

Development and Validation of a Clinical Risk Model for Predicting Malignancy in Patients with Thyroid Nodules

Shiva Borzouei¹ , Ali Safdari², Erfan Ayubi^{3*} 

1. Department of Internal Medicine, School of Medicine, Hamadan University of Medical Sciences, Hamadan, Iran
2. Department of Nursing, Malayer School of Medical Sciences, Chronic Diseases (Home Care) Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
3. Cancer Research Center, Institute of Cancer, Avicenna Health Research Institute, Hamadan University of Medical Sciences, Hamadan, Iran

Article history:

Received: 21 October 2024

Revised: 04 January 2025

Accepted: 06 February 2025

ePublished: 15 March 2025

*Corresponding author: Erfan Ayubi, Cancer Research Center, Institute of Cancer, Avicenna Health Research Institute, Hamadan University of Medical Sciences, Hamadan, Iran

E-mail: aubi65@gmail.com

Abstract

Background and Objective: Thyroid cancer is the most common malignancy of the endocrine system. Clinically, it is highly important to identify patients at risk of poor prognosis based on patient characteristics and the features of thyroid nodules. The purpose of the current study was to develop and validate a clinical risk model to predict malignancy in patients with thyroid nodules.

Materials and Methods: In the analytical cross-sectional study, the data of 650 patients (mean age: 42.36 ± 13.45 years, female: 86.15%) with thyroid nodules who underwent thyroidectomy were analyzed. The samples were patients referred to the specialized endocrinology clinic between 2014 and 2022. A multivariable model was built using demographic, clinical, and Bethesda System data through logistic regression as a generalized linear model (GLM). Interval validity of the model was checked using bootstrap resampling. The discrimination, calibration, and benefits of the model were evaluated using the area under the curve (AUC), Brier score, and decision curve analysis (DCA), respectively. The diagnostic performance of the GLM was compared with five machine learning (ML) algorithms, including linear discriminant analysis (LDA), random forest, neural network, support vector machine, and k-nearest neighbor.

Results: Out of 650 operated patients, 43% were benign and 57% malignant. The age, gender, history of thyroid diseases in first-degree relatives, type of thyroid disease, thyroid nodule focality, cervical adenopathy, and Bethesda system were significant features in constructing the prediction model based on GLM. The AUC and Brier score of the model were 0.89 and 0.12, respectively. The DCA also showed that the model performed well in clinical practice. Generally, there was no difference among the six ML algorithms in terms of prognostic performance; however, the prognostic parameters of GLM and LDA algorithms were higher than the others.

Conclusion: Developing and validating ML-based prognostic models using demographic, clinical, and Bethesda data may be useful for the treatment management of patients diagnosed with thyroid nodules.

Keywords: Machine Learning, Malignancy, Thyroid Nodule, Validation

Please cite this article as follows: Borzouei Sh, Safdari A, Ayubi E. Development and Validation of a Clinical Risk Model for Predicting Malignancy in Patients with Thyroid Nodules. 2025; 31(4): 219-227 DOI: 10.32592/ajcm.31.4.219

طراحی و اعتبارسنجی یک مدل خطر بالینی برای پیش‌بینی بدخیمی در بیماران مبتلا به ندول تیروئید

شیوا برزویی^۱ ID، علی صفدری^۲، عرفان ایوبی^۳ ID*

۱. گروه داخلی، دانشکده پزشکی، دانشگاه علوم پزشکی همدان، همدان، ایران
۲. گروه پرستاری، دانشکده علوم پزشکی ملایر، مرکز تحقیقات بیماری‌های مزمن (مراقبت در منزل)، دانشگاه علوم پزشکی همدان، همدان، ایران
۳. مرکز تحقیقات سرطان، پژوهشگاه سرطان، پژوهشگاه سلامت ابن سینا، دانشگاه علوم پزشکی همدان، همدان، ایران

چکیده

سابقه و هدف: سرطان تیروئید شایع‌ترین بدخیمی سیستم اندوکرین است. از نظر بالینی بسیار مهم است که با توجه به خصوصیات بیماران و ویژگی‌های ندول تیروئید بیماران مستعد پیش‌آگهی بد شناسایی شوند. هدف از مطالعه حاضر طراحی و اعتبارسنجی یک مدل خطر بالینی برای پیش‌بینی بدخیمی در بیماران مبتلا به ندول تیروئید بوده است.

مواد و روش‌ها: در این مطالعه مقطعی تحلیلی داده‌های ۶۵۰ بیمار (میانگین سنی: $45/13 \pm 42/36$ سال، جنسیت زن: ۸۶/۱۵ درصد) مبتلا به ندول تیروئیدی که تحت تیروئیدکتومی قرار گرفته بودند، بررسی شده‌اند. نمونه‌ها، بیماران مراجعه‌کننده به کلینیک تخصصی غدد در بازه زمانی ۱۳۹۳ تا ۱۴۰۱ بوده‌اند. یک مدل چندمتغیره با استفاده از داده‌های جمعیت‌شناختی، بالینی و همچنین داده‌های سیستم بتسدا از طریق تحلیل رگرسیون لجستیک به مثابه یک مدل خطی تعمیم‌یافته (generalized linear model; GLM) ساخته شده است. اعتبار درونی مدل با استفاده از نمونه‌گیری مجدد بوت‌استرپ ارزیابی شده است. توانایی تشخیصی، کالیبراسیون و سودمندی مدل به ترتیب با استفاده از سطح زیر منحنی راک (area under the ROC curve)، نمره Brier و تحلیل منحنی تصمیم‌گیری (decision curve analysis; DCA) ارزیابی شده است. عملکرد تشخیصی مدل GLM با پنج الگوریتم یادگیری ماشین شامل آنالیز تشخیصی خطی (linear discriminant analysis; LDA)، جنگل تصادفی (random forest; RF)، شبکه عصبی (neural network; NN)، ماشین بردار پشتیبان (support vector machine; SVM) و k-نزدیک‌ترین همسایگی (k-nearest neighbor; kNN) مقایسه شده است.

