JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

Differentiated Thyroid Cancer Recurrence Prediction Using Boosting Algorithms

Mucahid Mustafa Saritas[1], Muslume Beyza Yildiz[2], Talha Alperen Cengel[1], Murat Koklu[2,*]

[1]Graduate School of Natural and Applied Sciences, Selcuk University, Konya, Türkiye
mustafa.saritas@selcuk.edu.tr, Orcid:0000-0001-5451-9092
talhacengel@gmail.com, Orcid: 0009-0005-6196-6487

[2]Department of Computer Engineering, Faculty of Technology, Selcuk University, Konya, Türkiye
beyzayldz642@gmail.com, Orcid: 0009-0002-0231-687X
mkoklu@selcuk.edu.tr, Orcid:0000-0002-2737-2360
(*Email Corresponding Author: mkoklu@selcuk.edu.tr)
Received: ………….. | Revision: ………………… | Accepted: ………………

Abstract

This study aims to compare the performance of AdaBoost, Gradient Boosting, and CatBoost algorithms in predicting the recurrence risk of Differentiated Thyroid Cancer (DTC). DTC is the most common type of thyroid cancer, and due to its recurrence risk, accurate and effective prediction models are needed. In this study, a dataset containing clinical and pathological data of patients diagnosed with DTC was used. The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1 score. The results revealed that the CatBoost algorithm achieved the highest performance, with an accuracy of 98.70% and an F1 score of 98.69% on the test data. The Gradient Boosting algorithm ranked second with an accuracy of 97.40% and an F1 score of 97.40%, while the AdaBoost algorithm showed the lowest performance, with an accuracy of 96.10% and an F1 score of 96.14%. These findings indicate that the CatBoost algorithm outperforms the other algorithms in predicting DTC recurrence risk and is a suitable candidate for use in clinical decision support systems.

Keywords : Differentiated Thyroid Cancer, AdaBoost, Gradient Boosting, CatBoost, Machine Learning

1. INTRODUCTION

Differentiated thyroid cancer (DTC) is the most common type of thyroid cancer, arising as a result of the malignant transformation of thyroid cells [1, 2]. The thyroid gland is an endocrine organ located in the anterior part of the neck, playing a crucial role in hormone production. The uncontrolled growth and proliferation of thyroid cells lead to malignancy (cancerous transformation), resulting in thyroid cancer. DTC includes papillary and follicular thyroid cancers and is generally known for having a better prognosis compared to other types of thyroid cancer [3, 4].

Currently, various clinical and pathological factors are used to predict the recurrence risk of DTC. These factors include the patient's age, tumor size, lymph node involvement, and histological characteristics [3, 5]. However, traditional methods suffer from significant limitations, such as subjective assessment, challenges in data integration and processing, inability to analyze complex relationships, high time and resource consumption, and low predictive accuracy [6, 7]. These shortcomings highlight the necessity for more precise, objective, and rapid methods for identifying DTC recurrence risk. At this point, artificial intelligence (AI) emerges as a promising solution. AI can efficiently process large and complex datasets, providing objective and reliable predictions for clinical decision support systems [8-10].

Machine learning and AI techniques have made significant advancements in the medical field, introducing new approaches for disease diagnosis and treatment[11, 12]. Unlike traditional methods, AI-based systems are less subjective and possess the capability to analyze large datasets, accurately identifying complex relationships and predictions [13, 14]. These systems integrate data from various sources, such as electronic health records, biopsy results, and imaging data, to provide more comprehensive and precise predictions [15]. In this study, we aim to classify DTC recurrence risk using AdaBoost, CatBoost, and Gradient Boosting (GBoost) algorithms. These algorithms are boosting methods that combine multiple weak classifiers to build strong classifiers and are known for their high accuracy rates [16-18]. Boosting algorithms enhance classification accuracy by minimizing error rates in the dataset, thereby enabling the early and accurate detection of patients at risk of recurrence.

The primary objective of this study is to compare the performance of AdaBoost, CatBoost, and GBoost algorithms to determine which model predicts DTC recurrence risk with the highest accuracy. In this context, the developed models will be evaluated using performance metrics such as accuracy, precision, recall, and F1 score. The findings will demonstrate the effectiveness of machine learning-based approaches in predicting DTC recurrence risk, supporting their applicability in clinical settings.

The results of this study highlight the importance of utilizing machine learning techniques for predicting DTC recurrence risk and serve as a benchmark for future research in this field. Future studies should aim to improve the overall performance of these models by utilizing larger and more diverse datasets, as well as integrating different AI techniques for more comprehensive analyses. The classification results have been compared to identify the algorithm with the highest predictive accuracy.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

2. RESEARCH METHODOLOGY

This section provides a detailed explanation of the dataset used in the study, the machine learning algorithms applied, and the performance metrics employed to evaluate the models. The overall methodology followed in this study is illustrated in Figure 1.

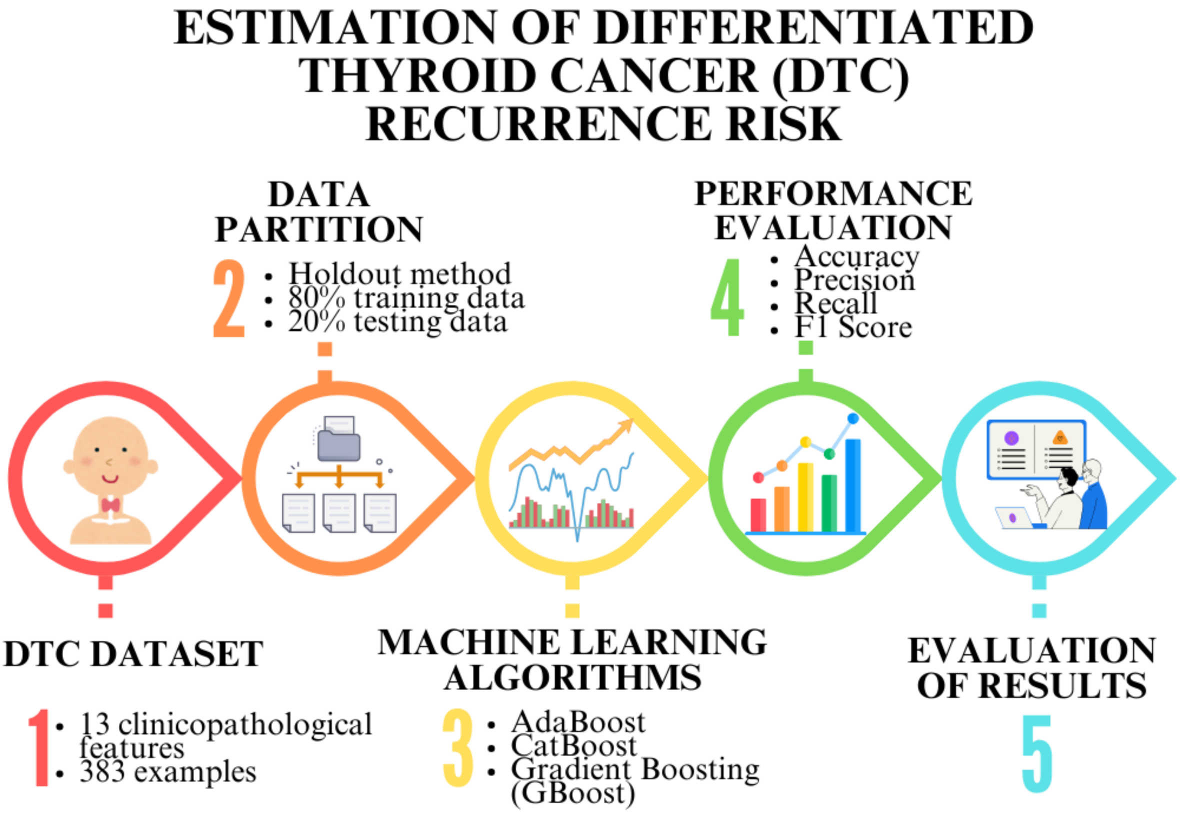


Figure 1. General Flow Diagram of Differentiated Thyroid Cancer Recurrence Prediction Using Boosting Algorithms

