Middle Black Sea Journal of Health Science / Mid Blac Sea J Health Sci
August 2023; 9(3): 440-452

RESEARCH ARTICLE
DOI: 10.19127/mbsjohs.1282265

Machine Learning Approach for Thyroid Cancer Diagnosis Using Clinical Data

İpek Balıkçı Çiçek[1(ID)], Zeynep Küçükakçalı[1(ID)]

[1]Inonu University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey.

Received: 13 April 2023, Accepted: 16 July 2023, Published online: 31 August 2023

Abstract

Objective: With an early diagnosis of thyroid cancer, one of the world's most significant health issues, it is feasible to treat the nodules before the spread of malignant thyroid gland cells. It has become crucial to develop models for predicting thyroid cancer. In light of this, the purpose of this study is to develop a clinical decision support model using the Bagged CART model, a machine learning (ML) model for the prediction of thyroid cancer.

Methods: Between 2010 and 2012, 724 patients who applied to China Median University Shengjing Hospital comprised the study's data set. The dataset comprises information on nodule malignancies, demographic characteristics, ultrasound characteristics, and blood test results for all patients who underwent thyroidectomy. Using this open-access data set, the Bagged CART modeling technique was applied. Negative predictive value (NPV), specificity (Spe), balanced accuracy (BACC), positive predictive value (PPV), accuracy (ACC), sensitivity (Sen), and F1-score performance metrics were used to evaluate the model's predictive performance. In addition, a 10-fold cross-validation method was used to determine the validity of the model. In addition, variable importance was established, which reveals how much the input variables impact the output variable.

Results: ACC, BACC, Sen, Spe, PPV, NPV, and F1-score obtained from the model performance metrics were calculated to 99.1%, 98.7%, 99.7%, 97.7%, 99.1%, 99.2%, and 99.4%, respectively, as a result of modeling. According to the variable importance values that were acquired for the input variables in the dataset that was investigated in this study, the seven variable that hold the greatest significance are as follows: size, TSH, blood flow: size, TSH, blood flow: enriched, multilateral: yes, FT4, site: isthmus, and age, in that order.

Conclusion: As a result, the Bagged CART model was found to be effective at predicting thyroid cancer based on the findings of this study. In addition, in this study, risk factors for thyroid cancer were evaluated and their importance values were given. With these results, the decision-making process about the disease will be able to accelerate and thus, it will be able to effective in preventive medicine practices.

Key words: Bagged CART, machine learning, thyroid cancer, risk factors, classification.

Suggested Citation: Balıkçı Çiçek İ., Küçükakçalı Z. Machine Learning Approach for Thyroid Cancer Diagnosis Using Clinical Data. Mid Blac Sea Journal of Health Sci, 2023;9(3):440-452.

Address for correspondence/reprints:

İpek Balikçi Çiçek

Telephone Number: +90 (553) 373 2404

E-mail: ipek.balikci@inonu.edu.tr

INTRODUCTION

Thyroid cancer is caused by the abnormal growth of malignant tumor cells in thyroid gland tissue. The development of thyroid cancer occurs when cells within a malignant tumor change or adapt their cellular structure. The frequency of thyroid cancer has gradually climbed during the previous three decades (1). The American Cancer Society estimates that there were around 43.800 newly diagnosed cases of thyroid cancer in 2022, with roughly 2.230 people losing their lives to the disease. Cancer of the thyroid is a solid tumor that most often manifests in the thyroid gland as a nodule or mass at the front base of the neck (2). Thyroid cancer is the most common type of cancer in teens and young adults, and it is the ninth most common type of cancer in women overall (3). The death rate for thyroid cancer is quite low; nevertheless, the sickness recurrence rate is rather high, which is directly correlated to a greater level of incurability, morbidity, and mortality among patients (4). Even though there are several different varieties of thyroid cancer, the two types that are most common account for around 95% of all occurrences of thyroid cancer. The two types of thyroid cancer are papillary and follicular thyroid cancer (5).

