



# A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis

## OPEN ACCESS

### Edited by:

Xin Gao,

King Abdullah University of Science and Technology, Saudi Arabia

### Reviewed by:

Eman Toraih,

Tulane University, United States

Lei Zhu,

Fifth Affiliated Hospital of Wenzhou Medical University, China

Hongzhou Liu,

Chinese PLA General Hospital, China

### \*Correspondence:

Lingyan Zhu

zhuling\_happy@126.com

Jingyuan Yang

yangjingyuan@gmc.edu.cn

†These authors have contributed equally to this work

### Specialty section:

This article was submitted to Computational Genomics, a section of the journal Frontiers in Genetics

Received: 15 March 2022

Accepted: 16 June 2022

Published: 08 July 2022

### Citation:

Jin S, Yang X, Zhong Q, Liu X, Zheng T, Zhu L and Yang J (2022) A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis. *Front. Genet.* 13:896805. doi: 10.3389/fgene.2022.896805

Shuai Jin<sup>1†</sup>, Xing Yang<sup>2†</sup>, Qiliang Zhong<sup>3†</sup>, Xiangmei Liu<sup>4</sup>, Tao Zheng<sup>1</sup>, Lingyan Zhu<sup>5\*†</sup> and Jingyuan Yang<sup>6\*†</sup>

<sup>1</sup>School of Big Health, Guizhou Medical University, Guiyang, China, <sup>2</sup>School of Medicine and Health Administration, Guizhou Medical University, Guiyang, China, <sup>3</sup>Department of Urology, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, <sup>4</sup>School of Clinical Medicine, Guizhou Medical University, Guiyang, China, <sup>5</sup>Health Management Center, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, <sup>6</sup>School of Public Health, Guizhou Medical University, Guiyang, China

**Purpose:** To explore clinical and non-clinical characteristics affecting the prognosis of patients with differentiated thyroid cancer with distant metastasis (DTCDM) and establish an accurate overall survival (OS) prognostic model.

**Patients and methods:** Study subjects and related information were obtained from the National Cancer Institute's surveillance, epidemiology, and results database (SEER). Kaplan-Meier analysis, log-rank test, and univariate and multivariate Cox analysis were used to screen for factors influencing the OS of patients with DTCDM. Nine variables were introduced to build a machine learning (ML) model, receiver operating characteristic (ROC) was used to evaluate the recognition ability of the model, calibration plots were used to obtain prediction accuracy, and decision curve analysis (DCA) was used to estimate clinical benefit.

**Results:** After applying the inclusion and exclusion criteria, a total of 3,060 patients with DTCDM were included in the survival analysis from 2004 to 2017. A machine learning prediction model was developed with nine variables: age at diagnosis, gender, race, tumor size, histology, regional lymph node metastasis, primary site surgery, radiotherapy, and chemotherapy. After excluding patients who survived <120 months, variables were sub-coded and machine learning was used to model OS prognosis in patients with DTCDM. Patients 6–50 years of age had the highest scores in the model. Other variables with high scores included small tumor size, male sex, and age 51–76. The AUC and calibration curves confirm that the XGBoost model has good performance. DCA shows that our model can be used to support clinical decision-making in a 10-years overall survival model.

**Conclusion:** An artificial intelligence model was constructed using the XGBoost algorithms to predict the 10-years overall survival rate of patients with DTCDM. After

model validation and evaluation, the model had good discriminative ability and high clinical value. This model could serve as a clinical tool to help inform treatment decisions for patients with DTCDM.

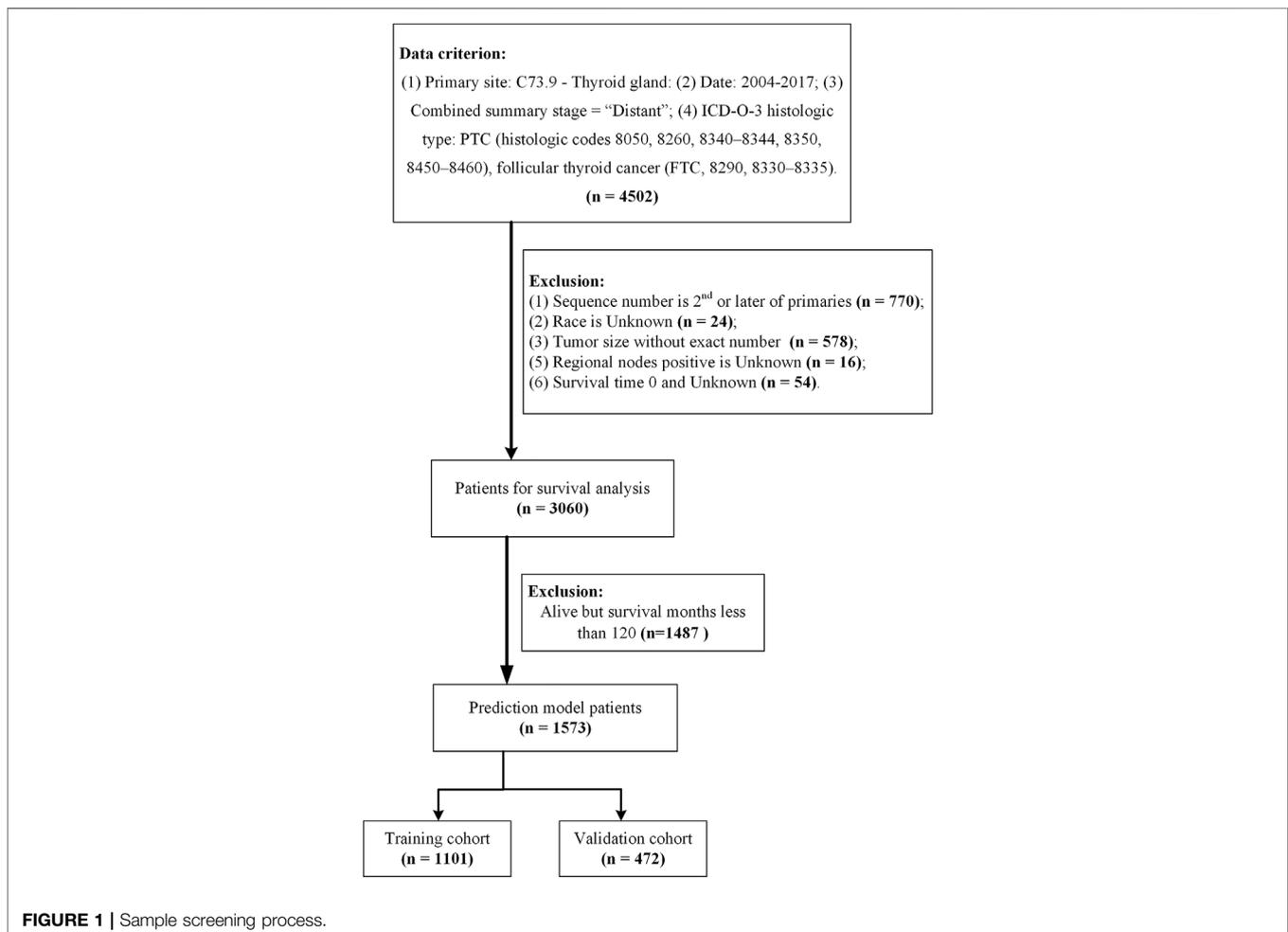
**Keywords:** differentiated thyroid cancer, Xgboost algorithm, machine learning, distant metastases, predictive model, SEER database

## INTRODUCTION

Differentiated thyroid cancer (DTC) is the most common endocrine malignancy, with the global incidence increasing dramatically in recent decades (Cabanillas et al., 2016; Siegel et al., 2020). Most DTC patients have a good long-term prognosis because of biological characteristics and effective responses to treatment modalities (Lim et al., 2017; Liu et al., 2018). For TC patients with distant metastasis (DM), however, the overall prognosis is significantly worse (Durante et al., 2006; Sugitani et al., 2008; Farooki et al., 2012; Nies et al., 2021).

