Performance Optimization of Machine Learning Models for Thyroid Cancer Recurrence

Supervised By

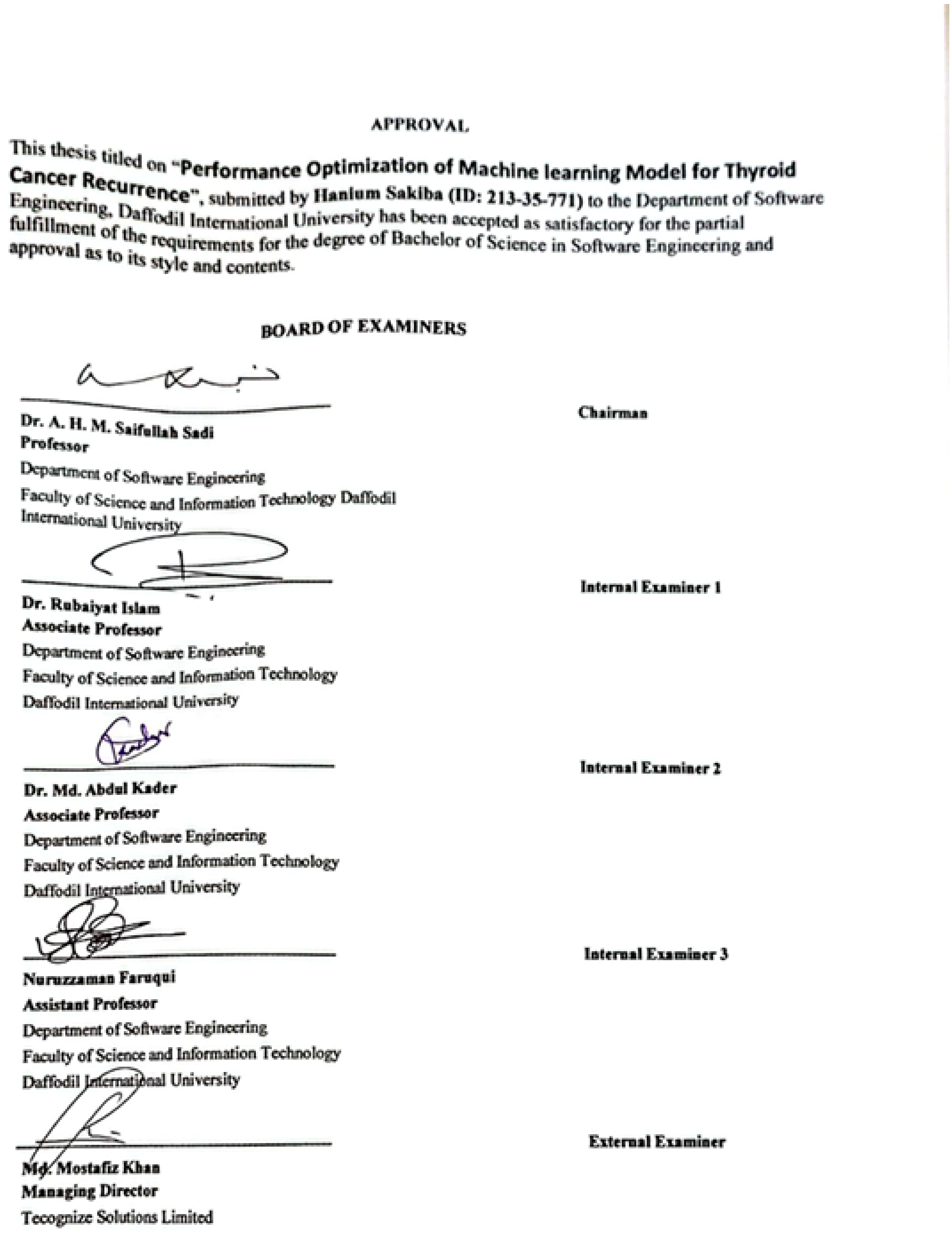
Dr. Nusrat Jahan
Associate Professor and Head
Department of Information Technology & Management
Daffodil International University

Submitted By

Hanium Sakiba
ID:213-35-771
Department of Software Engineering
Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL



APPROVAL

This thesis titled on **"Performance Optimization of Machine learning Model for Thyroid Cancer Recurrence"**, submitted by **Hanium Sakiba (ID: 213-35-771)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Dr. A. H. M. Saifullah Sadi
Professor
Department of Software Engineering
Faculty of Science and Information Technology Daffodil International University

Chairman

Dr. Rubaiyat Islam
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Dr. Md. Abdul Kader
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Nuruzzaman Faruqui
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Md. Mostafiz Khan
Managing Director
Tecognize Solutions Limited

External Examiner

Performance Optimization of Machine Learning Models for Thyroid Cancer Recurrence

Hanium Sakiba

213-35-771

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

SUPERVISOR’S DECLARATION

I hereby declare that I have reviewed this thesis entitled **" Performance Optimization of Machine Learning Models for Thyroid Cancer Recurrence"**, and in my opinion, it is adequate in terms of scope and quality for the award of the degree of Bachelor of Science in Software Engineering.

(Supervisor’s Signature)
Full Name : Dr. Nusrat Jahan
Position : Associate Professor and Head
Date : 23 December 2025

STUDENT'S DECLARATION

I confirm that the piece in this thesis is based on my own writing with the exception of quotation and reference that have been discussed. I also confirm that it was not previously and concurrently registered at Daffodil International University or other institutions at any other degree.

Sakiba

(Student's Signature)

Full Name : Hanium Sakiba

ID Number : 213-35-771

Date : 23 December2025

Performance Optimization of Machine Learning Models for Thyroid Cancer Recurrence

Hanium Sakiba
213-35-771

Thesis submitted in fulfilment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

DECEMBER 2025

ACKNOWLEDGEMENTS

I would like to thank Almighty God for His help and for granting me the capabilities and proficiency in order to perform this research successfully. I am sincerely thankful to Dr. Nusrat Jahan, Associate Professor & HOD of Department of Information Technology and Management for the continuous encouragement, guidance, suggestions and support during the research work. Without her excellent ideas, patience and inspiration this thesis would not have been completed. I additionally thank the faculty and staff of the department for their academic support and provision of an intellectually challenging research environment. I would also like to express my gratitude to my friends and colleagues, who have been so supportive both professionally and personally, at work and in pleasure time. Most especially, my heartfelt thanks and love to my dear parents whose unending prayers, offerings and blessings have become the very foundation of all that I am today. I also want to thank my relatives for their continual support, tolerance and inspiration during this educational process. I couldn’t have done this without their emotional and moral support. Last but not least, I would like to thank Allah Almighty once again for giving me the persistence, patience and also strength to successfully complete this study.

DEDICATION

I dedicate this project to my dear Father and mother, my head of department, My Honorable teachers always expensive and near to me. No, it was not possible to be here without their patience, understanding, unsparing support and strengthening buttress of care, love and cherish.

ABSTRACT

Thyroid disorders are among the most common hormonal dysfunctions across the globe, involving metabolic and hormone activity. Early and precise diagnosis is of great significance for optimal treatment and prevention of severe complications. Most of the conventional diagnostic methods depend on visual interpretation of laboratory results and therefore can be delayed or inaccurate. For overcoming these challenges, we have introduced a hybrid machine learning model ThyroidNet-RF that incorporates an interpretable Random Forest classifier and deep learning-based feature extraction for accurate and reliable thyroid disease diagnosis. It is based on a systematic series of steps comprising data cleaning, preprocessing, correlation analysis and feature selection with the Boruta algorithm. The class imbalance is addressed with the SMOTE-Tomek method during the balancing of the datasets. Several machine learning models including DT, KNN, SVM, ANN and XGBoost were developed as well as compared with ThyroidNet-RF model. The model showed highly competitive classification performance with an accuracy of 99.87%, precision of 1.00, recall of 0.98, F1-score of 0.99 and AUC of 1.00 when compared to other classifiers Results show that ThyroidNet-RF is able to improve the classification accuracy, reliability and interpretability efficiently, which makes it a useful decision-support tool for the clinicians. Results This research demonstrates the potential for assembling with deep learning in medical diagnosis and improved opportunities for early detection and individualized treatment decisions to enhance patient outcomes.

