Archives of Current Medical Research
Arch Curr Med Res 2025; 6(3): 280-287

Received: 16 April 2025
Revised: 27 May 2025
Accepted: 02 June 2025
DOI: 10.47482/acmr.1677545

ORIGINAL ARTICLE

Predicting Recurrence of Differentiated Thyroid Cancer with an Explainable Artificial Intelligence Model

Ahmet Cankat Öztürk[1] **Erkan Akkur**[2] **Serkan Çizmecioğulları**[3]

1 Presidency of The Republic of Türkiye Secretariat of Defence Industries, Ankara, Türkiye
2 Turkish Medicines and Medical Devices Agency, Ankara, Türkiye
3 Kırşehir Ahi Evran University, Vocational School of Technical Sciences, Electronics and Automation Biomedical Device Technology, Kırsehir, Türkiye

Abstract

Background: This study aimed to predict the recurrence of differentiated thyroid cancer and identify its most representative risk factors using an explainable artificial intelligence model.

Methods:: The publicly available Differentiated Thyroid Cancer Recurrence dataset from the University of California Irvine Machine Learning Repository, comprising 383 patients and 17 features, was employed. Five classifiers, -Random Forest, Gradient Boosting, AdaBoost, Support Vector Classifier and Logistic Regression-, were employed to predict the recurrence. Permutation feature importance (PFI) and SHapley Additive exPlanations (SHAP) explainable artificial intelligence methods were used to determine the features that had the most impact on the prediction result.

Results: The Random Forest algorithm outperformed others, achieving an accuracy of 97.39% and an Area under the Curve of 0.993. Response to treatment, ATA risk stratification, tumor stage and patient age were determined as the factors with the highest contribution to the model prediction process through SHAP and permutation importance analyses, and this finding was consistent with the prognostic markers stated in the literature.

Conclusion: The proposed explainable machine learning framework has shown satisfactory results in predicting DTC recurrence while identifying clinically important features. This approach can offer valuable support to clinicians in early identification of high-risk patients and personalization of surveillance strategies.

Keywords: Differentiated thyroid cancer recurrence prediction, machine learning, explainable artificial intelligence, SHAP, permutation feature importance

Corresponding Author:
Ahmet Cankat Öztürk, Presidency of The Republic of Türkiye Secretariat of Defence Industries, Ankara, Türkiye
E-mail: ahmetcankatozturk@gmail.com

INTRODUCTION

Differentiated thyroid cancer (DTC) arising from follicular thyroid cells represents approximately 90% of all thyroid cancers (1). Despite the excellent 10-year survival rate of over 90% for DTC, recurrence remains a significant concern for both patients and healthcare systems. The risk of recurrence, ranging from 5% to 30%, can result in additional health problems, increased treatment costs and a serious psychological burden on patients (2-3). Early and accurate prediction of recurrence can improve patient outcomes and the efficiency of health systems by helping clinicians deliver personalized care and timely interventions (4). However, the high number of clinical variables and the complex relationships between them limit the effectiveness of traditional statistical methods in predicting DTC recurrence (5, 6). Machine learning (ML) algorithms offer advantages in this context, as they can capture both linear and non-linear patterns in large clinical datasets. These models can support the development of data-driven decision support systems by revealing hidden patterns in medical data (7, 8). Despite this potential, the use of ML for DTC recurrence prediction remains limited. Most existing studies focus mainly on performance metrics, without adequately addressing how these models can be interpreted and integrated into clinical decision-making processes (9–12). Moreover, the decision-making processes of these models generally remain in the form of a "black box", making it difficult for clinicians to trust the model output and rely on it to make their decisions. Therefore, it is essential not only to develop high-performance models, but also to illuminate decision-making processes to increase clinical applicability and physician confidence in ML-enabled models. Explainable Artificial Intelligence (XAI) methods have been introduced in recent years to provide a more understandable and interpretable understanding of the decision processes of ML models. Such methods improve clinical reliability by clarifying which variables play a role in the predictions of the models and thus can contribute to clinicians making more informed, transparent and patient-specific decisions (13-15).

