

**A NOVEL EXPLAINABLE AI BASED APPROACH: EARLY-STAGE  
THYROID DISEASE CLASSIFICATION**

**BY**

**MD. NIZHUM  
ID: 201-15-3043**

This Report Presented in Partial Fulfillment of the Requirements for the  
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Nasima Islam Bithi**  
Lecturer  
Department of CSE  
Daffodil International University

Co-Supervised By

**Mr. Md Assaduzzaman**  
Lecturer (Senior Scale)  
Department of CSE  
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

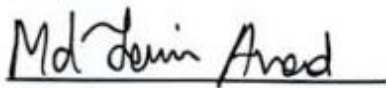
**DHAKA, BANGLADESH**

**JANUARY 2024**

## APPROVAL

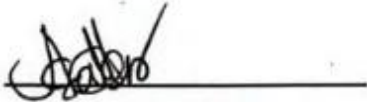
This research-based project titled “A Novel Explainable AI Based Approach: Early-Stage Thyroid disease Classification”, submitted by MD. Nizhum, ID No: 201-15-3043 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26 January 2024.

### BOARD OF EXAMINERS



**Chairman**

**Dr. Md. Taimur Ahad (MTA)**  
**Associate Professor & Associate Head**  
Department of Computer Science and Engineering  
Daffodil International University



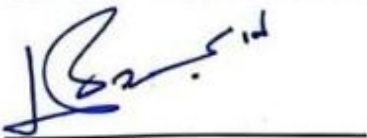
**Internal Examiner**

**Abdus Sattar (AS)**  
**Assistant Professor**  
Department of Computer Science and Engineering  
Daffodil International University



**Internal Examiner**

**Tapasy Rabeya (TRA)**  
**Senior Lecturer**  
Department of Computer Science and Engineering  
Daffodil International University



**External Examiner**

**Dr. Md. Sazzadur Rahman (MSR)**  
**Professor**  
Institute of Information Technology  
Jahangirnagar University

## DECLARATION

I hereby declare that this project has been done by us under the supervision of **Nasima Islam Bithi, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

**Supervised by:**

*For, [Signature] 21.01.24*

**Nasima Islam Bithi**

Lecturer

Department of CSE

Daffodil International University

**Co-Supervised by:**

*[Signature] 21.01.24*

**Mr. Md Assaduzzaman**

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

**Submitted by:**

*Nizhum*

**MD. Nizhum**

ID: 201-15-3043

Department of CSE

Daffodil International University

## ACKNOWLEDGEMENT

First, I express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year research-based project successfully.

I really grateful and wish our profound our indebtedness to **Nasima Islam Bithi, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Head**, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

I would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

## ABSTRACT

As of the most current information, thyroid illnesses remain widespread on a worldwide scale. The rate of hypothyroidism among Americans is around 4.6%, with a higher incidence seen in women and older individuals. Meanwhile, hyperthyroidism affects around 1.3% of the population. Globally, iodine deficiency continues to be a primary factor contributing to thyroid problems, especially in areas with limited availability of iodine-enriched diet or supplements. Concerns about thyroid disease should be raised for human health due to the thyroid gland's involvement in regulating human metabolism and its essential impact on overall well-being. Through this research on the thyroid, 3 types of thyroid disease can be identified (hyperthyroid, hypothyroid, negative). To do this research I used several machine learning models (Logistics Regression, Decision Tree, Naive Bayes, Random Forest, Bagging Classifier, XGBoost, SGD, AdaBoost, Grid Search Bagging, Grid Search Logistic, Support Vector Machine (SVM), ANN) etc. Among all the models some of the best performing models are: Bagging accuracy 99.04%, XGBoost accuracy 98.78%, Random Forest accuracy 98.67%, Grid Search Bagging accuracy 98.54%. Among these models, the Bagging classifier model performed best, so here selected this model as the best and final model. Then added Explainable AI to the final model. Explainable AI's job here is to explain how the model is making decisions based on which features to predict results. Previous papers have dealt with 1 or 2 classes. worked on 3 classes (hyperthyroid, hypothyroid, negative) at once to overcome that limitation, So, here can say that the time spent between these tasks is reasonable.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE NO</b>
Board of Examiners	i
Declaration	ii
Acknowledgements	ii
Abstract	iv
List of Figures	ix-x
List of Tables	xi
<b>CHAPTER</b>	
<b>CHAPTER-1: INTRODUCTION</b>	<b>1-6</b>
1.1 Introduction	1
1.2 Motivation	2
1.3 Problem Statements	4
1.4 Research Objectives	4
1.5 Expected Output	5
1.6 Research Questions	5
1.7 Report Layout	6
<b>CHAPTER 2: BACKGROUND</b>	<b>7-12</b>
2.1 Preliminaries/Terminologies	7
2.2 Related Works	8
2.3 Comparative Analysis and Summary	10
2.4 Scope of the Problem	11
2.5 Challenges	12
<b>Chapter 3: RESEARCH METHODOLOGY</b>	<b>13-35</b>

3.1 Research Subject and Instrumentation	13
3.2 Data Collection Procedure/Dataset Utilized	13
3.2.1 ThyroidDF	15
3.3 Data Pre-Processing	15
3.3.1 Null Value Handling	16
3.3.2 Outlier Handling	16
3.3.3 Scaling	16
3.3.4 Encoding	16
3.3.5 Label Encoding	16
3.3.6 Feature Selection	17
3.3.7 Train Test Split	17
3.3.8 Cross Validation	17
3.4 Proposed Methodology/Applied Mechanism	20
3.5 Model Apply	21
3.5.1 Decision Tree	21
3.5.2 K-Nearest Neighbors	22
3.5.3 Logistic Regression	23
3.5.4 Bagging	23
3.5.5 AdaBoost	24
3.5.6 Bernoulli Naïve Bayes	25
3.5.7 Random Forest	26
3.5.8 Support Vector Machine	27
3.5.9 Stochastic Gradient Descent	27
3.5.10 XGBoost	28
3.5.11 Grid Search Bagging	29

3.5.12 Grid Search Logistic	30
3.5.13 Artificial Neural Network	30
3.5.14 Explainable AI	32
3.5.15 Lime	32
3.5.16 Shap	33
3.5.17 Explainable AI's role in this research	33
3.6 Evaluation Metrics	33
3.7 Implementation Requirements	35

## **CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION 36-57**

4.1 Experimental Setup	36
4.2 Results & Analysis	36
4.2.1 Best & Final Model	37
4.2.2 Decision Tree	38
4.2.3 K-Nearest Neighbors	39
4.2.4 Logistic Regression	40
4.2.5 Bagging	42
4.2.6 AdaBoost	43
4.2.7 Bernoulli Naive Bayes	44
4.2.8 Random Forest	46
4.2.9 Support Vector Machine	47
4.2.10 Stochastic Gradient Descent	48
4.2.11 XGBoost	50
4.2.12 Grid Search Bagging	51
4.2.13 Grid Search Logistic Regression (GSLR)	52
4.2.14 Artificial Neural Network (ANN)	54



4.2.15 Explainable AI Visualization	55
4.3 Comparative Analysis	57
<b>CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY</b>	<b>58</b>
5.1 Impact on Society	58
5.2 Ethical Aspects	58
5.4 Sustainability Plan	58
<b>CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH</b>	<b>59-60</b>
6.1 Summary of the Study	59
6.2 Conclusions	59
6.3 Implication for Further Study	60
<b>REFERENCES</b>	<b>61-65</b>

## LIST OF FIGURES

<b>FIGURES</b>	<b>PAGE NO</b>
Figure 1.2.1: Thyroid patient	2
Figure 1.2.2: Thyroid patient	2
Figure 1.2.3: Thyroid disease ratio	2
Figure 1.2.4: Thyroid gland	3
Figure 3.2.1: CSV dataset.	13
Figure 3.2.2: Dataset feature list	14
Figure 3.3: Pre-Processing steps	15
Figure 3.3.8.1: Box plot visualization of some feature after preprocessing.	17
Figure 3.3.8.2: Sea born plots visualization of some feature after preprocessing.	18
Figure 3.3.8.3: Correlation heat map visualization after preprocessing.	19
Figure 3.4: Flow chart of proposed methodology	20
Figure 3.5.1: Decision Tree model design	22
Figure 3.5.2: K-Nearest Neighbors model design	22
Figure 3.5.3: Logistic Regression model design	23
Figure 3.5.4: Bagging model design	24
Figure 3.5.5: AdaBoost model design	25
Figure 3.5.6: Naïve Bayes model design	26
Figure 3.5.7: Random Forest model design	27
Figure 3.5.9: SGD model design	28
Figure 3.5.10: XGBoost model design	29
Figure 3.5.13: ANN model design	32
Figure 3.6: Confusion matrix	34
Figure 4.2.2.1: Confusion matrix (DT)	38
Figure 4.2.2.2: ROC curve (DT)	38
Figure 4.2.3.1: Confusion matrix (KNN)	39

Figure 4.2.3.2: ROC curve (KNN)	40
Figure 4.2.4.1: Confusion matrix (LR)	40
Figure 4.2.4.2: ROC curve (LR)	41
Figure 4.2.5.1: Confusion matrix (BG)	42
Figure 4.2.5.2: ROC curve (BG)	42
Figure 4.2.6.1: Confusion matrix (ADA)	43
Figure 4.2.6.2: ROC curve (ADA)	44
Figure 4.2.7.1: Confusion matrix (NB)	44
Figure 4.2.7.2: ROC curve (NB)	45
Figure 4.2.8.1: Confusion matrix (RF)	46
Figure 4.2.8.2: ROC curve (RF)	46
Figure 4.2.9.1: Confusion matrix (SVM)	47
Figure 4.2.9.2: ROC curve (SVM)	48
Figure 4.2.10.1: Confusion matrix (SGD)	48
Figure 4.2.10.2: ROC curve (SGD)	49
Figure 4.2.11.1: Confusion matrix (XGB)	50
Figure 4.2.11.2: ROC curve (XGB)	50
Figure 4.2.12.1: Confusion matrix (GSB)	51
Figure 4.2.12.2: ROC curve (GSB)	52
Figure 4.2.13.1: Confusion matrix (GSLR)	52
Figure 4.3.13.2: ROC curve (GSLR)	53
Figure 4.2.14.1: ANN Model Epoch vs Loss Plot-1	54
Figure 4.2.14.2: ANN Model Epoch vs Validation Accuracy Plot-2	54
Figure 4.2.15.1: Lime visualization of Bagging model	55
Figure 4.2.15.2: Shap visualization of Bagging model	56

## LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
Table 2.3: Comparative Table of Other Paper's	10-11
TABLE 3.2.1: Class and instances counts for ThyroidDF	15
Table 4.2: Showed all-model result	37
Table 4.3: Proposed model accuracy comparison with other's model for ThyroidDF	57

# CHAPTER 1

## Introduction

### 1.1 Introduction

Thyroid is an endocrine gland, it is an important part of our body, it is located in the middle of the throat and looks like butterfly wings. That is a small organ approximately 2 inches. It produces thyroxine hormone. Through this gland, our body produces the necessary energy from food. Hormones secreted by the thyroid gland help in digestion of food. Mainly thyroid hormones produce carbohydrates, proteins in the body. Also, thyroid hormone helps in blood pressure, heart rate, body temperature, growth and helps in reducing cholesterol. Due to thyroid diseases, many types of side effects occurred in our body such as weight loss, anxiety, nervousness, mood swings, hand tremors, increased heart rate, excessive heat, body weakness, irregular periods in girls, insomnia, poor eyesight. Dizziness, blurred vision, hair loss, and various other problems. This disease causes various problems on the thyroid, hypothyroidism, hyperthyroidism, thyroid nodules, thyroiditis. Hypothyroidism is when thyroid gland secretes less than required quantity of hormones. If thyroid gland secretes more hormones than required, the condition is called hyperthyroid problem. Thyroid nodules occur when the thyroid gland grows at an abnormal rate. If the thyroid gland is damaged due to various infections, it is called thyroiditis. May also cause thyroid cancer. To diagnose this disease, blood levels (T3, T4, TSH) are measured, also tests are done through CT-scans, MRIs. Currently, thyroid disease has become a common disease, this illness affects 40% of the global population. About 50 million humans affected from thyroid disease in Bangladesh and in between 30 million people not aware of their situation. Machine learning is crucial in the diagnosis of thyroid problems. Machine learning can work with large datasets, processing large datasets easily. Can analyze different types of data, machine data, imaging data, which humans cannot do. Machine learning algorithm can accurately detect diseases, symptoms in less time.

Machine learning models have been trained to recognize early indicators or risk factors linked to thyroid illnesses by examining past data from individuals with these problems. Before the illness worsens, prompt intervention and therapy are made possible by this early prognosis [1]. Based on unique patient data, machine learning models may assist in

developing individualized treatment regimens that take into account a variety of variables, including age, gender, genetic predisposition, environment, and more. This customized strategy decreases needless medical interventions while improving patient outcomes. Healthcare resources may be utilized more effectively with machine learning supporting early detection. In order to maximize the use of medical resources and experts' time, patients who are recognized as being at risk might be given priority for further testing or consultations with specialists. Through constant learning from fresh data, machine learning models are able to adjust and enhance their diagnostic precision over time. The models are kept up to speed with the most recent developments and medical knowledge thanks to this ongoing learning process.

## 1.2 Motivation

Machine learning helps in many ways in the diagnosis of thyroid diseases. By increasing early detection, accuracy and efficiency, each patient can be monitored separately, costs can be reduced, health care will be more accessible, health system will be improved.



Figure 1.2.1: Thyroid patient [2]



Figure 1.2.2: Thyroid patient [3]

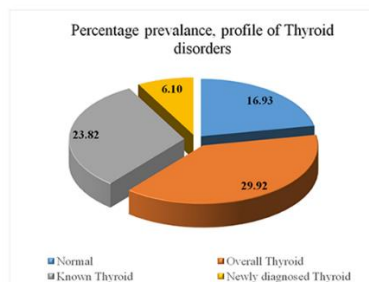


Figure 1.2.3: Thyroid disease ratio [4]

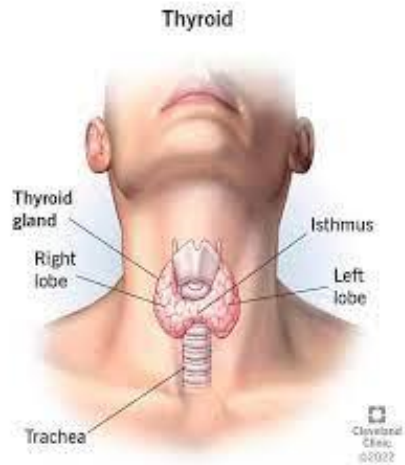


Figure 1.2.4: Thyroid gland [5]

- **Early Detection:** If thyroid problems are not identified or addressed, serious consequences may result. Through the analysis of many data points, machine learning algorithms may assist in early detection, which may result in prompt treatments and better patient outcomes.
- **Accuracy and Efficiency:** Large volumes of patient data, lab findings, imaging scans, and symptoms may all be processed more rapidly and correctly by machine learning models than by more conventional techniques. By increasing accuracy, misdiagnosis rates may be decreased and healthcare resource allocation can be optimized.
- **Regular Improvement:** Machine learning models have the capacity to continually learn from fresh data, which allows them to gradually adapt and enhance their diagnostic skills. The models remain current with the most recent developments in medicine because to this continuous learning process.
- **Personalized Patient Care:** By taking into account specific patient data, machine learning makes it possible to create individualized treatment plans that result in more specialized and successful treatments. Both patient happiness and health outcomes are enhanced by this strategy.
- **Accessible Healthcare:** Machine learning can assist in making healthcare more effective and accessible by supporting early detection and diagnosis for medical

personnel. This is especially useful in resource-constrained locations or situations where specialized knowledge may be hard to come by.

- **Research and Development:** Using machine learning methods for identification thyroid illness may further medically research lead to the creation of new understandings, therapies, and diagnostic instruments.
- **Reduced Healthcare Costs:** Because machine learning may help with early identification and action, long-term healthcare expenses may be lowered by averting problems and lowering the need for lengthy hospital stays or treatments.

All in all, diagnosing thyroid disease through machine learning is a breakthrough. This will open up new possibilities in our health sector. The number of patients suffering from thyroid disease can be reduced. Treatment of this disease will become easier.

### **1.3 Problem Statements**

After reading the papers of other researchers, there are still many limitations in their work. Some accuracy is low, some of the paper algorithms are not optimized properly, and some of paper's dataset is small. Few researchers worked with specific area or location example- (Iraq, Africa). Some models take longer time to get results. Some models not perform as expected/well, and also some of model correctly classification rate is low. Some models only work with 2 or 1 symptom of thyroid disease. In many cases the model cannot detect the model dataset well, resulting in not good output/accuracy from the model. There are also various limitations. I will try to overcome those limitations in my paper.

### **1.4 Research Objectives**

I will try to overcome the limitations that still exist in the diagnosis of thyroid disease through my research. First, I selected a dataset with globally collected data, so that my learned model works well everywhere. Then I will work to accurately detect 2 symptoms of thyroid disease, perfectly detect the disease in quick time and give good performance. It is capable of early illness detection. This model will be able to detect 3 stages (hyperthyroid, hypothyroid, negative) of thyroid disorder.

- Build a best and effective model for detect thyroid problems
- One system could be detected different types of thyroid disease.



- It will work for any thyroid patient in any area of the world.
- Early-stage detection.
- It can be easily detected in a short time
- Reduces medical expenses.
- It will make the work of doctors easier.
- It will help the doctor to reach a decision quickly.

It will help the doctors, to detect thyroid disorders in a very short time, and easily, making the work of the doctors easier.

### **1.5 Expected Output**

I will try to do a few things with this model. Through my model, 3 problems of thyroid disease can be diagnosed (hyperthyroid, hypothyroid and normal). Also, this model can detect the disease in very short time. It can also accurately detect thyroid problems. And this model will work for any adult patient anywhere in the world.

