**Department of Electrical and Computer Engineering**

**North South University**



**CSE445 Report**

**Thyroid Cancer Recurrence Prediction Using Machine Learning Techniques**

**Name: S. T. A. Mahmud Tonmoy ID: 2011105042**

**Name: Fazley Rabbi ID: 2022242042**

**Name: Shakil Mahmud Resvy ID: 2014319042**

**Faculty:**

**Riasat Khan**

**Assistant Professor**

**ECE Department**

**Spring, 2025**

**Individual Contribution Table**

|  |  |  |
| --- | --- | --- |
| **Section** | **Contributing Member Name** |  |
| IEEE Word/LaTEX formatting | Fazley Rabbi |  |
| Grammarly check | Fazley Rabbi, S. T. A. Mahmud Tonmoy | Grammarly Score: 95 |
| Abstract | S. T. A. Mahmud Tonmoy |  |
| Keywords | S. T. A. Mahmud Tonmoy |  |
| Introduction Motivation | S. T. A. Mahmud Tonmoy |  |
| Paper Review 1 | S. T. A. Mahmud Tonmoy | [5] |
| Paper Review 2 | Fazley Rabbi | [6] |
| Paper Review 3 | Shakil Mahmud Resvy | [7] |
| Introduction Second-Last Paragraph (describe your work) | Fazley Rabbi |  |
| Proposed System (Dataset and Preprocessing) | S. T. A. Mahmud Tonmoy |  |
| Proposed System (Model description) | Shakil Mahmud Resvy | Random Forest |
| Shakil Mahmud Resvy | SVM |
| Shakil Mahmud Resvy | Logistic Regression |
| S. T. A. Mahmud Tonmoy | XGBoost |
| S. T. A. Mahmud Tonmoy | Decision Tree |
| Fazley Rabbi | Gradient Boosting |
| Fazley Rabbi | Stacking |
| Fazley Rabbi | Blending |
| Results and Discussion | Shakil Mahmud Resvy,  S. T. A. Mahmud Tonmoy |  |
| Figure and Table Title Formatting | S. T. A. Mahmud Tonmoy,  Shakil Mahmud Resvy |  |
| Conclusions | Fazley Rabbi |  |
| Equations formatting | S. T. A. Mahmud Tonmoy |  |
| References Formatting in IEEE format | S. T. A. Mahmud Tonmoy |  |

Thyroid Cancer Recurrence Prediction Using Machine Learning Techniques

Mahmud Tonmoy   
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
mahmud.tonmoy@northsouth.edu

Fazley Rabbi  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
fazley.rabbi01@northsouth.edu

Shakil Resvy  
Electrical and Computer Engineering  
North South UniversityDhaka, Bangladesh  
shakil.resvy@northsouth.edu

*Abstract*—Thyroid Cancer is a malignant neoplasm of the thyroid gland that globally affects 586,202 people, where the death rate is 43,646, which increased to 821,173 later on. It has become common, and the major sufferers are women. Thyroid cancer can be treated with radioactive therapy, but still, there is a high chance of recurrence, which is very alarming. The recurrence of thyroid cancer depends on various factors like smoking, obesity, genetics, and others. The death rate can be reduced if the recurrence of thyroid cancer can be predicted early. In this paper, an automatic thyroid cancer recurrence prediction system has been developed using multiple machine-learning techniques. The authors have used a public dataset named Differentiated Thyroid Cancer Recurrence. Data preprocessing was done by applying feature selection, label encoding, one-hot encoding, and a train-test split to this dataset. After that, three types of feature selection algorithms have been applied to this work. The SMOTE technique has also been applied to manage the imbalance of classes. In this work, various machine learning approaches have been implemented where the random forest model gave the best performance with 0.9708 test accuracy and precision, recall, and F1-score of 0.97. LIME’s explainable AI approach is implemented to understand how the best-performing model works visually.