The main histological subtypes of DTC include papillary thyroid carcinoma (PTC), and follicular thyroid carcinoma (FTC), <10% of people with DTC will develop DM (Durante et al., 2006). Most DTCDM is asymptomatic and only detected

during systematic surveillance or systemic metastatic examination of malignant lymph nodes. The common site of distant metastases from thyroid cancer is the lung, followed by the bone, brain, and liver (Chesover et al., 2021; Liu et al., 2021). Because the incidence of DM is low and there is often an absence of symptoms, it is frequently overlooked at the time of initial TC diagnosis. Ten-year survival rates are often used to assess DTC treatment efficacy and characterize risk factors. For patients with metastatic thyroid cancer, treatment is often individualized and usually consists of thyroid surgery with adjuvant radiotherapy or chemotherapy (Sampson et al., 2007). While prognostic models have been developed to study factors influencing different primary cancer subtypes (Zhao et al., 2019; Zhou et al., 2020; Jin et al., 2021; Kong et al., 2021), the 10-years overall survival of patients with DTCDM is unclear, especially because OS has not



changed significantly in recent decades (Goffredo et al., 2013). In addition, there are few prognostic models for DTCDM that can help to inform patient follow-up and treatment decisions (Chen et al., 2021).

Machine learning (ML) is an emerging multi-disciplinary approach used to correlate multiple discrete variables and accurately predict outcomes. Following the development of evidence-based medicine and the need for more advanced tools to collect medical data with complex structures and large sample sizes, ML emerged as an alternative approach to disease diagnosis and prognosis, with high predictive performance and a wide range of applications (Goecks et al., 2020; May, 2021). ML algorithms are now being successfully applied to predict cancer survival (Angraa et al., 2020). The XGBoost algorithm (XGB), in particular, is shown to have excellent prediction performance in previous studies (Senders et al., 2020; Jiang et al., 2021).

The Surveillance, Epidemiology, and End Results (SEER) program is a population-based cancer registry system sponsored by the United States' National Cancer Institute (NCI) that currently covers about 28% of the population in 18 registered states (Noone et al., 2016). Using the considerations listed above, a prognostic modeling analysis of patients with DTCDM was conducted using SEER data. This study assessed the ability of clinical and non-clinical factors to predict DTCDM using the XGB model. The XGB model was also built to predict the 10-years OS rate of TC patients with DM. The performance of the XGBoost model using logistic regression (LR), random forests (RF), and support vector machine (SVM) models were compared.

## MATERIALS AND METHODS

### Study Population

This study used the SEER database (<https://seer.cancer.gov/>) developed by the National Cancer Institute, a free cancer registry in the United States. A data use agreement was signed with the SEER database official and all authors followed the specified conditions. Because SEER data is freely available to researchers and patient personal information is officially withheld, no moral or ethical support from the host institution was required for this study.

Data was downloaded from the SEER 18 regs plus database using SEER\*Stat (Version 8.3.9.2, <https://seer.cancer.gov/data-software/>). Data were selected from patients with histologically confirmed distant metastatic thyroid cancer using the following criteria: 1) primary site code is C73.9 - thyroid gland, 2) dates ranging from 2004 to 2017, 3) PTC (histologic codes 8050, 8260, 8340–8344, 8350, 8450–8460), FTC (8290, 8330–8335), and 4) diagnosis confirmation combined with a summary stage for the distant future. Exclusion criteria included 1) sequence numbers for second or later occurrences, 2) unknown race, 3) unclear tumor size, 4) unknown surgery, 5) unknown regional nodes positive (RN\_positive), and 6) survival time of 0 or unknown. A total of 3,060 patients were included in the survival analysis to clarify possible factors influencing the prognostic model. Of these, 1,487 patients who had survived <120 months by the follow-up

cut-off date were excluded and 1,573 patients were included in the prediction model and grouped in the training set ( $n = 1,101$ ; 70%) or the validation set ( $n = 472$ ; 30%) at a 7:3 ratio (Figure 1).

### Variable Selection and Endpoints

To take full advantage of ML, several established demographic and clinical characteristics were selected as independent variables for analysis. Pathology variables commonly used in thyroid cancer research, including tumor size, histological type, regional lymph node status, surgical modality, radiotherapy, and chemotherapy, and demographic indicators, including age, gender, and ethnicity were included. Age and tumor size were optimally stratified for processing using X-tile software (<https://medicine.yale.edu/lab/rimm/research/software/>) and included in the survival analysis. These analyses were performed before excluding patients who survived for <120 months. The 10-years OS rate for patients with DTCDM was defined as the model endpoint.

