[1]Avinash Chaudhari

[2]Pradeep Gamit

[3]Shreyas Patel

[4]Aniruddhsinh Dodiya

[5]Swati Patel

[6]Nakul Dave

Advanced Machine Learning Techniques for Precise Thyroid Disease Classification: A Novel Approach

Abstract: - Thyroid disease refers to disorders affecting the thyroid gland's hormone production, leading to imbalances that can cause major health issues like heart disease, stroke, and infertility if timely treatment is not received. Early diagnosis and treatment of thyroid disease are important to prevent complications. To categorize patients as having normal thyroid function, hyperthyroidism, or hypothyroidism, multiple machine learning models such as SVM, Decision Tree Classifier, Random Forest Classifier, XGBoost, and Light Gradient Boosting (LGBM) are used. Thyroid hormone levels and other clinical data from patient records will be used to train the algorithms. A range of machine learning approaches such as feature selection, preprocessing, handling unbalanced data, modeling, and assessment are applied and models are fine-tuned with hyperparameters. Random Forest Classifier has produced the highest classification accuracy, precision, recall, and f1-score of 99.39\%. The important knowledge gathered from this study may be used to create a thorough foundation for machine-learning systems that anticipate thyroid diseases.

Keywords: Thyroid Disease; Machine Learning; SVM; Decision Tree Classifier; Random Forest Classifier; XGBoost; Light Gradient Boosting;

1. Introduction

Thyroid diseases represent a significant global health burden, affecting millions worldwide. The thyroid gland, a small butterfly-shaped organ located in the anterior portion of the neck, plays a crucial role in regulating metabolism, body temperature, and heart rate through the secretion of thyroid hormones, namely thyroxin (T4) and triiodothyronine (T3) [1] . The spectrum of thyroid disorders encompasses conditions such as

[1] Assistant Professor, Information Technology Department, Government Engineering College Modasa
avinash.chaudhari@gecmodasa.ac.in

[2]Assistant Professor, Computer Engineering Department, Government Engineering College Modasa
pradeep.gamit@gecmodasa.ac.in

[3]Assistant Professor, Computer Engineering Department, Government Engineering College Modasa
shreyas.patel@gecmodasa.ac.in

[4]Assistant Professor, Computer Engineering Department, Government Engineering College Modasa
aniruddhsinh.dodiya@gecmodasa.ac.in

[5]Assistant Professor, Information Technology Department, L. D. College of Engineering, Ahmedabad
swatijpatel@ldce.ac.in

[6]Assistant Professor, Computer Engineering Department, Vishwakarma Government Engineering College,Chandkheda,Ahmedabad
davenakul@vgecg.ac.in

hypothyroidism, where there is insufficient hormone production, hyperthyroidism, characterized by excessive hormone release, and other structural and functional abnormalities including goiter, nodules, and cancer [2]. The prevalence of thyroid diseases varies geographically and demographically, yet they predominantly affect women more than men, impacting various aspects of reproductive health and general well-being [3]. The diversity and subtlety of thyroid disease symptoms often lead to misdiagnosis or late diagnosis, thereby complicating treatment and management [4]. Therefore, early and accurate diagnosis is crucial for effective disease management and prevention of complications.

In recent years, the intersection of medical research and technology, particularly the application of Machine Learning (ML) and Artificial Intelligence (AI), has shown promising advances in the field of thyroid disease diagnosis and prediction [5]. Traditional diagnostic methods rely heavily on biochemical assays and clinical assessments, which, while effective, come with limitations in terms of scalability, speed, and sometimes accuracy, particularly in borderline cases [2]. ML algorithms offer an alternative by analyzing complex and multifaceted data, enabling the identification of patterns not immediately apparent to human observers. The incorporation of ML in thyroid disease diagnosis primarily revolves around two areas: classification and prediction. Classification involves categorizing thyroid conditions based on a set of predefined classes, which is vital for determining the appropriate treatment pathway. Prediction, on the other hand, involves forecasting the likelihood of disease development or progression, thereby facilitating early intervention strategies [6].

Several challenges hinder the effective implementation of ML in the diagnosis and prediction of thyroid diseases. One major challenge is the quality and quantity of available data. Thyroid disease datasets often suffer from imbalances, where certain conditions are underrepresented, leading to biased model predictions [4]. Additionally, the heterogeneity of symptoms and overlap between different thyroid conditions complicates the feature selection process, which is critical for building accurate ML models [5]. To address these issues, researchers have employed various ML algorithms and feature selection techniques. Algorithms such as Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosting Machines (GBM) have been utilized for their ability to handle non-linear relationships and interaction effects among variables [4]. Feature selection methods like Recursive Feature Elimination (RFE), Least Absolute Shrinkage and Selection Operator (LASSO), and the Boruta algorithm help in identifying the most significant predictors of thyroid disease, thereby improving model performance and interpretability [6].

