Ahliya Journal of Allied Medico-Technology Science
[Published by Palestine Ahliya University]
Vol. 02 Issue 01 (2025)
ISSN: 3007-9489

Artificial Intelligence Fights Thyroid Cancer: Accurate Prediction of Thyroid Cancer Recurrence Risk through Machine Learning Models

Khaled Sabarna[1], Murad Zeer[*2], Mutaz Rasmi Abu Sara[**2]
[1] Nursing Department, Faculty of Allied Medical Sciences, Palestine Ahliya University (Palestine)
✉ k.sabarna@paluniv.edu.ps
[2] Faculty of Engineering and Information Technology, Palestine Ahliya University (Palestine)
[*]✉ muradzeer@paluniv.edu.ps
[**]✉ moutaz.a@paluniv.edu.ps

Received:30/03/2025 Accepted:25/04/2025 Published:30/04/2025

Abstract: ***The aim of this study is to determine the possible use of machine learning models in improving predictive accuracy of recurrence of thyroid cancer, particularly Differentiated Thyroid Cancer (DTC). A cohort of 383 patients who underwent radioactive iodine treatment was studied with six machine learning models, i.e., Random Forest, Decision Tree, and K-Nearest Neighbors. Experiments showed that Random Forest was the best-performing model, which had 98.7% test set accuracy, showing that it had strong predictive power with accuracy and reliability. Data rebalancing solved data imbalance issues, as well as the utilization of methods for hyperparameter tuning. Feature selection and the use of laboratory and clinical data in improving performance were also a high priority in the research. But more research will be required to enhance clinical transparency and develop models on more heterogeneous data, incorporating genetic and biochemical biomarkers.***

Keywords: Artificial Intelligence, Thyroid Cancer, Cancer Recurrence Prediction, Machine Learning, Predictive Modeling.

1. Introduction

Thyroid cancer is among the diseases that have witnessed a significant spike in incidence levels within the last few years. One of the biggest challenges of this disease's treatment is cancer relapse after treatment since cancer relapse can prove difficult to anticipate in some cases even without visible clinical symptoms [1]. Machine learning and artificial intelligence are among the newer methods that have proven to be very promising in increasing recurrence risk prediction accuracy and follow-up of patients [2], [3]. Through these technologies, most studies focus on building machine learning models for predicting relapse in thyroid cancer patients based on a mix of clinical and lab data.

Different techniques have been used in coming up with these models, and based on a study conducted by[4], machine learning algorithms such as bagging algorithms have been helpful in the 98.17% accuracy level of predicting recurrence of thyroid cancer. The rationale for such efficiency has been that the research emphasized especially on accurate feature selection together with extensive and diverse datasets. Moreover, empirical studies have already demonstrated that the use of methods such as Random Forest can be greatly enhanced for prediction accuracy, particularly in rich data environments [5].

Along with the continuation of the activity in this direction, it has become essential to mention the difficulties with the application of machine learning models in healthcare. Other studies like [6] have established that models derived using methods like multiple linear regression can be somewhat effective but with high rates of error, implying that it is possible to improve the models and use them on more varied sets of data so that their results can be generalized. On the other hand, model interpretability is essential to justify their use in the clinic. Studies like [7] have found that through the use of techniques like XGBoost and SHAP for assessing feature interpretative influence, model usability in clinical clinics can be enhanced, in the sense that it promotes higher confidence in the outputs and supports decision-making in a clinical context.

As explained in the above literature, it is evident how important it is to create machine learning models that can improve the integration of clinical, laboratory, and genetic data. This will assist in improving the potential for thyroid cancer recurrence prediction as well as enhancing healthcare plans for patients.

2. Literature Review

Recurrence of thyroid cancer, particularly differentiated thyroid cancer (DTC), is one of the biggest challenges of disease management since recurrence is difficult to forecast in some patients despite the absence of apparent clinical symptoms. Since the specialty has witnessed recent advancement using artificial intelligence, machine learning has emerged as a promising tool to enhance the predictive accuracy and fine-tune follow-up strategies in recurrent risk patients.