Keywords—malignant neoplasm, follicular, parafollicular

# Introduction

Thyroid cancer is a malignant neoplasm of the thyroid gland that originates from the follicular or parafollicular cells of the thyroid gland [2]. It is the most common endocrine malignancy [1, 2, 3, 4]. Thyroid cancer’s leading risk factors are obesity, smoking, exposure to radiation, environment, and genetics [1, 3, 4]. The number of thyroid cancer occurrences is more common in women than men. It is three times higher [1, 3]. The Global Cancer Statistics (GLOBOCAN) report shows that in 2020, there were 586,202 cases of Thyroid Cancer. The death number was 43,646 all over the world [3], which increased in 2022, and the number of cases went to 821,173 [4]. Thyroid cancer has become so common in the United States that the annual case is 40,000 [1]. Thyroid cancer can be divided into four groups histologically, which are papillary thyroid cancer (PTC), follicular thyroid cancer (FTC), medullary thyroid cancer (MTC), and anaplastic thyroid cancer (ATC) [2]. FNAB (fine-needle aspiration biopsy) with a cytological analysis is the most commonly used method to diagnose thyroid cancer [2]. Surgical removal of the thyroid gland with radioactive therapy is done for the treatment after cancer cells are investigated by the “OMICS” technology [2]. With a 5-year relative survival rate of 98.2% for all stages, most of the thyroid cancer patients have excellent prognosis [4].

In the following paragraph, some recent works on automatic thyroid cancer recurrence prediction systems have been described briefly.

Ozturk et al. [5] implemented various machine-learning models to predict the recurrence of thyroid cancer recurrence. They chose the Thyroid Cancer Recurrence Prediction dataset, which contains 383 samples and 17 features. The authors have used the K-Fold method to prevent overlearning and ensure accurate results. The XGBOOST model outclassed others by achieving the best numbers, respectively, with an accuracy of 0.953, an F1-score of 0.917, an F-05 score of 0.919, and an AUC value of 0.982.

Imtiaz et al. [6] predicted differentiated thyroid cancer recurrence using many machine-learning models. Among the models examined, the XGBoost model achieved the best performance score, precision, recall, and F1-score compared to the other models- F1 coefficient 97.4% and accuracy 97.5%. Impressive results on XGBoost due to handling complex feature interactions and overfitting, especially on structured clinical data.

Clark et al. [7] used various machine-learning models to predict the recurrence of differentiated thyroid cancer. They analyzed over 16 clinical risk factors to a robust prediction system. Their study consists of many algorithms and ensemble methods, most prominently Random Forest, an ensemble of decision trees resulting in better performance than others. The accuracy is 0.99, and the F1-score is 0.98. In order to handle imbalanced data, they used SMOTE along with hyperparameter tuning, thus increasing the reliability and accuracy of the models.

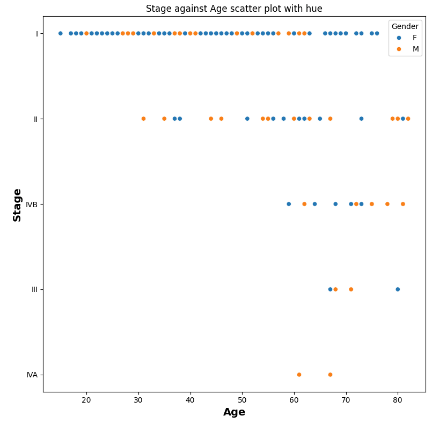
In this work, an automatic thyroid cancer recurrence prediction system has been developed employing a machine learning approach to detect thyroid cancer recurrence. We have used the Differentiated Thyroid Cancer Recurrence dataset, a public dataset. In the data preprocessing, we had to drop some values from the dataset as there were duplicate values. Then, we used the min-max scalar and standard scalar for the feature scaling to normalize the dataset. Also, we have used label and one-hot encoding to convert the categorical variables of the dataset into numerical variables. Then, we applied the holdout validation technique with Stratify to split the data into train and test. We used three algorithms for feature selection on training data: the variance threshold technique to drop constant features and Pearson’s correlation technique to figure out the statistical relation between random variables, and then we removed the features with high correlation. Then, we removed features with high correlation. Lastly, RFECV was used to select the features. After that, the SMOTE technique was used to remove the imbalance in the dataset. We have implemented machine learning models like Decision Tree, SVM, Logistic Regression, Random Forest, Gradient Boosting, XGBoost, and ensemble techniques such as Stacking and Blending. After that, the performance of these models was evaluated through test accuracy, precision, recall, F1-score, and AUC-ROC. After all, the best model prediction result is explained using the LIME explainable AI technique, which gives a visual image to understand better the model’s decision, where we can see which features were selected and were vital to predict the model’s output.