Early detection of malignant thyroid nodules can lead to efficient therapy and minimal harm if the nodules are treated before the cancerous cells in the thyroid gland spread (6). In addition, having a pre-surgical diagnosis that determines whether thyroid nodules are benign or malignant helps the procedure go more smoothly and lowers the chance of recurrence following surgery. Thyroid cancer screening is a process that enables the early identification of malignant thyroid nodules (7). Neck palpation during a physical examination and ultrasonography, which may identify both nonpalpable and palpable nodules, particularly those with a diameter of less than 1 cm, are the two main approaches for detecting thyroid cancer (8). Ultrasonography (US), as the primary diagnostic tool, is used to detect the features of thyroid nodules. These discovered traits assist in the classification of nodules as benign or malignant (9). Until date, the bulk of malignant nodule diagnoses are based on surgeons' and radiologists' clinical experience (10). Human judgment is slow and prone to mistakes in many circumstances. To enhance medical decisions and minimize labor effort, accurate and comprehensible prediction models are desperately needed (11).

As a novel technology, computer-aided diagnosis (CAD) has been used to diagnose thyroid nodules in recent decades. The use of artificial intelligence into CAD tools makes them more intelligent and enhances the quality and consistency of ultrasound feature interpretation, resulting in fewer needless biopsies. The basic methodologies of artificial intelligence-based CAD systems that have a significant influence on the medical field are ML (12).

In the medicine, like in all other sectors of human knowledge, novel approaches based on ML and collaborative reasoning are utilized in an attempt to succeed where traditional predicting methods have failed (13). ML algorithms uncover more complex associations among existing data than traditional regression models. A ML system is one that studies the design and operation of data-learning and prediction-capable algorithms. Based on input samples, these algorithms create a model to make and anticipate decisions (14).

CART is a continually improving nonparametric ML tool for regression and classification problems. CART recursively partitions data based on the binary partitioning technique to investigate the link between response variables and predictors (15). Bagged CART is an advancement on the CART algorithm that combines CART with bagging techniques in order to improve predictive model performance and reduce overfitting. Bagged CART was developed to improve the CART algorithm. (16). Each classifier in this approach generates and saves its model by categorizing a piece of the data. Eventually, based on vote intention among these categories, the class with the highest votes is chosen as the final classifier (17).

By employing the ML method of Bagged CART on an open access dataset consisting of open access with benign and malignant nodules patients, the purpose of this investigation is to classify thyroid cancer and identify the factors associated with it.

METHODS

Dataset and Variables

The data for this investigation were obtained from 724 patients admitted between 2010 and 2012 to the Shengjing Hospital of China Median University. The datasets comprise information on each patient's nodule malignancy, demographics, ultrasound characteristics, and blood test results. Each patient had a single or multiple nodules in three distinct areas: the right lobe, the left lobe, and the isthmus. If a patient had multiple nodules in a single region, only the largest one was included in the dataset. The "Thyroid" open-access dataset used in this study is available at https://zenodo.org/record/6465436#.Y06MQ3ZBxZ.In this dataset, there are a total of 724 patients, 204(28.2%) of whom are benign nodules patients and 520(71.8%) are malignant nodules patients. Table 1 shows the variables included in the dataset.

Table 1: The variables contained inside the dataset.

Variables	**Variable Types**	**Variable Roles**
Age: The Age of the Patient	Quantitative	Predictor
Gender: 0: Male, 1: Female	Qualitative	Predictor
FT3: Triiodothyronine Test Result	Quantitative	Predictor
FT4: Thyroxine Test Result	Quantitative	Predictor
TSH: Thyroid-Stimulating Hormone Test Result	Quantitative	Predictor
TPO: Thyroid Peroxidase Antibody Test Result	Quantitative	Predictor
TGAb: Thyroglobulin Antibodies Test Result	Quantitative	Predictor
Site: The Nodule Location, 0: Right, 1: Left, 2: Isthmus	Qualitative	Predictor
Echo Pattern: Thyroid Echogenicity, 0: Even, 1: Uneven	Qualitative	Predictor
Multifocality: If Multiple Nodules Exist in One Location, 0: No, 1: Yes	Qualitative	Predictor
Size: The Nodule Size in Cm	Quantitative	Predictor
Shape: The Nodule Shape, 0: Regular, 1: Irregular	Qualitative	Predictor
Margin: The Clarity of Nodule Margin, 0: Clear; 1: Unclear	Qualitative	Predictor
Calcification: The Nodule Calcification, 0: Absent, 1: Present	Qualitative	Predictor
Echo Strength: The Nodule Echogenicity, 0: None, 1: Isoechoic, 2: Medium-Echogenic, 3: Hyperechogenic, 4: Hypoechogenic	Qualitative	Predictor
Blood Flow: The Nodule Blood Flow, 0: Normal, 1: Enriched	Qualitative	Predictor
Composition: The Nodule Composition, 0: Cystic, 1: Mixed, 2: Solid	Qualitative	Predictor
Multilateral: If Nodules Occur in More Than One Location, 0: No, 1: Yes	Qualitative	Predictor
Mal: The Nodule Malignancy, 0: Benign, 1: Malignant	Qualitative	Output