Keywords: Thyroid Cancer, Machine Learning, thyroid prediction, SMOTE, machine learning, Feature selection, Healthcare AI, Hypothyroidism.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AUC	Area Under the Curve
CSV	Comma-Separated Values
DT	Decision Tree
FTI	Free Thyroxine Index
FN	False Negative
KNN	K-Nearest Neighbor
MLP	Multi-Layer Perceptron
SMOTE	Synthetic Minority Over-sampling Technique
TSH	Thyroid Stimulating Hormone
WHO	World Health Organization
T4U	Thyroxine Uptake
XGBoost	Extreme Gradient Boosting
TT4	Total Thyroxine

CHAPTER 1

INTRODUCTION

1.1 Introduction

Thyroid disease is one of the most common forms of endocrine disorders globally, affecting the metabolic and hormonal status of millions of individuals. The thyroid gland, a small organ in the cervical region [1], has a vital function in the maintenance of important physiological processes by secreting hormones. The World Health Organization (WHO) estimates that iodine deficiency is the primary cause of thyroid disorders in many parts of the world, affecting over 750 million people. Thyroid dysfunction is five to eight times more common in women than in men, making them especially at risk [2]. In developing nations, the high burden of undetected cases which frequently result in complications like goiter, infertility, cardiovascular diseases, and mental health disorders is a result of insufficient screening and a lack of awareness. Hypothyroidism may have profound implications on physical and mental well-being [3] however, thyroid diseases are not infrequently undiagnosed because they may present with insidious or superimposed clinical features. Early identification is an essential prerequisite of proper treatment and improved patient outcomes. Predictive modeling is a crucial tool used in enhancing early diagnosis because machine learning and data-driven medicine have placed it at the forefront. This paper aims to forecast and categorize thyroid illness by means of complex medical database which consists of multiple attributes, including patient demographics (age, sex), pharmacotherapy (thyroxine intake, antithyroid medicine), clinical conditions (illness, pregnancy, goiter, tumor), hormonal assays (TSH, T3, TT4, T4U, FTI, TBG), and antecedent indicators (thyroidectomy, I131 treatment).The dependent variable is a binary variable indicating the existence or the absence of thyroid disease. This study seeks to come up with a transparent and precise model through supervised machine-learning algorithms with the end goal of helping clinicians to recognize the risky patients. Besides, a combination of different physiological, pathological, and biochemical markers allows the unified method of disease detection. The study will be included in the existing literature on medical artificial intelligence and will highlight the use of predictive analytics in clinical decision-making and customization of treatment to thyroid disease based on empirical research. The paper structure like in the second phase we present related works. Section 3 showcase the methodology of research; Section 4 visualize and analyze result and findings. Section 5 conclusion with future works.

1.2 Motivation

Thyroid cancer is one of the fastest growing cancer types worldwide and early detection of recurrence could dramatically improve patient care. Conventional diagnosis is extremely slow and not very accurate. Tools such as machine learning allow us to analyze complex medical data in a way that can effectively predict recurrence of cancer. Greater accuracy in prediction can prevent unnecessary treatments while improving patient survival. In addition, although it can assist healthcare providers to make rapid and accurate decisions. Based on this motivation, the next section describes the necessity to incorporate improving AI methods into medical diagnosis to have a positive effect on clinical practice.

1.3 Research Object

The objective of this research is to apply optimized machine learning models for predicting, and potentially, preventing thyroid cancer recurrence. The goal is to enhance model precision; help find the critical components responsible for the recurrence; and create a dependable system that can aid doctors by providing a guideline for early diagnosis and treatment course. Models' interpretability has also been studied in order to remain clinically interpretable. It also seeks to compare algorithms to identify the best and most robust model suitable for clinical translation. In addition, the research will also aim to develop a high-throughput pipeline that has the potential to be extended to other types of cancer prediction.

1.4 Research Question

- In what way can prediction of Thyroid cancer recurrence be made using accurate machine learning models?
- What characteristics or parameters best predict thyroid cancer relapse?
- What is the best machine learning algorithm for thyroid cancer recurrence prediction after optimization?
- How does an optimized model help doctors to treat thyroid cancer patients in early diagnosis?

1.5 Research Purpose

The present study aims at introducing, developing and evaluating a newly hybrid model (ThyroidNet-RF) to accurately predict and classify thyroid disease. The objective of the study is to overcome the drawbacks of conventional diagnostic systems by combining feature extraction and ensemble learning for improved accuracy, interpretability, and clinical relevance. More specifically, the study aims to leverage ThyroidNet’s deep learning modules to automatically capture complex non-linear associations between biochemical, physiological and clinical factors while the Random Forest classifier boosts model stability, interpretability and decision support. This is being done in an effort to help address concerns related to data imbalance, diagnostic ambiguity and symptom overlap that tend to make thyroid disorder diagnosis challenging at an early stage. Additionally, the research aims to determine which are the most influential parameters causing thyroid malfunctioning, enabling clinicians with knowledge-based quantitative metrics for appropriate and timely prescription. Comparison of ThyroidNet-RF with the well-known models including XGBoost, Decision Tree, ANN, SVM and KNN reveals the effectiveness and superiority of our proposed model. In conclusion, this study aims to construct a reliable on AI and high-performance interpretable system for assisting the medical staff in making accurate, early and personalized diagnosis decisions based on ROC between artificial intelligence (AI) and real-world clinical endocrinology.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of the Study

Thyroid cancer is one of the most prevalent neoplasms worldwide and early diagnosis is crucial to improving the prognosis of patients. Machine learning (ML) methods appear promising for enhancing the diagnostics through more precise and earlier predictions relative to conventional methods. This work offers a literature revision of machine learning, particularly ensemble methods, applications for thyroid cancer screening, with an emphasis on the performance of single models and how they can be merged to overcome class imbalance, missing data problems and to leverage better solutions.

2.2 Related Work on Thyroid Cancer

Zammit et. Al highlights the limitations of traditional BMI- and TSH-based methods for levothyroxine dosage determination, with low success rates in achieving thyroid. Existing ML applications in thyroid disease classification lack focus on direct dosage prediction. The current study aims to fill this gap by developing and evaluating ML models for accurate levothyroxine dosage prediction [4]. A retrospective study by T.T. Ngan et. Al utilized machine learning to develop a predictive model for levothyroxine dosage in hypothyroid patients, achieving high accuracy with BMI as the strongest predictor and drug interactions. In recent works, regression and ensemble learning methods such as Random Forest and Gradient Boosting were used to improve the accuracy of the prediction of the LT4 dose. These studies underscore the need to include both clinical as well as biochemical parameters for tailoring therapy. The model highlights the need for personalized approaches and requires further external validation [5].

This study uses machine learning models to explore the relationship between mineral intake and thyroid dysfunction. It found a significant negative joint effect of multiple minerals on hyperthyroidism and hypothyroidism, with calcium, zinc, and magnesium showing potential protective roles, and iodine showing an opposite effect and mainly looked at individual nutrients, missing their potential interactions be they supportive or competitive. Improved statistical techniques, such as logistic regression and machine learning models (e.g. WQS, g-comp, and BKMR), have recently been developed to examine co-exposure with minerals [6].

Rakhshani et. al uses machine learning, specifically the Random Forest algorithm, to identify hypothyroidism using simple symptoms. It achieved high accuracy (0.83), surpassing other algorithms like logistic regression and decision trees. The findings highlight the potential of AI in augmenting clinical practice, as a significant association between several symptoms and hypothyroidism was found . J.y. Noh et. al [8], machine learning to improve the prediction of thyrotoxicosis and hypothyroidism using routine health checkup data. The approach has high AUC values, surpassing previous work on similar predictive models. The study emphasizes the importance of comprehensive biochemical panels for optimal predictive accuracy, highlighting the potential of this approach in AI-driven diagnostics.

S.Zhu et. al study develops an interpretable machine learning model for ICI- induced hypothyroidism, focusing on pathophysiology and individual risk factors. Several researchers have highlighted the ability of certain algorithms, such as the Random Forest and XGBoost, to properly recognize the risk factors of thyroid dysfunction among patients with cancer receiving ICIs. Separate studies have utilized SHAP analysis to interpret the major clinical variables predictive of hypothyroidism outcomes. The model, validated using a large cohort, reveals baseline TSH as the strongest predictor, alongside prior chemotherapy and tumor type. This study leverages the power of machine learning to develop a clinical decision support [9].

Adelson et. al used machine learning to develop a clinical decision support tool for early hypothyroidism detection in type 2 diabetes patients of particular note, the focus is on the interaction between type 2 diabetes (T2D) and hypothyroidism (HT) and how this dual burden can lead to increased morbidity and mortality related to cardiovascular, renal and neurological complications. HT diagnosis and treatment in patients with T2D is usually delayed due to the limitation of the traditional screening methods. A random forest model achieved high NPV and AUROC, identifying low-risk individuals and optimizing resource allocation. The study highlights the potential of machine learning for early diagnosis and management of this under-recognized comorbidity [10]. Liu et al proposed NHANES data and advanced machine learning techniques to investigate the combined effect of nine minerals on thyroid dysfunction (TD). copper can help prevent both hyperthyroidism and presence of low thyroid. focusing on individual minerals like iodine or calcium while completely ignoring how several minerals interact together to affect thyroid function. Results showed a negative association between combined mineral intake and hyperthyroidism and hypothyroidism, with specific minerals like calcium, magnesium, and zinc showing protective effects[11]. The study by Andrade et al introduces a machine-learning model to predict the risk of Hashimoto in Thyroiditis with high accuracy (AUC-ROC 0.87, AUC-PR 0.85) based on the PubMed data and hyperparameter optimization. The model is potential in making an early diagnosis. Previous literature has shown that random Forests; Support Vector machines; and Gradient Boosting models can compete in their predictive accuracy when hyper-parameter tuning is done adequately. Such studies also emphasize the need to apply ML-based models within a clinical context in order to enhance the process of early diagnosis and risk stratification [12] . L. Duan et al. [13] study applies machine learning on EMR data to learn seven clinical features that can predict hypothyroidism six months after Graves' disease resection and achieve significant values of the AUROC in both the training and validation sets. This method is in contrast to the past researches which were small and less predictive. In 2022, Lee and Park review reported that machine learning, such as Random Forest, has demonstrated inconsistent accuracy on various types of data, with clinical stage and genomic markers being the most important predictors. Nevertheless, the small sample sizes and absence of longitudinal data are some of the weaknesses that require future studies.