یافته‌ها: از ۶۵۰ بیمار جراحی شده ۴۳٪ خوش‌خیم و ۵۷٪ بدخیم بوده‌اند. سن، جنسیت، سابقه بیماری‌های تیروئید در بستگان درجه اول، نوع بیماری تیروئید، فوکل بودن ندول تیروئید، آدنوپاتی گردنی و سیستم بتسدا ویژگی‌های مهم برای ساخت مدل پیش‌بینی بر اساس GLM بوده‌اند. مقدار AUC و نمره Brier مدل به ترتیب ۰/۸۹ و ۰/۱۲ بوده است. همچنین نتایج DCA نشان‌دهنده سودمندی بالینی مدل است. به طور کلی، از نظر عملکرد پیشگویی بین شش الگوریتم یادگیری ماشین تفاوتی وجود نداشته است؛ با این حال، پارامترهای پیشگویی الگوریتم‌های GLM و LDA از سایرین بالاتر بوده است.

نتیجه‌گیری: طراحی و اعتبارسنجی مدل پیشگویی مبتنی بر الگوریتم‌های یادگیری ماشین با استفاده از داده‌های جمعیت‌شناختی، بالینی و داده‌های بتسدا ممکن است برای مدیریت درمان بیماران مبتلا به ندول تیروئید سودمند باشد.

واژگان کلیدی: اعتبارسنجی، بدخیمی، ندول‌های تیروئید، یادگیری ماشین

تاریخچه مقاله:

دریافت: ۱۴۰۳/۰۷/۳۰

ویرایش: ۱۴۰۳/۱۰/۱۵

پذیرش: ۱۴۰۳/۱۱/۱۸

انتشار: ۱۴۰۳/۱۲/۲۵

تمامی حقوق نشر برای دانشگاه علوم پزشکی همدان محفوظ است.

* نویسنده مسئول: عرفان ایوبی، مرکز تحقیقات سرطان، پژوهشگاه سرطان، پژوهشگاه سلامت ابن سینا، دانشگاه علوم پزشکی همدان، همدان، ایران

ایمیل: aubi65@gmail.com

استناد: برزویی، شیوا؛ صفدری، علی؛ ایوبی، عرفان. طراحی و اعتبارسنجی یک مدل خطر بالینی برای پیش‌بینی بدخیمی در بیماران مبتلا به ندول تیروئید. مجله پزشکی بالینی ابن سینا، زمستان ۱۴۰۳؛ ۳۱(۴): ۲۲۷-۲۱۹

طول دهه گذشته بخش جدایی‌ناپذیر تحقیقات پیش‌آگهی و پیش‌بینی سرطان‌ها بوده است [۲۱]. با در نظر گرفتن مسائل بیان‌شده، هدف از مطالعه حاضر طراحی و اعتبارسنجی یک مدل خطر بالینی برای پیش‌بینی بدخیمی در بیماران مبتلا به ندول تیروئید با استفاده از متغیرهای جمعیت‌شناختی، بالینی و نتایج تست بتسدا است.

روش کار

جمعیت مورد مطالعه و خصوصیات بیماران

این مطالعه مقطعی تحلیلی شامل تمام بیمارانی بوده است که طی سال‌های ۱۳۹۳ تا ۱۴۰۱ با تشخیص گواتر ندولر به کلینیک غدد همدان، واقع در غرب ایران، مراجعه کرده بودند. همه بیماران ابتدا تحت آسپیراسیون سوزنی نازک و سپس تحت عمل جراحی تیروئیدکتومی قرار گرفته‌اند. آسپیراسیون سوزنی نازک توسط متخصص اندوکراین انجام و نتایج در یک مرکز دانشگاهی بر اساس سیستم بتسدا تفسیر شده است. جراحی تیروئیدکتومی بیماران نیز در بیمارستان ریفراک تخصصی انجام شده است. مشخصات افراد شامل متغیرهای جمعیت‌شناختی (سن، جنسیت، تحصیلات)، سابقه بیماری‌های تیروئید در بستگان درجه یک، نوع اختلال تیروئید (یوتیروئید، هیپوتیروئیدی بالینی و زیربالینی، هیپرتیروئیدی بالینی و زیربالینی)، علائم بالینی (خیر و بله)، کانونی بودن (یک‌کانونی یا چندکانونی)، اندازه گره (≥ 4 سانتی‌متر و < 4 سانتی‌متر)، لنفادنوپاتی گردنی (خیر و بله) و نتایج سیستم بتسدا بوده است. بتسدا یک سیستم استاندارد برای گزارش نتایج سیتولوژی آسپیراسیون تیروئید است که در شش مورد تقسیم‌بندی شده است. این شش مورد شامل غیرتشخیصی / نامناسب (nondiagnostic/unsatisfactory; ND/US)، خوش‌خیم (benign; B)، آتیپی با اهمیت نامشخص / ضایعات فولیکولی با اهمیت نامشخص (atypia of undetermined significance/follicular lesions of undetermined significance; AUS/FLUS)، نئوپلاسم فولیکولی / مشکوک به نئوپلاسم فولیکولی (follicular neoplasm/suspicious of follicular neoplasm; FN/SF)، مشکوک به بدخیمی (suspicious of malignancy; SM) و بدخیمی (malignancy; M) است. ویژگی‌های بیان‌شده پیش‌بینی‌کننده و نتایج پاتولوژی (خوش‌خیم و بدخیم) پیامد مورد مطالعه در نظر گرفته شده است.

تجزیه و تحلیل آماری

داده‌های توصیفی به صورت تعداد (درصد) بر اساس ندول‌های بدخیم یا خوش‌خیم گزارش شده و با استفاده از آزمون کای‌اسکوئر آزمایش شده‌اند. در مرحله بعد متغیرهای معنادار وارد تحلیل رگرسیون لجستیک چندمتغیره به عنوان یک مدل خطی

سرطان تیروئید شایع‌ترین بدخیمی سیستم اندوکراین است [۱]. پیش‌بینی می‌شود که تا سال ۲۰۴۰ میلادی سرطان تیروئید در گروه سنی ۲۰ تا ۴۹ سال جزو چهار سرطان شایع باشد [۲]. در حال حاضر، یک روند افزایشی در شیوع در این نوع سرطان وجود دارد؛ اما این روند در همه مناطق جهان یکسان نیست. افزایش شیوع سرطان تیروئید بر حسب کشور، جنسیت و هیستولوژی متفاوت است [۳]. اگرچه بالاترین میزان بیماری و مرگ و میر ناشی از این سرطان در کشورهای توسعه‌یافته رخ می‌دهد، اما افزایش شیوع آن در کشورها و مناطق با درآمد کم و متوسط هشداردهنده است [۴]. برای مثال، بین سال‌های ۱۹۰۰ و ۲۰۱۹، افزایش تقریبی ۴۰۰٪ در شمال آفریقا و خاورمیانه رخ داده است [۵]. در این منطقه، ایران یکی از کشورهایی است که بیشترین بروز سرطان تیروئید را دارد. در سال ۲۰۱۹، تعداد موارد جدید آن در ایران به ۳۱۹۸ مورد رسید، در حالی که این رقم در سال ۱۹۹۰ حدود ۵۳۹ مورد بوده است [۵].

