Multiclass prediction of thyroid disease using machine learning models by applying BOO_ST balancing technique

Mrs. Geetha A M
Assistant Professor
Dept of Computer Science
Surana College - Autonomous
Bengaluru 560004

Ms. Hemavathi K
Research Scholar, MCA Department
Community Institute of Management Studies
Bengaluru 560004

ABSTRACT

A vital component of society is healthcare, which guarantees that patients receive the required care in a timely manner. One illness that is progressively affecting people all around the world is thyroid disease. In order to detect thyroid disease, this paper presents a machine learning framework that can be applied to precisely process data obtained from medical sector to diagnose thyroid illness. In this investigation, several machine learning and deep learning models are employed to forecast thyroid disease. This machine learning framework is trained with thyroidDF dataset obtained from Kaggle repository and applied for diagnosing multiple thyroid illness such as hypothyroid, hyperthyroid, subclinical hypothyroid, subclinical hyperthyroid, miscellaneous and general health, treatment and post treatment condition. The model employs BOO_ST technique primarily for data preprocessing and balancing. XGBOOST classifier showed better performance compared to other seven classifiers with highest accuracy of 98.23%. The proposed model helps inexperienced clinicians to accurately diagnose thyroid patients with correct classification of thyroid illness.

KEYWORDS: Machine Learning, Deep Learning, Thyroid illness, XGBOOST, multiclass classification

INTRODUCTION

Thyroid endocrine gland resembling butterfly shape is located at the front of the neck. Thyroid hormones produced by the thyroid gland is discharged into the bloodstream, where it is absorbed by all bodily tissues. The body needs thyroid hormones to use energy, staying warm, and to maintain proper functioning of heart, brain, muscles, and many other body organs (www.thyroid.org). Thyroxine (T4) hormone released by thyroid gland is then transformed into Triiodothyronine hormone (T3). Thyroid-stimulating hormones (TSH), produced in pituitary gland regulates the quantity of T4 hormone to be produced by the thyroid gland. Thyroid illnesses occurs when abnormal quantity of T3 and T4 hormone is released due to the dysfunction of thyroid gland. According to ChenLing, LiXue et al,[1], thyroid illnesses can cause depression, raised cholesterol, elevated blood pressure, cardiovascular problems, and impaired fertility. Monaco Fabrizio [2], the study used machine learning (ML) models to classify thyroid illnesses into three classes namely hypothyroid, hyperthyroid and euthyroid. Excessive thyroid hormone circulation results in hyperthyroidism, a condition that is frequently linked to illnesses such as goiter, toxic multinodular and Graves' disease. Conversely, hypothyroidism results from insufficient hormone production of thyroid. Frequent causes of hypothyroid include the deficiency of iodine and Hashimoto's thyroiditis. Ionita et al,[3] study showed abnormal growth of thyroid gland in goiter and in thyroid nodules. The thyroid lumps in these cases are classified as benign or autoimmune reactions or infections or malignancy. Thyroid cancer is an inflammation of the thyroid gland: a condition in which abnormal cells proliferate within the thyroid gland. For healthcare practitioners, accurately diagnosing these disorders is crucial. SM Gorade, A Deo, et al,[4], blood tests and clinical symptom assessment are crucial for diagnosing and treating thyroid disorders. A growing number of healthcare facilities are using ML models to detect illnesses more quickly and precisely and also in decision-making and lowering mortality rates. These ML models can manage massive volumes of data. Bichler M et al,[5] describes how to apply seven ML algorithms for binary

classification of thyroid disease using K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector (SVM), Decision Tree (DT), and Logistic Regression (LR) Classifier, Extreme Gradient Boost (XGBOOST), and Naive Bayes (NB) for categorizing the type of thyroid diseases (TD).

LITERATURE REVIEW

Alshanbari et al.,[6] used dynamic selection hybrid model to improve clinical decision-making. Though the study proposed an accuracy of 99.33% in predicting three classification of thyroid disease, it demands high computational resources and incurs huge computational costs compared to conventional single classifiers. The study Obaido et al, [7] achieves 84% of accuracy using the stacking ensemble feature selection method on a dataset containing 1232 samples. However, the study has limitations in addressing class data imbalance and capturing complex patterns. DT, RF, NB, XGBOOST, LR, KNN, SVM.

