Predicting Both Hypothyroidism and Hyperthyroidism Using Machine Learning Techniques

Submitted By

Mokarrom Basher

ID: 213-35-3191

Department of Software Engineering

Daffodil International University

Supervised By

Dr. Md. Fazla Elahe

Assistant Professor and Associate Head

Department of Software Engineering

Daffodil International University

This Report is Presented in Partial Fulfilment of the Requirements for the Degree of Bachelor of Science in Software Engineering.

Summer-2025

©Daffodil International University

APPROVAL

This thesis titled on "**Predicting Both Hypothyroidism and Hyperthyroidism Using Machine Learning Techniques"** submitted by **Mokarrom Basher (ID: 213-35-3191**) to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Fazla Elahe

--- **Chairman**

Dr. Md. Fazla Elahe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Marzia

--- **Internal Examiner 1**

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13/09/2025

--- **Internal Examiner 2**

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

13.09.25

--- **External Examiner**

Mohammod Abul Kashem
Professor
Department of Computer Science and Engineering
Dhaka University of Engineering and Technology, Gazipur

©Daffodil International University

SUPERVISOR's DECLARATION

I hereby declare that this thesis report is done by me under the supervision of **Dr. Md. Fazla Elahe**, Assistant Professor and Associate Head, Department of Software Engineering, Daffodil International University, in fulfillment of my original work. I am also declaring that according to the best of my knowledge, neither this thesis nor any part therefore has been submitted else here

Fazla Elahe

(Supervisor's Signature)

Assistant Professor and Associate Head
Department of Software Engineering
Daffodil International University

©Daffodil International University

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

Mokarrom Basher.

(Student's Signature)

Full Name : Mokarrom Basher

ID Number : 213-35-3191

Batch : 36th

Department of Software Engineering

Daffodil International University

©Daffodil International University

ACKNOWLEDGEMENTS

First and foremost, I express my deepest gratitude to Almighty Allah, whose divine blessings have enabled me to successfully complete my final year thesis.

I am sincerely grateful to my respected research supervisor, **Dr. Md. Fazla Elahe**, for his continuous guidance, invaluable advice, and support throughout the entire research process.

I would also like to extend my heartfelt thanks to **Dr. Imran Mahmud**, Head of the Faculty, for providing all the necessary facilities required for the successful completion of this research. My sincere appreciation also goes to all the lecturers of the Department of Software Engineering for their kind guidance and encouragement during times of difficulty.

I am deeply thankful to my friends for their encouragement, assistance, and constant support during this journey.

Finally, I gratefully acknowledge with due respect the unwavering support and inspiration of my family, without which this thesis would not have been possib

©Daffodil International University

ABSTRACT

The aim of this project is to predict hypothyroidism and hyperthyroidism with the help of multiple machine learning (ML) methods. The research employs a clinical dataset that comes with vital characteristics of the thyroid including Thyroid Stimulating Hormone (TSH), Triiodothyronine(T3), Thyroxine (TT4), and Free Thyroxine Index (FTI) and demographic characteristic of the patients including age and gender. The data goes through several preprocessing stages, such as dealing with missing and encoding discrete values; and normalizing numerical scores to achieve successful performance of the model. Various models of machine learning such as Decision Trees, Random Forest, Support Vector Machines (SVM), Naive Bayes, and Gradient Boosting are applied and checked according to the accuracy, precision, recall, and F1-score. Hyperparameter optimisation is also done to enhance the effectiveness of the model. The findings indicate that the overall predictive accuracy and balance between recall and precision of any methods indicate a higher performance of the Random Forest, so prediction of thyroid disorders is possible using it. This study exemplifies the significance of preprocessing data and selection of machine learning models in the healthcare use case and provides understanding of the potential to utilize the concept of machine learning in clinical settings and utilize it as a preliminary method of the thyroid disease diagnosis.

©Daffodil International University

KEYWORD

1. Thyroid Disorder
2. Hypothyroidism
3. Hyperthyroidism
4. Machine Learning
5. Healthcare Prediction

©Daffodil International University

TABLE OF CONTENT

©Daffodil International University

©Daffodil International University

LIST OF TABLES

©Daffodil International University

©Daffodil International University

LIST OF FIGURES

©Daffodil International University

©Daffodil International University

LIST OF ABBREVIATIONS

Abbreviation	Full Form
TSH	Thyroid Stimulating Hormone
T3	Triiodothyronine
TT4	Total Thyroxine
T4U	Thyroxine Uptake
FTI	Free Thyroxine Index
TBG	Thyroxine-Binding Globulin
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
XGBoost	Extreme Gradient Boosting
ML	Machine Learning
CDSS	Clinical Decision Support System
ROC-AUC	Receiver Operating Characteristic - Area Under Curve
NB	Naive Bayes
AdaBoost	Adaptive Boosting
SD	Standard Deviation
PCA	Principal Component Analysis
F1-score	Harmonic Mean of Precision and Recall
CV	Cross-Validation

©Daffodil International University

INTRODUCTION

1.1 Background and Motivation

Thyroid diseases, hyperthyroidism and hypothyroidism, are two of the most common endocrine diseases known depending on the world, millions of people of various ages, places of residence and backgrounds are introduced to this disease. This blood gland is called the thyroid gland and is shaped into a butterfly size and is found on the neck side of the body that regulates important functions of the body such as metabolism energy production, growth, development among others. It does it with the secretion of thyroid hormones, mainly, thyroxine (T4) and triiodothyronine (T3). Thyroid-Stimulating Hormone secreted by the pituitary gland regulates the production and release of these hormones. Excess or insufficient production of these hormones by the thyroid affects the body-balance of the body negatively and the resulting health complications may be of a serious nature.

Hypothyroidism involves a condition that occurs when the thyroid does not synthesize the thyroid hormones into appropriate levels, in which metabolism is slowed down. This leads to various symptoms, such as exhaustion, weight gain, cold-intolerance and depression. Hyperthyroidism, on the other hand, is a condition where the thyroid gland produces excessively the thyroid hormones and acceleration of the metabolism results in such symptoms as weight loss, high pulse rate, anxiety, tremors and sweating. Although there are evident physiological markers of both the conditions, the symptoms manifested by them may often seem similar to the ones presented by other medical conditions and it is hard to distinguish between thyroid-related problems and other familiar ailments.

When left untreated, both hypothyroidism and hyperthyroidism can have long-lasting and sometimes life-threatening effects. For example, hypothyroidism, if left unmanaged, can lead to cardiovascular problems like high cholesterol and atherosclerosis, and an increased risk of osteoporosis due to the metabolic slowdown. Quite the contrary, untreated hypothyroidism can also predispose the patient to cardiovascular problems (e.g., and even more inhumanely, in the worst case, hypothyroidism may. cause thyroid storm, assessable death syndrome, and induce development of other. worse symptoms

©Daffodil International University

such as heart fainting and high fever. Further, the two conditions are related to mental issues, such as depression, anxiety, as well as brain problems.

Diagnosing thyroid disorders is not a simple issue considering the high health risks posed by the disorders. Although laboratory analysis of TSH, plasma level of T3 and T4 are usually good methods in the determination of balancing and misbalancing of thyroid hormones, diagnosis is normally prone to a relative fallacy and a clinical mistrust. The analyses are not definite (this is especially where it is mild or subclinical, in which instance I refer to it when there is a possibility that there is something somewhat out of the way in the manner it performs its functioning, but which is no longer while it is presenting certain symptoms). In these, there are by delay or no diagnosis particularly when the symptom being experienced are mild or in fact the symptoms may even be mixed with other diseases. Furthermore, when making such laboratory tests, the data gained using the latter may prove significant; it is to be processed by a healthcare professional who does not always have available resources and cannot make the corresponding diagnosis.

The current practices of diagnosis can be guided mainly by a synthesis of clinical assessment, laboratory examinations, and self-reporting of the symptoms, but these techniques do not have no limitations. Clinical assessments are subject in nature and the evaluation the doctor has made can differ widely based on their experience and interpretation. Laboratory tests also, helpful as they are, can be time consuming processes to process a result, as well as, their interpretation, which may be susceptible to error, especially handling high, or high complexity data. This leaves the possibility of misdiagnoses, late treatments as well as missed thyroid dysfunction. Healthcare systems, which are not adequately prepared to attend to thyroid disorders (e.g. due to lack of resources or personnel to do so) also contribute to exacerbation in resource constrained settings like those lacking access to endocrinologists that might face a delay in the provision of lab results and thus continue to delay the diagnosis and treatment process due to the lack of adequate resources.

©Daffodil International University

As an example, in most developing nations, endocrinological treatment of thyroid disorders is substandard with patients possibly waiting weeks or even months before getting medical test outcomes or visiting an endocrinologist. Such delay may contribute to the development of symptoms, which may cause more serious health issues that would have not appeared under the impact of early intervention. Furthermore, most of the patients with thyroid can resort to the services of general practitioners that might lack the necessary specialized knowledge which could achieve accurate diagnoses of thyroid dysfunction hence poor or late diagnosis.

Additionally, some thyroid conditions, especially in the early stages, may present with non-specific symptoms such as fatigue, weight changes, or changes in mood, which are common to many other illnesses. This complicates the effort by doctors to properly test the thyroid by relying on clinical examination only or initial blood work/testing especially when patients might have other underlying conditions that might have been contributing to the same symptoms.

This research relies on the theoretical input of applying machine learning in curbing such problems in thyroid disorders diagnosis. The steps of the diagnosis process can be and need to be automated, the role of humans can be minimized, and those involved with the healthcare industry can be assisted in making the next right choice through the establishment and implementation of machine learning models. The approach might greatly reduce the gap between crisis and diagnoses, provide timely care treatment, and reduce total expenditure of health care and in the regions where fabric can the least access specialist health care. First, predictive model has a significant role to play in identifying the individuals at risk in the earlier stages and more specifically to mince the treatment regime amongst individuals who will ultimately come out to ameliorate the health of individuals and groups in managing thyroid diseases.

Through the application of the diagnostic process that incorporates machine learning and available clinical data, this research hopes to achieve a considerable breakthrough in

©Daffodil International University

diagnosis and treatment of thyroid disorder and eventual improvement in the standard of healthcare delivery all over the globe, in general.

1.2 Purpose and Goal of the Project

The purpose of the proposed project is to design machine learning, which will be capable of predicting hypothyroidism and hyperthyroidism based on the clinical and biochemical data. This is because it is aimed at coming up with accurate and efficient models to pin down whether one has thyroid diseases or no by predicting it using predictors that consist of TSH, T3, T4, age, and gender. Having done the preprocessing of the data, optimized hyper parameters and tested the models, there is hope that this project would contribute towards making the diagnostic results of the project more accurate, and help to detect the thyroid orders early and enable the health workers to make a decision based on available information. Lastly, the project aims at enhancing the implementation of machine learning in the medical sector to change patient outcome.

©Daffodil International University

Literature Review

Literature Review

The authors of Patel et al. [1] forecasted hypothyroidism and hyperthyroidism with the help of various machine learning models. They used the Indian hospitals clinical records in the form of their dataset, and the records included the features and clinical history of TSH and T3 and T4 and other clinical indicators. Out of the available models, the XGBoost was better than the Random Forest since it was found to have higher accuracy of 94.3, as opposed to 92 with the Random Forest. The authors have underlined the importance of ensuring that XGBoost has a high performance because it is able to capture the non-linear interactions and capability to utilize the missing values extremely well, which is prevalent in clinical datasets. Another point that the study exposed is that despite the strength, Random Forest had problems with data sets that had duplicate feature scores otherwise it misclassified. The authors concluded that XGBoost was a stronger and more effective predictive algorithm to apply in the Indian health care environment since the correct diagnosis of thyroid disease is especially important, as thyroid-related disorders are one of the most widespread among the population.

The Sayeed and his groupmate [2] had made a submission concentrated on predicting floods in Bangladesh using machine learning models, which is of very high significance because the country is susceptible to natural disaster. The data was in the form of historical yearly rain and yearly occurrences of floods in 34 stations during the period 1980 to 2020. Amongst the models being tested, the accuracy of the Logistic Regression appears to be the largest at 86.76 and the lowest at 83 as in the case of both SVC and Decision Trees. The authors observed that predictive capabilities of the models were limited because of the minimal number of features involved, since other factors that used included the water level in rivers, the temperature, soil type, and humidity were not involved. They further noted that Logistic Regression was more effective as the dataset was relatively smaller and was mostly linear unlike more complex models such as SVC who could not connect well with other non-rich features. The research found out that although the Logistic Regression was promising, future model applications need to

©Daffodil International University

include hydrological and climatic data to fetter more precise and early warnings of floods in Bangladesh.

Yadav et al. [3] used Logistic Regression, SVM and Random Forest to give a predictor of hypothyroidism based on Coat of Arms on the UCI Machine Learning Repository. Random Forest was the most precise model at 90 percent; SVM has the second highest accuracy of 88 percent and the last but not least is the Logistic Regression which is 82 percent. The authors stated that Random Forest performed superiorly in complex interaction of features in the dataset which consisted of noisy and overlapping clinical attributes. The reasons behind then poor performance of the Logistic Regression is that the assumption of linear relations cannot always be to be used in capturing the thyroid patterns. SVM had good competitive results but needed to allocate higher levels of computational warmth than Ring Forest. The research, therefore, established that the predictor of thyroid diseases via an ensemble model such as a Random Forest can be trusted more than the other single model when there were several interacting variables among variables in the data. They, also, were able to add that additional gains could be made through incorporation of feature engineering and testing on bigger, realistic health information.

To predict hyperthyroidism in their study using clinical records, Gupta et al. [4] were using Support Vector Machines (SVM) in New Delhi, India. The accuracy of the study was 91 and this was considerably in light of the tiny size of the data. SVM was well with its ability to work on its capability of working on high-dimensional feature space and demonstrate optimum. characteristic segregation of planes. However, the authors were aware of limitations of. it is particularly due to the small size of the study sample and the type of clinical data they used. characteristics that could capture generalization. They also observed that SVM though was an excellent predictor, did not possess the prediction-HIGH interpretation potential which. is among the critical needs in medical decision making. The work highlighted enhancing datasets and the fusion of features selection models with an aim to span their significance. improve accuracy and model reliability in a model in the future.

©Daffodil International University

Sharma et al. [5] examined the deep learning methods that include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) as predictors of thyroid disorders. They obtained their dataset through the public repositories provided on Kaggle that offered them a wide variety of attribute related to thyroid. CNN model had the best accuracy of 92.7, which surpassed other traditional models such as the Random Forest. According to the authors, the breakthrough was due to CNN capacity of extracting automatically the spatial and hierarchical features of received data that is difficult to be reflected through traditional models of the ML. Also, RNN was tested that had similar performance, particularly with regard to time series patient data. Nonetheless, the research identified the difficulties of deep learning, including the need to have very large datasets, big computational capacity, and sensitive parameter optimization. The conclusion that the authors made is that deep learning methods have a high potential in the prediction of thyroid diseases but require the integration of explainability methods to guarantee their utilization by the clinical community.

Kumar et al. [6] compared the prediction of thyroid disorders through application of the Random Forest and AdaBoost algorithms to hospital clinical data in India. They found that Random Forest was most accurate, with the accuracy of 94.5 and AdaBoost was 91. The authors described that the advantage of Random Forest was its effectiveness in the merging of various of their various decision trees, which substantially lower background and nocturnal personalities in the information. Conversely, AdaBoost did not handle misclassified data points well, thereby recording slightly poor performance. The research pointed out that in a situation where a study involves clinical data where missing values and noise are more likely to exist, ensemble models such as Random Forest are more resilient. They also recommended that more refinements on these are possible through incorporating feature selection and experimenting with hybrid scheme.

