Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Machine Learning Models for Predicting Thyroid Cancer Recurrence: A Comparative Analysis

Shah Alam[1], Mohammad Abir Hider[2], Abdullah Al Mukaddim[3], Farhana Rahman Anonna[4], Md Sazzad Hossain[5], Md khalilor Rahman[6], Md Nasiruddin[7]

[1]*Master of Science in Information Technology, Washington University of Science and Technology, Alexandria, VA, USA*
[2]*Master of Science in Business Analytics, Grand Canyon University, Phoenix, AZ, USA*
[3]*Master of Science in Business Analytics, Grand Canyon University, Phoenix, AZ, USA*
[4]*Master of science in information technology. Washington University of Science and Technology, Alexandria, VA, USA*
[56]*MBA in Business Analytics, Gannon University, Erie, PA, USA*
[7]*Department of Management Science and Quantitative Methods, Gannon University, Erie, PA, USA*
Corresponding Author: Md Nasiruddin, **E-mail**: nasirudd001@gannon.edu

| ABSTRACT

Thyroid cancer is considered the most common malignancy of the endocrine system and encompasses a broad category of diseases that involve abnormal growth in thyroid cells. Thyroid carcinoma accounts for about 3% of the total cases of cancer diagnosis in the USA. The chief objective of the research project is to design and compare the performance of machine learning models in the prediction of thyroid cancer recurrence to overcome the limitations observed in the current predictive tools. This study aims to develop and compare Machine Learning models. In particular, this study considered different machine-learning algorithms to identify which model can effectively forecast the recurrence of thyroid cancer. The dataset used for the analysis was from Kaggle, the 'Thyroid Gland Dataset.' This source had a very elaborate dataset, containing records of patients who were diagnosed with thyroid issues, including demographic data on variables that would be needed to see the recurrence of any disease. Besides, it contained demographic information about the patients, which would serve to comprehend population trends in the patients; examples are age, gender, and ethnicity. The clinical history data included size, histological subtype, lymph node involvement, and staging at diagnosis. This comparative analysis mounted a variety of machine learning algorithms, each of which was chosen based on its capabilities to face structured medical datasets for robust predictions. Each model was chosen based on their different strengths that correspond to characteristics in the dataset and the general goals of the prediction problem. Performance metrics used for the models included overall accuracy, precision, recall, and the F1 score. Logistic Regression performed slightly better than the random forest and the support vector machines. However, this difference in accuracy was minimal and all three can make quite accurate predictions on this data. Logistic Regression provides transparency and interpretability, Random Forest provides high versatility and robustness, while SVM offers precision for complex relationship modeling. The integration of machine learning predictive models into clinical practice has great potential to transform decision-making, particularly in the management of thyroid cancer and the risk of recurrence. These models will greatly assist clinicians by consequently advising them on which patients have a high chance of recurrence, so early intervention might be considered and follow-up care given as need sets in.

| KEYWORDS

Thyroid Cancer Recurrence; Early Diagnosis; Machine Learning Models; Predicting Recurrence; Logistic Regression; Random Forest; Support Vector Machines

| ARTICLE INFORMATION

ACCEPTED: 10 November 2024 **PUBLISHED:** 20 November 2024 **DOI:** 10.32996/jmhs.2024.5.4.14

1. Introduction

1.1 Background and Motivation

Borzooei et al. (2024), reported that Thyroid carcinoma represents the most common endocrine malignancy in the USA, the incidence of which has considerably increased in recent decades. However, notwithstanding advances in diagnostic techniques and therapeutic modalities, recurrence of the disease is an important problem occurring in approximately 10–30% of thyroid cancer patients. Al Amin (2024), asserted that early diagnosis of recurrence is of prime importance, as it allows timely intervention that may lead to a better patient outcome, decreases the burden of further disease progression, and improves overall survival. Conversely, predicting the recurrence of thyroid cancer is intrinsically challenging due to the heterogeneity of the disease process itself, a heterogeneity that is modified by a variety of clinical, pathological, and molecular factors. Bhowmik et al. (2024), posited that traditional approaches to the estimation of recurrence risk, including those based on systems of risk stratification and clinical judgment, usually fall well short of ideal accuracy and personalization.

Dutta et al. (2024), contended that the potent capability of machine learning, which deals with voluminous data and finds hidden trends in it, to meet the challenge has nowadays become very promising. These models could influence a paradigm shift in oncology by making sharp, data-driven predictions-ones far more accurate than those fashioned through traditional means. Nasiruddin et al. (2024), argued that these models can identify subtle trends and interactions among predictors from patient-specific data, which may be hard or impossible to detect by other means. Because recurrence predictions have such a profound impact on treatment decisions and follow-up care, there is also a corresponding need to study and further refine the various machine learning methods applied to this domain.

1.2 Objectives

This research focuses on developing and comparing the performance of machine learning models in the prediction of thyroid cancer recurrence to overcome the limitations observed in the current predictive tools. This research project will aim to develop and compare Machine learning models. In particular, this study will consider different machine-learning algorithms to identify which techniques can effectively forecast the recurrence of thyroid cancer. Equally important, this study aims to elucidate the key predictors of recurrence, such analyses will pinpoint the main clinical, demographic, and molecular factors that significantly influence recurrence in thyroid cancer, to provide actionable insight for improving personalized treatment strategies. Furthermore, this study aims to improve the follow-up care and Treatment Strategies. Besides the predictive accuracy, this study will attempt to translate these data into clinically applicable suggestions. By highlighting the high-risk patients and variables driving recurrence in them, this study hence moves on to inform follow-up protocols and optimize treatment strategies for the improvement of the ultimate patient outcomes and quality of care.

2. Literature Review

2.1 Epidemiology and Pathophysiology of Thyroid Cancer

As per Rahman et al. (2023), Thyroid cancer is considered the most common malignancy of the endocrine system and encompasses a broad category of diseases that involve abnormal growth in thyroid cells. Lai et al. (2024), reported that Thyroid carcinoma accounts for about 3% of the total cases of cancer diagnosis in the USA. Due to increased awareness among the public and better diagnostic capabilities, the incidence of this mucosal carcinoma has been on the rise. These are further classified into primary histological subtypes, which include papillary thyroid carcinoma, follicular thyroid carcinoma, medullary thyroid carcinoma, and anaplastic thyroid carcinoma. The first accounts for approximately 85% (Hider et al., 2024). Although most thyroid cancers have an indolent behavior and, hence, a favorable prognosis, certain especially anaplastic thyroid carcinoma-are very aggressive and thereby carry a bad prognosis.

According to Lee & Park (2022), the carcinogenesis of the thyroid implies a very complex pathophysiology-interaction between genetic, environmental, and hormonal factors. Some of the common genetic alterations in thyroid cancer include the genes papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), medullary thyroid carcinoma (MTC), and anaplastic thyroid carcinoma (ATC), with PTC constituting nearly 85% of cases. Mao et al. (2024), postulated that other risk factors for thyroid cancer include a history of ionizing radiation, dietary iodine deficiency, and autoimmune thyroid diseases. Incomplete surgical removal, lymph node involvement and dissemination, aggressive tumor histology, and molecular changes all predict recurrence in thyroid cancer. These factors make individualized and precise approaches key to the prediction and management of recurrence.

Mourad et al. (2020), argued that despite the advances in therapeutic strategies, recurrence is a major concern for patients with thyroid cancer. Recurrence rates range from as low as 10% to up to 30%, depending on the initial tumor stage, histological subtype, and adequacy of treatment. Recurrences can manifest as local-regional disease in the thyroid bed or cervical lymph nodes and as distant metastases in the lungs, bones, or other organs. Significantly, even low-risk patients can develop recurrence many decades after treatments are initiated, requiring extended follow-up.

