

Predicting Thyroid Disease Treatment Outcomes Using Machine Learning Approaches

NEHABANU H HARLAPUR

Professor, RTE Society's Rural Engineering College Hulkoti, Karnataka

Abstract— Thyroid diseases, including hypothyroidism, hyperthyroidism, and autoimmune thyroid disorders, require careful management and personalized treatment strategies. In this study, we explore the use of machine learning (ML) models to predict treatment outcomes for thyroid diseases based on clinical and diagnostic data. By applying various ML algorithms such as KNN (K-Nearest Neighbors), Logistic Regression, Random Forest, decision trees, support vector machines (SVM) and XGBoost neural networks to patient records, we aim to identify the most effective treatments for patients, improve decision-making processes, and reduce trial-and-error approaches in thyroid disease management. The findings demonstrate that machine learning can play a significant role in enhancing treatment planning and optimizing outcomes for patients with thyroid disorders. The paper presents the findings of the machine learning models, comparing the prediction accuracy and other performance metrics. It may show how models like SVM or random forests outperform traditional statistical models or expert-based predictions.

Indexed Terms— Machine learning, classification model, Thyroid diseases, KNN (K-Nearest Neighbors), Logistic Regression, Random Forest, Decision trees, Support vector machines (SVM) and XGBoost

I. INTRODUCTION

The thyroid is a vital endocrine gland located in the neck, responsible for producing hormones—primarily FT3 (triiodothyronine) and FT4 (thyroxine)—that regulate various body functions, including metabolism, heart rate, and body temperature. These hormones play a crucial role in determining how the body uses and consumes nutrients. When the thyroid functions abnormally, it can result in conditions like hyperthyroidism (overproduction of thyroid hormones) or hypothyroidism (underproduction of thyroid hormones), both of which can cause significant health issues. In addition to these dysfunctions, the thyroid can experience inflammation (thyroiditis) or enlargement due to nodules or goiters. Some nodules may even become malignant, leading to

thyroid cancer. Because of the complexity and potential severity of thyroid conditions, effective treatment is critical for patient health.

Levothyroxine Treatment for Thyroid Disorders: One of the most commonly prescribed treatments for thyroid diseases is levothyroxine (LT4), a synthetic form of the thyroid hormone used primarily for hypothyroidism. Levothyroxine is also used in various other scenarios, such as:

- Treatment of goiter.
- Prophylaxis to prevent relapses after the total or partial removal of a goiter.
- Hormone replacement therapy in cases of thyroid hypofunction.
- Treatment of thyroid inflammation.
- Therapy after treatment with antithyroid drugs.
- Post-surgery for malignant thyroid tumors to suppress tumor growth and compensate for hormone deficiencies.

However, determining the right dosage of levothyroxine is not straightforward. The dose varies widely depending on several factors, including the patient's residual thyroid function, body weight, and thyroid-stimulating hormone (TSH) levels. It is crucial to tailor the dose to the individual's needs, adjusting it over time as the patient's physiology and medical condition change (e.g., weight changes or pregnancy). As such, continuous monitoring and adjustment of the treatment based on clinical and laboratory assessments are necessary.

Machine Learning in Thyroid Disease Treatment Prediction. Given the complexity of managing thyroid diseases and the need for constant adjustments in treatment, machine learning (ML) presents a promising approach to support endocrinologists in monitoring and predicting treatment outcomes. Recent studies have demonstrated the effectiveness of ML

techniques in various medical domains, such as heart disease, diabetes, and Parkinson's disease. These techniques can significantly reduce the time and cost required for diagnosing and treating patients.

II. LITERATURE SURVEY

Umar Sidiq, Dr, Syed Mutahar Aaqib, and Rafi Ahmad Khan [15] Classification, which is used to characterize predefined data sets, is one of the most popular supervised learning data mining techniques. The information for this study was gathered from a well-known Kashmiri laboratory. The entire research project will be conducted on the ANACONDA3-5.2.0 platform. In an experimental analysis, classification methods such as k nearest neighbors, Support vector machine, Decision tree, and Nave bayes may be used. The Judgment Tree has the greatest accuracy of the other classes, at 98.89 percent.

Sindhya, Mrs K [16] Thyroid disorder is a chronic illness that affects people all over the world. Data mining in healthcare is producing excellent results in the prediction of different diseases. The accuracy of data mining techniques for prediction is high, and the cost of prediction is low. Another significant benefit is that prediction takes very little time. In this study, I used classification algorithms to analyze thyroid data and came up with a result. A model's efficacy is primarily determined by two factors. The first is prediction precision, and the second is prediction time. According to our findings, Nave Bayes took just 0.04 seconds to forecast. However, it is less accurate than J48 and Random Forest.

