INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Web Based Application for Early Detection of Thyroid Disorders in Nigeria

Peter S. Idoko[1*], Iyinoluwa T. Idowu[2], Emmanuel O. Ayodele[3], Peter A. Idowu[4]

[1]Department of Computer Science, Federal Polytechnic Orogun, Delta State

[2]Department of Nursing Science, Elizade University, Ilara-Mokin, Ondo State

[3,4]Department of Computer Science and Cybersecurity, Obafemi Awolowo University, Ile-Ife, Nigeria

[*]Corresponding Author

DOI: https://doi.org/10.51584/IJRIAS.2026.11030067

Received: 18 March 2026; Accepted: 24 March 2026; Published: 11 April 2026

ABSTRACT

Thyroid gland disorders represent a significant public health challenge globally, with a particularly pronounced burden in low- and middle-income countries like Nigeria. This paper focuses on selecting best features for early detection of Thyroid disorder in Nigeria using machine learning approach. In machine learning, feature selection is crucial to designing a good model and obtaining the best model performances. The redundant and undesired features may need to be removed from the original datasets to train the model faster, easily interpret the data, and avoid overfitting problems. This paper focuses on a robust ML-based selective features for prediction of early detection of thyroid gland disorders in Nigeria, leveraging clinical data (TSH, T3, T4, autoantibodies), ultrasound findings, demographic variables (age, sex, BMI), and environmental factors (iodine status, goitrogen exposure). This study employs a dual-pronged approach to feature selection, combining filter-based methods with Random Forest techniques to ensure comprehensive identification of the most predictive variables. The result showed that Random Forest and Gradient Boosting delivered superior results, with Random Forest slightly outperforming Gradient Boosting. Using all features, Random Forest achieved accuracy = 0.9978, precision = 0.9986, recall = 0.9971, F1-score = 0.9978, and ROC-AUC = 0.9999, indicating near-perfect discrimination. Gradient Boosting closely followed with similar metrics (accuracy = 0.9971, ROC-AUC = 0.9999). In conclusion, the comparative analysis confirms that Random Forest and Gradient Boosting offer the most reliable and accurate predictions, benefiting from their ensemble architecture and ability to model complex interactions.

Keywords: Selective Features, Thyroid Disorders, Random Forest, Gradient Boosting, Machine Learning.

INTRODUCTION

Thyroid gland disorders represent a significant public health challenge globally, with a particularly pronounced burden in low and middle-income countries like Nigeria, where healthcare resources are often limited, and diagnostic delays exacerbate disease outcomes. The thyroid gland, a small endocrine organ located in the neck, plays a critical role in regulating metabolism, growth, and development through the production of hormones such as triiodothyronine (T3) and thyroxine (T4), controlled by thyroid-stimulating hormone (TSH) from the pituitary gland [1]. Disorders of the thyroid, including hypothyroidism, hyperthyroidism, goiter, and thyroid cancer, arise from disruptions in this delicate hormonal balance, leading to a spectrum of clinical manifestations ranging from fatigue and weight changes to life-threatening cardiovascular complications and malignancy [2]. In Nigeria, the prevalence of thyroid disorders is estimated to range between 15% and 20% in the general population, with higher rates in specific regions due to environmental and dietary factors [3]. This high prevalence, coupled with limited access to advanced diagnostic tools and low public awareness, underscores the urgent need for innovative approaches to improve early detection and risk stratification.

Nigeria, as Africa's most populous nation, faces unique challenges in addressing thyroid disorders. The country's healthcare system is characterized by a shortage of specialized endocrinologists, limited availability of diagnostic imaging such as ultrasound, and inconsistent access to laboratory tests for thyroid function (TSH, T3, T4) and autoantibodies (anti-thyroid peroxidase [anti-TPO], anti-thyroglobulin [anti-Tg]) [4]. Furthermore, socioeconomic barriers, including poverty and inadequate healthcare infrastructure in rural areas, hinder timely diagnosis and treatment. For instance, fine-needle aspiration cytology (FNAC), a critical diagnostic tool for evaluating thyroid nodules, is primarily available in tertiary hospitals located in urban centers, leaving rural populations underserved [2]. These systemic challenges contribute to late-stage diagnoses, particularly for thyroid disorders, which have seen a rising incidence in Nigeria, with papillary thyroid carcinoma accounting for approximately 53% of cases [3]. Environmental and dietary factors play a significant role in the epidemiology of thyroid disorders in Nigeria. Iodine deficiency, a leading cause of goiter and hypothyroidism, remains prevalent in certain regions, particularly in the northern parts of the country where soil and water iodine levels are low [4]. Despite efforts to promote iodized salt consumption, compliance remains inconsistent, with studies reporting that only 60% of households in northern Nigeria use adequately iodized salt [5]. Additionally, the widespread consumption of goitrogenic foods, such as cassava, a staple food in many Nigerian diets, further exacerbates thyroid dysfunction by inhibiting iodine uptake and thyroid hormone synthesis [3]. Cassava, rich in cyanogenic glycosides, releases thiocyanate, which competes with iodine in the thyroid gland, leading to goiter formation, particularly in iodine-deficient populations [4]. Other environmental factors, such as exposure to radiation from natural or medical sources, and genetic predispositions, including autoimmune thyroid diseases like Hashimoto’s thyroiditis and Graves’ disease, also contribute to the burden of thyroid disorders in Nigeria [2].

Machine learning (ML) has emerged as a transformative tool in healthcare, offering the potential to address diagnostic gaps in resource-limited settings like Nigeria. ML algorithms, such as Random Forest, Support Vector Machines (SVM), and Neural Networks, can analyze complex datasets, including clinical, demographic, and environmental variables, to predict disease risk with high accuracy [6]. In the context of thyroid disorders, ML models have been successfully applied globally to classify patients as having hypothyroidism, hyperthyroidism, or thyroid cancer based on laboratory results, imaging features, and patient demographics [1]. For example, a study by [6] demonstrated that a Neural Network model achieved an F1-score of 0.92 in classifying thyroid disease using TSH, T3, T4, and autoantibody levels. Similarly, Random Forest models have been shown to effectively handle high-dimensional data, making them suitable for integrating diverse features such as ultrasound findings and dietary patterns [2].

In Nigeria, the application of ML to thyroid disorders risk assessment is particularly promising due to the ability of these models to work with heterogeneous and incomplete datasets, which are common in the country’s healthcare system. By incorporating locally relevant risk factors, such as iodine deficiency, goitrogenic diet, and socioeconomic variables, an ML-based classification model can provide a cost-effective and scalable solution for early detection [5]. Such a model could prioritize patients for further diagnostic workup, reducing the burden on over-stretched healthcare facilities and enabling timely interventions.

This paper focuses on a robust ML-based selective features for prediction of early detection of thyroid gland disorders in Nigeria, leveraging clinical data (TSH, T3, T4, autoantibodies), ultrasound findings, demographic variables (age, sex, BMI), and environmental factors (iodine status, goitrogen exposure).

Related works

[7] focused on Thyroid Disease Prediction based on Feature Selection and Machine Learning, in this paper, feature selection approach was used to eliminate certain irrelevant characteristics from the thyroid dataset (from the UCI machine learning repository) and to select optimal features. The dataset has three target classes named normal, hypothyroid, and hyperthyroid. The subjects were classified through seven different machine-learning algorithms. Random Forest classifier achieves the highest accuracy 99.58% which is better than the existing state-of-the-art methods. This study does not focus on risk factors for prediction of thyroid disorders in Nigeria which is the focus of this paper.

The study by [8] used deep learning approach for selecting of feature for thyroid disease classification .The study classified thyroid disease into three categories: hyperthyroidism, hypothyroidism and normal and obtained classification results are used for the diagnosis purposes. The study mainly concentrates on feature engineering and model optimization for deep learning. For getting better accuracy extra tree classifier based selected features are used for feature selection along with random forest classifier. As demonstrated by the results, the proposed system achieves relevant performances in terms of qualitative metrics for the thyroid nodule classification task, thus resulting in a great asset when formulating a diagnosis. K-fold validation technique along with F1 score corroborate the superior performance of the proposed. The study does not really search for best features to predict thyroid disease which is the focus of this paper.