2.1. Differentiated Thyroid Cancer Recurrence (DTC) Dataset

In this study, the Differentiated Thyroid Cancer Recurrence (DTC) dataset was utilized. This dataset includes 13 clinicopathological features of patients diagnosed with DTC. The dataset consists of a total of 383 samples, with each sample classified based on DTC recurrence risk [19]. The features of the dataset are presented in Table 1. The age distribution of individuals in the dataset is shown in Figure 2, while gender distribution, smoking status, and treatment response status are provided in Figure 3. Additionally, Figure 4 presents a violin plot illustrating the age distribution across three different risk categories (low, medium, and high risk).

Table 1. Differentiated Thyroid Cancer Recurrence (DTC) dataset

Attribute	Values
Age	15-82
Gender	Male, Female
Currently smoking	Yes, No
History of smoking	Yes, No
History of radiotion exposure	Yes, No
Thyroid function	Euthyroid, Subclinical hypothyroidism, Clinical hypothyroidism, Subclinical hyperthyroidism, Clinical hyperthyroidism
Pathology	Papillary, Micropapillary,

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

	Follicular, Hurthel cell
Focality	Uni-focal, Multi-focal
ATA Risk	Low, Intermediate, High
Adenopathy	None, Right-sided, Left-sided, Bilateral, Posterior, Extensive
Goiter	None, Single nodular right, Single nodular left, Multi nodular, Difuse
Tumor	T1a, T1b, T2a, T2b, T3a, T3b, T4a, T4b
Node	N0, N1a, N1b
Metastasis	M0, M1
Stage	1, 2, 3, 4a, 4b
Treatment response	Excellent, Indeterminate, Biochemical incomplete, Structural incomplete
Recurrence	Yes, No

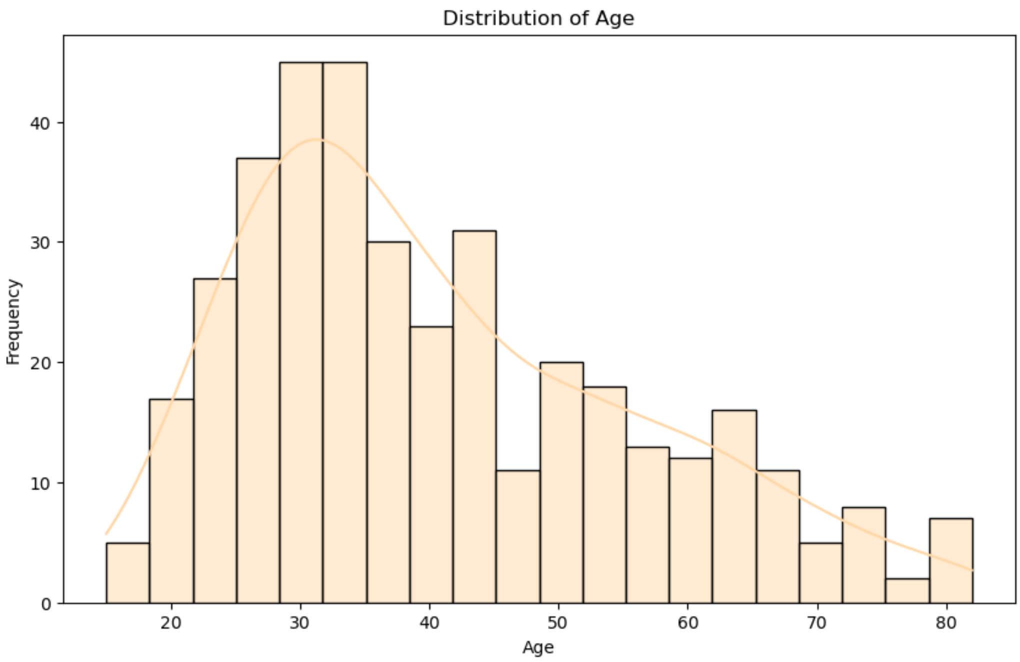
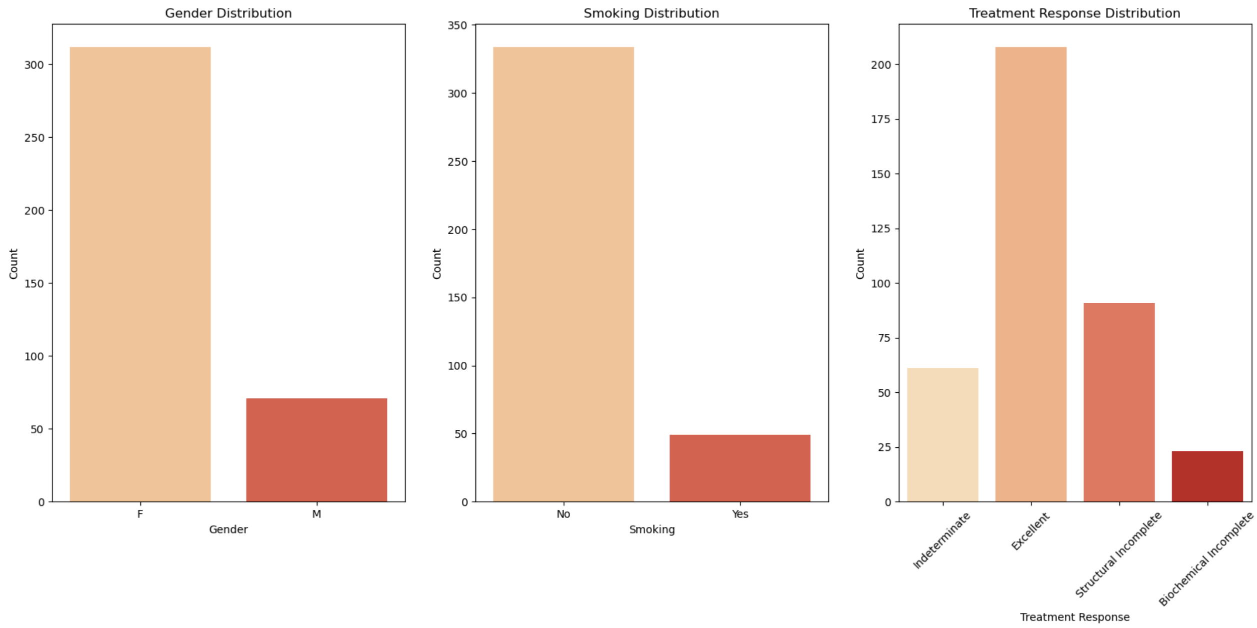


Figure 2. Age distributions of individuals in the data set

The age distribution histogram peaks within the 30-35 age range. The histogram exhibits a long right tail, indicating that the number of older individuals aged 60-80 is relatively low. The density curve overlaid on the histogram shows a peak around 30 years of age, gradually declining until 80 years. This distribution suggests that younger and middle-aged individuals constitute the majority, while older individuals are less represented in the dataset.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490



(a) **(b)** **(c)**

Figure 3. Gender Distribution (a), Smoking Status (b), and Treatment Response (c)

The majority of participants in the dataset are female, and a significant proportion do not smoke. Regarding treatment response, most participants have shown an excellent response to treatment. Additionally, the number of individuals with an indeterminate response is moderate, while those with a structurally incomplete or biochemically incomplete response are relatively fewer. These distributions provide a general summary of the demographic and health characteristics within the dataset.