Ref	Target Classes	Data Size	Best ML Classifier	Research Gaps
[6]	3	3772	DSHM	Limited dataset
[7]	3	1232	Stacked Ensemble	Limited dataset
[8]	3	3221	RF	Limited dataset
[9]	3	1464	Bagging	Limited dataset
[10]	5	9173	RF	Performed on limited dataset
[11]	5	9172	DT	Used limited dataset to perform
[12]	10	9172	AdaBoost	Performed on limited dataset
[13]	4	3772	DT	Limited dataset
[14]	2	-	Stacking	Binary class
[15]	3	7200	RF	Imbalanced dataset
[16]	3	7200	XGBOOST	Limited dataset
[17]	2	3076	RF	Limited dataset
[18]	3	3163	DT	Limited dataset

Shama et al [8] explored various ML algorithms on a UCI dataset containing 3221 samples for predicting three categories of thyroid disease and achieved an accuracy of 91.42%. Mir and Mittal [9] used SVM, NB, J48, Bagging, Boosting classifiers on the data collected from 1464 Indian patients. Bagging technique emerged as the top performer achieving 98.56% accuracy. However the study considered a limited dataset and ML model could classify only three categories of thyroid illness. Chaganti et al,[10] proposed ML model with RF, GBM, ADA, LR, SVM algorithms on a dataset consisting of 9173 patient samples. With 6771 samples belonging to normal patients not showing any sign of thyroid illness, an accuracy of 98% was achieved using RF. The proposed model predicted primary hypothyroid, increased binding protein, concurrent non-thyroidal illness and autoimmune thyroiditis.

Quaranti and Taleb [11] integrated ontology for interpretability with the DT classifier. The ontology integration classifier provided better insights in decision-making process of the ML algorithm. ML classifier was applied on 9172 observations with 31 features and was downloaded from Kaggle. Based on the findings, ontology integrated model outperformed decision trees. Gupta et al.,[12] focused on developing an optimized ML algorithm based on differential evolution to classify different types of hypothyroid disease. To predict hypothyroid illness, Guleria et al.,[13] used ML and deep learning (DL) algorithms such as NB, DT, Artificial Neural Network (ANN), RF and multiple class classification. Data repositories from WEKA simulator's contained limited samples of 3772 hypothyroidism-related records, each record containing 30 features. For compensated, negative, and primary hypothyroidism, the RF classifier demonstrated higher prediction accuracy of 99.71%, 97.93%, and 89.47%, respectively.The study [14] employed decision trees, random forest, XGBoost, Light GBM, extra trees, gradient boosting, and a stacking ensemble model for early diagnosis prediction of hypothyroidism disease using feature selection and explainable Artificial Intelligence (AI). The result demonstrated 99.16% accuracy using stacking model on the UCI dataset containing 30 features. The model predicts binary classification of normal and hypothyroid. The study by Shahid et al,[15] uses UCI thyroid disease dataset and ML classifiers like RF, SVM, and KNN in diagnosing thyroid disease. The results demonstrates RF classifier(98.50%) performs better than SVM (97.02%) and KNN (95.81%) for classification of thyroid disease into three classes namely hypothyroid, hyperthyroid, and normal. Mona et al,[16] focused on optimized gradient Boosting techniques to classify hyper, hypo

and normal TD on UCI dataset containing 7200 instances and 21 features. Okuboyejo et al,[17] employed 7 ML algorithms including RF, DT, KNN, SVM, ANN, XGBOOST, NB to predict hypothyroidism and euthyroid of a dataset containing 3076 records with 30 attributes. Almahsh et al.,[18] proposed another model using SVM, DT, NB and ensemble model to detect hypothyroidism disease with accuracy of 97.6% was obtained on UCI dataset containing 3163 samples.

RESEARCH METHODOLOGY

Fig (1) illustrates the research methodology used for multiclass TD prediction. The dataset thyroidDF.csv is acquired from Kaggle repository. This dataset consists of 9172 records and each record is represented with 31 features.