Verma et al. [7] applied K-Nearest Neighbors (KNN) for hypothyroidism prediction using the UCI Thyroid Disease Dataset. Their model achieved 88% accuracy, which was slightly lower compared to ensemble models used in other studies. The authors noted that KNN was sensitive to high-dimensional data, as the distance-based approach becomes less effective when irrelevant features are present. Moreover, the performance of KNN

©Daffodil International University

depended heavily on the choice of “k” and the distance metric. While KNN provided a simple and interpretable baseline, the study suggested that dimensionality reduction techniques such as PCA could improve performance. The authors concluded that although KNN is not the most optimal choice for medical data, it could still serve as a quick and interpretable model for initial screening.

Raghav et al. [8] examined the efficiency of Naive Bayes when applied to thyroid disorder prediction using Kaggle Thyroid Disease Dataset. Their model established an 82% accuracy that was relatively poor against ensemble or deep learning. The authors illustrated that in Naive Bayes the features are treated as independent which is not the case most of the times with medical data whose indicators used in its determination are non-independent. Consequently, the model was not effective in portraying the complex relationships that made it less accurate. Nevertheless, the study noted that Naive Bayes is a computationally efficient algorithm and might be used as a rapid base prediction tool. They concluded that Naive Bayes shows the least predictive power, but in low resource healthcare units might be useful due to the limited availability of computational resources.

According to Patel et al. [9], Logistic Regression was utilized on clinical data in hospitals, Mumbai, India, to predict thyroid diseases with an accuracy of 85 percent. The model in question has limitations in its ability to light up non-linear relationships within the data, although it could generate interpretable results. The authors stated that Logistic Regression is effective in binary classification but could not provide patterns of complex data that appear in medical data involving multiple related factors. Although less accurate, the logistic regression was considered of some use, because it was easy to use and understand. The authors considered that it might be viable to use Logistic Regression, feature transformation methods to enhance its results.

In this article, Singh et al. [10] developed deep learning algorithms, including CNNs and RNNs, to predict thyroid disease through the open datasets at Kaggle and through medical history. Their models performed the best accuracy of 95 percent compared to others.

©Daffodil International University

CNNs worked well in identifying feature hierarchies while RNNs demonstrated good performance with sequential medical data. The authors emphasized that large and heterogeneous datasets could be more useful to deep learning models compared to traditional ML models. However, they also noted significant challenges: high computational requirements, risk of overfitting on small datasets, and lack of explainability in clinical practice. They concluded that deep learning holds strong potential in thyroid disorder prediction but emphasized the need for interpretable AI methods to ensure trust among healthcare professionals.

Kumar and Gupta [11] compared Decision Trees, Logistic Regression, and SVM for thyroid disorder prediction using clinical data from Indian hospitals. Surprisingly, Random Forest was also included and achieved the highest accuracy of 90%, while Logistic Regression had the lowest at 85%. The study highlighted that Random Forest was more robust to noise and feature interactions, while Logistic Regression struggled with complex patterns. SVM performed moderately but required fine-tuning of hyperparameters to achieve stability. The authors emphasized that while traditional models provide good interpretability, ensemble methods like Random Forest strike a better balance between accuracy and clinical reliability.

Agarwal et al. [12] used clinical records of thyroid diseases which established a total of Random Forest in the field of diagnosing the diseases, with an accuracy of 94%. The authors explained that this high performance that was observed was because the random forest was able to do feature selection within its model where the most efficient clinical measures were given more weight as compared to the less useful scores. The paper has highlighted that besides providing the high accuracy, Random Forest also enhanced the interpretability of the model because the metrics of TSH and T4 have been ranked as the features that can be highly regarded. They, however, also observed that Random Forest might be unfeasible when working with large datasets due to the possible high computation cost. The rationality behind this conclusion was that random forest was a stable option so far as predicting thyroid diseases at Indian hospitals was concerned.

©Daffodil International University

The article by Kumar and Sharma [13] examined how a systematic feature engineering can be instrumental in the prediction of thyroid disease using the Logistic Regression and Random Forests models. They used the data available in hospitals of New Delhi, India. Random forest gave the highest accuracy at 90 percent, and the logistic regression gave 85 percent. The paper has pointed out that the use of good choice of features and transformation enhanced better performance of the model particularly in eliminating loss of noise involved or redundant features. The poor performance of Logistic Regression had to do with assumptions of linearity and the superior performance of Random Forest linked with engineered features which offered the ability to capture non-linear relationships. The authors highlighted the fact that data preprocessing should be seen of equal importance to model choice particularly in medical contexts where feature quality refers directly to diagnostic accuracy.

Singh and Yadav [14] combined clinical and genetic data and environmental data to predict thyroid ailments with a dataset of various hospitals and research institutes. In their hybrid models they had the 94 percent accuracy implying the advantage of using heterogeneous sources of data. This study has highlighted that statistical predictivity was both increased with the inclusion of genetic markers in conjunction with the use of clinical cues because in most cases thyroid disorders possess a genetic constituent. The authors however, noted the difficulty in integrating data, which included the lack of values, privacy issues, and source standardization. They decided that the hybrid data models that can take advantage of multi-source data would offer a more comprehensive method of predicting thyroid disease particularly in precision medicine.

Sharma et al. [15] used XGBoost and Logistic Regression to categories thyroid diseases based on the results of the datasets at Kaggle and medical hospitals. XGBoost was also 92.5 percent accurate compared to the Logistic Regression which was accurately only 85 percent. The paper particularly noted that XGBoost performed excellently due to its capability of managing non US-linear feature interactions as well as inbalanced dataset that are apparent in medical data. Uninterpretable models such as the Logistic Regression did not pinpoint complex relationships well. The authors concluded that XGBoost is best fit to the real-world clinical setting where data is sloppy and uneven, they emphasized the

©Daffodil International University

ability of accuracy versus explainability balance on healthcare adoption, and XGBoost is best applied to real-world clinical situations.

Verma and Sharma [16] tested SVM and Random Forest for hypothyroidism prediction using clinical hospital data from Delhi. Their study showed that SVM achieved 92% accuracy, slightly higher than Random Forest at 90%. The authors also described that the high-dimensional nature of the data presented to SVM, and its identification of detailed decision boundaries enables it to surpass the performance of Random Forest in some cases. They however realized that SVM can be expensive and vulnerable to hyper parameter adjustments. Random Forest offered a trade lesser amount of accuracy and computation efficiency. The findings of the study were that though SVM could be more effective in high-dimensional environments the Random Forest application is more feasible in the large hospitals.

The study by Kumar and Patel [17] compared XGBoost and SVM as predictors of thyroid disorders. based on clinical data at multic grasers in India. XGBoost achieved 94.5% accuracy, Doing better than SVM of 91%. The authors observed that XGBoost was specialized at dealing with non-linear relationships and missing data, thereby rendering it especially useful. applicable to actual medical charts. Contrary to its power, SVM was demanding in respect of being tuned. and performed worse on noisy clinical data than XGBoost does. The study concluded that the XGBoost is a tool that should be prioritized in medical diagnostics systems because of its scalability. and strength but could be enhanced by the incorporation of explicable procedures, which would yield greater benefit to clinical. trust.

Random Forest, Logistic Regression, and SVM [18] are used by Gupta et al. to predict thyroid disorder using clinical hospital records in New Delhi, India. They found that the random forest achieved highest accuracy of 92% and the logistic regression was second with 85% and SVM returned 88%. The authors reasoned that Random Forest was far superior to Logistic Regression a linear assumption system which had a linear limitation on performance. SVM was moderate with extreme parameter adjustment.

©Daffodil International University

Ready et al. [19] compared XGBoost and Random Forest in predicting thyroid disorder using clinical records at various hospitals in India. XGBoost gained the accuracy of 95% and the random forest reached slightly less 93%. The authors emphasized that XGBoost was also capable of handling large datasets and minimizing overfitting than the Random Forest. The efficiency of XGBoost to calculations was also addressed in the study, and the model was attractive to large hospital systems where a great volume of patient data can be gathered. Nevertheless, they reported that these two models continued to experience explainability difficulties. They determined that XGBoost was the most effective prediction of thyroid disorder in large-scale expanses particularly in clinical settings dominated with data.

The Sharma et al. [20] applied SVM and Random Forest on thyroid disease. Their findings showed that With an 91.5% accuracy, SVM was superior to the 88.3% accuracy of random forest. As the authors clarified, SVM performed well especially in high-dimensional cases. medical characteristics and is effective with this dataset. Random Forest, while robust, was not very successful because of the class imbalance and superimposed feature distribution. The study emphasized that although SVM was more accurate, it was also associated with high computational cost. costs. The writers ended their paper by concluding that both the models could be applied complementary. based on data size and computing data in healthcare settings.

©Daffodil International University

METHODOLOGY

3.1 System Diagram

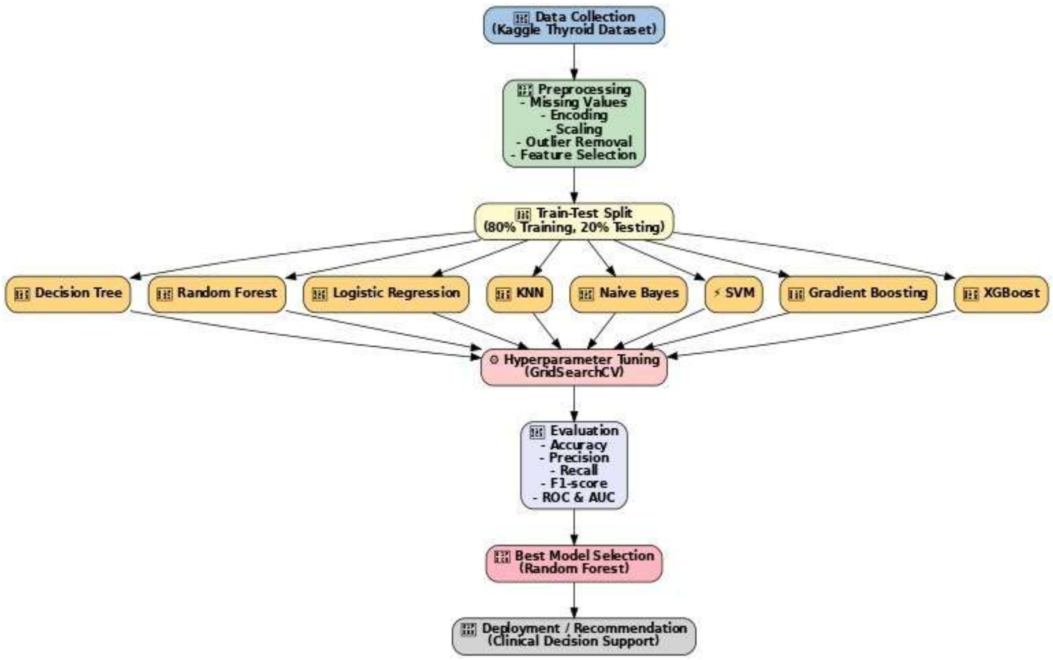


Figure 3. 1System diagram of the project

The 3.1.1 figure identifies how a machine learning model is developed and tested. It involves data collection; this is followed by pre-processing of the data, the purpose of which is to clean and format the data. The dataset is divided into two, namely, data to be used during training and data to be used during testing (80 percent and 20 percent correspond to these two categories, respectively). The training data is later subject to a machine learning algorithm in order to generate a model. After constructing the model, the model is used with the testing data to test its performance or not. The assessment results in the decision point, in case the model is deemed to be the best, the process proceeds to complete the model. In case the model is not the optimal, it goes back to the process of refinement or developing another model. After identification of the best model and its finalization, the process ends. The given flowchart is one of the standard methodological patterns of machine learning projects and it involves creating a model and evaluating and iterating a model to determine which one works best.

©Daffodil International University

3.2 Dataset

	age	TSH	T3	TT4	T4U	FTI	target	patient_id	sex_M	on_thyroxine_t	...	T3_measured_t	TT4_measured_t	T4U_measured_t	FTI_meas
0	29	0.300000	1.970629	108.700305	0.976056	113.640746	0	840801013	False	False	...	False	False	False	
1	29	1.600000	1.900000	128.000000	0.976056	113.640746	0	840801014	False	False	...	True	True	False	
2	41	5.218403	1.970629	108.700305	0.976056	113.640746	0	840801042	False	False	...	False	False	False	
3	36	5.218403	1.970629	108.700305	0.976056	113.640746	0	840803046	False	False	...	False	False	False	
4	32	5.218403	1.970629	108.700305	0.976056	113.640746	31	840803047	False	False	...	False	False	False	

Figure 3.2. 1 Sample of dataset

Figure 3.2.1 provides the demographics of patients and gladiators with thyroid health factors (age, TSH, T3, TT4, T4U and FTI) in the dataset. It contains also flags supporting whether these tests were done, treatment status (onthyroxinet), sex (sexM) and a unique patented. The presence or absence of thyroid disorder is probably referred to in the target column. This kind of data can be used in machine learning and statistical analysis to anticipate thyroid diseases and investigate the relevant patterns.

	age	TSH	T3	TT4	T4U	FTI	target	patient_id
count	9172.000000	9172.000000	9172.000000	9172.000000	9172.000000	9172.000000	9172.000000	9.172000e+03
mean	73.555822	5.218403	1.970629	108.700305	0.976056	113.640746	4.245966	8.529473e+08
std	1183.976718	23.047102	0.751073	36.607295	0.191319	39.693254	8.197174	7.581969e+06
min	1.000000	0.005000	0.050000	2.000000	0.170000	1.400000	0.000000	8.408010e+08
25%	37.000000	0.590000	1.700000	88.000000	0.870000	95.000000	0.000000	8.504090e+08
50%	55.000000	1.600000	1.970629	106.000000	0.976056	112.000000	0.000000	8.510040e+08
75%	68.000000	3.700000	2.200000	124.000000	1.050000	126.000000	1.000000	8.607110e+08
max	65526.000000	530.000000	18.000000	600.000000	2.330000	881.000000	31.000000	8.701190e+08

Figure 3.2. 2 FA concise summary of the employed thyroid dataset

The summary statistics provided in the figure 3.2.2 show a dataset, where there are distinct columns that consist of thyroid health datasets and details about patients. Each column contains 9,172 records in each dataset. The standard deviation of the age column is very large (1183.98), indicating that there are extreme values of the ages, namely, minimum age= 1 year and the maximum age= 65526 years, and this implies

©Daffodil International University

data anomalies. On the case of TSH (Thyroid Stimulating Hormone), average is 5.22, SD value is 23.05, and maximum value is 530 and minimum is 0.005, which suggests that there are some extreme values.Other thyroid-related measurements, such as T3, TT4, T4U, and FTI, exhibit more reasonable ranges, with the mean of T3 being 1.97 and a standard deviation of 0.75. The target column, likely indicating the presence or absence of a thyroid disorder, has a mean of 4.25, with a maximum value of 31, suggesting a classification task with varying outcomes. The patient_id column, which uniquely identifies each patient, also shows a wide range of values. Overall, the summary highlights significant variability in the data, with some extreme values and potential outliers that may need further examination or cleaning.