همان‌طور که انتظار می‌رود، سرطان تیروئید پیش‌آگهی مناسبی دارد. با این حال، در گروهی از بیماران متاستاز رخ می‌دهد که منجر به کاهش میزان بقای پنج ساله می‌شود [۶]. متاستاز در این نوع سرطان عمدتاً تابعی از ویژگی‌های ندول تیروئید، نوع و زیرنوع هیستولوژیکی و مرحله سرطان است [۷-۹]. سیتولوژی آسپیراسیون با سوزن باریک (Fine needle aspiration cytology; FNAC) همراه با سیستم بتسدا برای گزارش سیتوپاتولوژی تیروئید، روش اصلی برای ارزیابی ندول‌های تیروئید در نظر گرفته می‌شود [۱۰-۱۲]. در حالی که FNAC روش تشخیصی استاندارد طلایی برای ندول‌های تیروئید در نظر گرفته می‌شود، نتایج ممکن است در موارد ندول‌های تیروئید بزرگ‌تر از ۴ میلی‌لیتر یا زیرنوع فولیکولی قابل اعتماد نباشد [۱۳]. ارزیابی بدخیمی و تهاجمی ضایعات مشکوک مستلزم بررسی هیستوپاتولوژیکی پس از تیروئیدکتومی است [۱۴]. اگرچه ندول‌های تیروئید عمدتاً خوش‌خیم هستند، شناسایی بدخیمی در آن‌ها چالش‌برانگیز است، با این حال، برخی مطالعات نشان داده‌اند که پیش‌آگهی ندول‌های تیروئید را می‌توان با استفاده از مدل‌های چندمتغیره و الگوریتم‌های یادگیری ماشین (Machine learning; ML) با مجموعه‌ای از داده‌های جمعیت‌شناختی، بیوشیمیایی، سونوگرافی و کلینیکوپاتولوژیک به عنوان متغیرهای پیشگوکننده به خوبی تعیین کرد [۱۵-۱۸]. مدل‌های یادگیری ماشین می‌تواند خط مشی پردازش داده را برای تشخیص، پیش‌آگهی و پیش‌بینی ندول‌های تیروئید بهبود بخشد [۱۹]. طراحی و اعتبارسنجی مدل پیشگوکننده با استفاده از ویژگی‌های تاثیرگذار ممکن است به تشخیص بیماران پرخطر کمک کند و منجر به پیش‌آگهی بهتر شود [۲۰]. الگوریتم‌های ML با شناسایی الگوها و یادگیری از داده‌ها می‌توانند پیش‌بینی‌های دقیق و قابل اعتمادی ارائه دهند. استفاده از نوموگرام و الگوریتم‌های ML در

نرم‌افزارهای STATA 14 و R 4.1.3 تحلیل شدند.

نتایج

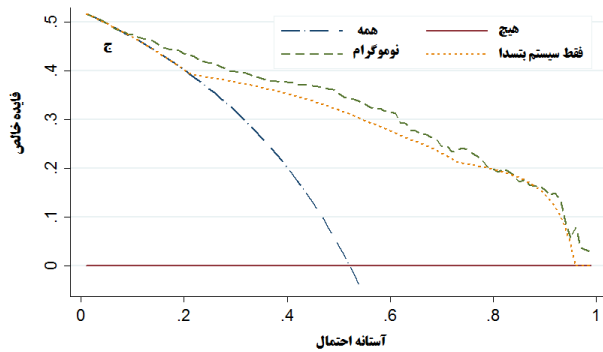
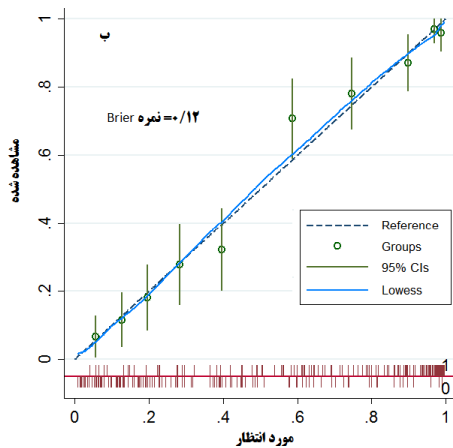
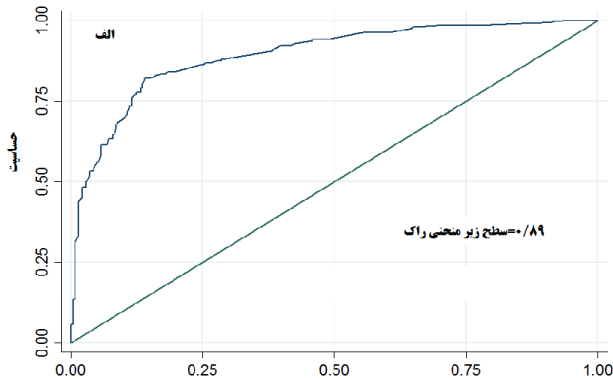
از ۶۵۰ بیمار مبتلا به ندول تیروئیدی که تحت جراحی تیروئیدکتومی قرار گرفته‌اند، ۴۳٪ خوش‌خیم و ۵۷٪ بدخیم بوده‌اند. ۸۶/۱۵٪ زن و ۱۳/۸۵٪ مرد بوده‌اند. گروه‌های سنی پایین‌تر، جنسیت زن، با سابقه خانوادگی سرطان تیروئید در بستگان درجه اول و آدنوپاتی گردنی بیشتر در بیماران با ندول بدخیم مشاهده شده است ($P < 0.05$). اکثر ندول‌های بدخیم یک‌کانونی بوده‌اند. بر اساس طبقه‌بندی سیستم بتسدا، FN/SFN و SM در بیماران با ندول بدخیم بیشتر بوده است ($P < 0.05$). خصوصیات با سطح معناداری کمتر از ۰/۰۵ وارد تحلیل چندمتغیره وارد شدند. شانس ابتلا به بدخیمی در بیماران زیر ۳۰ سال ۲/۷۰ برابر بیشتر از افراد ۴۵ تا ۶۰ سال است. وقتی ندول تیروئید یک‌کانونی باشد، شانس بدخیمی ۲/۷۷ برابر بیشتر از ندول تیروئید چندکانونی است. داشتن آدنوپاتی گردنی خطر بدخیمی را ۳/۸۹ برابر افزایش می‌دهد. علاوه بر این، شانس بدخیمی برای FN/SFN و SM سیستم بتسدا به ترتیب ۹/۵۰ و ۹۷/۶۹ برابر افزایش می‌یابد (جدول ۱).

