

Article

Not peer-reviewed version

---

# Predictive Analytics for Thyroid Cancer Recurrence: A Machine Learning Approach

---

[Elizabeth Clark](#) , [Samantha Price](#) , [Theresa Lucena](#) , [Bailey Haberlein](#) , [Abdullah Wahbeh](#) \* , [Raed Seetan](#)

Posted Date: 9 September 2024

doi: [10.20944/preprints202409.0667.v1](https://doi.org/10.20944/preprints202409.0667.v1)

Keywords: Differentiated Thyroid Cancer; SMOTE; Machine Learning; Predictive Analytics; Random Forest Classifier



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

## Article

# Predictive Analytics for Thyroid Cancer Recurrence: A Machine Learning Approach

Elizabeth Clark, Samantha Price, Theresa Lucena, Bailey Haberlein, Abdullah Wahbeh and Raed Seetan

\* Correspondence: abdullah.wahbeh@sru.edu

**Abstract:** Differentiated thyroid cancer (DTC), comprising of papillary and follicular thyroid cancers, is the most prevalent type of thyroid malignancy. Accurate prediction of DTC is crucial for improving patient outcomes. Machine learning (ML) offers a promising approach to analyze risk factors and predict cancer recurrence. In this study, we aimed to develop predictive models to identify patients at an elevated risk of DTC recurrence based on 16 risk factors. We developed six ML models and applied them to a DTC dataset. We evaluated the ML models using Synthetic Minority Over-Sampling Technique (SMOTE) and with hyperparameter tuning. We measured the models' performance using precision, recall, F1 score, and accuracy. Results showed that Random Forest consistently outperformed the other investigated models (KNN, SVM, Decision Tree, AdaBoost, and XGBoost) across all scenarios, demonstrating high accuracy and balanced precision and recall. The application of SMOTE improved model performance and hyperparameter tuning enhanced overall model effectiveness.

**Keywords:** differentiated thyroid cancer; SMOTE; machine learning; predictive analytics; random forest classifier

## 1. Introduction

Differentiated thyroid carcinoma (DTC), which encompasses both papillary and follicular thyroid cancers, represents the most prevalent type of thyroid malignancies [1]. The recurrence risk is influenced by many factors, necessitating accurate predictive models to enhance patient outcomes. Over the past two decades, incidence of thyroid cancers has increased internationally, however rate of mortality has stayed consistent [2]. This lack of increased relative mortality is thought to be associated with contemporary advancements in diagnosis and treatment approaches that emphasize the importance of accurate risk assessment in predicting recurrence and tailoring early interventions.

Traditional non-technical approaches to monitor for and detect recurrence rely heavily on evaluation of risk assessment data, including biological markers, genetic factors, imaging, and comorbidities, however assessment of all these factors in combination is a challenge in a busy medical environment. In recent years, research into and use of machine learning (ML) has become increasingly attractive for this purpose to assess complex data sets consisting of patient risk factors and use this analysis to predict risk for cancer occurrence and recurrence. Although these methods are promising, they are not yet widely used.

More research on this topic with larger and more diverse datasets are needed to thoroughly demonstrate the importance of ML in the healthcare setting regarding recurrent cancer. The aim of this study is to enrich existing literature by developing predictive models to identify patients at an elevated risk for recurrence of DTC based on 16 commonly accepted risk factors, including patient data, treatment types, and personal histories. More specifically, we aim to develop six machine learning algorithms and determine which one among them can provide the most accurate predictions of thyroid cancer recurrence.

We aim to utilize a dataset comprising 383 instances and 16 features, sourced from the Differentiated Thyroid Cancer Recurrence dataset in the UC Irvine Repository database [3]. This

dataset includes various factors such as age, gender, smoking status, history of radiotherapy, thyroid functioning, adenopathy status, pathology category, risk category, focality, primary tumor (T), regional lymph nodes (N), distant metastasis (M), cancer stage, and response to treatment. Supervised learning and ensemble learning algorithms will be employed to analyze the data and predict the risk of recurrence. The performance of each model will be evaluated using precision, recall, F1 score, and accuracy. This study aims to contribute to this critical area by developing robust predictive models to enhance the management and prognosis of patients with DTC.