This study presents a ML-based framework for predicting recurrence of differentiated thyroid cancer. Furthermore, it focuses on the utilization of XAI techniques to improve the clinical interpretability of the decision-making processes of ML models. In this way, it is aimed to identify the main risk factors contributing to the development of recurrence and to provide valuable contributions to clinical decision-making processes.

MATERIALS AND METHODS

Since this study was conducted on a publicly available clinical data set, Ethics Committee approval is not required. An XAI-based model was introduced to predict the probability of DTC recurrence. The proposed model starts with the data pre-processing phase, encompassing label encoding, and feature selection. Five distinct ML algorithms were employed in the prediction process. Diverse performance metrics were applied to obtain the optimal prediction outcome. Lastly, XAI techniques were utilized to determine the features that provide the most significant impacts on the probability of DTC recurrence. The proposed framework is illustrated in Figure 1.

Dataset

The dataset examined in this study was obtained from a previous study conducted by Borzooei et al. (10). The title of the dataset is Differentiated Thyroid Cancer Recurrence and it is accessible via the Machine Learning Repository at University of California Irvine, which is a publicly available (16). It contains 16 independent features and 1 target feature that pertain to 383 thyroid cancer patients who were monitored for a minimum of 10 years and up to 15 years. The target feature was categorized as no recurrence and recurrence. In the dataset used in this study, the age of patients with DTC recurrence was 47.11±18.27 years and 38±12.95 years for those without recurrence. The features in the dataset and their descriptions are presented in Table 1.

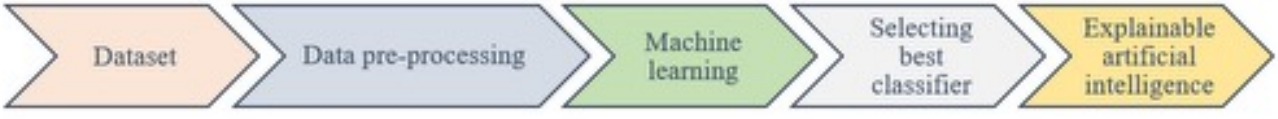

Figure 1: The framework of the proposed ML-based prediction model for DTC recurrence shows the steps of data preprocessing, model training, performance evaluation and XAI analysis.

Table 1. The features and descriptions of dataset

No	Features	Description	Unique V
1	Age	Represents the age of individuals in the dataset.	
2	Gender	Indicates the gender of individuals	[Female, Male]
3	Smoking	Possibly an attribute related to smoking behaviour.	[No, yes]
4	Hx Smoking	Indicates whether individuals have a history of smoking	[No, yes]
5	Hx Radiotherapy	Indicates whether individuals have a history of radiotherapy treatment	[No, yes]
6	Thyroid Function	Possibly indicates the status or function of the thyroid gland.	[Clinical, Euthyroid, Subclinical]
7	Physical Examination	Describes the results of a physical examination	[Single nodular goiter-left, Multinodular goiter, Normal, Single nodular goiter-right]
8	Adenopathy	Indicates the presence and location of adenopathy	[No, Right, Extensive, Left, Bilateral, Posterior]
9	Pathology	Describes the types of thyroid cancer based on pathology examinations	[Micropapillary, Papillary, Follicular, Hurthle cell]
10	Focality	Indicates whether the thyroid cancer is unifocal or multifocal.	[Uni-Focal, Multi-Focal]
11	Risk	Represents the risk category associated with thyroid cancer.	[Low, Intermediate, High]
12	Tumor (T)	Represents the tumor stage of thyroid cancer, indicating the size and extent of the primary tumor.	[T1a, T1b, T2, T3a, T3b, T4a, T4b]
13	Lymph Nodes (N)	Represents the N (Node) stage of thyroid cancer, indicating the involvement of nearby lymph nodes.	[N0, N1b, N1a]
14	Metastasis	Represents the M (Metastasis) stage of thyroid cancer, indicating whether the cancer has spread to distant organs.	[M0, M1]
15	Stage	Represents the overall stage of thyroid cancer based on the combination of T, N, and M stages.	[I, II, IVB, III, IVA]
16	Treatment Response	Describes the response to treatment, including categories	[Indeterminate, Excellent, Structural Incomplete, Biochemical Incomplete]
17	Recurred	Indicates whether thyroid cancer has recurred	[No, yes]