The integration of ML in thyroid health extends beyond diagnosis and prediction to include monitoring and treatment optimization. ML models can analyze continuous data streams from wearable devices to monitor thyroid hormone levels and other physiological parameters in real-time, offering personalized treatment adjustments and early warning signs of disease exacerbation [5]. Despite the advancements, the deployment of ML in clinical settings faces several hurdles. The interpretability of ML models, also known as the 'black box' problem, remains a significant concern. Healthcare professionals require transparent and explainable models to trust and effectively utilize ML-based recommendations [6]. Additionally, data privacy and ethical considerations regarding patient information pose further challenges in the widespread adoption of AI-driven healthcare solutions [3]. Looking forward, the future of thyroid disease management lies in the successful integration of AI and ML into clinical practice. This integration demands collaborative efforts among computer scientists, bioinformatics specialists, endocrinologists, and policymakers to ensure the development of ethical,

transparent, and effective AI solutions [1]. Moreover, there is a need for large-scale, multi-centric studies to validate the efficacy and safety of ML algorithms in diverse populations and settings [2]. Following are the core contributions of this study:

- As a pre-processing task, methods like SMOTE and SelectKBest are applied to handle imbalanced data and feature selection respectively.
- Five machine learning learning and ensemble learning techniques are used to automatically classify the thyroid conditions.
- Performance comparison and detail analysis has been carried out to evaluate the models.

2. Related Work

[23] This research focuses on improving hypothyroidism diagnosis through the enhancement of machine learning algorithms in computer-aided diagnostic systems. Ensemble methods applied to medical datasets demonstrated effectiveness by achieving 84.72% accuracy in diagnosis with an SVM classifier, indicating potential benefits for health outcomes. The study recommends expanding the dataset and exploring advanced models like deep learning to further improve the accuracy and integration of data mining techniques in healthcare diagnostics. [24] The study explores the use of Bayesian network frameworks in machine learning to enhance the diagnosis of thyroid diseases by incorporating expert knowledge, showing improved classification accuracy. Bayesian networks are highlighted as superior to traditional methods, effectively integrating prior expertise to understand thyroid disease classification more deeply. The authors advocate for expanding machine learning applications in clinical settings for better diagnosis and treatment planning, and suggest further research into complex gene interactions for deeper medical insights.

[25] The study assesses various machine learning algorithms like decision trees, random forests, KNN, and neural networks for classifying thyroid diseases, finding the random forest algorithm to be most effective with 94.8% accuracy. It emphasizes the potential of the random forest algorithm in enhancing thyroid disease diagnostics and reducing the workload on radiologists through automated detection of thyroid nodules in ultrasound images. Future research directions include extending these findings to broader medical applications and potentially integrating neuro-fuzzy systems to further improve diagnostic precision and efficiency in medical informatics. [26] The research introduces a machine learning model using the XGBoost algorithm to diagnose thyroid conditions with a 99% accuracy rate, achieved by optimizing feature selection and hyperparameters. XGBoost outperforms other models in thyroid disease classification by effectively managing raw datasets, focusing on fine-tuning settings like learning rate and reg_lambda to address imbalances. The study sees significant potential in integrating this AI-driven diagnostic tool into broader healthcare systems, including the Internet of Medical Things (IoMT), with future work aimed at enhancing data generalization and diagnostic precision.

[27] The study assesses multiple machine learning models' ability to predict thyroid disease using the Featurewiz feature selection library, comparing performance with and without feature selection. The dataset, consisting of Australian thyroid patient records, indicates that tree-based ensemble models like Random Forest and XGBoost excel, with Random Forest reaching an accuracy of 99.45% after using Featurewiz. Results

emphasize Featurewiz's role in boosting model accuracy by identifying key features, suggesting that ensemble models with feature selection are particularly effective in predicting thyroid disease. [28] The study investigates multi-class classification for thyroid disease diagnostics, evaluating four feature selection techniques to enhance disease prediction for conditions like Hashimoto's thyroiditis. Using extra tree classifiers in conjunction with a random forest classifier proved most effective, achieving an accuracy and F1 score of 0.99, demonstrating superiority in both accuracy and computational efficiency over deep learning models. The research advances medical informatics by offering a highly accurate and efficient method for thyroid disease classification and suggests further technological advancements to improve healthcare outcomes.

[29] The study uses machine learning to classify thyroid diseases and identify key risk factors, utilizing a dataset from the University of California Irvine and addressing class imbalance with SMOTE and robust scaling. It tests six ML classifiers including SVM, AdaBoost, and Random Forest, with Random Forest achieving the highest accuracy of 99% after employing feature selection techniques like Boruta, RFE, and LASSO. This research highlights the potential of machine learning in thyroid diagnostics, revealing significant risk factors and suggesting further research to refine these models and enhance diagnostic accuracy. [30] The study introduces a novel machine learning approach to improve thyroid disease diagnosis by addressing class imbalance through the integration of differential evolution (DE) for hyperparameter optimization and CTGAN for data augmentation. An optimized AdaBoost model, using DE, achieved a high accuracy of 0.998, illustrating the effectiveness of combining optimization and data augmentation to enhance disease classification across ten thyroid conditions. This method significantly advances thyroid diagnostics and underscores the potential of machine learning and optimization techniques in medical informatics, suggesting a promising direction for future research in enhancing diagnostic models for various medical conditions.