[9] The study focused on a thyroid disease prediction approach which utilizes random forest-based features to obtain high accuracy. The approach can obtain a 0.99 accuracy to predict ten thyroid diseases. The study does not really search for best features to predict thyroid disease which is the focus of this paper.

[10] carried out a study to analyze the use of filter-based (F-Score) and wrapper-based (Recursive Feature Elimination) feature selection algorithms on its effect on disease identification and classification. The analysis was also performed with Principle Component Analysis dimensionality reduction algorithms. Performance evaluation was performed with three metrics, namely, accuracy, sensitivity and specificity. Four classifiers, namely, MultiLayer Perceptron, Back Propagation Neural Network, Support Vector Machine and Extreme Learning Machine were used to analyze the selected algorithms. Experimental results showed that while both F-Score and Recursive Feature Elimination improved the performance of thyroid disease diagnosis, the wrapper-based algorithm produced maximum efficiency and produced a maximum accuracy of 98.14% with ELM classifier. The study does not really search for best features to predict thyroid disease which is the focus of this paper.

[11] carried out study on the Thyroid disorders and the articles claimed that Thyroid disorders occur due to the malfunctioning of the thyroid gland, which may result in an imbalanced metabolic rate due to inappropriate hormone levels synthesis. An overactive gland results in hyperthyroidism, whereas an underactive or sluggish thyroid lead to hypothyroidism. Both disorders, if not detected and managed timely, can lead to severe health complications. Early identification is crucial to delay or avoid debilitating complications and achieve a better quality of life through the right medical interventions and precise hormonal readjustments. The proposed hybrid algorithm method finds the best features for finding thyroid disease uses performance measures such as accuracy, F1-score, precision, and recall. The research demonstrates promising results with an accuracy of 98.91 % and an F1-score of 94.83, showcasing the robustness of the proposed algorithms on a benchmark dataset. The findings hold potential to improve clinical decision-making processes. This study advances medical diagnostics by combining machine learning algorithms with nature-inspired optimization techniques to detect thyroid illnesses in their early stages. This article proposes a novel hybrid algorithm that combines the Cuttlefish Optimization Algorithm (CFA) and Simulated Annealing (SA) to find the best features for finding thyroid disease. The study uses machine-learning models for classification. The integration of machine learning and nature-inspired optimization significantly enhances the diagnostic capabilities of healthcare systems, enabling prompt diagnosis and treatment planning for thyroid disorders. The study does not really search for best features to predict thyroid disease which is the focus of this paper.

According to research carried out by [12], the study introduces a novel approach for thyroid prediction by considering three various publicly available datasets. The input data from the dataset is preprocessed to ensure standardization and balance for mitigating the biased outcomes. Then, the proposed cascaded autoencoder-based simple recurrent model is employed for extracting significant spatio-temporal features. From the extracted features, the optimal feature is selected using the proposed Opposition Learning-based Red Panda Optimization (OL_RPO) algorithm, which enhances the efficiency of the predictive model. Finally, the thyroid prediction is performed using the Enhanced Transformer Model, which uses the selected features to achieve robust and accurate predictions. The analysis of the proposed model based on Accuracy, Specificity, Sensitivity, F-Score, positive predictive value (PPV), negative predictive value (NPV), and Error acquired the value of 99, 99.2, 99.01, 98.501, 98.1, 1.9, and 0.07689 respectively. The study focused on thyroid prediction using deep learning which is quite different from selection of features for thyroid disorder.

METHODS

The used dataset for this research is sourced from a publicly available repository and contains records related to thyroid disorders diagnosis and treatment. The dataset consists of 3,772 patient records with 26 attributes, which include both numerical and categorical variables. The data was originally stored in structured format and later converted into a .csv file, where the first row specifies the attribute names, followed by rows representing data from each patient. The class label for this dataset is binaryClass, which indicates the presence or absence of thyroid disorders. Out of the 26 attributes, 25 were used as input variables, while one (binaryClass) served as the target variable for building predictive models. The dataset contains comprehensive information about thyroid-related conditions, including hormone levels (e.g., TSH, TT4, T4U), treatment status (e.g., on thyroxine, antithyroid medication), and related medical history (e.g., goitre, psych, hypopituitary). The attributes and their characteristics are summarized in Table 3.1, which provides three key columns, that is, the variables which are the names of the features included in the dataset., the data type: indicates whether the data is numerical or categorical and the description: outlines the meaning or state of each feature (e.g., 1 shows presence of a condition or characteristic, and 0 shows absence).

This dataset was used to explore the relationship between various clinical and biochemical features and thyroid dysfunction, supporting the development models for machine learning for early and accurate detection.

Table 3.1: Description of identified variables

Variables	Data Type	Description
Age	float64	Patient age in years (numerical)
Sex	float64	Gender of patient (0 = male, 1 = female)
on thyroxine	int64	Currently taking thyroxine (1 = yes, 0 = no)
query on thyroxine	int64	Suspected thyroxine use (1 = yes, 0 = no)
on antithyroid medication	int64	On antithyroid meds (1 = yes, 0 = no)
Ill	int64	ill patient (1 = yes, 0 = no)
pregnant	int64	Pregnant patient (1 = yes, 0 = no)
thyroid surgery	int64	History of thyroid surgery (1 = yes, 0 = no)
I131 treatment	int64	Received I131 treatment (1 = yes, 0 = no)
query hypothyroid	int64	Suspected hypothyroidism (1 = yes, 0 = no)
query hyperthyroid	int64	Suspected hyperthyroidism (1 = yes, 0 = no)
Lithium	int64	On lithium medication (1 = yes, 0 = no)
goitre	int64	Presence of goitre (1 = yes, 0 = no)
tumor	int64	Presence of tumor (1 = yes, 0 = no)
hypopituitary	int64	Hypopituitary condition (1 = yes, 0 = no)
psych	int64	Psychological disorders (1 = yes, 0 = no)
TSH measured	int64	TSH test done (1 = yes, 0 = no)
T3 measured	int64	T3 test done (1 = yes, 0 = no)
TT4 measured	int64	Total T4 test done (1 = yes, 0 = no)
TT4	float64	Total thyroxine level (µg/dL)
Diet	Int64	Balance Diet(1 = yes, 0 =no)
binaryClass	int64	Target variable (1 = thyroid disorders, 0 = normal)

Method of preprocessing of collected Data

Identification and loading of the thyroid disorders dataset and its description as represented in Table 3.1 was used to guide preprocessing. This stage is a critical aspect in the development of a robust models classification for machine learning, as it ensures that data that are fed into the algorithms are validated and appropriately scaled. During the preprocessing stage, missing values identified within the dataset were addressed using appropriate imputation techniques. This helped to minimize the effect of data sparsity and ensure consistency in the dataset without introducing significant distortion to the distribution of values. Imputation was

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

particularly crucial for biochemical features such as TSH, TT4, and FTI, which occasionally contained missing entries due to unmeasured test results. In addition, standardization was performed using the Standard Scaler utility from the scikit-learn library. This transformation process rescaled all numerical features to have a mean of 0 and a standard deviation of 1. Standardization was essential for improving the performance of machine learning algorithms used in this study, such as Random Forest and Gradient Boost, especially when features varied across different scales.

The tools employed for data preprocessing in this study include:

i. Data manipulation for Pandas and structure handling,

ii. Scikit-learn for scaling, transformation, and data splitting (via train_test_split),

iii. and NumPy for efficient numerical operations.

The preprocessing pipeline ensured that the input data was adequately prepared for feature selection and model training phases, leading to improved accuracy and interpretability of the thyroid classification models.

Method of feature selection

The identification of optimal feature subsets represents a critical component in developing robust classification models for thyroid gland disorders. This study employed a dual-pronged approach to feature selection, combining filter-based methods with Random Forest techniques to ensure comprehensive identification of the most predictive variables.

Filtering methods operate independently for learning algorithm, evaluating features based on their statistical properties and relationships target variable. These techniques offer computational efficiency and model-agnostic applicability, making them particularly suitable for initial feature screening in medical datasets where numerous clinical parameters may exhibit varying degrees of relevance. The implementation of filter-based selection began with analyzing correlation aim of identifying redundant features that might introduce multi-collinearity issues. Features demonstrating correlation coefficients exceeding 0.9 were flagged for potential removal. It also applies univariate statistical tests were employed to assess individual feature significance. For continuous variables, the F-statistic was calculated to measure the linear relationship between each feature and the thyroid disorders classification. Meanwhile, categorical variables underwent chi-square testing to evaluate their independence from the target classes. Information gain served as another crucial metric in the filter selection process. This measure quantifies the reduction in entropy achieved by partitioning the dataset based on specific feature values. Features yielding higher information gain values were prioritized, as they demonstrated superior discriminative power in distinguishing between different thyroid disorders categories. The mutual information score complemented this approach by capturing both linear and non-linear relationships between features and target variables.