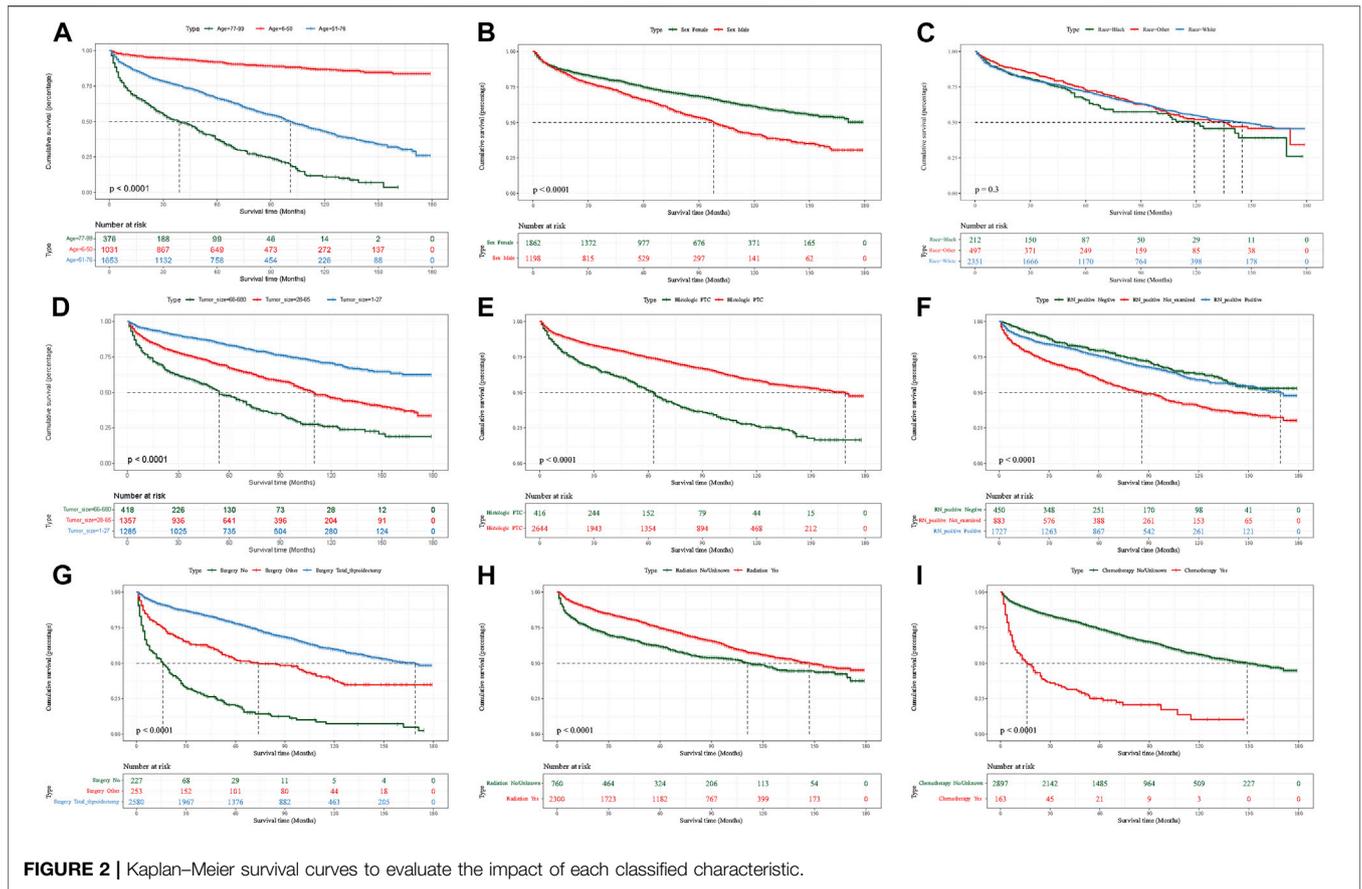
### Statistical Analysis

All statistical analyses and model building in this study were performed using R (version 4.1.2, <https://www.r-project.org/>). The Kaplan–Meier method, with both univariate and multifactorial Cox, was used to screen for OS prognostic factors in thyroid cancer patients, and all variables were used to construct prognostic models. The Chi-square test was then used to analyze differences between the training and validation cohorts. The training set was used to build the XGBoost model and the model was evaluated with the test set. The model capability evaluation includes the following three items: 1) receiver operating characteristic (ROC) curves were used to analyze model discrimination and the area under the ROC curve (AUC) was used to assess predictive model accuracy (Hanley and McNeil, 1982; Wolbers et al., 2009); 2) calibration plots were used to assess the performance of the model, which calibrates how well model predictions agree with the actual observations (Leonard et al., 2020); and 3) decision curve analysis (DCA) was used to assess the clinical usefulness and net benefit of predictive model performance by calculating the difference between the true and false positive rates, weighted by the probability of the chosen risk threshold to assess the net benefit of the model (Vickers and Elkin, 2006). Logistic regression, SVM, and random forest models were built for comparison to the XGBoost model.

## RESULTS

### Baseline Characteristics and Survival Analysis

A total of 3,060 patients with DTCDM were included in the survival analysis for variable screening. The best cut-off values for age were 50 and 76 years old, and the tumor size was 27 and 65 mm (Supplementary Figure S1). Most of the study population (54.02%) was 51–76 years of age with a higher proportion of females than males. Most patients' tumor size was <6.5 cm and the most common histological type was PTC



(86.41%), and FTC type has about 14%. Above half of patients had regional node metastases (56.44%) and total thyroidectomy, radiotherapy, and chemotherapy were used for 56.44%, 76.16%, and 5.33% of patients, respectively. After validation using the Kaplan-Meier method and log-rank test, only survival rates of DTCDM patients did not differ between races (Figure 2C), and all other variables differed at all scales (Figures 2A,B,D-I). We performed univariate and multivariate Cox models for age, gender, race, tumor size, histological type, RN\_positive, surgery, radiotherapy, and chemotherapy, respectively. The multivariate model showed that all variables were independent prognostic factors for DTCDM, except for the race (Table 1).

### Prognostic Model Construction

Using survival analysis, the following nine variables were included in the prognostic model: age, sex, race, tumor size, histological type, RN\_positive, surgery, radiation, and chemotherapy. Most patients (65.9%) with DTCDM  $\geq 10$  years were 51–76 years of age, and a higher proportion were women (59.0%) and had PTC histologic staging (81.8%), total thyroidectomy (76.7%), radiotherapy (72.3%), and chemotherapy (7.6%). Analysis of the differences between the two groups for each variable showed no statistical difference, indicating good comparability between the training and validation sets (Table 2).

Unlike associated studies (Hou et al., 2020; Wei et al., 2021), categorical variables were included in the prediction model as

“dummy variables”, age:  $>76$  years, sex: female, race: black, histological type: FTC, tumor size:  $>65$  mm, surgery: no, radiation: no, chemotherapy: no, and RN\_positive negative as a control. The XGB algorithm identifies the importance of features based on the magnitude of the gain value obtained for each variable (relative importance scores out of 100), with higher values indicating greater importance to the predicted target. The variables with the highest scores are: age: 6–50 years (39points), tumor size: 1–27 mm (11 points), sex: male (7 points), age: 51–76 years (6 points), RN\_positive: not examined (5 points), tumor size: 28–65 mm (5 points), and radiation: yes (5 points) (Figure 3). These variables were included in the LR, SVM, RF models, along with the XGB model for performance testing.

### Model Performance

The optimal model performance parameters were determined after several validations and debugging. To assess the ability of the XGB, SVM, LR, and RF models to identify the OS of patients with 10-years DTCDM, the training and validation sets were plotted, and the AUC was calculated.

In the training cohort, the AUC of the XGB model was 0.948, higher than the AUC of the other models (SVM AUC, 0.888; LR AUC 0.873, RF AUC,0.881), and The AUC of XGB in the test group was 0.864, slightly lower than the AUCs of LR and SVM (0.889 and 0.871) and higher than the AUC of the RF model (0.858) (Figure 4). The calibration plots of the 10-years TCDM OS for the training and

**TABLE 1** | The baseline characteristics, univariate and multivariate cox analysis.

Characteristics	Total (n, %)	Univariate		Multivariate	
		HR (95% CI)	p	HR (95% CI)	p
Age at diagnosis, years					
77–99	376 (12.29)	References		References	
6–50	1031 (33.69)	0.07 (0.06–0.09)	<0.001	0.10 (0.08–0.13)	<0.001
51–76	1653 (54.02)	0.41 (0.36–0.47)	<0.001	0.49 (0.43–0.57)	<0.001
Sex					
Female	1862 (60.85)	References		References	
Male	1198 (39.15)	1.59 (1.41–1.79)	<0.001	1.27 (1.12–1.43)	<0.001
Race					
Black	212 (6.93)	References		References	
Other	497 (16.24)	0.82 (0.64–1.07)	0.141	0.98 (0.75–1.27)	0.858
White	2351 (76.83)	0.85 (0.68–1.06)	0.143	1.04 (0.83–1.31)	0.714
Tumor size, mm					
66–680	418 (13.66)	References		References	
28–65	1357 (44.35)	0.53 (0.46–0.61)	<0.001	0.7 (0.70–0.82)	<0.001
1–27	1285 (41.99)	0.25 (0.21–0.29)	<0.001	0.41 (0.34–0.49)	<0.001
Histologic					
FTC	416 (13.59)	References		References	
PTC	2644 (86.41)	0.42 (0.36–0.48)	<0.001	0.78 (0.66–0.91)	<0.001
RN positive					
Negative	450 (14.71)	References		References	
Not examined	883 (28.86)	2.08 (1.71–2.52)	<0.001	1.34 (1.09–1.64)	0.005
Positive	1727 (56.44)	1.15 (0.95–1.39)	0.145	1.47 (1.21–1.79)	<0.001
Surgery					
No	227 (7.42)	References		References	
Other	253 (8.27)	0.33 (0.26–0.41)	<0.001	0.53 (0.42–0.67)	<0.001
Total thyroidectomy	2580 (84.31)	0.16 (0.13–0.18)	<0.001	0.35 (0.28–0.42)	<0.001
Radiation					
No	760 (24.84)	References		References	
Yes	2300 (76.16)	0.66 (0.58–0.75)	<0.001	0.86 (0.75–0.98)	0.029
Chemotherapy					
No	2897 (94.67)	References		References	
Yes	163 (5.33)	4.73 (3.89–5.75)	<0.001	2.55 (2.07–3.13)	<0.001

RN, positive, regional nodes metastases.

test sets showed good linear agreement between predicted and actual observations from the XGB model. The XGB and LR models fit better than the SVM and RF models (Figure 5). The DCA curves for the XGB, SVM, RF, and LR models were plotted for the training and test cohorts. The y-axis of the decision curve represents the net benefit, the decision analysis metric that determines whether the benefits of a particular clinical decision outweigh the harms. Each point on the x-axis represents a threshold probability that distinguishes between patients who die and those who survive. The analysis shows that the XGB, LR, RF, and SVM models all achieve a net clinical benefit (Figure 6).