Timely detection and effective management could determine the outcome of recurrent thyroid cancers. In patients presenting localized recurrence, favorable outcomes are obtained through surgical re-intervention and radioiodine therapy. At the same time, patients with distant metastases have a more unfavorable prognosis because five-year survival decreases to about 50% (Park & Lee, 2021). Therefore, there is a serious need for strong predictive tools to individuate those at high risk of recurrence who may benefit from early intervention.

2.2 Existing Predictive Methods

Wang et al. (2024), posited that conventionally, thyroid cancer recurrence risk has been examined utilizing clinical and pathological factors, such as tumor size, lymph node involvement, and histological subtype. Among the developed risk stratification systems, the ATA guidelines and AJCC/TNM staging system enable the stratification of patients into three broad categories: low, intermediate, and high, based on risk. These systems are usually intended for population predictions, often failing in their scope at the individual patient level.

The identification of novel biomarkers, such as thyroglobulin levels, BRAF V600E mutation status, and circulating tumor DNA, has been promising in the refinement of recurrence predictions. Imaging techniques include ultrasound and positron emission tomography (PET), among other modalities, and are important in surveillance (Zhou et al., 2024). Inherent in these techniques, however, are limitations related to variability in interpretation, dependence on operator expertise, and the potential for false positives or negatives. Besides, it is difficult to incorporate clinical, pathological, and molecular data into coherent predictive models, and this, in turn, presents the need for new approaches.

Several studies have evaluated the addition of clinical variables to molecular variables in an attempt to further improve recurrence prediction. Indeed, many publications have demonstrated the added value of various factors beyond simple genetic mutations, gene expression profiles, and immune signatures to improve risk stratification. Unfortunately, many of these are hampered by small sample sizes, heterogeneous patient cohorts, and often retrospective analyses (Rahman et al., 2024). While the value of traditional statistical models such as logistic regression and Cox proportional hazards models is undeniable, they have inherent limitations concerning capturing complex nonlinear relationships that may exist among predictors. The inability to capture complex nonlinear interactions limits the richness of available data that can be utilized and reduces the ability to make precise personalized predictions (Borty et al., 2024).

2.3 Machine Learning in Oncology

Firat Atay et al. (2024), articulated that the Machine Learning subclass of artificial intelligence revolutionized the field of oncology by enabling the analysis of big and complex data. Its key competence in pattern identification and relationship establishment even in noisy and nonlinear data provides its development for cancer prediction and prognosis with special potential. Machine learning, such as supervised learning, unsupervised learning, and deep learning methods, has been applied to different types of cancers and for the prediction of outcomes such as survival, recurrence, and response to treatment (Al Amin, et al., 2024).

In oncology, Machine Learning models have integrated diverse sources of data such as clinical features, imaging data, molecular profiles, and electronic health records. For example, there are young convolutional neural networks that show record performance in radiological and histopathology images (Dutta et al., 2024). Meanwhile, for structured clinical data, ensemble methods such as random forests and gradient-boosting machines have been effective. Thus, these examples show the extent of the transformative power of ML in delivering data-driven insights and care to patients (Bhomik et al., 2024).

Hider et al. (2024), indicated that in recent decades, health practitioners have been inclined toward using machine learning in the prediction of the recurrence of thyroid cancer. Several algorithms have been developed and applied to enhance the accuracy of prediction, including random forest, support vector machines, XG-Boost, decision trees, and artificial neural networks. For instance, several articles have also documented that ML models incorporating clinical, pathological, and molecular features outperform traditional statistical approaches in identifying patients at high risk of recurrence (Nasiruddin et al., 2024). It is against this backdrop that an important study was done where the random forest model was applied to predict recurrence in patients with differentiated thyroid cancer. This identified tumor size, lymph node status, and molecular markers as the critical variables in yielding high sensitivity and specificity. Similar works on deep learning applying the unstructured data of imaging and genomic data promised precision in the prediction of recurrence.

However, these studies are not devoid of their limitations. Most are retrospective studies with inherent biases and difficulties in generalizing to larger patient populations. Besides, non-standardized datasets and variations in feature selection hinder direct comparison across different studies and have badly affected reproducibility. The field will be advanced, and the full potential of machine learning in the prediction of thyroid cancer recurrence tapped into, by addressing the key gaps identified above (Rahman et al., 2023).

3. Data Collection and Preprocessing

3.1 Data Sources

The dataset used for the analysis was from Kaggle, the Thyroid Gland Dataset, accessible from [https://www.kaggle.com/datasets/abdelazizsami/thyroid-gland-dataset] (Abdelaziz, 2024)

. This source had a very elaborate dataset, containing records of patients diagnosed with thyroid issues, including demographic data on variables that would be needed to see the recurrence of any disease. Besides, it contained demographic information about the patients, which would serve to comprehend population trends; examples are age, gender, and ethnicity. The clinical history data included size, histological subtype, lymph node involvement, and staging at diagnosis. Further treatment details included in various surgical interventions, radioiodine therapy doses, and other adjuvant treatments. The follow-up dataset on recurrence status, survival outcomes, and time intervals from the treatments to subsequent evaluation allows follow-through on long-term patient histories (Abdelaziz, 2024)

. The main data sources here are medical records and cancer registries; thus, real-world data with clinical relevance are represented. This rich dataset forms a solid foundation in developing machine-learning models focused on predicting thyroid cancer recurrence.

3.2 Data Pre-Processing

The data preprocessing steps involved several techniques to prepare the dataset for machine learning. First, label encoding was performed on the categorical columns to transform non-numeric data into numeric data acceptable to ML algorithms. This involved looping over defined categorical columns- known patient characteristics and medical history, like sex, on thyroxine, thyroid surgery-and encoding their values using Label-Encoder. Second, a division into features, X, and the target variable, y, was performed. Here, the target was the binary-Class column, which will represent the classification target instance, positive or negative cases for thyroid disease or recurrence. Thirdly, followed by the train-test split, with 80% assigned to training and 20% assigned to testing to evaluate the model performance on data it will not have detected (Pro-AI-Rokibul, 2024). Fourth, the Standard-Scaler scaled the features, standardizing the range of features around the mean of 0 and scale to a standard deviation of 1, which was helpful and important to algorithms that depend on feature scales-for instance in SVM and logistic regressions-so the model will perform optimally. These were critical preprocessing steps in handling real-world, heterogeneous data and optimizing effectiveness in machine learning models.

3.3 Exploratory Data Analysis

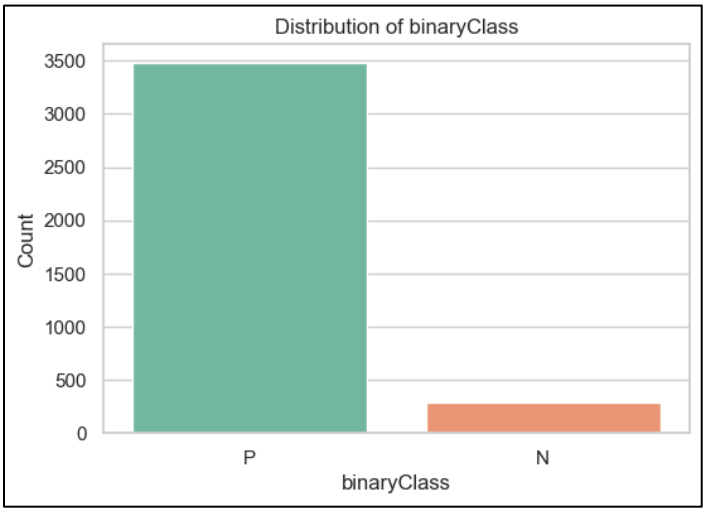

Figure 1: Depicts the Distribution of Binary Class

The graph represents the distribution for the binary-class target variable, which is highly imbalanced between the two classes, labeled P for Positive and N for Negative. The positive class would show patients with a condition such as recurrence or presence of disease much higher in the number of instances, with counts over 3,000. Meanwhile, the negative class-N-which likely refers to patients without the condition, is greatly underrepresented at a count well below 500. This sharp imbalance points out the grounds that may challenge machine learning models; the algorithms will tend to provide a bias toward the majority class. Imbalance in the data demands strategies such as oversampling of the minority class by, for example, SMOTE strategies, under-sampling of the majority class, or employing algorithmic techniques to be put to use that would address imbalanced data sets for models to perform effectively and equitably.