This is a selection leverages the ensemble learning paradigm to provide robust importance rankings. Unlike filter methods, this approach considers feature interactions and non-linear relationships inherent in complex medical datasets. Its algorithm generates multiple decision trees utilizing bootstrap sampling and random feature subsets, thereby reducing overfitting while maintaining predictive accuracy. Feature importance in Random Forest is calculated through two primary mechanisms. First, the mean decrease in impurity measures how much each feature contributes to homogeneous node splits across all trees in the forest. Features consistently producing pure splits receive higher importance scores. Second, the permutation importance evaluates the deterioration values which are randomly shuffled, providing insight into each variable's contribution to predictive accuracy. The Random Forest selection process involved training an ensemble of 100 decision trees on the preprocessed thyroid disorders dataset. Random subset of features was used to constructed each tree, with the square root of the total feature count serving as the selection criterion. Bootstrap sampling ensured that each tree trained on a different subset of observations, enhancing robustness of importance estimates.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

The integration of filter and Random Forest selection methods created a comprehensive feature evaluation framework. Features identified as significant by both approaches were prioritized for inclusion in the final model. This dual validation approach reduces the risk of selecting spurious features while ensuring that truly predictive variables are retained.A ranking system was established to integrate results from both selection methods. Filter-based metrics were normalized to a common scale, and Random Forest importance scores were standardized using z-score transformation. The combined ranking considered both statistical significance and predictive contribution, creating a robust foundation for feature subset selection.

The final feature subset was determined through iterative evaluation, where different combinations of top-ranked features were tested using cross-validation. This process balanced model complexity with predictive performance, ensuring that the selected features provided optimal discrimination between thyroid disorders classes while maintaining computational efficiency.

RESULTS AND DISCUSSION

Exploration data analysis of thyroid disorders dataset results

This report presents a thyroid function set of data containing 3772 patient records. The dataset includes variables such as age, sex, thyroid hormone levels (TSH, TT4, FTI), clinical indicators (e.g., on thyroxine), a binary classification of thyroid condition (0 = negative, 1 = positive), and a dietary indicator (0 = no specific diet, 1 = specific diet). The analysis aims to explore the distribution of thyroid function metrics, their relationships with thyroid conditions, and the influence of diet. The dataset was cleaned to ensure analytical accuracy, numeric fields were converted to appropriate formats, and binary variables (sex, on thyroxine, binaryClass, diet) were encoded as 0 or 1 and records with missing or invalid numeric fields were excluded and mean imputation was used to replace other missing values. Table 4.1 presents the summary statistics for continuous and categorical variables.

Table 4.1: Summary statistics of the features in dataset

Feature	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Age	3772.0000	51.7359	20.0823	1.0	36.0	54.0	67.0	99.0
Sex	3772.0000	0.3153	0.4554	0.0	0.0	0.0	1.0	1.0
on thyroxine	3772.0000	0.1230	0.3285	0.0	0.0	0.0	0.0	1.0
query on thyroxine	3772.0000	0.0133	0.1144	0.0	0.0	0.0	0.0	1.0
medication	3772.0000	0.0114	0.1062	0.0	0.0	0.0	0.0	1.0
Sick	3772.0000	0.0390	0.1936	0.0	0.0	0.0	0.0	1.0
Pregnant	3772.0000	0.0141	0.1177	0.0	0.0	0.0	0.0	1.0
thyroid surgery	3772.0000	0.0141	0.1177	0.0	0.0	0.0	0.0	1.0
I131 treatment	3772.0000	0.0156	0.1241	0.0	0.0	0.0	0.0	1.0
query hypothyroid	3772.0000	0.0620	0.2413	0.0	0.0	0.0	0.0	1.0
TSH	3772.0000	5.0868	23.2909	0.005	0.6	1.6	3.8	530.0
TT4	3772.0000	108.3193	34.4965	2.0	89.0	106.0	123.0	430.0
T4U	3772.0000	0.9950	0.1852	0.25	0.89	0.995	1.07	2.32
FTI	3772.0000	110.4696	31.3551	2.0	94.0	110.0	121.3	395.0
Diet	3772.0000	0.5000	0.5001	0.0	0.0	0.5	1.0	1.0
binaryClass	3772.0000	0.0771	0.2669	0.0	0.0	0.0	0.0	1.0

The age distribution shown in Figure 4.1, exhibits a wide distribution, ranging from 1 to 99 years, with a mean of 51.74 and a standard deviation (SD) of 20.08, suggesting substantial variability across age groups. The interquartile range (IQR) spans from 36 (25th percentile) to 67 (75th percentile), indicating that the majority of subjects are middle-aged or older adults., while the sex variable is binary, encoded as 0 and 1, with a mean of 0.315, indicating that approximately 31.5% of the cohort belongs to the encoded category (likely males or females, depending on coding), while the remainder represents the opposite sex. Similarly, categorical indicators for clinical conditions such as on thyroxine (12.3%), query on thyroxine (1.3%), on antithyroid medication (1.1%), sick (3.9%), pregnant (1.4%), thyroid surgery (1.4%), and I131 treatment (1.6%)—are relatively infrequent, reflecting the rarity of these conditions in the sample population. The feature query hypothyroid shows a slightly higher prevalence at 6.2%.

Biochemical markers demonstrate substantial variation. TSH (Thyroid Stimulating Hormone) exhibits a markedly skewed distribution, with a mean of 5.09 and SD of 23.29, spanning an extreme range from 0.005 to 530.0. This suggests the presence of outliers or cases of severe thyroid dysfunction. Similarly, TT4 (Total Thyroxine) and FTI (Free Thyroxine Index) display means of 108.32 and 110.47, with relatively narrower spreads (SD = 34.50 and 31.36, respectively), though the maximum values (430.0 for TT4 and 395.0 for FTI) indicate potential outlier effects. T4U (Thyroxine Uptake) shows a more constrained distribution, with a mean near 1.0 and limited dispersion (SD = 0.185). The diet variable is evenly distributed (mean ≈ 0.50, SD = 0.50), indicating a balanced representation between the two diet categories. The target variable, denoting the classification outcome, has a mean of 0.077, implying that approximately 7.7% of the samples belong to the positive class, signifying an imbalanced dataset a crucial consideration for predictive modeling.

Handling class imbalance with SMOTE

A pronounced imbalance class, with the minority class accounting for less than 10% of observations. This risk of imbalance posed a class bias toward majority during model training. In addressing this, the Synthetic Minority Over-sampling Technique was