- This model can detect three types of thyroid problems (hyperthyroid, hypothyroid and normal).
- It can detect the thyroid disease in very short time.
- It can detect thyroid problems accurately.
- This model will work every thyroid patient.

### **1.6 Research Questions**

- How effective is this model in diagnosing thyroid problems?
- Which machine learning model is best and effective in classifying thyroid disease?
- How much can this model help doctors in the diagnosis of thyroid disease?
- How is the accuracy rate of this model in the diagnosis of thyroid disease?
- Using this model, how will it affect medical expenses?

## **1.7 Report Layout**

Chapter 1: Represents introduction, motivation, problem statement, research objective, research question.

Chapter 2: Background, preliminaries, related work, comparative analysis & summary, challenges.

Chapter 3: Research methodology, data preprocessing, methodology, model, implementation requirements.

Chapter 4: Experimental results, discussion, experimental setup, result & analysis, final model.

Chapter 5: Impact on society, environment & sustainability.

Chapter 6: Summary, conclusion, recommendation and implication for future research.

References

## **CHAPTER 2**

### **Background**

#### **2.1 Preliminaries/Terminologies**

Here prioritized few topics in this research project. Diagnosing thyroid disease, identifying thyroid problems according to classification. Easy detection in less time. Diagnosing thyroid disease at an early stage.

In clinical practice, neck ultrasonography is a commonly used and effective diagnostic method for thyroid nodules, a common endocrine condition. As a result of localized aberrant thyroid cell development, thyroid nodules are a common indicator of endocrine diseases. Through this model, any patient problem can be identified. By training this model with different types of patient data, this model can easily diagnose the patient's disease by reviewing the patient data with the data of that dataset.

Also, this model can detect 3 types of thyroid problems (hyperthyroid, hypothyroid, negative). hyperthyroid, hyperthyroid is basically if a patient's thyroid gland produces more hormones, the condition is called hyperthyroid problem. hypothyroid, hypothyroid is when the patient's thyroid gland produces less hormones than required, the condition is called hypothyroid problem. Normal means if a thyroid gland is in normal condition, i.e., no problem is observed, that condition will be called normal.

Through this model, problems can be easily detected within short time. Consequently, a decision can be reached by analyzing the patient's condition quickly and treatment can be started quickly. By doing it through the model, the time and money needed to do various tests (Blood test, CT scan, ultrasonogram etc.) will be saved to a large extent. Treatment can be started as soon as possible through early detection. When thyroid abnormalities are treated early on, complications can be avoided and the effects on general health are reduced.

Also, with the help of this model, thyroid disorders can be detected in early stage. As a result, the patient's health condition can be cured before it deteriorates further. Organs can be protected from permanent damage. Health risks can be minimized. If thyroid diseases are not treated, it can result in a number of health concerns, including osteoporosis, heart

problems, infertility, and occasionally thyroid cancer. Starting treatment on time can reduce medical expenses.

## **2.2 Related Works**

This study tries to predict thyroid disease and try to get best accuracy by using different machine learning algorithms. Dataset collected from UCI. They used (Decision tree, Random Forest, LightGBM, XGBoost, GaussianNB, KNN, ANN, SVC, CatBoost, Extra-Trees) algorithms. Accuracy is 0.958%. Worked on a particular dataset. [6]

There aims to categorize this disease in few types like hyperthyroidism, hypothyroidism and normal. The data was taken from Iraqi citizens. Data was taken on around 1250 different people, both females and males, age group ranging from one year to one year. Used (SVM, LR, MLP, DT, NB, RF, KNN) algorithms. Accuracy is 98.98%. Worked on small dataset. [7]

Used different types of machine learning algorithm to predict hyperthyroidism, hypothyroidism. Dataset collected from UCI. They used (DT, RF, GB, NB, LR, KNN, SVM) algorithms. They get accuracy 84.48%, precision score 25, recall score 24, F1 score 25. There accuracy is too low. [8]

They try get find out a classification-based machine learning system for detect euthyroidism, hyperthyroidism, and hypothyroidism. Dataset collected from two hospitals in Haryana from January 2020 to July 2020. Used (LR, DT, NB, RF, Proposed Algorithm) models. Accuracy is 94%, precision score 96, recall 97, F1 Score 96. Work on small dataset. [9]

In this paper try to categorize thyroid problems into three types: hypothyroidism, hyperthyroidism, normal. Data collected from Iraqi hospital. Used (SVM, RF, DT, MLP, NB, LDA, KNN, LR) models. They get Accuracy 96.4%. Work on particular area dataset. [10]

This study tries to risk assessment of thyroid nodules through text-classification method in neural-networks. Dataset collected from random hospital. They used many models- (Text CNN, LSTM Network Model, Convolutional Neural Network+bidirectional cyclic

network). Get best accuracy on Bidirectional (96.3%). They have huge information gape to understand this paper. [11]

In this paper they try to develop a model for predict thyroid disease in primary stage by using machine leaning algorithm. Dataset collected from UCI repository. They compared their model with (Neural Network, Random Forest (RF)) algorithms. They get best accuracy on random forest classification accuracy-96.16%. [12]

Their target is to classified the thyroid disease into three type- normal, hypothyroid, hyperthyroid. Collected data from Iraqi people. Used (Decision Tree, Support Vector Machine, Random Forest, Naïve Bayes, Logistic Regression, K-Nearest Neighbors, Linear Discriminant Analysis). Get best accuracy on decision tree, accuracy- 98.4%. It will work only for a specific area. [13]

In this work they try to categorize the thyroid disease properly and easily. They collected data from random 500 thyroid patients. Used four models (Naïve Bayes, Support Vector Machine, Decision Tree, K-Nearest Neighbors). They get best accuracy on decision tree, accuracy-97.35%. Work on small dataset and applied only four models. [14]

The target of this study is to classify the thyroid disease by using machine learning algorithms. Dataset collected from unknown source. They used (Random Forest). They best accuracy get from two models, XGBoost and Bagging models, accuracy – 95.86%. They are developing the model for run on the website and iOS platform in future. [15]

## 2.3 Comparative Analysis and Summary

TABLE 2.3: COMPARATIVE TABLE OF OTHER PAPER'S

Author Name	References	Year	Algorithms	Accuracy	Limitations
Islam, et al	[6]	2022	(DT, RF, LightGBM, XGBoost, GaussianNB, KNN, ANN, SVC, CatBoost, Extra-Trees.) Best Model- (ANN)	Accuracy (95.8), Precision (95.7), Recall (95.9), F1-score (95.7)	Worked on a particular dataset
Singh, et al	[7]	2022	(RF, NB, SVM, DT, LR, KNN, MLP) Best Model- (RF)	Accuracy (98.98)	Small amount of dataset.
Shama, et al	[8]	2022	(RF, DT, NB, KNN, SVM, LR, Gradient Boosting) Best Model- (Random Forest)	Accuracy (91.92), Precision (25), Recall (24), F1-Score (25).	Lack of model optimization
Verma, et al	[9]	2022	(LR, DT, NB, RF, Proposed Algorithm) Best Model- (Proposed Algorithm)	Accuracy (94), Precision (96), Recall (97), F1-Score (96)	Work on a particular area dataset.
Kumar, et al	[10]	2023	(DT, SVM, RF, LDA, LR, MLP) Best Model- (MLP)	Accuracy (96.4)	Work on a particular area dataset
Wen, et al	[11]	2023	Text CNN, LSTM Network Model, Convolutional Neural Network+bidirectional cyclic network. Best Model-	Accuracy (96.3)	Relatively less efficient model.

			(Bidirectional)		
Pan, et al	[12]	2016	(DT, RF, SVM, NB, LR) Best Model- (RF)	Accuracy (96.16)	Relatively less efficient model.
Sonuç, et al	[13]	2021	(DT, SVM, RF, NB, LR, KNN, LDA) Best Model- (Decision Tree)	Accuracy- (98.4)	Work for a specific area
Dharmarajan, et al	[14]	2020	NB, SVM, DT, KNN Best Model- (Decision Tree)	Accuracy- (97.35)	Work on small dataset and applied only four models
Viswanatha, V	[15]	2023	Random Forest	Accuracy (95.86)	Relatively less efficient model.

**Contribution:** So, in this research will try to overcome these limitations such as: Work on a standard dataset. Take a global dataset which is contains worldwide patient data. And also focus on model performance and accuracy. Try to build a best and effective model which will work for any thyroid patient in any area of the world. Also, able to detect thyroid problems in early stage.

## 2.4 Scope of the Problem

Various researchers have tried to diagnose thyroid disease in different ways, but they have not been completely successful. Their model has several limitations. For example: the model does not work properly, the accuracy of the model is low, it cannot work for patients of all ages. Some models work by location or on patients in a specific location. I will try to solve all kinds of problems in my research and develop a complete self-contained model, this model will work on any patient anywhere in the world and it will be able to accurately identify and classify thyroid problems in a fast time, this model can detect three types of thyroid problems (hyperthyroid affected patient, hypothyroid affected patient and normal people). And also, it will work on pregnant patient. My research paper will examine the range of thyroid illnesses, with an emphasis on their epidemiology, effects, diagnosis,

therapy, and wider ramifications. It enhances our collective knowledge of thyroid problems and provide guidance for better patient treatment and public health campaigns.

## **2.5 Challenges**

The first challenge to face is finding a good dataset. The dataset can be used to properly train the model. Most datasets have various problems such as missing values, incomplete data, small datasets, inaccurate data, unbalance datasets, low attributes, data types. Various types of problems have to be overcome in data preprocessing method. Such as missing data handle, outliers handle, data cleaning, inappropriate data format handle, data scaling, data transform, unbalance data etc. Feature selection is the most important part of any research. A model's performance depends on feature selection. If the feature selection is done well then, the model can be trained well and the model can perform accurately. If the feature selection is not done properly, consequently the model's precision will decrease. If all methods are not done correctly, then the accuracy will decrease. Starting from data pre-processing to model selection, if the work goes wrong somewhere in between, the accuracy will decrease if the process is not completed properly.



# CHAPTER 3

## Research Methodology

### 3.1 Research Subject and Instrumentation

**Subject:** Basically, I am researching on health sector. I tried to identify different types of thyroid diseases.

**Instruments:** To do this research I used Google Colab as software. Hardware setup used for colab – Processor 2vCPU @ 2.2GHz, System RAM 13 GB, Storage HDD 108 GB.

### 3.2 Data Collection Procedure/Dataset Utilized

I have taken 3 class rich dataset, because I will detect the disease in 3 steps. hyperthyroid, hypothyroid and normal i.e., if there is disease it will detect it and if there is no disease it will also detect it. The dataset was collected from the UCI- machine learning repository. This dataset has 31 attributes. It contains several kinds of data, including- age, thyroxine, TSH, T3, T4, FTI, TBG, tumor, goitre, hyperthyroid, hypothyroid, sick etc.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC		
1	age	sex	on_thyro_query	on_antith_sick	pregnant	thyroid_sl131	treat_query_hy	query_hy	lithium	goitre	tumor	hypopituit	psych	TSH_meas	TSH	T3_meas	T3	TT4_meas	TT4	T4U_meas	T4U	FTI_meas	FTI	TBG_meas	TBG	referral_status					
2	29	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.3	f	t	1.6	f	1.9	t	128	f	f	f	other	-		
3	29	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.6	t	1.9	t	128	f	f	f	f	f	f	other	-		
4	41	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	11	other	-	
5	36	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	26	other	-
6	32	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	36	other	-
7	60	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	26	other	-
8	77	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	21	other	-
9	28	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.7	t	2.6	t	116	f	f	f	f	f	f	SVI	-		
10	28	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.2	t	1.8	t	76	f	f	f	f	f	f	other	-		
11	28	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.9	t	1.7	t	83	f	f	f	f	f	f	other	-		
12	54	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.9	t	2.3	t	133	f	f	f	f	f	f	other	-		
13	42	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1	t	1.8	t	105	f	f	f	f	f	f	other	-		
14	51	M	t	f	f	f	f	f	f	f	f	f	f	f	f	f	0.5	f	t	f	f	f	f	f	f	f	f	other	-		
15	51	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.7	t	2.4	t	116	f	f	f	f	f	f	other	-		
16	37	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.9	f	f	f	f	f	f	f	f	f	other	-		
17	18	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	2.6	f	t	f	f	f	f	f	f	f	f	other	-		
18	54	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1	t	2	t	122	f	f	f	f	f	f	SVHC	-		
19	43	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	2.1	t	116	f	f	f	f	f	f	f	other	-		
20	63	F	t	f	f	t	f	f	f	f	f	f	f	f	f	f	88	f	t	48	t	1.02	t	47	f	f	other	F			
21	36	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.5	t	2.4	t	90	t	1.06	t	85	f	other	-			
22	40	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	f	t	79	t	0.94	t	84	f	other	-				
23	40	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	1.2	t	2.3	t	104	f	1.08	t	96	f	other	-			
24	40	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	5.9	t	2.1	t	88	t	0.84	t	105	f	other	-			
25	77	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	t	2.4	t	107	t	1.13	t	95	f	other	-			
26	77	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	4	t	2	t	126	f	f	f	f	f	SVHC	-			
27	77	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.6	f	t	113	t	1.07	t	106	f	other	-				
28	77	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.8	t	2.4	t	150	f	f	f	f	f	SVHC	-			
29	51	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	t	2.1	t	93	t	0.87	t	106	f	SVI	-			
30	75	F	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	t	1.6	t	157	t	0.89	t	176	f	SVI	AM			
31	56	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.2	t	1.6	t	80	t	0.62	t	129	f	SVI	-			
32	42	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	3	t	2	t	91	t	0.91	t	100	f	SVI	-			
33	85	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	t	0.1	t	47	t	0.48	t	69	f	SVI	-			
34	41	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	t	1.6	t	39	t	1	t	39	f	other	R			
35	71	F	t	f	f	f	f	f	f	f	f	f	f	f	f	f	0.05	f	t	126	t	1.38	t	91	f	other	I				
36	67	M	f	f	f	f	f	f	f	f	f	f	f	f	f	f	0.2	f	t	71	t	0.79	t	90	f	other	-				

Figure 3.2.1: CSV dataset.

#	Column	Non-Null	Count	Dtype
0	age	9172	non-null	int64
1	sex	8865	non-null	object
2	on_thyroxine	9172	non-null	object
3	query_on_thyroxine	9172	non-null	object
4	on_antithyroid_meds	9172	non-null	object
5	sick	9172	non-null	object
6	pregnant	9172	non-null	object
7	thyroid_surgery	9172	non-null	object
8	I131_treatment	9172	non-null	object
9	query_hypothyroid	9172	non-null	object
10	query_hyperthyroid	9172	non-null	object
11	lithium	9172	non-null	object
12	goitre	9172	non-null	object
13	tumor	9172	non-null	object
14	hypopituitary	9172	non-null	object
15	psych	9172	non-null	object
16	TSH_measured	9172	non-null	object
17	TSH	8330	non-null	float64
18	T3_measured	9172	non-null	object
19	T3	6568	non-null	float64
20	TT4_measured	9172	non-null	object
21	TT4	8730	non-null	float64
22	T4U_measured	9172	non-null	object
23	T4U	8363	non-null	float64
24	FTI_measured	9172	non-null	object
25	FTI	8370	non-null	float64
26	TBG_measured	9172	non-null	object
27	TBG	349	non-null	float64
28	referral_source	9172	non-null	object
29	target	7546	non-null	object
30	patient_id	9172	non-null	int64

dtypes: float64(6), int64(2), object(23)

Figure 3.2.2: Dataset feature list

So, from the chart we can see that, in this dataset have 31 different attributes included 1 target attribute, which are contains different types of data values, and also this are different data type.

### 3.2.1 ThyroidDF

This dataset contains a total of 9173 data, and a total of 31 attributes including 1 target attribute. Target column has 182 negative values, 593 hypothyroid values, 6767 hyperthyroid values.

TABLE 3.2.1: CLASS AND INSTANCES COUNTS FOR THYROIDDF

Class	Encoded Class	Values
Negative	0	182
Hypothyroid	1	593
Hyperthyroid	2	6767

### 3.3 Data Pre-Processing

Data pre-processing is most important part of any research. A well processed dataset, helps to better perform a model. The pre-processing method prepares the dataset for use in the algorithm. There are various problems in the dataset, such as: missing data, error data, outlier data, imbalance data, null value, unsupported data type. We can solve these problems through different methods. null value handling, outlier handling, data resampling, data normalization, data transform, scaling etc. Also feature selection is done for the model in preprocessing method.

Pre-Processing steps for this thesis:

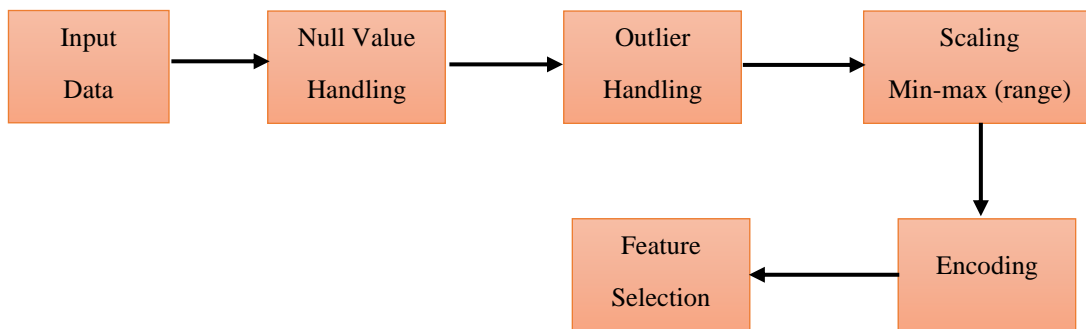


Fig: 3.3 Pre-Processing steps

### **3.3.1 Null Value Handling**

After taking the dataset input, first I checked the null value in the dataset and there were several null values in the dataset. I have adopted a couple of methods to handle null values. Some columns in the dataset had 80-90% data missing, in that case I removed those columns. After that, I have filled the null value in a column with the column's most value. Then I filled the missing value using median process of null value handling method in some columns.