Here is the organization of the report. Section Ⅱ describes the proposed system for an automatic thyroid cancer recurrence prediction system with figures and flowcharts. In Section Ⅲ, the results are displayed and discussed. Finally, Section Ⅳ concludes the paper with some ideas of future work that can be done to improve this work.

# Proposed System

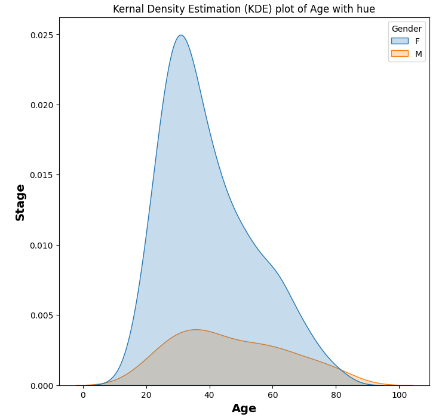
## Dataset

The Differentiated Thyroid Cancer Recurrence dataset is an open-source dataset [8]. It is a classification problem. The dataset consists of 383 instances and 16 features. Names of those features are - Age, Gender, Smoking, Hx Smoking, Hx Radiotherapy, Thyroid Function, Physical Examination, Adenopathy, Pathology, Focality, Risk, T, N, M, Stage, and Response. Among the 383 instances, 108 instances have developed thyroid cancer recurrence.



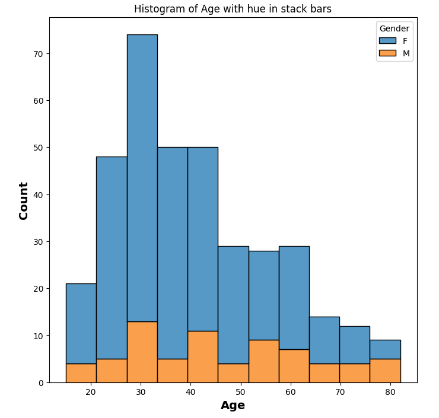
1. Scatter plot of Age with hue

Fig 1 illustrates the Scatterplot of feature Age with hue where hue is Gender.



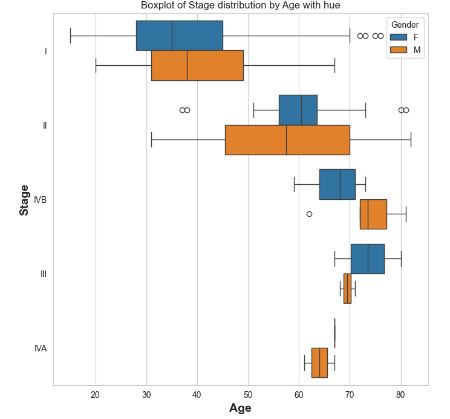
1. KDE plot of Age with hue

Fig 2 illustrates the Kernal Density Estimation (KDE) plot of feature Age with hue where hue is Gender.



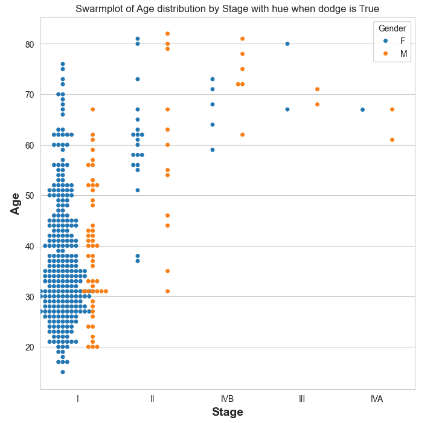
1. Histogram of Age with hue in stack bars

Fig 3 illustrates the Histogram plot of feature Age with hue in stack bars, where hue is Gender.