## 2. Background and Related Work

### 2.1. Overview of DTC and Thyroid Cancer

The thyroid is an organ in the neck that produces hormones that are key for body functions, the most important of which is thyroid hormone. Thyroid hormone, which is segmented into two configurations, thyroxine (T4) and triiodothyronine (T3), is particularly important for cell proliferation and differentiation as well as cellular metabolic processes (Bhattacharya et al., [5]). High levels of TSH are well established to be related to thyroid malignancy, but recent studies also find that T3 and T4 may also be involved in thyroid cancer development. Sasson et al., [4] found that high levels of FT4 were directly associated with DTC malignancy and a high FT4/FT3 ratio also significantly increased risk of malignancy. Additionally, heightened levels of thyroid stimulating hormone (TSH) are also known to be associated with increased likelihood of nodule malignancy [4].

Although there are several types of thyroid cancers, DTC, which consists of both papillary and follicular cancers of the thyroid, is by far the most prevalent kind of thyroid cancer, making up greater than 90% of all diagnoses [5]. Disease recurrence associated with DTC affects roughly 30% of patients within 10 years of initial diagnosis, making DTC dangerous in the long-term as well as short [1]. Papillary thyroid cancer (PTC) is the most prevalent form of thyroid cancer, and is characterized by slow growth, but often spreads to surrounding lymph nodes in the neck [6]. Follicular thyroid cancer (FTC), which represents approximately 10-15% of all thyroid cancers, is also characterized by slow growth, however it is also associated with a higher rate of metastasis to distant parts of the body due to invasion of the blood vessels [5]. When dissecting the two subcategories of DTC, it is apparent that though DTC is very treatable, it also poses significant concerns for recurrence due to its subcategories' proclivity for metastasis.

### 2.2. Current Methods for Monitoring Recurrence

DTC is most commonly recurrent in regional or cervical lymph nodes, making them prone to increased metastasis and mortality [1]. Classification systems are an important part of both treating DTC as well as predicting likelihood of recurrence in the future. The most widely used of these classification systems in clinics today include the American Thyroid Association (ATA) risk classification system and tumor node metastasis (TNM) staging. The ATA risk classification categorizes DTC into high, intermediate, and low risk of recurrence based on biological factors such as histological subtype, size and extent of the primary tumor, and presence of distant metastasis [7]. Staging, on the other hand, is the process of predicting prognosis of cancer patients. This has traditionally been done through TNM staging, which takes into consideration factors regarding the tumor characteristics, lymph node involvement, and metastasis of the cancer. Staging in this way is useful to plan current treatments and determine progression over time, however TNM staging is criticized for lacking important biological characteristics of malignant tumors [8].

### 2.3. Current Research on Machine Learning in Thyroid Cancer

In recent years, artificial intelligence (AI) has gained popularity for its potential uses in medicine, with cancer being a particularly prospective area of involvement. Cancers in general are highly multifactorial, in that various genetic, epigenetic, proteomic, and transcriptome changes can affect an individuals' likelihood of developing any number of different cancers, therefore multiple complex factors must be considered at the same time to predict outcomes and create best treatments [5]. This kind of complexity is consistent in thyroid cancers, which are likely associated with genetic factors

and environmental ones. Although it is very difficult to consider so many factors simultaneously and on an individual basis with non-technological means, AI can be used as a tool to do exactly that. By analyzing complex multi-omics data in an efficient way, AI has the potential to help diagnose current cancers, predict prognosis, discover new biomarkers for cancer, identify underlying mechanisms, and develop personalized treatments to more effectively eliminate thyroid cancer [5,9].

One of the most promising and currently implemented uses for Machine Learning (ML) in thyroid cancer healthcare is its use in diagnostics and screening. Classic ML algorithms are currently used in the computer-aided diagnosis (CAD) systems to improve accuracy of diagnosis and to reduce time required for image interpretation, though ultimately determination about diagnosis is still solely determined by the medical professional [10].