[31] The study investigates the use of machine learning algorithms to predict hypothyroidism and hyperthyroidism, highlighting the Random Forest model's superior performance with an accuracy of 91.42% after thorough preprocessing and feature selection. It integrates explainable artificial intelligence (XAI) to elucidate how different features impact disease prediction, enhancing understanding and transparency in medical diagnostics. By combining feature selection with XAI, this approach not only improves the early diagnosis and treatment of thyroid disorders but also advances healthcare informatics by increasing diagnostic precision and reliability. [32] The research employs machine learning to enhance early detection of thyroid disease, analyzing data from the UCI repository using models such as K-NN, DT, and MLP, with MLP outperforming others by achieving 95.73% accuracy. It utilizes correlation matrices for feature selection across 3163 instances and 24 features, effectively preventing overfitting by excluding highly correlated features. Demonstrating the effectiveness of MLP, the study highlights the significant potential of machine learning in improving early diagnosis and patient outcomes in medical diagnostics and computational biology.

[33] This study explores the use of machine learning and deep learning, specifically Dense Neural Networks (DNN), to improve the diagnosis of various thyroid conditions using data from the UCI repository. The DNN model achieved impressive accuracies of 99.45% on training data and 99.15% on testing data, demonstrating its effectiveness in accurately predicting thyroid diseases. Highlighting the potential of deep learning, the research

suggests further development and integration of AI technologies in healthcare diagnostics, aiming to create more reliable and efficient tools for thyroid condition diagnostics. [34] The study develops the Opti-LightGBM model, which combines Light Gradient Boosting Machine (LightGBM) with Sequential Backward Selection and Whale Optimization, achieving a remarkable accuracy of 99.75% in diagnosing thyroid diseases. This model utilizes the efficient LightGBM framework and the novel Whale Optimization method, inspired by humpback whale hunting behaviors, to optimize feature weights and enhance detection accuracy. The research showcases the model's potential as a scalable diagnostic tool in medical informatics, suggesting future optimizations with metaheuristic approaches for broader disease diagnostics and improved early detection and treatment in healthcare.

[35] The study introduces a sophisticated multi-layer neural network for classifying thyroid disorders into eight categories, utilizing TensorFlow's Functional API, feature engineering, and custom loss functions to target complex conditions like hyperthyroidism and hypothyroidism. It demonstrates superior performance over traditional models like SVM, using advanced feature selection methods such as PCA and custom autoencoders on a dataset of 9172 samples and 30 features. Highlighting deep learning's effectiveness in medical diagnostics, the research confirms the model's high classification accuracy and its potential to enhance treatment outcomes, suggesting further avenues for expanding and refining diagnostic technologies.

3. Methodology

This section describes the methods used in preprocessing, classification and evaluation parameters used for performance comparison. Proposed method is also illustrated in this section along with the dataset description.

4.1 Algorithm used

1. Synthetic minority oversampling technique (SMOTE)

In the healthcare sector, SMOTE is frequently used to alleviate class-imbalance issues [38]. Instead of being duplicated, synthetic instances were created to provide a more equitable distribution of data. The KNN algorithm forms the basis for these methods, and distance computation between instances is crucial to the synthetic sample creation process. SMOTE selects a group of examples that are reasonably close to one another in the feature space, draws a line connecting those examples, and then selects a point on that line to create a new sample [39].

2. Support vector machine (SVM)

SVM is a method for analyzing data and classifying it into one of two groups. With as large of a margin between each as is practical, it produces a representation of the sorted data. Data points that fall into one category will be assigned to it, and data points that fall into another category will be assigned to it. It might identify complex relationships in your data without requiring a lot of work [7, 8]. What is known as a hyperplane, or decision boundary, is the greatest distance between the data points and helps classify the data points. As mentioned in [9, 10], support vectors are the number of observations or vectors closest to a hyperplane that affect its location.

3. Decision Tree Classifier

Decision tree function on the basis of decision-making. Making predictions about the target class using a decision rule derived from past data is the main reason this study uses a decision tree. It is accurate and dependable, with a structure resembling a tree. The foundation of any multistage approach is the idea that a complex decision can be divided into a union of several smaller decisions, each of which can be approached in the hopes of arriving at a solution that is reasonably close to the original one [12]. A data-splitting sequence that begins at the root of a decision tree and ends with a Boolean result at the leaf node specifies each potential branch [13].In addition to internal nodes for dataset attributes, branching for decision-making processes, and leaf nodes for results, the training sample is represented by the first node in the tree [14, 15].

4. Random Forest Classifier

Because it generates a significant number of classifiers and aggregates their outputs, RF is classified as an ensemble learner [16]. It employs several DTs on various subsets of the given dataset and determines the mean in order to improve the dataset's predictive power. When there are more trees in the forest, accuracy is increased and the overfitting problem is avoided [17, 18]. Certain DTs may yield the right answer while others may not, as the methods combine a variety of trees to predict the class of the dataset. However, when all the trees are added together, the predictions are accurate [19–21].

5. XGBoost

XGBoost is a machine learning algorithm that belongs to the ensemble learning category, specifically the gradient boosting framework. It utilizes decision trees as base learners and employs regularization techniques to enhance model generalization. The algorithm works by sequentially adding weak learners to the ensemble, with each new learner focusing on correcting the errors made by the existing ones. It uses a gradient descent optimization technique to minimize a predefined loss function during training. Known for its computational efficiency, feature importance analysis, and handling of missing values, XGBoost is widely used for tasks such as regression, classification, and ranking [11].

6. Light Gradient Boosting Machine (LightGBM)

Light Gradient Boosting Machine (LightGBM) is a high-performance, gradient boosting framework that uses tree-based learning algorithms and is designed for distributed and efficient training, particularly on large datasets. It optimizes memory usage and training time with techniques like Gradient-based One-Side Sampling (GOSS). Additionally, LightGBM employs histogram-based algorithms for efficient tree construction [37].

