RESEARCH PAPER

Assessing Machine Learning Approaches for Recurrence Prediction in Differentiated Thyroid Cancer: A Clinicopathological Data Analysis

Prashasti Kanikar

Department of Computer Engineering, MPSTME, SVKM's Narsee Monjee Institute of Management Studies (NMIMS) Deemed-to-be-University, Mumbai, India

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT

The recurrence of differentiated thyroid cancer is a significant field of research, although the overall mortality is very low. This study compares various classification methods, including Random Forest, Logistic Model Trees, REP Tree, J48, Hoeffding Tree and Decision Stump, on a clinicopathological data set for the prediction of recurrence. The models are evaluated according to accuracy and errors defined as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Relative Absolute Error (RAE) and Root Relative Square Error (RRSE). The best performers, Random Forest, Logistic Model Trees and REP Tree achieved accuracy of more than 95% and reduced errors, indicating strong predictive value.

Keyword: *Thyroid Cancer, Machine Learning, Recurrence Prediction*

How to cite this article: Kanikar P , Assessing Machine Learning Approaches for Recurrence Prediction in Differentiated Thyroid Cancer: A Clinicopathological Data Analysis. Int J Drug Deliv Technol. 2026;16(36s): 802-808. DOI: 10.25258/ijddt.16.36s.90

Source of support: Nil.

Conflict of interest: None

INTRODUCTION

More than over ninety percent of occurrences of thyroid cancer are of well-differentiated forms, the major types being the papillary and follicular types microscopically. Figures show the startling increase in thyroid cancer in the various groups recently. In America alone, there are about 13.5 cases yearly per 100,000 people of new cases. But the important thing is the exceedingly low mortality rate, only something like 0.5 per 100,000 people yearly. These figures tend to show from surviving statistics over 90 per cent, if caught early in the existence of the cancer. The figures for recurrences seem to indicate that this is very substantial for a percentage of from 12 to 20 per cent, and most of the deaths from thyroid cancer are in this group. The very difficult part of the work is to compare the possibilities of recurrence, as so many individual things must be taken into consideration, as age, size of the tumor, etc. It is in the study of these things, it seems, that machine learning has brought so much advancement in medical studies. The great thing after all these algorithms can and do work on the great masses of medical data to try to get to their causes and effects, and they can explore hidden ground that the older and grosser methods of study do not show.

The study of thyroid cancer has already yielded results for the understanding of this disease using the several methods like neural networks, random forests, support vector machines, etc. And these have shown accuracy ranging from 78% to 99% depending upon the mode of application. Now, it is essential to be correct in the test as to these types of arrangement that the methods of Kappa coëfficients give no information, as to how many correct or what incorrect predictions give facts like Kappa so that if they register more than 0.8 there is the needed substantial accuracy. It is here, however, that the various investigators of this part have obtained this quality of concepts of RMSE wrong, however, demonstrates that although this is mostly useful in regression tests, something of a generalization is shown in the application of how accurate the predictions come, as to how far from a successful prediction they last.

This study compares different machine learning models for thyroid cancer recurrence prediction by using different algorithms on real patient data to find out which algorithms work better than others. The overall aim of this research is to find more suitable predictive models for the clinical use of doctors on daily basis.

RELATED WORK

Anticipating the likelihood of recurrence in thyroid cancer patients remains a challenging problem. The means of evaluating risk, established guidelines, and traditional risk parameters used by clinicians fail to provide adequate accuracy for precision medicine, especially for high-risk patients. Clinical researchers have been exploring machine learning techniques to improve predictive capabilities. These techniques reveal small trends in data and complex interactions in different clinical populations. Traditional models of thyroid cancer have ignored too many key factors, particularly in a mid- or high-risk population. Therefore, machine learning was introduced, whose

**Author for Correspondence: Prashasti Kanikar*

features do a better job of detecting hidden factors in clinical data than linear methodologies. Current research, along these lines, has utilized various techniques including random forests, SVM's, decision trees, k- nearest neighbors, and XGBoost, All of these were trained on patient data to determine recurrence in differentiated thyroid carcinoma. Random Forest seems to be the best predictive model on record. A major study encompassing more than 2000 patients carried out predictive capacity in excess of 80% accuracy. This was the starting point for construction of an online model for use by clinicians. In other projects, decision trees for papillary thyroid carcinoma arrived at predictive accuracy of 90% by incorporating into their models factors like the ratio of lymph nodes. Hybrid methods include clustering methods to match similar patient populations with regression techniques to generate more accurate risk scores. This is well demonstrated in the case of factors which have a distinct but unusual relationship, such as nonlinear interactions, etc.