SL NO	FEATURE	DESCRIPTION
1	Age	patient age
2	sex	patient sex identification
3	thyroxine	patient is on thyroxine?
4	query_on_thyroxine	check if patient is on thyroxine?
5	onantithyroid_meds	patient is on anthithyroid medication?
6	Sick	is patient sick
7	pregnant	is patient pregnant
8	thyroid_surgery	has patient underwenr thyroid surgery?
9	I131_treatment	has patient undergone I131 treatment?
10	query_hypothyroid	whether patient has hypothyroid?
11	query_hyperthyroid	whether patient has hyperthyroid?
12	lithium	whether patient has lithium?
13	goitre	if patient has goitre?
14	tumor	if patient has tumor?
15	hypopitutary	whether patient hypopitutary gland is normal?
16	psych	whether patient is psych?
17	TSHmeasured	if TSH measured in the blood?
18	TSH	TSH level in blood
19	T3measured	whether T3 measured in the blood?
20	T3	T3 level in blood
21	TT4measured	whether T4 was measured in the blood?
22	TT4	TT4 level in blood
23	TT4Umeasured	whether TT4U measured in the blood?
24	T4U	T4U level in blood
25	FTI measured	FTI measured in the blood
26	FTI	FTI level in blood
27	TBGmeasured	whether TBG measured in blood?
28	TBG	level of TBG in blood
29	referral source	source name
30	binaryclass	target value positive or negative

Significant features are obtained by performing feature based analysis. The age group of patients is between 1 – 97 years. The feature pregnancy does not allow to fill the missing age feature using statistical methods and hence instances with null as the age is dropped from the dataset, remaining features in the dataset having null values are filled with the mean value of the column. referral_source and patient_id features don't have any implication on target classes, hence they are dropped from the dataset. After preprocessing there are 8865 samples in the dataset. The target feature contains seven classes.

The dataset categorization in Table1 shows the total number of samples for each category of target class. From the Table1, it is clear that the proposed dataset is highly imbalanced. Out of 8865 patient samples, 6559 samples are having target class as negative. 'Negative' classification means a normal patient having no signs of TD.The dataset is balanced by increasing the record counts for other target classes using BOO_ST and SMOTE techniques[19].

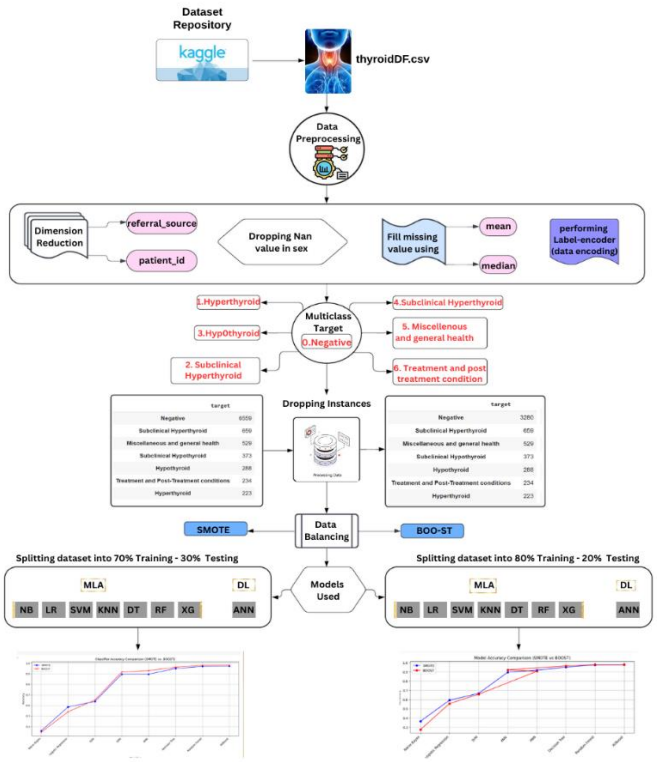


Figure 1. Proposed Methodology of the Research Study

For balancing the dataset, the pool of 6559 records with negative target class is reduced to 3280 records. For target classes like hypothyroid, hyperthyroid, subclinical hypothyroid, subclinical hyperthyroid, general health issues, treatment and post-treatment disorders the total count of samples are maintained same.

	target
Negative	6559
Subclinical Hyperthyroid	659
Miscellaneous and general health	529
Subclinical Hypothyroid	373
Hypothyroid	288
Treatment and Post-Treatment conditions	234
Hyperthyroid	223

dtype: int64

Table1. Unbalanced dataset

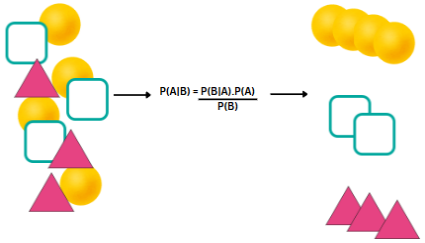
The resultant balanced dataset contained 18 duplicated records when BOO_ST techniques were applied. Similarly 32 duplicated records were found in the dataset with SMOTE technique. After removing duplicate records, the dataset was divided into 70%: 30% of training and testing dataset and also into 80%:20% split of training and testing dataset. Experiments using several ML and DL algorithms were carried on both the split types to determine the efficiency of the models.