Thyroid Imaging Reporting and Data System (TI-RAD) is a commonly used method for categorizing biopsied thyroid nodules today [11]. These categories are labelled 1-5, with TR1 being benign and TR5 being highly suspicious and are based on a points scale determined through nodule factors, such as composition, echogenicity, shape, margin, and echogenic foci. Gu et al., [9] constructed a classification model for thyroid cancer based on risk factors as well as a prediction model for metastasis based on risk factors. Authors argue that TI-RADS, though a useful screening tool, are not as accurate at determining malignancy and metastasis as it could be. They utilized ML to predict both malignancy and metastasis of 1,735 patients, aiming to improve early diagnosis and treatment by analyzing risk factors. Results showed that XGBoost achieved the highest performance, suggesting that ML can significantly aid in early diagnosis and treatment decisions compared to TI-RADS.

Fine-needle aspiration biopsy (FNAB) is often used for suspicious thyroid nodules found via ultrasound. However, up to 30% of these may be classified as indeterminate thyroid nodules (ITN), often necessitating further surgery to determine malignancy [12]. To develop a cost-effective, non-invasive ML model, Luong et al., [12] analyzed electronic health record data from 355 nodules classified as indeterminate by FNAB. They found that a random forest classifier had the best performance. Findings demonstrated the potential of ML models to aid in early clinical decision-making and reduce unnecessary procedures.

Ballester et al., [13] conducted a retrospective analysis of 5,351 thyroid tumors, investigating pathologic upstaging, where the final pathologic stage exceeds the initial clinical stage. They reported upstaging rates of 17.5% for tumor stage, 18% for nodal stage, and 10.9% for summary stage, identifying factors like Asian race, older age, and lymph vascular invasion as contributors to upstaging. This study underscores the importance of recognizing factors that contribute to upstaging for improved management and counseling of thyroid cancer patients.

Distant metastasis often indicates poor prognosis, as metastasis to other body parts can be more challenging to find and treat. Mao et al., [14] studied 5,809 patients to evaluate ML models for predicting distant metastasis in follicular thyroid carcinoma (FTC). They found that the XGBoost model had the best performance, with diagnosis age, race, extrathyroidal extension, and lymph node invasion being significant risk factors.

In a retrospective analysis, Medas et al., [15] investigated factors influencing recurrence in 579 DTC patients. They found a recurrence rate of 6.2% and a five-year disease-free survival rate of 94.1%. Multivariate analysis identified lymph node metastasis as a strong predictor of recurrence, with multifocality and extrathyroidal extension also associated with increased risk. Conversely, microcarcinoma (tumor size  $\leq 1\text{cm}$ ) was an independent protective factor, emphasizing the need for risk stratification in personalizing treatment plans. Findings suggest that high-risk patients may benefit from more aggressive follow-up and treatment to better prevent recurrence.

Jin et al., [16] developed an overall survival (OS) prognostic model for participants with differentiated thyroid cancer with distant metastasis. Nine variables were introduced to build a machine learning model, receiver operating characteristic (ROC) was used to evaluate the recognition ability of the model, calibration plots were used to obtain prediction accuracy, and decision curve analysis (DCA) was used to estimate clinical benefit. The proposed was found to have good discriminative ability and high clinical value in its 10-year survival predictions.

Tang et al., [17] developed a nomogram to predict cancer-specific survival (CSS) in patients with PTC. They utilized the Surveillance, Epidemiology, and End Results (SEER) database to procure participants for the study. COX regression analysis demonstrated that age, gender, marriage, tumor grade, TNM stage, surgery, radiotherapy, chemotherapy, and tumor size were significantly associated with CSS in middle-aged patients with PTC. These ten variables were then used to develop a prediction model that could predict and affect the CSS of middle-aged PTC. This tool was found to have good accuracy and discrimination, and better overall clinical value than traditional TNM staging for this population. Park and Lee [18] utilized five ML models to determine which best predicted recurrence of PTC in a cohort of 1040 patients. Results showed that the Decision Tree (DT) model achieved the best accuracy at 95% and the lightGBM and stacking models together achieved 93% accuracy.