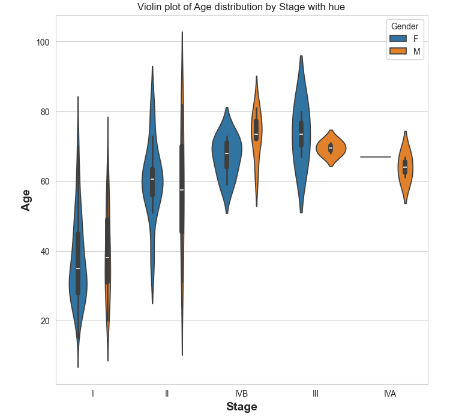
1. Boxplot of Stage distribution by Age with hue

Fig 4 illustrates the Boxplot of Stage feature distribution by Age feature with hue to detect outliers, where hue is Gender.



1. Swarmplot of Age distribution by Stage with hue when Dodge is True

Fig 5 illustrates the Swarmplot of feature Age distribution by Stage feature with hue when Dodge is true, where hue is Gender. By keeping Dodge true, we get a clear visual where the points do not intercept that much with each other.



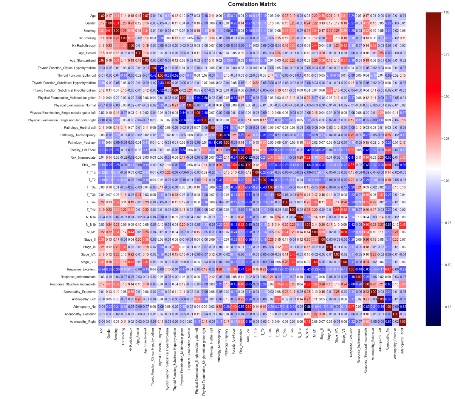
1. Violin plot of Age distribution by Stage with hue

Fig 6 illustrates the Violin plot of feature Age distribution by Stage feature with hue, where hue is Gender.

## Dataset Preprocessing

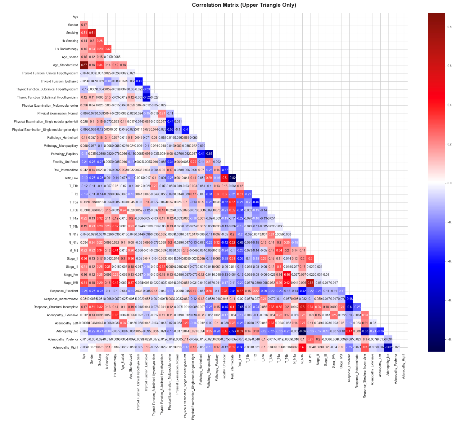
We have checked for duplicates and found some duplicate values. We dropped those duplicates. As a result, the instances of the dataset became 364, and the feature remained the same as before. Then, we investigated for null values in our dataset and found no null values. Then, we used the Min-Max scaler and Standard scaler for feature scaling to normalize the dataset. We have done Label encoding for the binary categorical columns in our dataset and One-hot encoding for the multi-class categorical columns in our dataset to convert the categorical variables into numerical values. Then, we used the holdout validation technique with Stratify to separate the train and test data to ensure the model performs well in unseen data. Then, we used three feature selection algorithms: the Variance threshold technique, Pearson’s correlation, and the Recursive Feature Elimination with cross-validation technique to remove irrelevant features from the data and improve the performance of the model.

Here are the figures of feature selection on training data and their description.



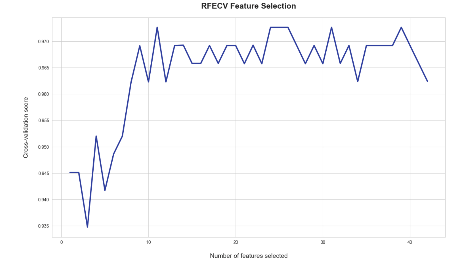
1. Correlation Matrix after using Pearson’s correlation technique for feature selection

Fig 7 illustrates the correlation matrix of training data for feature selection with Pearson’s correlation technique.



