published by Elsevier B.V. All rights reserved.

1. Introduction

Pelvic lymph node dissection (PLND) remains the gold standard for evaluation of lymph node involvement (LNI) and staging for patients with prostate cancer for whom radical prostatectomy is planned [1]. In contemporary cohorts, LNI is detected in up to 12% of patients undergoing PLND [2], meaning that the majority of patients undergoing the procedure will have pathologically negative nodes [3]. Although PLND provides valuable staging information, its therapeutic benefits are not clear and it is associated with nontrivial complications such as lymphocele development, thromboembolic events, and, less commonly, obturator nerve, vascular, and ureteric injury [4,5]. Thus, the optimal role of PLND remains a subject of ongoing debate.

The European Association of Urology (EAU) and European Society for Radiotherapy and Oncology recommend extended PLND (ePLND) for patients with a risk of LNI of >5% according to the Briganti 2012 nomogram [6]. The American Urological Association, American Society for Radiation Oncology, and Society of Urologic Oncology recommend PLND for patients with unfavorable intermediate- and high-risk disease [7]. Finally, the National Comprehensive Cancer Network guidelines recommend ePLND for patients with a risk of LNI of >2% [8]. Several elegant tools have been developed for prediction of the risk of LNI, with the Memorial Sloan Kettering Cancer Center (MSKCC) calculator, Roach formula, and Briganti nomogram used most widely [3,9,10]. While recent updates to these tools have yielded improvements in performance, they require additional imaging and pathologic data that may not be available in routine clinical practice.

Machine learning (ML) is well suited for the development of prediction tools owing to its ability to efficiently process vast amounts of data and discover complex relationships among variables [11]. ML holds much promise in health care, with numerous recent applications [12]. The purpose of this study was to develop an ML model to estimate the risk of LNI in patients with prostate cancer referred for radical prostatectomy and to investigate—through external validation—whether this model can outperform currently available nomograms using the same clinicopathologic variables.

2. Patients and methods

2.1. Study cohorts

The study population comprised patients from two academic institutions: University Medical Center Hamburg-Eppendorf in Hamburg, Germany (UKE) and University of California-San Francisco, San Francisco, CA (UCSF). All information was collected and submitted by the respective institutions. An institutional review board approved the study at each site. We report our methodology and findings according to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline [13]. Patients aged ≥18 yr were consecutively treated with radical prostatectomy and PLND between 1990 and 2020. Further details on patient selection are provided in the Supplementary material. Patients who received neoadjuvant hormone therapy, those with clinical stage T4 disease, and patients with no biopsy or outcome (nodal dissection) information were excluded from the analysis. Furthermore, owing to the inclusion criteria of previous nomograms, patients who had prostate-specific antigen (PSA) >50 ng/ml or <0.1 ng/ml were also excluded from the external validation data set [10,14].

2.2. Model development

The outcome of interest was positive lymph nodes on surgical pathology after radical prostatectomy (pN+). To ensure ease of use in routine practice, we intentionally selected variables that would be readily available to clinicians without the need for extensive pathologic information or magnetic resonance imaging (MRI) data. We therefore only included the following independent variables: age at diagnosis, PSA at diagnosis or before treatment (in ng/ml), percentage positive cores on biopsy, primary and secondary biopsy Gleason scores, and clinical T stage. Biopsy Gleason scores ranged from 1 to 5. To minimize the impact of varying definitions and practices over time and between the two institutions, clinical T stage was encoded as 0 for patients with cT1–2 and 1 for patients with cT3 disease. Our models were trained and internally validated using the data set provided by UKE, the larger data set, to increase accuracy and generalizability [13]. The data set provided by UCSF was used to externally validate the performance of our models

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006

and to compare their performance to that of the reference models.

Details on model development and prediction calculation are provided in the Supplementary material.

2.3. Model comparison and statistical analysis

Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC) [13,15]. The 95% confidence intervals (CIs) and p values for the difference in AUC between any two models were obtained using the method described by DeLong et al [16]. This was performed for internal validation (using 10-fold cross validation) and external validation. Calibration was compared using the guidelines of Van Calster et al [17,18]. Clinical utility, defined as net benefit (NB), was compared using decision curve analysis (DCA) [19] at the clinically relevant thresholds of 2%, 5%, and 7%.