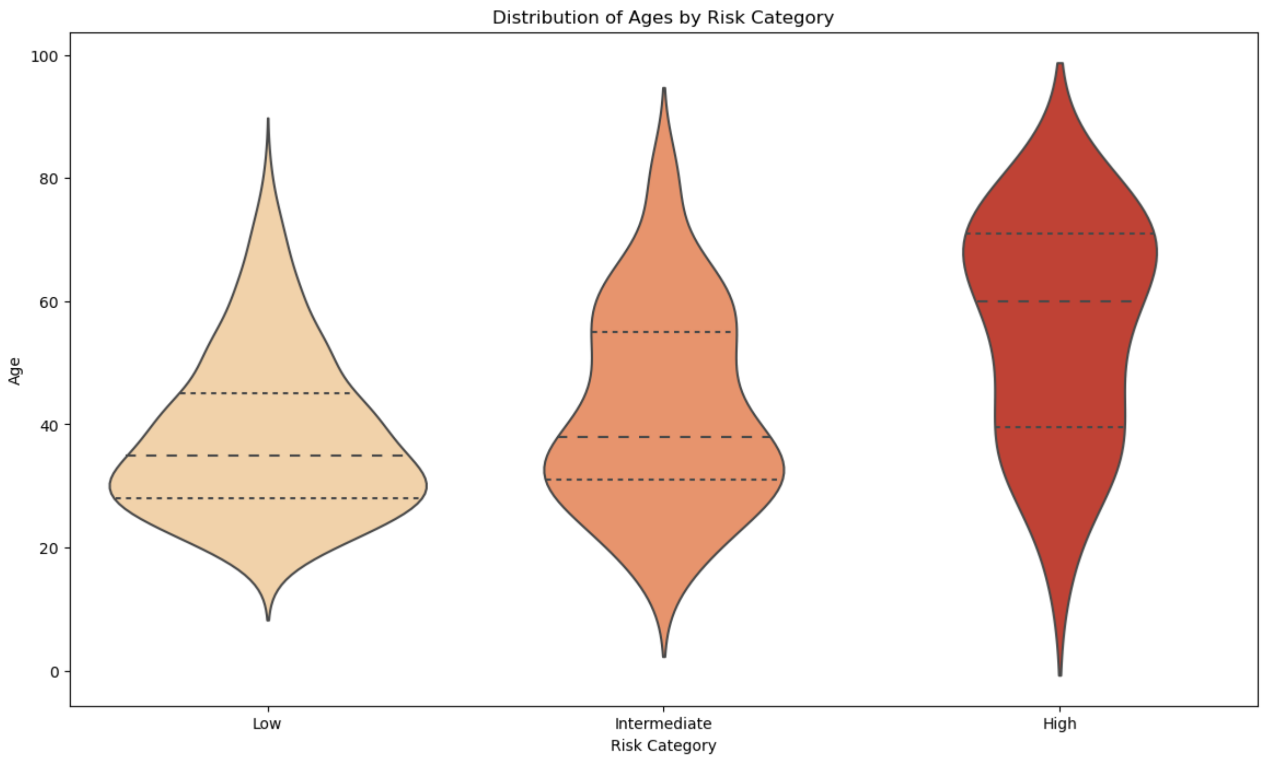


Figure 4. Risk Violin Plot of Age Distribution by Risk Category

Examining Figure 4, the dataset generally exhibits a wider age distribution in the low-risk category, while the age range is narrower in the intermediate and high-risk categories. Notably, individuals aged 30-50 years are more prevalent in the moderate and high-risk groups.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

2.2. Holdout Method

The Holdout Method is a simple and widely used validation technique in machine learning and statistical modeling, where the dataset is split into training and test subsets to evaluate model performance. In this approach, the dataset is randomly divided into two parts. The training set, which is used for the model's learning process[20]. The test set, which consists of previously unseen data used to measure the model's generalization performance. The splitting ratio is typically 70-30% or 80-20%. The Holdout Method is preferred for its simplicity and efficiency, as it allows for quick implementation and requires training and testing only once, making it computationally time-efficient. Due to these advantages, it is commonly applied in model validation and evaluation [21]. In this study, the Holdout Method was employed, with the dataset split into 80% training and 20% test sets, following standard practice in the literature.

2.3. Confusion Matrix

The Confusion Matrix is a tabular representation used to evaluate the performance of a classification model. It provides insight into the types of errors the model makes and the distribution of these errors [22-24]. By analyzing the Confusion Matrix, the strengths and weaknesses of the model can be identified [25, 26]. An example Confusion Matrix is presented in Figure 5.

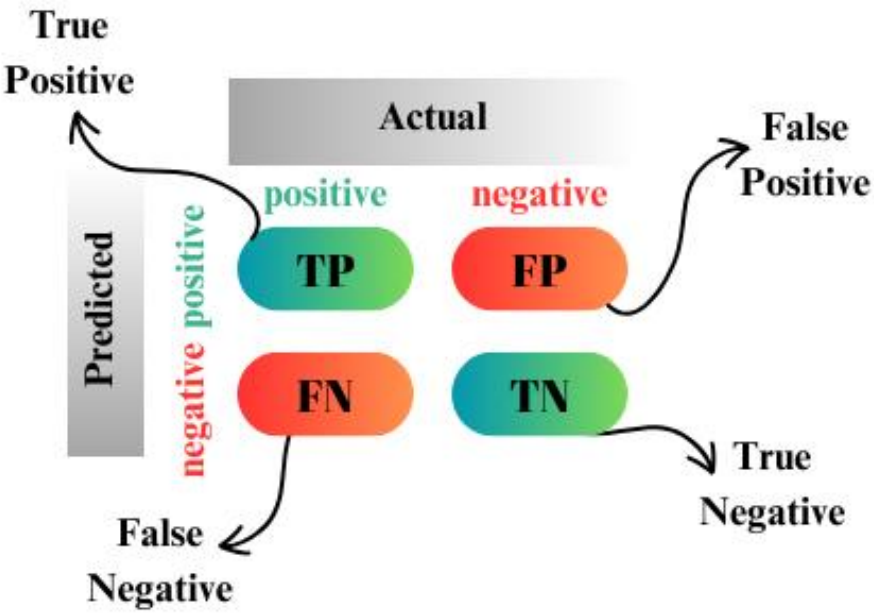


Figure 5. Confusion matrix

The Confusion Matrix presented in Figure 5 is a widely used tool for evaluating the performance of classification models.

True Positives (TP): Cases correctly classified as positive by the model.

True Negatives (TN): Cases correctly classified as negative by the model.

False Positives (FP): Cases incorrectly classified as positive when they are actually negative.

False Negatives (FN): Cases incorrectly classified as negative when they are actually positive.

The Confusion Matrix not only helps assess a model's overall accuracy but also provides valuable insight into the types of errors it makes.

2.4. Performance Measures

The performance of the models was evaluated using metrics such as accuracy, precision, recall, and F1 score[27]. These metrics were employed to assess how accurately the models predicted the risk of DTC recurrence. Through these components, various performance metrics can be calculated as follows:

Accuracy: The proportion of correctly classified samples out of the total samples. It is used to evaluate the general performance of the model [28]. It is calculated as: formula 1.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

Precision: The proportion of positively classified samples that are actually positive. It is used to reduce the impact of false positives [29] and is calculated as: formula 2.