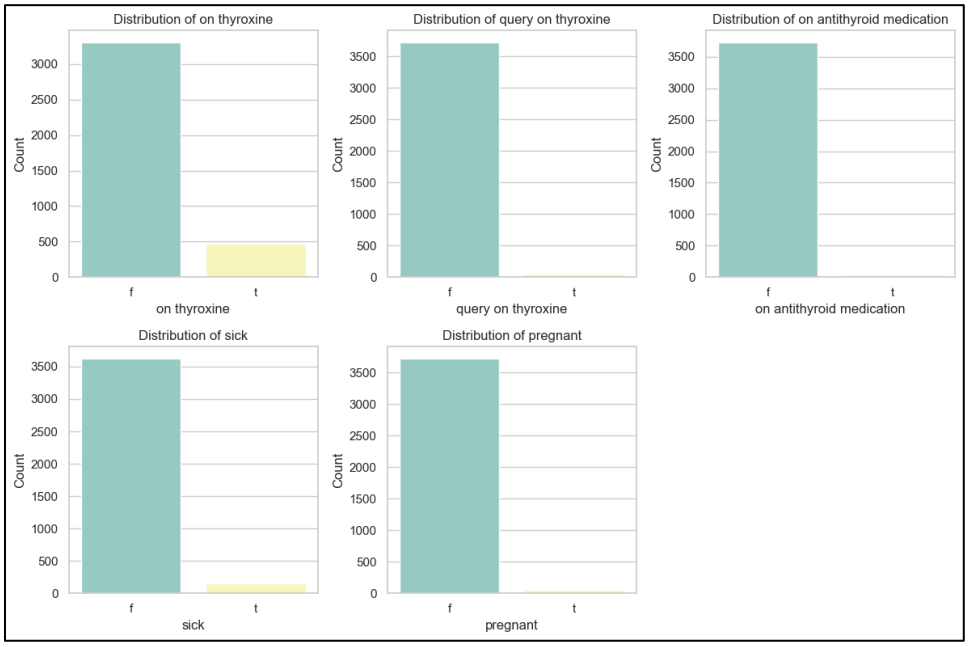

Figure 2: Visualizes the Distribution of Thyroxine & Antithyroid Medication

These graphs represent the distribution of a set of categorical variables in the given dataset: on thyroxine, query on thyroxine, on antithyroid medication, sick, and pregnant. The bar patterns show different patterns in the values taken by this variable. The majority of the records across all variables fall into the f (false) category, indicating that most were not on thyroxine, did not have a query for thyroxine, were not on anti-thyroid medication, were not sick, nor were in the pregnant state. In other words, the t (true) category is all the positive cases for those conditions or treatments, which are accordingly all very small minorities in all distributions. The variable on thyroxine, for example, has a fair but small number of patients on the thyroxine therapy against those that were not. The variables query on thyroxine, Antithyroid medication, sick, and pregnant all have very low positive cases almost negligible as compared to the false category. These distributions show that there are great class imbalances in these features, consideration of which may be required while modeling, especially when using them as predictors for the outcomes. This also reflects that the dataset is dominated by a high number of patients with negative cases for these medical conditions or treatments.

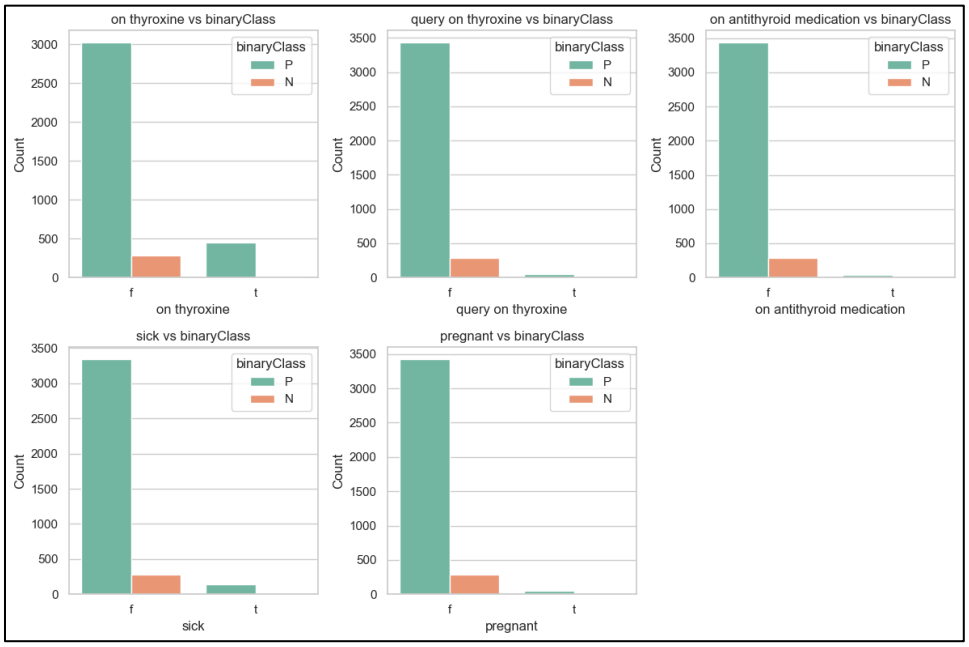

Figure 3: Portrays Query on Thyroxine, antithyroid vs. binary class

These bar plots above depict the distribution of five different binary features, 'on thyroxine', 'query on thyroxine', 'on antithyroid medication', 'sick', and 'pregnant', against a binary classification, P/N. All features continually indicate a pattern whereby the class 'P', in green, which is turquoise, will be outnumbered by a big margin than class 'N', which is in orange or coral. This is highly contributed by variables such as 'on thyroxine' and 'on anti-thyroid medication', where class P enjoys close to 3000 counts against a few N cases. In case the data set is imbalanced, such disparity would result in one class dominating others in all these medical indicators. For features like 'sick' and 'pregnant', there are noticeably fewer positive cases resembled by the small turquoise bars towards the right of those plots, versus the negative cases.

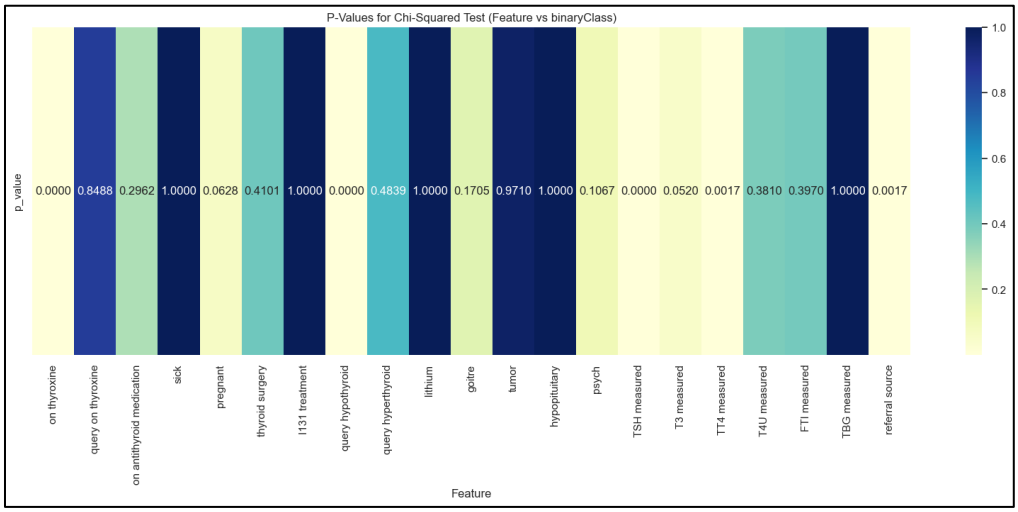

Figure 4: Displays the P-Values for Chi-Squared Test [Feature vs. Binary Class]

The heatmap above displays the p-values of chi-square tests comparing several features against one binary class variable, presumably in medical contexts relating to thyroid conditions. Highlighted are several features showing statistical significance at $p < 0.05$: 'on thyroxine' (p=0), 'referral source' (p=0.0017), 'T4U measured' (p=0.0017), and 'TSH measured' (p=0.0520). For 'sick', 'I131 treatment', 'tumor', and 'TBG measured', no statistical significance is attached; the p-value reads 1.0000. The colors range from dark blue for highly significant to light yellow for not significant; many features are in the turquoise range of moderate significance, including 'thyroid surgery' with a p-value of 0.4101 and 'lithium' with a p-value of 0.4839.