تعمیم‌یافته (GLM) شده‌اند. اعتبار داخلی مدل از طریق بوت‌استرپ با ۱۰۰۰ نمونه‌گیری مجدد ارزیابی و مدل نهایی از طریق نونموگرام نمایش داده شد. کالیبراسیون مدل از طریق کالیبراسیون و نمره Brier ارزیابی و توانایی تشخیصی مدل از طریق مساحت زیر منحنی راک (AUC) ارزیابی شد. سودمندی بالینی مدل با استفاده از تحلیل منحنی تصمیم‌گیری (DCA) ارزیابی شد. عملکرد مدل GLM با پنج الگوریتم یادگیری ماشین دیگر، شامل LDA، kNN، RF، NN و SVM مقایسه شد. در فرایند ساخت مدل، مجموعه داده به دو بخش تقسیم شد: ۸۰٪ برای آموزش و ۲۰٪ برای آزمایش. در طول مرحله مدل‌سازی، مدل در مجموعه داده آموزشی توسعه یافت و عملکرد آن به صورت داخلی در مجموعه داده آزمایشی ارزیابی شد. برای بهینه‌سازی مدل و تنظیم پارامترهای آن، از اعتبارسنجی متقابل (۱۰ فولد) با سه تکرار استفاده شد. عملکرد الگوریتم‌های یادگیری ماشین با استفاده از ماتریس سردرگمی و شاخص‌هایی مانند صحت، کاپا، حساسیت (بازخوانی)، ویژگی، دقت (ارزش اخباری مثبت)، ارزش اخباری منفی، امتیاز AUC، F1 و مساحت زیر منحنی دقت - بازخوانی (area under the precision-recall curve; AUPRC) ارزیابی شد. مقدار p-value کمتر از ۰/۰۵ آستانه معناداری آماری در نظر گرفته شد. داده‌ها با استفاده از

جدول ۱: خصوصیات شرکت‌کنندگان با ندول تیروئید و نتایج مدل رگرسیون لجستیک چندمتغیره

متغیرها	خوش‌خیم (n=۲۸۰)	بدخیم (n=۳۷۰)	p-value	نسبت شانس (فاصله اطمینان ۹۵ درصد)	p-value
سن (سال)					
کمتر مساوی ۳۰	۴۲ (۱۵)	۹۷ (۲۶/۲۲)		رفرنس	
۳۰ تا ۴۵	۱۰۴ (۳۷/۱۴)	۱۵۵ (۴۱/۸۹)	۰/۰۰۱	۰/۶۵ (۰/۳۵-۱/۱۹)	۰/۱۱۶
۴۵ تا ۶۰	۱۰۴ (۳۷/۱۴)	۷۲ (۱۹/۴۶)		۰/۳۷ (۰/۱۸-۰/۷۵)	۰/۰۰۶
بالتر از ۶۰	۳۰ (۱۰/۷۱)	۴۶ (۱۲/۴۳)		۰/۶۳ (۰/۲۶-۱/۴۹)	۰/۲۹
جنسیت					
خانم	۲۵۳ (۹۰/۳۶)	۳۰۷ (۸۲/۹۷)	۰/۰۰۷	رفرنس	
آقا	۲۷ (۹/۶۴)	۶۳ (۱۷/۰۳)		۰/۹۲ (۰/۴۷-۱/۸۰)	۰/۸۱
سطح تحصیلات					
بی‌سواد	۴۳ (۱۵/۳۶)	۵۹ (۱۵/۹۵)			
کمتر مساوی دیپلم	۱۷۷ (۶۳/۲۱)	۲۳۴ (۶۳/۲۴)	۰/۹۷		
بیشتر از دیپلم	۶۰ (۲۱/۴۳)	۷۲ (۲۰/۸۱)			
سابقه بیماری‌های تیروئید در بستگان درجه ۱					
خیر	۱۸۹ (۶۷/۵۰)	۲۸۳ (۷۶/۴۹)		رفرنس	
خوش‌خیم	۸۵ (۳۰/۳۶)	۶۷ (۱۸/۱۱)	< ۰/۰۰۱	۰/۴۵ (۰/۲۶-۰/۸۰)	۰/۰۰۷
بدخیم	۶ (۲/۱۴)	۲۰ (۵/۴۱)		۱/۳۷ (۰/۲۰-۹/۴۰)	۰/۷۴
نوع بیماری تیروئید					
یوتیروئید	۲۰۷ (۷۳/۹۳)	۳۳۰ (۸۹/۱۹)		رفرنس	
هیپوتیروئیدسم بالینی	۱۵ (۵/۳۶)	۱۲ (۳/۲۴)		۱/۱۱ (۰/۳۷-۳/۳۵)	۰/۸۵
هیپوتیروئیدسم تحت بالینی	۲۲ (۷/۸۶)	۱۱ (۲/۹۷)	< ۰/۰۰۱	۰/۷۱ (۰/۱۹-۲/۶۵)	۰/۶۱
هیپر تیروئیدسم بالینی	۲۱ (۷/۵۰)	۹ (۲/۴۳)		۰/۲۰ (۰/۰۴-۱/۱۰)	۰/۰۶
هیپر تیروئیدسم تحت بالینی	۱۵ (۵/۳۶)	۸ (۲/۱۶)		۱/۶۱ (۰/۴۱-۶/۳۱)	۰/۴۹

برای kNN عدد ۰/۸۳ و برای SVM برابر ۰/۸۱ بوده است.