Data Pre-processing

Data pre-processing is an essential step before building ML-based models. Making suitable data for the models has a significant impact on the prediction performances. In the initial preprocessing stage, the dataset was checked for missing data and no missing data was found in the data set. In the next stage, considering that most of the categorical variables in the dataset are unordered and the dataset is relatively small, label encoding—a method that converts each categorical value into a unique numeric label—was applied. This process transforms categorical variables into numeric values, enabling the model to process them more effectively. The subsequent stage is the feature selection process. Not all features in the dataset contribute equally to the performance of the model, and some may provide redundant information, increasing the risk of model overfitting. Therefore, feature selection is critical to ensure that only the most meaningful and informative features are included in the model. Recursive Feature Elimination (RFE) with Random Forest as the base estimator is an iterative feature selection method and identifies the most important features to improve the accuracy of the model. In this technique, a model is initially trained with all features. The model evaluates the contribution of each feature to the accuracy and, at each iteration, removes the least important features. This process is repeated one feature at a time. The removal of features is based on changes in the model's performance. This process continues until the features that most improve the accuracy of the model remain (17). This approach selected the 11 most prominent features (age, gender, thyroid function, physical examination, adenopathy, focality, risk, T, N, stage and response) out of 16 features in the dataset.

Machine Learning Algorithms

Five classifiers—Random Forest, Support Vector Classifier, AdaBoost, Gradient Boost, and Logistic Regression—were trained and evaluated to predict DTC recurrence. Random Forest (RF) is an ensemble learning algorithm consisting of a large number of decision trees and each tree makes decisions independently, then the results are combined (18). Support Vector Classifier (SVC) is a classification algorithm that tries to find the optimal hyperplane to classify the data (19). AdaBoost is an ensemble learning algorithm used to build a strong learner from weak learners, it works by weighting misclassified examples more. Gradient Boost (GB) is an algorithm for minimizing errors, where weak learners are successively combined to form a strong model (20). Finally, Logistic Regression (LR) is a regression model that makes probability estimates by modeling linear relationships and is widely used for classification problems (21). The ML algorithms were evaluated by analyzing their statistical significance through recall, precision, accuracy, and area under curve (AUC) performance criteria. It is essential to utilize a suitable evaluation technique. One such method is K-fold cross-validation, used to determine the mean accuracy of a model. The strategy of this method is to build a cross-validation approach where a certain k value is selected and the dataset is split into k subset of equal size. At each iteration, one of the k subsets is utilized as a test set while the remaining subsets are employed for model training. This process continues until all subsets are used as test sets once. This method utilizes the mean of the calculated values as a performance measure (22). A 5-fold cross-validation technique was applied to validate the predictive performance of the models. All analyses were conducted using Python 3.7.12 (Python Software Foundation), and relevant libraries for ML and XAI approaches.

The complex nature of ML-based models requires better explanations of how models make predictions and which input features contribute more to a model's decision. XAI models refer to a set of processes and techniques that are intended to provide a clear and understandable explanation for decisions generated by ML models (23). In this study, permutation feature importance (PFI) and SHapley Additive exPlanations (SHAP) methods are used as XAI methods. The PFI technique is utilized to retrieve the importance of features depending on their effects on the prediction of a trained a ML algorithm. PFI measures the decrease in model performance after permuting the values of each feature. Features are ranked in descending order based on their impact on the model's performance (24). SHAP approach yields important information regarding the model. It enables the identification of the most influential features contributing to the prediction process. It indicates how a feature influences an individual prediction in comparison to others. The SHAP importance is represented as an absolute value for model training, taking into account both direction and magnitude (25).