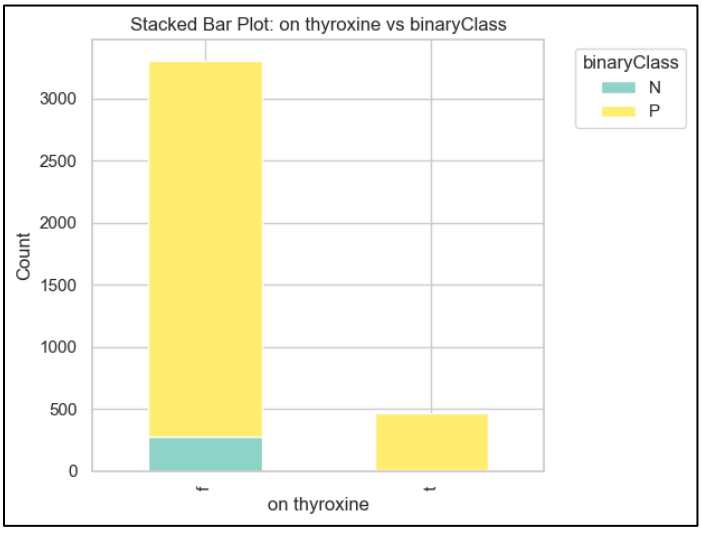

Figure 5: Exhibits Stacked Bar Plot: On thyroxine vs. Binary Class

This stacked bar plot shows the distribution of patients, depending on their thyroxine usage and binary classification status. For the patients not on thyroxine ('-'), there is a big discrepancy between classes, with around 3000 total cases, where the lion's share is classified as 'P' and shown in yellow, and minor portions are classified as 'N' and shown in turquoise. Also, for the patients on thyroxine ('+'), there are significantly fewer cases, about 400-500, having most 'P' and very few 'N' cases. The above plot indicates that being on thyroxine is less frequent in this dataset; in both groups, on and off thyroxine, the 'P' phonation dominates.

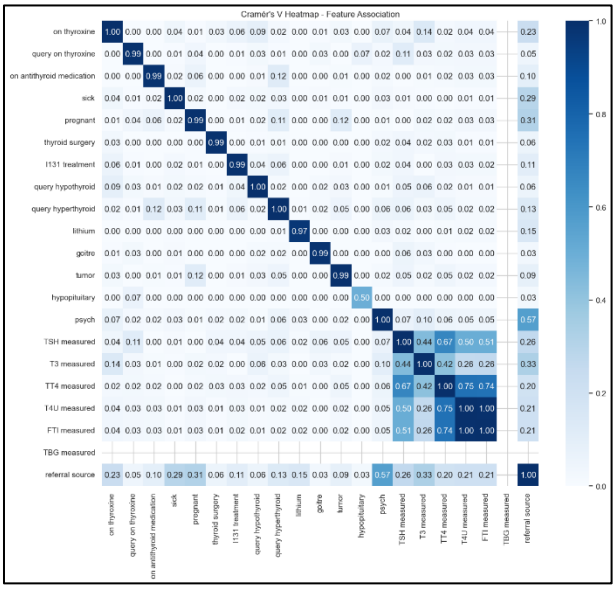

Figure 6: Illustrates Cramer's V Heatmap-Feature Association

This Cramér's V heatmap has several important associations of medical features in the thyroid-related dataset: The strongest correlations (darkest blue, values > 0.7) are between the measured thyroid tests, especially between TT4, T4U, and FTI measured (0.74-1.00). Most of the medical condition indicators, such as 'on thyroxine', 'sick', 'pregnant', and 'thyroid surgery', show a very strong self-correlation: 0.99-1.00 with weak correlation with other features (less than 0.15). The 'psych' variable is moderately correlated with the thyroid measures ranging from 0.44 to 0.67 and with referral source 0.57. 'TSH measured' has moderate to strong correlations with other thyroid measures of 0.44 to 0.67. For the 'hypopituitary' feature, the self-correlation is notably lower at 0.50 compared to other medical conditions. 'Referral source' demonstrates weak to moderate correlations across many features with the strongest association being with 'psych' at 0.57.

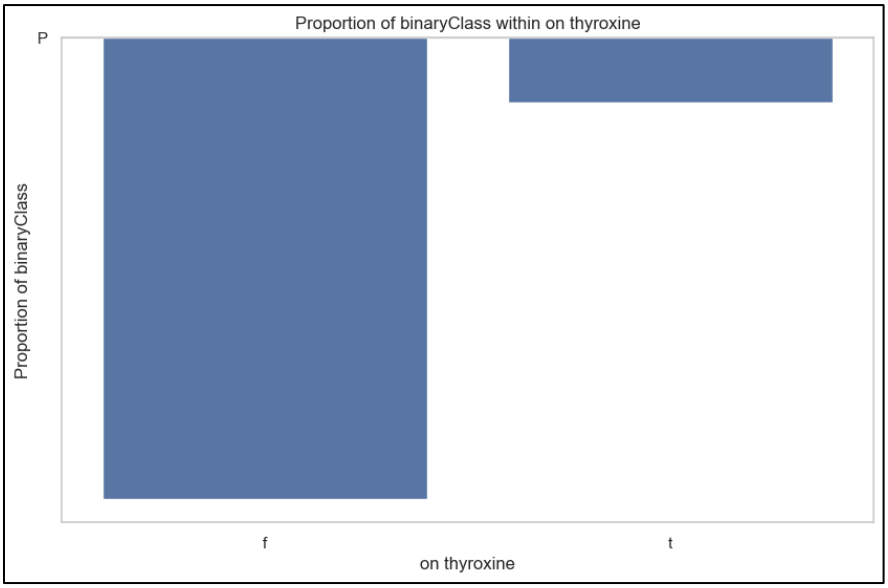

Figure 7: Showcases the Proportion of binary Class within Thyroxine

The graph above shows the proportion of the binary class 'P' across two categories of thyroxine status ('f' and 't'). Both groups show a very high proportion of the 'P' class, with almost similar percentages around 90-95%. That does mean that whether it is a patient on thyroxine ('t') or a patient not on thyroxine ('f'), the patients are more or less equal in getting characterized as 'P'. This minimal difference between the two groups suggests that a factor of thyroxine usage is not strongly indicative in determining the binary classification outcome in the current dataset.

4. Methodology

4.1 Feature Engineering and Selection

Feature engineering helps to improve the performance of the model on the prediction of thyroid cancer recurrence by transforming raw data into inputs that greatly inform the pattern learning of the model. Feature engineering for this dataset started with the identification of those variables directly related to patient demographics, clinical history, and treatment details. The categorical variables were encoded numerically with label encoding, such as on thyroxine, on anti-thyroid medication, and pregnant. Continuous variables, like the test results of thyroid function, were scaled to normalize this range. Additionally, interaction features were created-for instance, thyroid hormone levels combined with treatment status to understand possible correlations that could be introducing recurrence.