The model with the best performance for the full validation data set was then validated using patients specifically treated after 2005 and after 2012. As the International Society for Urological Pathology (ISUP) guidelines for Gleason grading were updated in 2005, and UCSF adopted a more extended approach to PLND around 2012, these analyses would validate our results for cohorts enriched with patients graded using the updated ISUP guidelines, and patients treated with ePLND, respectively.

Statistical and performance analyses were conducted using R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria; references are provided in the Supplementary material). All tests were two-sided, with the significance level set at $p < 0.05$.

3. Results

3.1. Participants

The training and validation data sets consisted of 20 267 and 1322 patients, respectively. CONSORT diagrams are shown in Supplementary Figures 1 and 2 and patient characteristics are summarized in Table 1. In the validation data set, 110 patients had data missing for the percentage positive cores on biopsy and were excluded from the analysis. A comparison of patients with and without missing data in the validation set is shown in Supplementary Table 3. Supplementary Table 4 compares patients with and without LNI in the combined data set.

Table 1 – Baseline patient characteristics for the training and external validation sets

	Training	Validation	p value [a]
Patients (n)	20 267	1322	
Median age, yr (IQR)	64 (59–69)	62 (57–67)	<0.001
Type of surgery, n (%)			<0.001
Open radical prostatectomy	12 946 (64)	484 (37)	
Robotic surgery	7321 (36)	821 (63)	
Data missing		17	
Median PPC, % (IQR)	40 (23–58)	41 (25–58)	0.4
Biopsy Gleason primary, n (%)			<0.001
1	0 (0.0)	1 (0.076)	
2	22 (0.11)	18 (1.4)	
3	13 100 (65)	586 (44)	
4	6571 (33)	658 (50)	
5	460 (2.3)	59 (4.5)	
Data missing	114		
Biopsy Gleason secondary, n (%)			<0.001
2	26 (0.13)	15 (1.1)	
3	8 108 (40)	607 (46)	
4	10 723 (53)	581 (44)	
5	1293 (6.4)	119 (9.0)	
Data missing	117		
Median PSA before treatment, ng/ml (IQR)	7.88 (5.48–12.0)	7.40 (5.20–11.3)	<0.001
LN involvement (pN+), n (%)	2 444 (12)	119 (9.0)	0.001
Median LNs removed, n (IQR)	13 (8–20)	11 (6–17)	<0.001
Clinical T stage, n (%)			<0.001
cT1	2 (0.0)	0 (0.0)	
cT1a	1 (0.0)	0 (0.0)	
cT1b	4 (0.0)	0 (0.0)	
cT1c	14 867 (75)	296 (22)	
cT2	7 (0.0)	607 (46)	
cT2a	3142 (16)	125 (9.5)	
cT2b	1391 (7.0)	47 (3.6)	
cT2c	314 (1.6)	31 (2.3)	
cT3	40 (0.19)	0 (0.0)	
cT3a	130 (0.64)	196 (15)	
cT3b	9 (0.0)	20 (1.5)	
cTx	1 (0.0)	0 (0.0)	
Data missing	359		

IQR = interquartile range; LN = lymph node; PPC = percentage positive cores on biopsy; PSA = prostate-specific antigen.
[a] p values were obtained using the Kruskal-Wallis rank-sum test for continuous variables and the χ^2 test for categorical variables.

3.2. Feature importance

Supplementary Table 5 shows the final optimal XGBoost model hyperparameters. Figure 1 shows a Shapley additive explanation (SHAP) summary plot for the XGBoost ensemble model. The coefficients and intercepts for the LR models are available in Supplementary Tables 6 and 7.

3.3. Model performance

3.3.1. Internal validation

Internal validation on the training set was performed using tenfold cross validation. All three models (LR models and XGBoost) had excellent discrimination and calibration. The average AUCs were 0.82 (95% CI 0.81–0.83) for XGBoost and 0.81 (95% CI 0.80–0.82) for both LR models. Calibration curves are shown in Supplementary Figures 3–5.

3.3.2. External validation

3.3.2.1. Discrimination. As shown in Table 2, the standard LR, ensemble LR, and XGBoost models significantly outperformed traditional models in terms of discrimination.

3.3.2.2. Calibration. Our models had good calibration for the external validation set, with slopes close to 1 (standard LR 1.10, ensemble LR 1.06, XGBoost 1.09) and intercepts close to 0 (standard LR −0.548, ensemble LR −0.622, XGBoost −0.509). The Roach formula (slope 1.36, intercept −0.425) and the MSKCC (slope 0.896, intercept −0.878) and Briganti (slope 0.786, intercept −1.01) nomograms were not as well calibrated (Supplementary Figs. 6–11).