III. METHODOLOGY

A. Data Collection

This dataset contains 13 clinicopathologic features aiming to predict the recurrence of well-differentiated thyroid cancer. The data was collected over 15 years, and each patient was followed for at least 10 years.

- i. Age: The age of the patient at the time of diagnosis or treatment.
- ii. Gender: The gender of the patient.
- iii. Smoking: Whether the patient is a smoker or not.
- iv. Hx Smoking: Smoking history of the patient.
- v. Hx Radiotherapy: History of radiotherapy treatment for any condition.

- vi. Thyroid function: The status of thyroid function, possibly indicating any abnormalities.
- vii. Physical examination: Findings from a physical examination of the patient, which may include palpation of the thyroid gland and surrounding structures.
- viii. Adenopathy: Presence or absence of enlarged lymph nodes in the neck region.
- ix. Pathology: Specific types of thyroid cancer as determined by pathology examination of biopsy samples.
- x. Focality: Whether the cancer is unifocal (limited to one location) or multifocal (present in multiple locations).
- xi. Risk: Risk of cancer.
- xii. T: Tumor classification based on the size and extent of invasion into nearby structures.
- xiii. N: Nodal classification indicating the involvement of lymph nodes.
- xiv. M: Metastasis classification indicating the presence or absence of distant metastases.
- xv. Stage: The overall stage of the cancer.
- xvi. Response: Response to treatment, indicating whether the cancer responds positively or remains stable after treatment.
- xvii. Recurred: Indicates whether the cancer has recurred after initial treatment (No = 0, Yes = 1).

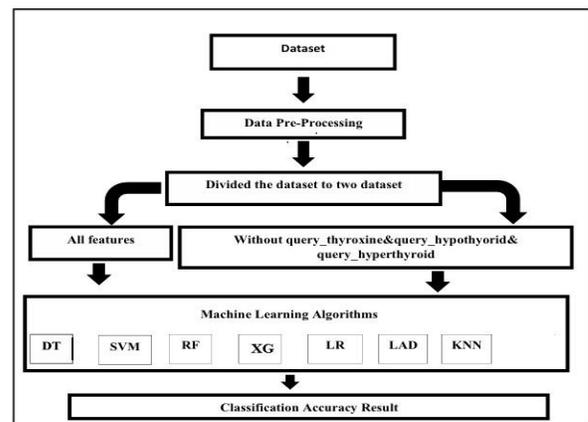
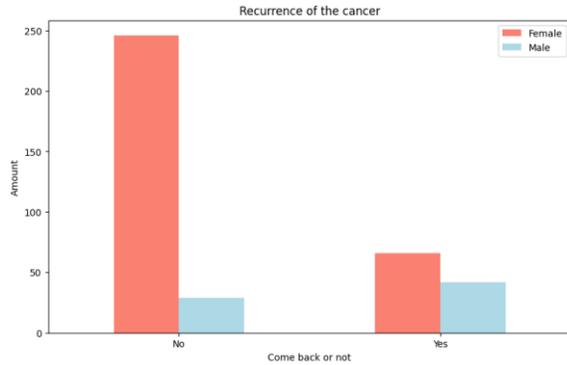


Figure1: Shows how data is entered and the operations that take place

B. Data Exploration

Gender Analysis of Thyroid Cancer Recurrence Visualization:

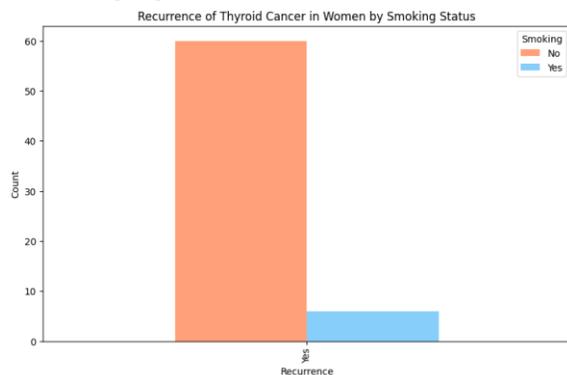
The following graph illustrates the distribution of thyroid cancer recurrence by gender:



By analyzing the data, we can observe that thyroid cancer recurrence is more prevalent among female patients compared to male patients.