Bagged Classification and Regression Trees (Bagged CART)

The non-parametric decision tree logging method known as CART has become quite popular. Breiman et al. devised this technique (1984) (18). Binary trees serve as the foundation for this strategy. This tree serves as the foundation for more complicated algorithms such as Random Forest, in addition to other trees. The CART decision tree method first divides the input into binary components before moving on to the construction of the decision tree. In order to identify which variables should be provided additional information on classification, the CART tree makes use of the Gini index. When classifying, variables that have lower Gini indices are given a larger amount of weight. The CART algorithm uses trial and error to figure out the best possible value for the separator point in each dimension or variable, which leads to a lower Gini index in the end (19).

The bagging approach has the potential to significantly enhance the accuracy of the CART, which is well recognized as an unstable model (20). The bagged CART improves classification performance, eliminates overfitting, and considerably reduces prediction variation. The CART algorithm begins by performing a recursive split on the training sample units using a predetermined number of variables. The method then assesses each of the predictive elements in order to ascertain which binary division of a predictive variable is most likely to diverge from the variable that was anticipated as the result of the analysis. In order to construct homogenous end nodes in a hierarchical tree, the method is frequently repeated for each of the initial split outcomes. CART prunes the trees to minimize overfitting when the results of cross validation give the lowest error rate (15, 21).

Biostatistical analysis

Qualitative data from the variables included in the study were summarized with number (percentage). The Kolmogorov-Smirnov test was utilized in order to investigate whether or not the quantitative data adhered to a normal distribution. Data that did not show normal distribution were summarized with the median (minimum-maximum). Normally distributed ones were summarized as mean±standard deviation. In the statistical analyses, the Pearson chi-square test, the Continuity Correction test, and the Mann-Whitney U test were utilized, depending on the circumstances, to determine whether or not there is a statistically significant difference between the target variable and the input variables. In the statistical analyses that were carried out, a value of p less than 0.05 ($p<0.05$) was regarded as statistically significant. All of the analyses were carried out with the help of IBM SPSS Statistics 26.0 for Windows (New York; USA).

Machine Learning Modeling and Performance Evaluation

During the modeling phase of the investigation pertaining to the aforementioned data set, the Bagged CART technique was employed. For modeling, the entire data set was utilized. For the analysis, the technique of n-fold cross validation was utilized. n-fold cross-validation divides the data into n segments and applies the model to n of them. One of the n components is utilized for testing, while the remaining n-minus-one components are used to educate the model. The study employed a 10-fold cross-validation method to increas the model's validity. As criteria for performance evaluation, the BACC, ACC, Spe, Sen, NPV, PPV, and F1-score were utilized. In addition, variable importance was determined, which represents the extent to which the input variables influence the target variable. Modeling was accomplished using R Studio 4.2.1. (22).

RESULTS

The study's data set contains 724 individuals, 204 (28.2%) of whom had benign nodules and 520 (71.8%) have malignant nodules. The patients' mean age was 45.59±12.609 years. The median age of patients with benign nodules was 48 (15-79) years, whereas the median age of patients with malignant nodules was 44 years (13-82 years). The study included 121 (16.7%) men and 603 (83.3%) women.

In addition, of the male patients, 27(22.3%) have benign nodules and 94 (77.7%) have malignant nodules. While 177(29.4%) of the female patients have benign nodules, 426(%70.6) have malignant nodules.

Table 2 displays the results of statistical analyses of the independent variables in terms of the dependent variable.