Wang et al., [19] used five ML models to predict structural recurrence in papillary thyroid cancer (PTC) patients, analyzing electronic medical records from 2,244 patients. The authors utilized the least absolute shrinkage and selection operator (LASSO) method to select nine variables for developing the prediction models, which included thyroglobulin (TG), lymph node (LN) variables (LN dissection, number of LNs dissected, lymph node metastasis ratio (LNR), and N stage), comorbidities and metabolic-related variables (comorbidity of hypertension, comorbidity of diabetes, BMI, and low-density lipoprotein (LDL)). Variable importance analysis showed that the most important variables across all models were TG, LNR, and N stage. The top performing models were SVM, XGBoost, and Random Forest (RF) models, all of which showed better discrimination than the ATA risk stratification according to the AUC values and corresponding indices. Furthermore, their RF model was found to have the most consistent calibration, as well as good discrimination and interpretability. Findings suggest that patients with recurrent disease are more likely to be older, male, cigarette smokers, alcohol drinkers, and have various comorbidities, highlighting the potential of ML in enhancing current risk stratification methods and assisting in personalized patient management.

Finally, Borzooei et al., [20] conducted a prospective study using the Differentiated Thyroid Cancer Recurrence dataset that is also being used in our study. They trained ML models on three distinct combinations of features: a data set with all features excluding ATA risk score (12 features), another with ATA risk alone, and a third with all features combined (13 features). Authors found that the model that combined the clinicopathologic features with ATA risk score outperformed the other two models. SVM was found to be the best performing ML model.

### 3. Design and Methodology

The Differentiated Thyroid Cancer Recurrence dataset from the University of California at Irvine Machine Learning Repository was used in this study [3]. This dataset consists of the retrospective clinical data for 383 patients diagnosed with DTC, each followed for a minimum of ten years. The collected clinical data included 16 features: age at diagnosis, gender, current smoking status, prior smoking history, history of head and neck radiation therapy, thyroid function, presence of goiter, presence of adenopathy on physical examination, pathological subtype of cancer, focality, ATA risk assessment, TNM staging, initial treatment response, and recurrence status. The dataset contains 312 females (81%) and 71 males (19%). The average age at diagnosis was 41. The pathological subtype breakdown was 287 Papillary (75%), 48 Micropapillary (13%), 28 Follicular (7%), and 20 Hurthel Cell (5%). According to the ATA risk classification, patients were classified as follows: 249 low risk (65%), 102 intermediate risk (27%), and 32 high risk (8%). Most cases (333) were classified as Stage 1 (87%). 208 patients had an excellent initial treatment response (54%). The remaining cases had the following initial treatment responses: 91 structural incomplete (24%), 61 indeterminate (16%), and 23 biochemical incomplete (6%). In terms of recurrence, 108 patients (28%) experienced recurrence. Notably, this dataset contained no missing values.

ML models were applied in this study to analyze and predict DTC recurrence. The ML models used were KNN, SVM, Decision Tree, Random Forest, AdaBoost, and XGBoost. KNN is a simple, instance-based learning algorithm that classifies data points based on the majority class among their

k-nearest neighbors in the feature space, using a chosen distance metric. SVM works by using a kernel trick to map the inputs into high-dimensional feature spaces and draw margins between the classes [21]. It is robust against outliers and works well with high dimensional small datasets [22]. Decision Tree represents choices and their results in a tree-shaped graph. The results are easy to interpret but can be prone to overfitting and tend to be sensitive to data changes [21,22]. Random Forest uses a parallel ensemble method to create multiple decision trees. It is generally more accurate than a single decision tree and can handle high dimensional data [23]. AdaBoost is an ensemble method that improves poor classifiers by learning from prior errors. It generally performs well and has better accuracy than other methodologies. This algorithm can be sensitive to noisy data and outliers [21,23]. XGBoost is an ensemble model that uses a sequential method to build multiple decision trees, and each future tree corrects the errors made by the previous one. It has high accuracy, efficiency, and can handle missing values well. It generally does require careful tuning of hyperparameters to perform optimally [23].