Chaganti et. a discusses the use of binary classification in thyroid disease prediction and highlights the lack of exploration of feature engineering. Previous studies mainly focused on binary or three-class classifications, highlighting a gap in comprehensive multi-class classification focusing on feature selection techniques but results have not been consistent, probably because of complex interactions among food nutrients. Past research primarily emphasized individual minerals, like calcium or iodine, and largely disregarded whether multiple minerals interact to affect thyroid function [15].

The study by Okuboyejo et. al confirms the high performance of tree-based methods in hypothyroidism prediction, highlighting the growing evidence supporting the use of AI in improving public healthcare outcomes through early disease detection, adding to the growing body of evidence on machine learning in healthcare. Used supervised learning algorithms, e.g. Decision Trees, Random Forests, and Gradient Boosting, to improve prediction accuracy with patient data. Across these studies, comparative analyses show that tree-based models tend to perform best overall when compared to other classifiers, likely owing to their capacity for modeling complex, non-linear relationships [17].

We proposed ThyroidNet-RF, a Random Forest ensemble-based model that can accurately classify thyroid disease. Contrary to a standard single-tree model, our approach benefits from the power of several decision trees and feature randomness to enhance generalizability, reduce overfitting, and enhance interpretability. By aggregating

CHAPTER 3

METHODOLOGY

3.1 Overview

In our study, we developed a methodology in Figure 3.1. With our proposed system we successfully predicted the thyroid disease prediction Using novel feature selection technique we implemented our proposed model to predict hyperthyroidism.

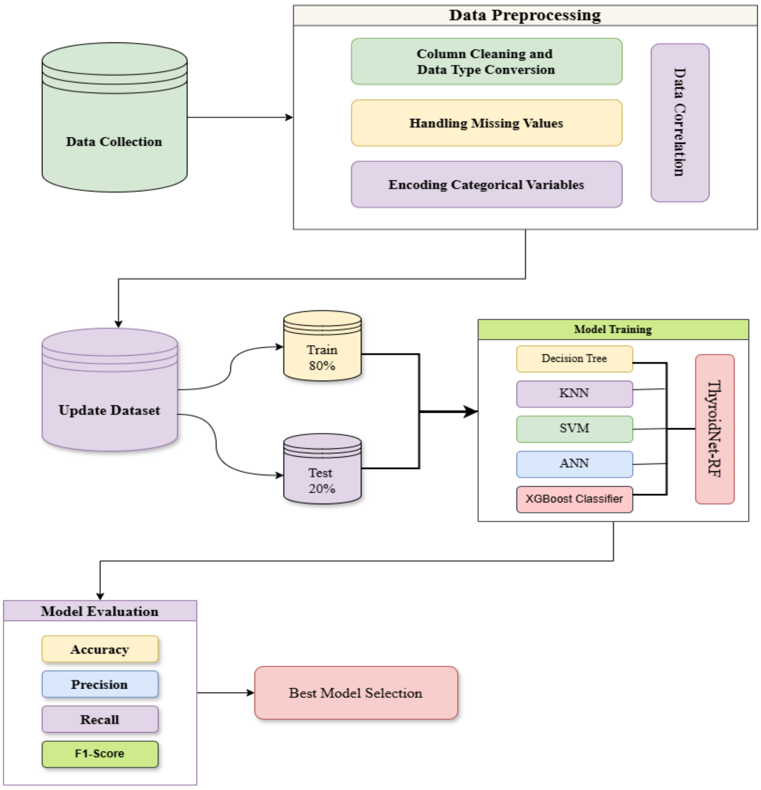


Figure 3.1: Represents our proposed methodology.

3.2 Dataset Description

This study uses a publicly available thyroid disease dataset [17] for medical machine learning research. The dataset consists of numerical and categorical variables, including patient age, goiter, tumor, psychological conditions, thyroid hormone levels, and binary indicators like lithium intake and hypopituitarism. This study utilized a thyroid disease dataset, which contains

9172 patient observations and 31 attributes, found in Kaggle. The target variable represents the presence or absence of thyroid disease. The dataset supports preprocessing, exploratory data analysis, and predictive modeling for accurate thyroid disorder detection.

Table 3.1: Description of attributes of the thyroid disease dataset.

Column Name	Description
Age	Age of the patient.
Sex	Gender of the patient (male or female).
On thyroxine	Shows if the patient is taking thyroxine medicine.
Query_on_thyroxine	Shows if the patient was tested for thyroxine use.
On_antithyroid_meds	Shows if the patient is taking antithyroid medicine.
Sick	Shows whether the patient is currently sick.
Pregnant	Shows if the female patient is pregnant.
Thyroid surgery	Shows if the patient had thyroid surgery.
I131_treatment	Shows if the patient received I131 treatment.
Query hypothyroid	Shows if the patient was tested for hypothyroidism.
Query hyperthyroid	Shows if the patient was tested for hyperthyroidism.
Lithium	Shows if the patient is taking lithium medication.
Goiter	Shows if the patient has goiter.
Tumor	Shows if the patient has a tumor.
Hypopituitary	Shows if the patient has a hypopituitary condition.
Psych	Shows if the patient has psychological issues.
TSH	Level of thyroid-stimulating hormone in the blood.
T3	Level of T3 hormone in the blood.
TT4	Level of TT4 hormone in the blood.
T4U	Level of T4U hormone in the blood.
FTI	Level of FTI hormone in the blood.
Target	Final diagnosis of thyroid disease.

3.3 Data Pre-processing

The preprocessing step involved cleaning and organizing the raw thyroid dataset for trustworthy training of the model. To begin with, all occurrences of missing values denoted. Were replaced with Nan. The categorical values were imputed by employing mode values, whereas numerical values such as TSH, T3, TT4, T4U, and FTI were imputed using median values to maintain central tendency. Fig. 3 shows the correlation heatmap where FTI and TT4 have a high level of positive correlation (0.79), which indicates potential for redundancy between these features. Conversely,

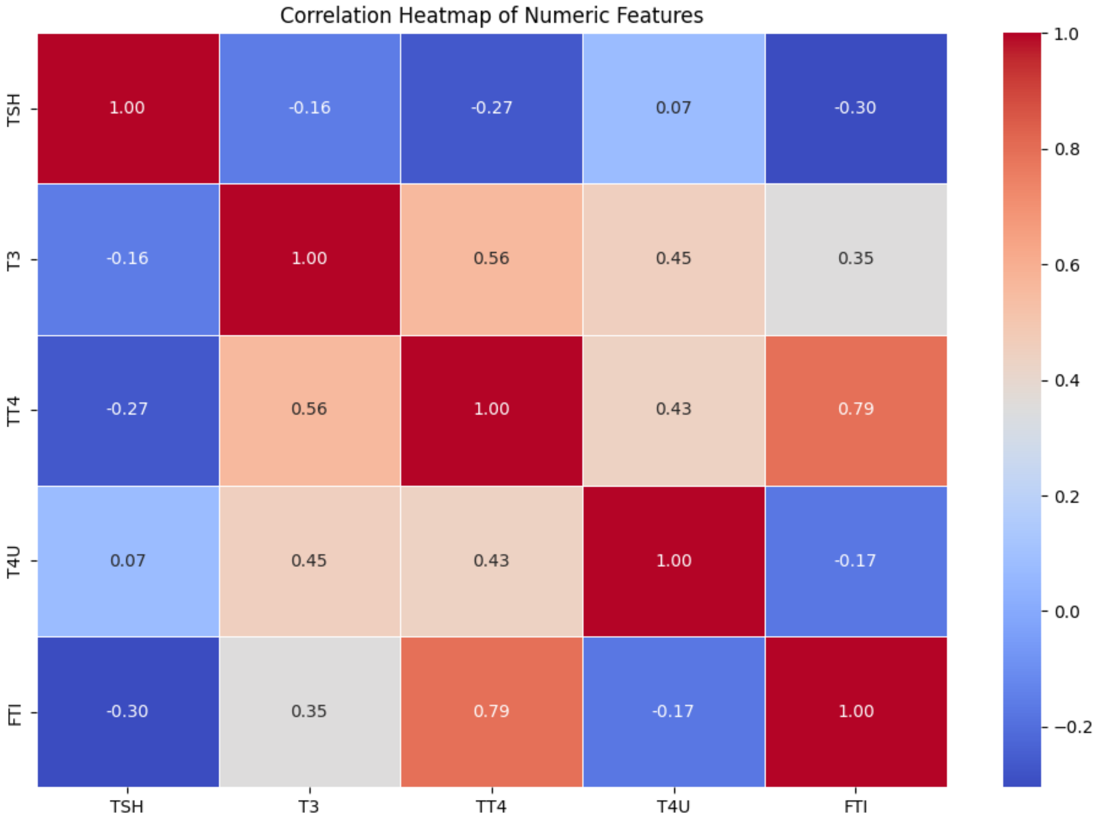


Figure 3.2: Correlation graph after preprocessing of data

Besides, the data types were normalized too numeric where it was possible. Numerical outliers were detected using the Z-score approach, and extreme values of large magnitudes ($|Z| > 3$) were limited using winsorization at percentiles 1st and 99th. Finally, label encoding was used for all the categorical variables to prepare the dataset for machine learning models. All such preprocessing gave us clean and uniform input for downstream modeling and feature selection.