Table 2. The results of the statistical analyses between the target variable and independent variables

Variables		The Nodule Malignancy		*p*
		Benign	Malignant	
Gender n(%)	Male	27 (13.24)	94 (18.08)	0.116**
	Female	177 (86.76)	426 (81.92)	
Site n(%)	Right	80 (39.22)	275 (52.88)	**<0.001****
	Left	82 (40.20)	230 (44.23)	
	Isthmus	42 (20.59)	15 (2.88)	
Echo Pattern n(%)	Even	187 (91.67)	449 (86.35)	0.065***
	Uneven	17 (8.33)	71 (13.65)	
Multifocality n(%)	No	117 (57.35)	310 (59.62)	0.578**
	Yes	87 (42.65)	210 (40.38)	
Shape n(%)	Regular	194 (95.10)	363 (69.81)	**<0.001*****
	Irregular	10 (4.90)	157 (30.19)	
Margin n(%)	Clear	93 (45.59)	131 (25.19)	**<0.001****
	Unclear	111 (54.41)	389 (74.81)	
Calcification n(%)	Absent	178 (87.25)	229 (44.04)	**<0.001****
	Present	26 (12.75)	291 (55.96)	
Echo Strength n(%)	None	5 (2.45)	3 (0.58)	**0.002****
	Isoechoic	5 (2.45)	7 (1.35)	
	Medium-echogenic	31 (15.20)	46 (8.85)	
	Hyperechogenic	3 (1.47)	1 (0.19)	
	Hypoechogenic	160 (78.43)	463 (89.04)	
Blood Flow n(%)	Normal	173 (84.80)	250 (48.08)	**<0.001****
	Enriched	31 (15.20)	270 (51.92)	
Composition n(%)	Cystic	10 (4.90)	8 (1.54)	**0.012****
	Mixed	19 (9.31)	34 (6.54)	
	Solid	175 (85.78)	478 (91.92)	
Multilateral n(%)	No	20 (9.80)	0 (0.00)	**<0.001*****
	Yes	184 (90.20)	520 (100.00)	
		Median(Min-Max)	Median(Min-Max)	
Age		48 (15-79)	44 (13-82)	**0.002***
FT3		4.44 (2.63-22.9)	4.32 (2.47-15.43)	0.196*
FT4		14.595 (6.65-59.08)	14.43 (5-28.76)	0.388*
TSH		1.327 (0.002-56.254)	1.561 (0-101)	**0.018***
TPO		0.595 (0-1001)	0.64 (0-1001)	0.950*
TGAb		2.695 (0-1001)	2.825 (0-1001)	0.980*
Size		0.9 (0.05-5.4)	1.6 (0-9)	**<0.001***

***: Mann Whitney U test, **:Pearson chi-square test, ***: Continuity Correction test, Min: Minimum, Max: Maximum**

ACC, BACC, Sen, Spe, PPV, NPV, and F1-score obtained from the Bagged CART model as a result of the modeling were 99.1%, 98.7%, 99.7%, 97.7%, 99.1%, 99.2%, and 99.4%, respectively.

In Figure 1, values of performance metrics obtained from Bagged CART model are shown.

Table 3: Performance metric values calculated from Bagged CART model.

Performance Metrics	Value (%)
ACC	99.1
BACC	98.7
Sen	99.7
Spe	97.7
PPV	99.1
NPV	99.2
F1-score	99.4

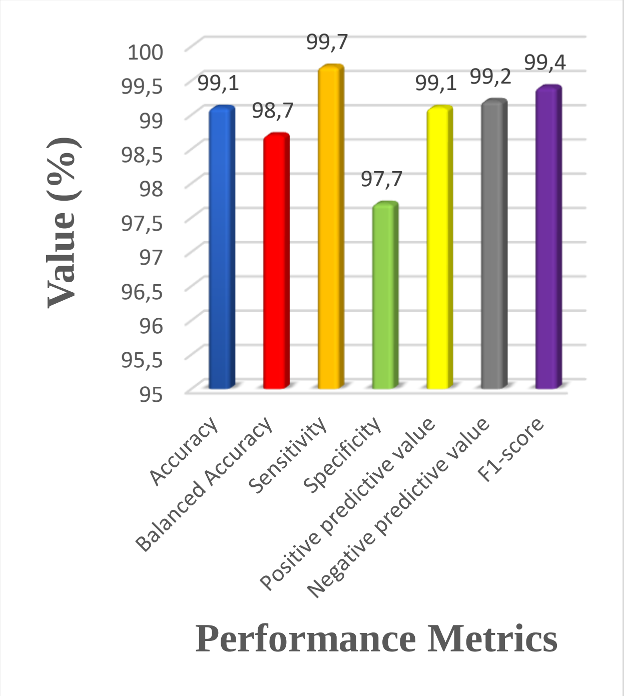

Figure 1: Performance metric values acquired from Bagged CART model.