RESULTS

A publicly available clinical dataset, titled **"Differentiated Thyroid Cancer Recurrence,"** was utilized to predict the probability of recurrence. For this purpose, different ML algorithms were used in the prediction process. The prediction performances of ML algorithms are tabulated in Table 2. In addition, AUC values showing the discriminative power of ML algorithms in general were also depicted in Figure 2. Among the ML algorithms, the RF algorithm predicted DTC recurrence probability as the most successful model with 97.39% of accuracy, 92.98% of precision, 98.15% of recall, and 95.50% of F1-Score. As highlighted in Figure 2, all models showed acceptable discriminative with AUC values ranging from 0.92 to 0.99, while the RF model performed the best with an AUC value of 0.993.

Table 2. Accuracy, precision, recall and F1-score values for ML algorithms

Models	Mean Accuracy	Mean Precision	Mean Recall	Mean F1-Score
LR	0.9191	0.8667	0.8426	0.8545
SVC	0.9373	0.8962	0.8796	0.8879
ADA	0.9452	0.9780	0.8241	0.8945
GB	0.9687	0.9615	0.9259	0.9434
RF	0.9739	0.9298	0.9815	0.9550

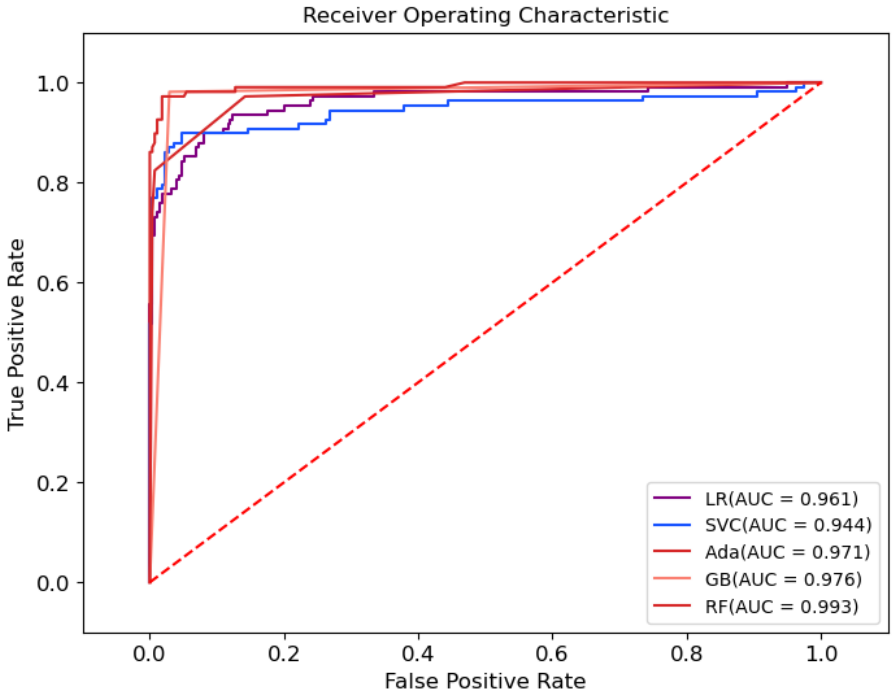

Figure 1I: ROC curves of ML algorithms

PFI and Shapley approaches were used to identify the features that contribute the most to the prediction performance of the RF algorithm, which has the highest prediction rate. The PFI plots for the test and training datasets and the results of the SHAP analysis are presented in Figure 3. Treatment response and ATA risk are the most critical features affecting the prediction process based on SHAP and PFI analysis.