	target
Negative	3280
Miscellaneous and general health	3280
Hypothyroid	3280
Hyperthyroid	3280
Treatment and Post-Treatment conditions	3280
Subclinical Hyperthyroid	3280
Subclinical Hypothyroid	3280

Table2. Balanced dataset.

Classifiers Description

Naive Bayes (NB): NB is a well-known probabilistic model for binary and multiclass classification task. NB operates under the assumption that features are conditionally independent given the class label. This means the presence or absence of a feature does not affect the other class label when the predictions are performed.

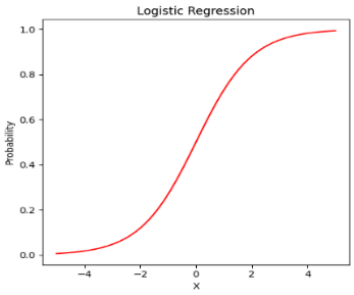


Formula: `P(A|B) = (P(B|A) * P(A)) / P(B)`
Where:

- `P(A|B) is the` **posterior probability** of class A given the features B.
- `P(B|A) is the` **likelihood** of the features B given the class A.
- P(A) is the **prior probability** of class A.
- P(B) is the **evidence** of the features B.

Logistic Regression (LR):

LR is an MLA model used for binary classification. It predicts the probability whether a given instance belongs to a particular class. To classify an instance LR uses a threshold (usually 0.3 – 0.5) to determine the class. Logistic function is based on the linear combination of input features.[20]

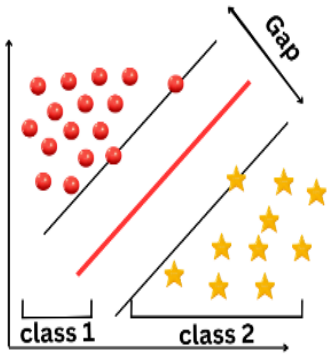


Formula: $\sigma(z) = 1/1+e^{-zl}$

Where: $z=\beta_0+\sum_{i=0}^{n} \beta_i \cdot x_i$ is the linear combination of the input features xi with corresponding weights βi and bias β_0.

Support Vector machine (SVM):

SVM is a supervised MLA that determines a hyperplane to maximize the margin between the classes. It is very effective in dividing the sample data. into different classes either using a hyperplane or using a set of hyperplanes for n dimension space. SVM works well in both linear and non-linear scenarios and performs tasks in both regression and classification. This SVM formula helps to decide the category by drawing a line boundary that separates the best data fit.

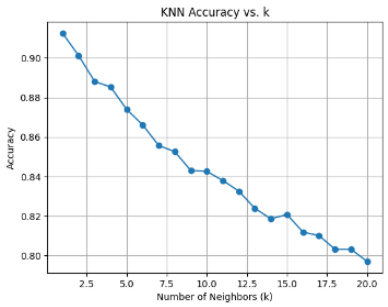


Formula: $f(X) = \text{sign}(w_1X_1 + w_2X_2 + ... + w_nX_n + b)$

Where: f(X) is the function of X features, w represents weight vector, b is the bias term, and X is features vector.

K-Nearest Neighbors (KNN):

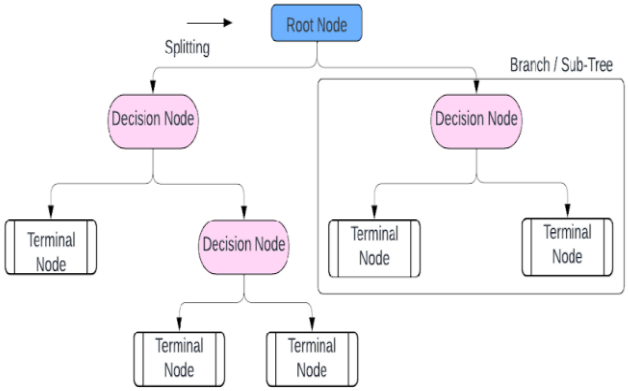
KNN is a supervised learning algorithm for predicting labelled data points by using elbow method and the majority in its closest neighbors. KNN is called lazy learner because the searching for the nearest neighbor from the whole training data makes the prediction very slow during the execution.