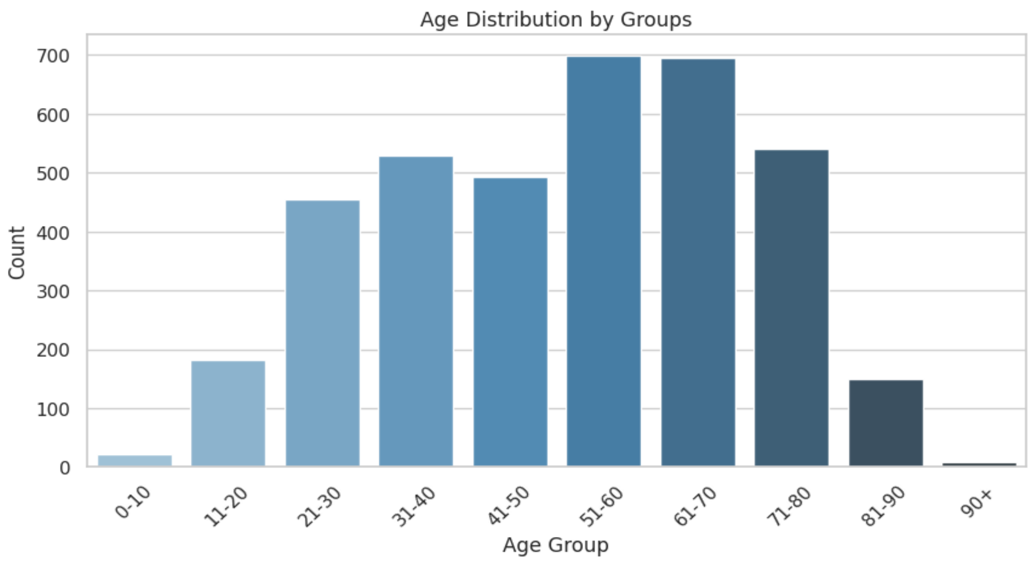

Figure 4.1 Age distribution of the dataset

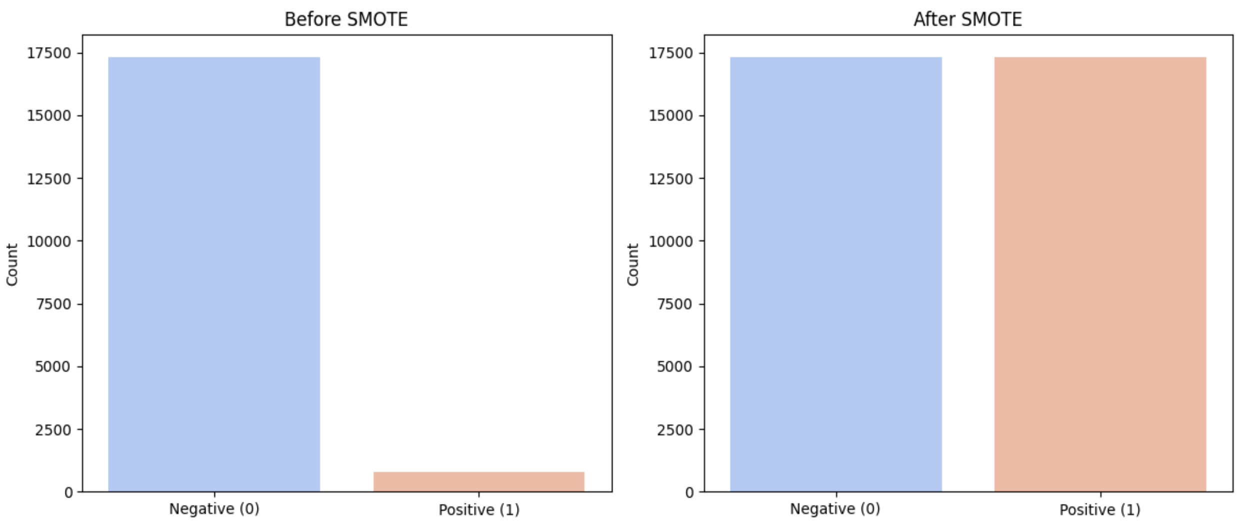

Figure 4.2: Class distribution before and after applying SMOTE

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

SMOTE generates synthetic examples for the minority class by interpolating between existing samples in feature space, thereby reducing overfitting associated with random oversampling. Figure 4.2, shows the distribution of classes before and after applying SMOTE. The transformation produced an approximately balanced dataset, ensuring both classes contribute equally to the learning process. This adjustment is expected to enhance model sensitivity and reduce misclassification of minority cases.

Results of feature Correlation Analysis

The correlation heat map (Figure 4.3) illustrates the pairwise linear relationships among dataset variables. Overall, most features exhibit weak correlations, indicating minimal multicollinearity—an advantageous property for predictive modeling. Notably, a distinct correlation cluster exists among thyroid-related biochemical markers, such as TT4, T4U, and FTI, which show strong positive associations consistent with their physiological interdependence. Similarly, measured features like T4U measured and FTI measured display strong correlations with their corresponding computed indices, reflecting measurement-based dependencies. Demographic and clinical variables—including age, sex, and treatment indicators (e.g., on thyroxine, I131 treatment)—exhibit negligible correlations with biochemical markers, suggesting that these factors contribute independently to thyroid function prediction.

The diet variable, representing malnutrition status, demonstrates near-zero correlation with both other predictors and the target variable (binaryClass), which encodes thyroid disorders status. This lack of linear association implies that malnutrition does not strongly influence thyroid disorders classification in a direct linear manner within this dataset. However, given the established clinical relevance of nutritional deficiencies in thyroid hormone regulation, the variable was retained for modeling to capture potential non-linear or interaction effects that correlation analysis cannot reveal.

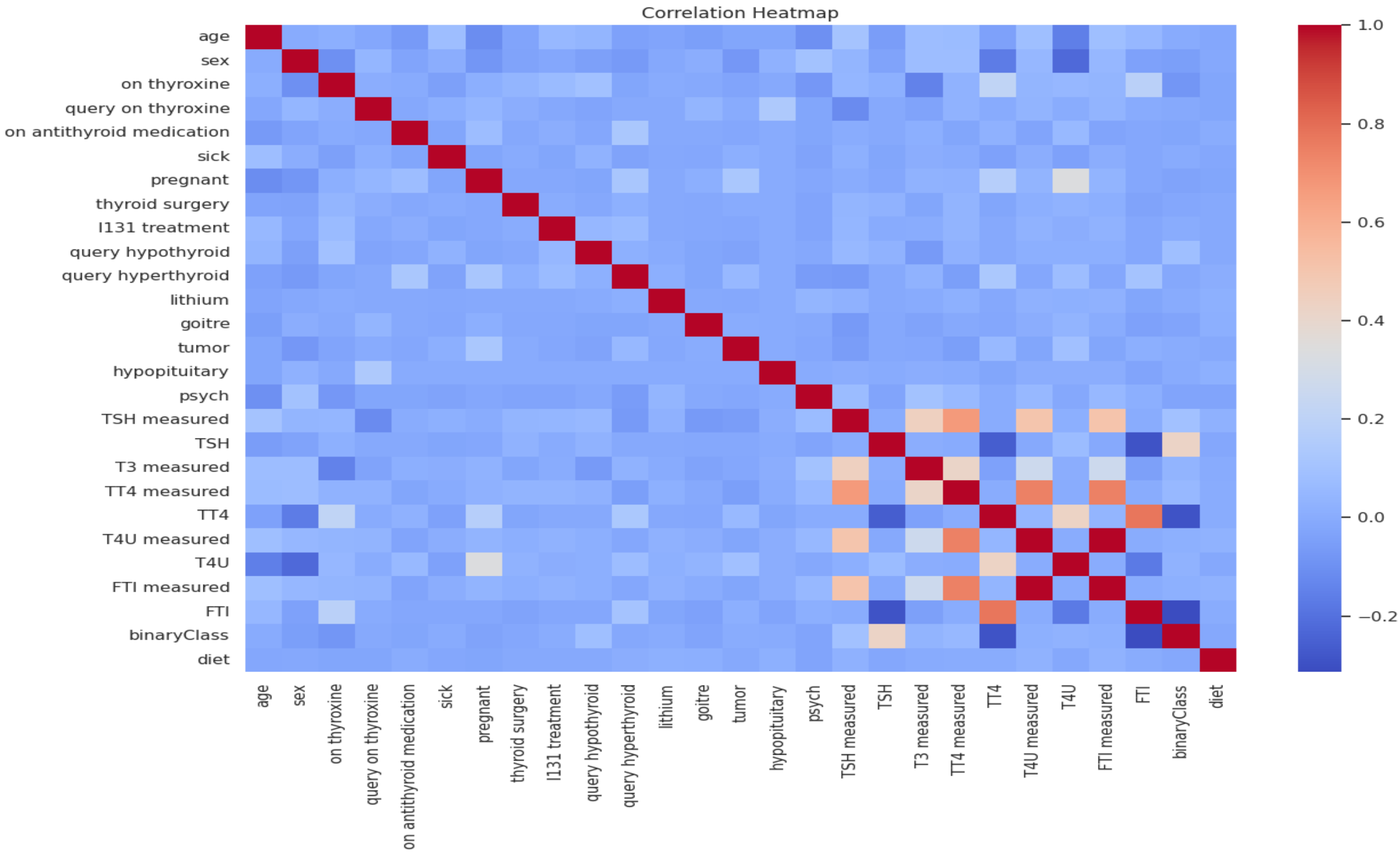

Figure 4.3: Correlation heatmap of the features

Finally, binaryClass shows only weak associations with individual predictors, confirming the absence of a single dominant linear predictor and highlighting the necessity for multivariate modeling approaches to achieve robust classification performance.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

Result of features selections

To construct a robust predictive model for thyroid condition classification (TargetClass: 0 = negative, 1 = positive), a hybrid feature selection approach was employed, integrating filter-based correlation analysis and embedded Random Forest importance scoring. This methodology aimed to distill a compact yet informative subset of features from the thyroid dataset, comprising 3772 patient records. Subsequently, four machine learning models Logistic Regression, Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP) Neural Network were evaluated using 5-fold stratified cross-validation. The feature selection process combined two complementary techniques to ensure robustness and relevance with Table 4.2 showing the results of the features selection

- **Filter method (correlation with target)**: Absolute Pearson correlations between each feature and the TargetClass target were computed. Features with correlations ≥ 0.05 were retained, yielding TSH measured, query hypothyroid, on thyroxine, TT4 measured, and sex. This method prioritized features with direct linear associations to thyroid condition.