شکل ۲. ارزیابی توان مدل تشخیصی برای برآورد خطر بدخیمی در بیماران با ندول تیروئید
سطح زیر منحنی راک (الف)، منحنی کالیبراسیون (ب)، تحلیل منحنی تصمیم‌گیری (ج)

مقدار AUC برای مدل رگرسیون لجستیک چندمتغیره با اعتبارسنجی بوت‌استرپ در تمییز ندول‌های بدخیم از خوشخیم ۸۹ درصد است (شکل ۲، الف). علاوه بر این، نمودار کالیبراسیون نشان می‌دهد که توافق بین مقادیر احتمال پیش‌بینی شده و مشاهده شده نسبتاً خوب است. نمره Brier نیز ۰/۱۲ بوده است (شکل ۲، ب). تحلیل منحنی تصمیم‌گیری نشان داده است که ارزش بالینی نوموگرام با همه متغیرهای معنادار بیشتر از عملکرد سیستم بتسدا به تنهایی است (شکل ۲، ج).

میانگین صحت الگوریتم‌های GLM، LDA، NN، kNN، SVM و RF در داده‌های آموزشی برآورد شده به ترتیب ۰/۷۹، ۰/۷۸، ۰/۷۷، ۰/۷۹، ۰/۷۸ و ۰/۷۸ بوده‌اند. مقادیر شاخص کاپا برای الگوریتم‌های بیان شده به ترتیب ۰/۵۸، ۰/۵۷، ۰/۵۸، ۰/۵۴، ۰/۵۶ و ۰/۵۶ بوده‌اند. تفاوت در صحت و کاپا بین الگوریتم‌های یادگیری ماشین کمتر از ۰/۰۱ بوده و معنادار آماری نبوده‌اند، به هر حال تفاوت بین مدل GLM و kNN (تفاوت صحت = ۰/۰۲، $P=0.008$) در مورد کاپا نیز، تنها تفاوت بین الگوریتم NN و kNN تفاوت کاپا = ۰/۰۴، $P=0.003$) و بین GLM و kNN معنادار بوده است (تفاوت کاپا = ۰/۰۴، $P=0.002$). مقدار AUC برای الگوریتم‌های GLM، LDA، NN، kNN، SVM، RF در داده‌های آموزش به ترتیب برابر ۰/۸۷، ۰/۸۷، ۰/۸۷، ۰/۸۴، ۰/۸۶ و ۰/۸۶ بوده‌اند. تفاوت‌های معنادار بین GLM و kNN و SVM و GLM ($P=0.003$)، LDA و kNN ($P=0.006$)، NN و kNN ($P=0.004$) و RF و kNN ($P=0.007$) مشاهده شده است. عملکرد تشخیصی مدل‌های توسعه‌یافته در داده‌های آزمایش در جدول (۲) نشان داده شده است. در مجموع، عملکرد مدل‌های LDA و GLM بهتر از مدل‌های دیگر بوده است. صحت (۰/۷۸) و کاپا (۰/۵۵) مدل LDA بالاتر از مدل‌های دیگر بوده است. مقدار AUC برای مدل GLM (۰/۸۴) بالاتر از مدل‌های دیگر یادگیری ماشین بوده است. کمترین ارزش‌های عملکرد تشخیصی برای الگوریتم‌های SVM و kNN مشاهده شده است. دقت و بازخوانی مدل GLM به ترتیب ۰/۸۱ و ۰/۷۸ بوده، در حالی که مقادیر متناظر دقت و بازخوانی برای مدل LDA به ترتیب ۰/۸۴ و ۰/۷۶ بوده است. مقادیر AUPRC برای مدل GLM برابر ۰/۸۶ مشاهده شده که بالاتر از مدل‌های دیگر بوده است. این رقم برای مدل LDA، NN، SVM و RF برابر ۰/۸۵ بوده،

جدول ۲. عملکرد شش مدل یادگیری ماشین در داده آزمایش

مدل‌های یادگیری ماشین						شاخص‌های عملکرد
kNN	SVM	NN	RF	LDA	GLM	
۰/۷۴	۰/۷۴	۰/۷۵	۰/۷۴	۰/۷۸	۰/۷۷	صحت
۰/۴۶	۰/۴۷	۰/۵۱	۰/۴۸	۰/۵۵	۰/۵۳	کاپا
۰/۷۹	۰/۷۱	۰/۸۰	۰/۷۹	۰/۸۰	۰/۷۵	ویژگی
۰/۶۶	۰/۶۹	۰/۶۸	۰/۶۷	۰/۷۱	۰/۷۲	ارزش اخباری منفی
۰/۸۱	۰/۷۸	۰/۸۳	۰/۸۱	۰/۸۴	۰/۸۱	دقت (ارزش اخباری مثبت)

۰/۷۸	۰/۷۶	۰/۷۰	۰/۷۲	۰/۷۶	۰/۶۹	یادآوری (حساسیت)
۰/۷۹	۰/۷۹	۰/۷۵	۰/۷۷	۰/۷۴	۰/۷۴	نمره F1
۰/۸۴	۰/۸۳	۰/۸۰	۰/۸۲	۰/۷۸	۰/۷۸	سطح زیر منحنی راک
۰/۸۶	۰/۸۵	۰/۸۵	۰/۸۵	۰/۸۱	۰/۸۳	سطح زیر منحنی دقت یادآوری

GLM: generalized linear model, LDA: linear discriminant analysis, RF: random forest, NN: neural network, SVM: support vector machine, kNN: k-nearest neighbors

بحث

این مطالعه نشان داده است که قدرت پیش‌بینی یک مدل شامل نتیجه سیستم بتسدا، سن، سابقه خانوادگی سرطان تیروئید در بستگان درجه اول، کانونی بودن ندول تیروئید و آدنوپاتی گردنی می‌تواند به ۸۹ درصد برسد. اگرچه شش الگوریتم یادگیری ماشین برتری خاصی نسبت به یکدیگر نداشته‌اند، عملکرد کلی الگوریتم‌های GLM و LDA بهتر از سایر مدل‌ها بوده‌اند. توانایی تشخیص ندول خوش‌خیم (بر اساس شاخص ویژگی) با الگوریتم LDA در مقایسه با GLM بهتر بوده است.