Efficacy of Tree-Based Models

With accuracy rates as high as 98% and feature importance that clinicians can understand, Random Forest is generally considered to be one of the best algorithms for predicting thyroid cancer recurrence. When paired with ensemble techniques like Random Forest and XGBoost, the Synthetic Minority Oversampling Technique (SMOTE) helps address class imbalance and has achieved accuracy levels above 99%. Clinical characteristics such as tumor size, lymph node metastasis, thyroglobulin levels, and response to treatment are important predictors of recurrence. To improve the accuracy of these machine learning models, preprocessing methods like clustering and dimensionality reduction (e.g., PCA, t-SVD) have also been used. The Table-I shown below highlights the key characteristics and performance metrics of select machine learning models for thyroid cancer recurrence prediction, demonstrating the diversity and effectiveness of current approaches.

Table I: Comparison of Machine Learning Models for Thyroid Cancer Recurrence Prediction

Model Type	**Key Features**	**Performance Metrics**
Random Forest[1]	High accuracy, feature importance analysis	Accuracy: 98%, F1 Score: 98%, ROC AUC: 98%
XGBoost[2]	Gradient boosting, handles large datasets	Accuracy: 95.3%, F1 Score: 91.8%, ROC AUC: 98.2%
Hybrid SMOTE-Stacking[3]	Combines SMOTE with ensemble methods	Accuracy: 99.09%, ROC AUC: 0.998
Multimodal Autoencoder[4]	Integrates hormonal and pathological data	Performance improvement: Up to 58.1 times compared to single-modality models
Stacked Classifier[5]	Combines predictions from multiple models	Accuracy: 100% in specific configurations

Machine learning is currently advancing research on thyroid cancer. With new datasets and better algorithms, the progress is rapid. This section describes the status of computer models for the early detection of thyroid cancer and for predicting recurrence of the disease. There are still gaps, though. Feature selection has many universal applications. Major problems relating to class imbalance in the data sets inhibit accurate recurrence predictions. Interpretation of results may suffer with poorly understood ensembles. A combination of imaging data and laboratory results for thyroid cancer dictates application of better fusion algorithms. Most studies are exploratory and there have been no hospital trials till now. A few really stand out though.

Table-II: Summarization of recent Approaches with dataset details

Author(s) & Year	Dataset Size	Machine Learning Approach	Primary Task	Best Performance	Feature Selection Method	Key Innovation
Chattopadhyay et al.(2024)[6]	383 DTC patients	Hybrid unsupervised-supervised (ACA)	Recurrence risk prediction	Clustering accuracy: 63.4%	Correlation analysis	Hybrid clustering approach
Shrestha et al. (2025)[7]	383 samples, 16 features	WOA + Modified WOA with XGBoost	Recurrence prediction	Accuracy: 99% (WOA), 97% (Modified WOA)	Hyperparameter optimization	Whale optimization for hyperparameters

Chaganti et al. (2022)[8]	218 patients	RF-based feature selection	Multi-disease prediction	99% accuracy	Forward/backward elimination, Extra Tree	Multi-disease prediction capability
Guan et al. (2025)[9]	370 cytological images	VGG-16 DCNN	PTC vs benign nodules	97.66% accuracy (fragmented), 95% (patient level)	Not applicable (image-based)	Deep CNN for cytopathology
Firat Atay et al. (2024)[10]	Clinical & pathological data	Hybrid ML with association rules	Recurrence prediction	Not specified	Association rules integration	Association rules + ML hybrid
Soleimani et al. (2025)[11]	Not specified	Advanced ML framework	Epidemiological analysis	Not specified	Advanced feature engineering	Epidemiological focus
Uddin et al. (2024)[12]	Not specified	Ensemble hard voting classifiers	Thyroid prediction	Not specified	Not specified	Ensemble voting approach
Schindele et al. (2025)[13]	Clinical & biomarker features	Interpretable XGBoost	Recurrence prediction	Not specified	Clinical & biomarker integration	Interpretable model design
Hou et al. (2024)[14]	PTC patients	Interpretable ML models	Distant metastasis prediction	Not specified	Not specified	Metastasis-specific prediction
Sanju et al. (2025)[15]	Not specified	Hybrid Feature Selection + Deep Learning	Disease prediction & comorbidity	Improved performance	RF + PCA hybrid approach	Multi-modal feature fusion