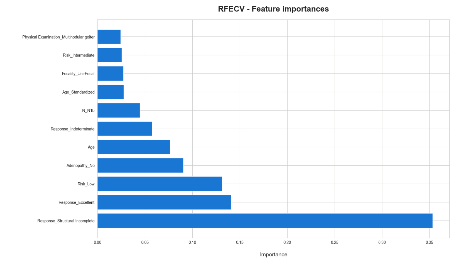
1. Correlation Matrix of the upper triangle only after using Pearson’s correlation technique for feature selection

Fig 8 illustrates the correlation matrix of the upper triangle for training data for feature selection with Pearson’s correlation technique.



1. RFECV Feature Selection

Fig 9 illustrates the RFECV Feature Selection using the Recursive Feature Elimination with cross-validation (RFECV) technique for training data, where several features are selected and their cross-validation score is shown.



1. RFECV- Feature Importances

Fig 10 illustrates the RFECV-Feature Importances using the Recursive Feature Elimination with cross-validation (RFECV) technique for training data, where selected features and their importance are shown.

In feature scaling, we have done the Min-Max scaler to normalize our data, and the data is scaled to the same range by using this equation:



Where denotes the original value, is the minimum value in the feature, is the maximum value in the feature, and is the normalized value.

After the Min-Max scaler, we have also done the Standard scaler in feature scaling to normalize our data using this equation:

(2)

Where denotes the original value, is the mean of the feature, is the standard deviation of the feature, and is the standardized value.

## Machine Learning Models

In this work, we have implemented multiple machine learning and ensemble techniques to develop the automatic thyroid cancer recurrence prediction system, briefly discussed here. Also, to prevent overfitting, we have implemented GridSearchCV to find the best hyperparameter for all the machine learning models.

### Random Forest: Random forest gives output prediction by combining the results of multiple decision trees. The data selected in the random forest for training is done randomly by using feature selection. In classification problems, the random forest gives output by using the majority voting technique, and in regression-type problems, it takes the average as the output.

### SVM: SVM can also be used for classification and regression-type problems. SVM tries to maximize the margin between the data points. It is a supervised machine learning model.

### Logistic Regression: For binary classification problems, we use the Logistic regression model. The output can be between 0 and 1. The decision boundary for logistic regression is, by default, set to 0.5.

### XGBoost: XGBoost works well in both classification and regression models. It uses the lasso and ridge technique to reduce the overfitting of the model.

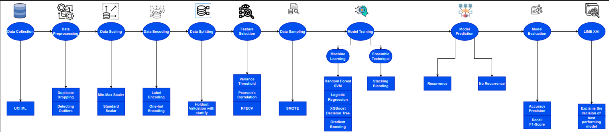
### Decision Tree: A decision tree can be used for classification and regression problems. It has a tree-like structure with three nodes. For classification problems, it uses Gini Impurity, and for regression-type problems, it finds the mean squared error value to minimize the difference between actual and predicted values.

### Gradient Boosting: Gradient Boosting uses gradient descent to minimize the loss function and reduce the gap between predicted and actual values. It can be used for classification and regression-type problems.

### Stacking: Stacking is an ensemble technique combining multiple base machine learning models, and it predicts the output. Stacking uses a cross-validation technique.

### Blending: Blending is similar to stacking, but it does not use cross-validation techniques, which makes the process faster. Blending uses the holdout validation technique.

We have added a flowchart to describe the working sequences of our thyroid cancer recurrence prediction system. The flowchart of the automatic thyroid cancer recurrence prediction system is demonstrated in Fig. 11.



1. Working sequences of the automatic thyroid cancer recurrence prediction system.

According to Fig. 11, we can see that: First, we collected our data from UCI ML, then we preprocessed the data where we dropped the duplicate values and detected outliers using Boxplot. In data scaling, we used the min-max scalar and standard scalar to normalize the dataset. Then, in data encoding, we used label and one-hot encoding to convert the categorical features into numerical values. After that, in data splitting, holdout validation is done to do train and test split. Three algorithms were applied in feature selection to remove unwanted data. In data sampling, SMOTE was applied to balance the dataset. Trained the data in eight Machine learning models, then these models predicted the output, which was later evaluated by the score of accuracy, F1, precision, recall, and AUC-ROC. LIME XAI then explains the best model found from the evaluation.

# Results and Discussion

We have done the data preprocessing and model training in Jupyter Notebook.