در مطالعه حاضر، نتایج سیستم بتسدا مهم‌ترین نقش را در پیش‌بینی بدخیمی بر اساس نسبت شانس‌های برآورد شده از مدل GLM داشته است. در مطالعه D'Andréa و همکاران [۲۲] از بین همه متغیرهای بالینی، زیستی، اولتراسوند و سیتولوژیک، سیستم بتسدا مهم‌ترین نقش را در نوموگرام برای پیش‌بینی بدخیمی داشته است. در مطالعه یاد شده اخیر، مقدار AUC برابر ۰/۸۴ برای مدل GLM و ۰/۸۲/۶ درصد برای الگوریتم RF در داده‌های آزمایش گزارش شده و آن‌ها نیز به این نتیجه رسیده‌اند که نوموگرام مبتنی بر یادگیری ماشین که شامل سیستم بتسدا و داده‌های بالینی باشد، مدیریت بیماران را هدایت می‌کند و از جراحی‌های غیرضروری در بیماران با خطر بدخیمی نامشخص جلوگیری می‌کند [۲۲].

در مطالعه Liu و همکاران [۲۳]، شانس بدخیمی میان بیماران با نتایج سیتولوژیک بینابینی (AUS، FN و SM) برابر ۰/۷۴ بوده و کمتر از گروه M پس از تطبیق برای سن، اندازه ندول و محتوای ندول بوده است. در مطالعه ما، گروه M از تحلیل چندمتغیره حذف شد؛ زیرا خطر بدخیمی در میان آن‌ها ۰/۱۰۰ بوده است. با این حال، شانس بدخیمی با کاهش دسته‌های بتسدا کمتر می‌شود. برای مثال، شانس بدخیمی میان بیماران با دسته FN/SFN تقریباً ۰/۹۰ کمتر از گروه سیتولوژیک SM بوده که خطر بدخیمی در آن حدود ۰/۹۶ بوده است. در مطالعه حاضر، خطر بدخیمی در همه دسته‌های بتسدا از آنچه در سیستم بتسدا برای گزارش سیتوپاتولوژی تیروئید (The Bethesda System for Reporting Thyroid) (Cytology; TBSRTC) پیشنهاد شده است، بالاتر مشاهده شد [۲۴]. در بیشتر مطالعات که با هدف تشخیص سرطان بر اساس دسته‌های بتسدا انجام شده‌اند، میزان تشخیص سرطان تیروئید به ویژه برای طبقه AUS/FLUS بیشتر از نقطه برش‌های استاندارد بوده است [۲۵، ۲۶].

مطالعات قبلی که از الگوریتم‌های یادگیری ماشین برای پیش‌بینی بدخیمی، عود یا متاستاز استفاده کرده‌اند، نشان داده‌اند

که هر یک از الگوریتم‌ها برای جنبه‌ای خاص از طبقه‌بندی مناسب هستند. در مطالعه حاضر، عملکرد کلی الگوریتم‌ها (بر اساس صحت و کاپا) مدل LDA از GLM برتری داشته است. با این حال، عملکرد در طبقه‌بندی در سراسر آستانه‌های مختلف (بر اساس AUC و AUCPR) نشان داده است که الگوریتم GLM بهتر از LDA عمل می‌کند. توانایی تشخیص ندول خوش‌خیم (بر اساس شاخص ویژگی) نیز با الگوریتم LDA در مقایسه با GLM بهتر بوده است. در مطالعه Xi و همکاران [۲۷]، عملکرد تشخیصی کلی برای تشخیص ندول‌های بدخیم برای دو الگوریتم RF و gradient boosting machine (GBM) در مقایسه با سایر الگوریتم‌های یادگیری ماشین بهتر مشاهده شده است. همچنین، توانایی تشخیص ندول خوش‌خیم برای مدل GLM بهتر از سایر الگوریتم‌های یادگیری ماشین بوده است. در مطالعه Gu و همکاران [۱۵]، با در نظر گرفتن شاخص‌های بیوشیمیایی و مشخصات جمعیت‌شناختی به عنوان پیش‌بینی‌کننده‌ها، الگوریتم eXtreme Gradient Boosting (XGBoost) در همه جنبه‌های تشخیص و طبقه‌بندی از سایر الگوریتم‌ها برتری داشته است. در مواردی که الگوریتم‌های یادگیری ماشین پارامترهای عملکرد تشخیصی متنوعی را نشان می‌دهند، استفاده از الگوریتم‌های یادگیری گروهی (Ensemble Learning) می‌تواند مفید باشد. یادگیری گروهی، به عنوان یک رویکرد فراگیر جامع به یادگیری ماشین، می‌تواند عملکرد پیش‌بینی را با ترکیب پیش‌بینی‌ها از چندین الگوریتم بهبود بخشد، حتی در مواردی که تعداد پیش‌گوکننده‌ها محدود است و مشکل طبقات نامتعادل از نظر اندازه وجود دارد [۲۸]؛ بنابراین، برای مثال، انتظار می‌رود که اگر نتایج دو مدل GLM و LDA در مطالعه حاضر به صورت گروهی با هم ترکیب شوند، پارامترهای عملکرد تشخیصی نیز افزایش خواهد یافت.

در مطالعه ما، علاوه بر سیستم بتسدا، متغیرهای دیگری مانند سن، کانونی بودن ندول تیروئید و آدنوپاتی گردنی نیز می‌توانند پیش‌بینی‌کننده‌های قدرتمندی برای بدخیمی در بیماران با ندول تیروئید باشند. می‌توان فرض کرد که مدل‌های اشباع شده با داده‌های جمعیت‌شناختی و بالینی نیز می‌توانند برای تشخیص، پیش‌آگهی و پیش‌بینی بدخیمی استفاده شوند. برای مثال، یک مطالعه نشان داده است که با اعمال الگوریتم شبکه عصبی مصنوعی (Artificial Neural Network; ANN) بر روی پیش‌بینی‌کننده‌های جمعیت‌شناختی مانند سن، جنسیت و نژاد و پیش‌بینی‌کننده‌های بالینی مانند اندازه تومور، محل گره یا تعداد غدد لنفاوی، عملکرد تشخیصی برای پیش‌آگهی سرطان تیروئید

درمانی در میان بیماران مبتلا به ندول تیروئید کمک کند.