3.3.3. Decision curve analysis

As shown in Figure 2, XGBoost provided the highest NB at the clinically relevant thresholds. Figure 3 shows the equiv-

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006

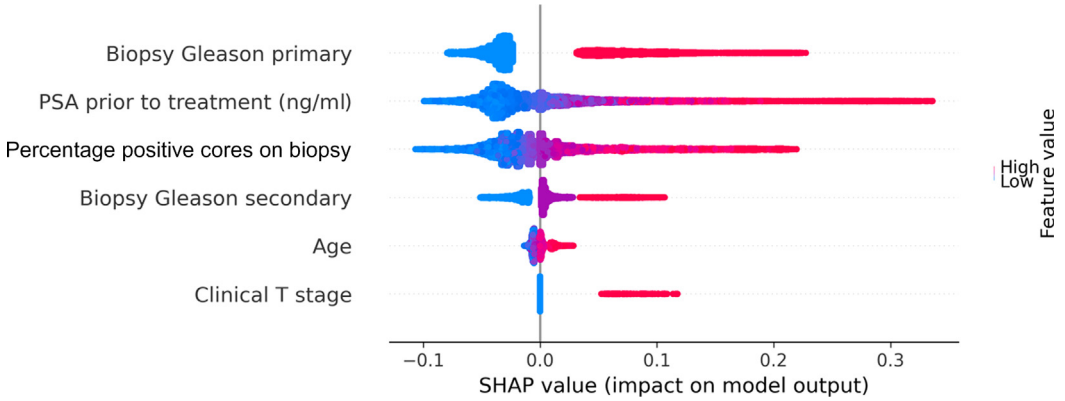

Fig. 1 – Shapley additive explanation (SHAP) summary plot showing XGBoost feature importance and contributions. Red represents higher feature values while blue represents lower values. PSA = prostate-specific antigen.

Table 2 – Discrimination measure and comparison of the different models

Model	AUC	ΔAUC					
		Standard LR		Ensemble LR		XGBoost	
		95% CI	*p* value [a]	95% CI	*p* value [a]	95% CI	*p* value [a]
Roach	0.74	0.029–0.11	<0.001	0.036–0.11	<0.001	0.042–0.12	<0.001
MSKCC	0.77	0.0057–0.061	0.018	0.013–0.065	0.003	0.016–0.070	0.002
Briganti 2012	0.79	0.0030–0.038	0.022	0.0079–0.044	0.005	0.0092–0.051	0.005
Standard LR	0.81	–		0.00044–0.011	0.033	−0.0025 to 0.022	0.12
Ensemble LR	0.81	−0.011 to 0.00044	0.033	–		−0.0064 to 0.015	0.4
XGBoost	0.82	−0.022 to 0.0025	0.12	−0.015 to 0.0064	0.4	–	

AUC = area under the receiver operating characteristic curve; ΔAUC = difference in AUC; CI = confidence interval; LR logistic regression.
[a] *p* value for the hypothesis H0: ΔAUC = 0.