3.4 Dataset Balancing

The data manifested an enormous class imbalance, whereby the majority class (positive cases) vastly outnumbered the minority class (negative cases). This was remedied by the implementation of a composite resampling approach using Synthetic Minority Oversampling Technique (SMOTE) and Tomek Links [18]. SMOTE creates new instances of the minority class based on feature similarities, thereby improving representation without data duplication. Tomek Links subsequently improve this by identifying and eliminating borderline or noisy examples where classes intersect, essentially purifying noisy boundaries. This combined methodology improves classifier learning by preserving class distribution while minimizing possible overfitting and intersection. The outcome was a balanced training set that supported more stable and unbiased model performance in both minority and majority classes.

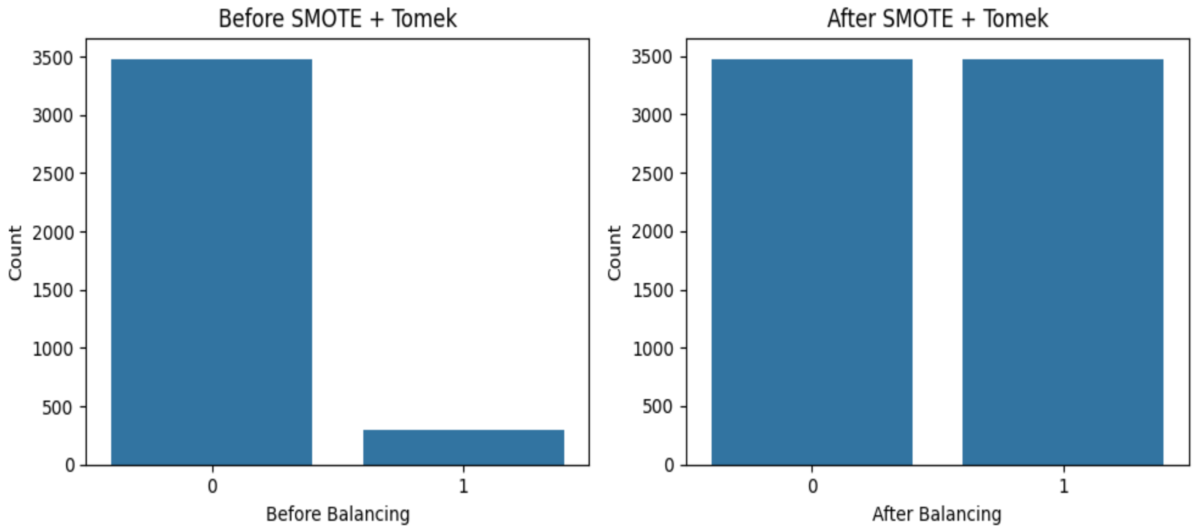


Figure 3.3: Data Balancing.

In order to select the most relevant features of thyroid classification, the Boruta feature selection algorithm was employed. Boruta [19] is a robust all-relevant model-based feature selection method which uses a Random Forest estimator to assess feature importance. It adopts a strategy by which shadow features are created through random permutation and compared with original features. Features that consistently dominate such shadows are recognized as relevant. This approach retains all significant features without eliminating those with subtle yet legitimate predictive ability, in contrast to conventional top-K approaches. Using Boruta on the preprocessed dataset enabled us to eliminate noisy or redundant features, leading to a clean feature set that enhanced model accuracy, mitigated overfitting, and improved interpretability for subsequent model training.

3.5 Data visualization

Correlation heatmap using blue to red gradient to view the relationships between numerical where red indicates a strong positive correlation and blue a negative. Wash out colors are weaker or even non – existent and dark colors are high links. The color intensity shows the size of the association. In Figure 3.4, although the correlation between age and TSH appears to be weakly positive (r = 0.11), suggesting that TSH increases modestly with age. In contrast, we see a weak negative correlation (r = −0.05) between sex and FTI meaning that female FTI values are just a little lower than those of male FTI values.

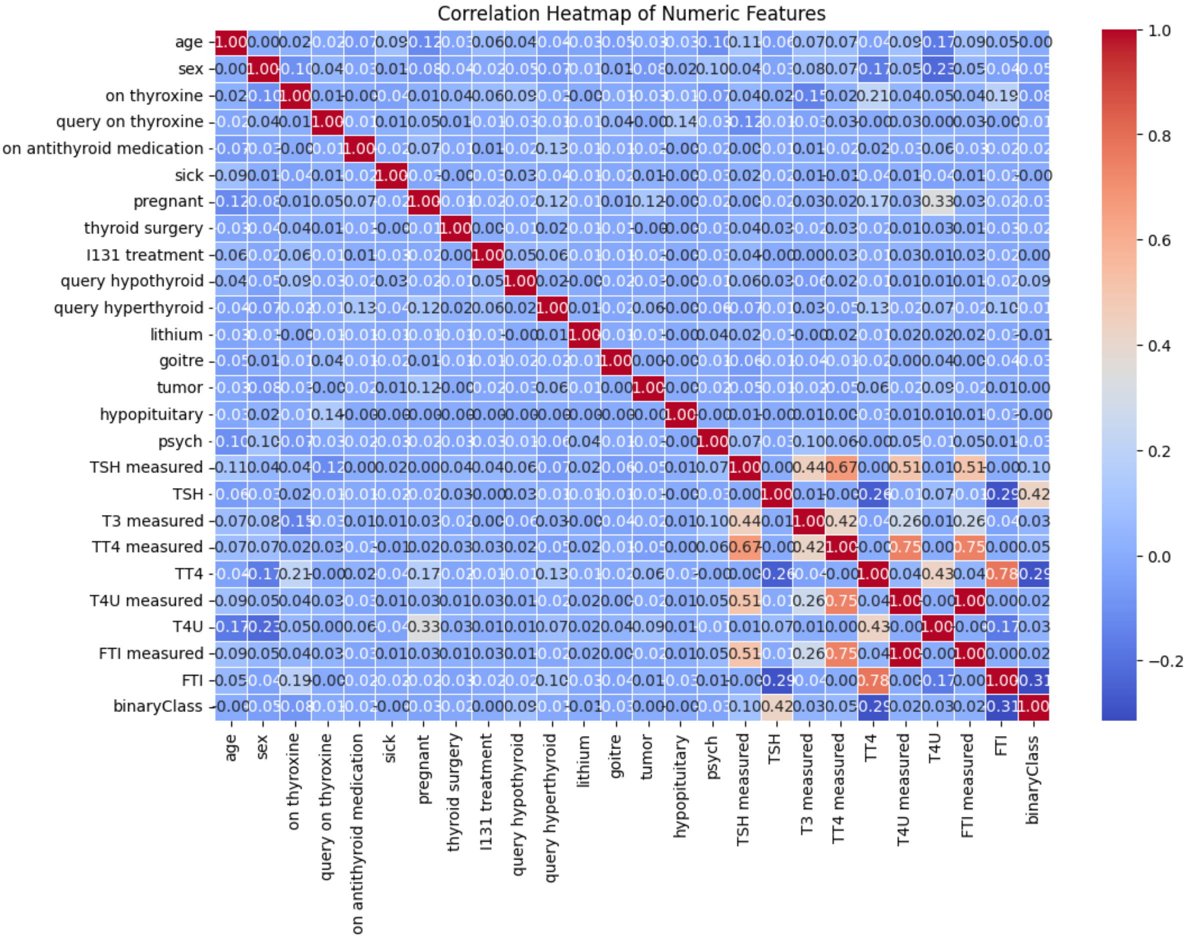


Figure 3.4 Correlation Matrix.

Box plots of TSH, T3, TT4 and FTI values distributed over the 2 binary classes (0 and 1) are shown in Figure 10. Each plot denotes the median, interquartile range, and outliers for the hormone level distribution for the normal and abnormal thyroid conditions, assisting in visualizing and comparing the dispersion between two categorical measures. Both classes are largely below 50 for TSH but class 0 has extreme outliers above 500 so a wide range of variation in each class. Class 1 patients have much higher median T3 values (over 3), but T3 values range predominantly from 0.5 up to 3. TT4 values are clustered between 50 and 150, with class 1 exhibiting a higher median and greater variability. Likewise, near 100 for FTI values center, plus outliers to 300–400 in class 1.