## DISCUSSION

Survival prediction is very difficult in malignancy but important for treatment planning and patient management (Cheon et al., 2016). Compared to the empirical predictions of clinicians, the XGBoost model provides a more reliable option for the 10-years survival status of patients with distant metastases from thyroid cancer. The current study found that XGB had a good predictive value and could help clinicians to develop a rational individualized treatment and management plan. Although thyroid cancer is a relatively

slow-growing cancer, once distant metastasis has occurred, the tumor grows exponentially at the location of metastasis, explaining why patients with DM have a worse prognosis (Rajan et al., 2020). While DTC generally has a favorable prognosis, the clinical course can be poor (See et al., 2017). Assessing the survival of patients with DTCDM is thus of great clinical importance.

This study used survival analysis to screen for factors that may affect the OS of patients with DTCDM. Although, the multiple Cox model we developed showed that race is not a prognostic factor for DTCDM patients. One study showed that race does affect the survival of patients with differentiated thyroid cancer and that treatment options need to be specific to different racial groups (Tang et al., 2018). However, Crepeau et al. (Crepeau et al., 2021) found that thyroid cancer prognosis and recurrence did not differ by race following the same surgical approach, especially when all patients receive the same quality of care. It is likely that differences in the prognosis of TC patients by race are the result of social and economic differences between racial groups.

The XGB algorithm is a new type of AI algorithm that is easy to use, efficient and accurate. This algorithm is becoming increasingly popular in the medical field and is now widely used for disease prediction and early diagnosis. The current study used the prognostic variables obtained from survival

**TABLE 2 |** The baseline characteristics of the training and validation sets used in the prognostic model.

Characteristics	Total cohort	Training cohort	Validation cohort	$\chi^2$	<i>p</i>
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)		
Age at diagnosis, years	1573 (100.0)	1101 (70.0)	472 (30.0)	2.011	0.366
77–99	281 (17.9)	202 (18.3)	79 (16.7)		
6–50	375 (23.8)	252 (22.9)	123 (26.1)		
51–76	917 (58.3)	647 (58.8)	270 (57.2)	0.003	0.959
Sex					
Female	928 (59.0)	650 (59.0)	278 (58.9)		
Male	645 (41.0)	451 (41.0)	194 (41.1)	1.895	0.388
Race					
Black	108 (6.9)	81 (7.4)	27 (5.7)		
Other	256 (16.3)	183 (16.6)	73 (15.5)	0.752	0.687
White	1209 (76.9)	837 (76.0)	372 (78.8)		
Tumor size, mm					
66–680	273 (17.4)	197 (17.9)	76 (16.1)	1.961	0.161
28–65	745 (47.4)	519 (47.1)	226 (47.9)		
1–27	555 (35.3)	385 (35.0)	170 (36.0)		
Histologic				0.18	0.914
FTC	286 (18.2)	210 (19.1)	76 (16.1)		
PTC	1287 (81.8)	891 (80.9)	396 (83.9)		
RN positive				0.944	0.624
Negative	220 (14.0)	154 (14.0)	66 (14.0)		
Not examined	585 (37.2)	413 (37.5)	172 (36.4)		
Positive	768 (48.8)	534 (48.5)	234 (49.6)	0.704	0.401
Surgery					
No	189 (12.0)	133 (12.1)	56 (11.9)		
Other	178 (11.3)	130 (11.8)	48 (10.2)	0.042	0.837
Total thyroidectomy	1206 (76.7)	838 (76.1)	368 (78.0)		
Radiation					
No	436 (27.7)	312 (28.3)	124 (26.3)		
Yes	1137 (72.3)	789 (71.7)	348 (73.7)		
Chemotherapy					
No	1453 (92.4)	1018 (92.5)	435 (92.2)		
Yes	120 (7.6)	83 (7.5)	37 (7.8)		

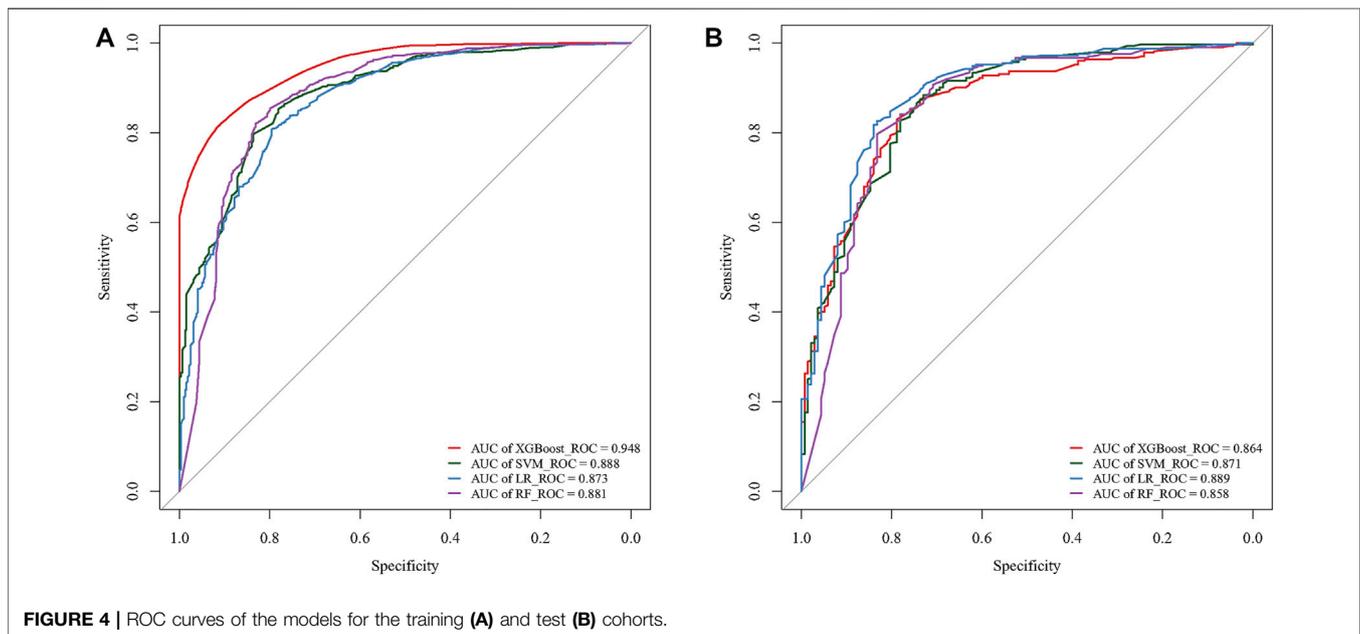
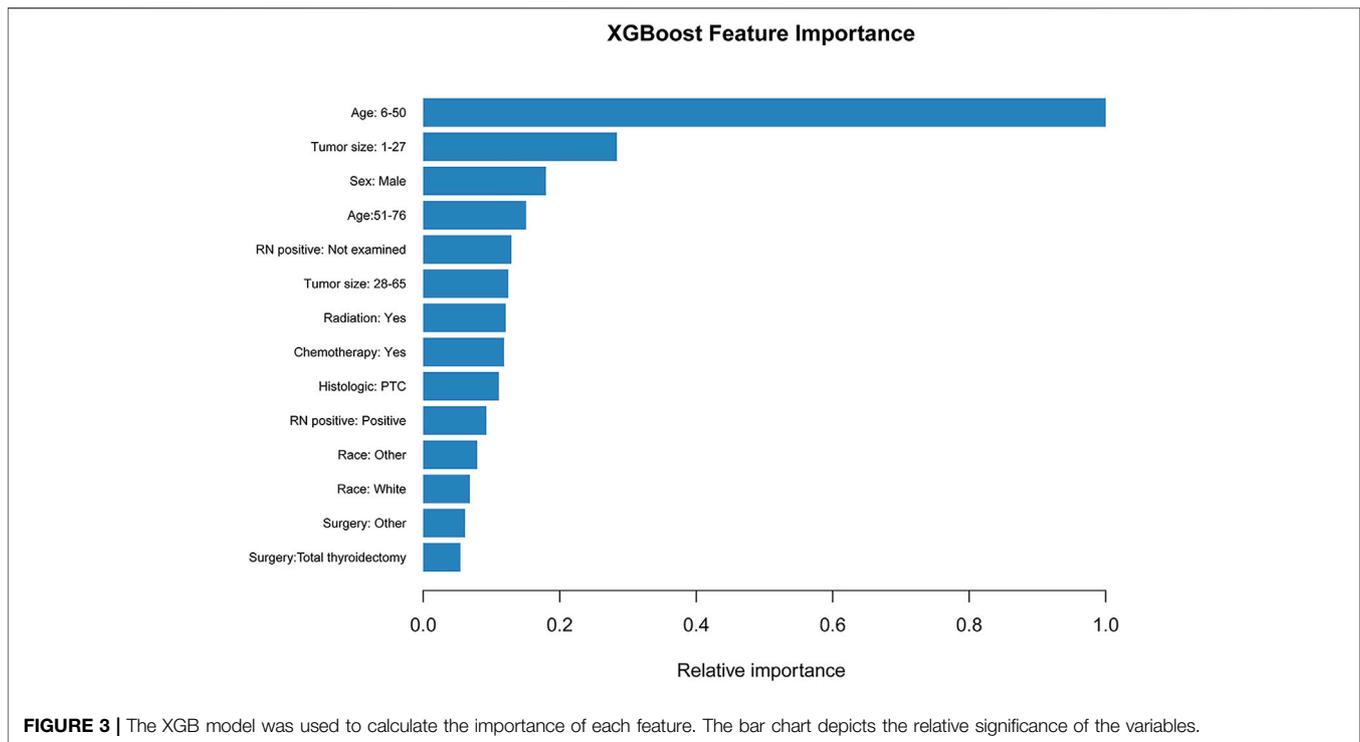
RN, positive, regional nodes metastases.