Recall: The proportion of actual positives that are correctly classified as positive by the model. It is used to reduce the impact of false negatives [30] and is calculated as: formula 3.

F1-Score: The harmonic mean of precision and recall. It provides a balanced performance measure [31] and is calculated as: formula 4.

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}} \quad (1)$$

$$Precision = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad (2)$$

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad (3)$$

$$\text{F1} - \text{Score} = 2 * \frac{\frac{\text{tp}}{\text{tp} + \text{fp}} * \frac{\text{tp}}{\text{tp} + \text{tn}}}{\frac{\text{tp}}{\text{tp} + \text{fp}} + \frac{\text{tp}}{\text{tp} + \text{tn}}} \quad (4)$$

2.5. Heatmap

A heatmap is a type of graph commonly used in data visualization and is particularly useful in analyzing complex datasets. This graph allows the visualization of data using a color scale through cells arranged in a matrix format. Each cell is represented by a color corresponding to a specific value; typically, cold colors are used for low values, while warm colors are used for high values. This method enables the easy identification of patterns, correlations, and densities within the data. Beyond data visualization, heatmaps are widely used in academic research and applications as a powerful tool supporting data-driven decision-making processes [32].

2.6. Machine Learning Algorithms

AdaBoost (Adaptive Boosting): AdaBoost, a machine learning algorithm and ensemble method, is a boosting algorithm where weak learners are trained sequentially with weighted training data to build a strong classifier [16]. The primary goal of AdaBoost is to increase the weight of misclassified examples, thereby creating more accurate classifiers in each cycle, ultimately leading to a robust classifier [33].

CatBoost: CatBoost, developed by Yandex, is a high-performance gradient boosting algorithm that works effectively with categorical data. The name "CatBoost" is an abbreviation of "Categorical Boosting," highlighting its superior performance in handling categorical data. CatBoost provides high accuracy and speed, especially for complex datasets and large-scale problems. It stands out from other gradient boosting algorithms due to its overfitting control, categorical data handling, and suitability for sequential data [34].

GBoost: GBoost is an abbreviation for the gradient boosting algorithm. This widely used machine learning method works by sequentially training a series of weak learners (usually decision trees). In each step, it attempts to correct the errors made in the previous steps, thereby ultimately forming a strong model. GBoost provides high accuracy, particularly in large and complex datasets [35].

3. RESULTS AND DISCUSSION

In this study, classification models were developed using AdaBoost, CatBoost, and Gradient Boosting (GBoost) algorithms to predict the recurrence risk of Differentiated Thyroid Cancer (DTC), and the performance of these models was compared. The performance evaluation was conducted using metrics such as accuracy, precision, recall, and F1 score. The results were evaluated separately for training and test datasets.

As shown in Table 3, according to the performance evaluations on the test data, the CatBoost algorithm achieved the highest accuracy (0.9870) and the best F1 score (0.9869). CatBoost provided the most accurate results by minimizing the number of false negatives and false positives on the test data.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

The Gradient Boosting algorithm also exhibited high accuracy (0.9740) and F1 score (0.9740), showing performance close to that of CatBoost. Although the AdaBoost algorithm showed slightly lower performance compared to the other two, it still achieved a high accuracy (0.9610) and F1 score (0.9614).

As seen in Table 2, the evaluation on the training data showed that the Gradient Boosting and CatBoost algorithms performed excellently (1.0000 accuracy and F1 score). The AdaBoost algorithm, on the other hand, achieved successful results with an accuracy and F1 score of 0.9869 on the training data.

Overall, this study demonstrates that the CatBoost algorithm outperforms the other two algorithms in predicting the recurrence risk of DTC.

Table 2. Results of training dataset

Algorithm	***Confusion Matrix***	***Accuracy***	***Precision***	***Recall***	***F1-score***
AdaBoost	216, 1 / 3, 86	0.9869	0.9869	0.9869	0.9869
GBoost	217, 0 / 0, 89	1.0000	1.0000	1.0000	1.0000
CatBoost	217, 0 / 0, 89	1.0000	1.0000	1.0000	1.0000

Table 3. Results of the test dataset

Algorithm	***Confusion Matrix***	***Accuracy***	***Precision***	***Recall***	***F1-score***
AdaBoost	56, 2 / 1, 18	0.9610	0.9621	0.9610	0.9614
GBoost	57, 1 / 1, 18	0.9740	0.9740	0.9740	0.9740
CatBoost	58, 0 / 1, 18	0.9870	0.9872	0.9870	0.9869

In classification with AdaBoost, the "feature importance" values reflecting the contribution of each feature to the model's prediction performance are shown in Figure 6. Upon examining the graph, it is evident that age, tumor size, and the presence of metastasis, denoted by T and response excellent, are significant features.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

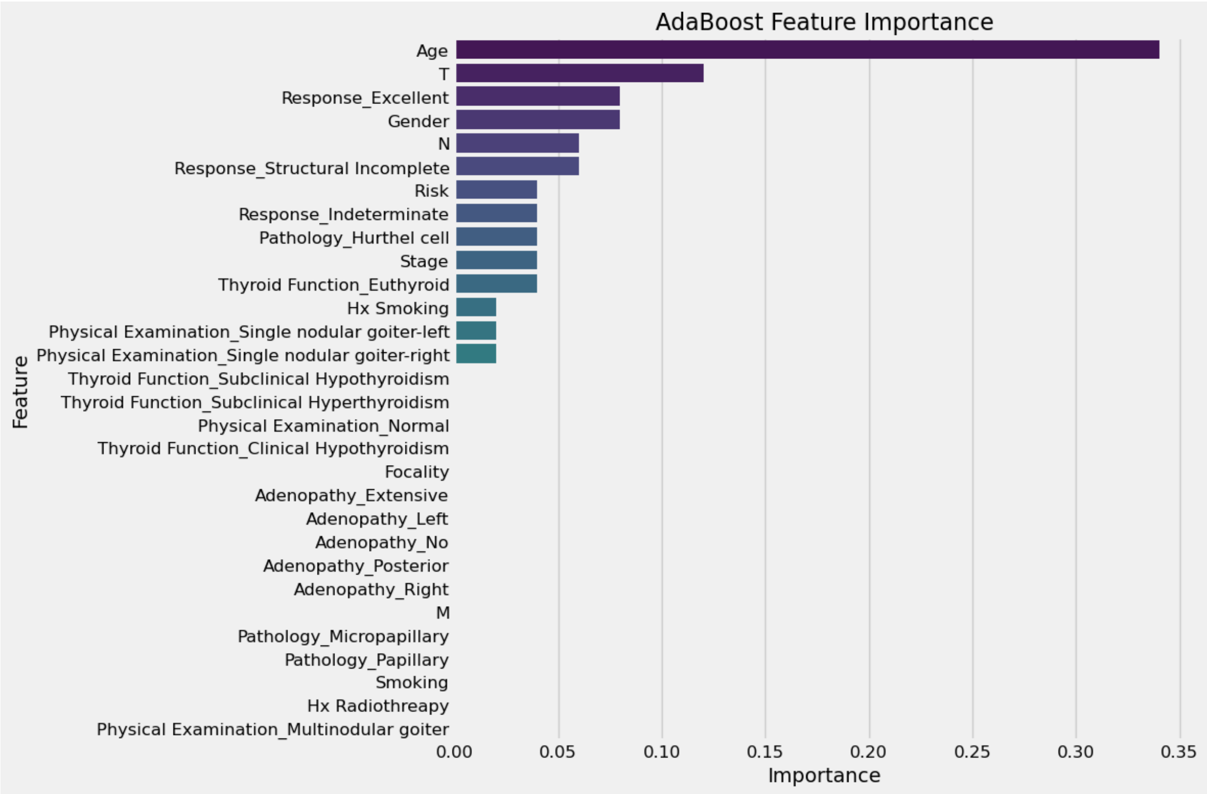


Figure 6. AdaBoost feature importance

In classification with GBoost, the "feature importance" values reflecting the contribution of each feature to the model's prediction performance are shown in Figure 7. Upon examining the graph, response structural incomplete, risk, age, tumor size, and the presence of metastasis, denoted by T, are identified as important features.