Here is the table illustrating all the features of various machine learning techniques for hyperparameter tuning; the values of the obtained optimized parameters are also added.

TABLE I. HYPERPARAMETER VALUES’ RANGES FOR VARIOUS ML MODELS

|  |  |  |
| --- | --- | --- |
| Model | Hyperparameter Value Range | Optimized value |
| Random Forest | n\_estimators: [200, 500], max\_features: [auto, sqrt, log2], max\_depth: [5, 10, 15] | n\_estimators: 100, max\_depth: 20, min\_samples\_split: 2 |
| SVM | C: [0.1, 1, 10, 1000], gamma: [0.1, 0.01, 0.001, 0.0001], kernel: [linear, RBF] | C: 0.1, Kernel: RBF, gamma: scale |
| Logistic Regression | C: [0.1, 1, 10], penalty: [l2] | C: 1, penalty: l2 |
| XGBoost | max\_depth: [3, 5, 7], learning\_rate: [0.01, 0.1, 0.3], n\_estimators: [100, 200] | max\_depth: 5, learning\_rate: 0.01, n\_estimators: 100 |
| Decision Tree | |  | | --- | |  |   criterion: [Gini, entropy], max\_depth: [None, 5, 10, 15] | criterion: entropy, max\_depth: 5 |
| Gradient Boosting | n\_estimators: [100, 200], learning\_rate: [0.01, 0.1, 0.5], max\_depth: [3, 5, 7]] | learning\_rate: 0.01, max\_depth: 3, n\_estimators: 100 |
| Stacking | Base models: Random Forest, SVM, Decision Tree, XGBoost | Final Estimator: Logistic Regression |
| Blending | Base models: Random Forest, SVM, Logistic Regression, XGBoost | Voting: Soft |

Table I illustrates the hyperparameter values’ ranges and the corresponding optimized hyperparameters for all the ML models implemented in this work.

We used accuracy, precision, recall, and F1-score to evaluate various Machine learning models. Equations of these metrics are:

(3)

(4)

.

(5)

(6)

Where denotes the True positive, is the true negative, is the False positive, and is the False negative.

True positive means the model’s prediction and the result are both positive. False positive means the model’s prediction is positive, but the result is negative. True negative means the model’s prediction and the result are both negative. False negative means the result is positive, but the model’s prediction is negative.

TABLE II. PERFORMANCE METRICS OF VARIOUS ML MODELS WITH DEFAULT HYPERPARAMETERS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| **Random Forest** | **0.9589** | **0.98** | **0.96** | **0.97** |
| SVM | 0.8082 | 0.84 | 0.90 | 0.87 |
| Logistic Regression | 0.9452 | 0.96 | 0.96 | 0.96 |
| XGBoost | 0.9315 | 0.96 | 0.94 | 0.95 |
| Decision Tree | 0.9041 | 0.92 | 0.94 | 0.93 |
| Gradient Boosting | 0.9315 | 0.96 | 0.94 | 0.95 |
| Stacking | 0.9452 | 0.96 | 0.96 | 0.96 |
| **Blending** | **0.9589** | **0.98** | **0.96** | **0.97** |

Performance metrics of various ML models with default hyperparameters that have been applied in this work are illustrated in Table II. According to this table, both Random Forest and Blending models perform well with default hyperparameters, and the accuracy is 0.9589, precision is 0.98, recall is 0.96, and F1-score is 0.97.