©Daffodil International University

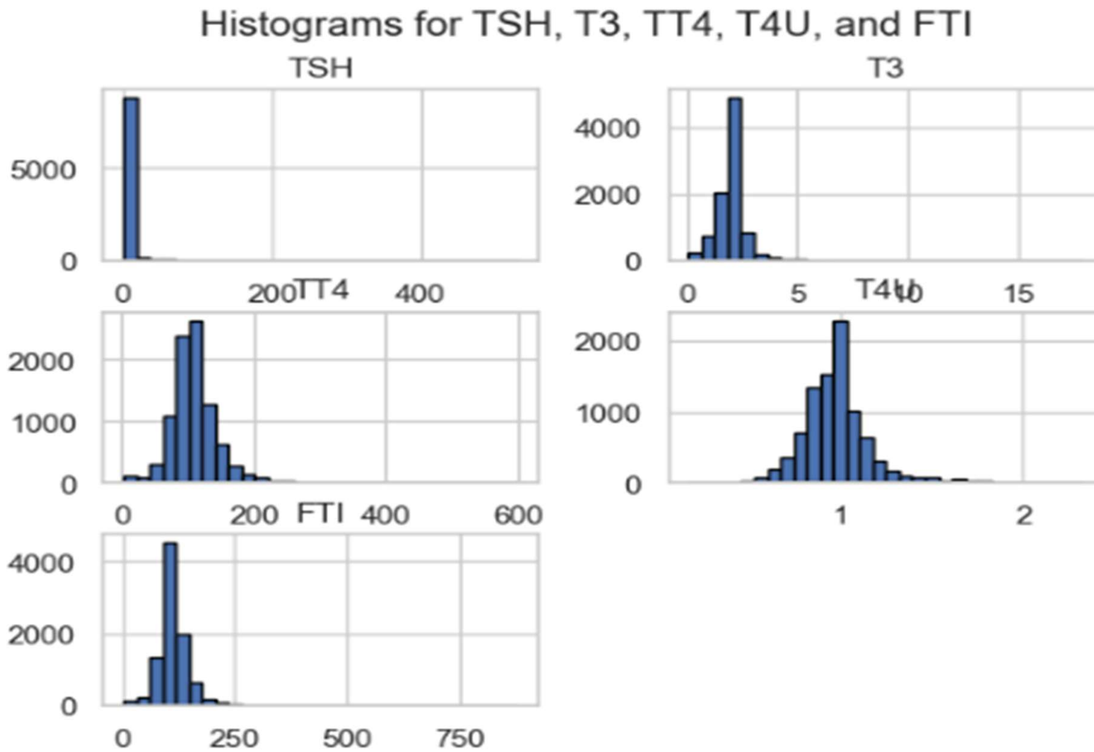


Figure 3.2.3: Histograms for TSH, T3, TT4, T4U, and FTI

The figure 3.2.3 presents histograms for the key variables—TSH, T3, TT4, T4U, and FTI—used in the machine learning model for predicting hypothyroidism and hyperthyroidism. Both histograms present the counts of these thyroid related. features. TSH is greatly skewed with majority of the values being found at low levels, revealing majority of normal levels of TSH. T3 also is skewed in distribution. and a zenith in the middle rank. TT4 is less extreme, and it is rather normal. distribution T4U is much gentler in its shape with a peak at lower values, which the use of outliers may at. the higher end. The distribution of FTI is similar to that of TT4 and the majority falls centrally. and an inclination toward greater values. These histograms can be used to understand the feature. distributions, which assist in choosing some model and preprocess before handing merchandise to the machine. learning analysis.

©Daffodil International University

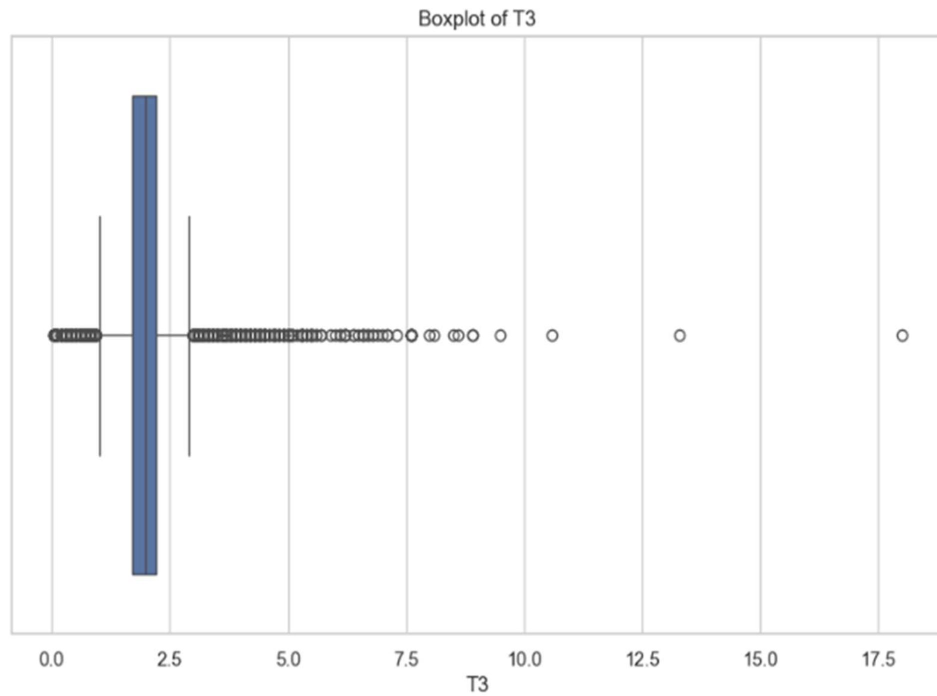


Figure 3.2. 4 : Boxplot for T3

TheT3variable figure 3.2.4 represents how T3 levels were spread in the data set that was used to predict the occurrence of hypothyroidism and hyperthyroidism. In the plot, there is a central box which denotes the interquartile range (IQR) and the median figure is indicated. The whiskers are then stretched to indicate the range of values in 1.5 times of the IQR and values out of the same values are regarded as outliers, which are marked with circles. As it can be seen in the plot, most of the points are concentrated in the lower part with few that are spread to higher values being outliers. This indicates that most of the T3 concentrations are also on the low sides of the scale with some extreme values that are likely to cause performance problems to the machine learning model unless managed adequately.

©Daffodil International University

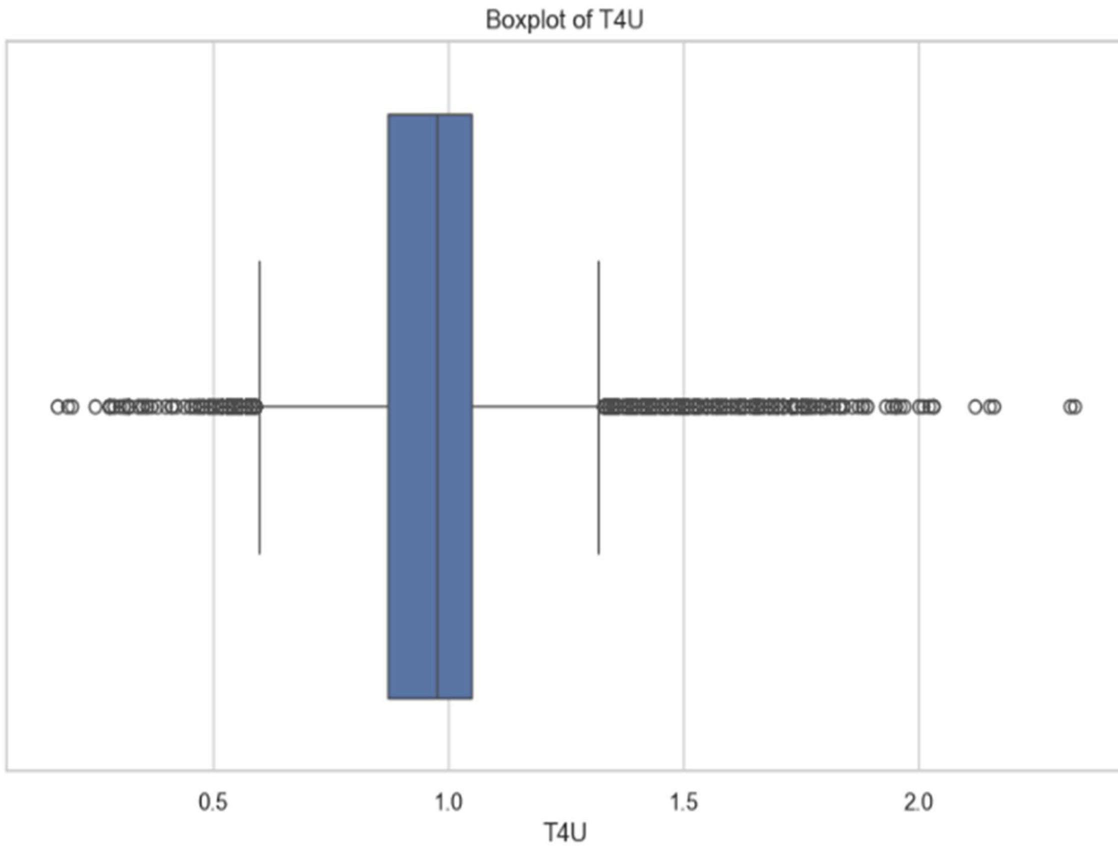


Figure 3.2. 5: Boxplot for T4U

The distribution of T4U levels in the dataset is provided in the figure -3.2.5 of the T4U variable. The box at the center that tells about the range of values (interquartile range IQR) shows that most of the values are between 0.5 and 1.5. The box indicates the central tendency of data with the median being indicated within the box. The whiskers increase to the values of 1.5 times IQR besides the points outside the extremes are shut as outliers as are indicated by circles. The plot depicts a few outliers on both sides particularly as the higher value and that can be attributed to the occurrence of captivated extreme values in the data. Most of the data points appear in the range of 1.0, with the existence of outliers that might impact the analysis and the performance of the model, hence may need further investigation or management as part of the preprocessing.

©Daffodil International University

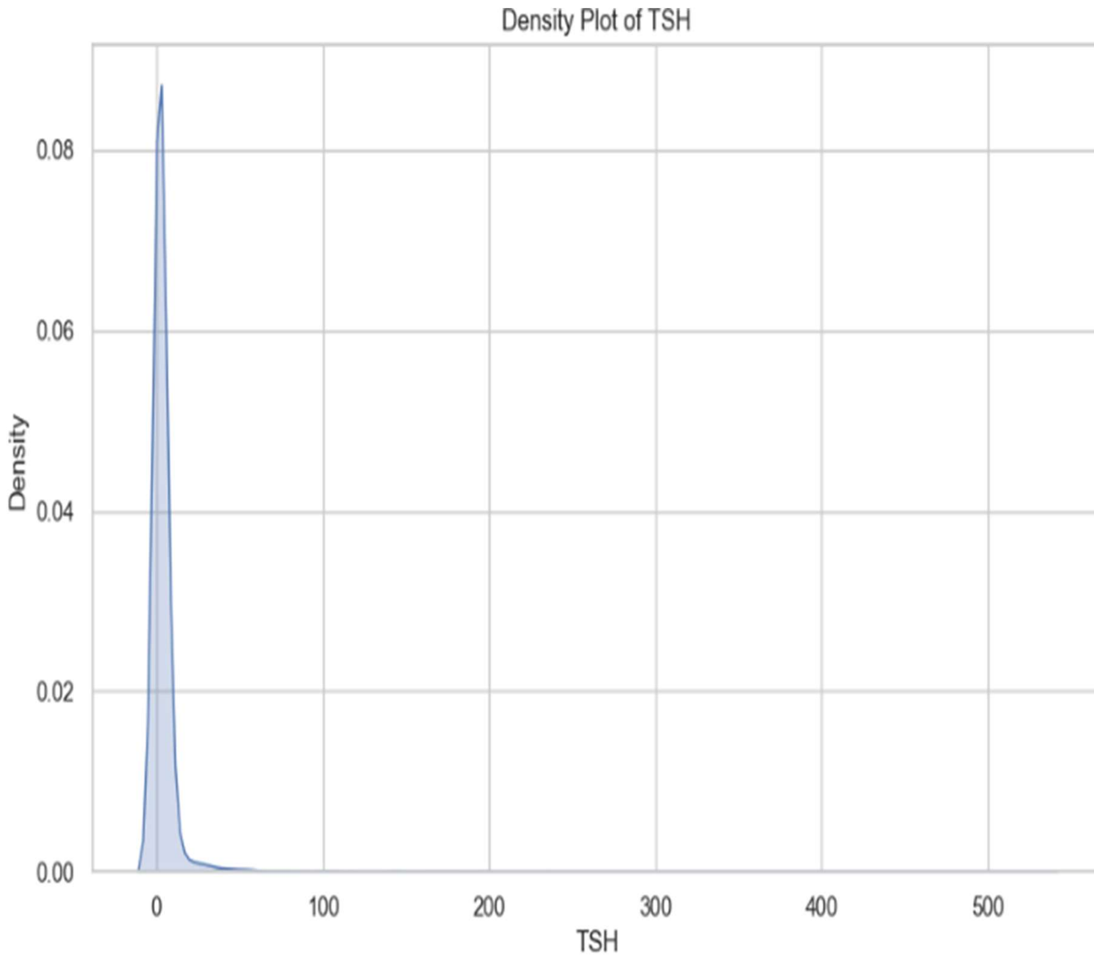


Figure 3.2. 6 : Density plot for TSH

A TSH (Thyroid Stimulating Hormone) variable that is represented as Figure 3.2.6 displays extremely skewed distribution. The sharpness of the peaks in the plot at 0 is valuable to report that, most of the data values are high concentrated around very low values of TSH. The concentrations are rapid decreasing with an increase in values, only a few data points are projected toward the lower values. This will imply that one has many samples with low TSH and very few with increasing values of TSH. The sharp marked peak and increase fall of the plot point out the data show a high level of skew, thus they might have to be dealt with in preprocessing a machine learning model, possibly by normalization or transformation.

©Daffodil International University

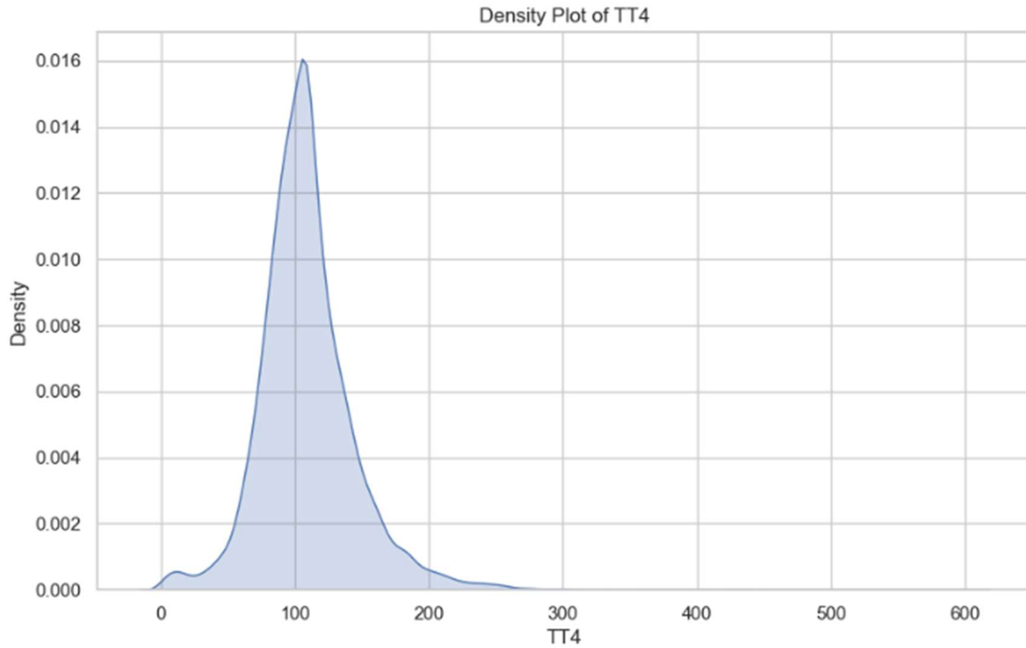


Figure 3.2. 7: Density plot for TT4

The Figure 3.2.7 of the TT4 (Total Thyroxine) value represents a right skewed distribution as the distribution is concentrated at 100. This implies that most of the data points are concentrated about this value and as the TT4 values go up, the density becomes more sizable. The distribution becomes smaller as the values go to higher regions, and few values are relatively above 200. That form of a plot implies that there are higher numbers of observations that lay near the lower to mid-range values with fewer observations that are in the higher range. This skew can be useful when processing data to machine learning models by normalizing the distribution possibly by transformation.