### **3.3.2 Outlier Handling**

In this step I checked the outlier values in the dataset. A column called age was showing a lot of values i.e., max age was 65526 which is not possible at all. In that case I have set the maximum age limit in the range of 110 using the outlier handling method.

### **3.3.3 Scaling**

As thyroid diseases are associated with various hormones and these hormones have certain levels. By scaling method, I have brought the values of different hormones within the specified range. TSH hormone levels range from 0.5 to 5.0, but there are more in this dataset, so here I set the specified max and min values, bringing all values within the range. TT4 hormone has a value of 5-12, as usual I have brought it within the range. The value of T4U hormone is between 0.7-1.2, which is the usual range.

### **3.3.4 Encoding**

Converted the values of each attribute of the dataset to unique category integer values through the encoding method, except for the value of the target column. Usually, the value in the dataset is in the form of string value, I applied the encoding method to convert it to unique integer value. Integer values are needed to train the model because the machine can only read 0,1. So in case of any machine learning, encoding has to be done.

### **3.3.5 Label Encoding**

Convert target column value to unique integer value using label encoding. The target value will be used for check model learning performance. After train model, how model is working that is checked and compared with the test value, the target value will be taken as test data.

### 3.3.6 Feature Selection

Feature selection becomes a crucial and essential phase for machine learning. Model learning and model performance, accuracy depends on feature selection. If you can't select the right features, the model doesn't learn properly. The performance of the model decreases. Every column in this dataset is almost equally important so here I will work on all the features.

### 3.3.7 Train Test Split

After the feature selection is complete, dataset split into 2 parts: train, test. Model will be trained with train data. The model's learning performance will be evaluated using test data, it will be checked how well the model is working. Train data taken 75% and test data taken 25%.

### 3.3.8 Cross Validation:

Cross-validation is a machine learning method utilized to evaluate a model's performance on data that it has not yet been exposed to. The process entails partitioning the accessible data into numerous folds or subsets, with one of those folds serving as the validation set, Meanwhile, the remaining folds are used to train the model. This procedure is iterated numerous times, with a distinct fold being utilized as validation set in each iteration. By averaging the outcomes of each validation phase, a more precise estimation of the model's performance is generated. [16]

### Box Plot Graph:

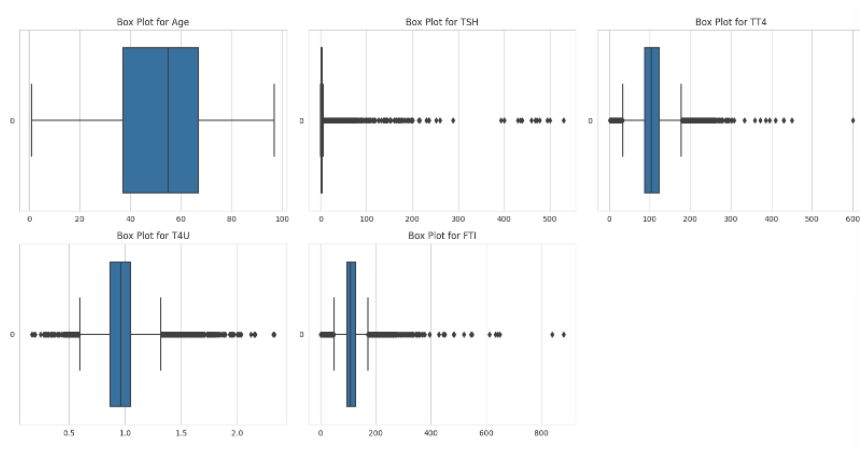


Figure 3.3.8.1: Box plot visualization of some feature after preprocessing.

**Seaborn plots:**

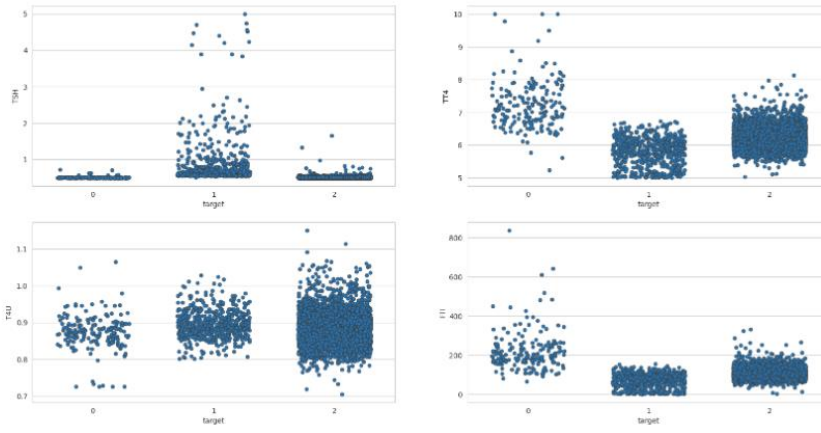


Figure 3.3.8.2: Sea born plots visualization of some feature after preprocessing.

### Correlation heat map:

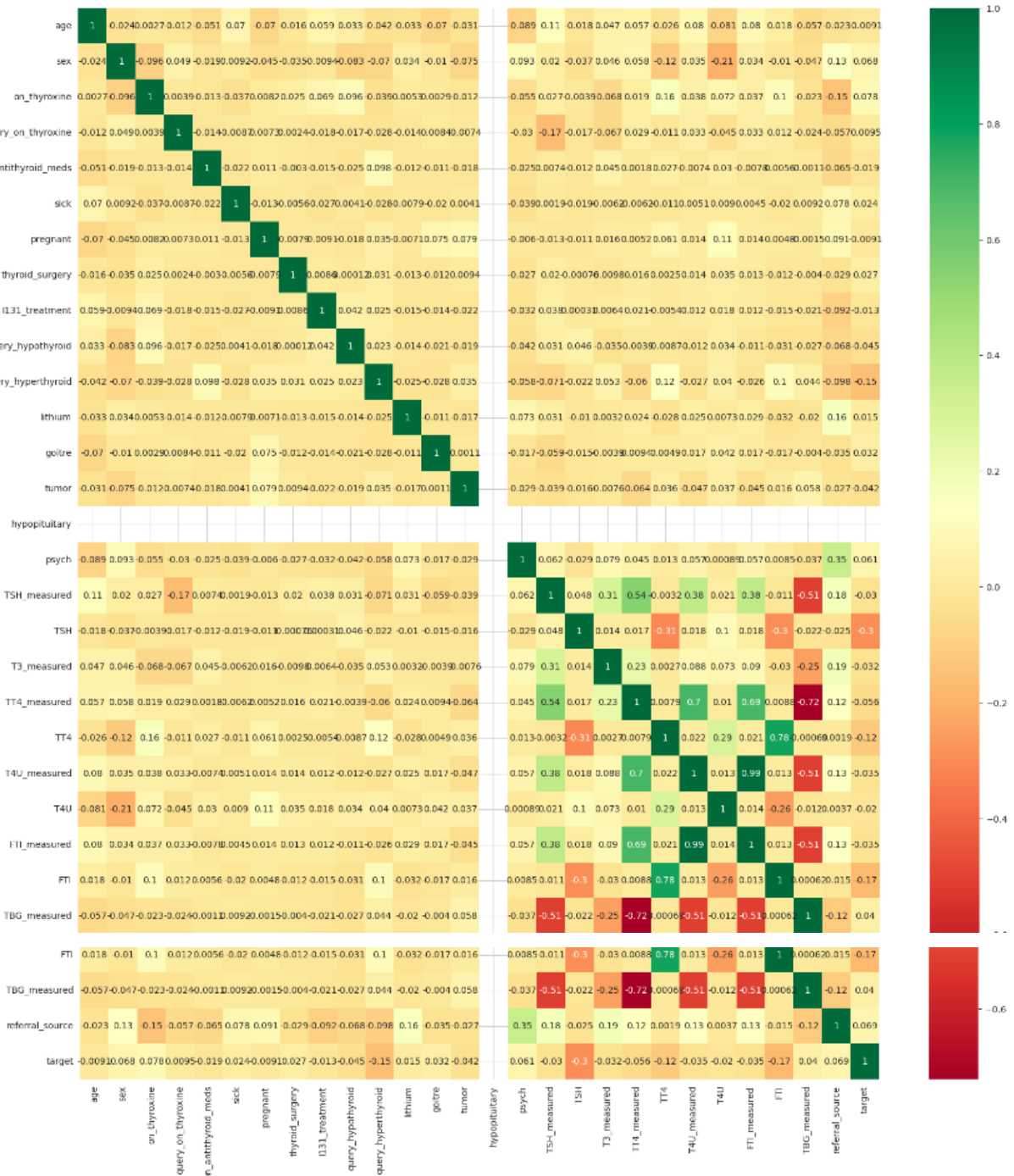


Figure 3.3.8.3: Correlation heat map visualization after preprocessing.

### 3.4 Proposed Methodology/Applied Mechanism

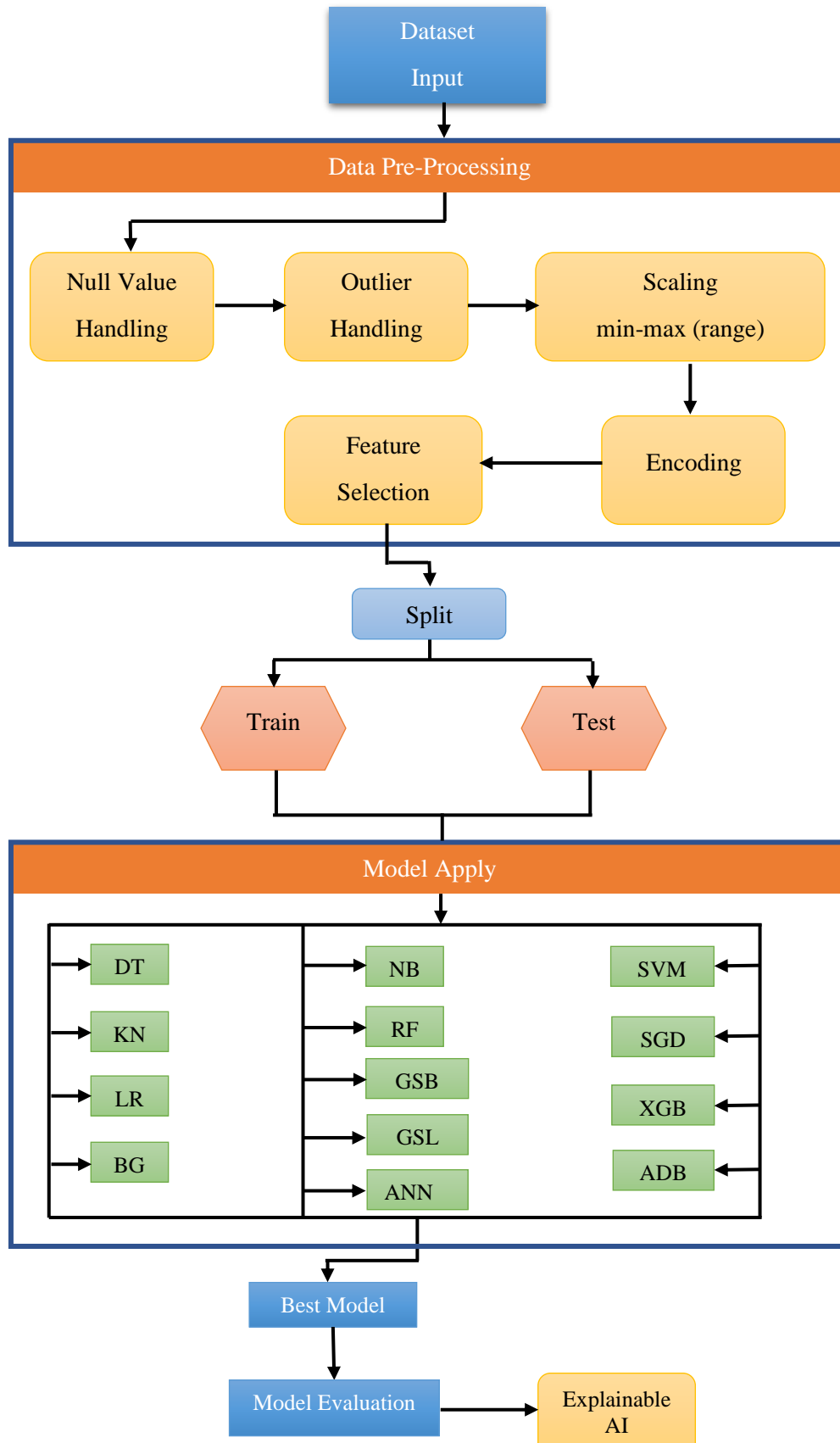




Figure 3.4: Flow chart of proposed methodology (ThyroidNet)

In the past, many researchers have worked differently through many models but none of the models could work perfectly. So, I will work with many models to get good results. In my research, I have applied about 13 models, among which the model that works best will be selected as the final model. The models are - Decision Tree, K-Neighbors, Logistic Regression, Bagging Classifier, AdaBoost Classifier, Bernoulli Naïve Bayes, Random Forest, Support Vector Machine, SGD Classifier, XGBoost, Grid Search Bagging, Grid Search Logistic, ANN.

### **3.5 Model Apply**

After data pre-processing and data train split test, I will start applying the ml model. I will apply 13 ml models in total, normal and ensemble, among which the best performing model will be selected as the final model.

#### **3.5.1 Decision Tree**

Decision Tree Algorithm is set up, trained on training data that is given, tested on both test data and the training data using accuracy metrics. A classification report is then generated [17]. Tried to predict thyroid disease using decision tree algorithm. It is currently one of the top algorithms. In case of over fitting dataset this model works very well, using tuning, this model can bring out good output. I extracted the result using the default parameters. I have trained the model through train data. And the 3 classes are predicted separately in the model. hyperthyroid, hypothyroid, normal or negative. For this model I applied the default parameter. Parameter details is given below- criterion: "gini", splitter: 'best', random\_state: "default", class\_weight: "default", min\_samples\_split: "2", min\_samples\_leaf: "1". The decision tree's construction and prediction process are influenced by each of these elements. The performance, overfitting, and generalization abilities of the model may all be significantly impacted by changing these parameters.

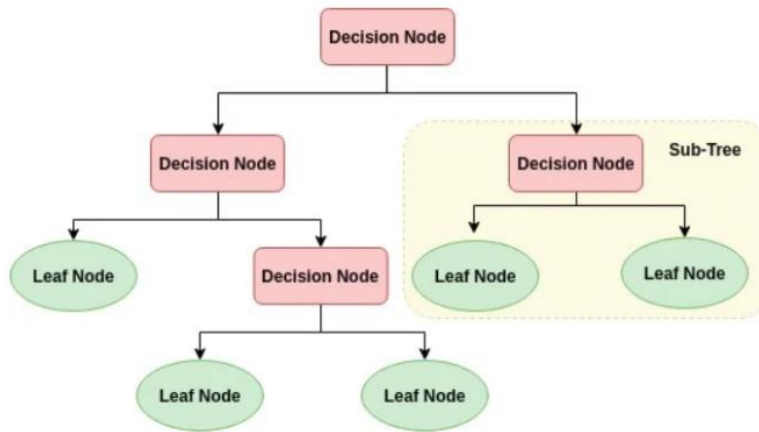


Figure 3.5.1: Decision Tree model design [18]

### 3.5.2 K-Nearest Neighbors

The supplied training data is used to train the KNN classifier. Through KNN algorithm I can predict thyroid disease and detect it according to classification. It is a classification-based model at the basic level of machine learning. It works mainly by predicting the result, depending on nearby data values. This algorithm works well for unbalanced datasets, selecting the closest data point depending on the distance. Then a specific range of values (K) is set, the data contained within it will fall into a class. For this model I adjusted the parameter as follows in the using all 30 features: I have set the parameter value in this model (n=5). As a result, the model will calculate the value of 5 values near the data and generate the result. Other parameter value is set to default value.

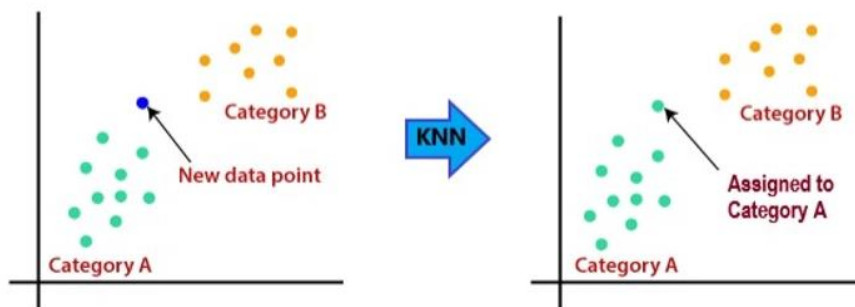


Figure 3.5.2: K-Nearest Neighbors model design [19]

### 3.5.3 Logistic Regression

The Logistic Regression used in order to forecast thyroid disease. When it comes to diagnosing illnesses, among the most used categorization schemes in the medical field is LR. Many optional parameters in the LR are essential to the operation of the model. The most important variables have been changed. LR settings adjusted as follows in the using all 30 features: {penalty: default='l2', fit\_intercept=True, max\_iter: int, default=100, class\_weight: default='none', solver: default='lbfgs', random\_state=default}. The parameter "random\_state" is used to regulate the level of randomness in the estimator. "Penalty" is used to establish the criterion for penalization. "Fit\_intercept" determines if a constant should be included in the decision function. "Solver" is a good tool for addressing optimization problems. [20]. RFE was adapted to the fair dataset in the second stage. Following that, LR was subjected to step 1 once again and its settings changed to match.