Impact of Smoking on Thyroid Cancer Recurrence

Visualization: The following graph illustrates the distribution of thyroid cancer recurrence by smoking status and gender:

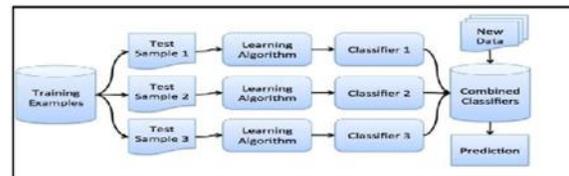


C. Modelling

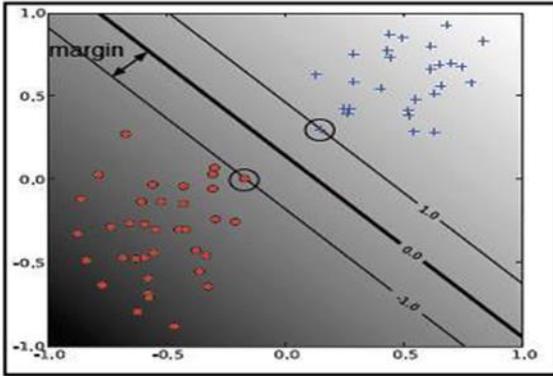
In this project we are going to implement six different classification models:

1. KNN (K-Nearest Neighbors)
 - KNN is a simple and versatile algorithm for classification tasks
 - It can capture complex patterns in the data
2. Logistic Regression
 - Logistic Regression is a classic and widely used model for binary classification problems
 - It provides a probabilistic interpretation of the relationship between independent variables and the dependent variable
3. Random Forest
 - Random Forest is an ensemble algorithm that combines multiple decision trees to improve accuracy and reduce overfitting. The random forest computes the mean response of every predictor for

energy consumption. Then, for each sample, a random forest adds the absolute distance each response was from the mean of each predictor for a total sum of the distance that each answer was from the means of the data. A high distance value will signify individuals who were consistently far away from the mean response in each sample. Detecting rates who repeatedly classify the samples was simple--a function that calculated the mode of each response was used. If the mode of a response was over 90% of the total number of questions, the research marked the response as potentially high in energy consumption. There are many responses marked. It was clear from a visual examination of these responses that the individuals had sampled with the same response.



- It is robust against outliers and missing data
4. Support Vector Machine (SVM)
 - SVM has good generalization capabilities and is resistant to overfitting. The support vector machine (SVM) is a machine learning and data mining algorithm to determine the strongest predictors of this variable for energy consumption. The research used popular classification methods to answer our question: best subset selection, boosting trees, and generalized additive models. Our first approach was to use forward, backward and best subset selection to obtain a subset of predictors that most strongly predicted consumption with a linear relationship. The SVM provided an approach that was to use a tree-based method to stratify the predictor space into sample regions using recursive binary splitting. The research decided to use the boosting tree method as this is known to be one of the most powerful tree based models. SVM also has a good ability to deal with high dimensionality data.



5. Decision Trees

- They can handle numerical and categorical data without extensive preprocessing

6. XGBoost

- XGBoost is an optimized implementation of gradient-boosting algorithms
- It provides high accuracy and speed in predictions

IV. VALIDATION

1. Accuracy

Accuracy measures the overall correctness of the model, representing the proportion of total correct predictions (both true positives and true negatives) out of all predictions made. The formula for accuracy is: $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

2. Precision

Precision focuses on the accuracy of positive predictions. It measures how many of the instances that the model predicted as positive are actually positive. It is defined as:

$$Precision = \frac{TP}{TP+FP}$$

3. Recall

Recall, also known as sensitivity or true positive rate, measures the ability of the model to identify all relevant positive instances. It is defined as:

$$Recall = \frac{TP}{TP+FN}$$

4. F-score (or F1-score)

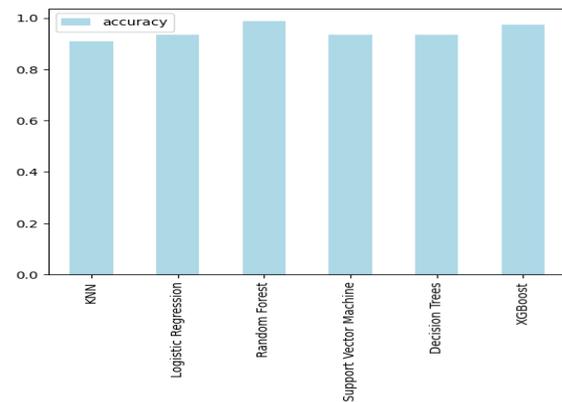
The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall, especially when you need a balance between the two, rather than favoring one over the other. The formula for F1-score is: $F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