- **Embedded method (random forest importance)**: A Random Forest classifier assessed feature importance, selecting the top 10 features: age, T4U, diet, sex, T3 measured, query hypothyroid, on thyroxine, sick, query hyperthyroid, and tumor. This approach captured non-linear and interactive effects.

Table 4.2: Results of feature selections

Feature	Filter-based	RF-based	Final Selection
Age	TSH measured	age	sex
Sex	query hypothyroid	T4U	query hypothyroid
Thyroxine on	Thyroxine on	diet	Thyroxine on
Thyroxine query on	TT4 measured	sex	
medication antithyroid on	sex	T3 measured	
Sick		query hypothyroid	
Pregnant		on thyroxine	
surgery thyroid		sick	
I131 treatment		Hyperthyroid query	
Hypothyroid query		tumor	
Hyperthyroid query			
Lithium			
Goiter			
Tumor			
Hypopituitary			
Psych			
measured TSH			
measured T3			
measured TT4			
TT4			
Diet			

Hybrid selection: The intersection of filter and embedded methods produced a final feature set: sex, query hypothyroid, and on thyroxine. This conservative strategy ensured only features with consistent predictive power across both methods were selected.

Results: performance metrics of models with feature selections

Machine learning performance algorithms employed to develop a predictive classification model for the early detection of thyroid disorders. The models Random Forest, Gradient Boosting, Multi-Layer Perceptron(MLP) Neural Network, and Logistic Regression were trained on feature subsets derived from Filter-based (Mutual Information, MI), Random Forest(RF)-based, and Hybrid Selected Features (intersection of Filter and RF-based: sex, query hypothyroid, on thyroxine) approaches. Performance is evaluated using accuracy, F1-score, and ROC-AUC, providing a multifaceted assessment of model efficacy in classifying thyroid conditions (binaryClass:0 = negative, 1 = positive) within a dataset of 3772 patient records.

Results of performance metrics of models with filter-based features

The Filter-based approach, utilizing Mutual Information(MI), prioritized features with robust statistical associations to thyroid condition, such as TSH measured and query hypo thyroid. As shown in Table 4.3 and illustrated in Figure 4.4, Random Forest excelled with an accuracy of0.9986, precision of 0.9971, recall of 1.0000, F1-score of 0.9986, and ROC-AUCof0.9999, demonstrating unparalleled proficiency in distinguishing thyroid disorders. The perfect recall (1.0000) underscores its ability to identify all positive cases, critical for early detection. Gradient Boosting followed closely, achieving an accuracy of 0.9957, precision of 0.9986, recall of 0.9928, F1-score of 0.9956 ,and ROC-AUC of 0.9985,reflecting robust performance. The MLP Neural Network recorded an accuracy of 0.9964.

Table 4.3: `Result of models with filter-based features

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**
Random Forest	0.9986	0.9971	1.0000	0.9986	0.9999
MLP Neural Network	0.9964	0.9986	0.9928	0.9964	0.9988
Gradient Boosting	0.9957	0.9957	0.9971	0.9957	0.9985
Logistic Regression	0.9742	0.9810	0.9670	0.9740	0.9948

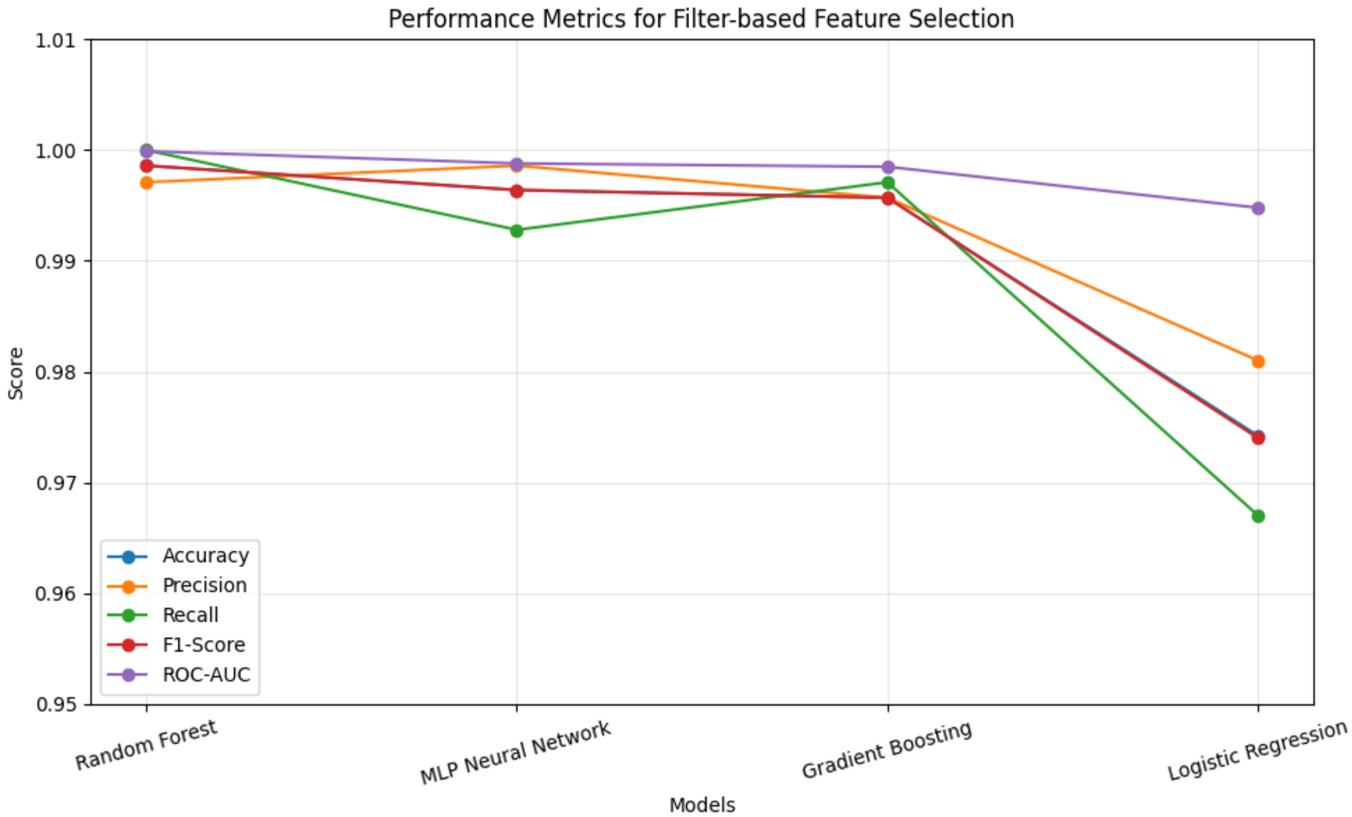

Figure 4.4: Performance Metrics of Filter Based Feature Selection

0.9957 for Precision, 0.9971 for recall, 0.9964 for F1-score and 0.9988 for ROC-AUC, show casing its non-linear modeling prowess. Logistic Regression, despite its linear nature, achieved 0.9742 for accuracy, 0.9810 for precision, 0.9670 for recall, 0.9739 for F1-score, and 0.9948 for ROC-AUC indicating solid discriminativ e power with 0.9957 accuracy, an 0.9957 for F1-score, and 0.9985 for ROC-AUC of, reflecting its effective optimization of the feature space. Logistic Regression, despite its linear framework, achieved a respectable accuracy of 0.9742, an F1-score of 0.9740, and an ROC-AUC of 0.9948, highlighting its discriminative power when supported by MI-selected features. Figure 4.5 displays the ROC-AUC curves, with Random Forest's near-unity score (0.9999) affirming its superior class separation. High precision and recall across models suggest effective handling of the dataset's imbalance, likely due to MI's selection of informative features. Random Forest's exceptional performance positions it as a prime candidate for clinical thyroid diagnostics.