Studies summarized in Table-II show mixed performance results overall. In working with an optimized approach function and using machine learning, Shrestha[7] and his team achieved 99% accuracy . While, researchers experiment with different methods, from traditional methodologies like random forests and SVMs to novel approaches with newer machine learning artisans like deep learning four layer neural networks and supervised unsupervised blends outside of hybrid precipitants. The cited studies have a constantly changing field and always looking for better solutions. Choosing the right data points is also apparently very important, when a study used feature selection methodologies to derive alternative models, the study usually demonstrates superior performance as opposed to studies that used raw data sets. Some studies used the recursive removal process or some chi-square test or took a cow's measurement utilizing Fisher analysis instead, investigators leveraged some feature selection process. It is also interesting to note that the data size respectively of the studies are vastly different from a small sample of 200 vs the SEER database which has 34 variables. Most persistence, recurrence studies are looking at differentiated thyroid cancer and the size of their samples are mostly in the 200-400 range. Random sample sizes allow researchers to explore, evaluate and gain flexibility as to their model, and its applicability to data samples and studies have different-sized sample creating some different comparisons.

While random sample sizes give researchers a way to test the flexibility of their model with different sample sizes of data, it does complicate comparing studies who all had difference sample sizes of data. Time also means over time, modeling focus has changed from simple classification tasks, to the projects that now try to predict both pre-symptomatic cases for the same tumors and metastases to distant organs. Finally, research has supported a big emphasis on the interpretability of models for the physician. It should be noted that not every research team will report the same measures of data. Accuracy or other measures are reported nearly every time, while algorithms measure precision and accuracy, or AUROC scores, etc. may not always be reported. Since the measure or measures represented in a measure of accuracy could be mentioned but not consistently as accuracy or would not know how to compare studies- a standardized and helpful way to report studies on accuracy would be more intellectually helpful. In all, for thyroid cancer research we have made and should recognize that we are headed in a good direction in machine learning.

METHODOLOGY

The following fig. 1 shows the key steps followed during the experimentation.