analysis to develop a prognostic model for DTCDM OS using the XGB algorithm. The model was validated by combining the clinical and non-clinical variables and was shown to be highly effective. Additional factors that may affect OS and provide more clinical information about DTCDM, were also reported. Traditional ML LR also performed well, possibly because of the study data or because LR performs just as well as ML in clinical prediction models (Christodoulou et al., 2019).

Age was the highest-scoring, and thus most significant, variable in the XGB model, Patients <51 years of age had significantly higher OS than patients >80 years of age. The role for age in predicting TC has been confirmed by other studies, with older patients having a poorer prognosis than younger patients (Sampson et al., 2007; Nixon et al., 2012). Compared with the 8th edition of the TNM staging system with a cut-off value of 55 years, the optimal cut-off values for age in this study were 50 and 77 (Amin et al., 2017). This implies that the cut-off value for age may need to be studied in depth for patients with distant metastatic thyroid cancer. In the current study, there was also a significant score for the 51–79 age group. These findings suggest that treatment should be tailored to patients in different age groups.

The results of one study confirm that patients with PTC with distant metastases have a good prognosis after treatment (Sampson et al., 2007). Another study found that sex is a prognostic factor for DTCDM, likely because estrogen production limits cancer progression (Suteau et al., 2021). The current study found that tumor size was an important factor affecting DTCDM OS, with relatively significant scores for both 1–27 mm and 28–65 mm, a finding supported by Nguyen et al. (Nguyen et al., 2018). Han et al. reported that 15–20 mm tumors do not affect the OS of TCDM patients (Han et al., 2017). Unexamined metastases and those localized in the lymph nodes also scored highly. The number of lymph node metastases correlates strongly with the presence of DM while the risk of DM can be assessed based on the number of lymph nodes (Jeon et al., 2016). These findings may help to resolve controversy over the indication of lateral lymph node dissection (Fan et al., 2018). The regional metastatic status of the lymph nodes should be assessed in all patients with DTCDM.

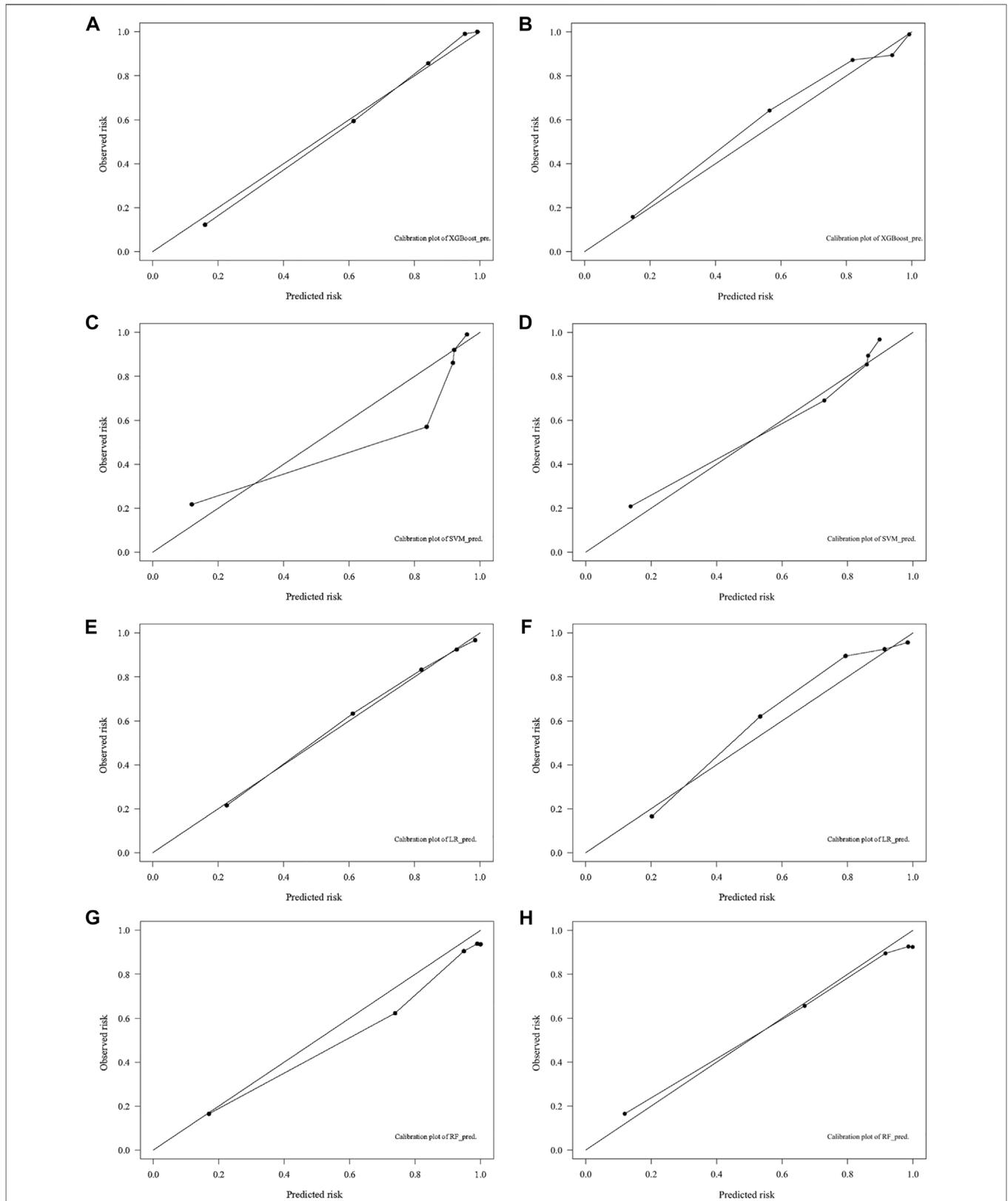
Radiotherapy had a relatively significant score in the OS prognostic model of patients with DTCDM. Studies indicate that radioactive iodine (RAI) treatment is very effective in patients with small metastases, indicating that early diagnosis improves outcomes (Durante et al., 2006; Sampson et al., 2007). Diagnosis of DM and