The experiment in [6] revealed the success of the integrated use of an unsupervised learning method through the application of ACA clustering algorithm and a supervised method through the use of multiple linear regression in predicting a recurrence risk of thyroid cancer. The data used in the study was gathered in the UCI repository and was normalized to be used with the help of Log-normalization and created a model to estimate the score of recidivism in the form of the recurrence risk score (RRS). However, despite the possibility of the model to classify the patients with accuracy of 63.4 with the label of high risk and non-high risk, the error rate of prediction was 21.68, which highlights the necessity of testing the model with more

diverse data to be generalized. In comparison, [4] provided the generalized overview of recent machine learning techniques used to predict the recurrence of DTC, which was supplemented with the accuracy rate of 98.17, outperforming the other algorithms in the test group. The paper has highlighted the proper choice of characteristics and the use of massive datasets to guarantee efficient predictive models to be used in the clinics.

[5] Six machine learning models were developed with 16 DTC risk factors and it was discovered that the Random Forest was the most successful one in terms of accuracy, recall and F1 score, especially when methodologies such as SMOTE were used to balance the classes. They also highlighted the need of hyperparameter optimization as a step towards global performance. The research did mention some weaknesses such as a small sample size and lack of clinical validation of its models.

[8] proposed a hybrid model based on association rule mining and classification algorithms based on 15-years-long clinical data. The accuracy of the RCAR was 96.7 percent and the predictors were found to be incomplete response and lymph node involvement, which were both found to be strongly predictive of recurrence. The model also stressed that medical algorithms should be transparent and interpretable because it will increase their acceptability in clinics.

[9] Examined the correlation of thyroglobulin levels stimulated with rhTSH and recurrence after thyroidectomy and radioactive iodine treatment. The results illustrated that a 2.5 ng/ml threshold level of the Tg marker may be employed in predicting recurrence risk with accuracy and emphasized the importance of integrating biochemical markers with AI-based prediction systems.

Finally, [7] used the XGBoost algorithm with SHAP (Shapley Additive Explanations) for interpretative analysis of the impact of each variable on the predictive models. This is an innovation in medical data science that offers explanations on the impact of each feature on the final result, thereby making the model more credible in clinic practice and a great decision-support tool.

The existing study demonstrates the huge potential of machine learning algorithms to forecast thyroid cancer recidivism from simple statistical models to rule-based hybrid models to complex interpretative models using algorithms like XGBoost. However, data imbalance problems, limited clinical verification, and needs for genetic and developmental data still remain. Therefore, the future of this branch which promises a lot lies in more clinically-focused studies with larger databases and interface between intelligent models and real biomarkers.

3.Methodology

3.1 Dataset

A Kaggle dataset is used as the basis of the study which has the history of 383 thyroid cancer patients who were treated by radioactive iodine (RAI). The dataset includes 13 clinical and medical characteristics including the age, gender, radiation history in the past, histologic tumor type, metastatic lymph nodes, grade of the tumor, staging, and tumor response to treatment. Worth mentioning, the data set is well-cleaned and there are no gaps, which contributes to the validity of the results related to the ability to predict the disease recurrence, evaluate risks, and key determinants.

3.2 Preprocessing

It initiated the study by means of the exploration data analysis (EDA) procedure, during which the descriptive statistics were explored in terms of variables like age and the frequency distribution of categorical variables. The target variable had a skewness, with the non-recurrence cases (No) being far more than those cases of recurrence (Yes). In order to overcome this problem, resampling was applied to improve the sample of the minority classes to the optimum level preventing the imbalance of both classes. Label Encoding was then used to encode the nominal features, and then StandardScaler was used to standardize the nominal features to achieve the same distribution of data as a precondition of distance-based algorithms. Finally, the dataset was split into training and test sets using an 80/20 ratio, and class balance was maintained through stratified sampling to ensure that both subsets preserved a representative distribution of all classes.

3.3 Algorithms Used

A variety of classification models were used in order to compare them in terms of prediction of cancer recurrence. They were also K-Nearest Neighbors (KNN) model, whose optimal performance was k=3 and the Support Vector Classifier (SVC) using the RBF kernel whose values were optimized by parameter (C=10 and gamma='auto'). Decision Tree and Naïve Bayes models were also trained for our base models; for the Decision Tree, the best results were achieved with the use of the "entropy" criterion at a depth of 5 and a minimum number of samples per node to avoid overfitting. The Naïve Bayes classifier and the Random Forest model were also employed to get comparative knowledge of various classification techniques.