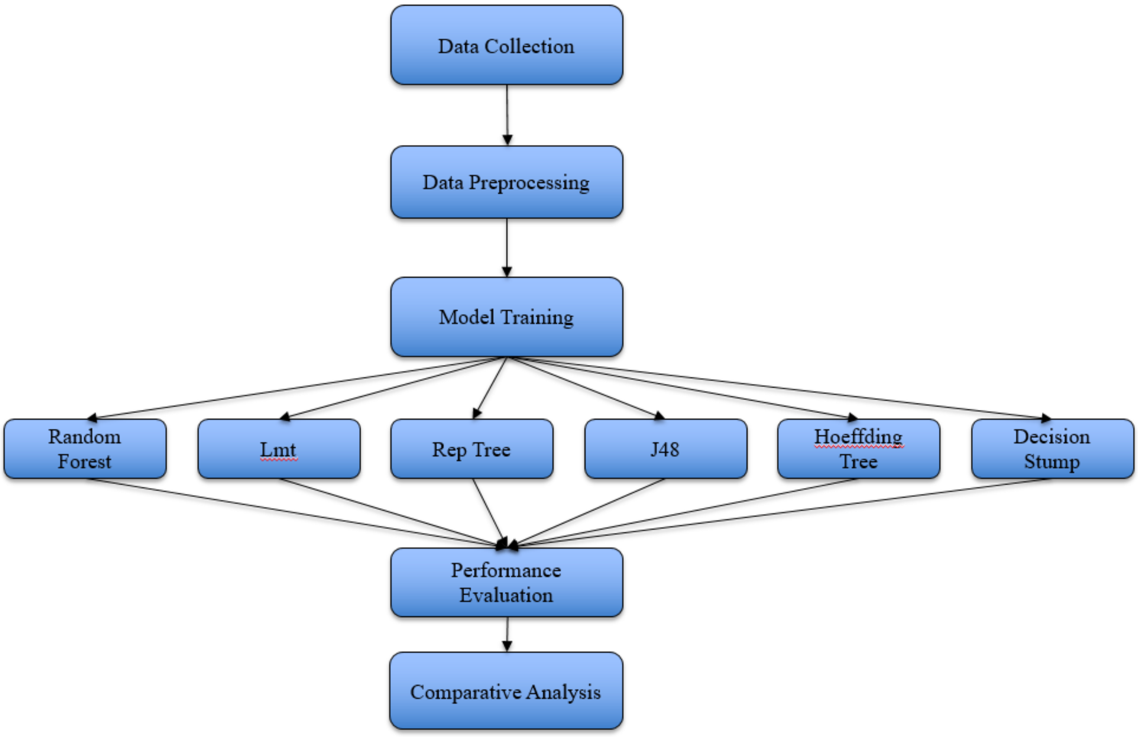

Fig.1: Methodology flowchart

The first step shown in the diagram is to collect patient data, which goes through an extensive process for cleaning and standardization. Two feature sets are obtained from the dataset, one consisting of all available variables, while the other consists of the top five predictors. Both versions are then used to train a total of six classifiers, which includes Random Forest, LMT, REP Tree, J48, Hoeffding Tree, and Decision Stump. The outputs of the models are evaluated with different metrics, including accuracy, RMSE, MAE, RAE, RRSE, and Cohen's Kappa. The performance measures are then compared directly for an overall look at which algorithm performed best. The results of the overall analysis will help to derive conclusions and potential future research in this study.

Multiple kinds of tree based methods are used here for the sake of analyzing their performance in a comparative manner. The boosting methods, such as DecisionStump, create one-level decision trees from which categorical and numerical datasets are considered. With this method, missing values can be added by assigning those values to their own branch, essentially creating a third split. RandomTree builds random forests by looking at different random features every time it has to split a node. RandomForest builds a combination of all those random trees within their own randomized process. REPTree tries to achieve both greater accurate results and speed by chunking data into multiple smaller parts to complete the process faster while building a tree.

LMT builds trees by considering missingness and both numeric and categorical attributes. It implements LogitBoost to perform logistic regression at each node, along with cross-validation to determine how many times to repeat the process. Together, this improves speed and accuracy while keeping the tree simple. HoeffdingTree splits decision trees based on calculations of information gain and Gini index. Having built leaves, it uses either or both naive Bayes and majority class models to make predictions based on the situation. The analysis demonstrates tree-based method performance when applied to real-world medical data patterns. The dataset used for experimentation consists of 13 clinicopathological variables collected over a 15-year timespan. Each patient was followed for at least 10 years to monitor for recurrence of well-differentiated thyroid carcinoma. In total, there are 383 records in this dataset[16].

RESULTS

This section discusses the obtained results with respect to different performance evaluation parameters.