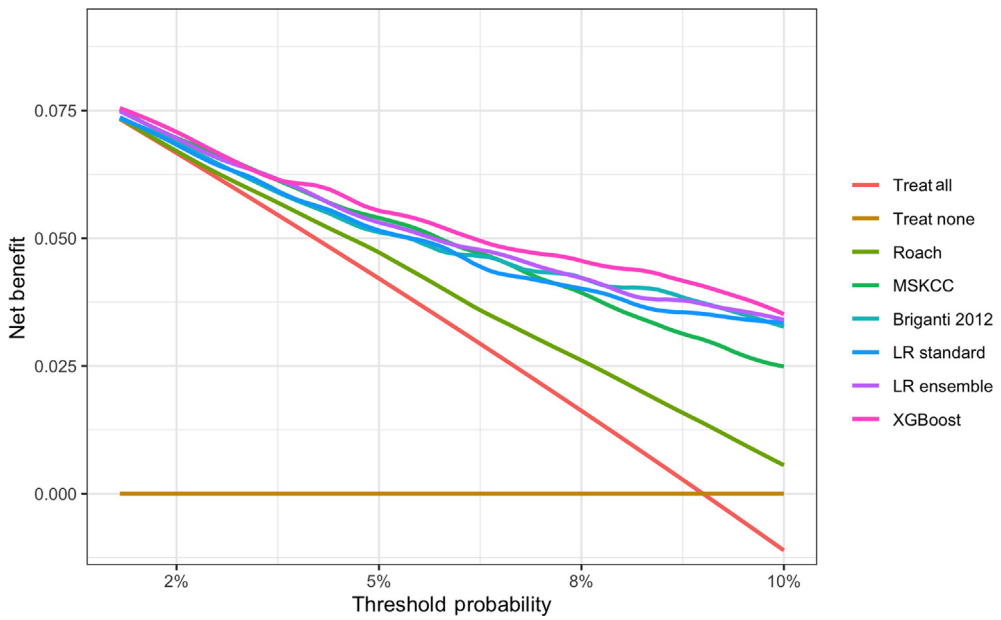

Fig. 2 – Decision curve analysis showing the net benefit of all models. A model outperforms another if it has a higher net benefit. An increase in net benefit of *x* % is equivalent to a strategy that detects lymph node involvement in *x* more patients in 100 without increasing the number of false positives. LR = logistic regression; MSKCC = Memorial Sloan Kettering Cancer Center.

alent number of unnecessary interventions avoided per 100 patients without missing any additional cases using each prediction tool at various thresholds [19], from which the superiority of the XGBoost model is evident.

4. Discussion

PLND remains the gold standard for nodal staging in patients with clinically localized prostate cancer undergo-

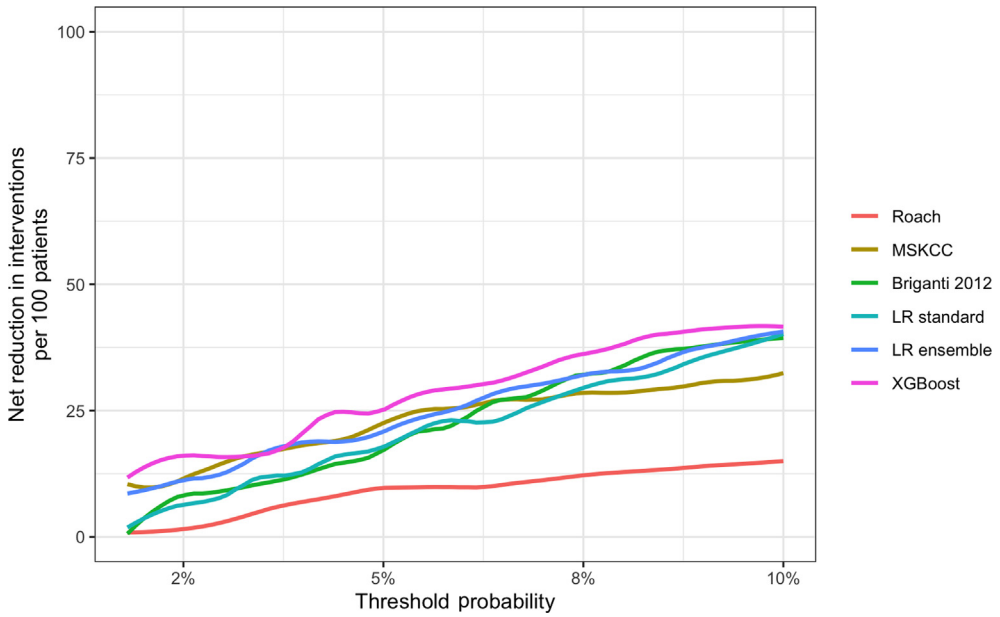

Fig. 3 – Net interventions avoided per 100 patients for all models. LR = logistic regression; MSKCC = Memorial Sloan Kettering Cancer Center.