Formula : $Y^{\wedge} = \text{mode}\ \{y_i \mid i \epsilon N_k(X)\}$

Where: The correct prediction is determined by majority class (mode) among the k nearest neighbors $N_k(X)$ to the input X.

Decision Trees (DT): DT, supervised MLA is used for both classification and regression problems. DT follows a set of if-else conditional statements to visualize the data and performs prediction according to the conditions. The important terminology DT works on root node, sub-tree, splitting, decision node, terminal node[21].



It mainly works on two attribute selective measures (ASM) used for data mining to reduce the data. The data analysis is very important because to make a better prediction of the target class.

1. Gini Index: $1-\sum_{i=1}^{n} (pi)^{\wedge}2$
2. Information Gain $E(s) = \sum_{i=1}^{c} -pilog2pi$

Random forest Tree (RF):

One of the popular MLA technique is RF. It is easy to use, combined with its ability to solve both classification and regression problems, have driven its popularity which aggregates the output of several decision trees to produce a single outcome. RF uses tree based classification method. Three primary hyperparameters for RF must be established prior to training. These include the number of trees, the size of the nodes, and the quantity of traits samples. RF slightly choose a portion of DT features [22].

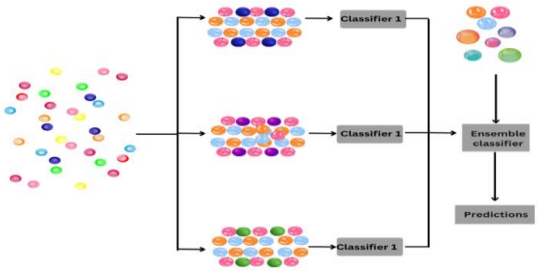
Formula: $y\hat{} = \frac{1}{m}\sum_{K=1}^{m} hk(X)$

Where:

- Y^ is the final prediction made by the RF.
- m is the total number of trees.
- $h_k(X)$ is the prediction made by the k-th DT for input X.
- X is the input feature vector.

Extreme-Gradient Boosting (XGBOOST):

XGBOOST [23] aims to reduce the discrepancy between the expected and actual values and at the same time as preventing the model from overfitting. In addition to using regression and classification to solve problems, this approach is the preferred choice for a variety of jobs.



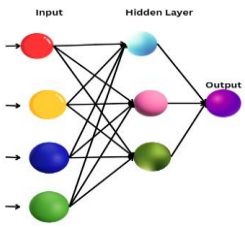
Objective = Loss + Regularization

Where:

- Loss – measures how the model's predictions match the actual values.
- Regularization – Penalty for making model too complex.

Artificial-Neural-Network (ANN):

ANN is part of supervised ML where we will be having multiple inputs as well as corresponding output in the dataset. The main aim of ANN[24] is to figure out a way of mapping the input layers of interconnected "neurons" that process data to its respective output. ANN(DL) is used to solve both classification as well as regression problems.



Formula: $Output=\sigma(i=1\sum nwi\cdot xi+b)$

Where:

- xi input features
- Wi are the weights corresponding to each input
- b is the bias term, which allows the model to shift the activation function
- σ is the activation function, which introduces non-linearity
- n is inputs to the neuron.

EXPERIMENTAL ANALYSIS

Synthetic Minority Over-sampling Technique (SMOTE): SMOTE technique operates by selecting the feature space that is close to one another. In this technique, a line is drawn between them, and a new sample is drawn at a point along the line. A minority class instance is first randomly selected via SMOTE[25], and then convexly created to find its k closest minority class neighbors.

Bootstrap Synthetic Minority Over-sampling Technique (BOO_ST): BOO_ST is an oversampling technique that generates instances of the minority class using k-nearest neighbors[6]. Boosting in conjunction with data sampling techniques might be a useful strategy for addressing issues related to class imbalance.

Validation:

The MLA models' efficacy and dependability are measured through various performance metrics. Every metric represents the performance of the model from different aspects. It is essential to assess model performance using metrics such as:

i) Accuracy: The most fundamental performance parameter is accuracy, which measures how accurate the model's predictions are overall.