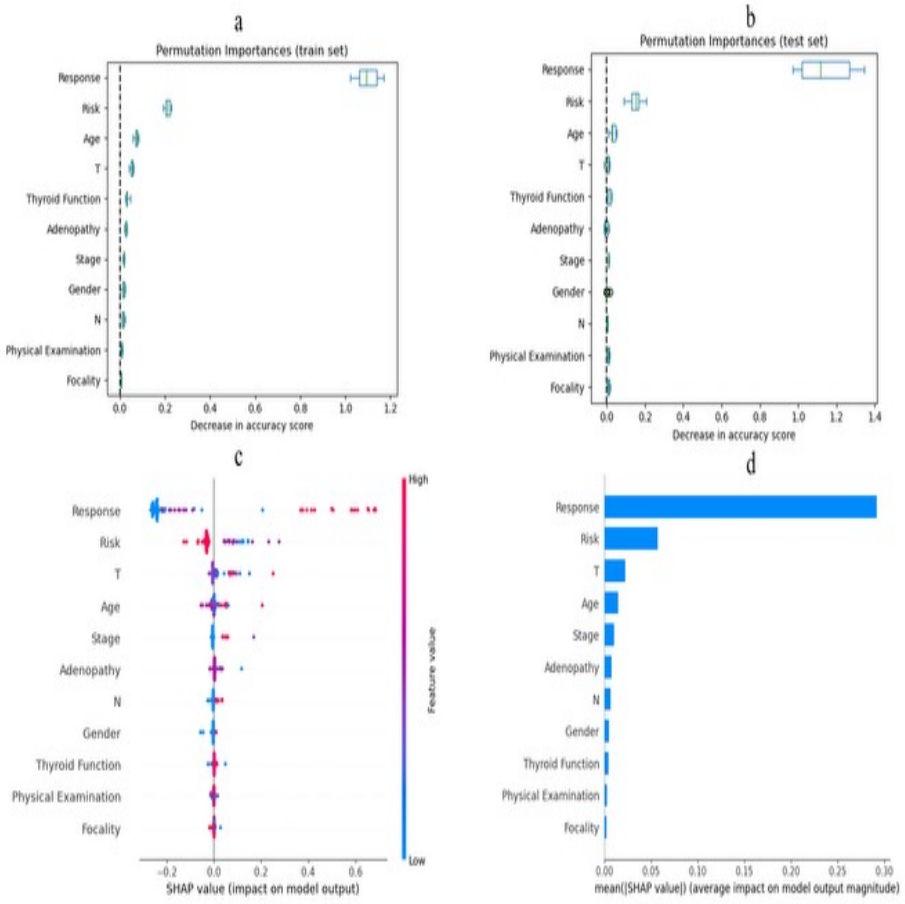

Figure 3: The most influential features for predicting differentiated thyroid cancer recurrence according to PFI and SHAP analyses in the Random Forest model. (a) PFI plot for training set, (b) PFI plot for test set, (c) SHAP Bee-swarm plot, and (d) SHAP summary plot.

DISCUSSION

This study adopted five different ML algorithms to predict DTC recurrence and leveraged XAI techniques to improve the interpretability of model results. As a result of the analysis, the RF algorithm stood out as the most effective model, achieving an accuracy of 97.39% and demonstrating superior discriminative ability with an AUC of 0.993, indicating excellent prediction accuracy in differentiating between recurrence and non-recurrence cases compared to other models. This algorithm performs as an ensemble method by combining a large number of decision trees and is particularly notable for its capacity to capture complex, nonlinear relationships, while reducing the risk of overfitting by training the trees on random subsets (26). In this respect, RF is particularly well-suited for analyzing complex, multidimensional medical datasets, such as the DTC recurrence dataset used in this study.

It is highly clinically important not only that the model performs well in prediction, but also that it can explain the properties on which these results are based. In this context, two different XAI approaches were adopted in the study to better understand the decision process of the model: PFI and SHAP. While PFI measures the

impact of an attribute on model accuracy by randomly mixing its values, the Shapley method computes the contribution of each feature to the prediction in a more detailed and fair way based on game theory. While PFI is faster and more practical, it can give misleading results in the case of highly correlated variables. SHAP method is more complex, but clearly shows the effect of features at the level of individual predictions (27). It was found that "response to treatment" and "ATA risk" variables were the most determinant factors in the predictions of the model in the analyses performed with both these methods. Ruben et al. (28) also emphasized that this classification was an effective variable in recurrence prediction. On the other hand, response to treatment presents dynamic information about the course of the disease during follow-up. Studies such as Tuttle et al. (29) and Park et al. (30) stated that this parameter could change over time and was a valuable indicator in understanding the likelihood of recurrence. These findings align with the key predictors identified through XAI in our study. Other significant variables highlighted by the model are T stage and age. The majority of patients with recurrence were identified to be in T3 and T4 stages. This is consistent with the literature that tumor size (especially above 4 cm) increases the risk of recurrence (31). Moreover, Altay et al. (32) reported that the likelihood of recurrence increases with increasing age. The model's significant identification of these two factors is consistent with clinical expectations.