ing radical prostatectomy [6,8]. Given the risk of complications associated with the procedure, it is crucial to appropriately select the patients who are most likely to benefit, while avoiding the unnecessary risk in those unlikely to harbor lymph node metastases. According to multiple published guidelines, the optimal risk cutoff varies; 2%, 5%, and 7% are commonly cited thresholds [6,8,20]. Our goals were (1) to develop and externally validate an ML model to predict LNI in patients with prostate cancer and (2) to investigate whether the ML model can outperform currently available nomograms using the same simple clinicopathologic variables. On external validation, our models exhibited better discrimination and calibration than the traditional models, whereas XGboost had similar discrimination and calibration but better NB than the LR models. We intentionally trained the LR models on our data to ensure that any increase in performance observed with XGBoost is in fact due to the capabilities of ML. Since our LR models outperformed traditional models in terms of discrimination and calibration, it is likely that this increase in performance is related to the training data set used rather than the choice of model. This is not a surprising finding, as our data set contained ~20 000 patients, whereas the MSKCC and Briganti nomograms were built using cohorts of ~12 000 and ~600 patients, respectively. However, discrimination and calibration are not the only evaluation metrics used. Indeed, XGBoost was the only model that outperformed all other nomograms in terms of clinical utility, even though the ensemble LR model was developed using the same pipeline as for XGBoost. It is clear, therefore, that while some of the increase in performance is attributable to the size of the training data and some to modeling techniques, there remains an increase in predictive performance that was only seen with XGBoost that is not attributable to either factor and is due to the predictive capabilities of ML. Finally, one might argue that the superior performance of the ML model observed on external validation might be because the UCSF and UKE data sets share higher similarity in baseline characteristics in comparison to the similarity shared between the UCSF set and the data sets used to train other models (eg, MSKCC). This is unlikely, however, as UCSF and MSKCC are both US institutions and are therefore more likely to follow similar treatment guidelines and practices than UCSF and UKE are. Furthermore, the UCSF and UKE cohorts have significantly different baseline characteristics (Table 1), making it even less likely that this can explain the improvement in performance. One caveat to keep in mind, however, is that a model developed at a specific institution is likely to exhibit better performance when used on patients from that institution. For example, the prediction of the MSKCC nomogram for a patient treated at MSKCC is likely to be more accurate than the XGBoost prediction.

The improvement in clinical utility means that XGBoost reduces the number of unnecessary interventions. For example, for every 100 patients screened using the 2% threshold, the difference in performance between XGBoost and the MSKCC nomogram is the equivalent of a scenario in which XGBoost would spare approximately five additional patients from undergoing PLND in comparison to the MSKCC nomogram, without missing any patients with LNI. This reduction in unnecessary procedures confers a decrease in the rate of adverse events and possibly monetary savings. As the ISUP guidelines were updated in 2005, an analysis was performed for patients diagnosed after 2005 (Supplementary Tables 10 and 11 and Supplementary Fig. 13). The results show that XGBoost maintained its superior performance, even with changes in clinical practice. Furthermore, our post-2012 analysis validates our results for patients treated specifically with ePLND (Supplementary Tables 8 and 9 and Supplementary Fig. 12).

The Briganti nomogram has been updated twice in recent years. The 2017 update improved performance at

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006

the cost of additional variables, that the authors emphasize are not readily available in daily practice [21]. By contrast, the 2019 update was developed using patients who underwent multiparametric MRI (mpMRI)-targeted biopsy [22]. This update was externally validated by the EAU Prostate Cancer Working Group and showed improvements in discrimination in comparison to the 2012 and 2017 nomograms [23] but was not superior to the 2012 nomogram in a different study [24]. Unfortunately, these additional variables are not available in our data sets so we could not compare our models to the newer Briganti nomograms. The main strength of our model is therefore its ability to provide better performance using simple variables in comparison to the updated versions of the Briganti nomogram, which rely on additional data such as MRI-targeted biopsy and detailed biopsy information. Thus, our model retains the ease of use afforded by other tools while significantly improving the performance. In addition, our model was developed using patients from Germany and externally validated for patients from the USA, further supporting its generalizability.

Other studies have used ML to calculate the risk of LNI in patients with prostate cancer. Wei et al [25] used the Surveillance, Epidemiology and End Results database to identify 2884 patients who underwent ePLND and to develop different ML models. Their XGBoost model had the highest accuracy and better discrimination than the Briganti nomogram. Unfortunately, it was not externally validated. Hou et al [26] used an institutional database of only 248 patients to develop ML algorithms to predict LNI using clinicopathologic variables and mpMRI-related features. The ML models had superior performance to the MSKCC nomogram and addition of mpMRI features significantly improved the performance. However, this work was limited by its extremely small sample size, its dependence on mpMRI features, and the absence of external validation. Lastly, Ötleş et al [27] developed an LR model to predict LNI in patients with prostate cancer. Their ML model showed excellent performance on external validation in terms of discrimination and calibration. However, while their model had better calibration than the Briganti and MSKCC tools, it was similar in terms of discrimination.