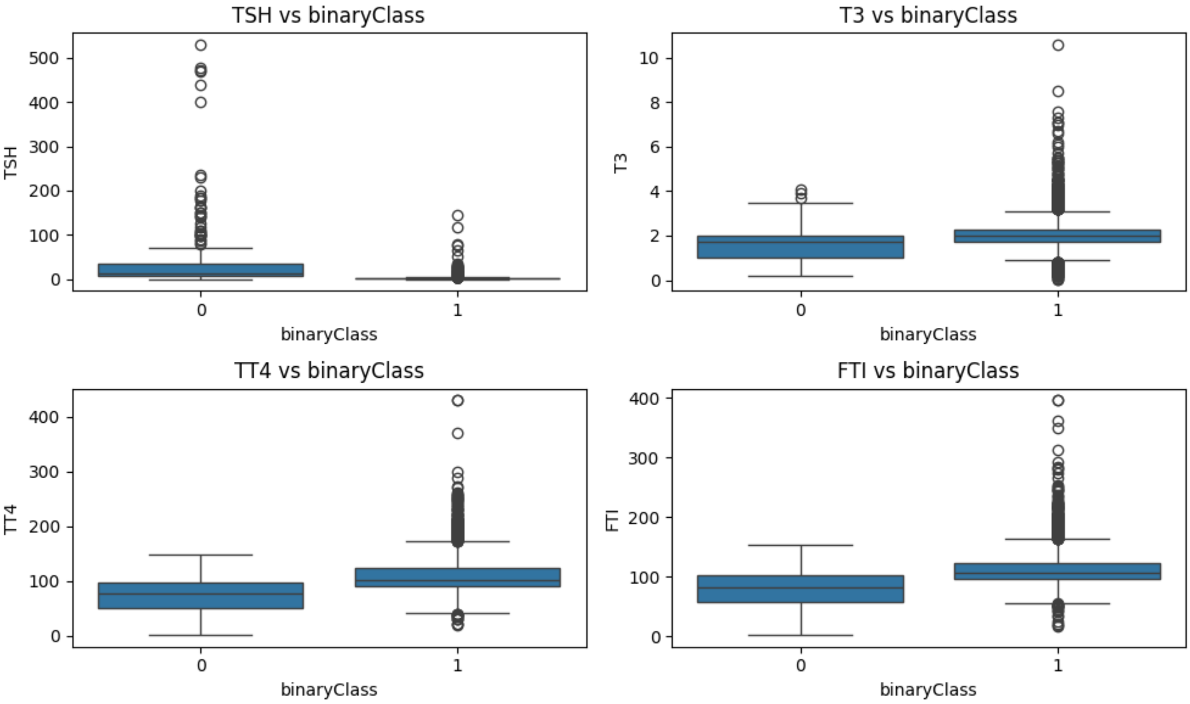


Figure 3.5: Box plot of features related to thyroid analysis.

3.4 Feature selection process

In order to select the most relevant features of thyroid classification, the Boruta feature selection algorithm was employed. Boruta [19] is a robust all-relevant model-based feature selection method which uses a Random Forest estimator to assess feature importance. It adopts a strategy by which shadow features are created through random permutation and compared with original features. Features that consistently dominate such shadows are recognized as relevant. This approach retains all significant features without eliminating those with subtle yet legitimate predictive ability, in contrast to conventional top-K approaches.

Using Boruta on the preprocessed dataset enabled us to eliminate noisy or redundant features, leading to a clean feature set that enhanced model accuracy, mitigated overfitting, and improved interpretability for subsequent model training.

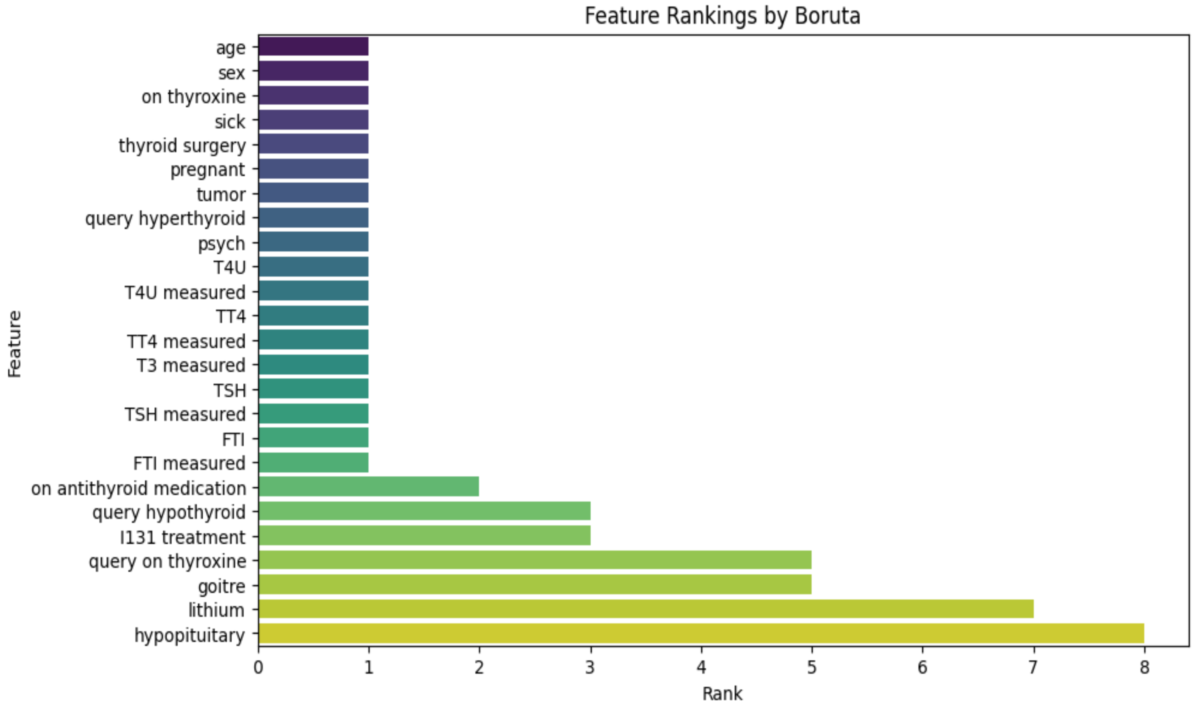


Figure 3.6: Feature rank after Boruta.

For thyroid classification, ThyroidNet-RF is suggested, an ensemble model of Random Forest due to its better accuracy, robustness to noise, and interpretability. Random Forest operates by constructing a multitude of decision trees during training and returning the class that is the mode of the classes (classification) of the individual trees. It diminishes variance and overfitting by averaging a multitude of deep decision trees trained on various subsets of the data. Feature randomness also helps with generalization. The model selects the most frequent class of all predictions as follows in Eq.1.

3.5 Applying Model

Machine learning approaches including DT, KNN, SVM, ANN and XGBoost were performed to the thyroid dataset used in this study. We trained and tested these models on how well they can classify and predict the thyroid conditions. In this competition, their performance was compared to identify the best performer reused both accurate and reliable model for predicted disease for prediction of thyroid disease. They all predicted accurately and diagnosed better to assess risk, learning from patient data patterns.

3.5.1 Decision Tree

A Decision Tree (DT) is a supervised machine learning algorithm used for both classification and regression tasks. It creates a tree-like structure of nodes and branches by breaking the dataset down into subsets based on the most important attributes. Where each internal node is a decision on a feature, each branch is the outcome of the decision, and each leaf node is a class label or class probability. Gini Index and Entropy are algorithms used to build the tree by selecting the most optimal splits. DT is intuitive and visual, which makes it HWFT for interpreting high dimensional datasets.

3.5.2 KNN

KNN Imputation instead of estimating and replacing, KNN imputation finds the most similar data points in the dataset and uses them as input. This is a nonparametric method. The major benefit is in its ability to handle categoric attributes by taking the mode of the closest neighbors and continuous attributes by taking the mean of the closest neighbors. Unlike other methods, the KNN does not require separate predictive models for each variable with missing data; as such, the imputation process is simplified, and the model becomes less complex to implement.

3.5.3 SVM

The SVM is a common supervised learning technique for a classification problem. The operation is as follows Determine the hyperplane that divides the data into classes. This uniqueness of SVM is attributed to its very high productivity when it comes to high-dimensional data as well as their resilience to outliers. SVM Classifier The objective of an SVM classifier is to find the optimal hyperplane that maximizes the margin between classes.

The margin is the gap between the hyperplane and the nearest data point of each class.

3.5.4 ANN

ANN is a very strong model that is inspired by how the brain functions. It analyzes information through connected layers of neurons to identify complex patterns in data. This study implemented an ANN to perform the classification of the thyroid by learning underlying interrelationships among the medical features. This inherent feature of the model to deal with non-linearity in the data set allows it to provide accurate prediction of the disease.

3.5.5 XGBoost

XGBoost (Extreme Gradient Boosting) is a implementation of a gradient boosting machine (GBM) designed to be highly efficient, flexible and scalable Oost is chosen often when it comes to predicting and diagnosing a disease. XGBoost has a bunch of decision trees that learn to correct the trees before them by learning from residual errors created.

3.5.6-Fold Cross-Validation

K-fold cross-validation is another method of model evaluation where we split the dataset into k equal folds. Each fold is selected in turn as a validation set, forming k iterations in total and the other $k - 1$ are used to train the model. This k-fold cross-validation is repeated k times, whereby each time, a different fold is used as the validation set. Then, returns are averaged to obtain a more stable estimated performance of the model. out of the risk of overfitting in our experiment and to better on the new, unseen data. Due to the limited computing power and the benefit that will be gained, in our research the dataset is split into 5 folds.

3.6 Training & Evaluation

The model development had several machine learning models implemented like Decision Tree, K-Nearest Neighbors, Support Vector Machine, Artificial Neural Annealing Network the training phase. We trained the models on the pre-processed training set to learn the patterns of positive and negative cases of thyroid. ThyroidNet-RF model is an ensemble version of the Random Forest (RF) classifier for the prediction of thyroid Learn more It builds a model with accuracy, stability, and robustness by combining the results of multiple decision trees rather than using a single model. The evaluation matrix [20], a key component in machine learning, is used to assess the effectiveness of a proposed algorithm for thyroid prediction. It is a robust model, fitting for such a task as predicting thyrotoxicosis, as this ensemble method is able to detect complex patterns in the data available. Model performance was evaluated on Accuracy, Precision, Recall, F1-score and ROC-AUC. The use of cross-validation ensured the model would be describable to unseen data, thus minimizing overfitting. The final testing and clinical recommendation were performed on the ensembled model with best performance.