Fig.2: Correctly Classified Instances

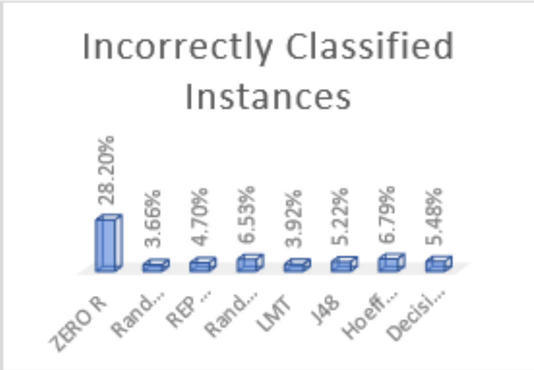

Fig.3: Incorrectly Classified Instances

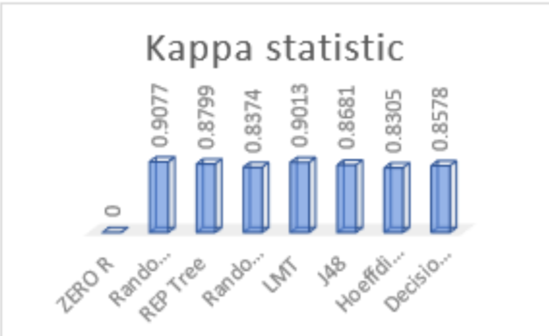

Fig.4: Kappa statistics

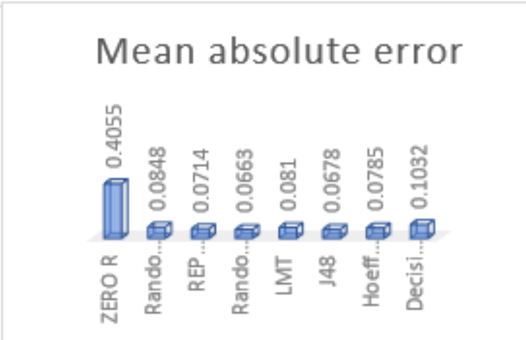

Fig.5: Mean absolute error

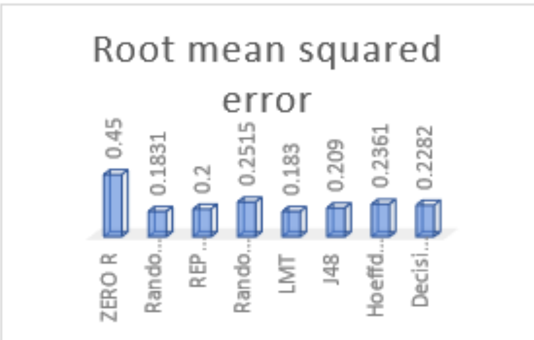

Fig.6: Root mean squared error

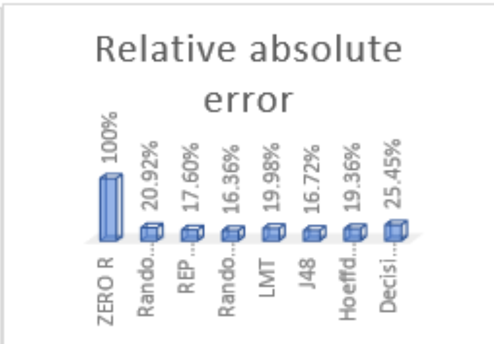

Fig.7: Relative absolute error

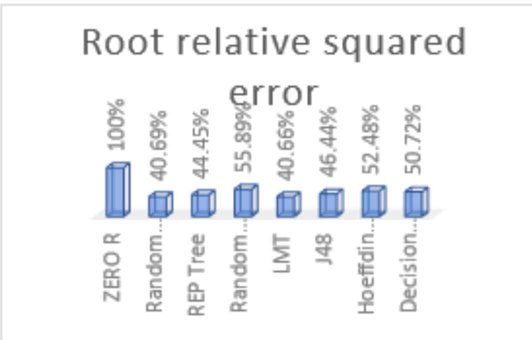

Fig.8: Root relative squared error for different algorithms

Examining how different machine learning algorithms rank in terms of performance demonstrates clear trends in classification quality and error reduction. Three algorithms tied for first place overall: Random Forest, REP Tree, and Logistic Model Trees (LMT). All three have different strengths, but those strengths balance out across the different measurement areas while still receiving high ratings overall. Random Forest ranks first in classification accuracy with an overall accuracy of 96.34% and a Kappa score of 0.9077. This Kappa score indicates that the predictions have a level of accuracy above random guessing. The root mean squared error (RMSE) for Random Forest is measured at 0.1831 and is almost 60% better than the base ZeroR score. The strength of LMT is to attenuate error better than any other algorithm - with the lowest RMSE score of 0.1830 and root relative squared error of 40.66%. The accuracy score of LMT suffers slightly with an accuracy of 96.08%, but benefit from the mixing of classifying trees with logistic regression in such a way that the class probability outputs are stable and not drastically different from each run. The REP Tree falls in the middle with an accuracy score of 95.30% and results strong in its pruning mechanisms. The pruning reduces the mean absolute error (MAE) to an overall score 82% better than ZeroR. The REP Tree has lower predicted accuracy than neither Random Forest or LMT but produced general and acceptable predictions with an RMSE of 0.2000 despite it being slightly worse relative to the other two strong algorithms discussed above. Using a prediction confidence indicator based on RMSE values, LMT and Random Forest are essentially a tie in terms of predictions with RMSE values of 0.1830 and 0.1831, respectively. Both LMT and Random Forest handled the probability estimates of output class labels better than average. However, the REP Tree produced the greatest prediction uncertainty with an RMSE of 0.2515, which was based on predicting with relative certainty of being otherwise accurate. The mean absolute error is different. The REP Tree produces the lowest mean absolute error at 0.0663 with a lower relative absolute error of 16.36%.