Python was used to run the different models. Python has a rich variety of libraries and tools which make python an excellent choice for implementing machine learning classification techniques. From data preparation and model training to evaluation and deployment, Python provides comprehensive support for every step of the machine learning workflow. The dataset was partitioned into a training set (80%) and a test set (20%). Data was run once with no modifications. Given the gender imbalance in this dataset, a second experiment was performed using SMOTE to address class imbalance. SMOTE generates synthetic samples for the minority class based on feature similarity with existing minority instances. This technique helps to alleviate the bias towards the majority class by increasing the representation of the minority class, thereby improving the performance of classifiers trained on imbalanced datasets. In hopes of improving ML model performance, a final run was conducted using hyperparameter tuning. This process involves selecting the best values for these parameters to achieve optimal model performance. Grid search was the method used for hyperparameter tuning which allows you to systematically find the best hyperparameter combination to optimize model performance. Hyper-parameter optimization techniques in machine learning encompass manual methods like trial-and-error tuning and exhaustive grid search. Random search efficiently explores broader spaces by random sampling [24].

The performance of the models was assessed using accuracy, precision, recall, and F1 score. These metrics provide a comprehensive view of each model's effectiveness. Accuracy shows how often the model was able to correctly predict if DTC recurred or not. It can be misleading at times if classes are imbalanced. Precision indicates how many patients predicted to have DTC recurrence actually had recurrence. High precision means that when the model predicts recurrence, it is likely to be correct. This is critical for minimizing false alarms which may cause unnecessary treatment. Recall shows how effectively the model predicts patients who will experience recurrence. High recall shows that the model will be able to predict most recurrences. This ensures that all at risk patients are identified. This minimizes false negatives and helps ensure that all patients receive the proper treatment needed to minimize their chances of recurrence. Finally, a high F1 score shows that the model has a good balance of precision and recall. This score can be especially helpful in cases of class imbalance [21].

#### 4. Results

In the initial run with no modifications (Table 1), KNN demonstrated strong overall performance with high accuracy (0.90) and precision (0.91). It did have a lower recall (0.81) which could mean that the model would miss some positive cases of recurrence.

**Table 1.** Summary of results without modification.

Model	Accuracy	Precision	Recall	F1
KNN	0.90	0.91	0.81	0.84
SVM	0.83	0.91	0.66	0.69

<b>Decision Tree</b>	0.92	0.89	0.91	0.90
<b>Random Forest</b>	0.99	0.99	0.97	0.98
<b>AdaBoost</b>	0.97	0.98	0.95	0.96
<b>XGBoost</b>	0.97	0.97	0.97	0.97

The SVM model had a lower accuracy (0.83), but still had good precision (0.91). The recall was low (0.66), which meant the model struggled to identify recurrence. The Decision Tree model had overall strong performance with high accuracy (0.92), precision (0.89), recall (0.91), and F1 score (0.98). A training score of 1.0 indicates there may have been overfitting to the training data. The test score of 0.92 indicates that the model still performs well with new data.

The Random Forest model had the best overall performance with near-perfect accuracy (0.99), precision (0.99), and F1 score (0.98). The recall (0.97) was only slightly lower, which indicates that this model was reliable for predicting DTC recurrence. The high training and test scores also indicate good model generalization and minimal overfitting.

AdaBoost also had high performance with high accuracy (0.97), precision (0.98), recall (0.95), and F1 score (0.96). The close training and test scores indicate that it is a robust choice. Finally, XGBoost also had consistently high metrics for accuracy (0.97), precision (0.97), recall (0.97), and F1 score (0.97). A training score of 1.0 does show the possibility of overfitting with this model.