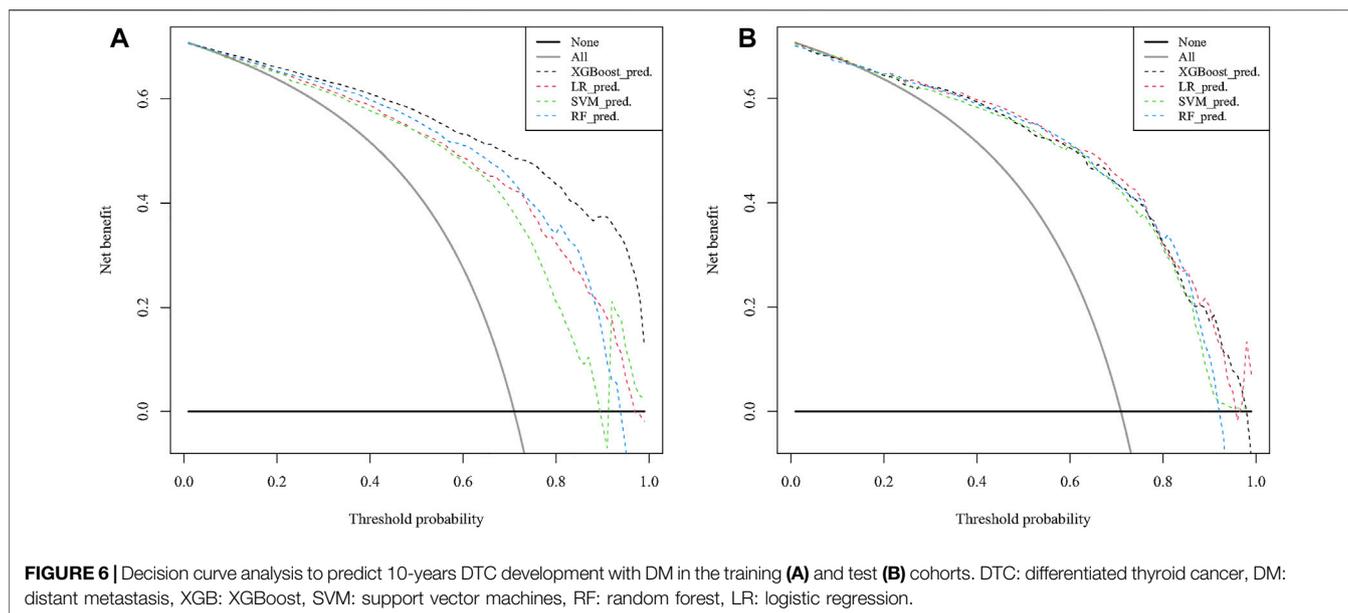
initiation of RAI therapy before overt metastases appear, especially in children and adolescents for whom selective treatment is more appropriate (Sugino et al., 2020). While RAI treatment is beneficial for TC survival, however, high-quality RAI accumulation may increase the risk of secondary tumor mutations and more aggressive disease, thus negatively impacting patient survival (Su et al., 2015; Pasqual et al., 2022). Nies et al. concluded that repeated RAI treatment is unlikely to

benefit TC patients and may do more harm than good over their lifespan (Nies et al., 2021). In summary, studies differ on whether and how to treat DTC patients with RAI (Jeon et al., 2016; Lin et al., 2018).

Total thyroidectomy and other surgical procedures were important in the prognostic model. Sampson et al. concluded that a total thyroidectomy should be performed alongside RAI treatment (Sampson et al., 2007). It is also suggested that, where



**FIGURE 5 |** Calibration plots for predicting 10-years DTC development with DM in the training (A,C,E,G) and test (B,D,F,H) cohorts. DTC: differentiated thyroid cancer, DM: distant metastasis, XGB: XGBoost, SVM: support vector machines, RF: random forest, LR: logistic regression.



possible, local curative surgery with RAI and thyroid hormone suppression should be performed in patients with DTCDM (Ito et al., 2010). However, the survival benefit of thyroid cancer surgery may vary depending on the site of metastasis (Besic et al., 2016). The current study found that chemotherapy was a strong predictor of OS in patients with DTCDM and a risk factor for OS in survival analysis. Chemotherapy is often administered to patients with large tumors who are no longer candidates for surgery or show iodine resistance. However, a study indicates that chemotherapy is highly toxic and is associated with a poor response rate (Schmidbauer et al., 2017). The high mortality rate of chemotherapy patients may be due to the relative severity of the disease in this group of patients, in addition to the toxic impact of the treatment. Recent studies have shown that targeted therapies such as tyrosine kinase inhibitors (TKIs) offer high survival rates and that patients may have a better outcome if targeted treatments are combined with chemotherapy (Kraeber-Bodéré et al., 2010; Carling and Udelsman, 2014; Viola et al., 2016; Lorusso et al., 2021).

Although the predictive model used in this study had a good performance, there were some limitations. First, the study relied on regression data and some samples with missing information were removed, which may have biased the model. Second, outcome data for individuals receiving targeted therapies were not included in the sample, which may have made the prediction model less comprehensive. Finally, more work needs to be done to explain the predictive efficacy of ML versus traditional statistical methods.

## CONCLUSION

This study analyzed the clinical characteristics and prognosis of patients with DTCDM and constructed prognostic models using four machine learning methods. The XGB model was effective at predicting the 10-years OS of patients with DTCDM and may help clinicians to make more accurate and personalized clinical

decisions. This is particularly important to improve the long-term prognosis of high-risk patients.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Ethical review and approval was not required for the animal study because The SEER database is free for researchers to download and has removed patient information.

## AUTHOR CONTRIBUTIONS

SJ, LZ and JY contributed to the conception and design of the study; SJ, QZ and TZ collected and analyzed data; XY, XL, LZ and QZ wrote and revised the manuscript; All authors reviewed and approved the final version of the manuscript.

## FUNDING

This study was supported by funds from the Ministry of Education Industry-Academia Cooperation Collaborative Education Project (202101311012, 202102576027).

## ACKNOWLEDGMENTS

We are grateful to all the experts who opened the research topic “Medical knowledge-assisted machine learning techniques in individualized medicine” in the journal *Frontiers in Genetics*.