V. RESULTS

4.1 Model Comparison

Modelo	Precisión
KNN	0.9091
Logistic Regression	0.9351
Random Forest	0.987
Support Vector Machine	0.9351
Decision Trees	0.9351
XGBoost	0.974

Model Comparison

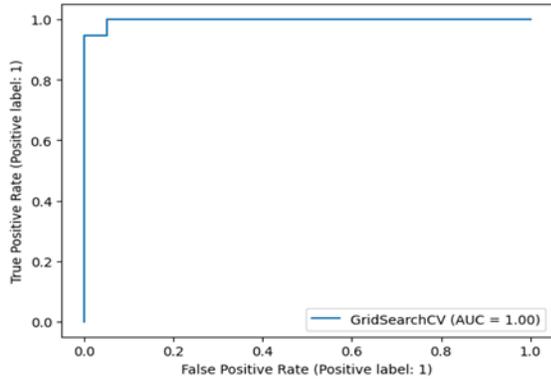


4.2 ROC Curve and AUC Scores

Receiving Operating Characteristic Curve(ROC)

acts as a probability curve that depicts two parameters at different threshold values: the false positive rate (FPR) on the horizontal axis and the true positive rate (TPR), or known as a recall on the vertical axis. Models that perform well possess ROC curves that sketch toward the top left corner. From the ROC curve, the Area Under the Curve (AUC), a measurement of a classifier’s capacity to distinguish between classes can be obtained. The AUC value is between 0 and 1, where $AUC = 1$ indicates that the model can accurately identify both the positive and the negative class points. On the contrary, when the AUC value equals 0, the model will give the wrong prediction: all positives will be negative, and vice versa. Meanwhile, the classifier is unable to distinguish between positive and negative class points when AUC is 0.5. Hence, it can be concluded that the higher the AUC value, the greater the prediction ability of the model.

$$FPR = \frac{FP}{TN+FP}$$

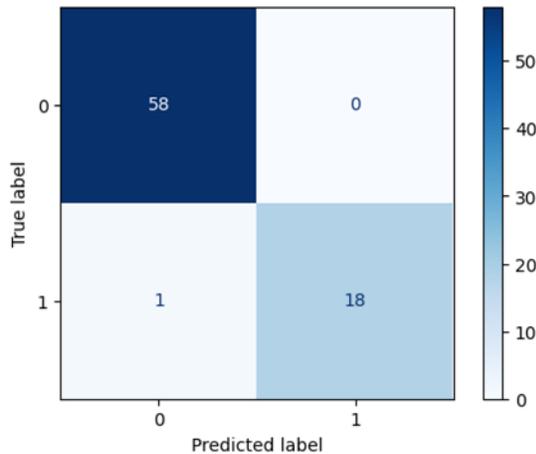


An AUC (Area Under the Curve) of 1.0 indicates that the model has perfect performance in distinguishing between positive and negative classes. This means that the model correctly classifies all instances without making any errors. However, it can be an indication of overfitting. This occurs when the model has learned the characteristics of the training dataset too well, including the noise, and may not generalize well to new and unseen data.

macro avg	0.99	0.97	0.98	77
weighted avg	0.99	0.99	0.99	77

- **High Precision and Recall:** The model has high precision and recall for both classes, with particularly notable perfect precision for the positive class (1.00) and perfect recall for the negative class (1.00). This indicates that the model is very reliable in its predictions and rarely makes errors.
- **High F1-Score:** The high F1-scores for both classes (0.99 for the negative class and 0.97 for the positive class) indicate that the model maintains a good balance between precision and recall.
- **Consistent Performance:** The macro and weighted averages indicate consistent and solid performance across both classes, which is especially important in cases of class imbalance.

4.3 Confusion matrix

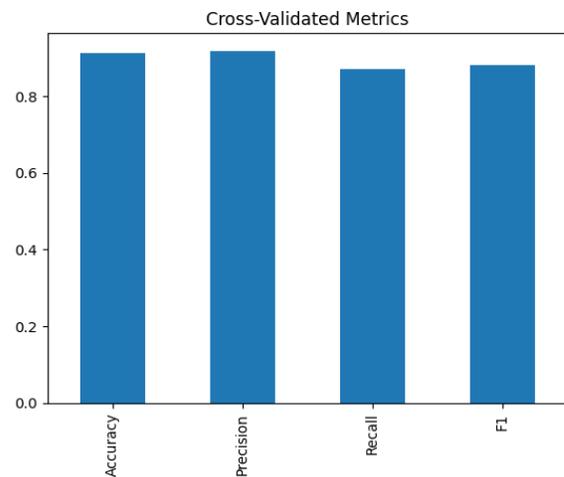


Evaluates classification accuracy by computing the confusion matrix with each row corresponding to the true class