Table 4 is a table of the variable importance values computed as a consequence of the Bagged CART model.

Figure 2 depicts a graph of the variable importance values obtained as consequence of the Bagged CART model. The seven most important variables related to multiple sclerosis were found as size, TSH, blood flow: enriched, multilateral: yes, FT4, site: isthmus, age.

Table 4: The variable importance values obtained as a result of the Bagged CART model

Variables	Importance Values
Size	100
TSH	63.292
Blood Flow: Enriched	59.741
Multilateral: Yes	59.293
FT4	56.479
Site: Isthmus	54.789
Age	53.434
FT3	53.179
Calcification: Present	50.695
TGAb	48.839
TPO	48.814
Shape: Irregular	33.286
Margin: Unclear	19.02
Multifocality: Yes	11.994
Site: Left	8.716
Echo Strength: Hypoechogenic	7.711
Echo Strength: Medium-Echogenic	6.201
Composition: Solid	5.554
Gender: Female	4.592
Echo Pattern: Uneven	4.573
Echo Strength: Isoechoic	2.882
Composition: Mixed	2.627

Figure 2 depicts a graph of the variable importance values obtained as consequence of the Bagged CART model. The seven most important variables related to multiple sclerosis were found as size, TSH, blood flow: enriched, multilateral: yes, FT4, site: isthmus, age.