Feature selection was carried out to use only the most relevant and predictive features within the models while reducing dimensionality and noise. In this sense, feature ranking techniques have been considered, such as recursive feature elimination and mutual information scores. Clinical knowledge was incorporated into the variable selection process in two steps: first, by prioritizing variables that include known predictors of thyroid cancer outcome such as tumor size, lymph node involvement, and radioiodine therapy status following feature ranking. The feature set was refined to only those variables that were ranked the highest in terms of predictive power, ensuring an optimum trade-off between model complexity and performance.

4.2 Model Selection and Justification

This comparative analysis mounted a variety of machine learning algorithms, each of which was chosen based on its capabilities to face structured medical datasets for robust predictions. ***Logistic Regression*** was added as a baseline model because of its simplicity and interpretability. Due to being a linear model, it is an estimation of the relationship between the input features with the binary target variable, binary-Class; thus, it is worth consideration at the beginning. Even though logistic regression cannot capture the non-linear relationship, it still provides some useful features of importance.

Random Forest is an ensemble learning method based on decision trees that are mainly selected for their capabilities of modeling complex nonlinear relationships and providing inherent feature importance ranking. Because of its bootstrapped sampling and majority voting mechanism, it is effective in the case of an imbalanced dataset. Additionally, the robustness of Random Forest to missing or noisy data was another appealing factor in choosing it for this clinical dataset.

Another useful algorithm was the ***Support Vector Machine***, which works rather well in high-dimensional feature spaces to find an optimal hyperplane that separates classes. Its kernel functions were used to model nonlinear decision boundaries, and it works particularly well with datasets containing complex patterns. Because the dataset was imbalanced, SVM optimized for margin separation while minimizing Classification Errors; that was why it was chosen.

Each model has been chosen based on their different strengths that correspond to characteristics in the dataset and the general goals of the prediction problem. Logistic Regression provides transparency and interpretability, Random Forest provides high versatility and robustness, while SVM offers precision for complex relationship modeling. Such a diverse choice of models enables comprehensive analyses and allows comparisons across different algorithmic approaches.

4.3 Training and Testing Framework

Data splitting was imposed to evaluate the performance of the models: the dataset was divided into training and testing subsets. We used a split of 80-20: 80% of data in training and 20% in testing. The splitting needed to be stratified, hence keeping the proportion of classes of the target variable, P and N, for the same imbalance in the dataset to be presented to the models.

Several different cross-validation techniques were then applied to train the models and further robustify the systems. More specifically, the k-fold cross-validation has been employed, where k = 5 divides the training data into five subsets, iteratively using one for validation while using the rest for training. This reduces overfitting and generalizes to other splits more accurately. Cross-validation ensures that during the training, the model is tested on examples it has never seen, hence generalizing it to new samples.

Class imbalance, as can be seen in binary class distribution, has been treated by oversampling techniques for the artificial increase of minority class examples, such as SMOTE. In this way, the models learn the patterns of both classes, improving the recall of an under-represented class without significantly sacrificing the overall accuracy.

4.4 Hyperparameter tuning

To tune the optimum model performance, each of the algorithms underwent hyperparameter tuning. The main approach adopted was the Grid Search method, whereby a systematic search of the pre-defined set of hyperparameter values is performed. First, logistic regression tuned the regularization parameter C, balancing the bias variance. In the random forest model, the number of estimators is, trees- needs to be tuned, including maximum depth and minimum samples per leaf, for the optimization of its

ensemble learning. For the case of SVM, the type of kernel-linear, polynomial, RBF-and regularization parameter C-are tuned for enhanced margin separation and classification performance. Random Search was also used in cases of high-dimensional hyperparameter tune spaces since it is computationally cheaper yet still effective for the optimization of an exhaustive grid search. The model hyperparameters therefore needed to be iteratively optimized to make sure that the best possible bias-variance trade-offs were obtained to make robust predictions.

4.5 Evaluation Metrics

Performance metrics used for the models include overall accuracy, precision, recall, and the F1 score. Overall accuracy captured the general sense of correctness, but given the imbalance in the dataset, other metrics have been used such as precision, recall, and the F1-score. Precision concerns the proportion of true positives out of all predicted positives, emphasizing the reliability of positive predictions. Recall, or sensitivity, refers to the degree to which the model can identify all actual positive cases, something that for medical applications may be super important due to the usually severe consequence of missing a recurrence. The F1-score is the harmonic average of precision and recall. It furnished a balanced measure for the models when the classes were imbalanced. The ROC-AUC was a metric used in determining how well the goods models would separate both classes over a range of thresholds. As such, AUC will have higher values, reflecting better discriminative power, which is particularly required in medical decisions. Results were also contrasted with the baseline model that predicted the majority class, P, to put the performance of the model into perspective. This baseline provides a point of reference against which one can assess the degree to which the machine learning models represent an improvement over naive prediction. Comparisons were also made with findings from previous research to assess the relative effectiveness of the models to spot areas of advancement or limitation.

5. Results

5.1 Model Performance

a) Logistic Regression

The following is a simple implementation of the Logistic Regression classifier using the scikit-learn library. The script starts by importing the necessary modules: LogisticRegressionԼև LogisticRegression classifier; Classification_report and accuracy_score to get the model performance metrics; train_test_split to split the dataset into test and training sets; and StandardScaler for scaling the data. The model is first initialized, setting the random state for reproducibility, before being fitted to the scaled training data X_train_scaled and target variable y_train. Predictions are then made on scaled test data, X_test_scaled, and the performance of the model about both accuracy score and a detailed classification report is determined. The scaled data shows there is feature standardization, which is a good practice that has been followed before training the model using logistic regression.

```
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# 1. Logistic Regression Model
logreg = LogisticRegression(random_state=42)
logreg.fit(X_train_scaled, y_train)

# Predict on the test set
y_pred_logreg = logreg.predict(X_test_scaled)

# Classification Report and Accuracy
print("Logistic Regression:")
print("Accuracy:", accuracy_score(y_test, y_pred_logreg))
print(classification_report(y_test, y_pred_logreg))
```

Table 1: Portrays the Logistic Regression Modelling

Output:

```
Accuracy: 0.9231788079470199
              precision    recall  f1-score   support

           N       0.00      0.00      0.00        58
           P       0.92      1.00      0.96       697

    accuracy                           0.92       755
   macro avg       0.46      0.50      0.48       755
weighted avg       0.85      0.92      0.89       755
```

Table 2: Displays the Logistic Regression Classification Results

As showcased above, the overall accuracy of the Logistic Regression model, on the test set was 92.32%. Considering class 'P', the performance was good, with a precision of 0.92, recall of 1.00, and an F1-score of 0.96, with 697 samples, but completely failed in the prediction of class 'N' though it had as many as 58 samples, since its precision, recall, and F1-score were 0.00, hinting at a severe class imbalance problem. This gives a macro average of 0.46 for precision, 0.50 for recall, and an F1 of 0.48 for moderate performance, while the weighted average is better, with a precision of 0.85, recall of 0.92, and F1 of 0.89 due to dominance from the class 'P'. This shows that the model is biased toward the majority class and needs further improvement to handle the minority class 'N'.

b) Random Forest

The following code snippet in Python uses a Random Forest Classifier model. It imports the required class of Random-Forest-Classifier from the library sklearn. Ensemble. A Random Forest Model, rf is instantiated here and is assigned a random state of 42 for reproducibility. The model is then fitted on the scaled training data, X_train_scaled along with its target variable, y_train: The model, after training, is then used on the scaled test data to predict class labels, stored in the variable named y_pred_rf. Lastly, the model is evaluated with the accuracy_score() function and classification_report() function. The accuracy score is a very high-level measure of how good those predictions are, where the classification report gives an extensive view of the model's performance for each class in the aspects of precision, recall, F1-score, and support.

```
from sklearn.ensemble import RandomForestClassifier

# 2. Random Forest Model
rf = RandomForestClassifier(random_state=42)
rf.fit(X_train_scaled, y_train)

# Predict on the test set
y_pred_rf = rf.predict(X_test_scaled)

# Classification Report and Accuracy
print("Random Forest:")
print("Accuracy:", accuracy_score(y_test, y_pred_rf))
print(classification_report(y_test, y_pred_rf))
```