The RF-based XAI model proposed in this study not only predicted with high accuracy, but also provided a clear and visual representation of the clinical variables that influence the prediction process. This approach can assist healthcare professionals to make more reliable and informed decisions.

However, the study had some limitations. Since the data set used was obtained from the UCI data repository, which is an open-access resource, and not from the real clinical setting, further studies should be conducted in different patient populations and real clinical conditions to test the generalizability of the proposed model. In the future, we are planning to further improve the accuracy, interpretability and clinical applicability of the model with local patient data in collaboration with expert physicians and to evaluate the use of advanced algorithms such as deep learning.

The proposed model performed well in predicting DTC recurrence. By combining high-performance algorithms and interpretability methods, the clinical variables most associated with recurrence were successfully identified. These results demonstrate the potential of the proposed XAI-integrated ML framework to enhance personalized follow-up strategies and to inform evidence-based clinical decision-making.

REFERENCES

1. Dralle H, Machens A, Basa J, Fatourechi S, Hay ID, et al. Follicular cell-derived thyroid cancer. Nat Rev Dis Primers. 2015 1: 15077.
2. Zhao H, Liu CH, Cao Y, Zhang LY, Zhao Y, Liu YW, et al. Survival prognostic factors for differentiated thyroid cancer patients, with pulmonary metastases: A systematic review and meta-analysis. Front Oncol. 2022 15:12:990154.
3. Na'ara S, Amit M, Fridman E, Gil Z. Contemporary management of recurrent nodal disease in differentiated thyroid carcinoma. Rambam Maimonides Med J. 2016 28: 7(1):e0006.
4. Clark E, Price S, Lucena T, Haberlein B, Wahbeh A, Seetan R. Predictive analytics for thyroid cancer recurrence: a machine learning approach. Knowledge. 2024 4(4):557-570.
5. Schindele A, Krebold A, Heiß U, Nimptsch K, Pfaehler E, Berr C, et al. Interpretable machine learning for thyroid cancer recurrence predicton: Leveraging XGBoost and SHAP analysis. European Journal of Radiology, 2025 186, 112049.
6. Medas F, Canu GL, Boi F, Lai ML, Erdas E, Calò PG. Predictive factors of recurrence in patients with differentiated thyroid carcinoma: A retrospective analysis on 579 patients. Cancers (Basel). 2019 Aug 22 11(9): 1230.
7. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med. 2021 2(6):642-665.
8. Taha K. Machine learning in biomedical and health big data: a comprehensive survey with empirical and experimental insights. J Big Data. 2025 12(1):61.
9. Wang H, Zhang C, Li Q, Tian T, Huang R, Qiu J, et al. Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches. BMC Cancer. 2024 24(1):427.
10. 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.
11. Setiawan KE. Predicting recurrence in differentiated thyroid cancer: a comparative analysis of various machine learning models including ensemble methods with chi-squared feature selection. Commun Math Biol Neurosci. 2024, 2024, 55.
12. Yasar S. Determination of possible biomarkers for predicting well-differentiated thyroid cancer recurrence by different ensemble machine learning methods. Middle Black Sea J Health Sci. 2024 10(3):255-265.
13. Frasca M, La Torre D, Pravettoni G, Cutica I. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. Discov Artif Intell. 2024 4(1):15.

14. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020, 20:1-9.

15. Minh D, Wang HX, Li YF, Nguyen TN. Explainable artificial intelligence: a comprehensive review. Artif Intell Rev. 202, 1-66.

16. UCI Machine Learning Repository. Differentiated thyroid cancer recurrence dataset. Available from: https://archive.ics.uci.edu/dataset/915/differentiated+thyroid+cancer+recurrence [Accessed 2024 Jun 4].

17. Yin Y, Jang-Jaccard J, Xu W, Singh A, Zhu J, Sabrina F, et al. IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. J Big Data. 2023 10(1):15.

18. Mohandoss DP, Shi Y, Suo K. Outlier prediction using random forest classifier. In: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC); January 2021. p. 27-33. IEEE.