Results of performance metrics of models with RF-Based features

The RF-based approach, leveraging Random Forest importance scores, selected features like age, T4U, and diet. Table 4.4 and Figure 4.6 reveal Random Forest leading with an accuracy of 0.9978, precision of 0.9986,recall of 0.9971, F1-score of 0.9978,and ROC-AUC of 0.9999, showcasing its adeptness at capturing complex feature interactions. Gradient Boosting followed with an accuracy, precision, re-call, and F1-score of 0.9971 and an ROC-AUC of 0.9990, reflecting near-equivalent performance. The MLP Neural Network achieved an accuracy of 0.9943, precision of 0.9957, recall of 0.9928, F1-score of 0.9942, and ROC-AUC of 0.9971, maintaining strong predictive power. Logistic Regression recorded an accuracy and precision of 0.9799 ,recall of 0.9799,F1-score of 0.9798, and ROC-AUC of 0.9932, demonstrating robust linear performance.

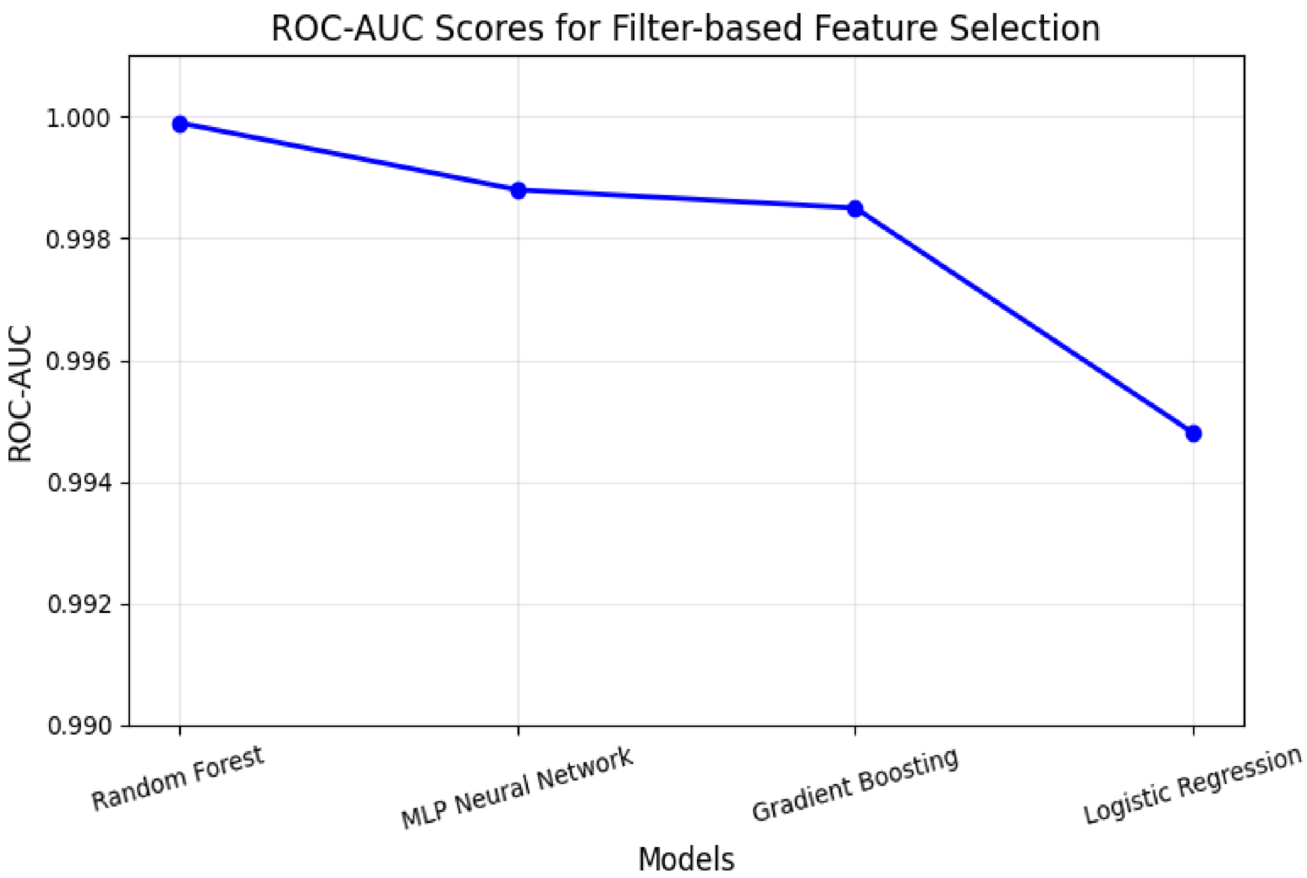

Figure 4.5: ROC AUC curve of filter based feature selection

Table 4.4: Result of models with RF-Based features

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random Forest	0.9986	0.9971	1.0000	0.9986	0.9999
MLP Neural Network	0.9964	0.9986	0.9928	0.9964	0.9988
Gradient Boosting	0.9957	0.9957	0.9971	0.9957	0.9985
Logistic Regression	0.9742	0.9810	0.9670	0.9740	0.9948