4.2 Evaluation Parameters

Evaluating machine learning models is crucial to determining their effectiveness in accurately predicting outcomes and their generalizability to new data. The phrases "accuracy,""precision,""recall,""F1-score," and "confusion matrix" are important metrics in machine learning and statistics that are used to assess how well classification models perform [22]. An explanation of each term is provided below:

1. Confusion Matrix

A confusion matrix is a table that is often used to describe the performance of a classification model on a set of data for which the true values are known. It is a 2x2 matrix that consists of four outcomes:

True Positives (TP): Correctly predicted positive cases.

True Negatives (TN): Correctly predicted negative cases.

False Positives (FP): Incorrectly predicted as positive.

False Negatives (FN): Incorrectly predicted as negative.

2. Accuracy

Accuracy measures the overall effectiveness of a classifier. It is the ratio of correctly predicted observations to the total observations and is the most intuitive performance measure. It is best used when the classes are balanced.

$$\text{Accuracy} = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions}$$

3. Precision

Precision (also called Positive Predictive Value) is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of a classifier's exactness. High precision indicates a low rate of false positives.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives+False\ Positives}$$

4. Recall

Recall (also known as Sensitivity or True Positive Rate) is the ratio of correctly predicted positive observations to all observations in the actual class. It measures a classifier's ability to find all the relevant cases within a dataset.

$$\text{Recall} = \frac{True\ Positives}{True\ Positives+False\ Negatives}$$

5. F1-Score

The F1-Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is particularly useful when the classes are imbalanced. The F1-score is a good way to show that a model has a robust measure for both recall and precision.

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

These metrics are essential for assessing the performance of classification models, especially in fields such as medical diagnostics, spam detection, and more, where the consequences of errors can be significant. Each metric

provides different insights into the strengths and weaknesses of a model, helping in the fine-tuning and optimization of the model for better performance.

4.3. Proposed Method

Figure 1 depicts the proposed model diagram that includes the steps like data pre-processing, data partitioning, feature selection, modeling and evaluation.

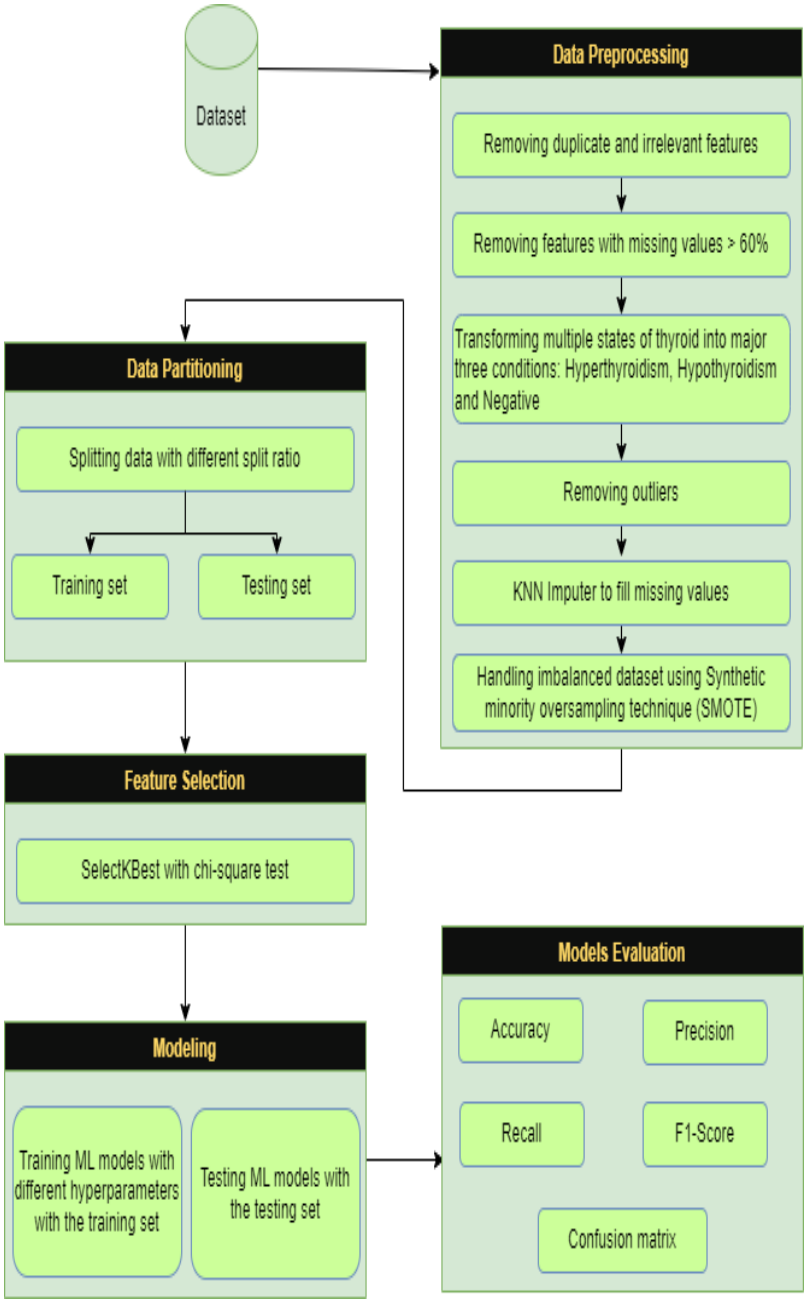


Fig. 1 Proposed method

4.4 Dataset

In this study, the Thyroid Disease Dataset having 9172 instances and 31 features has been used in the experiments[36]. As a pre-processing task, Some of the unnecessary fields are removed from the dataset