Accuracy = (TP + TN) / (TP+TN+FP+FN)

ii) Precision: The model refers to how "specific" the model's positive predictions are measured out of all positive predictions made by the model.

Precision= TP / (TP + FP)

iii) Recall: Recall gauges how well the models are able to locate every instance of positivity inside the dataset.

Recall = TP / (TP + FN)

iv)F1 score: The model's total performance is determined by its harmonic mean recall and precisions. It is especially useful when dealing with imbalanced classes.

F1 score = 2 * (Precision * Recall) / (Precision + Recall)

	Model	BOOST Accuracy	BOOST Precision	BOOST Recall	BOOST F1-Score
0	KNN	0.917829	0.918621	0.917829	0.916059
1	Logistic Regression	0.555798	0.550824	0.555798	0.549166
2	Decision Tree	0.967088	0.967011	0.967088	0.967013
3	Random Forest	0.981037	0.981092	0.981037	0.980948
4	SVM	0.652354	0.652378	0.652354	0.644786
5	Naive Bayes	0.282258	0.533502	0.282258	0.234955
6	ANN	0.935702	0.936346	0.935702	0.935290
7	XGBoost	0.981255	0.981341	0.981255	0.981183

Table 3(a) .Performance metrics for 70%-30% split

	Model	Accuracy	Precision	Recall	F1-Score
0	KNN	0.909566	0.910410	0.909566	0.907595
1	Logistic Regression	0.541512	0.527143	0.541512	0.527785
2	Decision Tree	0.961647	0.961658	0.961647	0.961614
3	Random Forest	0.980388	0.980344	0.980388	0.980302
4	SVM	0.636958	0.635439	0.636958	0.626828
5	Naive Bayes	0.284376	0.605828	0.284376	0.238773
6	ANN	0.937241	0.937756	0.937241	0.937183
7	XGBoost	0.982349	0.982342	0.982349	0.982301

Table 3(b) .Performance metrics for 80%-20% split

RESULT

Table 4(a) and 4(b) presents the accuracy of ML and DL models used in the study. The accuracy of the model remains all the same for the data split of 70%-30% and 80%-20% split ratio of training and test dataset.

	Model	SMOTE Accuracy	BOOST Accuracy
0	KNN	0.884850	0.917829
1	Logistic Regression	0.586362	0.555798
2	Decision Tree	0.951003	0.967088
3	Random Forest	0.973103	0.981037
4	SVM	0.642919	0.652354
5	Naive Bayes	0.339052	0.282258
6	ANN	0.912329	0.935702
7	XGBoost	0.972521	0.981255

Table 4(a) . Accuracy of classifiers for 70%-30% split

It is clear from the data that BOOST improves models' performance to a great degree, particularly for classifiers like XGBOOST, RF, and DT. The difference between SMOTE and BOOST between training and testing data splits of 70% - 30% and 80% - 20% is quite noticeable in this comparison. Further BOOST technique's efficacy in enhancing accuracy is demonstrated in KNN, DT, and ANN models.

	Model	SMOTE Accuracy	BOOST Accuracy
0	KNN	0.890754	0.909566
1	Logistic Regression	0.581771	0.541512
2	Decision Tree	0.956171	0.961647
3	Random Forest	0.972961	0.980388
4	SVM	0.639119	0.636958
5	Naive Bayes	0.351068	0.284376
6	ANN	0.910816	0.937241
7	XGBoost	0.972961	0.982349

Table 4(b). Accuracy of classifiers for 80%-20% split

NB and LR are examples of binary models that only display lower marginal gains and subpar BOOST performance. According to our findings, BOOST has a smaller effect on simpler classifiers than on more complex models, even though it can significantly improve their performance. All things considered, the BOOST approach helped XGBOOST and RF attain the greatest accuracy rates of 98.23% and 98.10%.