Our study is not without limitations. First, XGBoost is not as easy to interpret as LR or traditionally developed models. While SHAP values provide important insights into the model's decision-making process, they cannot be used to recreate the probability estimates of the model. Nonetheless, we aim to make our model available through a web-based interface so that clinicians will be able to easily use it in their practice. Second, as we lacked the number of positive and negative biopsy cores in our validation set, we could not compare our model to the MSKCC model that incorporates these predictors. Third, the number of lymph nodes removed differed significantly between patients with and without LNI. This could potentially have biased our results, as more patients with LNI might have been missed. Fourth, the criteria for PLND were not consistent across institutions, potentially introducing an element of selection bias into our results. Fifth, the median percentage positive cores was relatively higher for our patients than for other cohorts. Sixth, many changes occurred in the screening for and management of prostate cancer during the study period, including changes in PSA screening guidelines, surgical approaches, and Gleason grading, among others. Such changes in practice may result in "data drift", whereby the performance of an ML model degrades over time. While our post-2005 and post-2012 analyses show that this is not a concern for the predictive performance of our model for contemporary patients, major changes in practice may affect this finding. Nonetheless, this limitation applies to all predictive models, including conventional tools such as the MSKCC and Briganti nomograms. Seventh, our data sets did not include MRI or prostate-specific membrane antigen positron emission tomography data. This is a potential drawback, as guidelines currently recommend use of such imaging in the initial workup for patients with prostate cancer. Finally, our study was conducted using retrospective data, and thus is at risk of the biases inherent to such studies, namely missing data, inconsistent reporting, and incorrect data entry, among others.

5. Conclusions

In conclusion, we developed and validated an ML model that outperforms the Roach formula, the MSKCC calculator, and the Briganti 2012 nomogram in predicting LNI using similar clinicopathologic features, which will facilitate better patient selection.

Author contributions: Osama Mohamad had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Mohamad, Sabbagh, J. Feng, Tilki, Carroll, D'Amico.
Acquisition of data: Tilki, Huland, Böhmer, Graefen, Wiegel, Partin, Wu, Cowan, Carroll, Cooperberg, Chen, D'Amico, F.Y. Feng, Trock, Mohamad.
Analysis and interpretation of data: Sabbagh, Mohamad, J. Feng, Hong, Valdes, Huland, Böhmer, Graefen, Wiegel, Partin, Wu, Cooperberg, Chen, F.Y. Feng, Trock.
Drafting of the manuscript: Sabbagh, Mohamad, J. Feng, Hong.
Critical revision of the manuscript for important intellectual content: Mohamad, Sabbagh, J. Feng, Hong, Washington, Valdes.
Statistical analysis: Sabbagh, J. Feng, Hong, Valdes.
Obtaining funding: Mohamad.
Administrative, technical, or material support: Cowan.
Supervision: Mohamad, Washington, Tilki, Carroll.
Other: None.

Financial disclosures: Osama Mohamad certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

***Funding/Support and role of the sponsor*:** This work was funded by the UCSF Prostate Cancer Program Pilot Research Award.

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euo.2023.02.006.

References

[1] Mottet N, van den Bergh RCN, Briers E, et al. EAU-EANM-ESTRO-ESUR-SIOG guidelines on prostate cancer—2020 update. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2021;79:243–62. https://doi.org/10.1016/j.eururo.2020.09.042.

[2] Wilczak W, Wittmer C, Clauditz T, et al. Marked prognostic impact of minimal lymphatic tumor spread in prostate cancer. Eur Urol 2018;74:376–86. https://doi.org/10.1016/j.eururo.2018.05.034.

[3] Briganti A, Larcher A, Abdollah F, et al. Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: the essential importance of percentage of positive cores. Eur Urol 2012;61:480–7. https://doi.org/10.1016/j.eururo.2011.10.044.

[4] Fossati N, Willemse PPM, Van den Broeck T, et al. The benefits and harms of different extents of lymph node dissection during radical prostatectomy for prostate cancer: a systematic review. Eur Urol 2017;72:84–109. https://doi.org/10.1016/j.eururo.2016.12.003.

[5] Lestingi JFP, Guglielmetti GB, Trinh QD, et al. Extended versus limited pelvic lymph node dissection during radical prostatectomy for intermediate- and high-risk prostate cancer: early oncological outcomes from a randomized phase 3 trial. Eur Urol 2021;79:595–604. https://doi.org/10.1016/j.eururo.2020.11.040.