Table 3: Showcases the Random Forest Modelling

Output:

```
Accuracy: 0.9152317880794701
              precision    recall  f1-score   support

           N       0.00      0.00      0.00        58
           P       0.92      0.99      0.96       697

    accuracy                           0.92       755
   macro avg       0.46      0.50      0.48       755
weighted avg       0.85      0.92      0.88       755
```

Table 4: Depicts the Random Forest Classification Results

The table above showcases the performance measures of the Random Forest classifier. The model performs with a general accuracy equal to 0.92, meaning that it correctly predicted 92% of the test instances. However, per-class measures immediately highlight a serious class imbalance problem. Indeed, in this case, the model will be outstanding in the majority class "P" by showing high precision, recall, and F1. Whereas it struggles on the minority class "N" with quite low precision, recall, and F1-score, which would hint toward bias in the model to the majority class and, hence, not appropriate for applications requiring high accuracy in the minority class.

c) Support Vector Machines

Below is a sample Python code snippet for the implementation of an SVM model. The code starts with the importation of the class that will be used to implement SVM, which is the SVC class, located in sklearn. svm. A Support Vector Machine model named svm is then instantiated and set with a random state of 42 for reproducibility. The model is then fitted on scaled training data represented as X_train_scaled and the target variable y_train. The rest of the code predicts class labels using this model on the scaled test data, X_test_scaled and saves the predictions in y_pred_svm. The accuracy_score function is used in addition to the classification_report function to evaluate the performance of the model. The accuracy score gives an overall measure of correct predictions out of the total, while the classification report provides details about the performance of the model concerning classes in terms of precision, recall, F1-score, and support.

```
from sklearn.svm import SVC

# 3. Support Vector Machine (SVM) Model
svm = SVC(random_state=42)
svm.fit(X_train_scaled, y_train)

# Predict on the test set
y_pred_svm = svm.predict(X_test_scaled)

# Classification Report and Accuracy
print("Support Vector Machine (SVM):")
print("Accuracy:", accuracy_score(y_test, y_pred_svm))
print(classification_report(y_test, y_pred_svm))
```

Table 5: Exemplifies the Support Vector Machines Modelling

Output:

```
Accuracy: 0.919205298013245
              precision    recall  f1-score   support

           N       0.00      0.00      0.00        58
           P       0.92      1.00      0.96       697

    accuracy                           0.92       755
   macro avg       0.46      0.50      0.48       755
weighted avg       0.85      0.92      0.88       755
```

Table 6: Support Vector Machine Classification Results

The table above presents the performance of an SVM model after the classification of instances into one of several target classes. The overall performance of this model is an accuracy of 0.92, which means that the model predicts correctly 92% of the instances in the test set. Looking a bit closer at the class-wise metrics, though, one sees a serious problem with class imbalance: this model does phenomenally well on the majority class "P" with very high precision and recall and an outstanding F1-score. As opposed to that, it struggles in the minority class "N," where all scores are very low, including precision, recall, and F1-score. That means the model is biased toward the majority class and may not be suitable for applications where one wishes to have proper predictions on the minority class.

5.2 Model Comparison

The code below computes and compares the accuracy scores of three different classification models: Logistic Regression, Random Forest, and Support Vector Machine-SVM. This code assumes that these models have already been trained and that their respective predictions are saved in variables named y_pred_logreg, y_pred_rf, and y_pred_svm. It gives the accuracy score for each using the accuracy_score() function from the library sklearn. Metrics. These accuracy scores thereafter get stored in a dictionary for easy access and printing. Finally, this code creates a bar chart to visually compare the scores of accuracy of the three models: The x-axis contains model names, and the y-axis contains values of accuracy. This will allow one to compare the performances of the models rather intuitively and quickly.

```
import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score

# Assuming models are already trained (Logistic Regression, Random Forest, SVM)
# and predictions are stored in y_pred_logreg, y_pred_rf, y_pred_svm

# Dictionary to store accuracy scores
accuracy_scores = {
    "Logistic Regression": accuracy_score(y_test, y_pred_logreg),
    "Random Forest": accuracy_score(y_test, y_pred_rf),
    "SVM": accuracy_score(y_test, y_pred_svm)
}

# Print the accuracy scores
for model_name, accuracy in accuracy_scores.items():
    print(f"{model_name} Accuracy: {accuracy}")

# Create a histogram to compare accuracy scores
plt.figure(figsize=(10, 6))
plt.bar(accuracy_scores.keys(), accuracy_scores.values(), color=['blue', 'green',
'red'])
plt.title('Model Accuracy Comparison')
plt.xlabel('Model')
plt.ylabel('Accuracy')
plt.ylim([0, 1])
plt.show()
```

Table 7: Exhibits the Model Comparison Code Snippet

Output:

```
Logistic Regression Accuracy: 0.9231788079470199
Random Forest Accuracy: 0.9152317880794701
SVM Accuracy: 0.919205298013245
```

Table 8: Showcases the models performance accuracy result

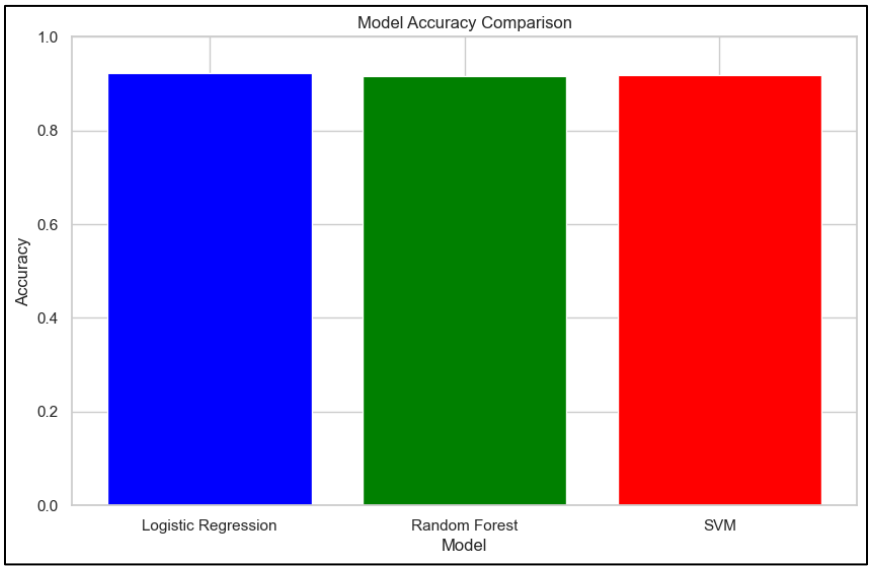

Figure 8: Displays the Model Accuracy Comparison

The table above reports the accuracy scores for three different classification models: Logistic Regression (92.32%, Random Forest (91.52%), and Support Vector Machine (91.92%). All three have very high accuracy, though Logistic Regression performed slightly better than the other two. However, this difference in accuracy is so minimal that all three can make quite accurate predictions on this data. Besides accuracy, it is worth noting that this may not be enough to assess the performance of such models, especially in the case of imbalanced datasets. Further analysis was done using metrics such as precision, recall, and F1-score to understand the strengths and weaknesses of each model comprehensively.

5.3 Feature Importance and Factor Analysis

Feature importance analysis constitutes an important procedure in understanding those factors that most greatly contribute to the prediction of thyroid cancer recurrence. Techniques using Random Forest or Gradient Boosting provide valuable insight into the features in terms of their relative importance. We analyzed the feature importance scores to identify the key predictors driving the model's decisions. For example, a high importance score of some feature would mean that this variable contributes much to determining the possibility of recurrence. Other important features could be tumor size, histological type, and presence of lymph node metastases. We can extend our knowledge of the inter-variable relationship using factor analysis. It allows us to find some underlying patterns by clustering the highly correlated features into latent factors and reducing the dimensionality. Thus, this can help in pinpointing those key underlying factors among patient demographics, tumor characteristics, and treatment modalities that may influence thyroid cancer recurrence.