3.4 Evaluation Metrics

In order to measure the performance of the models, the study employed a collection of effective evaluation measures that consider the overall accuracy and the distribution of errors. One of such measures was classification accuracy, which counted the correct predictions on all instances. Precision and Recall were the two measures to quantify models in terms of true positive classification and not failing to detect such instances. F1-Score, which is a combination of a combination of precision and recall, was also determined. Confusion Matrix was also used to compare the fine-grained error distribution amongst different classes. All these measurements constitute an evaluation framework, which is why it is reasonable to choose the most suitable model to predict thyroid cancer recurrence.

4. Results

The performance of six ML algorithms was assessed using classification metrics. Each model was evaluated on both the training and test sets to examine its generalization capability and to verify that overfitting was not occurring.

Table 1. Study Results

		KNN	SVC	LR	DT	NB	RF
Accuracy	Train	0.941	0.990	0.915	0.973	0.882	1.000
	Test	0.909	0.935	0.909	0.948	0.870	0.987
Precision	Train	0.943	0.993	0.899	0.982	0.885	1.000
	Test	0.904	0.939	0.892	0.966	0.871	0.991
Recall	Train	0.909	0.982	0.887	0.953	0.815	1.000
	Test	0.868	0.900	0.881	0.909	0.800	0.977
F1	Train	0.924	0.987	0.893	0.966	0.841	1.000
	Test	0.883	0.916	0.887	0.932	0.825	0.983

The Random Forest model was effective on all the measures on training and test sets. It reached 100 percent accuracy, precision, recall, and F1-score on the training set and this shows that it has an excellent ability to learn. The model scored 98.7, 99.11, and 97.73 in terms of accuracy, precision and recall respectively on the test set, indicating an exemplary ability to determine the result accurately and reliably.

The Decision Tree model performed excellently with a test set accuracy of 94.81, a recall of 90.91 and an F1-score of 93.25. The performance is an indicator that the model can render well on the data at fairly reduced complexity, particularly with the exploitation of optimally educated parameters including the criterion of entropy and tree depth limitation.

In the third place was the Support Vector Classifier (SVC) that obtained a testing accuracy of 93.51 and a F1-score of 91.69. It is also worthy to mention that the model demonstrated a consistent performance between training and testing with little variance, which is a good signifier of generalization. The large values of recall and precision are a pointer of the good generalization of the model in terms of false positives and false negatives.

The KNN model was relatively good with the test accuracy and test F1-score of 90.91 and 88.37 respectively. The model reported high score in terms of a substantial decrease in recall (86.82%), which implies that the model can possibly lose some of the positive cases of interest when making predictions on the reoccurrence of the disease.

The Logistic Regression model performed as well as KNN at test accuracy at 90.91% and an F1-score at 88.71%. Although these results are reasonably strong, the lower precision and recall relative to the best-performing models suggest that it may not be the optimal choice in situations where a high degree of balance—particularly in minimizing both false positives and false negatives—is essential.

Last, the Naïve Bayes model performed poorest with 87.01% testing accuracy, 82.53% F1-score, and poor 80.00% recall. This implies that the model underperformed with handling the features of the data being complicated even though the model itself has low complexity but is fast in its processing speed.

Based on these results, Random Forest is the optimal choice in this case according to its highest performance on all metrics, followed by Decision Tree and SVC. Naïve Bayes, however, performed relatively worse and thus is less appropriate for sensitive applications such as thyroid cancer recurrence prediction.

5. Discussion

The implications of the present study are educative in their discovery of the predictive capacity of the machine learning models in the forecasting of the recurrence of thyroid cancer, i.e., DTC. In line with the present trends in the field unraveled in the literature, this is in line with the increasing role of the machine learning in enhancing cancer recurrence prediction in terms of effectiveness and accuracy, particularly in the absence of clinical symptoms.

Compared to the previous ones, e.g., [4], [6], the findings of the given study suggest the high level of development of the machine learning algorithms related to the prediction of cancer recurrence. Chattopadhyay model is only recent but could only achieve a predictive accuracy of 63.4 which is indicative of the nature of the challenge which exists in making regular predictions on complex medical data. The test set accuracy of the Random Forest model used in this study was, nonetheless, 98.7% of the test set or, that is, the model performed better. The tricks to the success of machine learning-based medical applications such as the right selection of features, hyperparameter optimization, and balancing the unbalanced dataset are perhaps the reason behind such success.