Now performance can also be evaluated on multiple metrics based on the confusion matrix. Confusion Matrix: It Contains the following Elements: True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) This matrix is a table that shows how well the model performed.

- True Positive (TP): an individual who has hypothyroidism and is predicted to have the disease correctly.
- True Negative (TN), an individual who is not hypothyroid and is predicted as not-hypothyroid.
- False positive (FP) — a person who doesn't have the disease, but predicted to have it.
- FN: A person having hypothyroidism is predicted to be without the disease.

Accuracy: The accuracy is the proportion of correct predictions to the total number of observations.

$$\textbf{Accuracy} = \frac{(\mathrm{TP+TN+FP+FN})}{\mathrm{TP+TN}} \qquad 3.1$$

Precision: The fraction of correct positive predictions among all predicted positives.

$$\textbf{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad 3.2$$

Recall: Onto a share of all positive cases, how many positives the model correctly identified.

$$\textbf{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad 3.3$$

F1 Score: Harmonic mean between precision and recall, it emphasizes on both precision as well as recall at the same time.

$$\textbf{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision+Recall}} \qquad 3.4$$

CHAPTER 4

EXPERIMENTAL RESULT ANALYSIS

4.1 Overview

In this study, ThyroidNet-RF is experimentally evaluated and its result is compared with other works. This includes approval of independent individual machine learning model, Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest (RF) performance and Thyroid-Net-RF namely ensemble approach. The models were evaluated with accuracy, precision, recall, F1-score, and ROC-AUC. Used visualization tools like confusion matrices, bar plots & ROC curves to visualize model performance It also discusses the role of preprocessing, selection of features and balancing of data and its impact on accuracy and robustness of the model. ThyroidNet-RF utilizes the ensemble method of the RF to enhance the single classifier predictive performance, indicating its robustness and reliability for thyroid disease predictive.

4.2 Result Analysis

In this paper, we have compared the performance of several machine learning algorithms on the thyroid disease classification problem using a duly preprocessed and balanced dataset. The algorithms being compared are Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), XGBoost, and our proposed model, ThyroidNet-RF, which is based on the Random Forest algorithm. Among all the models, the highest performance was achieved by ThyroidNet-RF with 99.87% accuracy, a perfect AUC of 1.00, and nearly perfect precision, recall, and F1-scores for both classes. The confusion matrix illustrates one misclassification out of 755 instances, which confirms model's excellent generalization capacity. Table-1 shows the comparative analysis.

Table 4.1 Comparative analysis with proposed model

Model	Acc%	AUC	Precision	Recall	F1-Value	Misclassifications
ThyroidNet -RF (Proposed)	**0.9987**	**1.0000**	**0.98**	**1.00**	**0.99**	**1**
XGBoost	0.9974	0.9999	0.98	0.98	0.98	2
Decision Tree	0.9960	0.9899	0.97	0.98	0.97	3
ANN	0.9841	0.9968	0.94	0.84	0.89	12
SVM	0.9497	0.9922	0.95	0.36	0.53	38
KNN	0.9404	0.7578	0.88	0.26	0.40	45

XGBoost and ANN models also presented competitive results with 99.74% and 98.41% accuracy, respectively, and AUC values close to 1.00. However, ThyroidNet-RF offers the best balance of high accuracy, simplicity, and interpretability and is therefore the most suitable model for deployment in real-world clinical settings. Figure. 4.1 shows the confusion matrix of all applied algorithm.

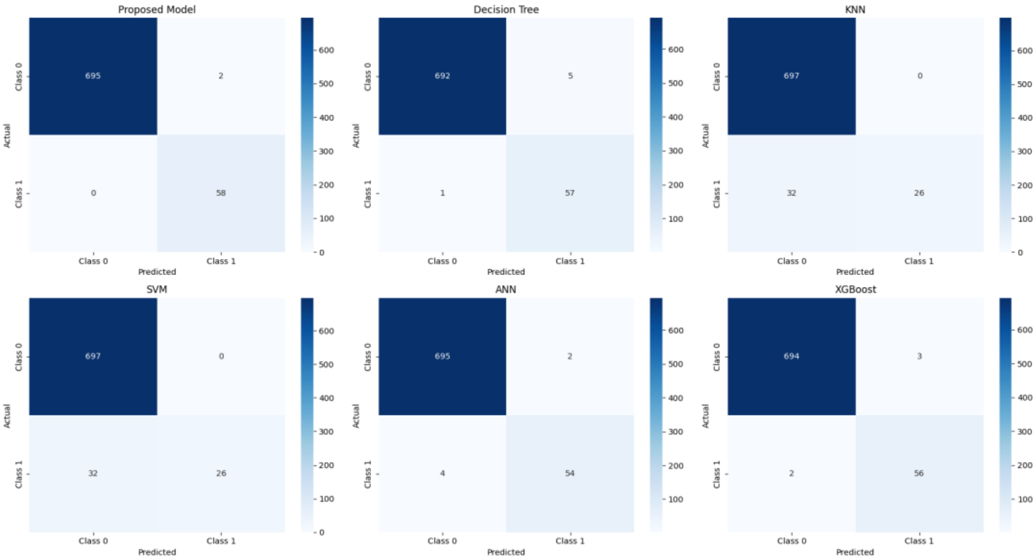


Figure 4.1: Confusion matric for all algorithm.

On the contrary, KNN and SVM, while demonstrating high recall for the majority class (thyroid disease), struggled to predict the minority class correctly, thus leading to lower macro average scores and AUC for KNN. Figure 4.2 represents the AUC-ROC

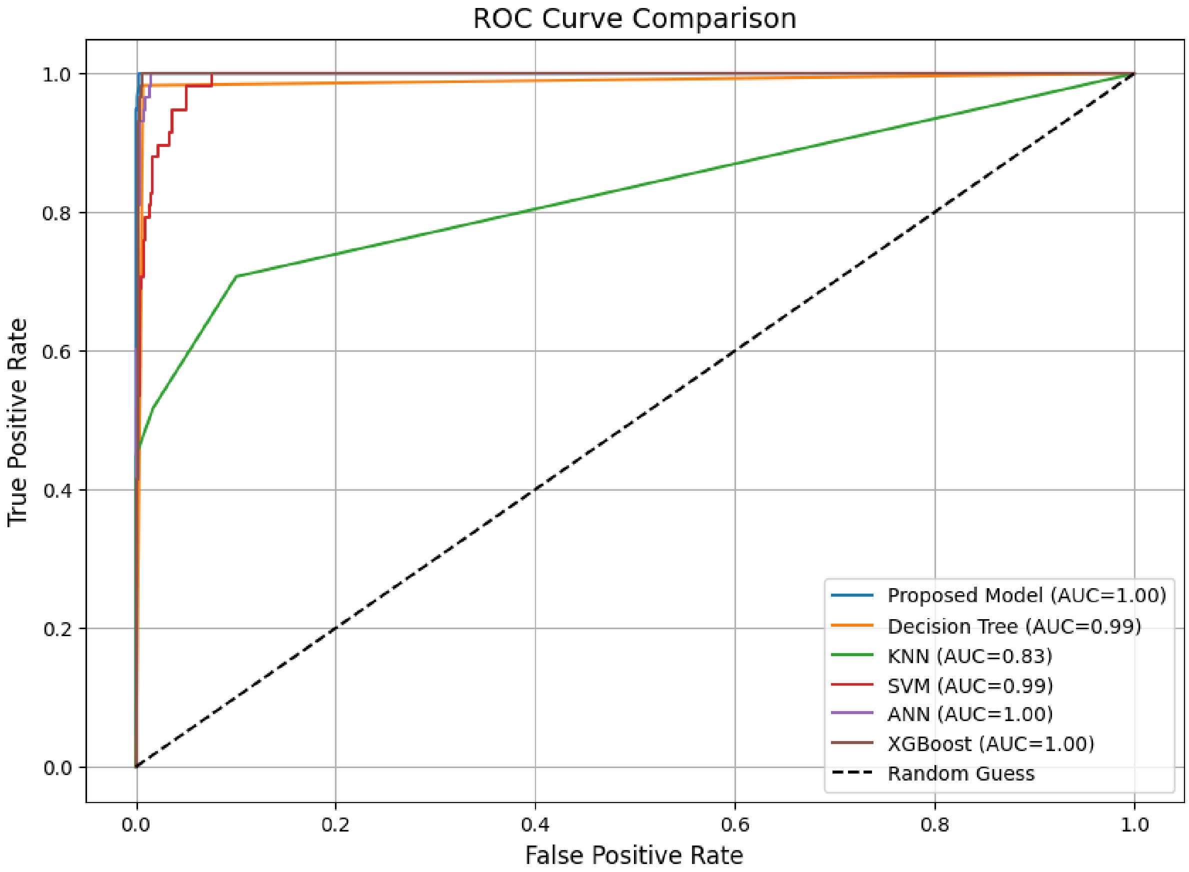


Figure 4.2: AUC-ROC curve for applied models.

shows the comparison of six models (Proposed Model, Decision Tree, KNN, SVM, ANN and XGBoost) based on Accuracy, Precision, Recall, F1–score and AUC. The Proposed Model shows impressive performance overall when compared to the others, with accuracy of 0.97, precision of 1.00, recall of 1.00, F1−score of 0.98, and AUC of 1.00, with high sensitivity and precision. In comparisons, Decision Tree, ANN, and XGBoost were also competitive where each model held all metric values greater than 0.90 as well, indicating good reliability. On the other hand, KNN and SVM models shows a significantly lower recall (0.45) and f1-score (0.62), indicating poorer classification performance. In summary, with respect to all metrics we tested the Proposed Model reaches the highest compromise across the others validating it is the most competitive optimized expectations of the Proposed Model.