The data was run again with the application of SMOTE (Table 2) to address class imbalances. Overall, the models showed good improvement with the application of SMOTE. KNN showed significant improvement, achieving high accuracy (0.97), precision (0.97), recall (0.97), and F1 score (0.97).

**Table 2.** Summary of results using SMOTE.

<b>Model</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1</b>
<b>KNN</b>	0.97	0.97	0.97	0.97
<b>SVM</b>	0.94	0.94	0.94	0.94
<b>Decision Tree</b>	0.94	0.94	0.94	0.94
<b>Random Forest</b>	0.95	0.95	0.96	0.95
<b>AdaBoost</b>	0.95	0.95	0.96	0.95
<b>XGBoost</b>	0.96	0.96	0.96	0.96

SVM greatly benefited from SMOTE, indicating that the model was likely influenced by the class imbalance. SVM had balanced performance with good accuracy (0.94), precision (0.94), recall (0.94), and F1 score (0.94). The Decision Tree model also showed improvements with consistent performance for accuracy (0.94), precision (0.94), recall (0.94), and F1 score (0.94). Once again, the perfect training score indicated that there could be a potential of overfitting with this model.

Random Forest did see a slight decrease in performance when SMOTE was applied. This could indicate that there had been some overfitting to the initial model. Since Random Forest is an ensemble method it can typically handle some degree of class imbalance. Introducing the synthetic samples to the data can add noise to the data which can also explain the decreased performance.

AdaBoost did see a slight decrease to both accuracy (0.95) and precision (0.95). Recall (0.96) did see some slight improvement, but the model is less apt to favor one class disproportionately. XGBoost saw similar results to the initial run and once again showed the potential for some overfitting with this model.

The final test was conducted using hyperparameter tuning (Table 3). The purpose of running hyperparameter tuning was to improve each algorithm's performance by finding the parameters which would ultimately maximize the performance of each algorithm. KNN had a consistent performance with no significant changes.

**Table 3.** Summary of results using Hyperparameter Tuning.

Model	Accuracy	Precision	Recall	F1	Support
KNN	0.90	0.91	0.81	0.84	77
SVM	0.94	0.94	0.89	0.91	77
Decision Tree	0.97	0.98	0.95	0.96	77
Random Forest	0.99	0.99	0.97	0.98	77
AdaBoost	0.96	0.98	0.92	0.94	77
XGBoost	0.97	0.97	0.97	0.97	77

Despite the hyperparameter tuning, recall remained lower (0.81), suggesting that the model could miss some cases of potential recurrence. SVM showed good improvement from the initial results, indicating hyperparameter tuning enhanced the model's ability to predict DTC recurrence. Decision Tree showed marked improvement with accuracy (0.97), precision (0.98), recall (0.95), and F1 score (0.96). This shows that hyperparameter tuning helped the Decision Tree model reduce overfitting and improve generalization.

Random Forest once again had a very high performance across the board for accuracy (0.99), precision (0.99), recall (0.97), and F1 score (0.98). Hyperparameter tuning helped to solidify its performance and make this model very reliable for predicting DTC recurrence. AdaBoost saw an improvement with precision (0.98) but did see a slight drop with recall (0.92).

Finally, XGBoost showed consistent performance with accuracy (0.97), precision (0.97), recall (0.97), and F1 score (0.97). This demonstrates that hyperparameter tuning helps enhance this model allowing it to maintain high performance across all metrics.

## 5. Discussion

The goal of this study was to determine what the best method was to predict likelihood of recurrence of DTC. To accomplish this, six different ML models were utilized, along with three different sets of parameters. SMOTE was used to handle any class imbalance in the training set. In order to clarify its use, we chose to run the ML algorithms both without using SMOTE and then once again with it. The purpose of running it without it was to create a baseline performance marker for each of the individual algorithms. After that, hyperparameter tuning was used as a third method to help analyze this dataset. Hyperparameter tuning selects the best possible values for each machine learning model to achieve the optimal performance. Through these three different parameters and six different machine learning models it was determined that the best machine learning model was Random Forest, which consistently outperformed or matched the other models across all three scenarios.