The study also conformed [4], whereby it was observed that bagging algorithms can be good in their accuracy (98.17) when their estimates are on DTC recurrence. This is in line with the fact that, ensemble learning models, like Random Forest, which entails a composite of numerous decision trees, can overpredict other models by a significant margin when it comes to forecasting complex, high-dimensional medical data. This research also confirms the strength of big, diverse datasets and the role of feature engineering in improving predictability, as highlighted by both [4] [5].

From the interpretability standpoint, this study is in line with how model choices need to be explained in healthcare contexts, as discussed by [7], where they used SHAP values to provide insights into the prediction. This research did not directly use SHAP or similar interpretability methods, but since Random Forest performed better and is easy to interpret, it may be a good choice to be used clinically without further modification.

One of the principal issues brought up within this research, after reports by [5], is that the dataset was small and heterogeneous. While Random Forest is very high-performing, its accuracy remains limited by the lack of clinical validation and real-world variability in data. Thus, while the model is a good performer, application to larger and more variable sets as well as clinical biomarkers is required to render it more practical in real life.

Finally, [9] work that incorporates biochemical markers like thyroglobulin level in predictive models implies that future research could advantageously incorporate standard clinical biomarkers into AI models. The inclusion may overcome some of the present model constraints while making the prediction more accurate through representation of more patient-specific

determinants.

The paper contributes to the growing body of literature on the application of machine learning models to predict recurrence of thyroid cancer. The results show that the Random Forest algorithm is the most effective one with high precision, recall, and accuracy. Nonetheless, the research also presents the necessity of conducting additional research, especially the integration of genetic and biochemical data, the increase in model interpretability, and the complete validation of the models by clinic.

6. Conclusion

The capabilities of machine learning, in this case, the Random Forest model, to enhance the predictive accuracy of thyroid cancer recurrence, were proven in this research. We arrive at the conclusions that the current clinical practice should be improved to allow the well-informed and clear predictions based on the application of the key clinical practice and artificial intelligence. However, when applying these models completely, it must be considered that additional research will be required in resolving the problems of variability and quantity of data and model readability, so that they can be used in conforming to the requirements of the clinical environment. The integration of the genetic and biochemical biomarkers into the current models can cause such a great change to the personalized medicine and can further enhance the plans of patient healthcare. According to this, we would recommend that a future research should be conducted to improve the generalizability and clinical applicability of these models in order to make them applicable in real healthcare practice.

References

[1] D. S. Ross et al., “2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis,” Thyroid, vol. 26, no. 10, pp. 1343–1421, 2016.

[2] F. Bini et al., “Artificial intelligence in thyroid field—a comprehensive review,” Cancers, vol. 13, no. 19, p. 4740, 2021.

[3] M. Zeer, M. R. A. Sara, A. Sbeih, and K. Sabarna, “Predicting the Risk of Myocardial Infarction (MI) using Machine Learning (ML),” Ahliya J. Allied Medico-Technol. Sci., vol. 1, no. 1, pp. 26–29, 2024.

[4] I. Imtiaz et al., “The Future of Differentiated Thyroid Cancer Recurrence Prediction Using a Machine Learning Framework Advancements, Challenges, and Prospects,” in 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), IEEE, 2024, pp. 518–525.

[5] E. Clark, S. Price, T. Lucena, B. Haberlein, A. Wahbeh, and R. Seetan, “Predictive analytics for thyroid cancer recurrence: A machine learning approach,” Knowledge, vol. 4, no. 4, pp. 557–570, 2024.

[6] S. Chattopadhyay, “Towards Predicting Recurrence Risk of Differentiated Thyroid Cancer with a Hybrid Machine Learning Model,” Medinformatics, 2024.

[7] A. Schindele et al., “Interpretable machine learning for thyroid cancer recurrence predicton: leveraging XGBoost and SHAP analysis,” Eur. J. Radiol., vol. 186, p. 112049, 2025.

[8] F. Firat Atay et al., “A hybrid machine learning model combining association rule mining and classification algorithms to predict differentiated thyroid cancer recurrence,” Front. Med., vol. 11, p. 1461372, 2024.

[9] R. T. Kloos, “Thyroid cancer recurrence in patients clinically free of disease with undetectable or very low serum thyroglobulin values,” J. Clin. Endocrinol. Metab., vol. 95, no. 12, pp. 5241–5248, 2010.