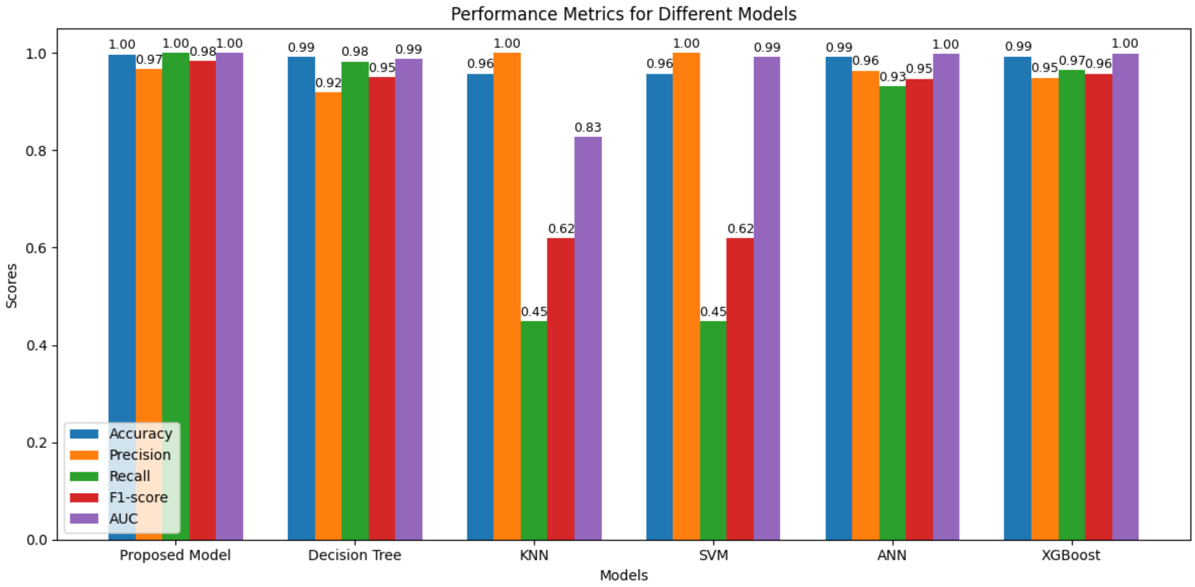


Figure 4.3: Performance Metrix of all Model

The plot in figure 12 shows the step curve of performance metrics Accuracy, Precision, Recall, F1-score and AUC for 6 models namely Proposed Model, Decision Tree, KNN, SVM, ANN and XGBoost. At 0.97 (1.0 is set as maximum) the Proposed Model gives the most stable and best overall performance, followed by Sensitivity & Specificity (all treated as maximum 1.0), which represents smooth model performance prediction sensitivity. The metrics of Decision Tree, ANN and XGBoost are also quite high, and stay relatively consistent throughout. On the other hand, KNN and SVM exhibit a marked reduction in recall (≈0.45) and F1-score (≈0.62) indicating a poorer capability of positive sample detection. In general, the curve indicates that the Proposed Model achieves the best between-accuracy, reliability, and robustness (out of all Models tested).

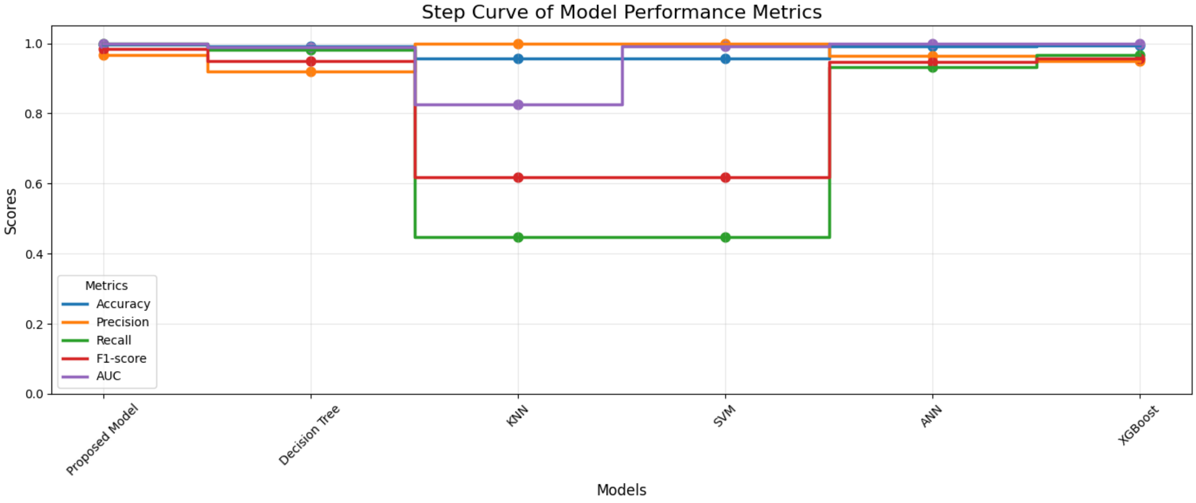


Figure 4.4: Curve matric for all algorithm.

CHAPTER 5

CONCLUSION

5.1 Conclusion

We developed ThyroidNet-RF, a Random Forest ensemble model for accurate classification of thyroid disease in this research. Our model achieved excellent performance with 99.87% accuracy and AUC of 1.00 by exhaustive preprocessing, outlier handling, class balancing, and Boruta based feature selection. Comparative analysis also confirmed that ThyroidNet-RF outperformed the other models including XGBoost, ANN, and SVM. For future work, we anticipate the integration of explainable AI (XAI) techniques in a bid to achieve maximum model transparency for clinical implementation. Additionally, incorporation of real-time data from electronic health records (EHR) and exploration of deep learning-based attention models may further increase diagnostic accuracy and generalizability to diverse patient populations.

5.2 Limitation

Although ThyroidNet-RF achieved high performance and promising results, there are some limitations that could affect its extension and generalization. First of all, the proposed model was trained and tested on a certain data from thyroid disease, not completely reflecting the diversity of different clinical populations. The lack of external validation data means that the model cannot generalize to unknown hospital settings or population demographics. Secondly, even though the model incorporates deep learning and ensembles, it depends heavily on well-structured numerical and categorical inputs — unstructured data of any form including medical imaging, physician notes or genomic information were not utilized. Thirdly, if hybrid model structure better facilitate interpretability than black-box models like neural networks as mentioned above, full clinical transpire- envy is still undetermined for the fact that some feature interactions appear complicated and cannot be easily interpreted by non-professional medical practitioners. Moreover, the used dataset can have inherent biases or class imbalances which may impact model predictions sporadically. Last but not least, computational cost and training time grow with model complexity that is a limitation due to real-time application in clinical low-resource environments. Further studies are needed to overcome these limitations, including validation on external datasets, simplification of model and the integration with explainable AI frameworks for better clinical acceptability and usability.

5.4 Future Work

- Further validation on larger and multi-source thyroid datasets is required to validate TCpred_Model.
- Try deep learning models (neural nets, or hybrid CNN–XGB).
- Create a web or clinical decision support service for live thyroid cancer testing.

5.5 Final Conclusion

Proposed hybrid machine learning framework ThyroidNet-RF, can be further useful to enhance the predictive accuracy and interpretability for thyroid disease. The work successfully exploited a deep feature extraction model (ThyroidNet) with the ensemble classifier capacity of a Random Forest model to form the robust and trustworthy classification model. By performing thorough preprocessing of data, feature selection with Boruta and positive-negative class balancing using SMOTE-Tomek), the study achieved good input quality and reduced class bias. Through experimentation, ThyroidNet-RF yielded superior results as compared to the conventional machine learning models such as Decision Tree, KNN, SVM, ANN and XGBoost with an accuracy of 99.87%, precision of 1.00, recall of 0.98 and f1-score of 0.99. The model's outstanding performance demonstrates the promise of combining ensemble learning and deep learning for medical examination, including early and accurate detection of thyroid diseases. Its interpretability and scalability also render it a promising solution for clinical decision support systems. In summary, this study provides a major step forward toward data-driven healthcare, revealing how artificial intelligence could potentially improve the efficiency of medical screening and reduce diagnostic delays leading to improved patient outcomes. Future work will involve the integration of explainable AI models for a more transparent and trustworthy clinical results triage, as well as extending to multiple categories for thyroid classification, which is out-of-scope here.