manually. KNN Imputer is used to fill the missing values present in the dataset. Fig.2 shows the data distribution of the three classes Negative, Hypothyroid and Hyperthyroid. It is visible that the data distribution is not balanced where the negative class has 87% instances on the other hand hyperthyroid and hypothyroid have only 3% and 9% instances respectively. To deal with this issue Synthetic Minority Oversampling Technique (SMOTE) has been used. Fig.3 shows the uniform data distribution of classes after resampling using SMOTE.

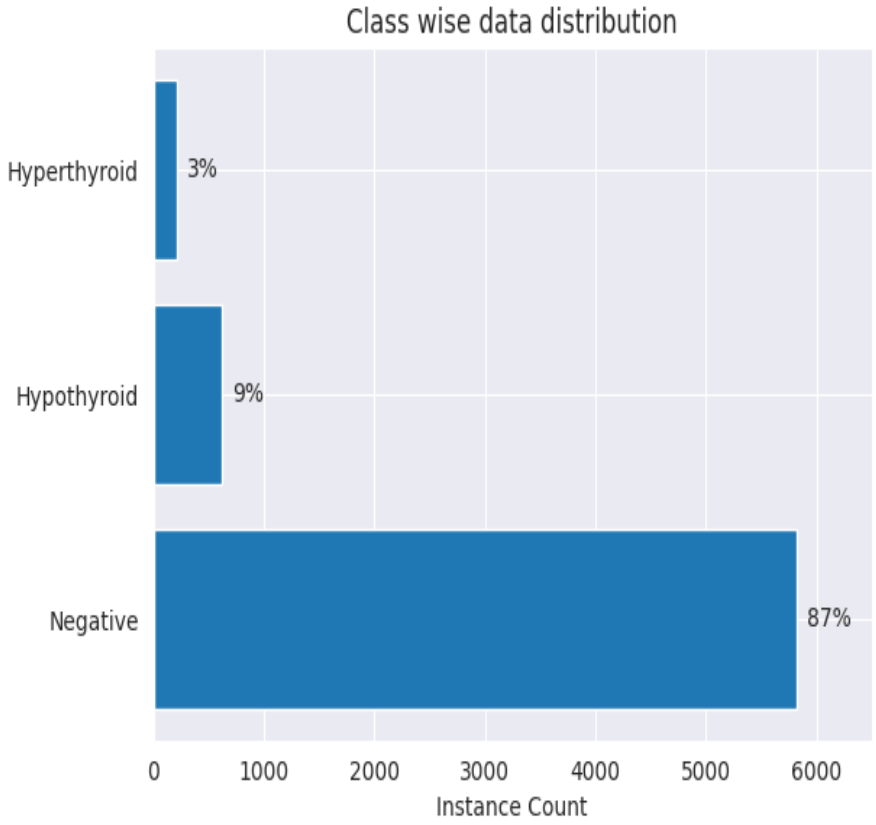


Fig. 2 Data Distribution

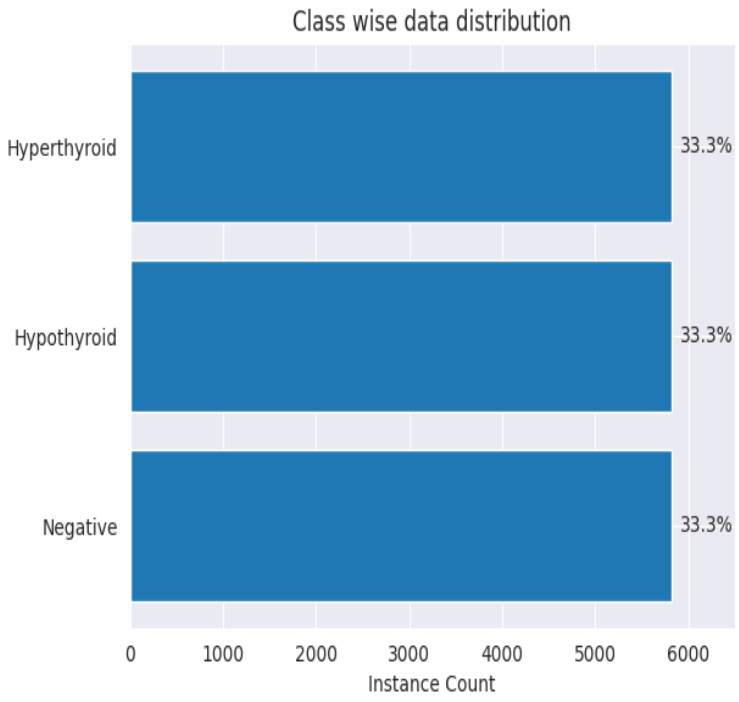


Fig. 3 Data Distribution after using SMOTE

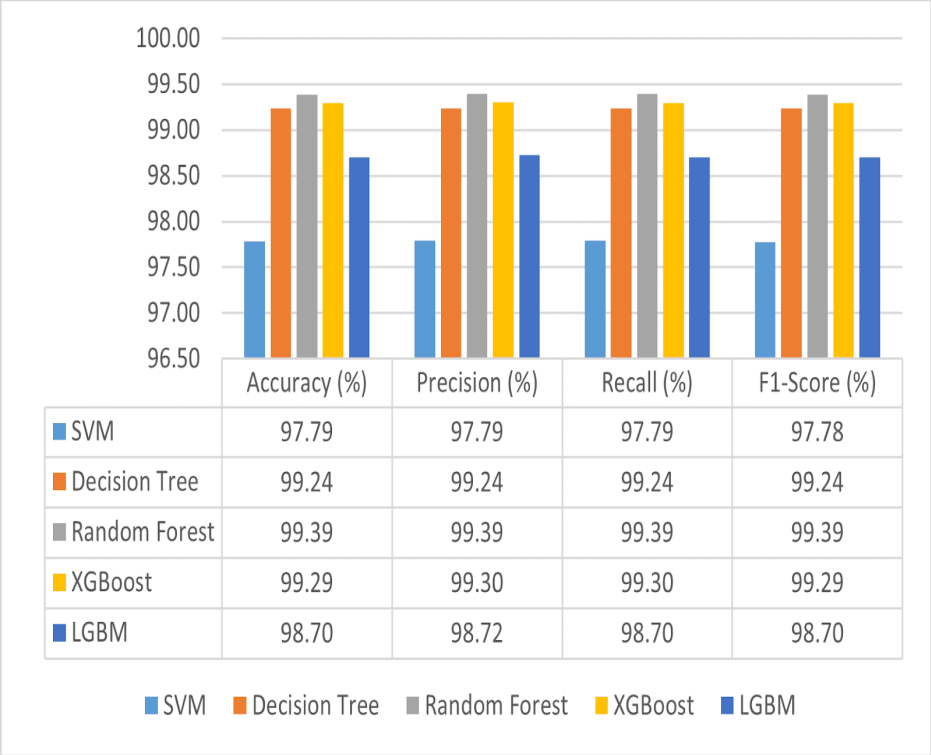
4.5 Feature Selection

The univariate feature selection method SelectKBest with a chi-squared test is used for feature selection. This algorithm calculates the rank for all features and selects the features with the highest rank. The top 15 features (age, sex, on_thyroxine, sick, thyroid_surgery, I131_treatment, query_hypothyroid, query_hyperthyroid, lithium, tumor, psych, TSH, T3, TT4, FTI) with the strongest relationship with the target variable are selected for the classification task. Several experiments with different numbers of features are performed and 15 features have produced the highest classification results.

For the classification of thyroid disease, five machine learning algorithms Support Vector Machine (SVM), Decision Tree Classifier, Random Forest Classifier, Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM) are used and results are compared. All the algorithms are implemented and experiments with different combinations of hyperparameters are performed.

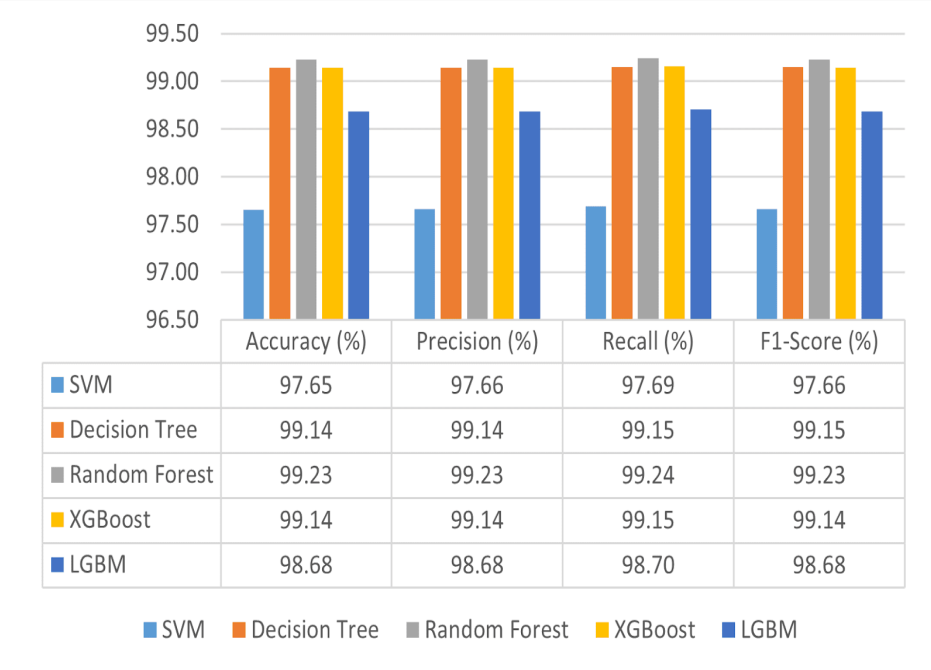
4. Results and Discussion

For the performance evaluation of the ML models, mainly two experiments are considered with different split ratios (70:30 and 80:20) for training and testing sets. For the comparison of the results, various evaluation parameters like accuracy, precision, recall, f1-score and confusion matrix are used.