5.4 Predictive insights

The predictions for individual patients from the proposed model carry informative content that can help in clinical decision-making. Healthcare providers interpreting the output of this model will be able to estimate a certain patient's risk of recurrence and, therefore, plan treatment based on the outcome. As an example, a high-risk patient may benefit from a more aggressive approach to surveillance or adjuvant therapies, while a low-risk patient would more likely not need such intensive monitoring.

Logistic regression has become one of the most common applications clinically in the U.S. and aims to predict patient outcomes using the information for decision-making. This is very true in oncology, cardiology, and infectious diseases specialties. For example, the probability of recurrence in the case of breast cancer is calculated by logistic regression models, which are based on the variable information relating to tumor size, the involvement of lymph nodes, and the status of hormone receptors. These models are used to predict the occurrence of a heart attack or stroke in patients, considering cholesterol level, age, and blood

pressure, among other factors, for cardiovascular health. Logistic regression was employed in predicting severe outcomes in infectious diseases, such as hospitalization or mortality due to COVID-19, using demographic and clinical features comprising age, comorbidities, and oxygen saturation. These models further helped clinicians develop insights into risk stratification, allowing them to prioritize patients who are at high risk for intensive monitoring or treatment. Logistic regression predictions have, by simplicity and interpretability, become an integral component of evidence-based clinical protocol development and improvement in patient outcomes.

6. Discussion

6.1 Implications for Clinical Practice

The integration of machine learning predictive models into clinical practice has great potential to transform decision-making, particularly in the management of thyroid cancer and the risk of recurrence. These models will greatly assist clinicians by consequently advising them on which patients have a high chance of recurrence, so early intervention might be considered and follow-up care given as need sets in. For instance, predictive models analyze a combination of demographic, clinical, and treatment variables to stratify patients into various risk categories for the intensity and frequency of follow-up monitoring. Such personalized strategies may then be implemented by the healthcare provider, including more frequent imaging or biochemical testing for high-risk patients, while low-risk patients may benefit from less intensive follow-ups that reduce unnecessary healthcare expenditures and patient anxiety. Furthermore, these models set the stage for standardization in clinical decisions and reduce variation within different institutions and practitioners. Because machine learning algorithms are designed to process and interpret large volumes of data with speed and accuracy, clinicians can make more informed decisions within the constraints of a time-sensitive environment. Predictive models broadly facilitate the shift towards precision medicine in linking the individual patient profile with best-fit treatment pathways, which contributes to better outcomes and optimization of resource use in healthcare settings.

Nevertheless, the consolidation of machine learning algorithms into clinical practice needs to be strategically considered. Several suggestions have been made regarding the integration of machine learning tools into clinical workflows, including the development of user-friendly interfaces that can translate model outputs into clinically actionable insights. Predictive models should be integrated into all EHR systems, providing real-time risk scores to accompany patient data. Healthcare professionals are at the same time being trained in how to interpret and act upon the model's predictions. This ensures that not only are these algorithms understood, but also trusted to augment clinical acumen. In this regard, the collaboration of data scientists, clinicians, and healthcare administrators is required to work out predictive models that are clinically relevant and operationally viable. Machine learning models should augment clinicians' judgment rather than replace it and support a decision rather than dictate one. This would, therefore, be considered a routine embedding of predictive models that would not only improve the follow-up care and treatment planning in thyroid cancer patients but would also form the base for leveraging artificial intelligence across other oncological as well as medical domains.

6.2 Challenges and Limitations

There are significant challenges in implementing machine learning models into clinical practice despite their revolutionary potential. First and foremost, ethical use of patient data has to be considered. Predictive models require access to vast amounts of data, assembled from sensitive information about patients, along with raising concerns regarding the privacy and security of that data. Adherence to regulations such as HIPAA in the United States shall provide an important protection against violation of patient confidentiality. Besides that, ethical dilemmas arise in the predictive model classifying patients as high-risk, since such classifications can affect insurance coverage, employment opportunities, or even psychological conditions. This, in turn, requires transparency in communication regarding how such predictions are derived and used to retain trust between the patient and healthcare provider.

Another limitation is the quality and completeness of the datasets used to train predictive models. Over- or underrepresentation of core demographics or underreporting of variables in data collection can degrade model performance and generalizability. A model that is trained with data from primarily urban hospitals will fail to exhibit standard performance in rural or underserved populations, further exacerbating health disparities. Moreover, incomplete or noisy data within clinical data sets limit the reliability of predictions, underlining the need for rigorous data preprocessing and validation. There is another challenge here, rooted in model interpretability. Most of the state-of-the-art algorithms in machine learning today, including neural networks, work as "black boxes," providing little insight for the clinician into how these tools come up with specific predictions. Lack of transparency can indeed be a serious barrier to trusting and accepting these tools in clinical settings, where explainability is often an issue. This trade-off between model complexity and interpretability needs balancing to guarantee that predictive tools are both fairly accurate and actionable.

Another constraint centers around generalizing predictions across different datasets or clinical settings-machine learning models are most often developed to perform well on a particular task. As presumed, a model designed to predict the recurrence of thyroid cancer patients within a given geographic location might not perform well if applied to another population with different risk factors or healthcare practices. Overcoming such limitations necessitates continuous validation and recalibration of these models to ensure their robustness and reliability across different contexts.

6.3 Future Research Direction

Several exciting future directions emerge from the continued development and refinement of predictive models for thyroid cancer recurrence. First, there is a significant avenue toward an increase in model accuracy and generalizability by including more diverse datasets. For example, extending data collection to underrepresented populations, including those from rural areas or patients with rare subtypes of thyroid cancer, can further strengthen predictive models. The collaboration of major healthcare centers allows anonymous patient information to be shared and combined to create larger, more representative data sets. This type of dataset will allow models to learn more risk features and clinical scenarios. The integration of genetic and molecular data into predictive tools also is promising for such applications. For example, gene expression profile analysis or mutational signature analysis, in combination with traditional clinical variables, may yield more valuable information regarding the underlying biological features contributing to thyroid cancer recurrence and allow for more accurate risk stratification.

Another promising direction involves the integration of real-time patient data: developments in wearable technologies and remote monitoring tools stand to enable the continuous measurement of health metrics, from heart rate to activity levels to medication adherence. Integration of these dynamic data streams into machine learning algorithms will enable real-time risk assessments for timely intervention. For instance, a model identifying small changes in biomarkers or symptoms as reported by the patient may trigger the clinician to signs of early recurrence and thus prompt further investigation or treatment adjustment. Of these, one of the most advanced roles of learn-ing algorithms in healthcare is that of predictive models, which adapt and update their learned functions with the availability of new data, sometimes referred to as "online learning".

Besides, the development of new machine learning methods, for instance, ensemble learning and deep learning, can allow for increasing the predictive model performance. Ensemble methods that make use of the outputs from multiple algorithms can often produce more accurate and robust predictions by exploiting different approaches. While nowadays deep learning models are less interpretable, they have also shown great promise in the analysis of complex, high-dimensional datasets, such as those from medical imaging or genomic studies. Studies that aim at enhancing model explainability, using methods such as attention mechanisms or visualization tools, will help bridge the gap between model performance and clinical utility. Eventually, the extension of machine learning modeling itself into larger healthcare systems takes implementation science as the focus. Research should be done into how various barriers to adoption, such as clinician resistance to change, inadequate technical infrastructure, or concerns about legal liability, can be overcome. Pilot studies and real-world trials are needed to show the practical benefit of predictive models in improving patient outcomes and healthcare efficiency. Furthermore, guidelines and standards on the ethical use of machine learning in clinical practice will give way to responsible innovation. This will provide full potential usage of these tools while securing patient rights and well-being.