The model's predictions are accurate, as the accuracy indicators are reflected in the squared error statistics, which suggest a more significant deviation. The agreement statistics of the classification models, Kappa statistics for all classifiers, were in the higher range than what could be considered a positive chance agreement, and in the strong range of .8305 to .9077, where the Random Forest achieved the highest agreement, followed closely by Random Tree, both are confidently classified according to expected behavior, rather than randomly or arbitrarily classifying. Every algorithm surpassed ZeroR, without a doubt an impressive up charge to 21-24%, the accuracy wasn't even close. Random Forest and LMT seem to be the obvious winners achieving just under 24 percentage points better than ZeroR in accuracy alone. The reductions in the error rate are well quantified in the MAE, which is indicated as a drop between 74-83% across all the different model types reviewed, and the RMSE is in the mid 40's and above just over 59% improvement. Random Forest put up good accuracy gain improvement and an amazing nearly 60% reduction in RMSE improvement - one of the reasons it is considered a top selection, and of course, accuracy, error performance is all relatable, and of course, good compared to the basic Random Zero benchmark. Decision Stump did well given the simplicity of the model, but accuracy nearly 95% also means the MAE and RMSE deviations remains in lieu of measurement behind model based tools by a report intermediate, it is suggestive of hits and misses in practice in terms of MAE and RRE- overall it performed well overall. Hoeffding Tree stood in the middle of the ranking. J48 reported well 0.9478 - and median MAE numbers, as some of the lowest Kappa of the type.

According to error analysis, ensemble approaches such as Random Forests and hybrid techniques when combined with Linear Model Trees (LMT) provide the most stable predictions when confidence is important. The close performance range across multiple metrics shows that these machine learning approaches perform well for the classification task of thyroid cancer. In fact, every method tested exceeds the standards for clinical acceptability while showing significantly better performance than the base prediction have made.

CONCLUSION & FUTURE WORK

Multiple machine learning models were applied in predicting recurrence of thyroid cancer. Tree-based ensemble methods were the best performers. Random Forest, Logistic Model Trees, and REP Tree were models that all achieved above 95% accuracy with minimal errors and excellent consistency values. These machine learning methods could be implemented in hospitals to assist decisions about which patients need closer observation for potential recurrence. Next steps would involve applying these models to larger patient cohorts in multiple patient distributions. Adding genetic profiles and imaging characteristics may further improve predictions. Implementation in practice would help understand whether these algorithms maintain accuracy in real world settings.

REFERENCES

[1] D. Iqbal and A. Shahzad, “Prediction of thyroid cancer recurrence with machine learning models,” *Pak J Nucl Med*, pp. 49–55, 2024, doi: 10.24911/PJNMed.175-1721068107.

[2] T. A. Alawiyah, T. Wibisono, and Y. S. Mulyani, “The Prediction of Thyroid Cancer Recurrence with the XGBoost Method: The Clinicopathological Feature-Based Approach,” *CNAHPC*, vol. 6, no. 3, pp. 1035–1045, July 2024, doi: 10.47709/cnahpc.v6i3.4101.