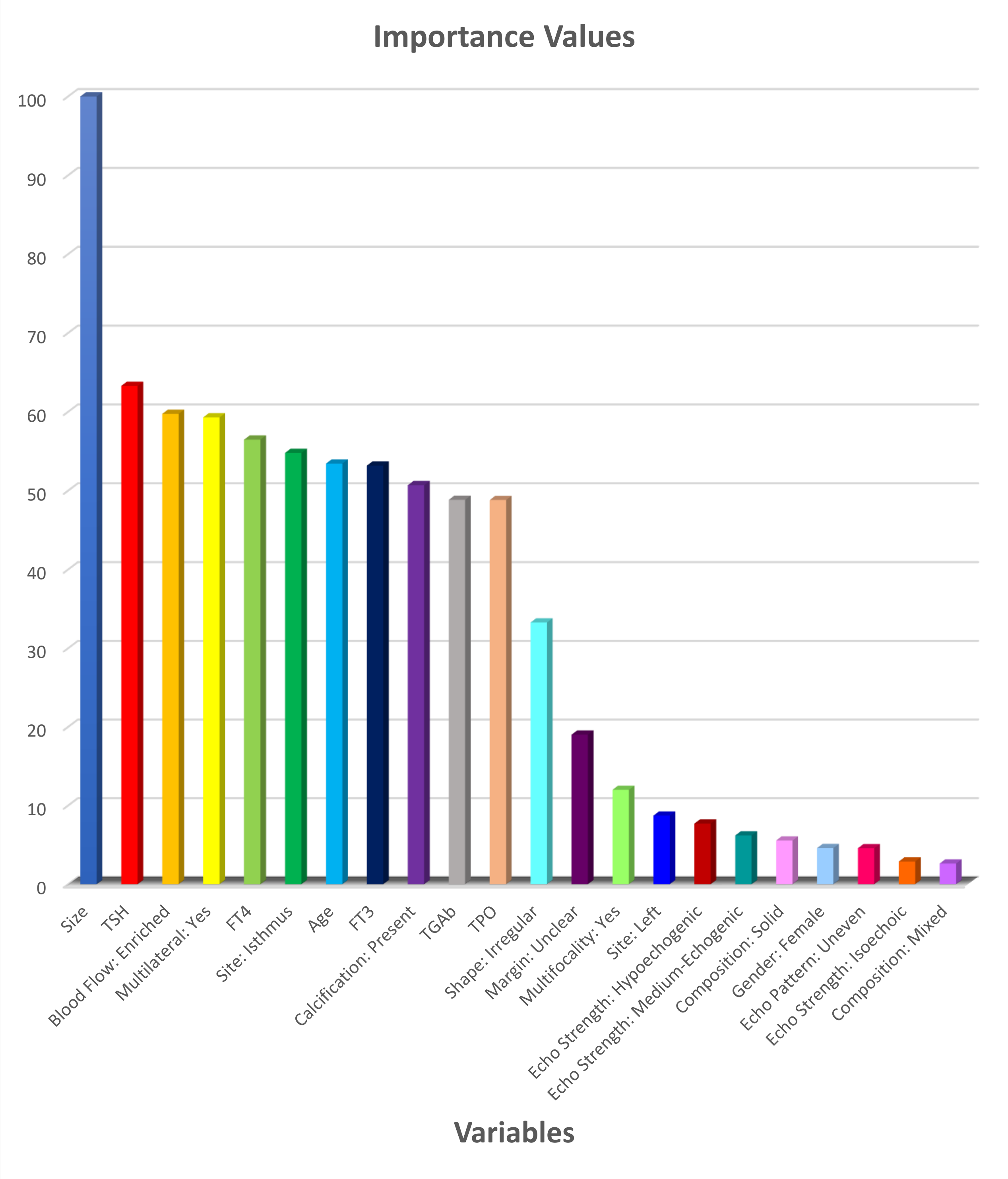

Figure 2: The variables' importance values as a consequence of the Bagged CART mode

DISCUSSION

As a crucial component of the human body, the thyroid generates a variety of hormones that perform several vital functions in the human body. Thyroid illness therefore endangers the human body in all physiological systems, including the endocrine, circulatory, neurological, respiratory, digestive, muscular, and reproductive systems (23). Thyroid diseases and disorders are widespread hormonal conditions that affect the overwhelming majority of the global population. Diseases and conditions affecting the thyroid include thyroiditis, thyroid nodules, and thyroid carcinoma. The prevalence of thyroid nodules and thyroid cancer is rising globally, predominantly owing to improved diagnostic procedures. The widespread use of US has exponentially increased thyroid nodule detection to approximately 20–67% (24).

It is essential to distinguish between benign and malignant thyroid nodules in order to prevent performing unneeded fine-needle aspiration biopsies and overtreating the condition, such as through surgery (25). The intricate nature of the nodules leads in complicated ultrasound pictures and perhaps mixed signals between benign and malignant nodules. As a result, the US cannot distinguish between cancerous and benign nodules. Early detection and classification of benign and malignant thyroid nodules is critical for directing clinical therapy and choosing surgical methods (26).

It has been demonstrated that machine-learning algorithms produce much more accurate predictions than human experts. ML models are becoming more popular and widely employed in a variety of fields. These models' main aim is to determine the effective factors and their relationships, and they may also be used to forecast. These models are a branch of artificial intelligence that may be utilized as a study and application area in a variety of fields. Additionally, ML techniques are frequently employed and applied in medical science for illness detection (27-29). For this reason, in the study, the classification of thyroid cancer using the Bagged CART model, which is an ML approach, and the predictive features related with the diagnosis of the disease were revealed with variable importance values.

The performance criteria obtained from the Bagged CART method result, accuracy, balanced ACC, BACC, Sen, Spe, PPV, NPV, and F1-score were obtained as 99.1%, 98.7%, 99.7%, 97.7%, 99.1%, 99.2%, and 99.4%, respectively. Successful findings were achieved for the diagnosis of thyroid cancer, and according to the variable importance produced as a consequence of the model, the seven variables most related with the diagnosis were size, TSH, blood flow: enriched, multilateral: yes, FT4, site: isthmus, and age, respectively.

Numerous investigations on various thyroid datasets have been conducted to date. Parikh et al. (2015), describe the two prediction models they built to address their multiclass classification challenge. They employed artificial neural networks and support vector machines, and the ANN obtained an accuracy of 97.17% (30). Ionita et al. (2016), investigated hybrid medical datasets and outlined a variety of applications of Naive

Bayes, Decision trees, MLP, and RBF networks. For classification accuracy, all classifiers classify and provide separate outcomes; however, a decision tree obtained 97.35% accuracy (31). Chaurasia et al. (2018), used a number of machine learning methods to examine data. Using Naive Bayes, they achieved 97.37% accuracy (32). Talasila et al. (2020), exhibited and evaluated numerous ways before determining that LightGBM gave more accurate predictions than other methods available (33). Kumar et al. (2020), used SVM to diagnose thyroid stage with an accuracy of 83.37%. Other classification techniques have been shown to be less efficient than Multiclass SVM. Furthermore, the model accurately differentiates between the four thyroid states (34). Aversano et al. used a range of machine learning methods to examine data. They specifically compared the output of 10 different classifiers. The other algorithms' performance is encouraging, notably the Extra-TreeClassifier, which obtains an accuracy of 84%. They also used a catboost classifier and attained a precision of 71% (35).