TABLE III. PERFORMANCE METRICS OF VARIOUS ML MODELS WITH OPTIMIZED HYPERPARAMETERS

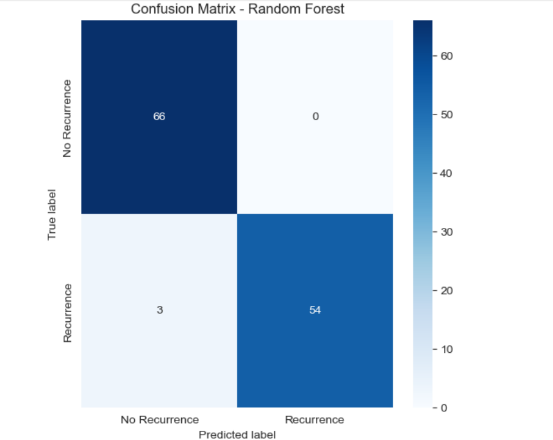
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| Random forest | 0.9708 | 0.97 | 0.97 | 0.97 |
| SVM | 0.9683 | 0.97 | 0.96 | 0.96 |
| Logistic regression | 0.9659 | 0.97 | 0.96 | 0.97 |
| XGBoost | 0.9659 | 0.97 | 0.96 | 0.96 |
| Decision Tree | 0.9610 | 0.96 | 0.95 | 0.96 |
| **Gradient Boosting** | **0.9732** | **0.97** | **0.97** | **0.97** |
| Stacking | 0.9452 | 0.96 | 0.96 | 0.96 |
| Blending | 0.9589 | 0.98 | 0.96 | 0.97 |

Performance metrics of various ML models with optimized hyperparameters that have been applied in this work are illustrated in Table III. This table shows that the gradient Boosting model performs well with optimized hyperparameters, with an accuracy of 0.9732, precision, recall, and F1-score of 0.97.

TABLE IV. PERFORMANCE METRICS OF VARIOUS ML MODELS WITH OPTIMIZED HYPERPARAMETERS AND FEATURE SELECTION

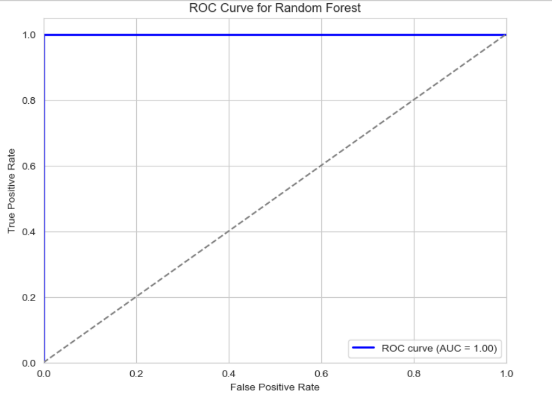
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| **Random Forest** | **0.9635** | **0.970** | **0.96** | **0.96** |
| SVM | 0.9200 | 0.93 | 0.90 | 0.91 |
| Logistic regression | 0.9500 | 0.95 | 0.94 | 0.94 |
| XGBoost | 0.9600 | 0.97 | 0.96 | 0.96 |
| Decision Tree | 0.9400 | 0.93 | 0.92 | 0.92 |
| Gradient Boosting | 0.9550 | 0.96 | 0.95 | 0.95 |
| Stacking | 0.9500 | 0.95 | 0.94 | 0.94 |
| Blending | 0.9550 | 0.96 | 0.95 | 0.95 |

Performance metrics of various ML models with optimized hyperparameters and feature selection that have been applied in this work are illustrated in Table III. From this table, we see that random Forest is the best-performing model with optimized hyperparameters and feature selection; the results for accuracy are 0.9635, precision 0.970, recall, and F1-score is 0.96.



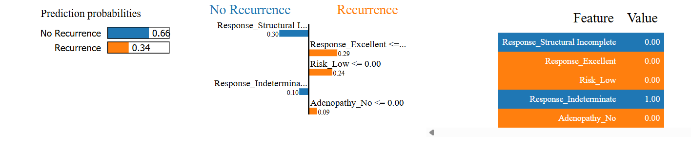
1. Confusion matrix of the best performing model: Random Forest.

Fig. 12 represents the performance of the best-performing Random Forest model in classifying Recurrence and No Recurrence for thyroid cancer. From the confusion matrix, we can see that for 54 instances, the model predicted recurrence correctly, and for 66 instances, the model predicted no recurrence correctly. There were just 3 cases where the model predicted recurrence when there was no recurrence.



1. ROC curve of the best performing model: Random Forest.

Fig. 13 represents the diagnostic ability of the best-performing model: Random Forest. True Positive measures the model’s correctness in measuring the positive cases, and False Negative denotes incorrectness. The blue line shows the model’s perfect performance as 1.00 AUC, which denotes the perfect classification ability of the model.