4.4 Classification report

	precision	recall	f1-score	support
0	0.98	1.00	0.99	58
1	1.00	0.95	0.97	19
accuracy			0.99	77

4.5 Cross validation metrics



CONCLUSION

Thyroid disease is one of the diseases that afflict the world's population, and the number of cases of this disease is increasing. Because of medical reports that show serious imbalances in thyroid diseases, our study deals with the classification of thyroid disease between hyperthyroidism and hypothyroidism. This disease was classified using algorithms. Machine learning

showed us good results .By applying various ML algorithms such as KNN (K-Nearest Neighbors), Logistic Regression, Random Forest, decision trees, support vector machines (SVM) and XGBoost neural networks. For the current model, a minimum precision of 95% was sought due to its medical use, achieving a precision of 99% with the XGBoost model trained after hyperparameter analysis. In the first model, all the characteristics consisting of 16 inputs and one output were taken, and the result of the accuracy of the random forest algorithm was 98.93, which is the highest accuracy among the other algorithms.

REFERENCES

(Periodicalstyle)

- [1] Azar, a.T, Hassanien, A.E. and Kim, T. Expert system based on neural fuzzy rules for thyroid diseases diagnosis, Computer Science, Artificial Intelligence, arXiv:1403.0522, Pp. 1-12,2012.
- [2] Keles, A. ESTDD: Expert system for thyroid diseases diagnosis, Expert Syst Appl., Vol. 34, No.1, Pp.242–246,2008.
- [3] a. c.c.Heuck, "World Health Organization," 2000. [Online]. Available: <https://www.who.int/>.
- [4] Kouroua, K., Exarchosa, T.P. Exarchosa, K.P., Karamouzisc, M.V. andFotiadisa, D.I. (2015) Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal, Vol. 13, Pp.8–17.
- [5] Shukla, A. & Kaur, P. (2009). Diagnosis of thyroid disorders using artificial neural networks, IEEE International Advance computing Conference (IACC 2009)– Patiala, India, pp 1016-1020.
- [6] Aswad, Salma Abdullah, and Emrullah Sonuç. "Classification of VPN Network Traffic Flow Using Time Related Features on Apache Spark." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.
- [7] Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
- [8] Chandio, Jamil Ahmed, et al. "TDV: Intelligent system for thyroid disease visualization." 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). IEEE, 2016.
- [9] Travis B Murdoch and Allan S Detsky. The inevitable application of big data to health care. *Jama*, 309(13):1351–1352, 2013.
- [10] Dr. Srinivasan B, Pavya K "Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue: 11 | Nov – 2016
- [11] Aytürk Keleş and Keleş, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov -2017 34.1 (2017): 242- 246
- [12] Khushboo Taneja, Parveen Sehgal, Prerana "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology (E-ISSN: 2321- 3264) Vol. 3, No. 2, April 2016
- [13] Chandel, Khushboo, et al. "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques." *CSI transactions on ICT* 4.2-4 (2016): 313-319.
- [14] Banu, G. Rasitha. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
- [15] Umar Sidiq, Dr, Syed Mutahar Aaqib, and Rafi Ahmad Khan. "Diagnosis of various thyroid ailments using data mining classification techniques." *Int J Sci Res Coput Sci Inf Technol* 5 (2019): 131-6.
- [16] Sindhya, Mrs K. "EFFECTIVE PREDICTION OF HYPOTHYROID USING VARIOUS DATA MINING TECHNIQUES."
- [17] AKGÜL, Göksu, et al. "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı." *Bilişim Teknolojileri Dergisi* 13.3 (2020): 255-268.

- [18] VijayaKumar, K., et al. "Random Forest Algorithm for the Prediction of Diabetes." 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2019.
- [19] Chaurasia, Vikas, Saurabh Pal, and B. B. Tiwari. "Prediction of benign and malignant breast cancer using data mining techniques." *Journal of Algorithms & Computational Technology* 12.2 (2018): 119-126.
- [20] Begum, Amina, and A. Parkavi. "Prediction of thyroid disease using data mining techniques." 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019.
- [21] C. Fan, F. Xiao, Z. Li, J. Wang. Unsupervised data analytics in mining big building operational data for energy efficiency enhancement: A review. *Energy Build.* 2018, 159, 296–308.