In addition, the machine learning method used is a tree-based method and includes some arrangements to increase model performance. Therefore, it gives better results than known basic machine learning methods. Thus, the model performance from this study demonstrates that clinically, it can reliably distinguish benign from malignant nodules.

As a result; early detection of malignant nodules is essential for effective disease management and reducing mortality rates. In the past ten years, the development of CAD systems utilizing artificial intelligence for the early detection of thyroid cancer has been extremely rapid. Thanks to these technologies, thyroid nodule treatment will develop. This article provides an review of the use of ML in the diagnosis of thyroid nodules. The overall result of this study showed that it will significantly benefit thyroid tumor classification with the latest advances in ML approaches and high specificity, sensitivity, accuracy and other performance metrics.

Ethical Approval: Ethics committee approval is not required in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept- İBÇ, Design- İBÇ, ZK, Supervision- Z K, Literature Review- İBÇ, Z K, Critical Review- İBÇ

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: There are no financial supports.

REFERENCES

1. Rossi ED, Pantanowitz L, Hornick JL. A worldwide journey of thyroid cancer incidence centred on tumour histology. The Lancet Diabetes and Endocrinology. 2021;9(4):193-4.

2. Anari S, Tataei Sarshar N, Mahjoori N, Dorosti S, Rezaie A. Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. Mathematical Problems in Engineering. 2022;2022.
3. Araque DVP, Bleyer A, Brito JP. Thyroid cancer in adolescents and young adults. Future Oncology. 2017;13(14):1253-61.
4. Tuttle RM, Ball DW, Byrd D, Dilawari RA, Doherty GM, Duh Q-Y, et al. Thyroid carcinoma. Journal of the National Comprehensive Cancer Network. 2010;8(11):1228-74.
5. Carcangiu ML, Steeper T, Zampi G, Rosai J. Anaplastic thyroid carcinoma: a study of 70 cases. American journal of clinical pathology. 1985;83(2):135-58.
6. Olson E, Wintheiser G, Wolfe KM, Droessler J, Silberstein PT. Epidemiology of thyroid cancer: a review of the National Cancer Database, 2000-2013. Cureus. 2019;11(2).
7. Lamartina L, Grani G, Durante C, Filetti S, Cooper DS. Screening for differentiated thyroid cancer in selected populations. The Lancet Diabetes & Endocrinology. 2020;8(1):81-8.
8. Lin JS, Bowles EJA, Williams SB, Morrison CC. Screening for thyroid cancer: updated evidence report and systematic review for the US Preventive Services Task Force. Jama. 2017;317(18):1888-903.
9. Keramidas EG, Iakovidis DK, Maroulis D, Karkanis S, editors. Efficient and effective ultrasound image analysis scheme for thyroid nodule detection. International Conference Image Analysis and Recognition; 2007: Springer.
10. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The diagnosis and management of thyroid nodules: a review. Jama. 2018;319(9):914-24.
11. Li T, Sheng J, Li W, Zhang X, Yu H, Chen X, et al. A new computational model for human thyroid cancer enhances the preoperative diagnostic efficacy. Oncotarget. 2015;6(29):28463.
12. Jin Z, Zhu Y, Zhang S, Xie F, Zhang M, Zhang Y, et al. Ultrasound computer-aided diagnosis (CAD) based on the thyroid imaging reporting and data system (TI-RADS) to distinguish benign from malignant thyroid nodules and the diagnostic performance of radiologists with different diagnostic experience. Medical Science Monitor: International Medical Journal of Experimental and Clinical Research. 2020;26:e918452-1.
13. Zhao Y, Healy BC, Rotstein D, Guttmann CR, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. PloS one. 2017;12(4):e0174866.
14. Chen JIZ, Hengjinda P. Early prediction of coronary artery disease (CAD) by machine learning method-a comparative study. Journal of Artificial Intelligence. 2021;3(01):17-33.