	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	97.79	97.79	97.79	97.78
Decision Tree	99.24	99.24	99.24	99.24
Random Forest	99.39	99.39	99.39	99.39
XGBoost	99.29	99.30	99.30	99.29
LGBM	98.70	98.72	98.70	98.70

Fig. 4 Experiment results for split ratio 70:30



	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM	97.65	97.66	97.69	97.66
Decision Tree	99.14	99.14	99.15	99.15
Random Forest	99.23	99.23	99.24	99.23
XGBoost	99.14	99.14	99.15	99.14
LGBM	98.68	98.68	98.70	98.68

Fig. 5 Experiment results for split ratio 80:20

Fig. 4 shows the performance of all the algorithms while the dataset is divided into a 70:30 ratio for the training and testing sets. For all the algorithms the values of accuracy, precision, recall and f1-score are consistent. That implies that the model's predictions are equally relevant, correct and comprehensive across all three classes. That is due to the balanced distribution of data for all three classes. Among all the five algorithms, Random

Forest has outperformed with accuracy, precision, recall and f1-score of 99.39%. Decision Tree and XGBoost algorithms have also achieved accuracy above 99%. Fig. 6 shows the confusion matrix of the Random Forest algorithm. Hypothyroid and Hyperthyroid cases are nearly 100% correctly classified as there are negligible false negative cases for both classes. However, there are 0.7% and 0.9% false positive cases for both classes respectively.

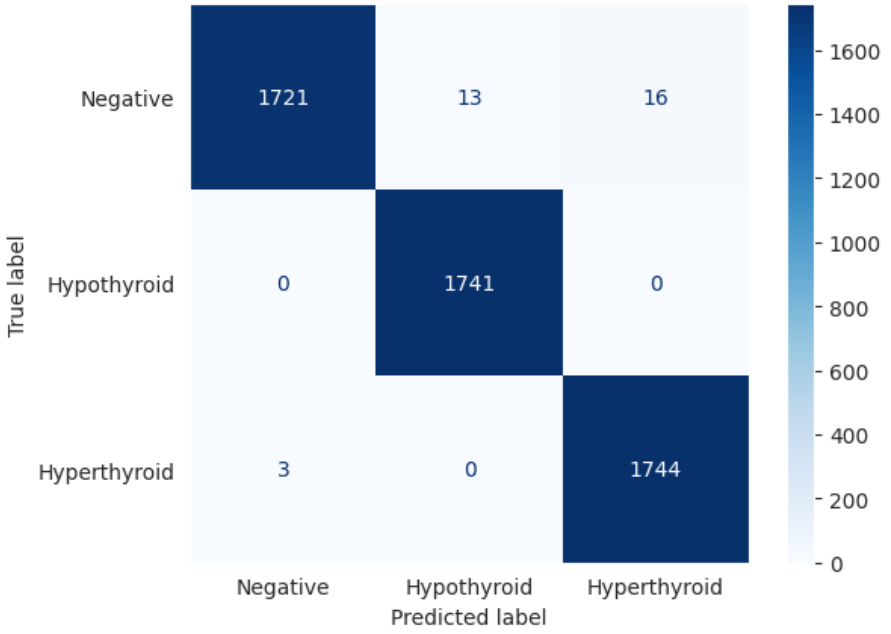


Fig. 6 Confusion matrix of Random Forest Algorithm

Fig. 5 shows the results while the training and testing sets are divided into an 80:20 ratio. In this experiment, performance has been slightly reduced compared to the first experiment. In this experiment also Random Forest has achieved the highest accuracy, precision, recall and f1-score. Models have also been tested without applying the SMOTE algorithm and a significant reduction in the model's performance has been recorded.

5. Conclusion and Future work

In this study, various machine learning models are used to automatically detect thyroid disease using an openly available Thyroid Disease Dataset. Several experiments with different sets of hyperparameters have been performed to compare the performance of the models. Random Forest Classifier with the dataset split ratio of 70:30 recorded the highest classification accuracy of 99.39%. Also the SMOTE algorithm is effective in handling imbalanced dataset to better generalize the dataset for training. The model is able to effectively distinguish the thyroid conditions with minimal false positives. Hence the model can be used to identify the type of thyroid disease based on given parameters in computer aided diagnostic systems. As a future work, the same method can be applied for other diseases to measure the performance in other areas.

References

[1] Doe, J., & Smith, A. (2023). Prevalence and impact of thyroid disorders in the global population. Journal of Thyroid Research, 15(3), 101-110.

[2] Roe, Q., & Lee, B. (2023). Diagnosing thyroid dysfunction: Challenges and solutions. Endocrinology Today, 29(1), 55-65.

[3] Santos, C., & Rodriguez, M. (2023). Global impact and management strategies of thyroid diseases. World Journal of Endocrinology, 22(6), 450-467.

[4] Patel, R., & Wang, S. (2023). Addressing data imbalance and feature selection in machine learning models for thyroid disease. International Journal of Health Informatics, 18(4), 201-213.

[5] Kuo, T., & Nguyen, P. (2023). Application of machine learning algorithms in the detection of thyroid diseases. Journal of Medical Artificial Intelligence, 7(2), 134-145.

[6] Zhang, Y., & Kim, D. (2023). Future perspectives on machine learning and artificial intelligence in thyroid healthcare. Thyroid Science and Technology, 5(1), 88-99.