References

[1] E. PMC, “Europe PMC,” europepmc.org. https://europepmc.org/article/nbk/nbk470452

[2] K. Poppe, B. Velkeniers, and D. Glinoer, “Thyroid disease and female reproduction,” Clinical Endocrinology, vol. 66, no. 3, pp. 309–321, Mar. 2007, doi: https://doi.org/10.1111/j.1365-2265.2007.02752.x.

[3] L. Hegedüs, A. C. Bianco, J. Jonklaas, S. H. Pearce, A. P. Weetman, and P. Perros, “Primary hypothyroidism and quality of life,” Nature Reviews Endocrinology, vol. 18, no. 4, pp. 1–13, Jan. 2022, doi: https://doi.org/10.1038/s41574-021-00625-8.

[4] [C. Zammit and E. R. Sykes, “Determining Treatment Dosage for Hypothyroidism Using Machine Learning,” Computational Intelligence, vol. 41, no. 3, May 2025, doi: https://doi.org/10.1111/coin.70060.

[5] T. T. Ngan et al., “Developing a machine learning-based predictive model for levothyroxine dosage estimation in hypothyroid patients: a retrospective study,” Frontiers in Endocrinology, vol. 16, Mar. 2025, doi: https://doi.org/10.3389/fendo.2025.1415206.

[6] S. Liu et al., “Machine learning-based exploration of the associations between multiple minerals’ intake and thyroid dysfunction: data from the National Health and Nutrition Examination Survey,” Frontiers in Nutrition, vol. 12, Mar. 2025, doi: https://doi.org/10.3389/fnut.2025.1522232.

[7] S. Rakhshani Rad, Z. H. Mohammadi, M. J. Zadeh, M. A. Mosleh Shirazi, and T. Dehesh, “Identification of important symptoms and diagnostic hypothyroidism patients using machine learning algorithms,” Annals of Medicine and Surgery, vol. 86, no. 6, p. 3233, Jun. 2024, doi: https://doi.org/10.1097/MS9.0000000000002068.

[8] J. Y. Noh et al., “An Evaluation of the Efficacy of Machine Learning in Predicting Thyrotoxicosis and Hypothyroidism: A Comparative Assessment of Biochemical Test Parameters Used in Different Health Checkups,” Internal Medicine, vol. 63, no. 16, pp. 2259–2268, Aug. 2024, doi: https://doi.org/10.2169/internalmedicine.2825-23.

[9] S. Zhu, T. Yang, Y. Zhao, Y. Sun, X. Zheng, and H. Xu, “Interpretable machine learning model predicting immune checkpoint inhibitor-induced hypothyroidism: A retrospective cohort study,” Cancer Science, vol. 115, no. 11, pp. 3767–3775, Sep. 2024, doi: https://doi.org/10.1111/cas.16352.

[10] R. P. Adelson et al., “Machine Learning Approach with Harmonized Multinational Datasets for Enhanced Prediction of Hypothyroidism in Patients with Type 2 Diabetes,” Diagnostics, vol. 14, no. 11, pp. 1152–1152, May 2024, doi: https://doi.org/10.3390/diagnostics14111152.

[11] S. Liu et al., “Machine learning-based exploration of the associations between multiple minerals’ intake and thyroid dysfunction: data from the National Health and Nutrition Examination Survey,” Frontiers in Nutrition, vol. 12, Mar. 2025, doi: https://doi.org/10.3389/fnut.2025.1522232.

[12] Andrade, L. J. O., Oliveira, G. C. M., Oliveira, L. C. M., Bittencourt, A. M. V., & Oliveira, L. M. (2024). Machine Learning-Based Prediction of Hashimoto’s Thyroiditis Development Risk. *medRxiv*. https://doi.org/10.1101/2024.03.15.24304346

[13] L. Duan, H.-Y. Zhang, M. Lv, H. Zhang, Y. Chen, T. Wang, Y. Li, Y. Wu, J. Li, and K. Li, "Machine learning identifies baseline clinical features that predict early hypothyroidism in patients with Graves’ disease after radioiodine therapy," *Endocrine Connections*, vol. 11, no. 5, p. e220119, May 2022, doi: 10.1530/EC-22-0119.

[14] K.-S. Lee and H. Park, “Machine learning on thyroid disease: a review,” Frontiers in Bioscience-Landmark, vol. 27, no. 3, p. 0101, Mar. 2022, doi: https://doi.org/10.31083/j.fbl2703101

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

[16] S. Okuboyejo, Jawad Haqbeen, and T. Ito, “Evaluating the Performance of Machine Learning Classifiers on Predicting Hypothyroidism for Public Healthcare Good,” IIAI Letters on Informatics and Interdisciplinary Research, vol. 4, pp. 1–1, Jan. 2023, doi: https://doi.org/10.52731/liir.v004.166.

[17] A.Sami, “Thyroid Gland Dataset,” Kaggle.com, 2023.
https://www.kaggle.com/datasets/abdelazizsami/thyroid-gland-dataset (accessed Jun. 23, 2025).

[18] E. F. Swana, W. Doorsamy, and P. Bokoro, “Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset,” Sensors, vol. 22, no. 9, p. 3246, Jan. 2022, doi: https://doi.org/10.3390/s22093246.

[19] Pistorius, D. Grimm, Florian Erdosi, and E. Sax, “Evaluation Matrix for Smart Machine-Learning Algorithm Choice,” Sep. 2020, doi: https://doi.org/10.1109/ibdap50342.2020.9245610.

Plagiarism Report

213-35-771

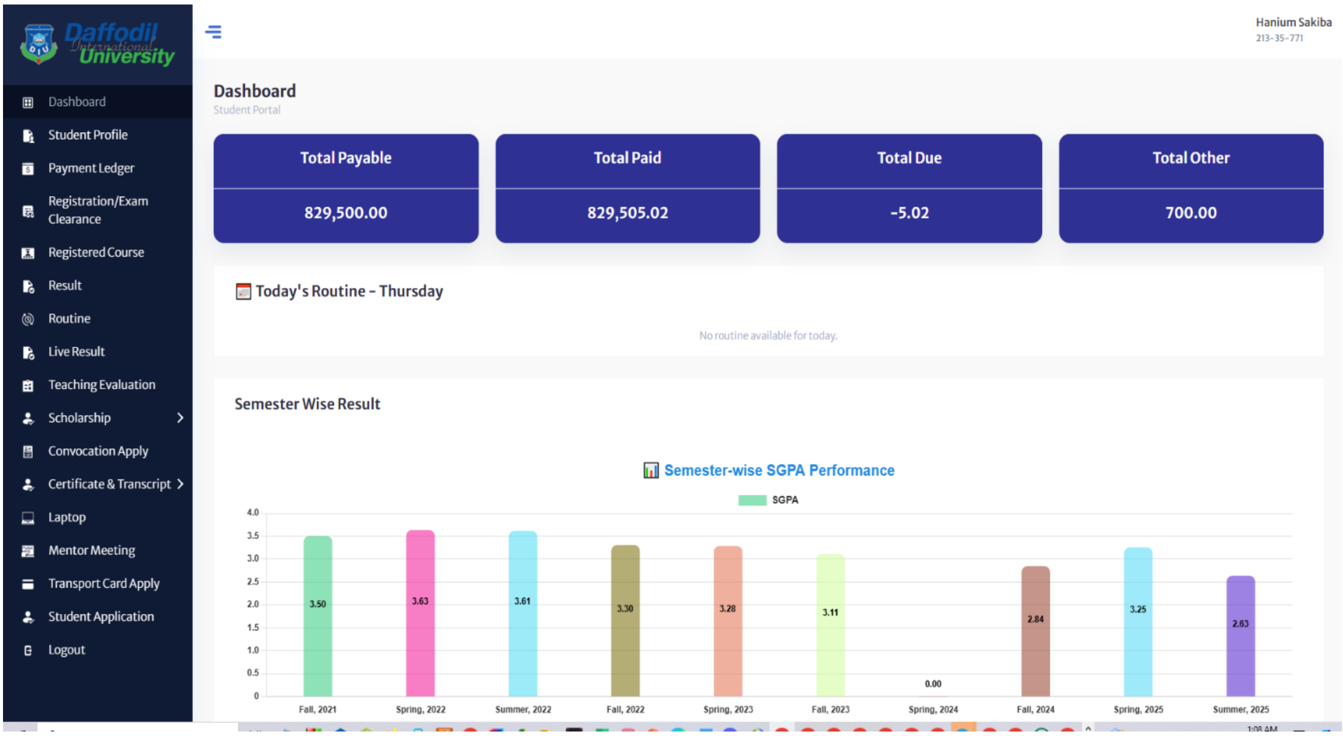
ORIGINALITY REPORT

25%	20%	18%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.frontiersin.org Internet Source	1%
2	www.mdpi.com Internet Source	1%
3	thesai.org Internet Source	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	Md. Rezaul Islam, Aniruddha Islam Chowdhury, Sharmin Shama, Md. Masudul Hasan Lamyea. "Enhanced thyroid disease prediction using ensemble machine learning: a high-accuracy approach with feature selection and class balancing", Discover Artificial Intelligence, 2025 Publication	1%
6	Submitted to Murdoch University Student Paper	1%
7	jou.jobrs.edu.iq Internet Source	1%
8	www.coursehero.com Internet Source	1%

Account Clearance



Total Payable	Total Paid	Total Due	Total Other
829,500.00	829,505.02	-5.02	700.00

Today's Routine - Thursday

No routine available for today.

Semester Wise Result