©Daffodil International University

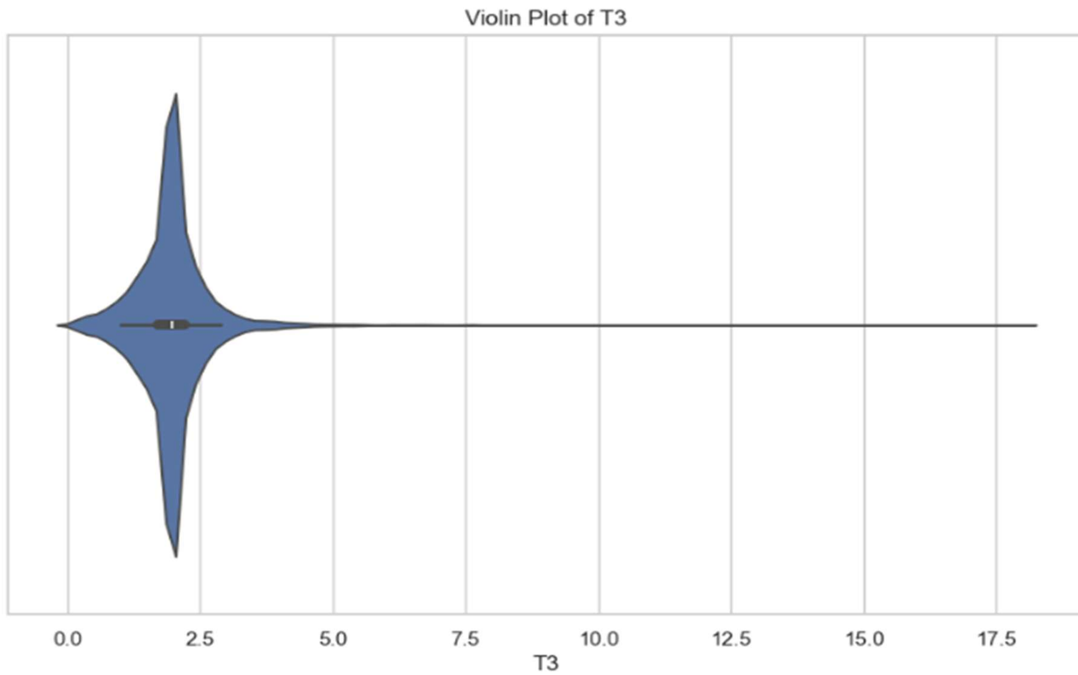


Figure 3.2. 8: Violin plot for T3

In the Figure 3.2.8 of T3 variable, the fact that at the left end, much of the data concentration at the left frequency is very skewed indicates that most of the data values are concentrated in the lower range, especially in the range 0 to 2.5. The shape of the plot suggests that the data features within this band with the broader part of the violin depicting the dispersion of the values. The straight line within the violin will show the median, which in this case, lies towards the bottom hence indicating the fact that most of the data is close to the lower T3 measurements. The elongated tail at the far right would indicate that there are some few high outlier values which may affect the performance of machine learning models to relate to it unless particular measures are taken. Plot shows some visual interpretation of skewness and variability of the data, which can be used to make preprocessing choices.

©Daffodil International University

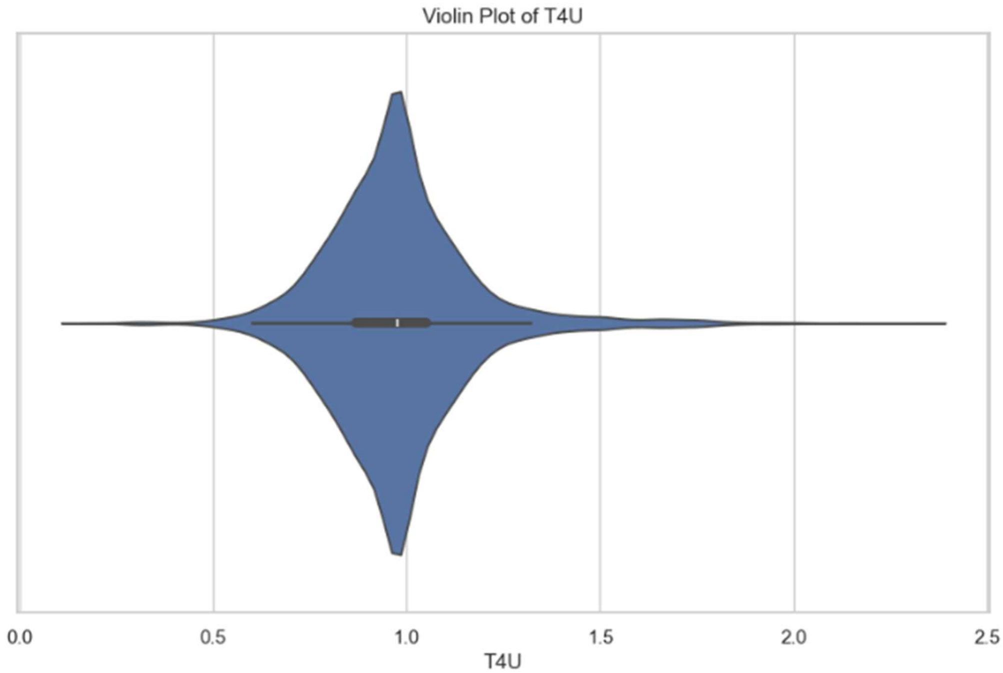


Figure 3.2. 9: Violin plot for T4U

Figure 3.2.9 of T4U variable represents a symmetrical distribution and the data tends to be concentrated in the middle. There is a centre area in the plot that is widest which means that majority of the values fall around the 1.0 point. The median is represented as the horizontal line within the "violin" and it is pretty much centrally placed. The shape is that we have a balanced distribution and there are not many extreme values in the data since we can see that it does not have any indicative outlier or extreme in either end. It is this distribution to mean that the values of T4U are reasonably uniform, and that most of the responses take values near the median, which may be indicative of a reducing degree of variability in this attribute compared to others. Based on this knowledge, an exceptional analysis or model training decision can be made.

©Daffodil International University

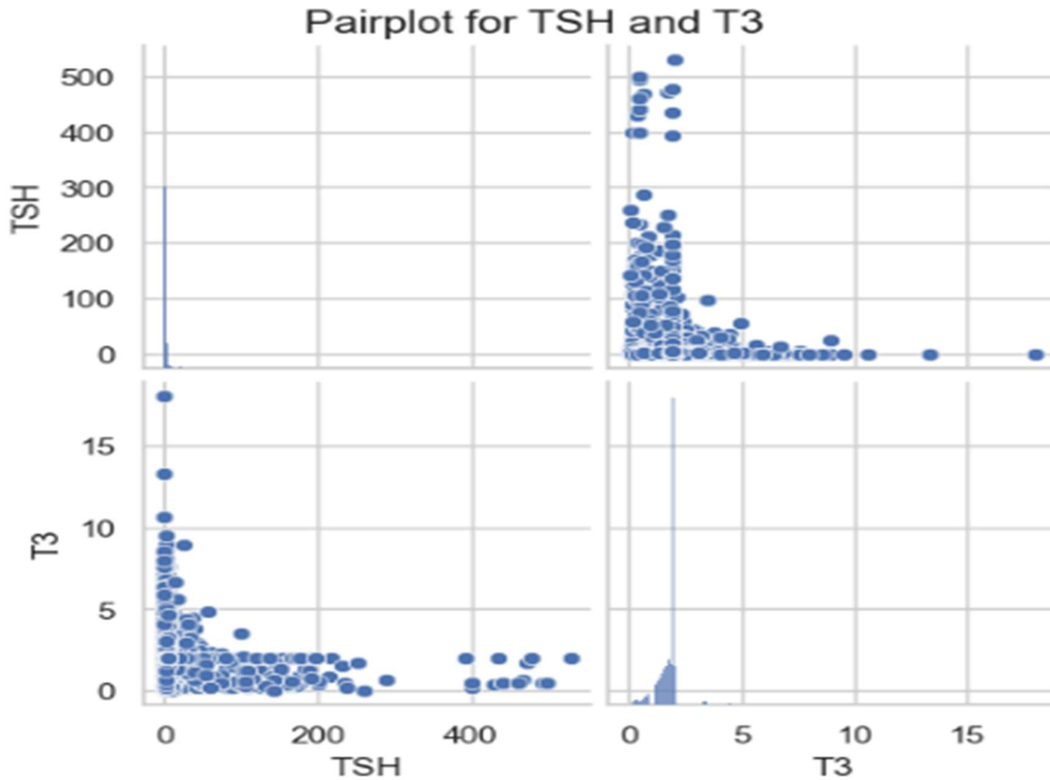


Figure 3.2. 10: Pairplot plot for T3,TSH

Figure 3.2.10 of TSH (Thyroid Stimulating Hormone) and T3 (Triiodothyronine) explain the allocation and correlation of the two aspects. There is the upwardly skewed prevalence of both variables where the curve TSH is more concentrated or cantered towards zero and the T3 is skewed mitochondrial around zero with a broader spread. The trends on the scatterplot indicate that the scatterplots of both TSH and T3 exhibit most of the data concentrated quickly on the smaller values with of some outliers on the larger values. It implies that TSH and T3 do not correlate clearly with each other because there is no specific linear or non-linear tendency in the correlation. The storytelling narrates on the necessity of study or changes to such attributes prior to utilizing them in any machine learning mechanisms.

©Daffodil International University

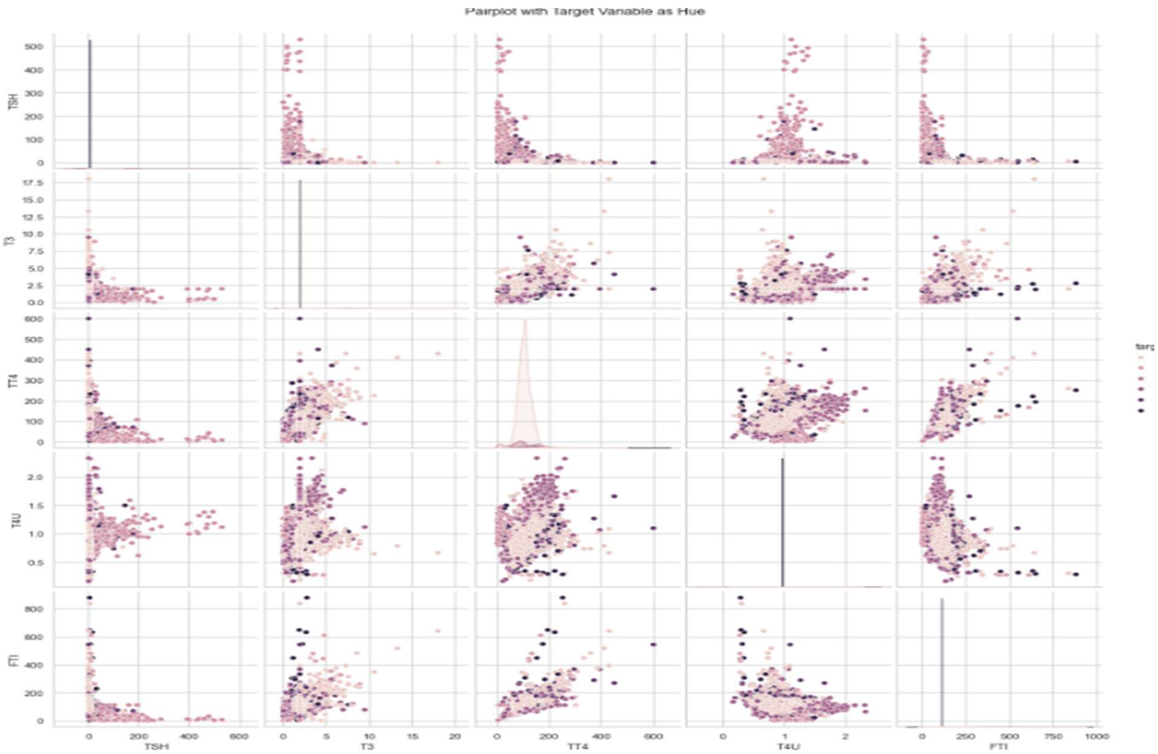


Figure 3.2. 11: Pairplot plot for T3,TT4,T4U,FTI,TSH

The above figure 3.2.11 illustrates that there existed relationships with the different thyroid-related aspects (TSH, T3, TT4, T4U, and FTI) and the target variable, which is the hue. The diagonal plots can give the personal distributions of each of the features whereas the scatter plots can be used to display the findings between two features. The colour gradient substitutes the target variable, whereby the light extends are used to show one variable, and darker to show the other. The appearance of most scatterplots depicts a disconnected and non-linear connection of the variables, whereby there are some groups of darker spots, which show the distribution of the target variable. Such patterns may give information of possible correlations and the relationship of the features with the target variable, which is essential to understand feature influence with the machine learning models. Because there are skewed distributions and concentration of the clusters in some areas, there is a probability that preprocessing needs handling of the outliers to enhance the performance of the model, hence the careful consideration of such processing is warranted.

©Daffodil International University

3.3 Dataset Preprocessing

This research project to create a predictive machine learning model, one that would predict hypothyroidism as well as hyperthyroidism, but based on the strength of a sophisticated classification software. Our project started by collecting the data, in which we used the publicly accessible dataset of clinical and biochemical data of persons diagnosed with thyroid disorders. The data would contain such important variables like TSH (Thyroid Stimulating Hormone) levels, T3, T4, age, Gender, and other medical variables that have been known to be a correlation with thyroid disease. This information formed the basis of model training and testing.

Preprocessing refers to the data-contents that are taken prior to commencing the machine learning in the raw data. In the beginning, the missing data were managed through imputation methods. The numerical attributes that had missing data were replaced by mean of the respective columns and the categorical ones by mode (frequency of occurrence). This made sure that no data was discarded thus avoiding biases and guaranteeing a complete dataset. The second activity in preprocessing was muting the categorical variables. Dependent variables like sex, on thyroxine, pregnant among others were one-hot coded, meaning that they were changed to binary variables to be understood by machine learning programs. Besides that, feature scaling was done to normalise numerical features with the help of min-max scaling. This is of special concern to such models as SVM and KNN, which are also sensitive to the size of the input data. Lastly, feature selection techniques were used (Random Forest feature importance and Select Best), and the best features to predict thyroid disorders were selected. Such characteristics as TSH, age, and T3 were identified as the most significant, whereas the irrelevant ones were eliminated, which minimizes the complexity of the model and maximizes its performance.

The mean imputation formula for replacing missing values with the mean of the available data for a particular feature can be expressed as:

©Daffodil International University

$$x_{\text{imputed}} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Where:

X_{imputed} is the imputed value for the missing data point.

x_i are the observed values of the feature.

n is the total number of observations (including missing values).

In words, to impute a missing value, you take the sum of all observed values for that feature and divide it by the total number of non-missing values. This gives you the mean value, which is then used to replace the missing value.

However, in practice, you might only consider the non-missing values when calculating the mean, so the formula could be simplified to:

$$x_{\text{imputed}} = \frac{\sum_{i=1}^{m} x_i}{m}$$

m is the number of observed (non-missing) values for the feature.