Overall, the application of SMOTE did improve the performance of most models, particularly those that struggled with imbalanced data, such as KNN and SVM. Hyperparameter tuning further enhanced performance, especially for models like SVM and Decision Tree. Random Forest, AdaBoost, and XGBoost demonstrated strong performance in all scenarios, making them reliable choices for handling imbalanced datasets. Hyperparameter tuning further enhanced the performance of most models, particularly SVM, Decision Tree, and AdaBoost. In all three scenarios Random Forest seems to be the best and most reliable algorithm. Random Forest consistently scored higher than the other algorithms in each scenario and would lead to the most reliable results when determining likelihood of recurrence of thyroid cancer. These findings underscore the importance of addressing data imbalance and optimizing model parameters to achieve the best predictive performance in medical diagnosis tasks. Future studies should explore these models in larger and more diverse datasets to validate and generalize these findings.

Results of this study imply interesting clinical applications, particularly advancements in personalized treatment development. This tool in combination with oversight from clinical experts could be utilized to better predict chances of recurrent DTC in patients following primary treatment methodologies. Understanding risk levels for recurrence allows clinicians to create more effective screening and monitoring programs at regular intervals to better detect recurrence in earlier stages

and to monitor more closely for distant metastasis that often increases risk of mortality in recurrent DTC. Overall, this study demonstrates some promising implications for clinical use, with the caveat that classic ML algorithms such as this one requires ongoing oversight from experts in the field to ascertain ongoing accuracy, particularly with application to larger and larger datasets.

## 6. Conclusions and Limitations

Random Forest is the superior algorithm when determining the recurrence of thyroid cancer. Whether the data is imbalanced, balanced, or ran with optimal settings, Random Forest continuously outperforms the other five investigated ML algorithms (KNN, SVM, Decision Tree, AdaBoost, and XGBoost). This study aimed to correct the imbalance that was seen in the previous study using this dataset. While it was corrected and proved to have a better algorithm to use in predicting recurrence of thyroid cancer, there still were limitations. Ideally having more data to work with would give a better or more accurate prediction. The recurrence of DTC is typically seen more in women than in men, so while the data may still be imbalanced if there was more of it, a stronger conclusion could be made utilizing a larger, more diverse dataset.