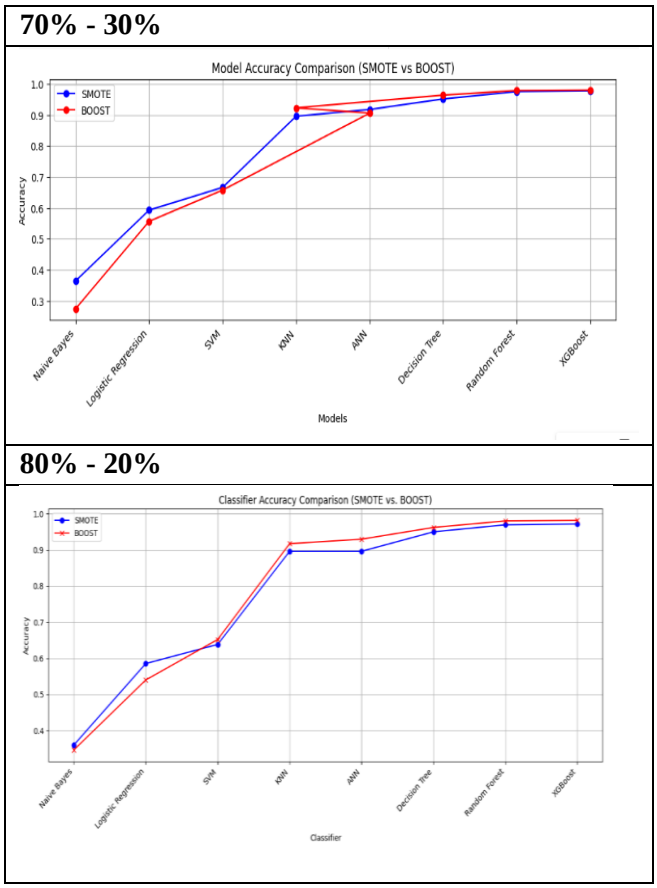


Figure 4 Model accuracy for 70%-30% split

CONCLUSION

This study predicts multiclass classification of TD by employing ML and DL algorithms. ML algorithms such as NB, DT, LR, KNN, SVM ,RF, XGBOOST and DL algorithm such as ANN is used in this model for predicting seven target classes of TD. SMOTE and BOO_ST techniques are applied for data balancing the

dataset. The dataset was divided into two combination 70% - 30% and 80% - 20% split of training and test data. This experiment helped to determine if the efficiency of the model improves with a different split ratio. The performance of ML and DL classifiers are determined using the standard performance metrics such as accuracy, precision, recall and F1-Score. The performance metrics are evaluated for both 70%-30% and 80%-20% split. The investigation demonstrates that boosting approaches greatly improves the performance of complex machine learning models, particularly when combined with data augmentation techniques like SMOTE. The virtues of ensemble approaches are demonstrated by sophisticated algorithms like RF and XGBOOST, which show significant gains in their prediction power through boosting. Conversely, less sophisticated algorithms such as LR and NB demonstrate limited or negative impacts from boosting, suggesting that these techniques are less useful for simpler models. According to these observations, boosting can significantly increase the efficacy of complicated models, but its benefits are less clear for simpler models. In order to attain optimal performance, the overall results emphasize how crucial it is to choose the right technique based on the model's complexity. The experimental study shows that the 70%-30% or 80%-20% split of training and testing dataset does not affect the performance of the model significantly.

The quantity of the dataset in the suggested study is limited . We need to have a better dataset with more data instances and attributes in the upcoming future in order to better generalize our findings. The training method will probably perform better with more data. Future research will investigate creating other classifications for thyroid disorders with improved accuracy.

REFERENCES

1.ChenLing, LiXue, ShengQuanZ, PengW-C(2016)Mining health examination records—a graph-based approach. IEEE TransKnowlDiscovEng28:2423–2437

2. Monaco Fabrizio (2003) Classification of thyroid diseases: suggestions for a revision. J Clin Endocrinol Metab 88:1428–1432

3. Ionita I, Ionita L (2016) Prediction of thyroid disease using data mining techniques. Broad Res Artif Intell Neurosci 7(3):115–124

4. Gorade SM, Deo A, Purohit P (2017) A study of some data mining classification technique. Int Res J Eng Technology 4(4):3112–3115

5. Bichler M, Kiss C (2004) A comparison of logistic regression, k-nearest neighbor, and decision tree induction for campaign management. In: Proceedings of the tenth Americas conference on information systems, New York

6. Khalid, M., Shamrat, F. J. M., Alshanbari, H., Farrash, M., & Qadah, T. M. (2024). A Dynamic Selection Hybrid Model for Advancing Thyroid Care with BOO-ST Balancing Method. IEEE Access.