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue III March 2026

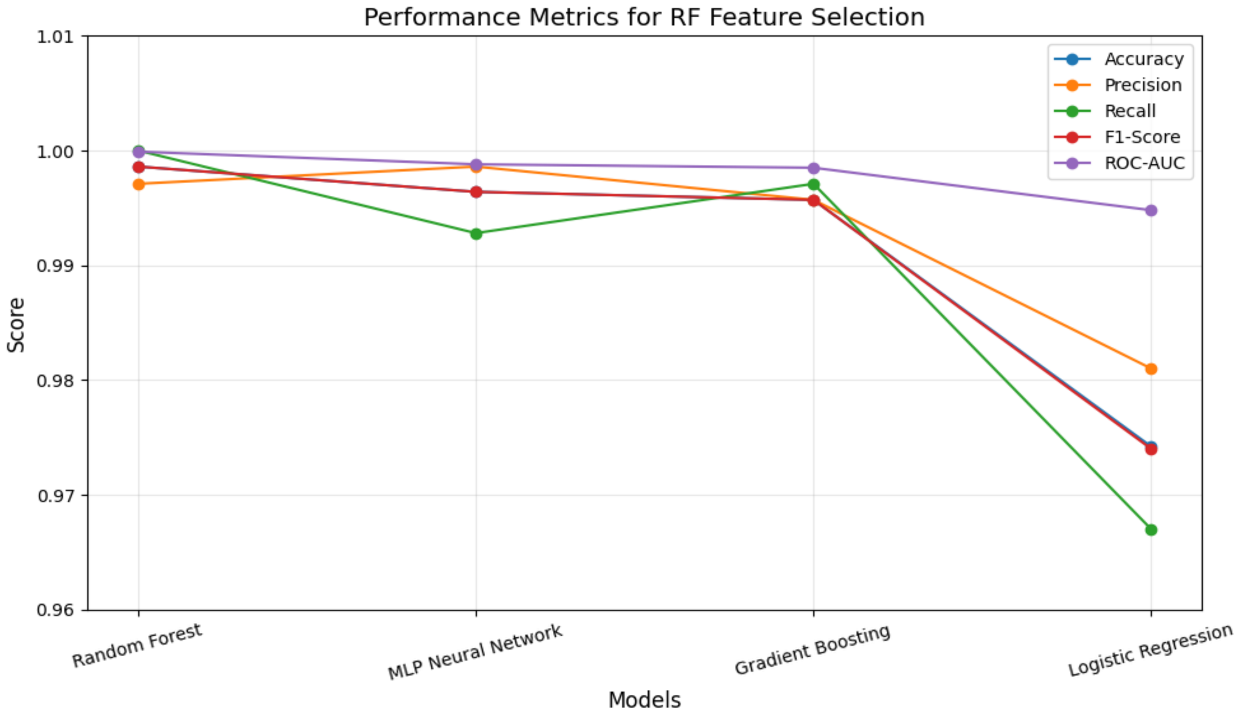

Figure 4.6: Performance metrics result of RF feature selection

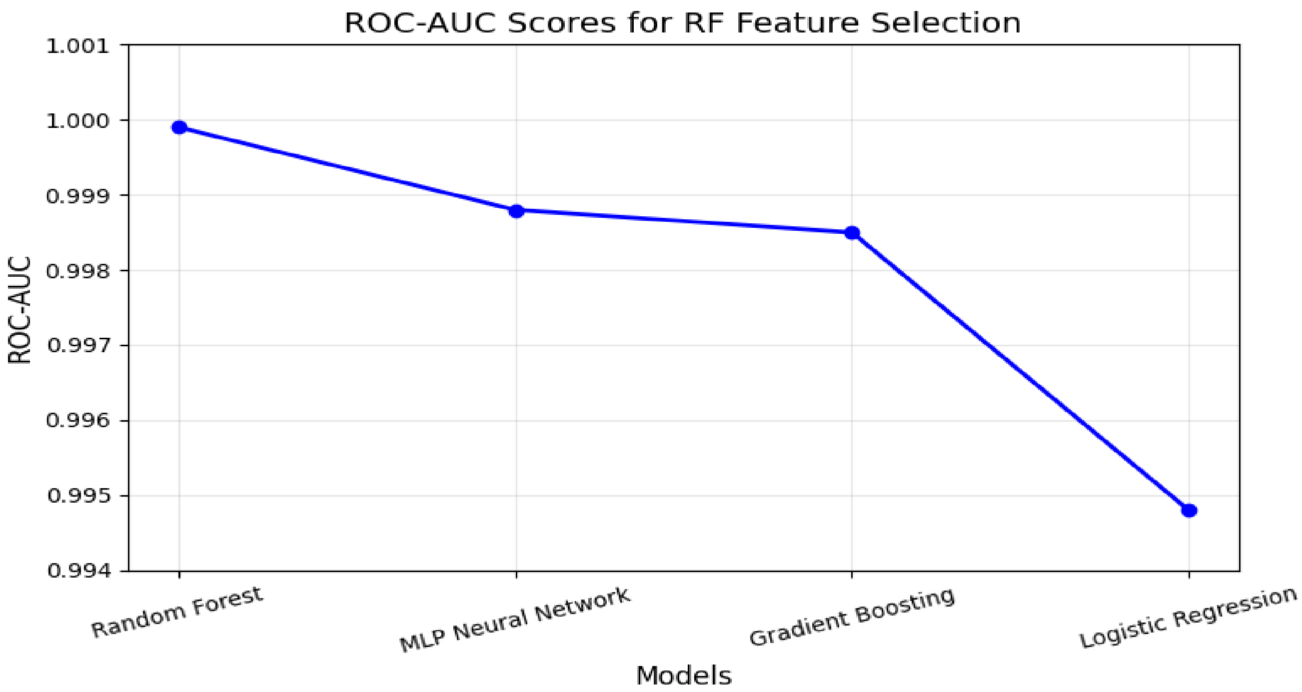

Figure 4.7: ROC AUC curve of RF feature selection

Figure 4.7, illustrates the ROC-AU curves, with Random Forest and Gradient Boosting achieving near-perfect scores. The high recall values, particularly Random Forest's 0.9971, highlight their sensitivity to positive cases, crucial for early thyroid detection.

Results of performance metrics of models with hybrid selected features

The hybrid selected features sex, query hypothyroid, and on thyroxine represent the intersection of the Filter and RF-based methods, prioritizing clinical relevance. As shown in Table 4.5, performance declined significantly. Random Forest and Gradient Boosting tied with an accuracy of 0.6863, precision of 0.6328, recall of 0.8865, F1-score of 0.7385, and ROC-AUCs of 0.7824 and 0.7825, respectively. The high recall (0.8865) indicates strong sensitivity to positive cases, but low precision suggests many false positives. The MLP Neural Network recorded an accuracy of 0.6648, precision of 0.6151, recall of 0.8793, F1-score of 0.7238, and ROC-AUC of 0.7548. Logistic Regression performed poorest, with an accuracy of 0.5951, precision of 0.5737, recall of 0.7385, F1-score of 0.6457, and ROC-AUC of 0.6019.

Results of comparison feature selection strategies

Figure 4.8, compares the predictive performance of different feature selection strategies using Random Forest across three core metrics: Accuracy, F1-score, and ROC-AUC. The results demonstrate a clear performance gradient among the feature selection approaches. Models utilizing All Features and RF-based Features exhibit nearly identical metrics, with Accuracy and F1-score approaching 0.998 and ROC-AUC nearly perfect at 0.9999. Similarly, Filter-based Feature Selection delivers competitive performance, slightly outperforming other strategies in certain metrics (Accuracy and F1-score ≈ 0.9986). These findings indicate that both filter and embedded methods preserve essential discriminatory information necessary for robust classification. Conversely, the Final Selected Features (intersection method) display a sharp decline in performance across all metrics (Accuracy = 0.6863, F1-score = 0.7385, ROC-AUC = 0.7824). This dramatic reduction underscores the limitations of using a minimal consensus feature set in complex clinical prediction tasks. While interpretability improves with fewer variables, the removal of high-variance, physiologically relevant attributes (e.g., TSH, TT4) restricts the model's ability to capture nonlinear relationships, leading to underfitting.

Table 4.5: Performance metric of hybrid feature selection

Models	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0.5951	0.5737	0.7385	0.6457	0.6019
Random Forest	0.6863	0.6328	0.8865	0.7385	0.7824
Gradient Boosting	0.6863	0.6328	0.8865	0.7385	0.7825
MLP Neural Network	0.6648	0.6151	0.8793	0.7238	0.7548

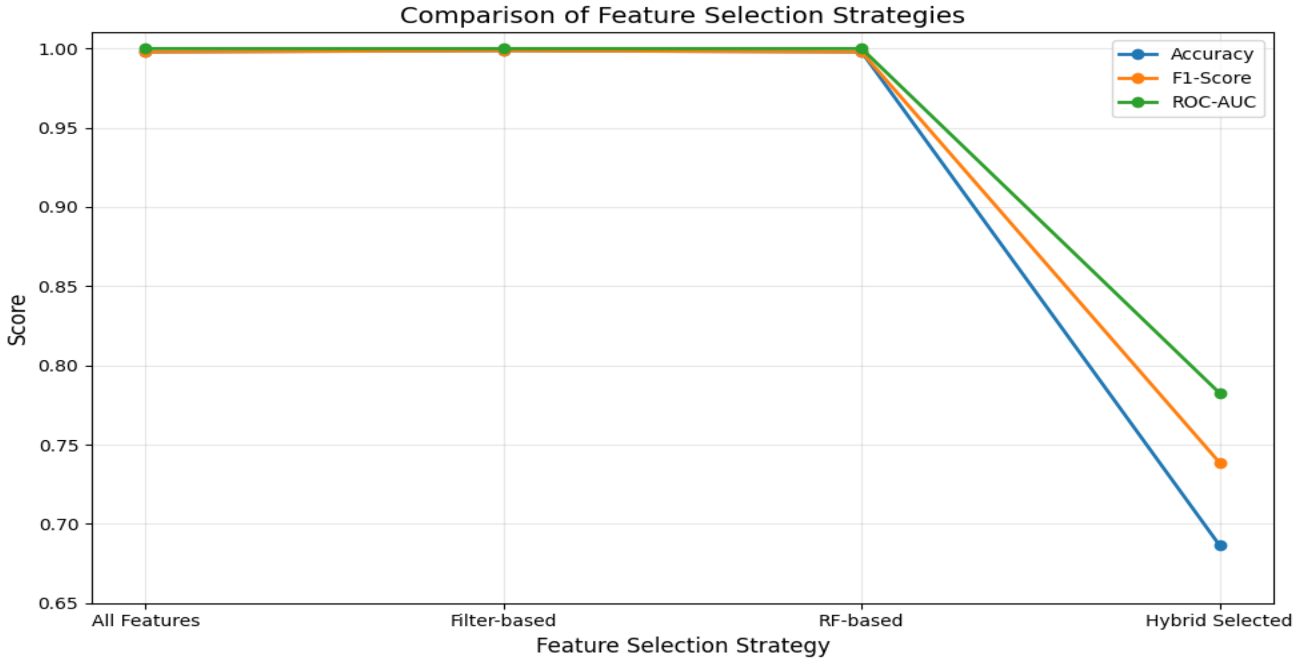

Figure 4.8: Comparison feature selection results

Overall, the analysis highlights a fundamental trade-off between dimensionality reduction and predictive accuracy. For high-stakes medical applications, strategies such as filter-based or RF-based selection, which balance efficiency with informational richness, are preferable over strict intersection-based approaches.