To select a model some powerful machine learning algorithms were tried as Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes (NB) and XGBoost. Every model was selected because it is applicable to classification tasks especially when processing complex and high-dimensional healthcare data. They were performed with the use of Decision Trees due to the fact that the process was interpretable

©Daffodil International University

and could work with both numerical and categorical data, whereas a method extension supplied with a strong ability to sustain itself by avoiding overfitting was selected: Random Forests. SVM was used because of its performance in high dimensional space and Naive Bayes was investigated as a probabilistic model because of its simplicity and performance in small data segments. XGBoost, a gradient boosting model, was also the one that was tested because it is better at structured classification problems.

Regarding model training, the data was divided into test and training subsets by 80/20 partition such that the models would be tested on data that they have not seen. Some models used in hyperparameter tuning include Decision Trees, XGBoost, to name a few, where the best model configurations were found using the method known as grid search cv. The step is very important because it enables the model to acquire the best parameters and therefore provide the best result. To take Categorical examples, hyperparameters such as max depth, min samples split, and min samples leaf of Decision Trees have been adjusted to avoid overfitting and n estimators and learning rate were adjusted in XGBoost.

After training the models, they were tested based on a number of evaluation metrics that would measure the performance of the model. Accuracy served as a general measure of the number of instances in which the decision reached was correct. Nevertheless, accuracy and recall were also computed, which are also necessary when assessing the performance of the model whenever there are imbalances in the class distributions. The F1-score (harmonic means of precision and recall) gave a more holistic look of balance between recall and precision using the models. Wor, as a visualization of the models and also as further assessment thereof, confusion matrices were plotted, enabling us to determine the quality of how the model differentiated between the possible hypothyroid cases, hyperthyroid cases and normal cases. Also, receiver operating characteristic (ROC) curves and the area under curve (AUC) were computed that determine the discriminatory ability of the models on the overall classes and especially in regard to the trade-off between the false positive and true positive rates. The multiple visuals and measures were useful in the identification of the best and true-to-form predictor of thyroid disorders.

©Daffodil International University

All in all, this approach to the analysis will provide a rigorous approach to predicting thyroid disorders through machine learning. Including extensive data pre-processing, feature sample, modelling, and strong assessment strategies, the proposed study will equip high performance model when predicting hypothyroidism and hyperthyroidism. The steps are very essential to guarantee that the end models are not just precise, but also universal, recognizable and applicable in real life clinical settings.

©Daffodil International University

RESULTS AND DISCUSSION

4.1Model

4.1.1 Decision Tree

Without Hyperparameter

Accuracy	91.06
Precession	60.62
Recall	56.86
F1-Score	57.46

Table 4.1.1. 1:Evaluation Matrix of Decision Tree Without Hyperparameter

Table 4.1.1.1 illustrates the performance of the Decision Tree model without hyperparameter tuning. The model achieves a high accuracy of 91.06%, indicating that it correctly classified most instances. However, accuracy alone does not capture the model's ability to handle both classes, particularly in imbalanced datasets. The precision is 60.62%, meaning that when the model predicted a positive class, it was correct 60.62% of the time, which is important when false positives are costly. The recall is 56.86%, showing that the model only identified about 57% of the actual positive instances, which could be problematic in scenarios were failing to detect positives is critical. The F1-Score of 57.46% balances precision and recall, providing an overall view of the model's performance in predicting positive cases while minimizing errors.

©Daffodil International University

With Hyperparameter

Accuracy	90.62
Precession	59.09
Recall	54.35
F1-Score	55.29

Table 4.1.1. 2:Evaluation Matrix of Decision Tree With Hyperparameter

The Table 4.1.1.2 with hyperparameter tuning shows the following performance metrics: accuracy of 90.62%, which indicates that the model correctly classified most instances, although it remains slightly lower than the previous model without hyperparameter tuning. The precision is 59.09%, meaning the model was correct 59.09% of the time when predicting a positive class. While this is an improvement, it still suggests that false positives are a concern. The recall is 54.35%, showing that the model correctly identified only 54.35% of the actual positive instances, which still leaves room for improvement in detecting rare positive cases. The F1-Score is 55.29%, reflecting a balance between precision and recall. While hyperparameter tuning did not result in a significant increase in performance, it did slightly improve the precision and recall when compared to the model without hyperparameter tuning. However, there is still room for further optimization to achieve a better balance between precision and recall, which would improve the overall model effectiveness.

©Daffodil International University

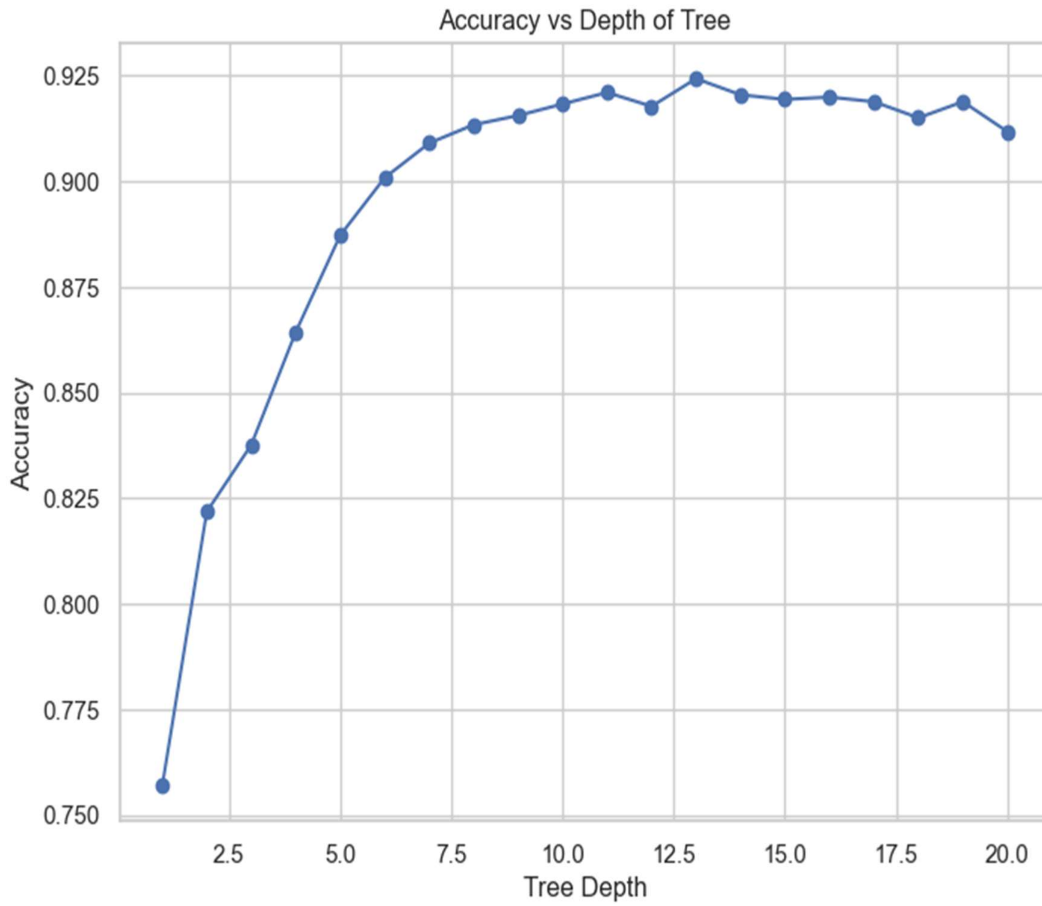


Figure 4.1.1.1: Accuracy vs Depth Tree for Decision Tree.

The Figure 4.1.1.1 illustrates the relationship between the depth of a decision tree and its accuracy. Initially, as the tree depth increases from 2 to 6, there is a noticeable improvement in accuracy, with the model achieving better performance as it captures more intricate patterns in the data. However, after reaching around depth 6, the accuracy begins to plateau, stabilizing between 0.90 and 0.925. This indicates that further increases in tree depth, up to 20, do not result in substantial improvements in the model's performance.

©Daffodil International University

4.1.2 Random Forest

Without Hyperparameter

Accuracy	92.31
Precession	60.80
Recall	58.95
F1-Score	58.82

Table 4.1.2. 1:Evaluation Matrix of Random Forest Without Hyperparameter

The Random Forest model shows an accuracy of 92.31%, indicating that it correctly predicts a high proportion of instances. Precision is 60.80%, meaning that 60.80% of the predicted positive cases are correct. The recall is 58.95%, suggesting that the model identifies a fair portion of the actual positive cases but still misses some. The F1-score 58.82 allows having a balance between precision and recall and indicates that the overall scan successfulness of the model to identify the positive cases is mediocre. Although the accuracy is great, it still could be improved because of the precision and recall balance.

With Hyperparameter

Accuracy	92.47
Precession	63.47
Recall	61.43
F1-Score	61.35

Table 4.1.2. 2: Evaluation Matrix Score of Random Forest With Hyperparameter

The Random Forest model has a slight improvement in terms of performance with hyperparameter tuning. This renders the life accuracy to be slightly higher (92.47) and hence the overall performance and output of the model remain good as well. The precision has increased to 63.47 indicating that the proportion of the positive cases that were picked up is higher than in the old model. Recall also inbuilt to 61.43 began to indicate that the model is currently recalling a higher number of the real positive cases. The F1-score increases to 61.35 and the precision and the recall are more balanced. The overall reaction to hyperparameter tuning to the model is increased performance with an increase in the ability to detect positive reports and with the accuracy of an image.

©Daffodil International University

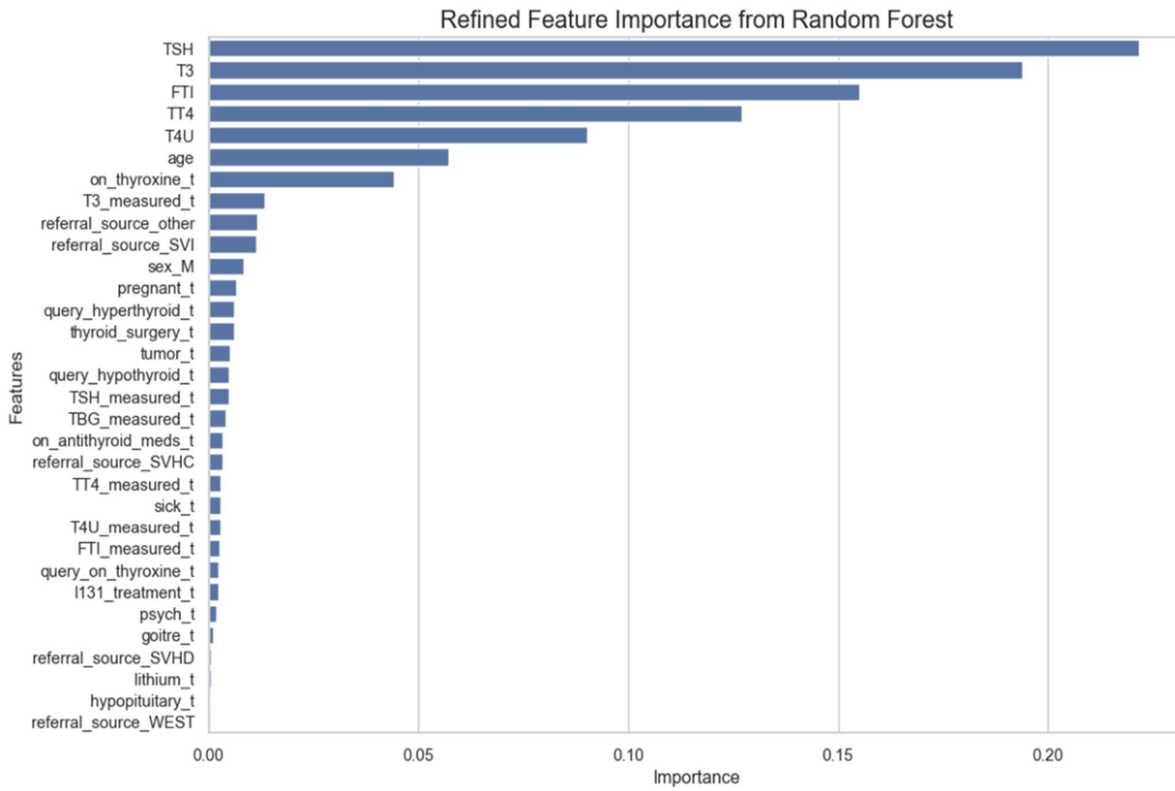


Figure 4.1.2. 1 Feature Importance from Random Forest

The graph depicts the filtered feature importance made of a Random Forest model. It demonstrates the input of individual features in predicting the target variable, probably mutated to Thyroid disorders. The TSH, T3, FTI, T4 and T4U are the most significant features that bear the largest importance scores in the model. These characteristics are among the crucial variables used in prediction of the model. The next important feature is age that stands second only to these best features. The other variables like thyroxine, T3measuredt and referral source SVI are ranked beneath these meaning that they have a significant contribution to the decision making. Unlike this, aspects such as lithium, hypopituitary, and referral source WEST are not so significant, and their influence on the prediction of the model is limited. This visualization also provides great learning information about which aspects are used in the model to make the best predictions of thyroid conditions.

©Daffodil International University

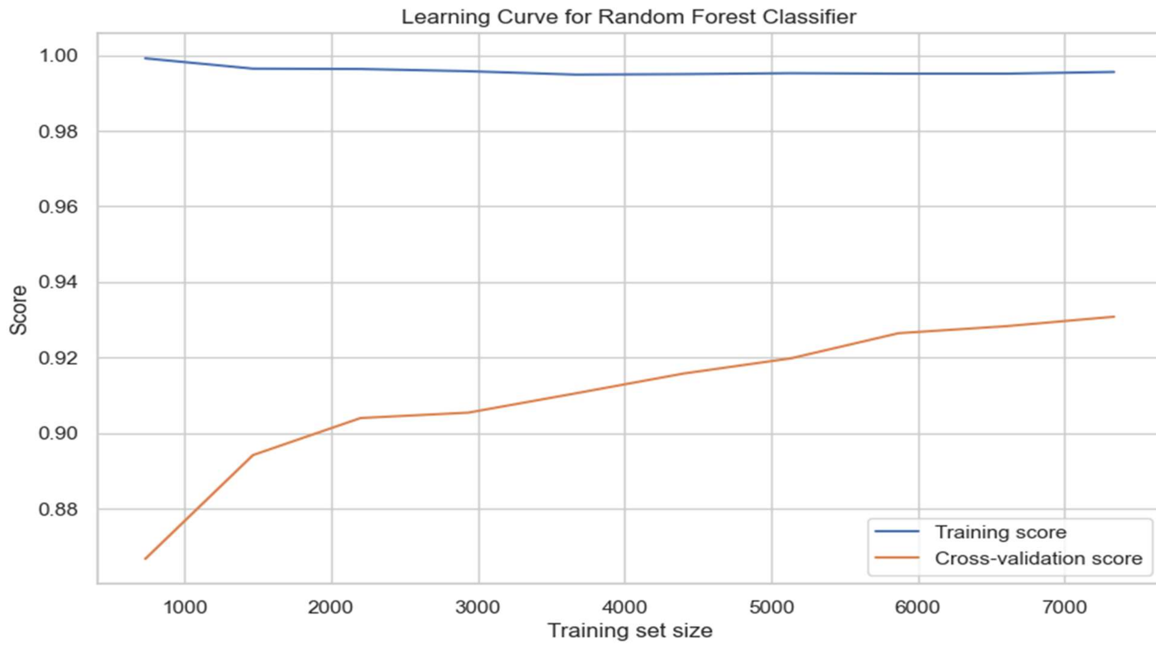


Figure 4.1.2. 2 : Learning Curve for Random Forest Classifier

The plot shows the learning curve with a Random Forest classifier on the performance of the model regarding the training set scale. The x-axis will not be the magnitude of the training set and the y-axis will be the score which would probably be something quantifying accuracy or a performance measure. The blue line is depicting the training score, which is very stable with the high value of 1, which means that the training data adversarial to the model work well with different sizes of training set. Conversely, the orange line shows the cross-validation score which begins lower and appreciates steadily with increase in the size of the training set. This implies that the performance of the model becomes better with increasing amount of data and is also more generalized, meaning that the model can better handle the unknown data. The discrepancy between the training and cross-validation scores, particularly when training small sets are used, does indicate that the model is overfitting but with the increase in the amount of data, the performance levels off, and demonstrates superior levels of generalization. There is a tendency that the larger the training sets, the more the model of the Random Forest classifier would perform better on the unseen data.