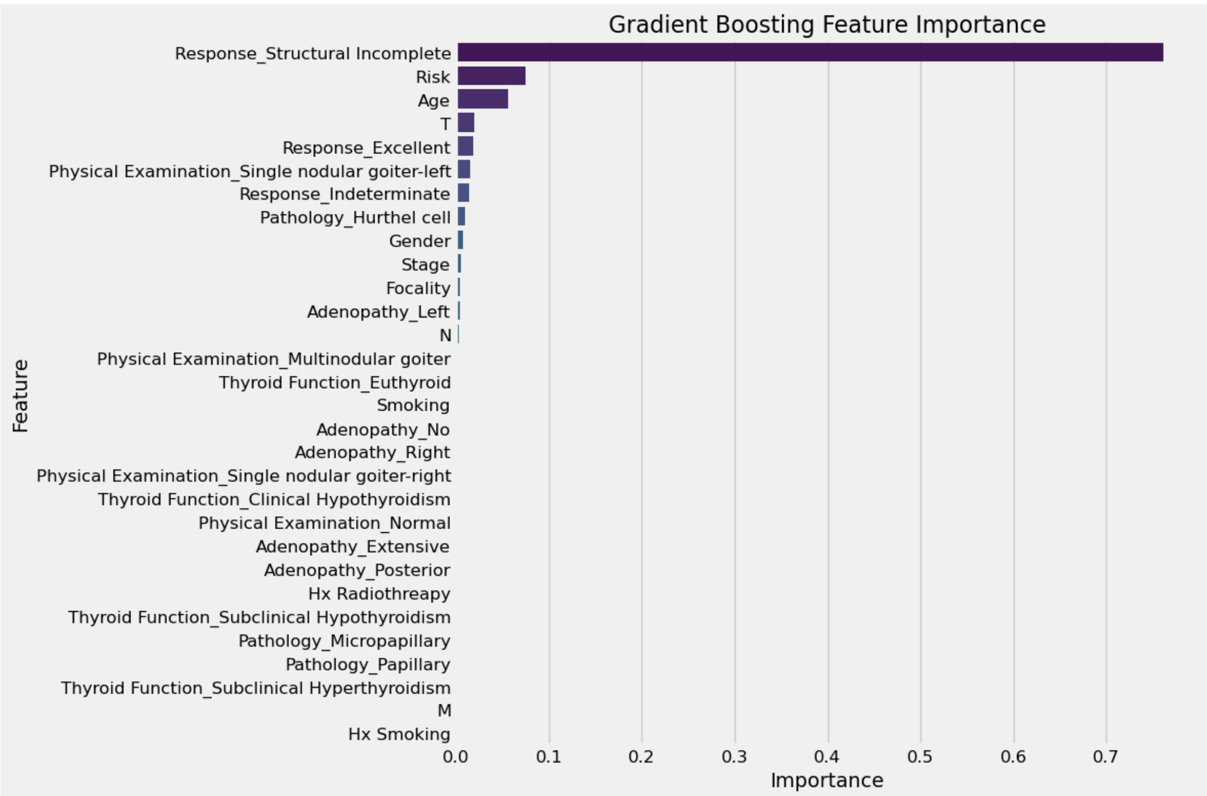


Figure 7. Gradient Boosting feature importance

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

In classification with CatBoost, the "feature importance" values reflecting the contribution of each feature to the model's prediction performance are shown in Figure 8. Upon examining the graph, response structural incomplete, response excellent, risk, age, tumor size, and the presence of metastasis, denoted by T, are identified as significant features.

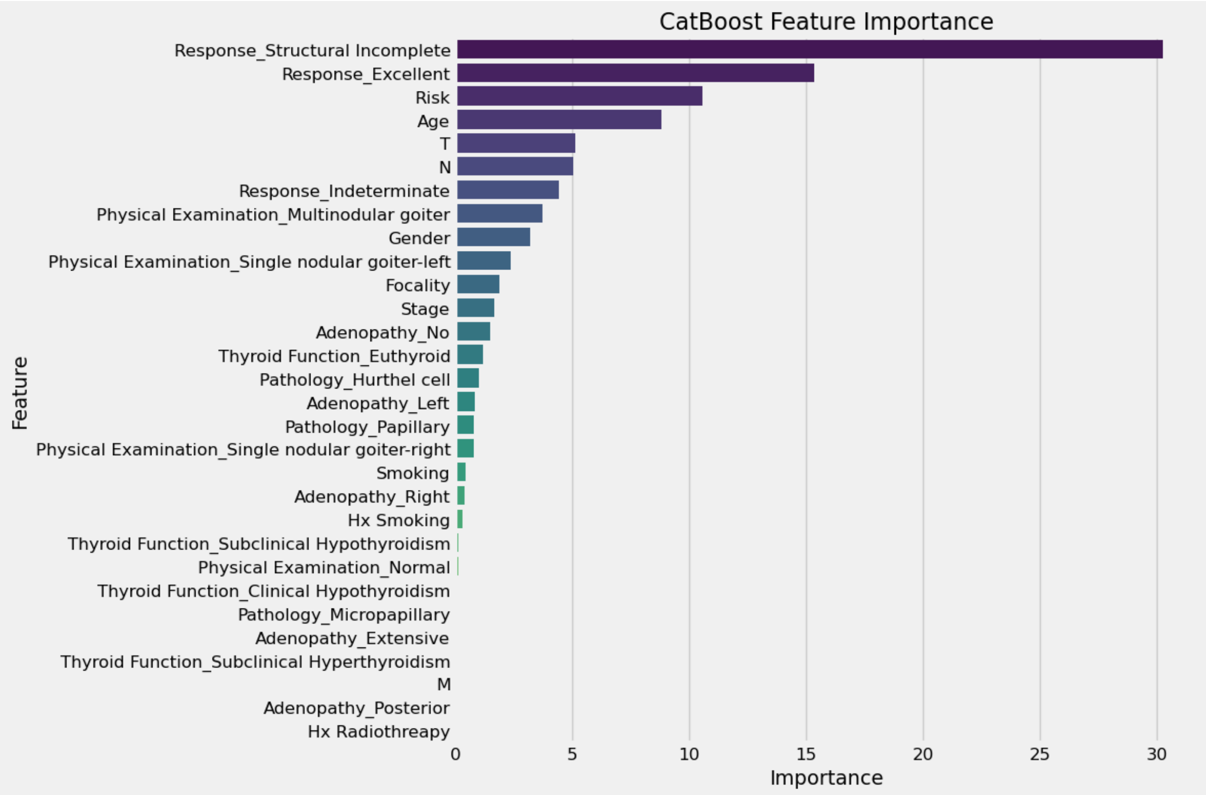


Figure 8. CatBoost feature importance

A heatmap depicting the correlations between clinical and biological parameters in the DTC dataset has been obtained and is shown in Figure 9. When examining the relationships between these parameters in Figure 9, the correlation coefficient between Stage and M (Metastasis) is 0.68, indicating a strong positive relationship between the staging of thyroid cancer and the presence of metastasis. This suggests that metastasis is more frequently observed in advanced stages of cancer, and these two parameters are closely related. The correlation coefficient between T and Stage is 0.54, indicating that the size or depth of the tumor is directly related to the cancer staging. This highlights that tumor size is a significant factor in determining cancer stage. The correlation coefficient between Recurred and Response is 0.71, suggesting that patients who respond positively to treatment have a lower likelihood of disease recurrence. This underscores the critical importance of effective treatment protocols in reducing recurrence risk. Hx Radiotherapy shows low correlations, especially with thyroid function, suggesting that previous radiotherapy treatments may have long-term effects on thyroid function.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