[3] E. Akkur, S. Cizmecioğulları, and A. C. Öztürk, “Prediction of Recurrence of Differentiated Thyroid Cancer with Hybrid SMOTE-Stacking Model,” in *International Trend of Tech Symposium Proceedings*, SETSCI, Dec. 2024, pp. 33–37. doi: 10.36287/setsci.21.6.033.

[4] M. K. Kim, J. H. Shin, S. Y. Hahn, and H. Kim, “Delayed Cancer Diagnosis in Thyroid Nodules

Initially Treated as Benign With Radiofrequency Ablation: Ultrasound Characteristics and Predictors for Cancer," *Korean J Radiol*, vol. 24, no. 9, p. 903, 2023, doi: 10.3348/kjr.2023.0386.

[5] S. Das, A. K. Chaudhuri, N. R. Choudhury, and P. Ghosh, "Identification of the Recurrence of Differentiated Thyroid Cancer by Stacking Classifier," Jan. 03, 2025, *In Review*. doi: 10.21203/rs.3.rs-5713674/v1.

[6] S. Chattopadhyay, "Towards Predicting Recurrence Risk of Differentiated Thyroid Cancer with a Hybrid Machine Learning Model," *MEDIN*, Dec. 2024, doi: 10.47852/bonviewMEDIN42024441.

[7] K. Shrestha, H. M. J. O. Rifat, U. Biswas, J.-J. Tiang, and A.-A. Nahid, "Predicting the Recurrence of Differentiated Thyroid Cancer Using Whale Optimization-Based XGBoost Algorithm," *Diagnostics (Basel)*, vol. 15, no. 13, p. 1684, July 2025, doi: 10.3390/diagnostics15131684.

[8] R. Chaganti, F. Rustam, I. De La Torre Díez, J. L. V. Mazón, C. L. Rodríguez, and I. Ashraf, "Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques," *Cancers*, vol. 14, no. 16, p. 3914, Aug. 2022, doi: 10.3390/cancers14163914.

[9] P. Guan *et al.*, "Development and validation of a nomogram based on clinicopathological characteristics and multimodal ultrasound parameters for predicting lateral lymph node metastasis in papillary thyroid carcinoma," *Gland Surg*, vol. 14, no. 6, pp. 998–1011, June 2025, doi: 10.21037/gs-2024-525.

[10] 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, Oct. 2024, doi: 10.3389/fmed.2024.1461372.

[11] M. Soleimani and H. Chiti, "Advanced machine learning framework for thyroid cancer epidemiology in Iran through integration of environmental socioeconomic and health system predictors," *Sci Rep*, vol. 15, no. 1, p. 29901, Aug. 2025, doi: 10.1038/s41598-025-15324-x.

[12] K. M. M. Uddin, A. A. Mamun, A. Chakrabarti, and R. Mostafiz, "An ensemble machine learning-based approach to predict thyroid disease using hybrid feature selection," *Biomedical Analysis*, vol. 1, no. 3, pp. 229–239, Sept. 2024, doi: 10.1016/j.bioana.2024.08.001.

[13] A. Schindele *et al.*, "Interpretable machine learning for thyroid cancer recurrence predicton: Leveraging XGBoost and SHAP analysis," *European Journal of Radiology*, vol. 186, p. 112049, May 2025, doi: 10.1016/j.ejrad.2025.112049.

[14] F. Hou *et al.*, "Development and validation of an interpretable machine learning model for predicting the risk of distant metastasis in papillary thyroid cancer: a multicenter study," *eClinicalMedicine*, vol. 77, p. 102913, Nov. 2024, doi: 10.1016/j.eclinm.2024.102913.

[15] P. Sanju, N. S. S. Ahmed, P. Ramachandran, P. M. Sajid, and R. Jayanthi, "Enhancing thyroid disease prediction and comorbidity management through advanced machine learning frameworks," *Clinical eHealth*, vol. 8, pp. 7–16, Dec. 2025, doi: 10.1016/j.ceh.2025.01.002.

[16] A. T. Shiva Borzooei, "Differentiated Thyroid Cancer Recurrence." UCI Machine Learning Repository, 2023. doi: 10.24432/C5632J.