©Daffodil International University

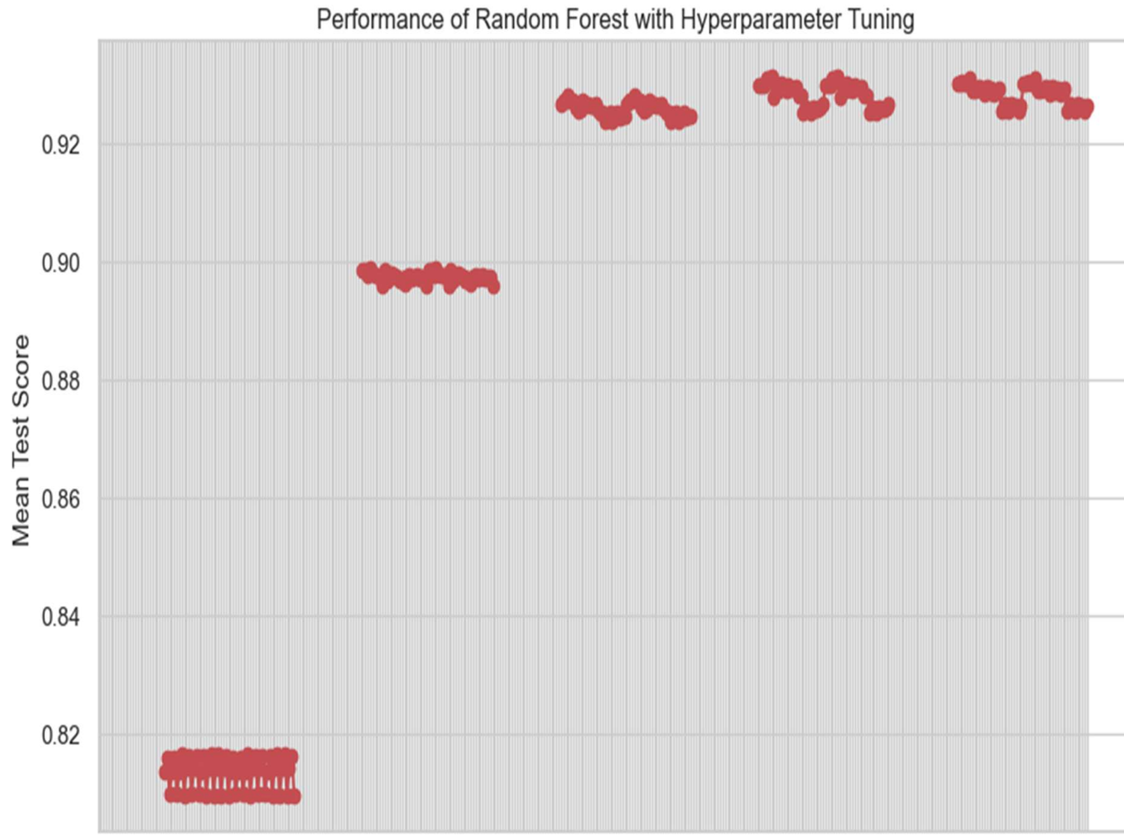


Figure 4.1.2. 3 : Performance vs Mean Test Score for Random Forest with Hyperparameter Tuning

The graph depicts the performance of a Random Forest model with hyperparameter tuning. The x-axis represents different configurations or iterations of hyperparameter tuning, while the y-axis shows the mean test score, likely reflecting the model's accuracy or performance on the validation set. The chart shows several data points clustered around 0.90 to 0.92, indicating that hyperparameter tuning has led to a consistent improvement in performance, with most test scores falling within this range. However, there is also a lower cluster around 0.82, suggesting that some configurations result in poorer performance. This scatter plot highlights the variability in model performance with different hyperparameter settings, but overall, it shows that fine-tuning the hyperparameters leads to improved accuracy and model performance.

©Daffodil International University

4.1.3 k-nearest neighbors (KNN)

Without Hyperparameter

Accuracy	78.74
Precision	34.97
Recall	24.57
F1-Score	26.94

Table 4.1.3. 1: Evaluation Matrix of k-nearest neighbour (KNN) Without Hyperparameter

The K-Nearest Neighbors (KNN) model, without hyperparameter tuning, shows an accuracy of 78.74%, indicating that it correctly classifies most instances. However, the precision is relatively low at 34.97%, meaning that less than a third of the predicted positive cases are correct. The recall is even lower at 24.57%, suggesting that the model misses a significant number of true positive cases, failing to identify many positive instances. F1-score of 26.94 is an indicator of trade-off between precision and recall, as despite the model having low capacity to generate results that include positive cases, it has a low accuracy to do so. The reasonable accuracy, however, together with the low level of precision, recall, and F1-score indicate that the model needs further improvements to achieve a higher level of performance, especially in recognizing instances of high odds.

With Hyperparameter

Accuracy	79.61
Precision	33.93
Recall	27.11
F1-Score	28.72

Table 4.1.3. 2: Evaluation Matrix of k-nearest neighbor (KNN) With Hyperparameter

Using hyperparameter tuning, K-Nearest Neighbor (KNN) model demonstrates the accuracy of 79.61 percent, which equals the performance in the absence of hyperparameter tuning, which means that the evaluated model accurately assigns the correct labels to an essential fraction of examples. The accuracy is low of 33.93 which implies that the defense of the predicted positive cases is only 1 in every 3 successful cases only. It is a recall of 27.11 which indicates that the model continues to miss many of the true positive cases in that there are many positive instances that are not identified by it. The F1 score is 28.72 which means does not achieve a good balance between

©Daffodil International University

precision and recall. Although this accuracy is similar, the poor precision, recall, and F1-score indicate that more improvement and refinement can be done to the model, especially detecting positive cases more accurately.

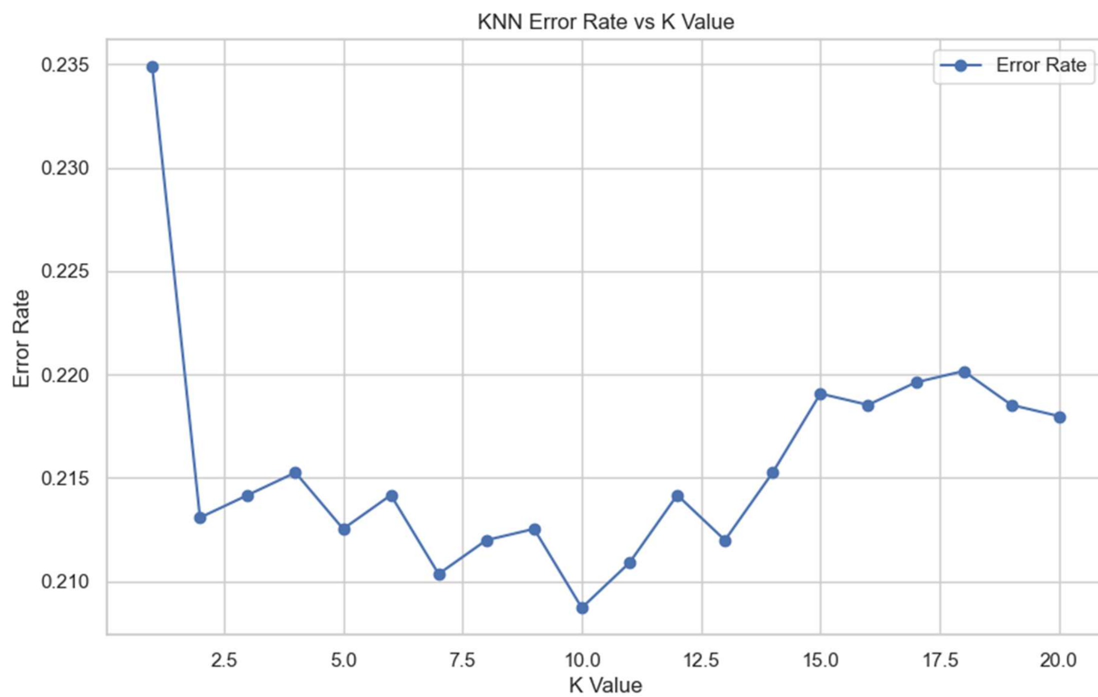


Figure 4.1.3. 1: Error Rate vs K Value for KNN

The figure depicts the dependence between the error rate of the K-Nearest neighbours (KNN) and K value that defines the number of neighbours used when performing the classification. The K value, which takes the value between 2 and 20, is given as the x-axis, whereas the error rate is given as the y-axis. The error at a value of K is very large start at K value of 2 with error rate of approximate 0.235 and as the K value increases, it reduces at a rapid rate till the K value of about 10 that the error rate is of about 0.210. Some fluctuation in the error level has taken place after this point with the values of the error level increasing as the K resembles close to 20 on the one hand. This indicates that the best K value (one of about 10) causes lowest error rate, and increasing the K values causes a slight increase in error rate, which could be an indication that increased K values lead to overgeneralization of the model.

©Daffodil International University

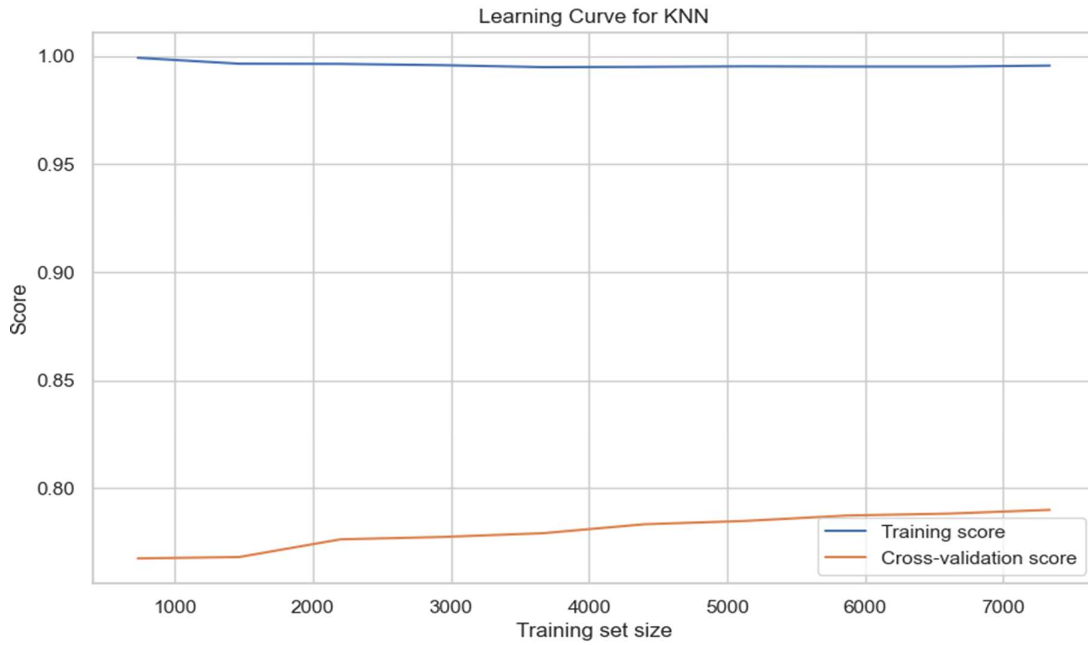


Figure 4.1.3. 2: Learning Curve for KNN.

The graph shows the learning curve for a K-Nearest Neighbors (KNN) model, illustrating how the model's performance changes as the training set size increases. The x-axis represents the training set size, while the y-axis shows the score, which likely reflects accuracy or model performance. The blue line represents the training score, which remains nearly constant at a high value of 1, indicating that the model performs well on the training data as the size increases. On the other hand, the orange line represents the cross-validation score, which starts lower and gradually increases as the training set size grows. This suggests that the model performs better on unseen data as it sees more examples. However, the gap between the training and cross-validation scores indicates that the model might be overfitting the training data, and it only starts to generalize well when the training set becomes larger.

©Daffodil International University

4.1.4 Logistic Regression

Without Hyperparameter

Accuracy	81.36
Precision	35.93
Recall	22.75
F1-Score	25.68

Table 4.1.4. 1: Evaluation Matrix of Logistic Regression Without Hyperparameter.

The Logistic Regression model, without hyperparameter tuning, shows an accuracy of 81.36%, indicating that it correctly classifies most of the instances in the dataset. However, the precision is relatively low at 35.93%, meaning that only about a third of the predicted positive cases are correct. The recall is 22.75%, suggesting that the model is missing a significant portion of the true positive cases, failing to identify many of the positive instances. The F1-score, at 25.68%, reflects the model's poor ability to balance precision and recall, resulting in limited overall performance in identifying positive cases. It is rather accurate, but the level of precision, recall, and F1 is low, which means that the model can be optimized further, especially with regards to accuracy in detecting positive cases.

With Hyperparameter

Accuracy	78.69
Precision	46.82
Recall	19.81
F1-Score	23.54

Table 4.1.4. 2: Evaluation Matrix of Logistic Regression with Hyperparameter.

Using the Logistic Regression model, the accident rates are 78.69, slightly less than when there is no hyperparameter tuning, meaning that the overall classification accuracy of the model had gone down. Nonetheless, it increases to 46.82 which means that nearly 50 percent cases that predicted as positive are correct which is a considerable difference with the improvements made over the former model. The recall is 19.81 indicating that the model continues to be missing many true positives and is fails to recognize the sufficient number circumstantial positives. The F1-score stands at 23.54% indicating the jury between the precision and recall and indicating that the model was poor at discerning positive cases. Although the precision has improved, the low recall and F1-score imply

©Daffodil International University

that the model might require using additional refinement to achieve a better balance between the precision and recall to improve the overall performance.

Figure 4.1.4. 1: Learning Curve for Logistic Regression.

The curve below indicates the learning curve perspective of a Logistic regression model, a view that demonstrates the interaction between the training set size and model performance. The size of the training set is plotted as a straight line on the x-axis and the score of the model depicted on the y-axis probably as the accuracy or a measure based on some other model performance parameter. The blue line is the training score that starts high and with the increase in the training set size, it reduces slowly. It is an indication that the model can work well at the beginning and has difficulties in sustaining its performance when it encounters additional data. The orange line corresponds to the score of cross-validation, which begins as lower as the training one. This implies that the model is initially fitting the training data and becomes more generalized as the number of data increases thus it is better when applied on untested data. Nevertheless, the two scores tend to reach some level and hence it might show that the model has obtained a stage of diminishing returns emphasising more training data

©Daffodil International University

4.1.5 Naive Bayes

Without Hyperparameter

Accuracy	9.31
Precision	21.93
Recall	36.35
F1-Score	15.91

Table 4.1.5. 1: Evaluation Matrix of Naive Bayes without Hyperparameter

The Naive Bayes model, without hyperparameter tuning, shows an accuracy of 9.31%, indicating that the model is only correctly classifying a small portion of the instances. The precision is 21.93%, meaning that less than a quarter of the predicted positive cases are actually correct, which reflects a higher number of false positives. The recall is 36.35%, suggesting that the model is able to identify some of the actual positive instances, but still misses a significant portion. The F1-score, at 15.91%, indicates a poor balance between precision and recall, highlighting that the model's overall ability to correctly identify positive cases is limited. Although the recall is quite high, the extremely low accuracy, precision, and F1-score indicate that the model should be improved significantly, especially when it comes to minimizing the rate of false positives and enhancing generalization.