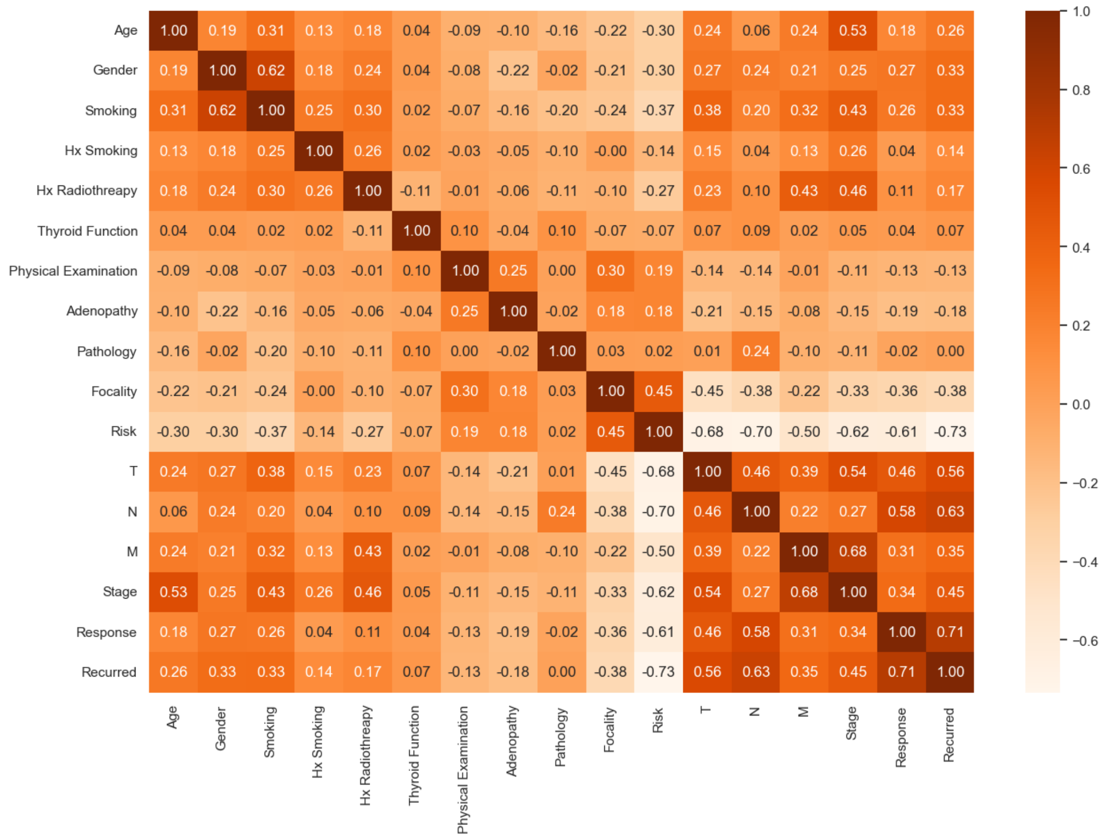


Figure 9. Correlation heatmap of clinical and pathological variables used in the prediction of Differentiated Thyroid Cancer recurrence.

3.1. Discussion

In this study, the performances of three different boosting algorithms (AdaBoost, Gradient Boosting, and CatBoost) were compared to predict the recurrence risk of Differentiated Thyroid Cancer (DTC). The results obtained show that the CatBoost algorithm outperforms the other two algorithms. This finding can be explained by CatBoost's superior generalization ability, particularly in high-dimensional and complex datasets.

The CatBoost algorithm achieved the highest performance with 98.70% accuracy and 98.69% F1 score on the test data. These results suggest that CatBoost can be effectively used in clinical decision support systems. The Gradient Boosting algorithm ranks second with 97.40% accuracy and 97.40% F1 score, while the AdaBoost algorithm, with 96.10% accuracy and 96.14% F1 score, showed the lowest performance. These findings indicate that CatBoost is a more reliable and effective model in predicting the recurrence risk of DTC.

Table 4. Comparison of this study with studies in literature

Reference	Dataset Values	Methods	Performance
Nguyen, et al., 2020 [36]	450 images	ResNet50	Accuracy: 92.05%
Ye, et al., 2020 [37]	1810 images	VGG19	Accuracy: 85.32%
Zhou, et al., 2020 [38]	1750 images	CNN	AUC: 0.97
Tsantis et al., 2009 [39]	85 images	SVM	AUC: 0.96
Guo et al., 2022 [40]	2423 patient	Random forest	AUC: 0.874
Borzooei et al., 2023 [41]	383 patient	KNN SVM DT RF ANN	Accuracy: 93% 96% 96% 96% 96%
Our Study, 2025	383 patient	AdaBoost GradientBoost CatBoost	Accuracy: 96.1%

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

	97.4%
	98.7%

The findings of this study are consistent with those of other studies in the literature. For instance, in similar disease prediction studies, the highest accuracy reported was 96% [41]. In our study, however, CatBoost demonstrated the highest performance with an accuracy of 98.7%. This result can be attributed to CatBoost's optimization techniques and its proficiency in handling categorical data.

However, there are some limitations to our study. First, the size and diversity of our dataset may affect the generalization ability of the model. Studies conducted with larger and more diverse datasets obtained from different populations could improve the model's performance and reliability.

In conclusion, the CatBoost algorithm was found to outperform other boosting algorithms in predicting the recurrence risk of DTC, and it was shown to be a suitable candidate for use in clinical decision support systems. These findings could enable more accurate and timely interventions during the treatment and follow-up processes for DTC patients.

In this study, the performances of AdaBoost, Gradient Boosting, and CatBoost algorithms were compared to predict the recurrence risk of Differentiated Thyroid Cancer (DTC). The results indicate that all three algorithms achieved high accuracy rates; however, the CatBoost algorithm demonstrated superior performance compared to the other two algorithms.

CatBoost Algorithm: CatBoost achieved the highest performance with an accuracy of 98.70% and an F1 score of 98.69% on the test data. This algorithm provides an advantage due to its ability to effectively handle categorical data and minimize the risk of overfitting.

Gradient Boosting Algorithm: Gradient Boosting exhibited performance close to CatBoost with an accuracy of 97.40% and an F1 score of 97.40%. This algorithm improves model performance by sequentially correcting errors.

AdaBoost Algorithm: AdaBoost showed slightly lower performance compared to the other two algorithms with an accuracy of 96.10% and an F1 score of 96.14%. Nevertheless, it still achieved high accuracy and was effective in predicting the recurrence risk of DTC.