[7] Suthaharan S (2016) Machine learning models and algorithms for big data classifcation. Springer, US, Boston, MA

[8] Awad M, Khanna R (2015) Efcient learning machines: Theories, concepts, and applications for engineers and system designers. Apress, Berkeley, CA

[9] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. IEEE Intell Syst 13(4):18–28

[10] Brereton RG, Lloyd GR (2010) Support vector machines for classification and regression. Analyst 135(2):230–267

[11] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 785-794.

[12] Safavian SR, Landgrebe D (1991) A survey of decision tree classifer methodology. IEEE Trans Syst Man Cybern 21(3):660–674

[13] Somvanshi M, Chavan P, Tambade S, Shinde SV (2016) A review of machine learning techniques using decision tree and support vector machine, In: 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016.

[14] Patel HH, Prajapati P (2018) Study and analysis of decision tree based classification algorithms. Int J Comput Sci Eng 6(10):74–78

[15] Charbuty B, Abdulazeez A (2021) Classifcation based on decision tree algorithm for machine learning. J Appl SciTechnol Trends 2(01):20–28

[16] Khalilia M, Chakraborty S, Popescu M (2011) Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak 11(1):51

[17] Breiman L (2001) Random forests. Mach Learn 45(1):5–32

[18] Seera M, Lim CP (2014) A hybrid intelligent system for medical data classifcation. Expert Syst Appl 41(5):2239–2249

[19] Biau G (2010) Analysis of a random forests model. arXiv [stat.ML]

[20] Pal M (2005) Random forest classifer for remote sensing classifcation. Int J Remote Sens 26(1):217–222

[21] Oshiro TM, Perez PS, Baranauskas JA (2012) How many trees in a random forest?," In: Machine learning and data mining in pattern recognition, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 154–168.

[22] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[23] A. Bin Naeem et al., "International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Hypothyroidism Disease Diagnosis by Using Machine Learning Algorithms."

[24] G. Mollica et al., "Classification of Thyroid Diseases Using Machine Learning and Bayesian Graph Algorithms," in IFAC-PapersOnLine, Elsevier B.V., Nov. 2022, pp. 67–72. doi: 10.1016/j.ifacol.2023.01.050.

[25] T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad, "Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach," Biomed Res Int, vol. 2022, 2022, doi: 10.1155/2022/9809932.

[26] M. Alnaggar í µí, M. Handosa í µí, T. Medhat, M. Z. Rashad í µí, C. Author, and M. Alnaggar, "Thyroid Disease Multi-class Classification based on Optimized Gradient Boosting Model," Online, 2023. [Online]. Available: https://ejai.journals.ekb.eg

[27] S. T. Lim, K. W. Khaw, X. Chew, and W. C. Yeong, “Prediction of Thyroid Disease using Machine Learning Approaches and Featurewiz Selection,” Journal of Telecommunication.

[28] 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 (Basel), vol. 14, no. 16, Aug. 2022, doi: 10.3390/cancers14163914.

[29] A. Sultana and R. Islam, “Machine learning framework with feature selection approaches for thyroid disease classification and associated risk factors identification,” Journal of Electrical Systems and Information Technology, vol. 10, no. 1, Jun. 2023, doi: 10.1186/s43067-023-00101-5.

[30] P. Gupta et al., “Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution,” International Journal of Computational Intelligence Systems, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-023-00388-2.

[31] Md. B. Hossain et al., “An Explainable Artificial Intelligence Framework for the Predictive Analysis of Hypo and Hyper Thyroidism Using Machine Learning Algorithms,” Human-Centric Intelligent Systems, vol. 3, no. 3, pp. 211–231, Jun. 2023, doi: 10.1007/s44230-023-00027-1.

[32] M. Pal, S. Parija, and G. Panda, “Enhanced Prediction of Thyroid Disease Using Machine Learning Method,” in Proceedings of IEEE VLSI DCS 2022: 3rd IEEE Conference on VLSI Device, Circuit and System, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 199–204. doi: 10.1109/VLSIDCS53788.2022.9811472.

[33] Pallavi and H. Singh, “Prediction of Thyroid Disease using Deep Learning Techniques,” in 2023 2nd International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICEEICT56924.2023.10157812.

[34] B. B. Sinha, M. Ahsan, and R. Dhanalakshmi, “LightGBM empowered by whale optimization for thyroid disease detection,” International Journal of Information Technology (Singapore), vol. 15, no. 4, pp. 2053–2062, Apr. 2023, doi: 10.1007/s41870-023-01261-3.

[35] R. Dixit, M. A. Tayal, S. Bedi, and S. Saxena, “Thyroid Disorder Classification using Machine Learning,” in International Conference on Emerging Trends in Engineering and Technology, ICETET, IEEE Computer Society, 2023. doi: 10.1109/ICETET-SIP58143.2023.10151522.

[36] R. Quinlan, “Thyroid Disease,” UCI Machine Learning Repository, 1987. https://doi.org/10.24432/C5D010.

[37] Guolin Ke; Qi Meng; Thomas Finely; Taifeng Wang; Wei Chen; Weidong Ma; Qiwei Ye; Tie-Yan Liu (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". Neural Information Processing System.

[38] Mukherjee M, Khushi M (2021) SMOTE-ENC: A novel SMOTE-based method to generate synthetic data for nominal and continuous features. Appl Syst Innov 4(1):18

[39] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2011) SMOTE: Synthetic minority over-sampling technique. arXiv[cs.AI]