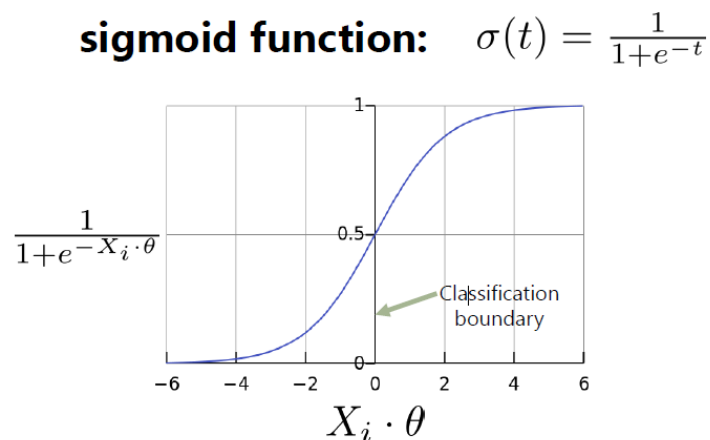


Figure 3.5.3: Logistic Regression model design [21]

### 3.5.4 Bagging

An ensemble machine learning method called a bagging classifier combines many classifier instances' predictions after they have been trained in using different portions of the training dataset in an attempt to enhance the efficiency of a base classifier. It works in several steps like: Reduced Overfitting and Variance, Hyperparameters, Parallelization, averaging, Bootstrap Sampling etc. All things considered, Bagging Classifier represents a strong ensemble method that raises the stability and precision of machine learning algorithms by harnessing the knowledge of many distinct models trained on various data subsets [22]. In this model, I used most of the default parameter. Only one parameter

customized ( $n\_estimators = 12$ ). After using this parameter, the model will calculate 12 estimators at a time.

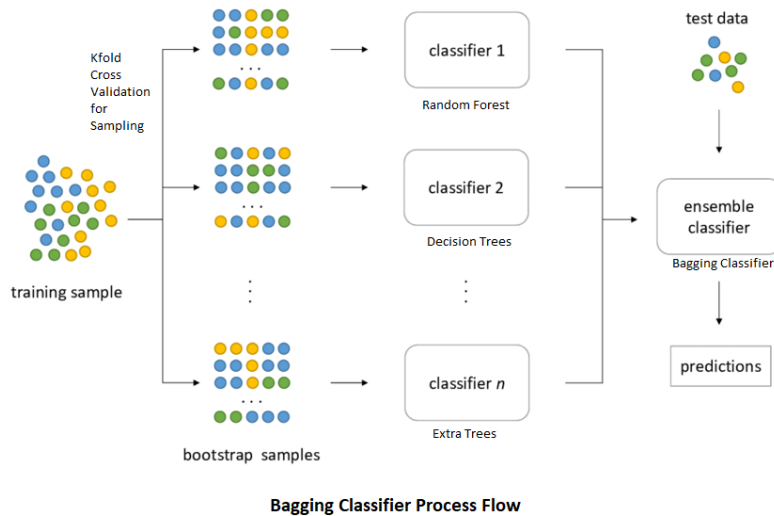


Figure 3.5.4: Bagging model design [23]

### 3.5.5 AdaBoost

Adaptive Boosting, or Adaboost, is a shorthand for an ensemble learning technique used in regression and classification. In order to produce a strong classifier, it combines many "weak" learners, which are usually decision trees or deep classifiers. The system trains each weak learner in turn, giving greater weight to incorrectly categorized data points at each iteration so that following weak learners may concentrate more on the challenging situations [24]. The characteristics or variables utilized to train those who are weak are referred to as features in Adaboost. The dataset and the issue being solved determine which particular features are used in an Adaboost model. These features might be any number of the data's qualities or traits, including: numerical feature, textual feature, categorical feature, image feature. It also tries to predict the target value through various other features. For this model I worked in 2 steps, in the first step I trained the model with default parameters and tried to classify thyroid diseases. In the 2nd step, I used hyper tune mode separately in this model, i.e., tried to get the best performance out of the model by auto-tuning all the parameters.

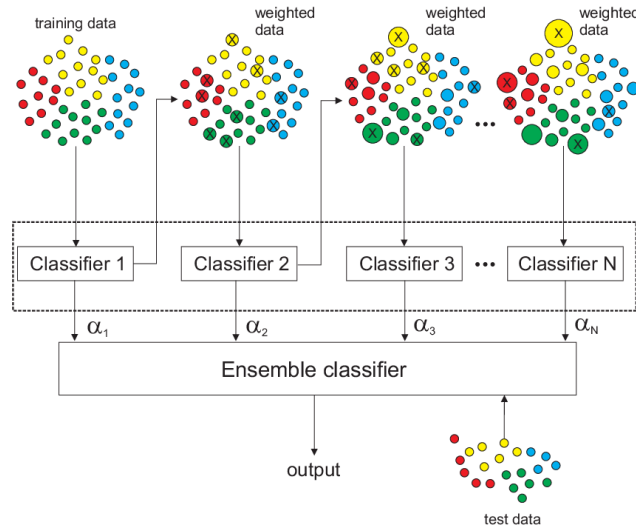


Figure 3.5.5: AdaBoost model design [25]

### 3.5.6 Bernoulli Naïve Bayes

Bernoulli Naïve Bayes classification technique is effective when applied to binary data and is especially made for characteristics that have a binary value (i.e., exist or do not exist). It is predicated on the "naive" presumption of feature independence in Bayes' theorem. This model has various features such as: binary feature denoting the existence or nonexistence of a specific attribute or characteristic and it shows (0,1). Feature Independence Given the class identifier, it assumes that the characteristics are conditionally independent. Although this assumption may not universally apply to all datasets, it frequently streamlines computations and proves to be effective in practical applications. The model works in several phases (training phase, prediction phase, evaluation phase). In the first step, the algorithm preprocesses data obtained from dataset and converts data according to the model's convenience. The next step is to predict the probability by selecting the model class. And the third step shows the result. Naive Bayes is highly computational and often performs well for classification of text problems (e.g., sentiment evaluation, spam detection) using binary characteristics (existence or nonexistence of words) and when the independence assumption holds pretty well. However, if an independent assumption is strongly broken or features are not correctly preprocessed, its performance may suffer [26].

In this research, at first input the balanced dataset into this model to strengthen the thyroid disease. Then I trained the model using the default parameters. RFE was implemented on

the balanced dataset during the second phase. Subsequently, step 1 was reapplied to NB and fine-tuned using the same settings.

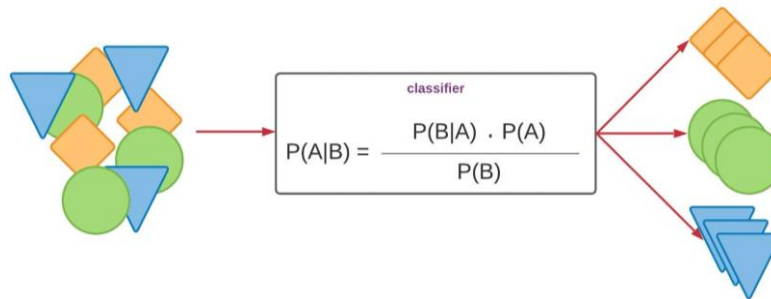


Figure 3.5.6: Naive Bayes model design [27]

### 3.5.7 Random Forest

As an ensemble model, Random Forest is a machine learning approach that uses a collection of decision trees to address regression and classification issues. Based on the principle of decision trees, the algorithm trains by creating a number of decision trees. It then determines the class that appears most often among the trees or calculates the average prediction of each of the trees. Features of the model - Ensemble Method, Bagging, Feature Randomness, Voting or Averaging etc. Features of the model are Ensemble Method, Bagging, Feature Randomness, Voting or Averaging etc. Working Procedure of this model (Bootstrapping, Feature Selection, Tree Building, Voting, Averaging, Final Prediction) etc. [28]

In this research purpose I used the default settings for RF model parameter and also used hyper tune mode for custom parameter. The variable `max_depth` represents the maximum depth of the tree. The term "`n_estimators`" refers to the total quantity of trees on the forest, whereas "`criteria`" is a factor that quantifies the significance of a split.

When splitting an internal node, the `min_samples_split` argument indicates the bare minimum of samples needed. An instance of "`min_samples_leaf`" is a leaf node that requires the bare minimum of samples. To be included for evaluation, a split position at any level must provide a minimum leaf training sample number for each branch. This might lead to a more seamless model, especially in the context of regression analysis. The

random\_state parameter controls the level of randomness in the bootstrapping samples used for tree construction.

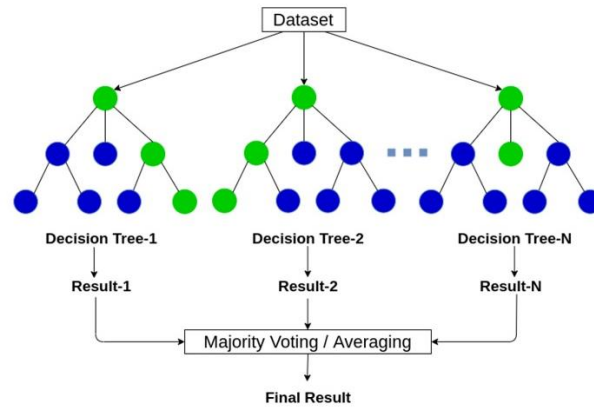


Figure 3.5.7: Random Forest model design [29]

### 3.5.8 Support Vector Machine

The SVM algorithm was applied for the purpose of thyroid disease prediction. It ranks high among the most effective classification models that are available today [20]. Support Vector Machines (SVM) has many optional parameters. Initially, the balanced dataset was subjected to SVM using all available features. For this model set the parameter such as: gamma='auto', probability='true', random\_state='42', other parameter set to default. Probability refers to the decision of whether or not to utilize probability estimates, whereas random state is employed to control the generation of random numbers to the purpose of combining data for likelihood predictions [20].

### 3.5.9 Stochastic Gradient Descent

Stochastic Gradient Descent, or SGD for short, that is popular optimization technique for deep learning and machine learning applications for training models, particularly in situations requiring huge datasets. The goal of this iterative optimization technique is to identify the ideal parameters for a particular machine learning model by minimizing the loss function. Stochastic Gradient Descent (SGD) models are characterized less by conventional data characteristics used in machine learning and more by their functionality,

setup, and parameters. An SGD model mostly consists of the following elements: optimization, stochasticity, iterative updating, learning rate, batch variants, parameter initialization, regularization techniques, parallelization etc.

At first inserted the balanced dataset and converted into suitable format for this model. After that I set the {max\_iter parameter value '1000'} and other parameter value is set to default for this model.

The procedure starts by setting the model's parameters, which include weights and biases, to zeros or random values. Calculating the loss function establishes the model's correctness with respect to the input data. Utilizing the loss function is one approach to quantify the disparity between anticipated and actual outcomes. SGD determines the loss function's gradient with regard to every model parameter. A function's gradient indicates the direction of its sharpest ascent.

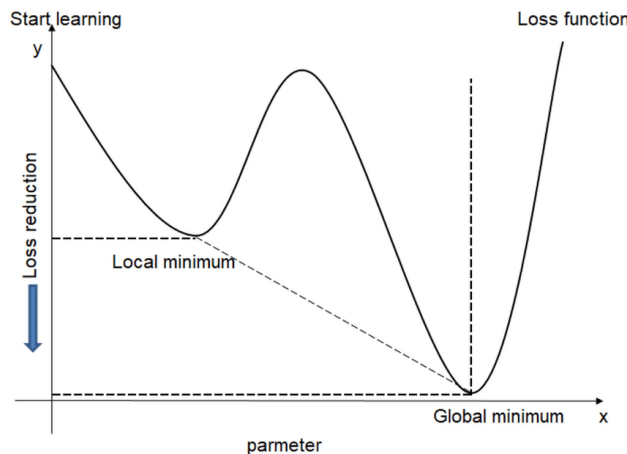


Figure 3.5.9: SGD model design [30]

### 3.5.10 XGBoost

XGBoost has shown remarkable efficacy as a boosting technique for tree models. The foundation of XGBoost is the gradient-boosted machine architecture proposed by (Friedman) [20]. To train the model, we utilized the same dataset and the XGBRegressor model from Python's XGBoost library. The first phase was specifying the XGBoost settings, including all available features. For this model only customized the n\_estimators' parameter which is set to (n=12), other's parameter is set to default value. The term "max\_depth" refers to the maximum depth of a tree, which limits the number of nodes in



the tree. The learning rate diminishes the impact of each individual tree. `Reg_lambda` refers to the L2 regularization term used to regularize weights. By augmenting this quantity, the model will exhibit a greater degree of caution. The user has the ability to include many evaluation metrics by using the `eval_metric` attribute. Python users should be mindful of providing metrics as a set of parameter pairs instead of a map. This ensures that the following `eval_metric` does not replace the previous one [20].

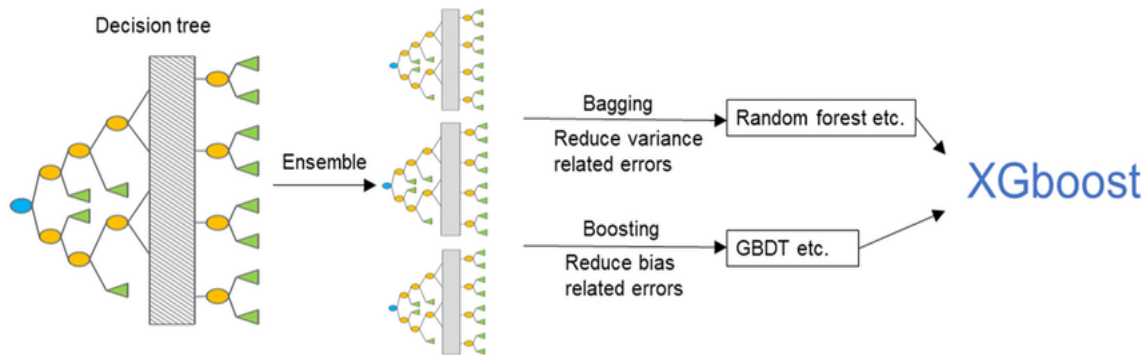


Figure 3.5.10: XGBoost model design [31]

### 3.5.11 Grid Search Bagging

Grid search is a method for determining the optimal hyperparameter combination in a machine learning algorithm. Bagging, which is an abbreviation to aggregating bootstraps, is an ensemble method utilized to enhance the precision and stability of machine learning models. It has many features such as: hyperparameter optimization, ensemble stability, performance enhancement, cross validation, generalization, model selection, evaluation etc. When you combine grid search with bagging, you are effectively attempting to optimize each of the hyperparameters in a bagging algorithm by doing an exhaustive search over a grid in hyperparameter values. This is done in order to find the best possible solution. Process of this model is understanding bagging, grid search, grid search with bagging, implementation and selection. [32]

In this model I used custom parameter for get best result. The parameter I speeched: {n\_estimator='2-16', for final bagg\_classifier n\_estimator=12, random\_state=22}. Other's parameter is set to default settings.

In essence, the integration of Grid Search with Bagging allows for a methodical examination of hyperparameters inside the ensemble framework, with the goal of enhancing the accuracy, balance, and generalization capabilities of the model. This methodology facilitates the construction of resilient and high-achieving machine learning algorithms for diverse tasks.

### **3.5.12 Grid Search Logistic**

The hyperparameter optimization method known as "grid search" implemented to identify optimal hyperparameter combination for machine learning algorithm. Grid search, in the context of logistic regression models, aids in the determination of optimal values for configuration options that impact the performance of the model, including normalization strength, penalizing type, and other parameters. Features of grid search algorithm- binary classification, probability, linear decision boundary, regularization, hyperparameter tuning, cross validation, scalability, evaluation etc.

For this model first input balanced dataset, then data converted into model suitable format. After that adjusted the parameter's value for get best result from the model. Parameter specified: {max\_iter=10000, random\_state=42, params:'C=2'}, other's parameter settings is set to default.

The objective of practitioners who integrate grid search & logistic regression to enhance the performance of model through a methodical investigation of different hyperparameter combinations. The optimal set of hyperparameters is then determined by selecting an evaluation metric.

### **3.5.13 Artificial Neural Network**

A class of machine learning models known as Artificial Neural Networks (ANNs) takes its cues from the complex network architecture seen in the human brain. Artificial neural networks (ANNs) are composed of interconnected nodes, also known as "neurons," which

are organised in the following layers: input, one or more concealed, and output. The input layer receives data, which is then processed via the hidden layers of the network using connections with weights and activation functions. Finally, an output is generated in the final layer. During the training phase, artificial neural networks (ANNs) modify the amount of weight of connections among neurons in order to minimize the discrepancy between expected and actual outcomes. This adjustment process is known as "backpropagation" and is helped by optimization techniques such as gradient descent. Artificial neural networks (ANNs) has the ability to acquire intricate patterns and correlations from data, making them valuable for a multitude of tasks such as regression, classification, pattern recognition, and others. Nevertheless, these models need significant quantities of data, fine-tuning, and computational resources, and may be susceptible to overfitting or lacking interpretability in certain scenarios. [33]

For this model at first inserted the balanced dataset, then converted data into model suitable format. After that model applied scaling to normalize the data into model required range. For define the ANN model, parameter specified: {dense (64, activation='relu'), drop=0.5, dense (32, activation='relu'), dropout=0.5, dense (3, activation='softmax'), epoch=50, batch size=32}. The architecture of this model such as: input layer- dense layer (neurons 64), dropout layer (dropout rate 50%), hidden layer- dense layer (neurons 32), dropout layer (dropout rate 50%), output layer- dense layer (neurons 3).