DISCUSSION OF RESULTS

The analysis of the thyroid disorders dataset, consisting of 3,772 patient records, offers critical insights into the interplay of demographic, clinical, and biochemical features in developing predictive models for early thyroid dysfunction detection. This section synthesizes the findings from Exploratory Data Analysis (EDA), feature selection strategies, and model evaluation, highlighting their implications for predictive accuracy and clinical applicability. The dataset exhibited significant class imbalance, with the positive class (thyroid disorders) representing only 7.7% of the observations. This imbalance posed a risk of bias toward the majority class, potentially leading to poor sensitivity for detecting true positive cases. To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied, generating synthetic minority samples and achieving a balanced distribution. This preprocessing step proved essential in improving model sensitivity, as reflected in higher recall and F1-scores across models. The demographic analysis revealed a broad age range (1–99 years) with a mean of 51.74 (SD = 20.08), emphasizing the need for models robust to age-related variability. Clinical features such as on thyroxine (12.3%) and antithyroid medication (1.1%) were sparsely distributed, indicating their limited standalone predictive power but potential utility in interaction effects. Biochemical markers demonstrated high variability, with TSH showing extreme skewness (mean = 5.09, max = 530.0), underscoring the importance of normalization to prevent outlier bias. TT4 and FTI exhibited moderate dispersion but included high-end outliers, reinforcing the need for preprocessing to ensure model stability.

A hybrid feature selection strategy integrating filter-based correlation analysis and Random Forest (RF) feature importance was explored alongside alternative approaches (all features, filter-based only, RF-based only). The hybrid method produced a minimal feature set (sex, query hypothyroid, on thyroxine), prioritizing interpretability but significantly reducing predictive power (Random Forest accuracy = 0.6863, ROC-AUC = 0.7824). Conversely, models utilizing all features or those selected via filter-based or RF-based strategies achieved near-perfect performance (Accuracy ≈ 0.998, ROC-AUC ≈ 0.9999). These results highlight a critical trade-off: while feature reduction enhances interpretability, it can omit physiologically relevant attributes (e.g., TSH, TT4), undermining classification accuracy. In the model performance analysis, four machine learning models of Logistic Regression, Random Forest, Gradient Boosting, and Multi-Layer Perceptron (MLP) Neural Network were evaluated across the chosen feature sets. The performance hierarchy was consistent across metrics:

- Random Forest and Gradient Boosting delivered superior results, with Random Forest slightly outperforming Gradient Boosting. Using all features, Random Forest achieved accuracy = 0.9978, precision = 0.9986, recall = 0.9971, F1-score = 0.9978, and ROC-AUC = 0.9999, indicating near-perfect discrimination. Gradient Boosting closely followed with similar metrics (accuracy = 0.9971, ROC-AUC = 0.9999).

- The MLP Neural Network demonstrated strong but slightly lower performance (accuracy = 0.9856, ROC-AUC = 0.9936), reflecting its ability to model complex non-linear relationships while being more sensitive to hyper parameter tuning and computational cost.

- Logistic Regression, despite being a linear model, achieved commendable results (accuracy = 0.9727, ROC-AUC = 0.9907), validating its role as a reliable baseline for interpretable modeling.

The confusion matrix further affirmed the models' superiority, with minimal false positives and negatives a critical advantage in clinical applications where diagnostic errors can have severe consequences. The results underscore the clinical relevance of Random Forest and Gradient Boosting models for thyroid disorders detection. Their robustness in handling heterogeneous features and class imbalance positions them as the preferred choice for deployment in diagnostic systems. However, interpretability remains a concern, particularly in clinical settings requiring transparency. While feature selection can improve interpretability, excessive reduction such as in the hybrid approach compromises predictive fidelity. Thus, approaches that balance dimensionality reduction and informational richness with filter-based or RF-based methods in this case, are recommended for real-world applications.

CONCLUSION

In conclusion, the comparative analysis confirms that Random Forest and Gradient Boosting offer the most reliable and accurate predictions, benefiting from their ensemble architecture and ability to model complex interactions. While MLP Neural Network and Logistic Regression remain viable alternatives, especially where interpretability or computational simplicity is prioritized, their performance is slightly inferior. These findings advocate for the adoption of ensemble techniques in clinical decision support systems for thyroid diagnostics, leveraging their superior accuracy, precision, and robustness to support early detection and improve patient outcomes.

The study identified key predictors (e.g., TSH, query hypothyroid, on thyroxine) tailored to the Nigerian dataset, emphasizing the importance of context-specific feature selection. The limited linear correlation of the diet variable suggests potential non-linear interactions relevant to Nigeria's nutritional landscape, warranting further exploration.

REFERENCES

1. Z. W. Baloch, S. L. Asa, J. Barletta, R. A. Ghossein, and O. Mete, "The 2022 WHO classification of thyroid tumors," Endocrine-Related Cancer, vol. 29, no. 12, pp. 133–150, 2022.
2. S. A. Kareem, A. A. Adeyemo, and O. A. Ojo, "The pattern of thyroid cancers in Nigeria," Indian J. Surg. Oncol., vol. 15, no. 2, pp. 245–253, 2024.
3. A. O. Ogbera, C. N. Okoro, and O. O. Balogun, "Epidemiology of thyroid disorders in Nigeria," African Health Sciences, vol. 23, no. 4, pp. 89–97, 2023.
4. A. O. Afolabi, T. A. Oluwasola, and A. M. Adebayo, "Iodine deficiency and thyroid disorders in northern Nigeria," African J. Endocrinol. Metab., vol. 14, no. 3, pp. 112–119, 2022.
5. M. A. Yusuf, A. B. Ibrahim, and S. Mohammed, "Iodized salt consumption and thyroid health in Nigeria," J. Public Health Africa, vol. 1 4, no. 7, pp. 56–63, 2023.
6. T. Alyas, J. A. Qazi, Y. Alsaawy, and M. Alshehri, "Empirical method for thyroid disease classification using ML," BioMed Res. Int., vol. 20, no. 1, pp. 34–56, 2022.
7. Z. Peya, J. M. S. Islam, and M. K. N. Chumki, "Thyroid Disease Prediction based on Feature Selection," in Proc. 25th Int. Conf. Computer and Information Technology (ICCIT), 2022, pp. 495–500.
8. H. R. Abhishek, A. Mura, B. J. Mathews, and A. T. Sashidharan, "Selective Feature Based Thyroid Disease Classification Using Deep Learning," Int. J. Eng. Res. Technol., 2023.
9. 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," Cancers (Basel), vol. 14, no. 16, p. 3914, 2022.
10. K. Pavya and B. Srinivasan, "Feature selection algorithms to improve thyroid disease diagnosis," in Proc. Int. Conf. Innovations in Green Energy and Healthcare Technologies (IGEHT), 2017, pp. 1–5.
11. K. Shrivastava, S. Pandey, R. Dubey, M. Namdev, V. Tiwari, and A. Sharma, "A novel hybrid approach for thyroid disease detection," MethodsX, vol. 15, p. 103558, 2025.
12. K. H. Priya and K. Valarmathi, "Deep learning based thyroid prediction with Red Panda Optimization," Sci. Rep., vol. 16, p. 2993, 2026.
13. Badridinova, Azimova, Iskandarova, Majidova, Abdullaev, Urinov, and Tokhirova, "Early detection of thyroid disease using hybrid ML," Health Leadership and Quality of Life, vol. 3, pp