تشکر و قدردانی

این مقاله برگرفته از طرح تحقیقاتی مصوب مرکز تحقیقات سرطان، دانشگاه علوم پزشکی همدان با شماره ۱۴۰۲۱۵۱۱۰۶۴ است. از مسئولان محترم مرکز تحقیقات سرطان و معاونت تحقیقات و فناوری دانشگاه به دلیل همکاری و حمایتشان، تشکر و قدردانی می‌شود.

تضاد منافع

نتایج این مطالعه با منافع نویسندگان در تعارض نیست.

ملاحظات اخلاقی

کمیته اخلاق دانشگاه علوم پزشکی همدان با شناسه 1402/776 IR.UMSHA.REC. انجام این مطالعه را تایید کرده است.

سهام نویسندگان

نویسنده اول (پژوهشگر اصلی): تدوین چهارچوب اصلی طرح، نظارت و مدیریت پروژه، تدوین بخش‌های مختلف طرح، جمع‌آوری داده‌ها، ویرایش علمی مقاله (۴۰ درصد)، نویسنده دوم (همکار) تنظیم پیشنهاد، جمع‌آوری داده‌ها، مشارکت در تدوین بخش‌های مختلف طرح، مرور مقاله (۲۰ درصد)، نویسنده سوم (پژوهشگر اصلی): مسئول مکاتبات، تنظیم پیشنهاد، مشارکت در طراحی مطالعه، نظارت بر حسن اجرای پروژه، تحلیلگر آماری، نگارش بخش روش‌شناسی طرح، نگارش مقاله (۴۰ درصد).

حمایت مالی

دانشگاه علوم پزشکی همدان از این طرح حمایت مالی کرده است.

می‌تواند بیش از ۹۰٪ باشد [۲۹]. در مطالعه‌ای دیگر با استفاده از سیزده ویژگی کلینیکوپاتولوژیک برای پیش‌بینی عود در بیماران مبتلا به سرطان تیروئید خوب‌تمایز شده، AUC الگوریتم‌های یادگیری ماشین مانند SVM، kNN، RF، ANN و درخت تصمیم‌گیری می‌تواند به حدود ۹۹٪ برسد [۳۰].

چندین محدودیت باید هنگام تفسیر نتایج مد نظر باشد. اول، انتظار می‌رفت که الگوریتم‌های پیشرفته یادگیری ماشین با یادگیری نظارت‌شده از داده و قابلیت‌های محاسباتی بهبود یافته، عملکرد تشخیصی بیشتری از مدل GLM کلاسیک داشته باشند. با این حال، چنین بهبودهایی در عمل مشاهده نشد. یکی از دلایل ممکن است تعداد محدود پیشگوکننده‌ها باشد. مدل‌های پیش‌بینی باید شامل متغیرهای مهمی مانند مرحله TNM، داده‌های بیوشیمیایی و داده‌های مولکولی شوند. با این حال، چنین داده‌هایی در دسترس ما نبوده است. دوم، مطالعات بزرگ‌تر و اعتبارسنجی خارجی مدل پیش‌آگهی با تعداد بیشتر پیشگوکننده برای تشخیص و پیش‌بینی دقیق و صحیح نیاز است که انجام گیرد. سوم، تورش ناشی از داده‌های تنک نیز بایستی در نظر گرفته شود.

نتیجه‌گیری

یک مدل پیش‌آگهی مبتنی بر یادگیری ماشین را با استفاده از داده‌های جمعیت‌شناختی، بالینی و سیستم بتسدا طراحی و اعتبارسنجی شد که ممکن است به راهنمایی و مدیریت گزینه‌های

REFERENCES

- Seib CD, Sosa JA. Evolving Understanding of the Epidemiology of Thyroid Cancer. *Endocrinol Metab Clin North Am*. 2019;**48**(1):23-35. PMID: 30717905 DOI: 10.1016/j.ecl.2018.10.002
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin*. 2018;**68**(6):394-424. PMID: 30207593 DOI: 10.3322/caac.21492
- Miranda-Filho A, Lortet-Tieulent J, Bray F, Cao B, Franceschi S, Vaccarella S, et al. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. *Lancet Diabetes Endocrinol*. 2021;**9**(4):225-34. PMID: 33662333 DOI: 10.1016/S2213-8587(21)00027-9
- Pizzato M, Li M, Vignat J, Laversanne M, Singh D, La Vecchia C, et al. The epidemiological landscape of thyroid cancer worldwide: GLOBOCAN estimates for incidence and mortality rates in 2020. *Lancet Diabetes Endocrinol*. 2022;**10**(4):264-72. PMID: 35271818 DOI: 10.1016/S2213-8587(22)00035-3
- Nejadghaderi SA, Moghaddam SS, Azadnajafabad S, Rezaei N, Rezaei N, Tavangar SM, et al. Burden of thyroid cancer in North Africa and Middle East 1990-2019. *Front Oncol*. 2022;**12**:955358. PMID: 36212501 DOI: 10.3389/fonc.2022.955358
- Wang Y, Guan Q, Xiang J. Nomogram for predicting central lymph node metastasis in papillary thyroid microcarcinoma: A retrospective cohort study of 8668 patients. *Int J Surg*. 2018;**55**:98-102. PMID: 29803769 DOI: 10.1016/j.ijsu.2018.05.023
- Chung SR, Baek JH, Choi YJ, Sung TY, Song DE, Kim TY, et al. The relationship of thyroid nodule size on malignancy risk according to histological type of thyroid cancer. *Acta Radiol*. 2020;**61**(5):620-628. PMID: 31554409 DOI: 10.1177/0284185119875642
- Parameswaran R, Shulin Hu J, Min En N, Tan WB, Yuan NK. Patterns of metastasis in follicular thyroid carcinoma and the difference between early and delayed presentation. *Ann R Coll Surg Engl*. 2017;**99**(2):151-154. PMID: 27659362 DOI: 10.1308/rcsann.2016.0300
- Wu MH, Lee YY, Lu YL, Lin SF. Risk Factors and Prognosis for Metastatic Follicular Thyroid Cancer. *Front Endocrinol*. 2022;**13**:791826. PMID: 35299967 DOI: 10.3389/fendo.2022.791826
- Baloch ZW, LiVolsi VA, Asa SL, Rosai J, Merino MJ, Randolph G, et al. Diagnostic terminology and morphologic criteria for cytologic diagnosis of thyroid lesions: a synopsis of the National Cancer Institute Thyroid Fine-Needle Aspiration State of the Science Conference. *Diag Cytopathol*. 2008;**36**(6):425-37. PMID: 18478609 DOI: 10.1002/dc.20830
- Kant R, Davis A, Verma V. Thyroid Nodules: Advances in Evaluation and Management. *Am Fam Physician*. 2020;**102**(5):298-304. PMID: 32866364
- Linhares SM, Handelsman R, Picado O, Farrá JC, Lew JI. Fine needle aspiration and the Bethesda system: Correlation with histopathology in 1,228 surgical patients. *Surgery*. 2021;**170**(5):1364-8. PMID: 34134896 DOI: 10.1016/j.surg.2021.05.016
- Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid*. 2016;**26**(1):1-133. PMID: 26462967 DOI: 10.1089/thy.2015.0020