Artificial neural networks (ANNs) contain the capacity to acquire intricate patterns and correlations in data, rendering them adaptable for diverse undertakings such as picture and audio identification, natural language comprehension, predictive analysis, and other applications. Nevertheless, the effectiveness of their performance is greatly dependent on the proper design of architecture, the quality of data, and the tuning of parameters. [34]

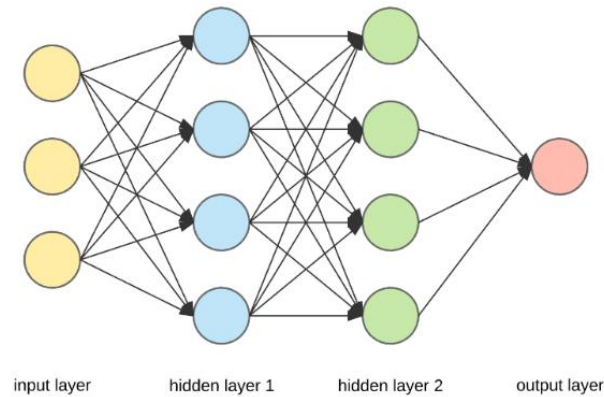


Figure 3.5.13: ANN model design [35]

### 3.5.14 Explainable AI

The development of AI systems capable of providing transparent explanations for their decisions and behaviors is referred to as "Explainable AI" (XAI). It attempts to help people understand the logic behind AI models, especially end users, stakeholders, and subject matter experts who may not be trained in data science or machine learning. Explainable AI uses a multidisciplinary strategy that integrates knowledge and skills from cognitive psychology, ethics, machine learning, and human-computer interaction. Practitioners and researchers collaborate to develop XAI techniques that strike a balance between precision, complexity, and comprehensibility. This encompasses the development of visualization tools to depict model behavior, the creation of algorithms that deliver significant insights while maintaining model performance, and the execution of user studies to assess the efficacy of explanations offered by AI systems. [36]

### 3.5.15 Lime

Lime, which stands for Local Interpretable Model-agnostic Explanations, is a widely used method in the Explainable AI (XAI) domain. Its primary objective is to provide interpretable explanation for individual predictions generated by intricate machine learning models. The objective is to improve the clarity and comprehension of AI forecasts on a regional scale. Lime has become popular because of its versatility, enabling users to get insights into predictions made by black-box models at a specific level. AI predictions on particular instances have been successfully used in diverse fields such as healthcare,

banking, and natural language processing [48]. This application has allowed practitioners & domain experts to comprehend and have confidence in the accuracy of these predictions, even when using intricate models.[36]

### **3.5.16 Shap**

In general, the term "shape" of Explainable AI (XAI) pertains to the methodological and structural framework it employs to ensure transparency and interpretability in its models of artificial intelligence [49]. Although no particular method or concept is referred to as a "shape of XAI." In brief, XAI is characterized by a comprehensive methodology that incorporates human-centered design, interpretability techniques, computational transparency, ethics, and ongoing improvement. Its objective is enhance the comprehensibility, reliability, and accountability of AI systems towards end-users and stakeholders.[36]

### **3.5.17 Explainable AI's role in this research**

After selecting the final model, added Explainable AI to the final model. Explainable AI helps people understand the results obtained from machines. It is very difficult for humans to understand the data or results obtained from the machine, because the machine uses high level language. explainable simply represents the data received from this machine. Explains what the machine is based on, on what data it is giving this result. In the case of this research, explain what the model is giving the prediction result based on, what feature the model is predicting this result based on. So, can say that explainable AI plays a very important role for research.

## **3.6 Evaluation Metrics**

The primary objective of machine learning is to construct a model that can effectively forecast outcomes based on freshly acquired data. Evaluation metrics allow us to evaluate the model's success in accomplishing this goal. By use metrics to assess the efficacy of a model, we may identify specific areas of mistake and make modifications that will improve the model's effectiveness and precision. We assessed the effectiveness of our proposed technique for intrusion detection utilizing a variety of evaluation metrics [37].

		Predicted	
		Negative (N) -	Positive (P) +
Actual	Negative -	True Negative (TN)	False Positive (FP) Type I Error
	Positive +	False Negative (FN) Type II Error	True Positive (TP)

Figure 3.6: Confusion matrix [38]

**Equation 1:**

**Accuracy:** It is an important method for evaluating the proportion of accurate classification predictions. Formula of accuracy:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)}$$

**Equation 2:**

**Precision:** It is an important method for evaluating the accurate positive result. Formula of precision:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

**Equation 3:**

**Recall:** It is an important method for evaluating the correctly predicted true positive rate. Formula of recall:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

**Equation 4:**

**F1-Score:** It Calculates the average value of precision and recall. Formula of f1-score:

$$\text{F1 - Score} = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

**Equation 5:**

**False Positive Rate (FPR):** The ratio of false positive is calculated using this formula:

$$\text{FPR} = \frac{\text{FP}}{(\text{FP}+\text{TN})}$$

**3.7 Implementation Requirements**

We have several requirements to implement this research. For example: to understand why this disease is happening, diagnostic advancements, treatment innovation, impact on public health, quality of life improvement, exploration of unknowns, personalized medicine, reducing economic burden etc. Research on thyroid illnesses is crucial for advancing our understanding, enhancing diagnosis and therapeutic approaches, mitigating the consequences on people and society, and facilitating the development of more effective management and preventive initiatives.

## CHAPTER 4

### Experimental Results and Discussion

#### 4.1 Experimental Setup

First, I took a dataset of thyroid disorders, from UCI-machine learning repository. This dataset contains data of 30 features and a target value. To do this research I used Google Colab as software. Hardware setup used for colab – Processor 2vCPU @ 2.2GHz, System RAM 13 GB, Storage HDD 108 GB.

#### 4.2 Results & Analysis

I have applied about 13 models (Decision tree, KNN, LR, bagging classifier, Adaboost, Bernoulli Naïve Bayes, RF, SVM, SGD, XGBoost, GSB, GSL, ANN) in my research, to make the research better and more effective. Among these, the model that performed best, got good accuracy, was selected as the final and best model. All the model confusion matrix, ROC curve and results is given below.

For this thesis, I applied cross-validation to the final model. I have divided the train data into 5-fold. To check train data accuracy. To check whether the dataset is overfitted. If there is a significant disparity in accuracy between the training data and the testing data, it might be inferred that the dataset was overfitted. So, I checked that through cross validation. And the accuracy in cross validation is (98.80%, 98.67%, 98.20%, 98.01%) avg accuracy- 98.40%.



TABLE 4.2: SHOWED ALL-MODEL RESULT

Model	Accuracy	Class	Precision	Recall	F1-score
Decision Tree	98.01%	0	77%	59%	67%
		1	89%	96%	92%
		2	99%	99%	99%
K-Neighbors Classifier	91.12%	0	58%	41%	48%
		1	63%	39%	48%
		2	94%	97%	96%
Logistic Regression	91.52%	0	50%	24%	32%
		1	83%	45%	58%
		2	94%	99%	96%
Bagging Classifier	99.09%	0	71%	59%	65%
		1	88%	100%	93%
		2	99%	98%	99%
AdaBoost Classifier	97.50%	0	62%	65%	56%
		1	94%	93%	89%
		2	99%	97%	98%
Bernoulli Naïve Bayes	88.34%	0	0%	0%	0%
		1	0%	0%	0%
		2	90%	100%	95%
Random Forest	98.67%	0	89%	59%	69%
		1	93%	98%	92%
		2	99%	99%	99%
Support Vector Machine	90.33%	0	62%	29%	40%
		1	76%	34%	47%
		2	93%	99%	96%
Stochastic Gradient Descent	89.13%	0	64%	41%	50%
		1	49%	32%	39%
		2	93%	97%	95%
XGBoost Classifier	98.78%	0	97%	59%	69%
		1	93%	98%	92%
		2	99%	98%	99%
Grid Search Bagging Classifier	98.54%	0	83%	59%	69%
		1	87%	98%	92%
		2	99%	99%	99%
Grid Search Logistic Classifier	94.03%	0	86%	35%	50%
		1	88%	66%	76%
		2	96%	99%	97%
Artificial Neural Network	97%	97%		97%	97%

**4.2.1 Best & Final Model:** Reviewing the results from the 13 models, I found that the bagging classifier model in my research performed the best. Using bagging classifier model, I got the results - Accuracy 99.04%, Precision 99%, Recall 99%, F1-score 99%. So, I will select this model as the best and final model.

## 4.2.2 Decision Tree

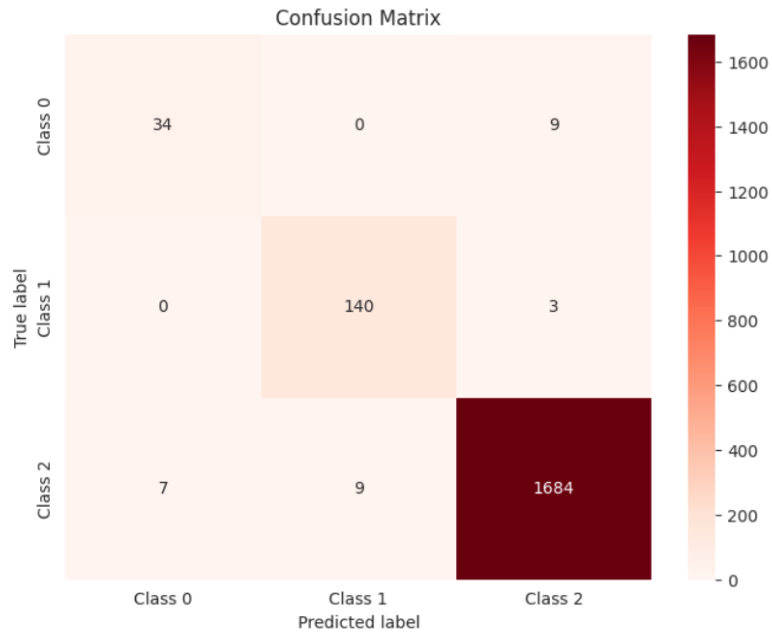


Figure 4.2.2.1: Confusion matrix (DT)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1858 and incorrectly predicted 28.

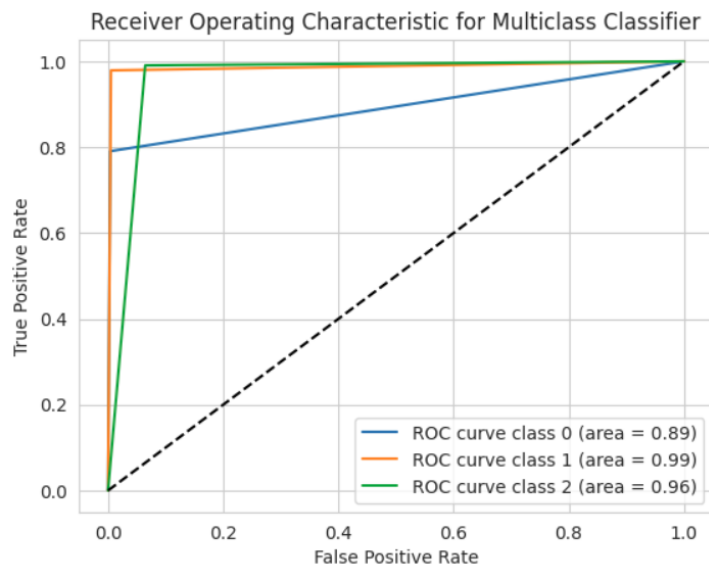


Figure 4.2.2.2: ROC curve (DT)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 89% and false positive rate 11%, and class 1, true positive rate 99% and false positive rate 1%, class 2, true positive rate 96% and false positive rate 4%).

### 4.2.3 K-Nearest Neighbors

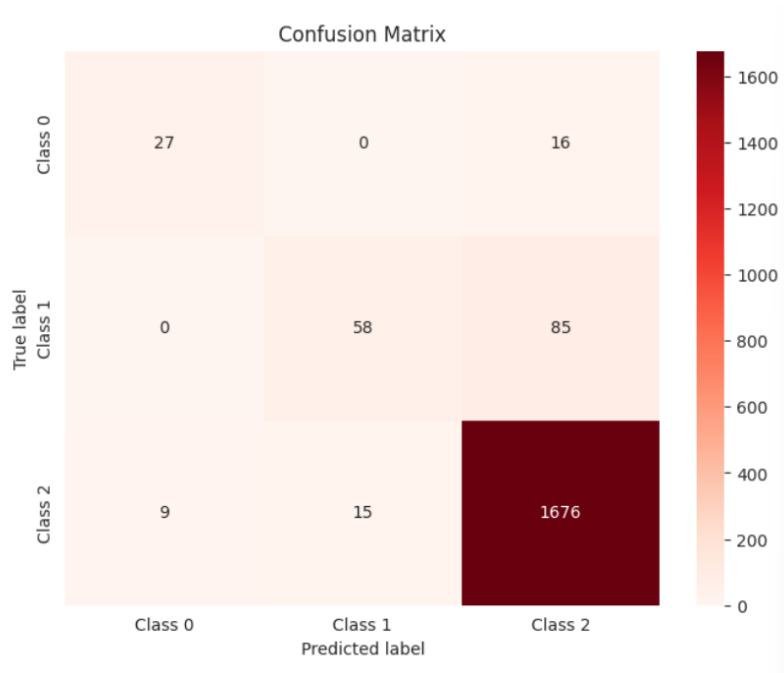


Figure 4.2.3.1: Confusion matrix (KNN)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1761 and incorrectly predicted 125.

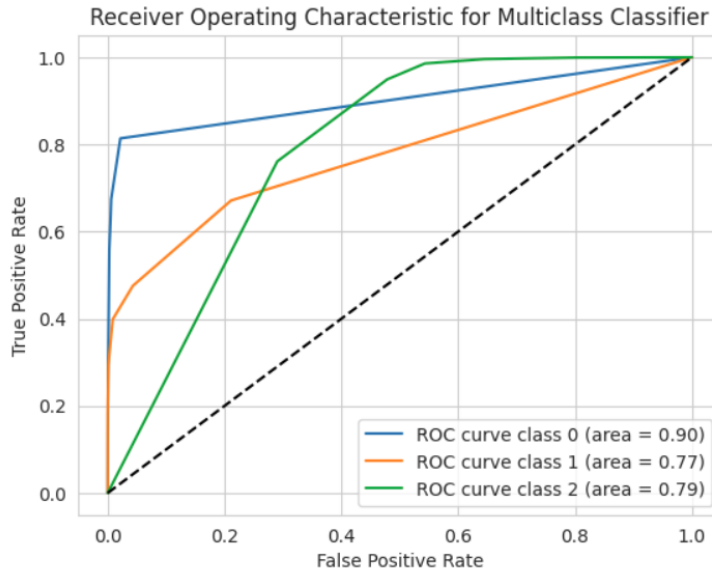


Figure 4.2.3.2: ROC curve (KNN)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 90% and false positive rate 10%, and class 1, true positive rate 77% and false positive rate 23%, class 2, true positive rate 79% and false positive rate 21%).

#### 4.2.4 Logistic Regression

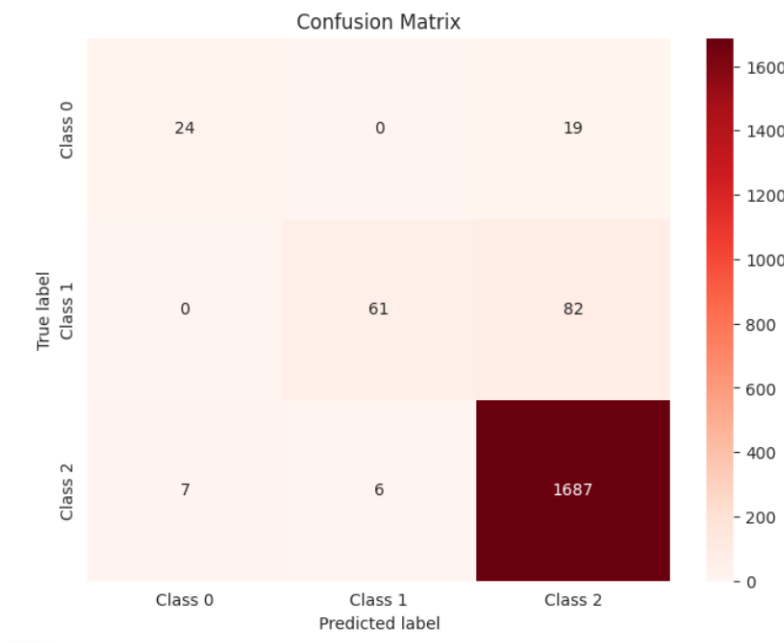


Figure 4.2.4.1: Confusion matrix (LR)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1952 and incorrectly predicted 114.

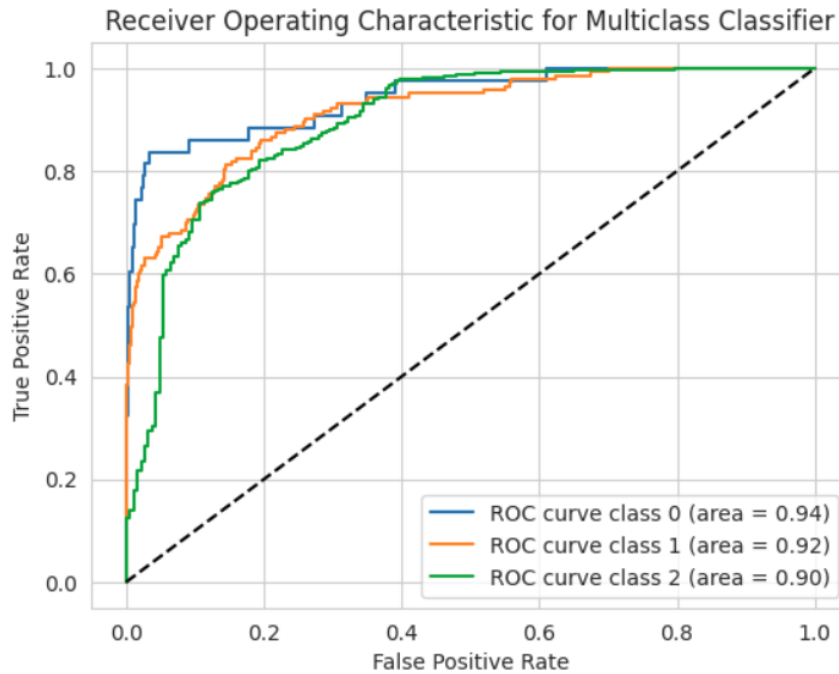


Figure 4.2.4.2: ROC curve (LR)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 94% and false positive rate 6%, and class 1, true positive rate 92% and false positive rate 8%, class 2, true positive rate 90% and false positive rate 10%).