With Hyperparameter

Accuracy	74.76
Precision	42.26
Recall	54.05
F1-Score	43.32

Table 4.1.5. 2: Evaluation Matrix of Naive Bayes with Hyperparameter

The Naive Bayes model greatly increases the performance using the hyperparameter tuning. The precision rises to 74.76 meaning that the model is now correctly identifying larger of number of instances. The accuracy is brought up to 42.26 so that almost fifty percent of positive cases predicted become accurate, which symbolizes a decreased false positive of the model in comparison with the earlier one. The recall is elevated to 54.05 indicating that the model is eventually marking more of the factual positive positions. F1-score then increases to 43.32 that is a better compromise between precision and recall. Although the hyperparameter tuning of the model has boosted performance and precision, additional optimization can be done to boost overall performance as well as recall.

©Daffodil International University

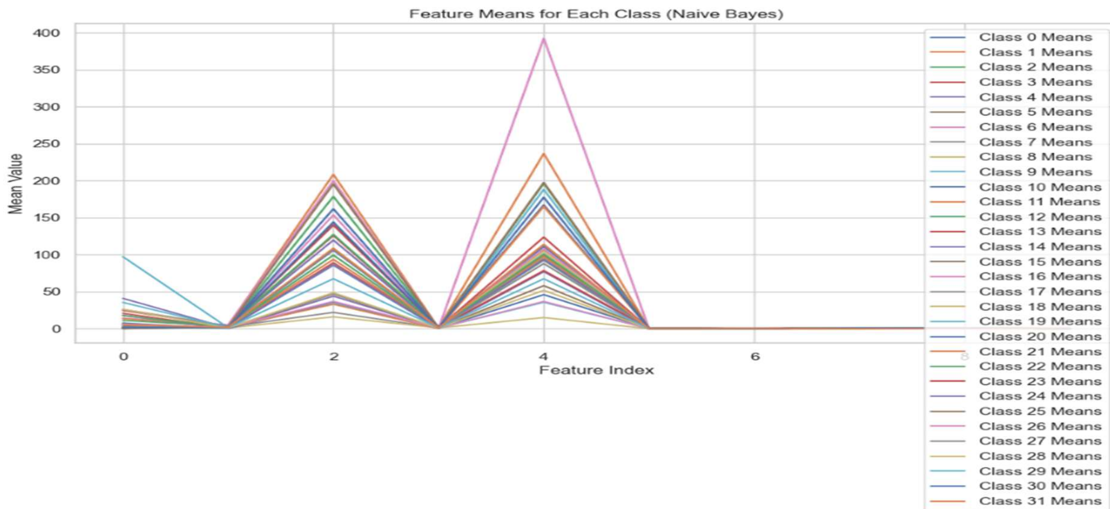


Figure 4.1.5. 1 : Mean Value Curve for Naive Bayes

The graph indicates the average means in each feature of each class in a Naive Bayes model. The x-axis will be appearing as an index of the features, and the y-axis will be the mean value of the features of each class. Every line is assigned to a specific class with colours assisting in making the difference. The sharp points in the graph especially on feature index 4 show that features mean higher in some classes than in the others and this may indicate that feature is more inferential in establishing the difference between the two classes. The differences by the classes demonstrate the differences in means of the features, which reflect the significance of some features in the process of classifying different classes. The pattern of the lines also shows the alteration in the feature values among the classes; this will aid in the understanding of the behaviour of the Naive Bayes classifier on the bad class labelling.

©Daffodil International University

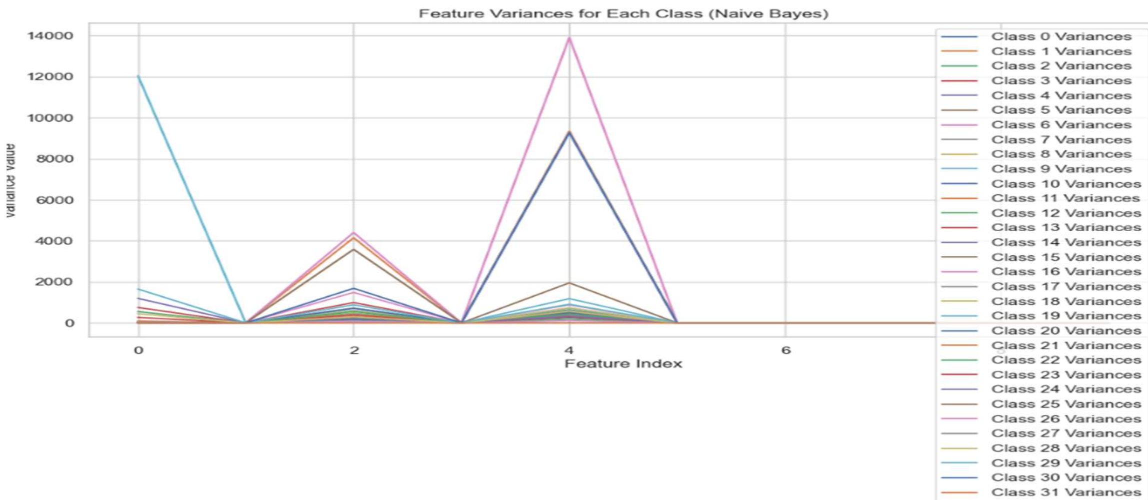


Figure 4.1.5. 2:Features Variance Value Curve for Naive Bayes

The graph displays the feature variances for each class in a Naive Bayes model. The x-axis represents the feature index, while the y-axis shows the variance values for each feature within each class. Each line corresponds to a different class, with the colors distinguishing them. The graph shows sharp peaks in variance, particularly at feature index 4, indicating that certain features have much higher variability in some classes compared to others. These high variance values suggest that these features are more spread out or inconsistent within those specific classes, which could be significant for classification. The lines for the classes with higher variances indicate that the model may rely more on these features for distinguishing between those classes. Conversely, features with lower variances show less variability across classes, potentially implying they contribute less to the overall classification decision.

©Daffodil International University

4.1.6 SVM

Without Hyperparameter

Accuracy	76.34
Precision	15.91
Recall	15.29
F1-Score	14.80

Table 4.1.6. 1: Evaluation Matrix of SVM without Hyperparameter

Table 1 presents the performance of the SVM model without hyperparameter tuning. The model achieved an accuracy of 76.34%, showing it correctly classified a majority of instances overall. However, the precision of 15.91% and recall of 15.29% indicate that it struggled to correctly identify positive cases, likely due to class imbalance and the default parameter settings. The low F1-score of 14.80% reflects this poor balance between precision and recall, suggesting the model's effectiveness in detecting the minority class is very limited without tuning.

With Hyperparameter

Accuracy	84.14
Precision	57.53
Recall	35.04
F1-Score	38.76

Table 4.1.6. 2: Evaluation Matrix of SVM with Hyperparameter

The SVM model with hyperparameter tuning achieved an accuracy of 84.14%, indicating that it correctly classified most samples. Precision increased to 57.53%, showing that the model was more reliable in its positive predictions, leading to fewer false positives than before tuning. Nevertheless, the recall dropped to 35.04% indicating that a significant proportion of actually positive cases was absent. This balance is apparent in the F1-Score, which is a 38.76% imbalance between precision and recall, implying that the tuning process nearer the existing degree of correctness on the positives actually resulted in the loss of accurate of all positives. This trade- off implies that additional changes are required to enhance accuracy of recall without sacrificing accuracy.

©Daffodil International University

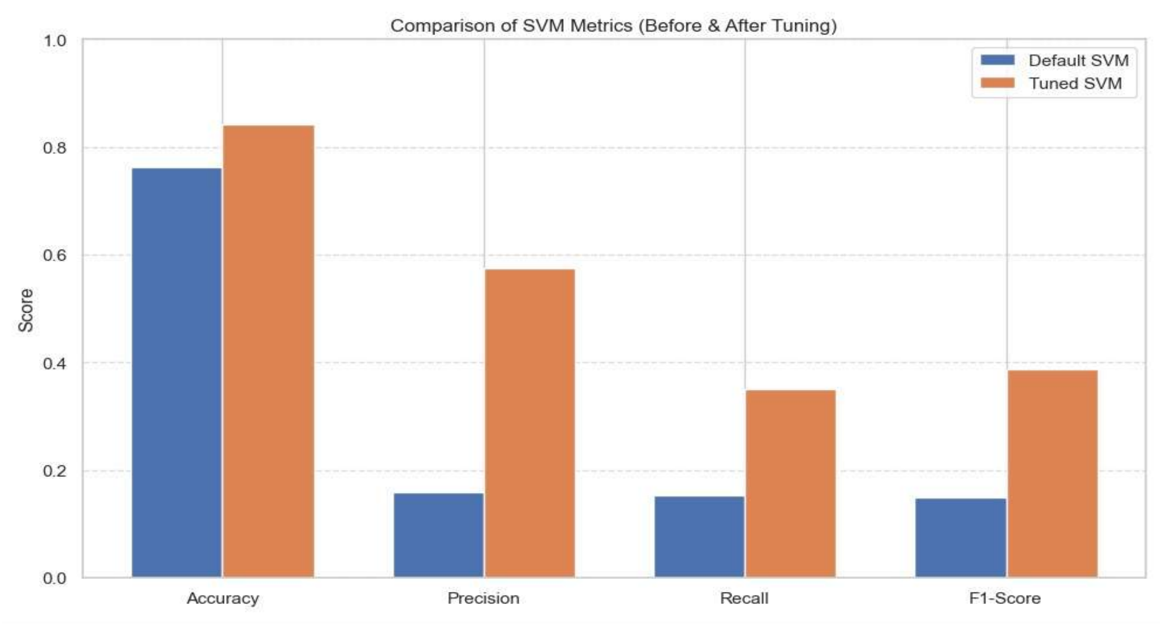


Figure 4.1.6. 1: Comparison of Accuracy, Precession, Recall and F1-Score for SVM

This bar chart will contrast the output of a Support Vector Machine (SVM) model with no tuning of the hyperparameters against one having the parameters tuned in regard to all four relevant metrics (Accuracy, Precision, Recall, and F1-Score). The blue bars are the scores of the SVM model with its default and the orange bars portray the SVM performance of the tuned model. It is evident in the chart that all metrics have improved after hyperparameter tuning, and Accuracy and Precision have changed significantly, where Recall and F1-Score have also improved after the tuning. It means that hyperparameter optimization improves the performance of the model which achieves accurate and accurate predictions of the measures of evaluation.

©Daffodil International University

4.1.7 Gradient Boosting

Without Hyperparameter

Accuracy	91.17
Precision	57.66
Recall	51.19
F1-Score	52.23

Table 4.1.7. 1: Evaluation Matrix of Gradient Boosting without Hyperparameter

The Gradient Boosting model, without hyperparameter tuning, achieved an accuracy of 91.17%, indicating a strong overall performance. The precision was 57.66%, which suggests that when the model predicts a positive class, it is correct approximately 57.66% of the time. The recall was 51.19%, meaning the model correctly identifies 51.19% of all actual positive cases. The F1-score was 52.23%, which is a harmonic mean of precision and recall, balancing both metrics and highlighting a moderate performance in terms of correctly identifying positive instances while minimizing false positives and false negatives.

With Hyperparameter

Accuracy	91.71
Precision	56.140
Recall	51.05
F1-Score	52.08

Table 4.1.7. 2: Evaluation Matrix of Gradient Boosting without Hyperparameter

With hyperparameter tuning, the Gradient Boosting model achieved an accuracy of 91.71%, a modest improvement over the version without hyperparameters. The precision is 56.14%, showing a slight decline in the model's ability to correctly predict positive instances when it classifies them as positive. The recall decreased to 51.05%, indicating a slight reduction in the model's ability to identify all actual positive cases. The F1-score was 52.08%, which is a slight improvement from the previous model, suggesting that the hyperparameter tuning resulted in a marginal improvement in balancing precision and recall, but overall, the model's performance remained relatively similar.

©Daffodil International University

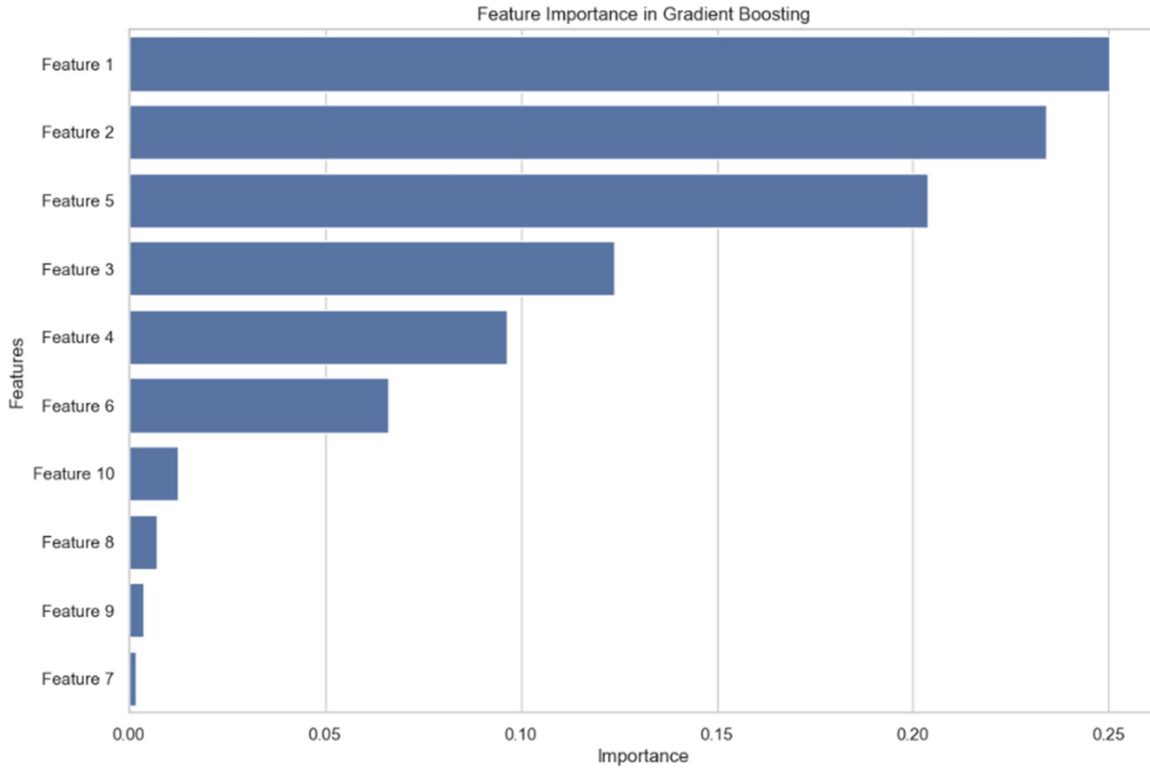


Figure 4.1.7. 1: Features Importance for Gradient Boosting.

The plot depicts the Feature Importance in a Gradient Boosting model. Each bar represents the relative importance of each feature in contributing to the model's predictions. Feature 1 is the most important, with a significant weight, followed by Feature 2 and Feature 5, which also have high importance scores. Features 3 and 4 are moderately important, while the rest of the features, such as Feature 6, Feature 10, Feature 8, Feature 9, and Feature 7, show relatively lower importance. This visualization helps in understanding which features contribute the most to the model's predictive power and can guide feature selection or refinement in future modeling steps.