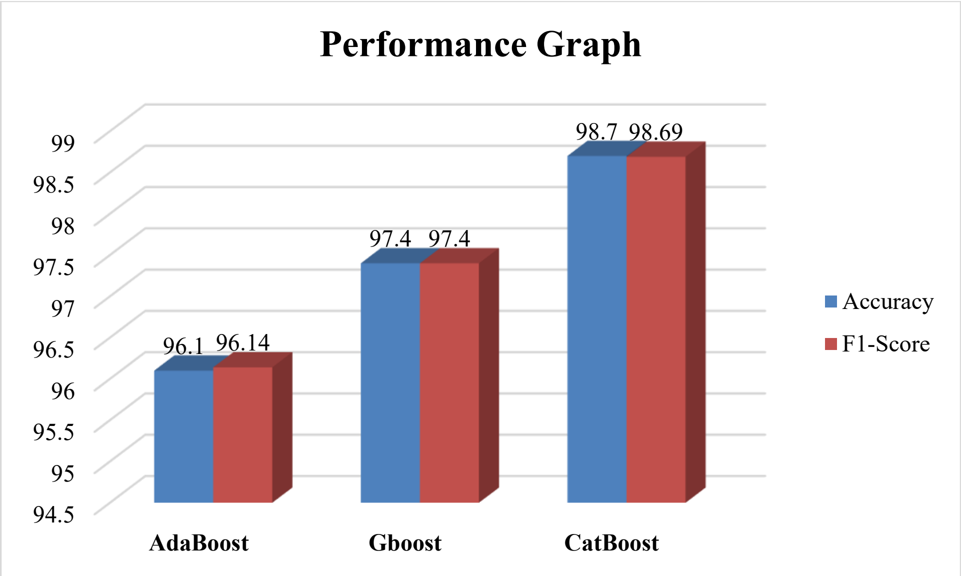


Figure 10. Performance graph of CatBoost, GBoost and AdaBoost algorithms

According to the results obtained in this study, as shown in Figure 10, the CatBoost algorithm provides the highest accuracy in predicting the recurrence risk of DTC. The ability of CatBoost to process categorical data and automatically optimize model parameters contributed to this high performance. The Gradient Boosting algorithm also stands out due to its high accuracy and generalization capability. Although AdaBoost demonstrated relatively lower performance, it can still be considered an effective classification algorithm.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

In this study, the performances of the AdaBoost, Gradient Boosting, and CatBoost algorithms were compared in predicting DTC recurrence risk. As shown in Figure 10, the CatBoost algorithm achieved the highest accuracy and F1 score on the test data, outperforming the other algorithms. The Gradient Boosting algorithm also performed well and outperformed AdaBoost in comparison.

The findings of this study demonstrate that artificial intelligence and machine learning algorithms have effective classification capabilities on medical data and can be a significant tool in predicting the recurrence risk of DTC. Particularly, the ability of CatBoost to handle categorical data and its overall performance suggest that it is a suitable candidate for use in clinical decision support systems.

4. CONCLUSION

This study demonstrated the effectiveness of machine learning algorithms, particularly boosting-based classifiers, in predicting the recurrence risk of Differentiated Thyroid Cancer (DTC) using clinicopathological features. Through a comparative evaluation of three widely used ensemble learning methods—CatBoost, Gradient Boosting (GBoost), and AdaBoost—it was observed that all algorithms achieved high predictive performance, with CatBoost outperforming the others. Specifically, CatBoost achieved an accuracy of 98.70% and an F1 score of 98.69%, underscoring its robustness in handling categorical variables and minimizing overfitting. Gradient Boosting also delivered a strong performance, with a 97.40% accuracy rate, while AdaBoost followed closely with 96.10% accuracy. These results suggest that boosting-based algorithms, when properly tuned and applied to structured clinical data, can serve as reliable tools for risk stratification in oncology.

The correlation heatmap further supported the model findings by identifying strong associations between DTC recurrence and specific clinical indicators, including risk score, treatment response, and tumor staging parameters (T, N, M). These results align with existing clinical knowledge and enhance the interpretability of the machine learning models, which is critical for real-world implementation in healthcare settings. In particular, the strong correlation between recurrence and treatment response ($r = 0.71$) highlights the importance of longitudinal patient monitoring in predictive modeling.

From a broader perspective, the findings of this study underscore the potential of machine learning-based decision support systems in enhancing clinical workflows. The integration of high-performing models such as CatBoost into clinical practice could aid clinicians in early identification of high-risk patients, facilitating timely intervention and personalized treatment planning. Moreover, the transparency and reproducibility of the modeling process, coupled with performance metrics based on standard evaluation criteria (accuracy, precision, recall, F1-score), further validate the applicability of these models in medical domains.

In conclusion, boosting-based machine learning algorithms offer promising performance in the prediction of DTC recurrence. Future work should focus on validating these findings across larger and more diverse patient populations, incorporating additional clinical and molecular features, and exploring hybrid modeling approaches to further improve prediction accuracy and clinical relevance.

5. DATA AVAILABILITY

Contacting the corresponding author Shiva Borzooei or accessing the study's dataset through the UCI Data website here; https://archive.ics.uci.edu/dataset/915/differentiated+thyroid+cancer+recurrence

6. ACKNOWLEDGEMENT

We extend our genuine gratitude to the Scientific Research Coordinator at Selcuk University for their significant aid and backing during this research.

REFERENCES

[1] A. Shokoohi *et al.*, "Treatment for recurrent differentiated thyroid cancer: a Canadian population based experience," *Cureus,* vol. 12, no. 2, 2020.

[2] M. E. Cabanillas, D. G. McFadden, and C. Durante, "Thyroid cancer," *The Lancet,* vol. 388, no. 10061, pp. 2783-2795, 2016.

[3] W. R. Burns and M. A. Zeiger, "Differentiated thyroid cancer," in *Seminars in oncology*, 2010, vol. 37, no. 6: Elsevier, pp. 557-566, doi: 10.1053/j.seminoncol.2010.10.008.

[4] M. Schlumberger and S. Leboulleux, "Current practice in patients with differentiated thyroid cancer," *Nature Reviews Endocrinology,* vol. 17, no. 3, pp. 176-188, 2021, doi: 10.1038/s41574-020-00448-z.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

[5] B. Schmidbauer, K. Menhart, D. Hellwig, and J. Grosse, "Differentiated thyroid cancer—treatment: state of the art," *International journal of molecular sciences,* vol. 18, no. 6, p. 1292, 2017, doi: 10.3390/ijms18061292.

[6] Y. Habchi *et al.*, "Ai in thyroid cancer diagnosis: Techniques, trends, and future directions," *Systems,* vol. 11, no. 10, p. 519, 2023, doi: 10.3390/systems11100519.

[7] H. K. ORNEK, B. YILMAZ, E. YASIN, and M. KOKLU, "Deep learning-based classification of skin lesion dermoscopic images for melanoma diagnosis," *Intelligent Methods In Engineering Sciences,* vol. 3, no. 2, pp. 70-81, 2024.

[8] Y. S. Taspinar, I. Cinar, and M. Koklu, "Classification by a stacking model using CNN features for COVID-19 infection diagnosis," *Journal of X-ray science and technology,* vol. 30, no. 1, pp. 73-88, 2022, doi: 10.3233/XST-211031.

[9] M. Koklu and K. Tutuncu, "Classification of chronic kidney disease with most known data mining methods," *Int. J. Adv. Sci. Eng. Technol,* vol. 5, no. 2, pp. 14-18, 2017.

[10] M. M. Saritas, M. Koklu, and I. A. Ozkan, "A Novel Embedded System on Cold Box Design for The Cold Chain," *International Journal of Applied Mathematics Electronics and Computers,* vol. 5, no. 4, pp. 67-70, 2017, doi: 10.18100/ijamec.2017436077.