## 4.2.5 Bagging

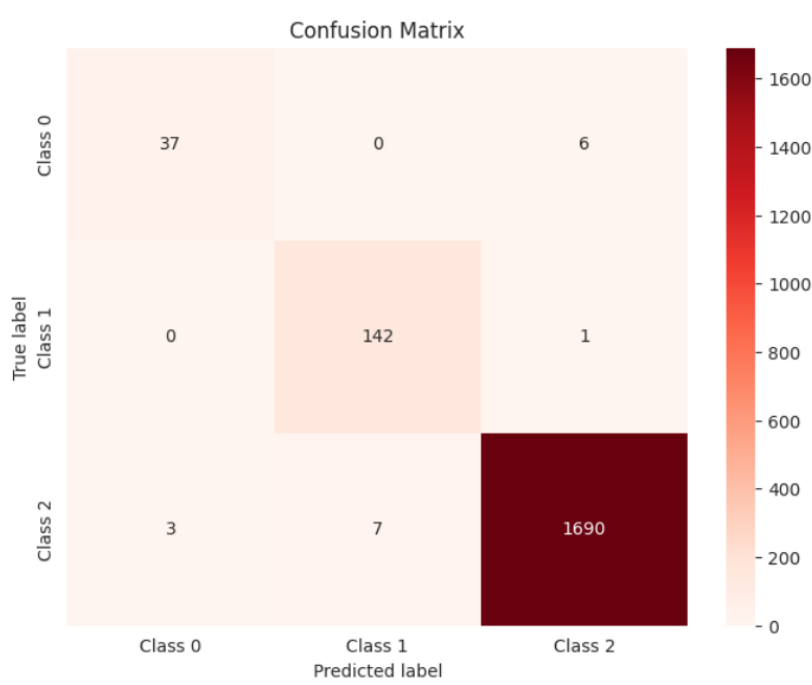


Figure 4.2.5.1: Confusion matrix (BG)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1869 and incorrectly predicted 17.

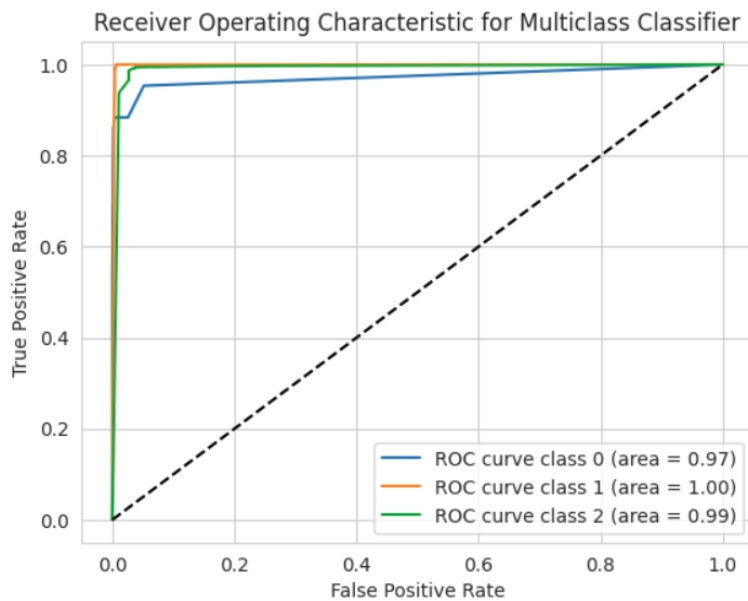


Figure 4.2.5.2: ROC curve (BG)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 97% and false positive rate 3%, and class 1, true positive rate 100% and false positive rate 0%, class 2, true positive rate 99% and false positive rate 1%).

#### 4.2.6 AdaBoost

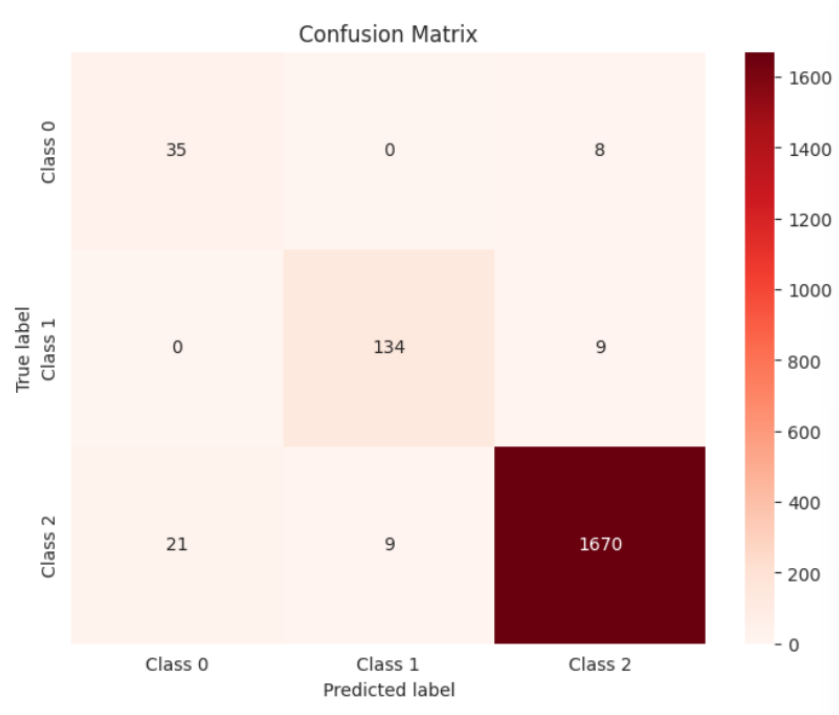


Figure 4.2.6.1: Confusion matrix (ADA)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1839 and incorrectly predicted 47.

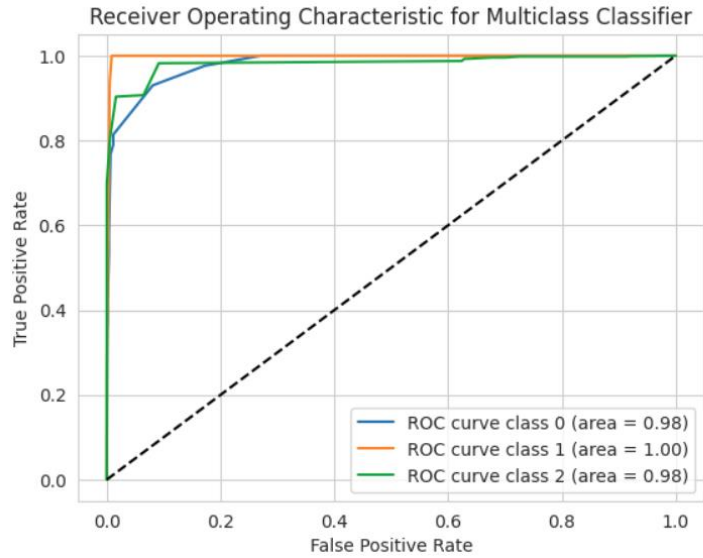


Figure 4.2.6.2: ROC curve (ADA)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 98% and false positive rate 2%, and class 1, true positive rate 100% and false positive rate 0%, class 2, true positive rate 98% and false positive rate 2%).

#### 4.2.7 Bernoulli Naive Bayes

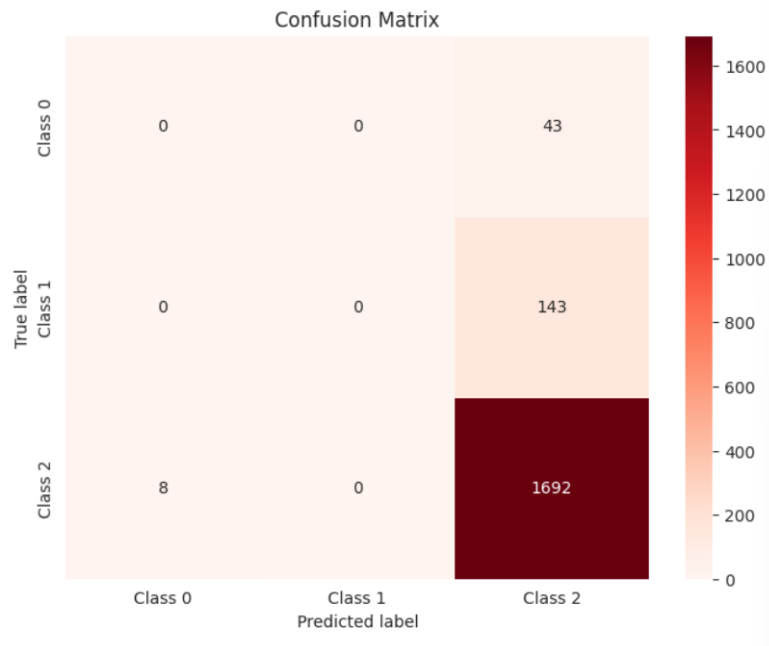


Figure 4.2.7.1: Confusion matrix (NB)



In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1692 and incorrectly predicted 194.

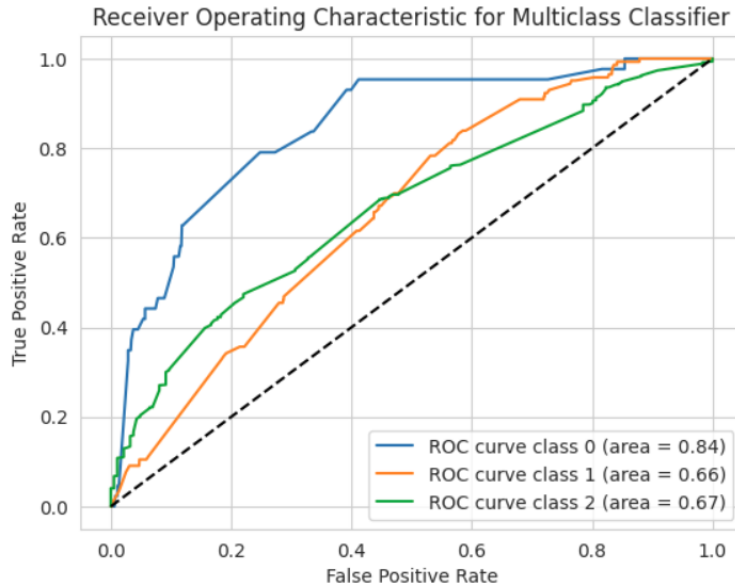


Figure 4.2.7.2: ROC curve (NB)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 84% and false positive rate 16%, and class 1, true positive rate 66% and false positive rate 34%, class 2, true positive rate 67% and false positive rate 33%).

## 4.2.8 Random Forest

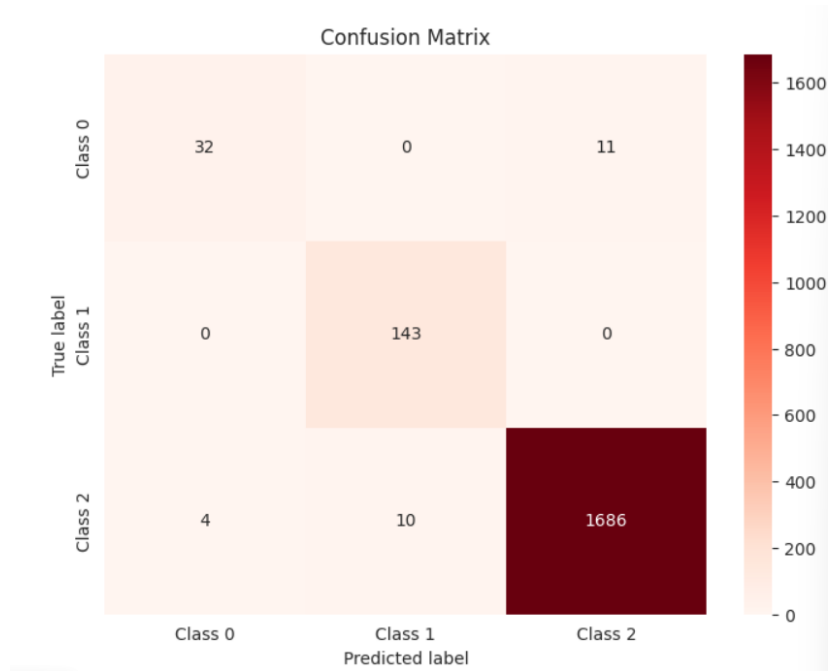


Figure 4.2.8.1: Confusion matrix (RF)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1861 and incorrectly predicted 25.

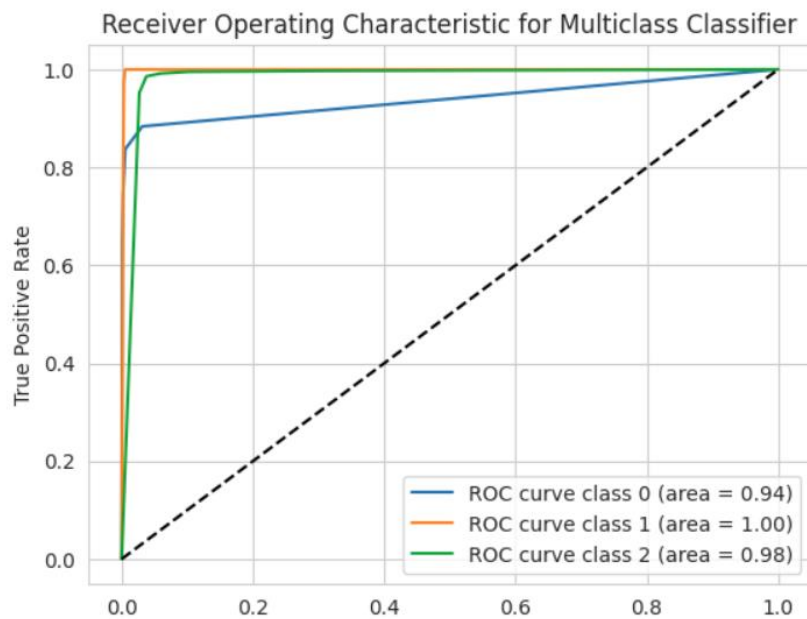


Figure 4.2.8.2: ROC curve (RF)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 94% and false positive rate 6%, and class 1, true positive rate 100% and false positive rate 0%, class 2, true positive rate 98% and false positive rate 2%).

#### 4.2.9 Support Vector Machine

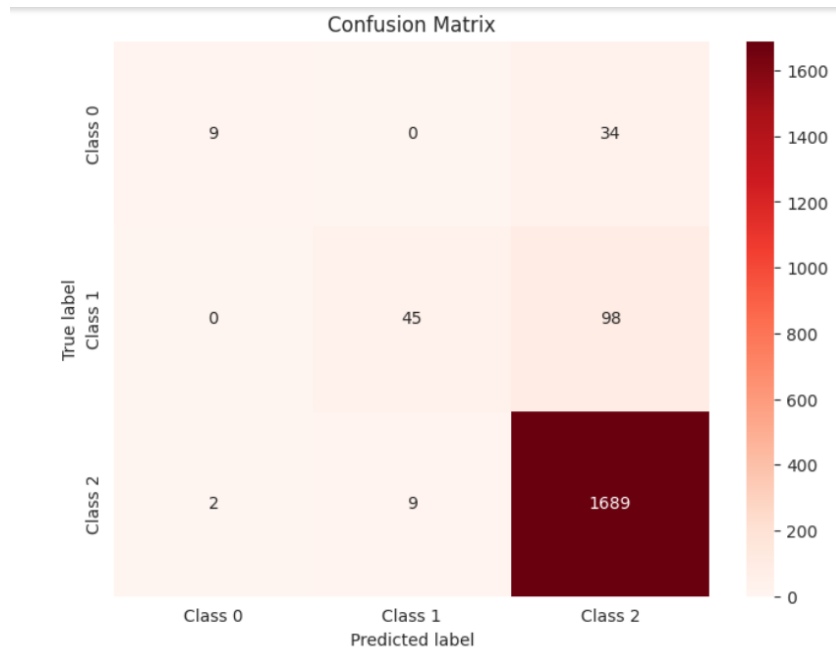


Figure 4.2.9.1: Confusion matrix (SVM)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1743 and incorrectly predicted 143.

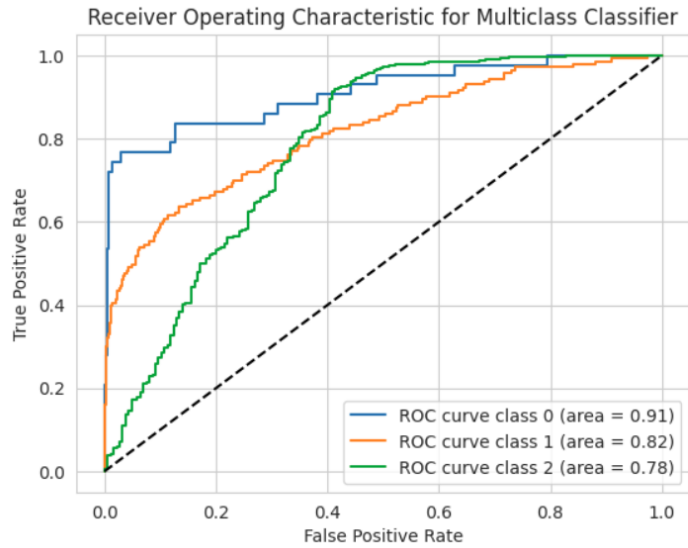


Figure 4.2.9.2: ROC curve (SVM)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 91% and false positive rate 9%, and class 1, true positive rate 82% and false positive rate 18%, class 2, true positive rate 78% and false positive rate 22%).