## References

1. A. Shokoohi *et al.*, "Treatment for Recurrent Differentiated Thyroid Cancer: A Canadian Population Based Experience," *Cureus*, vol. 12, no. 2, p. e7122, 2020, doi: 10.7759/cureus.7122.
2. A. Coca-Pelaz *et al.*, "Recurrent Differentiated Thyroid Cancer: The Current Treatment Options," *Cancers*, vol. 15, no. 10, Art. no. 10, Jan. 2023, doi: 10.3390/cancers15102692.
3. A. T. Shiva Borzooei, "Differentiated Thyroid Cancer Recurrence." UCI Machine Learning Repository, 2023. doi: 10.24432/C5632J.
4. M. Sasson *et al.*, "The T4/T3 quotient as a risk factor for differentiated thyroid cancer: A case control study," *Journal of Otolaryngology - Head & Neck Surgery*, vol. 46, no. 1, p. 28, Jan. 2017, doi: 10.1186/s40463-017-0208-0.
5. S. Bhattacharya, R. K. Mahato, S. Singh, G. K. Bhatti, S. S. Mastana, and J. S. Bhatti, "Advances and challenges in thyroid cancer: The interplay of genetic modulators, targeted therapies, and AI-driven approaches," *Life Sciences*, vol. 332, p. 122110, Nov. 2023, doi: 10.1016/j.lfs.2023.122110.
6. Y. Habchi *et al.*, "AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions," *Systems*, vol. 11, no. 10, Art. no. 10, Oct. 2023, doi: 10.3390/systems11100519.
7. -Brown Peter Watson and D. Anderson, "Differentiated thyroid cancer: A guide to survivorship care," *Australian Journal of General Practice*, vol. 52, no. 1/2, pp. 47–51, Jan. 2023, doi: 10.3316/informit.896817871599081.
8. N. Tang, C. Yang, J. Fan, and L. Cao, "VerifAI: Verified Generative AI," *arXiv preprint arXiv:2307.02796*, 2023.
9. J. Gu *et al.*, "A machine learning-based approach to predicting the malignant and metastasis of thyroid cancer," *Front. Oncol.*, vol. 12, Dec. 2022, doi: 10.3389/fonc.2022.938292.
10. L. Nagendra, J. M. Pappachan, and C. J. Fernandez, "Artificial intelligence in the diagnosis of thyroid cancer: Recent advances and future directions," *Artificial Intelligence in Cancer*, vol. 4, no. 1, pp. 1–10, 2023.
11. H. Ahmad and A. Van Der Lugt, "The Radiology Assistant : TI-RADS - Thyroid Imaging Reporting and Data System." Accessed: Aug. 12, 2024. [Online]. Available: <https://radiologyassistant.nl/head-neck/ti-rads/ti-rads>
12. G. Luong, A. J. Idarraga, V. Hsiao, and D. F. Schneider, "Risk Stratifying Indeterminate Thyroid Nodules With Machine Learning," *Journal of Surgical Research*, vol. 270, pp. 214–220, Feb. 2022, doi: 10.1016/j.jss.2021.09.015.
13. J. Ballester, Finn, Caitlin, S. Ginzberg, R. Kelz, and H. Wachtel, "Thyroid cancer pathologic upstaging: Frequency and related factors," *The American Journal of Surgery*, vol. 226, no. 2, pp. 171–175, Aug. 2023, doi: 10.1016/j.amjsurg.2023.03.023.
14. Y. Mao *et al.*, "Machine learning algorithms are comparable to conventional regression models in predicting distant metastasis of follicular thyroid carcinoma," *Clinical Endocrinology*, vol. 98, no. 1, pp. 98–109, 2023, doi: 10.1111/cen.14693.
15. F. Medas, G. L. Canu, F. Boi, M. L. Lai, E. Erdas, and P. G. Calò, "Predictive Factors of Recurrence in Patients with Differentiated Thyroid Carcinoma: A Retrospective Analysis on 579 Patients," *Cancers*, vol. 11, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/cancers11091230.
16. S. Jin *et al.*, "A Predictive Model for the 10-year Overall Survival Status of Patients With Distant Metastases From Differentiated Thyroid Cancer Using XGBoost Algorithm-A Population-Based Analysis," *Front. Genet.*, vol. 13, Jul. 2022, doi: 10.3389/fgene.2022.896805.

17. J. Tang *et al.*, "Development and validation of a nomogram to predict cancer-specific survival in middle-aged patients with papillary thyroid cancer: A SEER database study," *Helijon*, vol. 9, no. 2, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13665.
18. Y. M. Park and B.-J. Lee, "Machine learning-based prediction model using clinico-pathologic factors for papillary thyroid carcinoma recurrence," *Sci Rep*, vol. 11, no. 1, p. 4948, Mar. 2021, doi: 10.1038/s41598-021-84504-2.
19. H. Wang *et al.*, "Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches," *BMC Cancer*, vol. 24, no. 1, p. 427, Apr. 2024, doi: 10.1186/s12885-024-12146-4.
20. 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," *Eur Arch Otorhinolaryngol*, vol. 281, no. 4, pp. 2095–2104, Apr. 2024, doi: 10.1007/s00405-023-08299-w.
21. B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR). [Internet]*, vol. 9, no. 1, pp. 381–386, 2020.
22. H. Lickert, A. Wewer, S. Dittmann, P. Bilge, and F. Dietrich, "Selection of Suitable Machine Learning Algorithms for Classification Tasks in Reverse Logistics," *Procedia CIRP*, vol. 96, pp. 272–277, Jan. 2021, doi: 10.1016/j.procir.2021.01.086.
23. I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN COMPUT. SCI.*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.
24. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.