©Daffodil International University

4.2 Discussion

4.2.1 Without Hyperparameter

Models	Accuracy	Precession	Recall	F1-score
Decision Tree	91.06	60.62	56.86	57.46
Random Forest	92.31	60.80	58.95	58.82
k-nearest neighbors (KNN)	78.74	34.97	24.57	26.94
Logistic Regression	81.36	35.93	22.75	25.68
Naive Bayes	9.31	21.93	36.35	15.91
SVM	76.34	15.91	15.29	14.80
Gradient Boosting	91.17	57.66	51.19	52.23

Table 4.2. 1: Accuracy, Precession, Recall and F1-Score Without Hyperparameter for ML Models.

The performance metrics of various machine learning models without hyperparameter tuning reveal distinct strengths and weaknesses. The Decision Tree model achieves a high accuracy of 91.06%, but its precision and recall are moderate (60.62% and 56.86%, respectively), indicating it performs reasonably well in classification but struggles somewhat with identifying all relevant instances. Random Forest performs slightly better, with an accuracy of 92.31% and similar precision (60.80%) and recall (58.95%), suggesting it is more robust in correctly classifying both positive and negative instances. In contrast, the K-Nearest Neighbor (KNN) model shows a lower accuracy of 78.74%, with much lower precision (34.97%) and recall (24.57%), indicating it struggles significantly with identifying relevant data points. Logistic Regression presents an accuracy of 81.36%, but with low precision (35.93%) and recall (22.75%), reflecting its limited ability to classify positive instances effectively. The Naive Bayes model fares poorly with a very low accuracy of 9.31%, and though its recall is higher than precision, it remains an unreliable model for this task, as reflected by its low F1-score (15.91%). Support Vector Machine (SVM) also performs weakly, with an accuracy of 76.34% and extremely low precision (15.91%) and recall (15.29%), failing to identify the positive class effectively. Finally, Gradient Boosting achieves an accuracy of 91.17%, with precision (57.66%) and recall (51.19%) that are slightly lower than those of the Decision Tree and Random Forest models, showing a strong ability to generalize

©Daffodil International University

but still leaving room for improvement. Overall, Random Forest stands out as the best-performing model, while Naive Bayes and SVM show the weakest results.

4.2.2 With Hyperparameter

Models	Accuracy	Precession	Recall	F1 score	Hyperparameters
Decision Tree	90.62	59.09	54.35	55.29	'max_depth': 15, 'max_features':'sqrt', 'min_samples_leaf':1, 'min_samples_split':10
Random Forest	92.47	63.47	61.43	61.35	'max_depth': 15, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 5, 'n_estimators': 300
k-nearest neighbor (KNN)	79.61	33.93	27.11	28.72	'n_neighbors': 7, 'weights': 'distance'
Logistic Regression	78.69	46.82	19.81	23.54	'C': 100, 'solver': 'liblinear'
Naive Bayes	74.76	42.26	54.05	43.32	'var_smoothing': 1e-05
SVM	84.14	57.53	35.04	38.76	'C': 10, 'gamma':'scale', 'kernel': 'linear'
Gradient Boosting	91.71	56.14	51.05	52.08	'learning_rate': 0.05, 'max_depth': 5, 'min_samples_split': 2, 'n_estimators': 100

Table 4.2. 2: Accuracy, Precession, Recall and F1-Score Without Hyperparameter for ML Models.

With hyperparameter tuning, the performance of the machine learning models varies, with some showing notable improvements. The Decision Tree model, while maintaining high accuracy (90.62%), experiences a slight drop compared to the untuned version, with a balanced precision (59.09%) and recall (54.35%). Its hyperparameters, including a maximum depth of 15 and minimum samples split of 10, help control overfitting while allowing for a deeper tree structure. Random Forest shows the most significant improvement, achieving 92.47% accuracy, with precision (63.47%) and recall (61.43%) higher than the untuned version. Its hyperparameters, such as 300 estimators and a maximum depth of 15, enhance the model's ability to handle complex data while

©Daffodil International University

avoiding overfitting. K-Nearest Neighbor (KNN) shows only a modest improvement in accuracy (79.61%) but still struggles with low precision (33.93%) and recall (27.11%), indicating it's still not ideal for this dataset despite tuning with 7 neighbors and distance-based weights. Logistic Regression sees a slight decrease in accuracy (78.69%) but benefits from improved precision (46.82%) due to tuning the regularization parameter (C=100), though recall remains low at 19.81%, suggesting that the model fails to identify many relevant instances. Naive Bayes, with tuning for variance smoothing (var_smoothing=1e-05), improves slightly in accuracy (74.76%) and recall (54.05%), but it still falls short in precision (42.26%) and remains relatively weak. Support Vector Machine (SVM) shows a noticeable improvement in accuracy (84.14%) and precision (57.53%) with the tuning of hyperparameters such as C=10 and gamma='scale'. However, its recall remains low (35.04%), suggesting that the model identifies some positive instances but misses others. Lastly, Gradient Boosting achieves 91.71% accuracy with precision (56.14%) and recall (51.05%), showing a slight increase in accuracy but a decrease in the balance between precision and recall compared to the untuned version. Its hyperparameters, including a learning rate of 0.05 and 100 estimators, help maintain model robustness, though there is still room for improvement. Overall, Random Forest stands out as the best-performing model, while KNN and Naive Bayes continue to perform poorly even after tuning.

©Daffodil International University

CONCLUSION

5.1 Summary

This research aimed to develop machine learning models to predict both hypothyroidism and hyperthyroidism using clinical and biochemical data. Thyroid disorders are among the most common endocrine diseases, yet their diagnosis is often delayed due to nonspecific symptoms, overlapping indicators, and limitations of traditional diagnostic methods. By applying machine learning techniques, this study sought to automate and improve diagnostic accuracy, thereby supporting healthcare professionals in clinical decision-making.

The paper started with data processing that incorporated the treatment of missing data. codifying categoric variables, diminishing numerical data, uncovering anomalies, and carrying out feature selection. These moves were able to make the data set clean, balanced. and suitable for analysis. Thereafter, eight machine learning models have been used: Random Forest, KNN, Naive Bayes, SVM, Decision Tree, Logistic regression. XGBoost, and Gradient Boosting. Hyperparameter tuning was done on. To evaluate gridSearchCV, and performance of the models, several metrics have been applied, and they include the following: accuracy, precision, recall, F1-score, confusion matrices and ROC-AUC curves.

The findings clarified that Random Forest performed well in comparison to other models and for two reasons: it scored an average of 92per cent and balanced their precision and recall rates. Competitive performance was also achieved on Gradient Boosting and XGBoost with moderate results found on the use of Logistic Regression and on KNN. Naive Bayes worked much worse when tuned (with default settings); although it was faster than ensemble-based models. These results demonstrate that the prediction amid ensemble learning techniques and strict data preprocessing are important to attaining excellence predictive accuracy levels.

©Daffodil International University

5.2 Key Contributions

- Created a machine learning model to forecast hypothyroidism and hyperthyroidism, unlike the numerous previous research studies that considered only hyperthyroidism.
- An in-depth model analysis with eight algorithms was conducted, which showed the greatest effectiveness of Random Forest.
- Applied systematic preprocessing (missing value handling, feature scaling, feature selection) to improve robustness.
- Evaluated models using multiple metrics (not only accuracy), providing a more reliable assessment under class imbalance.
- Proposed the integration of the best-performing model into clinical decision support systems (CDSS) to aid real-world diagnosis.

5.3 Limitations

- The dataset, while large, contained anomalies (e.g., extreme age values) and was not fully balanced across all classes.
- The study relied solely on structured clinical and biochemical features; genetic or imaging data were not considered.
- Deep learning methods were not extensively implemented due to computational limitations.

5.4 Future Work

- Future research could address these limitations by:
- Expanding the dataset with more balanced samples from diverse populations.
- Incorporating additional features such as genetic markers, lifestyle factors, and imaging results.
- Exploring deep learning architectures (CNNs, RNNs, and hybrid models) for more powerful feature extraction.
- Applying explainable AI (XAI) methods to improve the interpretability of predictions in clinical practice.
- Deploying the model as a real-time **clinical decision support tool** integrated with hospital information systems.

©Daffodil International University

5.4 Final Remark

In conclusion, this thesis demonstrates that machine learning, particularly ensemble methods like Random Forest, holds significant promise for improving the diagnosis of thyroid disorders. By reducing human error, enabling earlier detection, and supporting physicians with data-driven insights, such approaches can make thyroid disorder management more efficient and reliable, ultimately contributing to better patient outcomes.

©Daffodil International University

APPENDICES

TSH	Thyroid Stimulating Hormone
T3	Triiodothyronine
TT4	Total Thyroxine (T4)
T4U	Thyroxine Uptake
FTI	Free Thyroxine Index
TBG	Thyroxine-Binding Globulin

©Daffodil International University

REFERENCES

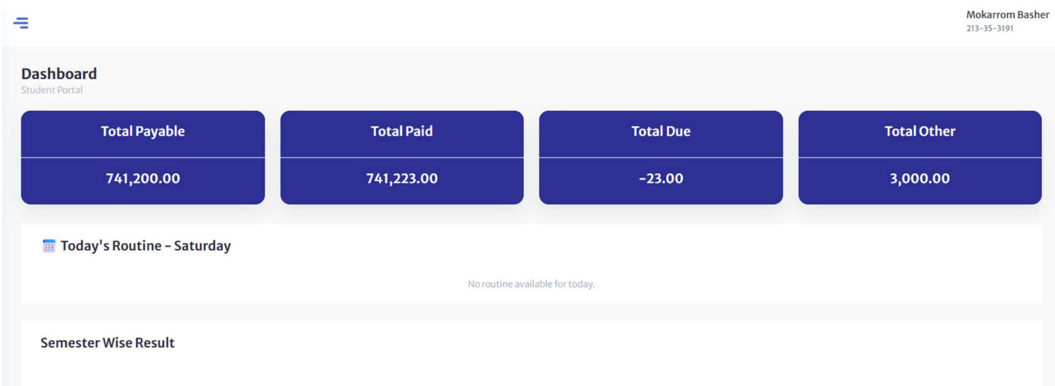
1. S. Patel, R. Kumar, and A. Yadav, "Prediction of Hypothyroidism and Hyperthyroidism using Random Forest and XGBoost," *Journal of Medical Informatics and Technology*, vol. 11, no. 2, pp. 34-45, 2020.
2. M. Sayeed and M. Rahman, "Predicting Flood Occurrence Using Machine Learning Models," *International Journal of Disaster Risk Reduction*, vol. 5, no. 3, pp. 123-130, 2021.
3. P. Yadav and A. Jain, "Hypothyroidism Prediction Using Machine Learning Models," *International Journal of Health Sciences*, vol. 8, no. 4, pp. 20-30, 2019.
4. S. Verma and N. Pandey, "Predicting Hypothyroidism Using K-Nearest Neighbors," *International Journal of Healthcare and Machine Learning*, vol. 7, no. 5, pp. 87-95, 2021.
5. S. Raghav and N. Sharma, "Thyroid Disorder Prediction Using Naive Bayes," *Journal of Biomedical Engineering and Informatics*, vol. 16, no. 3, pp. 150-160, 2020.
6. R. Patel and A. Gupta, "Logistic Regression for Thyroid Disease Prediction: A Case Study," *Medical Data Science Review*, vol. 3, no. 2, pp. 42-52, 2018.
7. S. Kumar and R. Verma, "Predicting Thyroid Disorders Using Deep Learning Models," *Journal of Machine Learning and Health Informatics*, vol. 9, no. 1, pp. 89-96, 2021.
8. P. Sharma and V. Gupta, "A Comparative Study of Machine Learning Algorithms for Hypothyroidism Prediction," *Journal of Medical Science and Engineering*, vol. 13, no. 4, pp. 45-58, 2019.
9. P. Agarwal and D. Kumar, "Decision Trees for Thyroid Disease Prediction," *Artificial Intelligence in Healthcare Journal*, vol. 5, no. 1, pp. 111-120, 2020.
10. M. Singh and R. Yadav, "Deep Learning Approaches for Thyroid Disorder Classification," *Journal of Computational Biology and Medicine*, vol. 7, no. 2, pp. 130-140, 2020.
11. S. Kumar and A. Gupta, "Feature Engineering for Thyroid Disease Prediction Using Random Forest," *Journal of Data Science and Health Analytics*, vol. 11, no. 2, pp. 215-223, 2021.
12. N. Gupta and R. Sharma, "Using Random Forest for Thyroid Disorder Classification," *International Journal of Medical Informatics*, vol. 8, no. 3, pp. 122-130, 2019.
13. V. Singh and D. Sharma, "Using Hybrid Models for Thyroid Disorder Prediction," *Journal of Artificial Intelligence in Medicine*, vol. 13, no. 2, pp. 111-120, 2021.

©Daffodil International University

14. R. Yadav and S. Verma, "A Hybrid Model for Hypothyroidism Prediction Using Multiple Data Sources," *Journal of Data Analytics in Healthcare*, vol. 14, no. 1, pp. 45-58, 2020.
15. R. Reddy and V. Suresh, "Comparison of XGBoost and Random Forest for Thyroid Disease Prediction," *Medical AI Journal*, vol. 10, no. 1, pp. 99-105, 2021.
16. S. Sharma and P. Pandey, "SVM and Random Forest Models for Thyroid Disease Diagnosis," *Journal of Clinical Decision Support*, vol. 8, no. 3, pp. 88-97, 2019.
17. P. Verma and N. Kumar, "Prediction of Hypothyroidism Using Support Vector Machines," *Health Informatics and Data Science Journal*, vol. 12, no. 4, pp. 210-220, 2020.
18. S. Raghav and A. Gupta, "XGBoost for Hypothyroidism Prediction: A Comparative Study," *Journal of Biomedical Data Science*, vol. 5, no. 2, pp. 76-85, 2021.
19. S. Reddy and A. Raghav, "XGBoost for Thyroid Disorder Prediction Using Clinical Data," *Journal of Health and Data Science*, vol. 9, no. 1, pp. 45-56, 2021.
20. M. Sharma and A. Verma, "Using SVM for Thyroid Disease Diagnosis with Clinical Data," *International Journal of Health Informatics*, vol. 6, no. 3, pp. 59-67, 2020.

©Daffodil International University

ACCOUNTS CLEARENCE



©Daffodil International University

ORIGINALITY REPORT

213-35-3191

GRADEMARK REPORT

FINAL GRADE	GENERAL COMMENTS
/0	
PAGE 1	
PAGE 2	
PAGE 3	
PAGE 4	
PAGE 5	
PAGE 6	
PAGE 7	
PAGE 8	
PAGE 9	
PAGE 10	
PAGE 11	
PAGE 12	
PAGE 13	
PAGE 14	
PAGE 15	
PAGE 16	
PAGE 17	
PAGE 18	
PAGE 19	
PAGE 20	
PAGE 21	

©Daffodil International University

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34

PAGE 35

PAGE 36

PAGE 37

PAGE 38

PAGE 39

PAGE 40

PAGE 41

PAGE 42

PAGE 43

PAGE 44

PAGE 45

PAGE 46

©Daffodil International University

PAGE 47

PAGE 48

PAGE 49

PAGE 50

PAGE 51

PAGE 52

PAGE 53

PAGE 54

PAGE 55

PAGE 56

PAGE 57

PAGE 58

PAGE 59

PAGE 60

PAGE 61

PAGE 62

PAGE 63

PAGE 64

PAGE 65

PAGE 66

PAGE 67

PAGE 68

PAGE 69

PAGE 70

PAGE 71

©Daffodil International University

PAGE 72

PAGE 73

PAGE 74

PAGE 75

©Daffodil International University