19. Zhang C, Shao X, Li D. Knowledge-based support vector classification based on C-SVC. Procedia Comput Sci. 2013 17:1083-1090.

20. Schapire R. E. The boosting approach to machine learning: An overview. In: Denison D.D., Hansen M.H., Holmes C.C., Mallick B., Yu B., eds. Nonlinear Estimation and Classification. Lecture Notes in Statistics, vol 171. Springer; 2003, 203-225.

21. LaValley MP. Logistic regression. Circulation. 2008 117(18):2395-2399.

22. Baghdadi NA, Farghaly Abdelaliem SM, Malki A, Gad I, Ewis A, Atlam E. Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. J Big Data. 2023 10(1):144.

23. Band SS, Yarahmadi A, Hsu CC, Biyari M, Sookhak M, Ameri R, et al. Application of explainable artificial intelligence in medical health: a systematic review of interpretability methods. Informatics Med Unlocked. 2023 40:101286.

24. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010 26(10):1340-1347.

25. Mosca E, Szigeti F, Tragianni S, Gallagher D, Groh G. SHAP-based explanation methods: a review for NLP interpretability. In: Proceedings of the 29th International Conference on Computational Linguistics; October 2022, 4593-4603.

26. Cutler A, Cutler DR, Stevens JR. Random forests. In: Ensemble Machine Learning: Methods and Applications. Springer. 2012, 157-175.

27. Kök I, Okay FY, Muyanlı O, Ozdemir S. Explainable artificial intelligence (XAI) for Internet of Things: a survey. IEEE Internet Things J. 2023 10(16):14764-14779.

28. Ruben R, Pavithran PV, Menon VU, Nair V, Kumar H. Performance of ATA risk stratification systems, response to therapy, and outcome in an Indian cohort of differentiated thyroid carcinoma patients: a retrospective study. Eur Thyroid J. 2019 8(6):312-318.

29. Tuttle RM, Tala H, Shah J, Leboeuf R, Ghossein R, Gonen M, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. Thyroid. 2010 20(12):1341-1349.

30. Park S, Kim WG, Song E, Oh HS, Kim M, Kwon H, et al. Dynamic risk stratification for predicting recurrence in patients with differentiated thyroid cancer treated without radioactive iodine remnant ablation therapy. Thyroid. 2017 27(4):524-530.

31. Ito Y, Miyauchi A, Kihara M, Fukushima M, Higashiyama T, Miya A. Overall survival of papillary thyroid carcinoma patients: a single-institution long-term follow-up of 5897 patients. World J Surg. 2018 42:615-622.

32. Altay FP, Cicek O, Demirkan E, Taşkaldiran I, Bozkus Y, Turhan Iyidir O, et al. Evaluation of prognosis and risk factors of differentiated thyroid cancer in a geriatric population. Turk J Geriatr. 2023 26:118-123.

Abbreviations list

DTC Differentiated Thyroid Cancer
ML Machine Learning
XAI Explainable Artificial Intelligence
SHAP Shapley Additive Explanations
PFI Permutation Feature Importance
RF Random Forest
SVC Support Vector Classifier
ADA AdaBoost
GB Gradient Boosting
LR Logistic Regression
AUC Area Under the Curve
ATA American Thyroid Association
RFE Recursive Feature Elimination
ROC Receiver Operating Characteristic
UCI University of California Irvine

Ethics approval and consent to participate

Since this study was conducted on a publicly available clinical data set, Ethics Committee approval is not required.

Consent for publication

Not applicable.

Availability of data and materials

Data is available at https://archive.ics.uci.edu/ (accessed on 04 June 2024).

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

Conceptualization, A.C.O, E.A, S.C.; methodology, A.C.O and E.A; software, A.C.O and E.A; formal analysis, A.C.O and E.A; data curation, A.C.O, E.A and S.C.; writing—original draft preparation, A.C.O, E.A and S.C.; writing—review and editing, A.C.O, E.A and S.C.; supervision, A.C.O and E.A.

Acknowledgements

The authors would like to thank the UCI Machine Learning Repository for providing the publicly available dataset used in this study.