#### 4.2.10 Stochastic Gradient Descent

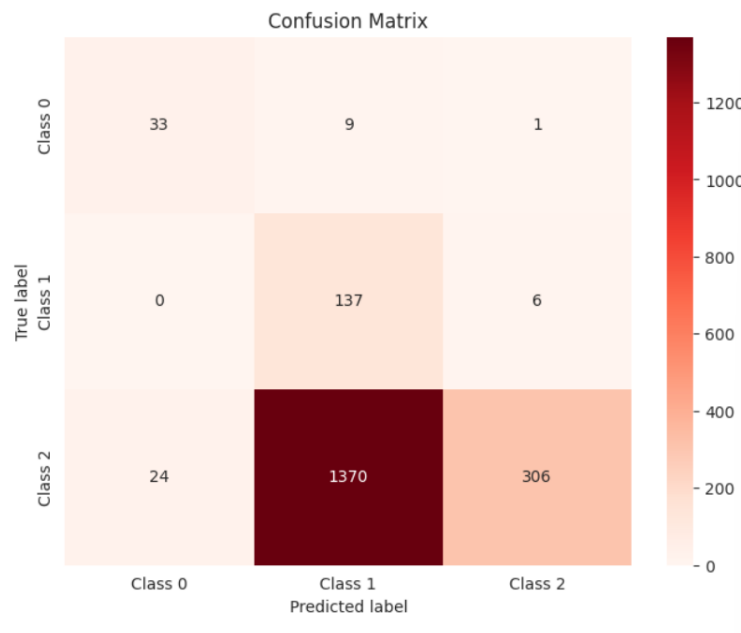


Figure 4.2.10.1: Confusion matrix (SGD)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 476 and incorrectly predicted 1440.

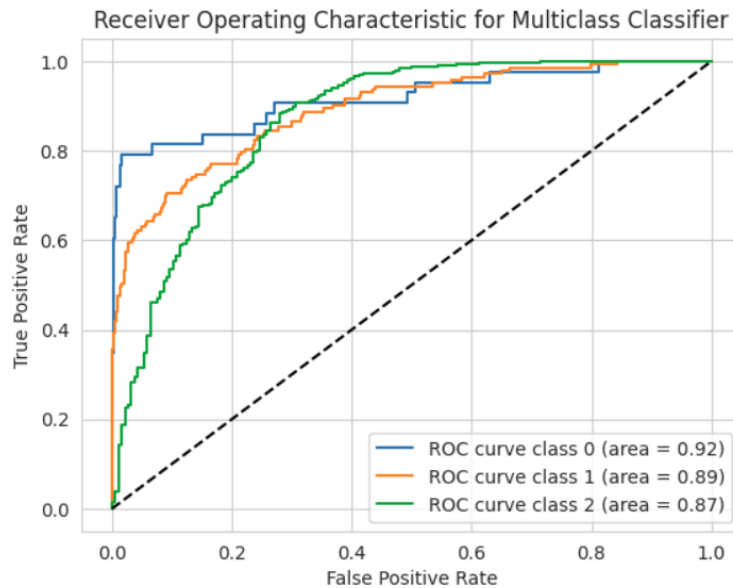


Figure 4.2.10.2: ROC curve (SGD)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 92% and false positive rate 8%, and class 1, true positive rate 89% and false positive rate 11%, class 2, true positive rate 87% and false positive rate 13%).

## 4.2.11 XGBoost

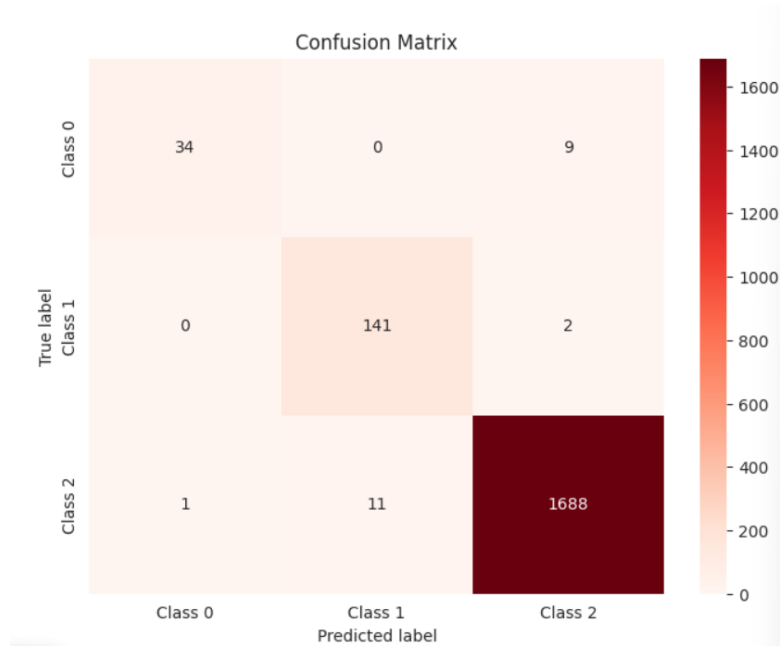


Figure 4.2.11.1: Confusion matrix (XGB)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1863 and incorrectly predicted 23.

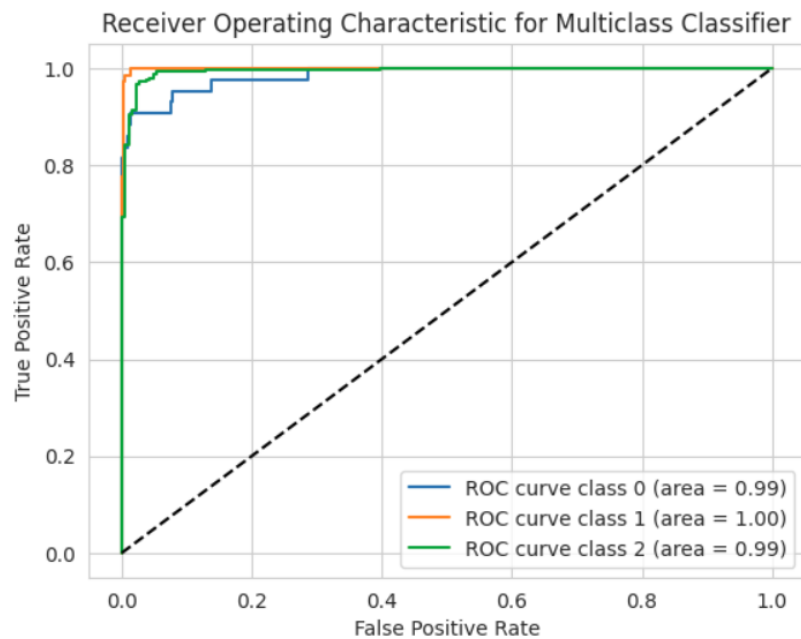


Figure 4.2.11.2: ROC curve (XGB)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 99% and false positive rate 1%, and class 1, true positive rate 100% and false positive rate 0%, class 2, true positive rate 99% and false positive rate 1%).

#### 4.2.12 Grid Search Bagging

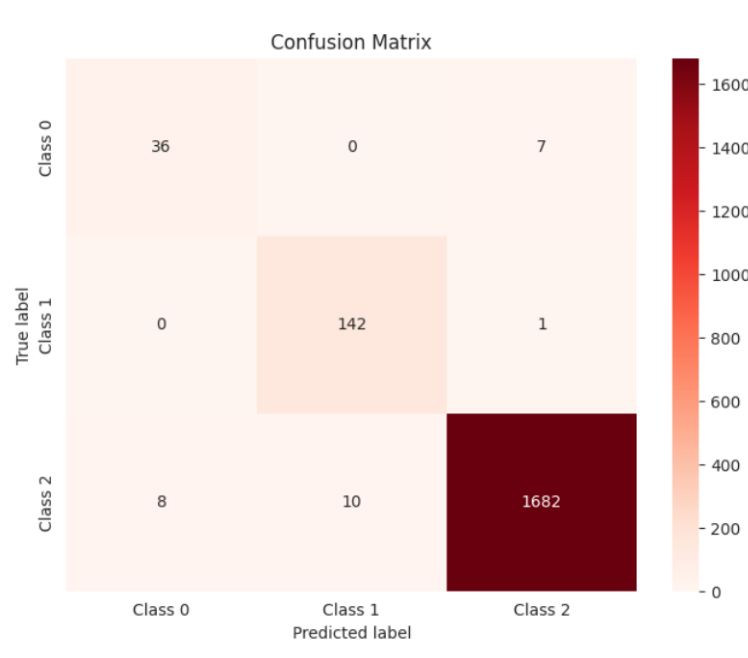


Figure 4.2.12.1: Confusion matrix (GSB)

In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1860 and incorrectly predicted 26.

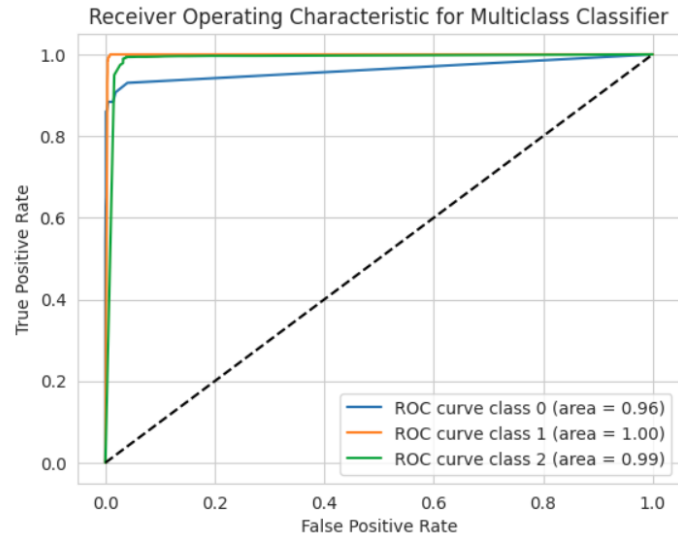


Figure 4.2.12.2: ROC curve (GSB)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 96% and false positive rate 4%, and class 1, true positive rate 100% and false positive rate 0%, class 2, true positive rate 99% and false positive rate 1%).

#### 4.2.13 Grid Search Logistic Regression (GSLR)

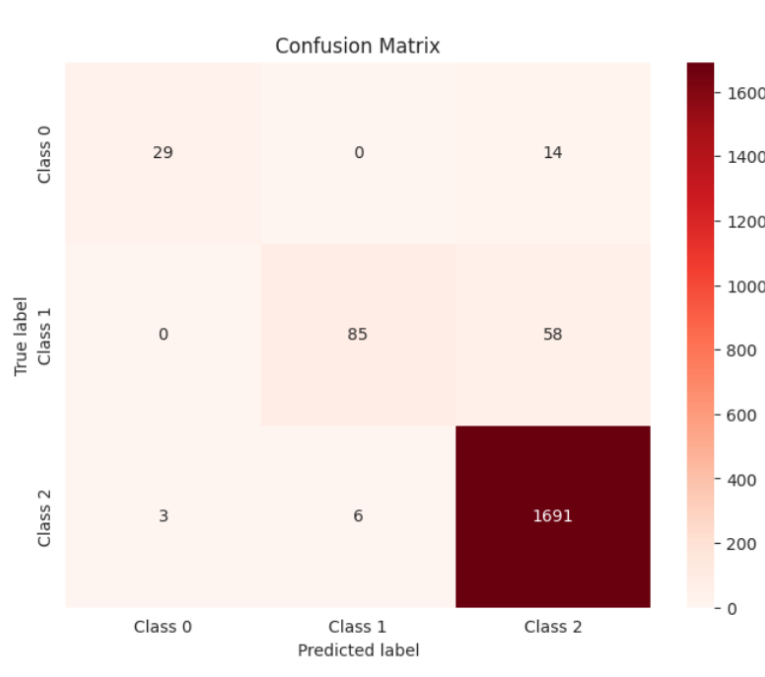


Figure 4.2.13.1: Confusion matrix (GSLR)



In this confusion matrix graph showing how many correct predictions the model made and how many incorrect predictions it made. Here we can see the model correctly predicted 1805 and incorrectly predicted 81.

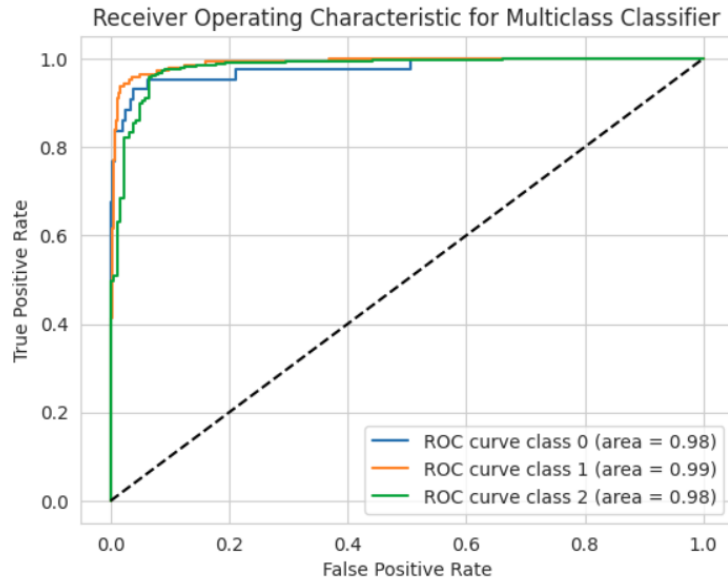


Figure 4.2.13.2: ROC curve (GSLR)

In this ROC curve shows all class accuracy rate with true positive rate. Here we can see that (class 0, true positive rate 98% and false positive rate 2%, and class 1, true positive rate 99% and false positive rate 1%, class 2, true positive rate 98% and false positive rate 2%).

## 4.2.14 Artificial Neural Network (ANN)

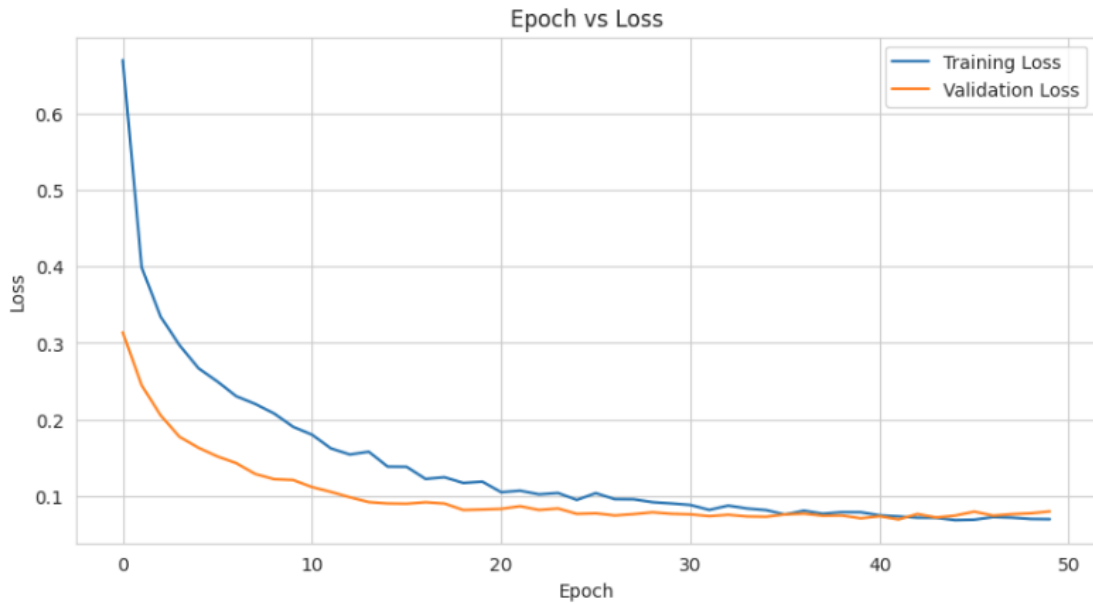


Figure 4.2.14.1: ANN Model Epoch vs Loss Plot-1

The above graph shows the statics of epoch and loss after applying the deep learning model. And it is seen that as the epoch number increases, the loss decreases.

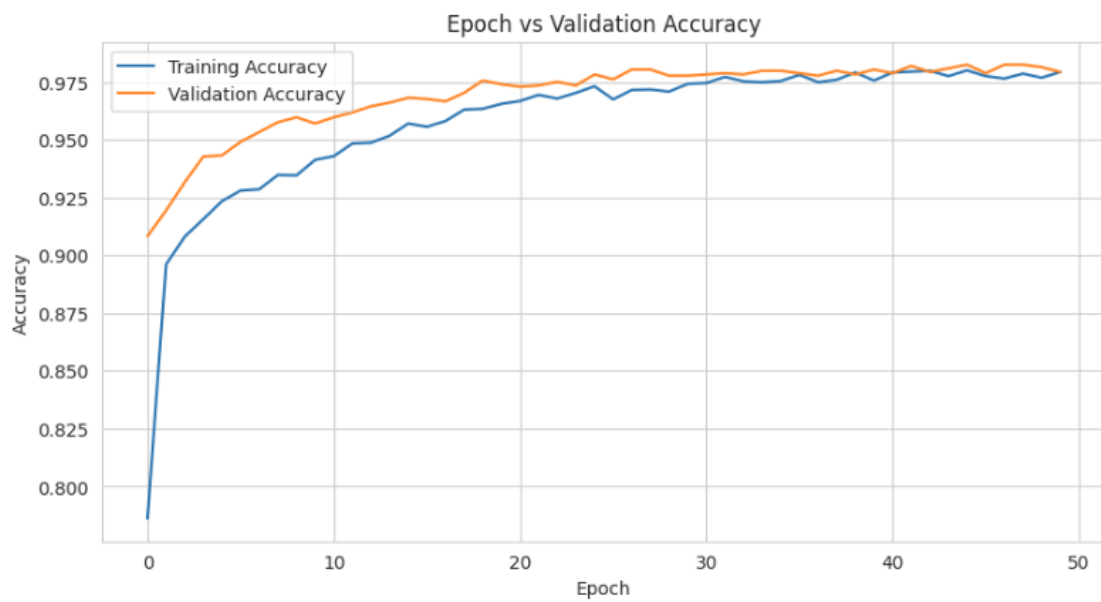


Figure 4.2.14.2: ANN Model Epoch vs Validation Plot-2

This graph shows the epoch and validation accuracy statistics. Here we can see that the model accuracy increases as the epoch number increases.