[6] Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. Eur Urol 2017;71:618–29. https://doi.org/10.1016/j.eururo.2016.08.003.

[7] Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part II: recommended approaches and details of specific care options. J Urol 2018;199:990–7. https://doi.org/10.1016/j.juro.2018.01.002.

[8] National Comprehensive Cancer Network. NCCN guidelines: prostate cancer. https://www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.

[9] Roach M, Marquez C, Yuo HS, et al. Predicting the risk of lymph node involvement using the pre-treatment prostate specific antigen and Gleason score in men with clinically localized prostate cancer. Int J Radiat Oncol Biol Phys 1994;28:33–7. https://doi.org/10.1016/0360-3016(94)90138-4.

[10] Memorial Sloan Kettering Cancer Center. Dynamic prostate cancer nomogram: coefficients. https://www.mskcc.org/nomograms/prostate/pre_op/coefficients.

[11] Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol 2019;188:2222–39. https://doi.org/10.1093/aje/kwz189.

[12] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262–73. https://doi.org/10.1016/S1470-2045(19)30149-4.

[13] Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73. https://doi.org/10.7326/M14-0698.

[14] Evidencio. Briganti 2012 nomogram: "Prediction of lymph node involvement in patients with prostate cancer". https://www.evidencio.com/models/show/670.

[15] Fenlon C, O'Grady L, Doherty ML, Dunnion J. A discussion of calibration techniques for evaluating binary and categorical predictive models. Prev Vet Med 2018;149:107–14. https://doi.org/10.1016/j.prevetmed.2017.11.018.

[16] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.

[17] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016;74:167–76. https://doi.org/10.1016/j.jclinepi.2015.12.005.

[18] Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230. https://doi.org/10.1186/s12916-019-1466-7.

[19] Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18. https://doi.org/10.1186/s41512-019-0064-7.

[20] Małkiewicz B, Ptaszkowski K, Knecht K, et al. External validation of the Briganti nomogram to predict lymph node invasion in prostate cancer—setting a new threshold value. Life 2021;11:479. https://doi.org/10.3390/life11060479.

[21] Gandaglia G, Fossati N, Zaffuto E, et al. Development and internal validation of a novel model to identify the candidates for extended pelvic lymph node dissection in prostate cancer. Eur Urol 2017;72:632–40. https://doi.org/10.1016/j.eururo.2017.03.049.

[22] Gandaglia G, Ploussard G, Valerio M, et al. A novel nomogram to identify candidates for extended pelvic lymph node dissection among patients with clinically localized prostate cancer diagnosed with magnetic resonance imaging–targeted and systematic biopsies. Eur Urol 2019;75:506–14. https://doi.org/10.1016/j.eururo.2018.10.012.

[23] Gandaglia G, Martini A, Ploussard G, et al. External validation of the 2019 Briganti nomogram for the identification of prostate cancer patients who should be considered for an extended pelvic lymph node dissection. Eur Urol 2020;78:138–42. https://doi.org/10.1016/j.eururo.2020.03.023.

[24] Frego N, Paciotti M, Buffi NM, et al. External validation and comparison of two nomograms predicting the probability of lymph node involvement in patients subjected to robot-assisted radical prostatectomy and concomitant lymph node dissection: a single tertiary center experience in the MRI-era. Front Surg 2022;9:829515. https://doi.org/10.3389/fsurg.2022.829515.

[25] Wei L, Huang Y, Chen Z, et al. Artificial intelligence combined with big data to predict lymph node involvement in prostate cancer: a population-based study. Front Oncol 2021;11:763381. https://doi.org/10.3389/fonc.2021.763381.

[26] Hou Y, Bao ML, Wu CJ, Zhang J, Zhang YD, Shi HB. A machine learning-assisted decision-support model to better identify patients with prostate cancer requiring an extended pelvic lymph node dissection. BJU Int 2019;124:972–83. https://doi.org/10.1111/bju.14892.

[27] Ötleş E, Denton BT, Qu B, et al. Development and validation of models to predict pathological outcomes of radical prostatectomy in regional and national cohorts. J Urol 2022;207:358–66. https://doi.org/10.1097/JU.0000000000002230.

Please cite this article as: A. Sabbagh, S.L. Washington 3rd, D. Tilki et al., Development and External Validation of a Machine Learning Model for Prediction of Lymph Node Metastasis in Patients with Prostate Cancer, Eur Urol Oncol (2023), https://doi.org/10.1016/j.euo.2023.02.006