[11] K. Erdem, E. Yasin, M. B. Yıldız, and M. Koklu, "Classification of Heart Diseases with Ensemble Learning Algorithms," *Sinop Üniversitesi Fen Bilimleri Dergisi,* vol. 9, no. 2, pp. 369-387, 2024.

[12] K. ERDEM, M. B. YILDIZ, E. T. YASIN, and M. KÖKLÜ, "A Detailed analysis of detecting heart diseases using artificial intelligence methods," *Intelligent methods in engineering sciences,* vol. 2, no. 4, pp. 115-124, 2023.

[13] N. Schwalbe and B. Wahl, "Artificial intelligence and the future of global health," *The Lancet,* vol. 395, no. 10236, pp. 1579-1586, 2020.

[14] I. Saritas, M. Koklu, and K. Tutuncu, "Performance of Classification Techniques on Parkinson's Disease," *Int. J. Adv. Sci. Eng. Technol,* vol. 5, no. 2, pp. 9-13, 2017.

[15] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," *Journal of global health,* vol. 8, no. 2, 2018, doi: 10.7189/jogh.08.020303.

[16] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review,* vol. 54, pp. 1937-1967, 2021, doi: 10.1007/s10462-020-09896-5.

[17] C. Ying, M. Qi-Guang, L. Jia-Chen, and G. Lin, "Advance and prospects of AdaBoost algorithm," *Acta Automatica Sinica,* vol. 39, no. 6, pp. 745-758, 2013, doi: 10.1016/S1874-1029(13)60052-X.

[18] C. Zhang, X. Chen, S. Wang, J. Hu, C. Wang, and X. Liu, "Using CatBoost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018," *Psychiatry Research,* vol. 306, p. 114261, 2021, doi: 10.1016/j.psychres.2021.114261.

[19] S. Borzooei and A. Tarokhian. *Differentiated Thyroid Cancer Recurrence*, UC Irvine Machine Learning Repository, doi: 10.24432/C5632J.

[20] I. Cinar, Y. S. Taspinar, and M. Koklu, "Development of early stage diabetes prediction model based on stacking approach," *Tehnički glasnik,* vol. 17, no. 2, pp. 153-159, 2023.

[21] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *2016 IEEE 6th International conference on advanced computing (IACC)*, 2016: IEEE, pp. 78-83, doi: 10.1109/IACC.2016.25.

[22] Y. S. Taspinar, M. Koklu, and M. Altin, "Identification of the english accent spoken in different countries by the k-nearest neighbor method," *International Journal of Intelligent Systems and Applications in Engineering,* 2020. [Online]. Available: https://hdl.handle.net/20.500.13091/1351.

[23] Y. S. Taspinar, M. M. Saritas, İ. Cinar, and M. Koklu, "Gender determination using voice data," *International Journal of Applied Mathematics Electronics and Computers,* vol. 8, no. 4, pp. 232-235, 2020, doi: 10.18100/ijamec.809476.

[24] M. M. Saritas, Y. S. Taspinar, I. Cinar, and M. Koklu, "Railway Track Fault Detection with ResNet Deep Learning Models," in *2023 International Conference on Intelligent Systems and New Applications (ICISNA'23)*, 2023.

[25] M. Koklu and K. Tutuncu, "Tree based classification methods for occupancy detection," in *IOP Conference Series: Materials Science and Engineering*, 2019, vol. 675, no. 1: IOP Publishing, p. 012032, doi: 10.1088/1757-899X/675/1/012032.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 663-676
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.490

[26] M. B. Yildiz, E. T. Yasin, and M. Koklu, "Fisheye freshness detection using common deep learning algorithms and machine learning methods with a developed mobile application," *European Food Research and Technology,* vol. 250, no. 7, pp. 1919-1932, 2024.

[27] R. KURSUN, M. M. SARITAS, and M. KOKLU, "Machine Learning-Based Kidney Disease Detection Using Deep Features from SqueezeNet," 2025.

[28] I. Cinar, Y. S. Taspinar, R. Kursun, and M. Koklu, "Identification of corneal ulcers with pre-trained AlexNet based on transfer learning," in *2022 11th Mediterranean conference on embedded computing (MECO)*, 2022: IEEE, pp. 1-4, doi: 10.1109/MECO55406.2022.9797218.

[29] R. Kursun, I. Cinar, Y. S. Taspinar, and M. Koklu, "Flower recognition system with optimized features for deep features," in *2022 11th Mediterranean conference on embedded computing (MECO)*, 2022: IEEE, pp. 1-4, doi: 10.1109/MECO55406.2022.9797103.

[30] H. Isik *et al.*, "Maize seeds forecasting with hybrid directional and bi-directional long short-term memory models," *Food Science & Nutrition,* vol. 12, no. 2, pp. 786-803, 2024, doi: 10.1002/fsn3.3783.

[31] R. Kursun, E. T. Yasin, and M. Koklu, "Machine learning-based classification of infected date palm leaves caused by dubas insects: a comparative analysis of feature extraction methods and classification algorithms," in *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2023: IEEE, pp. 1-6, doi: 10.1109/ASYU58738.2023.10296641.

[32] R. Schönhof, A. Werner, J. Elstner, B. Zopcsak, R. Awad, and M. Huber, "Feature visualization within an automated design assessment leveraging explainable artificial intelligence methods," *Procedia CIRP,* vol. 100, pp. 331-336, 2021, doi: 10.1016/j.procir.2021.05.075.

[33] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and its Interface,* vol. 2, no. 3, pp. 349-360, 2009.

[34] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *Journal of big data,* vol. 7, no. 1, p. 94, 2020, doi: 10.1186/s40537-020-00369-8.

[35] S. Shastri, P. Kour, S. Kumar, K. Singh, and V. Mansotra, "GBoost: A novel Grading-AdaBoost ensemble approach for automatic identification of erythemato-squamous disease," *International Journal of Information Technology,* vol. 13, pp. 959-971, 2021, doi: 10.1007/s41870-020-00589-4.

[36] D. T. Nguyen, J. K. Kang, T. D. Pham, G. Batchuluun, and K. R. Park, "Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence," *Sensors,* vol. 20, no. 7, p. 1822, 2020, doi: 10.3390/s20071822.

[37] H. Ye *et al.*, "An intelligent platform for ultrasound diagnosis of thyroid nodules," *Scientific Reports,* vol. 10, no. 1, p. 13223, 2020, doi: 10.1038/s41598-020-70159-y.

[38] H. Zhou *et al.*, "Differential diagnosis of benign and malignant thyroid nodules using deep learning radiomics of thyroid ultrasound images," *European Journal of Radiology,* vol. 127, p. 108992, 2020, doi: 10.1016/j.ejrad.2020.108992.

[39] S. Tsantis, N. Dimitropoulos, D. Cavouras, and G. Nikiforidis, "Morphological and wavelet features towards sonographic thyroid nodules evaluation," *Computerized Medical Imaging and Graphics,* vol. 33, no. 2, pp. 91-99, 2009, doi: 10.1016/j.compmedimag.2008.10.010.

[40] Y.-y. Guo *et al.*, "Machine learning for identifying benign and malignant of thyroid tumors: A retrospective study of 2,423 patients," *Frontiers in Public Health,* vol. 10, p. 960740, 2022, doi: 10.3389/fpubh.2022.960740.

[41] S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian, "Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study," *European Archives of Oto-Rhino-Laryngology,* vol. 281, no. 4, pp. 2095-2104, 2024, doi: 10.1007/s00405-023-08299-w.