7. Obaido, G., Achilonu, O., Ogbuokiri, B., Amadi, C. S., Habeebullahi, L., Ohalloran, T., ... & Aruleba, K. (2024). An Improved Framework for Detecting Thyroid Disease Using Filter-Based Feature Selection and Stacking Ensemble. IEEE Access

8. Shama, A., Hossain, M. B., Adhikary, A., Uddin, K. A., & Hossain, M. A. (2022). Prediction of Hypothyroidism and Hyperthyroidism Using Machine Learning Algorithms.

9. Mir, Y. I., & Mittal, S. (2020). Thyroid disease prediction using hybrid machine learning techniques: An effective framework. International Journal of Scientific & Technology Research, 9(2), 2868-2874.

10. Chaganti, R., Rustam, F., De La Torre Díez, I., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). Thyroid disease prediction using selective features and machine learning techniques. Cancers, 14(16), 3914.

11. OUARTANI, S., & TALEB, N. Decision Support System For Thyroid Disease Prediction Using Decision Tree Algorithm And Ontology.

12. Gupta, P., Rustam, F., Kanwal, K., Aljedaani, W., Alfarhood, S., Safran, M., & Ashraf, I. (2024). Detecting thyroid disease using optimized machine learning model based on differential evolution. International Journal of Computational Intelligence Systems, 17(1), 3.

13. Guleria, K., Sharma, S., Kumar, S., & Tiwari, S. (2022). Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. Measurement: Sensors, 24, 100482.

14. Cavalcante, C. M., & Rego, R. C. (2024, June). Early prediction of hypothyroidism based on feature selection and explainable artificial intelligence. In Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS) (pp. 49-60). SBC.

15. Shahid, A. H., Singh, M. P., Raj, R. K., Suman, R., Jawaid, D., & Alam, M. (2019, July). A study on label TSH, T3, T4U, TT4, FTI in hyperthyroidism and hypothyroidism using machine learning techniques. In 2019 International conference on communication and electronics systems (ICCES) (pp. 930-933). IEEE.

16. Alnaggar, M., Handosa, M., Medhat, T., & Z Rashad, M. (2023). Thyroid disease multi-class classification based on optimized gradient boosting model. Egyptian Journal of Artificial Intelligence, 2(1), 1-14.

17. Okuboyejo, S., Haqbeen, J., & Ito, T. (2023). Evaluating the Performance of Machine Learning Classifiers on Predicting Hypothyroidism for Public Healthcare Good. IIAI Letters on Informatics and Interdisciplinary Research, 4.

18. Almahshi, H. M., Almasri, E. A., Alquran, H., Mustafa, W. A., & Alkhayyat, A. (2022, May). Hypothyroidism prediction and detection using machine learning. In 2022 5th international conference on engineering technology and its applications (IICETA) (pp. 159-163). IEEE.

19. Srivastava, R., & Kumar, P. (2021). BL_SMOTE ensemble method for prediction of thyroid disease on imbalanced classification problem. In *Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020* (pp. 731-741). Springer Singapore.

20. Brindha, V., & Muthukumaravel, A. (2023). Efficient Method for the prediction of Thyroid Disease Classification Using Support Vector Machine and Logistic Regression. In *Computational Intelligence for Clinical Diagnosis* (pp. 37-45). Cham: Springer International Publishing.

21. Yadav, D.C., Pal, S. Prediction of thyroid disease using decision tree ensemble method. *Hum.-Intell. Syst. Integr.* **2**, 89–95 (2020). https://doi.org/10.1007/s42454-020-00006-y

22. Priya, V. V., Subashini, R., & Priya, S. H. (2023, February). Thyroid Disease Prediction using Random Forest Algorithm. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 794-799). IEEE.

23. Sankar, S., Potti, A., Chandrika, G. N., & Ramasubbareddy, S. (2022). Thyroid disease prediction using XGBoost algorithms. *J. Mob. Multimed*, *18*(3), 1-18.

24. Vasan, C. R. C., DSU, B., MS, C., & Devikarani, H. S. (2018). Thyroid detection using machine learning. *Computing (PDGC-2018)*, *20*, 22.

25. Khan, T. (2021, January). Application of two-class neural network-based classification model to predict the onset of thyroid disease. In *2021 11th International conference on cloud computing, data science & engineering (Confluence)* (pp. 114-118). IEEE.

https://www.thyroid.org/thyroid-function-tests