## REFERENCES

- Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, G. S., et al. (2017). The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More “personalized” Approach to Cancer Staging. *CA A Cancer J. Clin.* 67, 93–99. doi:10.3322/caac.21388
- Angraal, S., Mortazavi, B. J., Gupta, A., Khera, R., Ahmad, T., Desai, N. R., et al. (2020). Machine Learning Prediction of Mortality and Hospitalization in Heart Failure with Preserved Ejection Fraction. *JACC Heart Fail.* 8 (1), 12–21. doi:10.1016/j.jchf.2019.06.013
- Besic, N., Schwarzbartl-Pevcec, A., Videgar-Kralj, B., Crnic, T., Gazic, B., and Marolt Music, M. (2016). Treatment and Outcome of 32 Patients with Distant Metastases of Hürthle Cell Thyroid Carcinoma: a Single-Institution Experience. *BMC Cancer* 16, 162. doi:10.1186/s12885-016-2179-3
- Cabanillas, M. E., McFadden, D. G., and Durante, C. (2016). Thyroid Cancer. *Lancet* 388 (10061), 2783–2795. doi:10.1016/s0140-6736(16)30172-6
- Carling, T., and Udelsman, R. (2014). Thyroid Cancer. *Annu. Rev. Med.* 65, 125–137. doi:10.1146/annurev-med-061512-105739
- Chen, S., Jiang, L., Zhang, E., Hu, S., Wang, T., Gao, F., et al. (2021). A Novel Nomogram Based on Machine Learning-Pathomics Signature and Neutrophil to Lymphocyte Ratio for Survival Prediction of Bladder Cancer Patients. *Front. Oncol.* 11, 703033. doi:10.3389/fonc.2021.703033
- Cheon, S., Agarwal, A., Popovic, M., Milakovic, M., Lam, M., Fu, W., et al. (2016). The Accuracy of Clinicians’ Predictions of Survival in Advanced Cancer: a Review. *Ann. Palliat. Med.* 5 (1), 22–29. doi:10.3978/j.issn.2224-5820.2015.08.04
- Chesover, A. D., Vali, R., Hemmati, S. H., and Wasserman, J. D. (2021). Lung Metastasis in Children with Differentiated Thyroid Cancer: Factors Associated with Diagnosis and Outcomes of Therapy. *Thyroid* 31 (1), 50–60. doi:10.1089/thy.2020.0002
- Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., and Van Calster, B. (2019). A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models. *J. Clin. Epidemiol.* 110, 12–22. doi:10.1016/j.jclinepi.2019.02.004
- Crepeau, P. K., Kulkarni, K., Martucci, J., and Lai, V. (2021). Comparing Surgical Thoroughness and Recurrence in Thyroid Cancer Patients across Race/ethnicity. *Surgery* 170 (4), 1099–1104. doi:10.1016/j.surg.2021.05.001
- Durante, C., Haddy, N., Baudin, E., Leboulleux, S., Hartl, D., Travagli, J. P., et al. (2006). Long-term Outcome of 444 Patients with Distant Metastases from Papillary and Follicular Thyroid Carcinoma: Benefits and Limits of Radioiodine Therapy. *J. Clin. Endocrinol. Metab.* 91 (8), 2892–2899. doi:10.1210/jc.2005-2838
- Fan, W., Xiao, C., and Wu, F. (2018). Analysis of Risk Factors for Cervical Lymph Node Metastases in Patients with Sporadic Medullary Thyroid Carcinoma. *J. Int. Med. Res.* 46 (5), 1982–1989. doi:10.1177/0300060518762684
- Farooki, A., Leung, V., Tala, H., and Tuttle, R. M. (2012). Skeletal-related Events Due to Bone Metastases from Differentiated Thyroid Cancer. *J. Clin. Endocrinol. Metabolism* 97 (7), 2433–2439. doi:10.1210/jc.2012-1169
- Goecks, J., Jalili, V., Heiser, L. M., and Gray, J. W. (2020). How Machine Learning Will Transform Biomedicine. *Cell.* 181 (1), 92–101. doi:10.1016/j.cell.2020.03.022
- Goffredo, P., Sosa, J. A., and Roman, S. A. (2013). Differentiated Thyroid Cancer Presenting with Distant Metastases: a Population Analysis over Two Decades. *World J. Surg.* 37 (7), 1599–1605. doi:10.1007/s00268-013-2006-9
- Han, K., Kim, E.-K., and Kwak, J. Y. (2017). 1.5–2 Cm Tumor Size Was Not Associated with Distant Metastasis and Mortality in Small Thyroid Cancer: A Population-Based Study. *Sci. Rep.* 7, 46298. doi:10.1038/srep46298

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.896805/full#supplementary-material>

- Hanley, J. A., and McNeil, B. J. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143 (1), 29–36. doi:10.1148/radiology.143.1.7063747
- Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., et al. (2020). Predicting 30-days Mortality for MIMIC-III Patients with Sepsis-3: a Machine Learning Approach Using XGboost. *J. Transl. Med.* 18 (1), 462. doi:10.1186/s12967-020-02620-5
- Ito, Y., Masuoka, H., Fukushima, M., Inoue, H., Kihara, M., Tomoda, C., et al. (2010). Prognosis and Prognostic Factors of Patients with Papillary Carcinoma Showing Distant Metastasis at Surgery (M1 Patients) in Japan. *Endocr. J.* 57 (6), 523–531. doi:10.1507/endocrj.k10e-019
- Jeon, M. J., Kim, W. G., Choi, Y. M., Kwon, H., Lee, Y.-M., Sung, T.-Y., et al. (2016). Features Predictive of Distant Metastasis in Papillary Thyroid Microcarcinomas. *Thyroid* 26 (1), 161–168. doi:10.1089/thy.2015.0375
- Jiang, J., Pan, H., Li, M., Qian, B., Lin, X., and Fan, S. (2021). Predictive Model for the 5-year Survival Status of Osteosarcoma Patients Based on the SEER Database and XGBoost Algorithm. *Sci. Rep.* 11 (1), 5542. doi:10.1038/s41598-021-85223-4
- Jin, M., Megwalu, U. C., and Noel, J. E. (2021). External Beam Radiotherapy for Medullary Thyroid Cancer Following Total or Near-Total Thyroidectomy. *Otolaryngol. Head. Neck Surg.* 164 (1), 97–103. doi:10.1177/0194599820947696
- Kong, N., Xu, Q., Zhang, Z., Cui, A., Tan, S., and Bai, N. (2021). Age Influences the Prognosis of Anaplastic Thyroid Cancer Patients. *Front. Endocrinol.* 12, 704596. doi:10.3389/fendo.2021.704596
- Kraeber-Bodéré, F., Salaun, P.-Y., Oudoux, A., Goldenberg, D. M., Chatal, J.-F., and Barbet, J. (2010). Pretargeted Radioimmunotherapy in Rapidly Progressing, Metastatic, Medullary Thyroid Cancer. *Cancer* 116 (S4), 1118–1125. doi:10.1002/cncr.24800
- Leonard, S. A., Kennedy, C. J., Carmichael, S. L., Lyell, D. J., and Main, E. K. (2020). An Expanded Obstetric Comorbidity Scoring System for Predicting Severe Maternal Morbidity. *Obstet. Gynecol.* 136 (3), 440–449. doi:10.1097/aog.0000000000004022
- Lim, H., Devesa, S. S., Sosa, J. A., Check, D., and Kitahara, C. M. (2017). Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974–2013. *Jama* 317 (13), 1338–1348. doi:10.1001/jama.2017.2719
- Lin, J.-D., Kuo, S.-F., Huang, B.-Y., Lin, S.-F., and Chen, S.-T. (2018). The Efficacy of Radioactive Iodine for the Treatment of Well-Differentiated Thyroid Cancer with Distant Metastasis. *Nucl. Med. Commun.* 39 (12), 1091–1096. doi:10.1097/MNM.0000000000000897
- Liu, W. C., Li, Z. Q., Luo, Z. W., Liao, W. J., Liu, Z. L., and Liu, J. M. (2021). Machine Learning for the Prediction of Bone Metastasis in Patients with Newly Diagnosed Thyroid Cancer. *Cancer Med.* 10 (8), 2802–2811. doi:10.1002/cam4.3776
- Liu, Z., Zeng, W., Huang, L., Wang, Z., Wang, M., Zhou, L., et al. (2018). Prognosis of FTC Compared to PTC and FVPTC: Findings Based on SEER Database Using Propensity Score Matching Analysis. *Am. J. Cancer Res.* 8 (8), 1440
- Lorusso, L., Cappagli, V., Valerio, L., Giani, C., Viola, D., Puleo, L., et al. (2021). Thyroid Cancers: From Surgery to Current and Future Systemic Therapies through Their Molecular Identities. *Ijms* 22 (6), 3117. doi:10.3390/ijms22063117
- May, M. (2021). Eight Ways Machine Learning Is Assisting Medicine. *Nat. Med.* 27 (1), 2–3. doi:10.1038/s41591-020-01197-2
- Nguyen, X. V., Roy Choudhury, K., Tessler, F. N., and Hoang, J. K. (2018). Effect of Tumor Size on Risk of Metastatic Disease and Survival for Thyroid Cancer: Implications for Biopsy Guidelines. *Thyroid* 28 (3), 295–300. doi:10.1089/thy.2017.0526
- Nies, M., Vassilopoulou-Sellin, R., Bassett, R. L., Yedururi, S., Zafereo, M. E., Cabanillas, M. E., et al. (2021). Distant Metastases from Childhood Differentiated Thyroid Carcinoma: Clinical Course and Mutational