1. Machine learning model prediction interpretation by LIME explainable AI library.

Fig. 14 shows that Random Forest comes out as the best machine learning model after evaluating all the machine learning models, and Random Forest’s prediction interpretation is explained here by the LIME explainable AI library.

Table V. COMPARISON OF THE PROPOSED SYSTEM WITH EXISTING WORKS

|  |  |  |  |
| --- | --- | --- | --- |
| Ref. | Model | Accuracy | F1-score |
| [5] | XGBoost | 0.953 | 0.917 |
| [6] | XGBoost | 0.975 | 0.974 |
| [7] | Random Forest | 0.99 | 0.98 |
| This work | Random Forest | 0.9708 | 0.97 |

Table V illustrates a comparison of the proposed automatic thyroid cancer recurrence prediction system with other existing works on the Differentiated Thyroid Cancer Recurrence dataset, where we can see that our model Random Forest came close to the model implemented in the paper [6] and [7] in terms of accuracy and F1-score.

# Conclusions

Thyroid cancer is a common endocrine malignancy, and it has a high death rate; though it can be treated with lengthy medical treatments, there is still a considerable chance of recurring. So, predicting the recurrence of thyroid cancer earlier can reduce the death rate a lot, and it will be a great addition to medical science. This paper proposes an automatic thyroid cancer recurrence prediction system by applying various machine-learning techniques. An open-source dataset named Differentiated Thyroid Cancer Recurrence has been used for this work. The SMOTE technique is applied here to balance the dataset. This paper used accuracy, precision, recall, and F1-score performance metrics for the implemented machine-learning models. Random Forest achieved the best performance with an accuracy score of 0.9708, precision, recall, and F1-score of 0.97. The results obtained from the model can help predict the recurrence of thyroid cancer, making the treatment process quick and alerting patients about the disease. In the future, this work will have immense scope. We can use various Deep Learning models, which will improve the model result, and we can also use a private dataset, which will be beneficial from the perspective of Bangladesh.

##### References

1. G. R. Gimblet, P.Reddy, M. M. Holland, H. A. Houson, J. Whitt, J. A. Copland, S. S. Kenderian, R. Jaskula-Sztul and S. E. Lapi, “PET imaging of differentiated thyroid cancer with thyrotropin-alfa,” Scientific Reports, 2025.
2. S. Kumari, A. Makarewicz and J. Klubo-Gwiezdzinska, “Emerging Potential of Metabolomics in Thyroid Cancer—A Comprehensive Review,” Cancers, pp. 1017–1017, 2025.
3. M. Ghojazadeh, M. Mobasseri, H. Mostafaei, M. Asadizadeh-Azar, N. Kabiri, A. Kazemi, A. Lotfi, R. Aletaha, A. Akbari-Khoei and H. Salehi-Pourmehr, “The Global Trends of Thyroid Cancer Research: A Scientometric Study,” Journal of Cancer Epidemiology, 2024.
4. T. Alonso-Gordoa, P. Jimenez-Fonseca, J. Martinez-Trufero, M. Navarro, I. Porras, J. Rubió-Casadevall, M. A. Valles, N. Basté, J. Hernando and L. I. Docampo, “SEOM-GETNE-TTCC Clinical guideline thyroid cancer (2023),” Clinical & Translational Oncology, pp. 2902-2916, 2024.
5. C. Ozturk, O. Sagir and U. Vural, “Machine Learning Approaches to Predict Thyroid Cancer Recurrence: A Comparative Study,” International Conference on Computer Science and Engineering, pp. 1-6, 2024.
6. Imtiaz et al., “The Future of Differentiated Thyroid Cancer Recurrence Prediction Using a Machine Learning Framework Advancements, Challenges, and Prospects,” International Conference on Emerging Trends in Networks and Computer Communications, 2024, pp. 518-525
7. E. Clark, S. Price, T. Lucena, B. Haberlein, A. Wahbeh, and R. Seetan, “Predictive Analytics for Thyroid Cancer Recurrence: A Machine Learning Approach,” Knowledge, vol. 4, pp. 557–570, 2024.
8. S. Borzooei and A. Tarokhian. “Differentiated Thyroid Cancer Recurrence,” UCI Machine Learning Repository, 2023.