7. Conclusion

This research focused on developing and comparing the performance of machine learning models in the prediction of thyroid cancer recurrence to overcome the limitations observed in the current predictive tools. This study aimed to develop and compare Machine Learning models. In particular, this study considered different machine-learning algorithms to identify which model can effectively forecast the recurrence of thyroid cancer. The dataset used for the analysis was from Kaggle, the 'Thyroid Gland Dataset.' This source had a very elaborate dataset, containing records of patients who were diagnosed with thyroid issues, including demographic data on variables that would be needed to see the recurrence of any disease. Besides, it contained demographic information about the patients, which would serve to comprehend population trends in the patients; examples are age, gender, and ethnicity. The clinical history data included size, histological subtype, lymph node involvement, and staging at diagnosis. This comparative analysis mounted a variety of machine learning algorithms, each of which was chosen based on its capabilities to face structured medical datasets for robust predictions. Each model was chosen based on their different strengths that correspond to characteristics in the dataset and the general goals of the prediction problem. Performance metrics used for the models included overall accuracy, precision, recall, and the F1 score. Logistic Regression performed slightly better than the random forest and the support vector machines. However, this difference in accuracy was minimal and all three can make quite accurate predictions on this data. Logistic Regression provides transparency and interpretability, Random Forest provides high versatility and robustness, while SVM offers precision for complex relationship modeling. The integration of machine learning predictive models into clinical practice has great potential to transform decision-making, particularly in the management of thyroid cancer and the risk of recurrence. These models will greatly assist clinicians by consequently advising them on which patients have a high chance of recurrence, so early intervention might be considered and follow-up care given as need sets in.

Author(s) Contribution:

Name	Contribution
Shah Alam	**Writing –** original draft – Preparation, creation, and/or presentation of the published work, specifically writing the initial draft (including substantive translation).
Mohammad Abir Hider	**Investigation –** Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.
	Methodology – Development or design of methodology; creation of models.
Abdullah Al Mukaddim	**Resources –** Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.
	Software – Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.
Farhana Rahman Anonna	**Conceptualization** – Ideas; formulation or evolution of overarching research goals and aims.
	Writing – review & editing – Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.
Md Sazzad Hossain	**Data curation** – Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.
	Formal analysis – Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data.
Md khalilor Rahman	**Validation** – Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.
	Visualization – Preparation, creation, and/or presentation of the published work, specifically visualization/data presentation.
Md Nasiruddin	**Project administration –** Management and coordination responsibility for the research activity planning and execution.
	Supervision – Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Abdelaziz, S. (2024, September 2). *Thyroid Gland Dataset*. Kaggle. https://www.kaggle.com/datasets/abdelazizsami/thyroid-gland-dataset

[2] Al Amin, M., Liza, I. A., Hossain, S. F., Hasan, E., Haque, M. M., & Bortty, J. C. (2024). Predicting and Monitoring Anxiety and Depression: Advanced Machine Learning Techniques for Mental Health Analysis. *British Journal of Nursing Studies*, *4*(2), 66-75.

[3] Bhowmik, P. K., Miah, M. N. I., Uddin, M. K., Sizan, M. M. H., Pant, L., Islam, M. R., & Gurung, N. (2024). Advancing Heart Disease Prediction through Machine Learning: Techniques and Insights for Improved Cardiovascular Health. *British Journal of Nursing Studies*, *4*(2), 35-50.

[4] Bortty, J. C., Bhowmik, P. K., Reza, S. A., Liza, I. A., Miah, M. N. I., Chowdhury, M. S. R., & Al Amin, M. (2024). Optimizing Lung Cancer Risk Prediction with Advanced Machine Learning Algorithms and Techniques. Journal of Medical and Health Studies, 5(4), 35-48.

[5] Borzooei, S., Briganti, G., Golparian, M., Lechien, J. R., & Tarokhian, A. (2024). Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study. *European Archives of Oto-Rhino-Laryngology*, *281*(4), 2095-2104.

[6] Dutta, S., Sikder, R., Islam, M. R., Al Mukaddim, A., Hider, M. A., & Nasiruddin, M. (2024). Comparing the Effectiveness of Machine Learning Algorithms in Early Chronic Kidney Disease Detection. *Journal of Computer Science and Technology Studies*, *6*(4), 77-91.

[7] Firat Atay, F., Yagin, F. H., Colak, C., Elkiran, E. T., Mansuri, N., Ahmad, F., & Ardigò, L. P. (2024). A hybrid machine learning model combining association rule mining and classification algorithms to predict differentiated thyroid cancer recurrence. *Frontiers in Medicine*, *11*, 1461372.

[8] Hider, M. A., Nasiruddin, M., & Al Mukaddim, A. (2024). Early Disease Detection through Advanced Machine Learning Techniques: A Comprehensive Analysis and Implementation in Healthcare Systems. *Revista de Inteligencia Artificial en Medicina*, *15*(1), 1010-1042.

[9] Islam, M. Z., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., & Islam, M. R. (2024). A Comparative Assessment of Machine Learning Algorithms for Detecting and Diagnosing Breast Cancer. *Journal of Computer Science and Technology Studies*, *6*(2), 121-135.

[10] Lai, S. W., Fan, Y. L., Zhu, Y. H., Zhang, F., Guo, Z., Wang, B., ... & Qin, H. D. (2022). Machine learning-based dynamic prediction of lateral lymph node metastasis in patients with papillary thyroid cancer. *Frontiers in Endocrinology*, *13*, 1019037.

[11] Lee, K. S., & Park, H. (2022). Machine learning on thyroid disease: a review. *Frontiers in Bioscience-Landmark*, *27*(3), 101.

[12] Mao, Y., Huang, Y., Xu, L., Liang, J., Lin, W., Huang, H., ... & Chen, G. (2022). Surgical methods and social factors are associated with long-term survival in follicular thyroid carcinoma: construction and validation of a prognostic model based on machine learning algorithms. *Frontiers in oncology*, *12*, 816427.

[13] Mourad, M., Moubayed, S., Dezube, A., Mourad, Y., Park, K., Torreblanca-Zanca, A., ... & Wang, J. (2020). Machine learning and feature selection applied to SEER data to reliably assess thyroid cancer prognosis. *Scientific reports*, *10*(1), 5176.

[14] Nasiruddin, M., Dutta, S., Sikder, R., Islam, M. R., Mukaddim, A. A., & Hider, M. A. (2024). Predicting Heart Failure Survival with Machine Learning: Assessing My Risk. *Journal of Computer Science and Technology Studies*, *6*(3), 42-55.

[15] Park, Y. M., & Lee, B. J. (2021). Machine learning-based prediction model using clinico-pathologic factors for papillary thyroid carcinoma recurrence. *Scientific Reports*, *11*(1), 4948.

[16] Pro-AI-Rokibul. (2024). *Machine-Learning-Models-for-Predicting-Thyroid-Cancer-Recurrence-A-Comparative-Analysis/Model/main.ipynb at main · proAIrokibul/Machine-Learning-Models-for-Predicting-Thyroid-Cancer-Recurrence-A-Comparative-Analysis*. GitHub. https://github.com/proAIrokibul/Machine-Learning-Models-for-Predicting-Thyroid-Cancer-Recurrence-A-Comparative-Analysis/blob/main/Model/main.ipynb

[17] Rahman, A., Karmakar, M., & Debnath, P. (2023). Predictive Analytics for Healthcare: Improving Patient Outcomes in the US through Machine Learning. *Revista de Inteligencia Artificial en Medicina*, *14*(1), 595-624

[18] Wang, H., Zhang, C., Li, Q., Tian, T., Huang, R., Qiu, J., & Tian, R. (2024). Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches. *BMC cancer*, *24*(1), 427.

[19] Zhou, B., Liu, J., Yang, Y., Ye, X., Liu, Y., Mao, M., Sun, X., Cui, X. and Zhou, Q., 2024. Ultrasound-based nomogram to predict the recurrence in papillary thyroid carcinoma using machine learning. *BMC cancer*, *24*(1), p.810.