14. Aliyev A, Aliyeva I, Giammarile F, Talibova N, Aliyeva G, Novruzov F. Diagnostic accuracy of fine needle aspiration biopsy versus postoperative histopathology for diagnosing thyroid malignancy. *Endocrinol Diabetes Metab.* 2022;**5**(6):e373. [PMID: 36149057](#) [DOI: 10.1002/edm2.373](#)
15. Gu J, Xie R, Zhao Y, Zhao Z, Xu D, Ding M, et al. A machine learning-based approach to predicting the malignant and metastasis of thyroid cancer. *Front oncol.* 2022;**12**:938292. [PMID: 36601485](#) [DOI: 10.3389/fonc.2022.938292](#)
16. Liu WC, Li ZQ, Luo ZW, Liao WJ, Liu ZL, Liu JM. Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer. *Cancer Med.* 2021;**10**(8):2802-11. [PMID: 33709570](#) [DOI: 10.1002/cam4.3776](#)
17. Lu J, Liao J, Chen Y, Li J, Huang X, Zhang H, et al. Risk factor analysis and prediction model for papillary thyroid carcinoma with lymph node metastasis. *Frontiers Endocrinol.* 2023;**14**:1287593. [PMID: 38027220](#) [DOI: 10.3389/fendo.2023.1287593](#)
18. Zhang TT, Zeng J, Yang Y, Wang JJ, Kang YJ, Zhang DH, et al. A visualized dynamic prediction model for survival of patients with geriatric thyroid cancer: A population-based study. *Frontiers Endocrinol.* 2022;**13**:1038041. [PMID: 36568078](#) [DOI: 10.3389/fendo.2022.1038041](#)
19. Slabaugh G, Beltran L, Rizvi H, Deloukas P, Marouli E. Applications of machine and deep learning to thyroid cytology and histopathology: a review. *Front Oncol.* 2023;**13**:958310. [PMID: 38023130](#) [DOI: 10.3389/fonc.2023.958310](#)
20. Liu W, Wang S, Ye Z, Xu P, Xia X, Guo M. Prediction of lung metastases in thyroid cancer using machine learning based on SEER database. *Cancer Med.* 2022;**11**(12):2503-15. [PMID: 35191613](#) [DOI: 10.1002/cam4.4617](#)
21. Kourou K, Exarchos KP, Papaloukas C, Sakaloglou P, Exarchos T, Fotiadis DI. Applied machine learning in cancer research: A systematic review for patient diagnosis, classification and prognosis. *Comput Struct Biotechnol J.* 2021;**19**:5546-55. [PMID: 34712399](#) [DOI: 10.1016/j.csbj.2021.10.006](#)
22. D'Andréa G, Gal J, Mandine L, Dassonville O, Vandersteen C, Guevara N, et al. Application of machine learning methods to guide patient management by predicting the risk of malignancy of Bethesda III-V thyroid nodules. *Eur J Endocrinol.* 2022;**188**(3):lvad017. [PMID: 36799885](#) [DOI: 10.1093/ejendo/lvad017](#)
23. Liu X, Medici M, Kwong N, Angell TE, Marqusee E, Kim MI, et al. Bethesda Categorization of Thyroid Nodule Cytology and Prediction of Thyroid Cancer Type and Prognosis. *Thyroid.* 2016;**26**(2):256-61. [PMID: 26563459](#) [DOI: 10.1089/thy.2015.0376](#)
24. Cibas ES, Ali SZ. The 2017 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid.* 2017;**27**(11):1341-1346. [PMID: 29091573](#) [DOI: 10.1089/thy.2017.0500](#)
25. Gweon HM, Son EJ, Youk JH, Kim JA. Thyroid nodules with Bethesda system III cytology: can ultrasonography guide the next step? *Ann Surg Oncol.* 2013;**20**(9):3083-8. [PMID: 23700214](#) [DOI: 10.1245/s10434-013-2990-x](#)
26. Mileva M, Stoilovska B, Jovanovska A, Ugrinska A, Petrushevska G, Kostadinova-Kunovska S, et al. Thyroid cancer detection rate and associated risk factors in patients with thyroid nodules classified as Bethesda category III. *Radiol Oncol.* 2018;**52**(4):370-6. [PMID: 30265655](#) [DOI: 10.2478/raon-2018-0039](#)
27. Xi NM, Wang L, Yang C. Improving the diagnosis of thyroid cancer by machine learning and clinical data. *Sci Rep.* 2022;**12**(1):11143. [DOI: 10.1038/s41598-022-15342-z](#) [PMID: 35778428](#)
28. Yaghoobi Notash A, Yaghoobi Notash A, Omidi Z, Haghghat S. Prediction of lymphedema occurrence in patients with breast cancer using the optimized combination of ensemble learning algorithm and feature selection. *PMC Med Inform Decis Mak.* 2022;**22**(1):195. [PMID: 35879760](#) [DOI: 10.1186/s12911-022-01937-z](#)
29. Mourad M, Moubayed S, Dezube A, Mourad Y, Park K, Torreblanca-Zanca A, et al. Machine Learning and Feature Selection Applied to SEER Data to Reliably Assess Thyroid Cancer Prognosis. *Sci Rep.* 2020;**10**(1):5176. [PMID: 32198433](#) [DOI: 10.1038/s41598-020-62023-w](#)
30. Borzooei S, Briganti G, Golparian M, Lechien JR, Tarokhian A. Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study. *Eur Arch Otorhinolaryngol.* 2024;**281**(4):2095-2104. [PMID: 37902840](#) [DOI: 10.1007/s00405-023-08299-w](#)