- Landscape. *J. Clin. Endocrinol. Metab.* 106 (4), 1683–1697. doi:10.1210/clinem/dgaa935
- Nixon, I. J., Whitcher, M. M., Palmer, F. L., Tuttle, R. M., Shaha, A. R., Shah, J. P., et al. (2012). The Impact of Distant Metastases at Presentation on Prognosis in Patients with Differentiated Carcinoma of the Thyroid Gland. *Thyroid* 22 (9), 884–889. doi:10.1089/thy.2011.0535
- Noone, A.-M., Lund, J. L., Mariotto, A., Cronin, K., McNeel, T., Deapen, D., et al. (2016). Comparison of SEER Treatment Data with Medicare Claims. *Med. Care* 54 (9), e55–e64. doi:10.1097/mlr.0000000000000073
- Pasqual, E., Schonfeld, S., Morton, L. M., Villoing, D., Lee, C., Berrington de Gonzalez, A., et al. (2022). Association between Radioactive Iodine Treatment for Pediatric and Young Adulthood Differentiated Thyroid Cancer and Risk of Second Primary Malignancies. *Jco* 40 (13), 1439–1449. doi:10.1200/jco.21.01841
- Rajan, N., Khanal, T., and Ringel, M. D. (2020). Progression and Dormancy in Metastatic Thyroid Cancer: Concepts and Clinical Implications. *Endocrine* 70 (1), 24–35. doi:10.1007/s12020-020-02453-8
- Sampson, E., Brierley, J. D., Le, L. W., Rotstein, L., and Tsang, R. W. (2007). Clinical Management and Outcome of Papillary and Follicular (Differentiated) Thyroid Cancer Presenting with Distant Metastasis at Diagnosis. *Cancer* 110 (7), 1451–1456. doi:10.1002/cncr.22956
- Schmidbauer, B., Menhart, K., Hellwig, D., and Grosse, J. (2017). Differentiated Thyroid Cancer-Treatment: State of the Art. *Ijms* 18 (6), 1292. doi:10.3390/ijms18061292
- See, A., See, N. G., Tan, N. C., Teo, C., Ng, J., Soo, K. C., et al. (2017). Distant Metastasis as the Sole Initial Manifestation of Well-Differentiated Thyroid Carcinoma. *Eur. Arch. Otorhinolaryngol.* 274 (7), 2877–2882. doi:10.1007/s00405-017-4532-9
- Senders, J. T., Staples, P., Mehrtash, A., Cote, D. J., Taphoorn, M. J. B., Reardon, D. A., et al. (2020). An Online Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning. *Neurosurgery* 86 (2), E184–e192. doi:10.1093/neuros/nyz403
- Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590
- Su, D. H., Chang, S. H., and Chang, T. C. (2015). The Impact of Locoregional Recurrences and Distant Metastases on the Survival of Patients with Papillary Thyroid Carcinoma. *Clin. Endocrinol.* 82 (2), 286–294. doi:10.1111/cen.12511
- Sugino, K., Nagahama, M., Kitagawa, W., Ohkuwa, K., Uruno, T., Matsuzu, K., et al. (2020). Distant Metastasis in Pediatric and Adolescent Differentiated Thyroid Cancer: Clinical Outcomes and Risk Factor Analyses. *J. Clin. Endocrinol. Metab.* 105 (11), e3981–e3988. doi:10.1210/clinem/dgaa545
- Sugitani, I., Fujimoto, Y., and Yamamoto, N. (2008). Papillary Thyroid Carcinoma with Distant Metastases: Survival Predictors and the Importance of Local Control. *Surgery* 143 (1), 35–42. doi:10.1016/j.surg.2007.06.011
- Suteau, V., Munier, M., Briet, C., and Rodien, P. (2021). Sex Bias in Differentiated Thyroid Cancer. *Ijms* 22 (23), 12992. doi:10.3390/ijms222312992
- Tang, J., Kong, D., Cui, Q., Wang, K., Zhang, D., Liao, X., et al. (2018). Racial Disparities of Differentiated Thyroid Carcinoma: Clinical Behavior, Treatments, and Long-Term Outcomes. *World J. Surg. Onc.* 16 (1), 45. doi:10.1186/s12957-018-1340-7
- Vickers, A. J., and Elkin, E. B. (2006). Decision Curve Analysis: a Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* 26 (6), 565–574. doi:10.1177/0272989x06295361
- Viola, D., Valerio, L., Molinaro, E., Agate, L., Bottici, V., Biagini, A., et al. (2016). Treatment of Advanced Thyroid Cancer with Targeted Therapies: Ten Years of Experience. *Endocr. Relat. Cancer* 23 (4), R185–R205. doi:10.1530/erc-15-0555
- Wei, L., Huang, Y., Chen, Z., Li, J., Huang, G., Qin, X., et al. (2021). A Novel Machine Learning Algorithm Combined with Multivariate Analysis for the Prognosis of Renal Collecting Duct Carcinoma. *Front. Oncol.* 11, 777735. doi:10.3389/fonc.2021.777735
- Wolbers, M., Koller, M. T., Witteman, J. C. M., and Steyerberg, E. W. (2009). Prognostic Models with Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology* 20 (4), 555–561. doi:10.1097/EDE.0b013e3181a39056
- Zhao, H., Huang, T., and Li, H. (2019). Risk Factors for Skip Metastasis and Lateral Lymph Node Metastasis of Papillary Thyroid Cancer. *Surgery* 166 (1), 55–60. doi:10.1016/j.surg.2019.01.025
- Zhou, L., Li, Q., Chen, S., Huang, Y., Wei, W., Zhang, C., et al. (2020). Synergic Effects of Histology Subtype, Lymph Node Metastasis, and Distant Metastasis on Prognosis in Differentiated Thyroid Carcinoma Using the SEER Database. *Gland. Surg.* 9 (4), 907–918. doi:10.21037/gs-20-273

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Jin, Yang, Zhong, Liu, Zheng, Zhu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.