15. Hamze-Ziabari S, Bakhshpoori T. Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5′ and CART algorithms. Applied Soft Computing. 2018;68:147-61.
16. Deng H, Diao Y, Wu W, Zhang J, Ma M, Zhong X. A high-speed D-CART online fault diagnosis algorithm for rotor systems. Applied Intelligence. 2020;50(1):29-41.
17. Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, et al. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment. 2020;701:134474.
18. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees: Routledge; 2017.
19. Timofeev R. Classification and regression trees (CART) theory and applications. Humboldt University, Berlin. 2004;54.
20. Murphree DH, Arabmakki E, Ngufor C, Storlie CB, McCoy RG. Stacked classifiers for individualized prediction of glycemic control following initiation of metformin therapy in type 2 diabetes. Computers in biology and medicine. 2018;103:109-15.
21. Duan H, Deng Z, Deng F, Wang D. Assessment of groundwater potential based on multicriteria decision making model and decision tree algorithms. Mathematical Problems in Engineering. 2016;2016.
22. Ismay C, Kennedy PC. Getting Used to r, RStudio, and r Markdown. 2016.
23. Asif MA-A-R, Nishat MM, Faisal F, Shikder MF, Udoy MH, Dip RR, et al., editors. Computer aided diagnosis of thyroid disease using machine learning algorithms. 2020 11th International Conference on Electrical and Computer Engineering (ICECE); 2020: IEEE.
24. Yoon JH, Lee HS, Kim E-K, Moon HJ, Kwak JY. Malignancy risk stratification of thyroid nodules: comparison between the thyroid imaging reporting and data system and the 2014 American Thyroid Association management guidelines. Radiology. 2016;278(3):917-24.
25. Kaur K, Sonkhya N, Bapna A, Mital P. A comparative study of fine needle aspiration cytology, ultrasonography and radionuclide scan in the management of solitary thyroid nodule: A prospective analysis of fifty cases. Indian Journal of Otolaryngology and Head and Neck Surgery. 2002;54(2):96-101.
26. Shin JH, Baek JH, Chung J, Ha EJ, Kim J-h, Lee YH, et al. Ultrasonography diagnosis and imaging-based management of thyroid nodules: revised Korean Society of Thyroid Radiology consensus statement and recommendations. Korean journal of radiology. 2016;17(3):370-95.
27. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J-F, et al. Data mining in healthcare and biomedicine: a

survey of the literature. Journal of medical systems. 2012;36(4):2431-48.

28. Mansour R, Eghbal K, Amirhossein H. Comparison of artificial neural network, logistic regression and discriminant analysis efficiency in determining risk factors of type 2 diabetes. 2013.
29. Wang C, Li L, Wang L, Ping Z, Flory MT, Wang G, et al. Evaluating the risk of type 2 diabetes mellitus using artificial neural network: An effective classification approach. Diabetes research and clinical practice. 2013;100(1):111-8.
30. Parikh KS, Shah TP, Kota R, Vora R. Diagnosing common skin diseases using soft computing techniques. International Journal of Bio-Science and Bio-Technology. 2015;7(6):275-86.
31. Ioniţă I, Ioniţă L. Prediction of thyroid disease using data mining techniques. BRAIN Broad Research in Artificial Intelligence and Neuroscience. 2016;7(3):115-24.
32. Chaurasia V, Pal S, Tiwari B. Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology. 2018;12(2):119-26.
33. Talasila V, Madhubabu K, Mahadasyam MC, Atchala NJ, Kande LS. The Prediction of Diseases Using Rough Set Theory with Recurrent Neural Network in Big Data Analytics. International Journal of Intelligent Engineering & Systems. 2020;13(5).
34. Kumar HH, editor A novel approach of SVM based classification on thyroid disease stage detection. 2020 third international conference on smart systems and inventive technology (ICSSIT); 2020: IEEE.
35. Aversano L, Bernardi ML, Cimitile M, Iammarino M, Macchia PE, Nettore IC, et al. Thyroid disease treatment prediction with machine learning approaches. Procedia Computer Science. 2021;192:1031-40.