#### 4.2.15 Explainable AI Visualization

**Lime:**

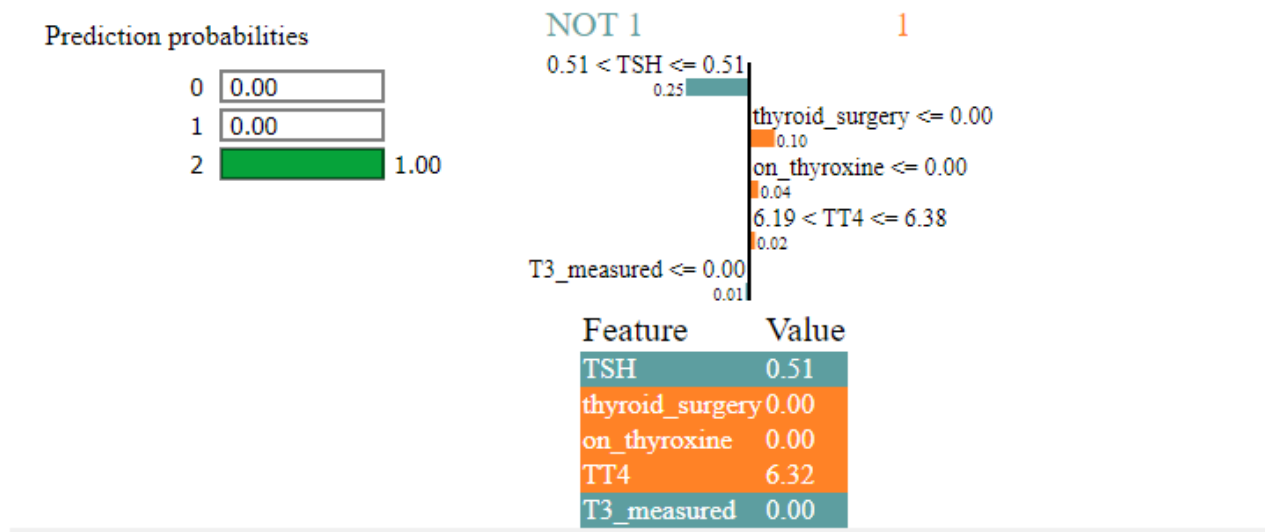


Figure: 4.2.15.1: Lime visualization of Bagging model

As seen in the Lime graph above, (XAI) explains the reason for the hyperthyroid result. It highlights the role of attributes behind this result. It gives globally entire impact summery of model.

## Shap:

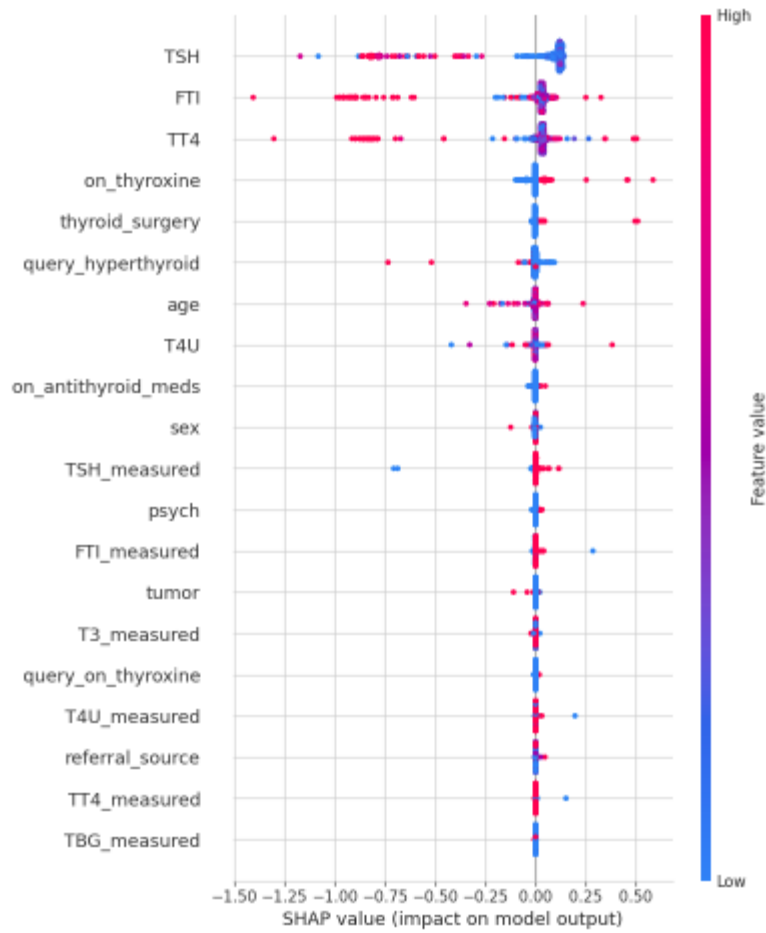


Figure 4.2.15.2: Shap visualization of Bagging model

The shape graph basically shows how much of an attribute plays a role in predicting hyperthyroid. And it shows sequentially (high to low). Also showing positive and negative values.

### 4.3 Comparative Analysis

TABLE 4.3: PROPOSED MODEL ACCURACY COMPARISON WITH OTHER'S MODEL FOR THYROIDDF

References	Model	Accuracy
Proposed Method	ThyroidNet	99.09%
Islam, et al [6]	ANN	95.87%
Shama, et al [8]	RF	91.92%
Gummadi, et al [39]	RF	97.05%
Duggal, et al [40]	SVM with (RFE)	92.92%
Ahmed et al [41]	SVM	95.7%

Here we can see that different researcher's get different accuracy. Comparing with all other researcher's model accuracy, it is found that our research model has the best accuracy and performs more effectively.

## **CHAPTER 5**

### **Impact on Society, Environment and Sustainability**

#### **5.1 Impact on Society**

This work will play an important role in our society. Using this system, thyroid problems can be detected at an early stage. Which type of disease is affected can also be identified. A proper understanding of the disease will be obtained. The patient's condition can be accurately monitored. No need to do different tests. Both time and money will be saved. Can start treatment quickly. It will help the doctor to diagnose the disease, time to time reporting of the condition will also help in prompt treatment. So, I can say that this work will have an important role on society, especially in the medical field.

#### **5.2 Ethical Aspects**

This work has several good qualities. Nowadays, going to the doctor requires a long treatment, many tests, a lot of expenses and a lot of time. But using this system does not require testing to detect the disease. The disease can be detected at an early stage. Can be treated quickly. Costs will decrease. Can be cured quickly, get a proper idea about the disease etc. So, it can be said that this system has several good qualities.

#### **5.4 Sustainability Plan**

We need to make a sustainable plan. Which can provide a permanent solution to thyroid disease. Thyroid disease is now the second most common disease in the world, and in the future this disease will become a threat to people all over the world. So, we need to control this disease by taking necessary steps now.

## CHAPTER 6

### Summary, Conclusion, Recommendation and Implication for Future Research

#### 6.1 Summary of the Study

Let me describe my entire work process. First, I took the dataset input and then preprocessed the dataset properly, because data preprocessing is most important to make the model learn well. At the beginning of preprocessing, I checked the dataset for null values and handled the null values through the null value handling method. In the next step, outliers were checked in the dataset and outliers were removed using outlier handling methods. Then using the scaling method in the dataset, I brought the imbalance data into a range and balanced the dataset, then did feature selection, and selected the important features to learn the model. Also applied some visualization to observe the state of the dataset after preprocessing.

After finishing the preprocessing, I started applying the model, since I decided to apply 13 models, applied all the models step by step, then selected the best performing model (Bagging Classifier) as the final model and that Added Explainable AI to the model.

#### 6.2 Conclusions

Thyroid disease has emerged as the second most prevalent endocrine disorder, trailing only diabetes, which has garnered significant global attention. Accurate disease classification is a critical component in the process of clinical diagnosis. In research tried to develop a proper working system which can detect thyroid disease very easily and this system can detect three types of disease (hyperthyroid, hypothyroid, negative). It can also detect thyroid problems at an early stage. Selected a good dataset from UCI-repository and preprocessed it, balanced the dataset and converted it to model suitable format. Balanced and fitted the dataset by removing missing values, error data, outliers, scaling, encoding, to train the model well. Then applied different models such as- bagging classifier model worked best. The results of this model are -Accuracy 99.04%, Precision 99%, Recall 99%, F1-score 99%. Then added Explainable AI on the best model, to explain the prediction

result of the model. Comparing with other works, we can see that our model gives the best accuracy.

### **6.3 Implication for Further Study**

There are plans to do more work on this research in the future. Wants to work on real-time or visual data in future. And add more classes of thyroid disease. There are plans to work with more large datasets. Also want to work on new problems of thyroid disease.



## REFERENCES

- [1] M. Surks, E. Ortiz, G. Daniels, C. Sawin, N. C.- Jama, and undefined 2004, “Subclinical thyroid disease: scientific review and guidelines for diagnosis and management,” *jamanetwork.com*, Accessed: Jan. 16, 2024. [Online]. Available: <https://jamanetwork.com/journals/jama/article-abstract/197994>
- [2] “Goiter | Endocrine Society.” Accessed: Jan. 16, 2024. [Online]. Available: <https://www.endocrine.org/patient-engagement/endocrine-library/goiter>
- [3] “Thyroid Disease: Everything You Need To Know - Aether Heath - Kingwood ER.” Accessed: Jan. 16, 2024. [Online]. Available: <https://kingwooder.com/thyroid-disease-everything-you-need-to-know/>
- [4] “(9) (PDF) ENDOCRINE HEALTH IN SCHOOL TEACHERS FROM PRIMARY, HIGH, AND HIGHER SECONDARY SCHOOLS OF SOUTH TAMILNADU, INDIA.” Accessed: Jan. 16, 2024. [Online]. Available: [https://www.researchgate.net/publication/334249641\\_ENDOCRINE\\_HEALTH\\_IN\\_SCHOOL\\_TEACHERS\\_FROM\\_PRIMARY\\_HIGH\\_AND\\_HIGHER\\_SECONDARY\\_SCHOOLS\\_OF\\_SOUTH\\_TAMILNADU\\_INDIA](https://www.researchgate.net/publication/334249641_ENDOCRINE_HEALTH_IN_SCHOOL_TEACHERS_FROM_PRIMARY_HIGH_AND_HIGHER_SECONDARY_SCHOOLS_OF_SOUTH_TAMILNADU_INDIA)
- [5] “Thyroid: What It Is, Function & Problems.” Accessed: Jan. 16, 2024. [Online]. Available: <https://my.clevelandclinic.org/health/body/23188-thyroid>
- [6] S. Islam, M. Haque, M. Miah, ... T. S.-P. C., and undefined 2022, “Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study,” *peerj.com*SS Islam, MS Haque, MSU Miah, TB Sarwar, R NugrahaPeerJ Computer Science, 2022•peerj.com, Accessed: Jan. 16, 2024. [Online]. Available: <https://peerj.com/articles/cs-898/>
- [7] T. Singh, A. Sahu, S. Dubey, ... M. S.-I. J. of, and undefined 2022, “Treatment of thyroid disease through machine learning predictive model,” *academia.edu*T Singh, AK Sahu, S Dubey, MP Sharma, S Verma, C KumarInternational Journal of Health Sciences, 2022•academia.edu, Accessed: Jan. 16, 2024. [Online]. Available: [https://www.academia.edu/download/93461396/IJHS-12813\\_3176-3188.pdf\\_filename\\_UTF-8IJHS-12813\\_3176-3188.pdf](https://www.academia.edu/download/93461396/IJHS-12813_3176-3188.pdf_filename_UTF-8IJHS-12813_3176-3188.pdf)
- [8] A. Shama, M. B. Hossain, A. Adhikary, K. M. A. Uddin, and M. A. Hossain, “Prediction of Hypothyroidism and Hyperthyroidism Using Machine Learning Algorithms,” 2022, doi: 10.21203/rs.3.rs-1486798/v2.
- [9] S. Verma, R. Popli, and ... H. K.-I. J. of, “Classification of thyroid diseases using machine learning frameworks,” *pdfs.semanticscholar.org*S Verma, R Popli, H Kumar, A SrivastavaInternational

*Journal of Health Sciences*•pdfs.semanticscholar.org, Accessed: Jan. 16, 2024. [Online]. Available: <https://pdfs.semanticscholar.org/70cb/f0aec6ac8c10035d98957006b32a7289f4a5.pdf>

- [10] E. S.-J. of P. C. Series and undefined 2021, “Thyroid disease classification using machine learning algorithms,” *iopscience.iop.orgE SonuçJournal of Physics: Conference Series, 2021*•iopscience.iop.org, doi: 10.1088/1742-6596/1963/1/012140.
- [11] J. Wen, L. Peng, H. Ji, and , li Peng, “Thyroid nodule risk assessment based on neural network text classification method,” 2023, doi: 10.21203/rs.3.rs-3599555/v1.
- [12] Q. Pan, Y. Zhang, M. Zuo, ... L. X.-2016 8th I., and undefined 2016, “Improved ensemble classification method of thyroid disease based on random forest,” *ieeexplore.ieee.orgQ Pan, Y Zhang, M Zuo, L Xiang, D Chen2016 8th International conference on information technology in, 2016*•ieeexplore.ieee.org, Accessed: Jan. 16, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7976548/>
- [13] E. S.-J. of P. C. Series and undefined 2021, “Thyroid disease classification using machine learning algorithms,” *iopscience.iop.orgE SonuçJournal of Physics: Conference Series, 2021*•iopscience.iop.org, doi: 10.1088/1742-6596/1963/1/012140.
- [14] V. Surwade, B. A. Al-Sum, S. Kumar Mishra, D. Srivastava, and R. Sharma, “Thyroid disease classification using decision tree and SVM,” *scholar.cu.edu.egK Dharmarajan, K Balasree, AS Arunachalam, K AbirmaiIndian Journal of Public Health Research & Development, 2020*•scholar.cu.edu.eg, Accessed: Jan. 16, 2024. [Online]. Available: [https://scholar.cu.edu.eg/sites/default/files/randaerfan/files/ijphrd\\_march\\_2020\\_0.pdf#page=263](https://scholar.cu.edu.eg/sites/default/files/randaerfan/files/ijphrd_march_2020_0.pdf#page=263)
- [15] V. Viswanatha, “Thyroid Disease Detection Using Machine Learning approach,” 2023, Accessed: Jan. 16, 2024. [Online]. Available: <https://osf.io/preprints/peha2/>
- [16] “Cross Validation in Machine Learning - GeeksforGeeks.” Accessed: Jan. 16, 2024. [Online]. Available: <https://www.geeksforgeeks.org/cross-validation-machine-learning/>
- [17] M. Hossin, & M. S.-I. journal of data mining, and undefined 2015, “A review on evaluation metrics for data classification evaluations,” *academia.eduM Hossin, MN SulaimanInternational journal of data mining & knowledge management process, 2015*•academia.edu, Accessed: Jan. 16, 2024. [Online]. Available: <https://www.academia.edu/download/37219940/5215ijdkp01.pdf>
- [18] “Decision Trees for Machine Learning.” Accessed: Jan. 16, 2024. [Online]. Available: <https://devopedia.org/decision-trees-for-machine-learning>
- [19] “K-Nearest Neighbors (KNN) in Depth | by Tech-AI-Math | Artificial Intelligence in Plain English.” Accessed: Jan. 16, 2024. [Online]. Available: <https://ai.plainenglish.io/k-nearest-neighbors-knn-769bd39514c6>

- [20] M. ALSAADAWI, “DETECTION OF THYROID DISEASE USING MACHINE LEARNING MODELS,” 2023, Accessed: Jan. 16, 2024. [Online]. Available: <http://acikerisim.karabuk.edu.tr:8080/xmlui/handle/123456789/2444>
- [21] “Logistic Regression | How to Implement Logistic Regression in Python?” Accessed: Jan. 16, 2024. [Online]. Available: <https://www.analyticsvidhya.com/blog/2021/05/how-can-we-implement-logistic-regression/>
- [22] O. Sagi and L. Rokach, “Ensemble learning: A survey,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/WIDM.1249.
- [23] “Bagging Classifier Python Code Example.” Accessed: Jan. 16, 2024. [Online]. Available: <https://vitalflux.com/bagging-classifier-python-code-example/>
- [24] A. Wyner, M. Olson, J. Bleich, D. M.-T. J. of M. Learning, and undefined 2017, “Explaining the success of adaboost and random forests as interpolating classifiers,” *jmlr.org* AJ Wyner, M Olson, J Bleich, D Mease *The Journal of Machine Learning Research*, 2017•*jmlr.org*, vol. 18, pp. 1–33, 2017, Accessed: Jan. 16, 2024. [Online]. Available: <https://www.jmlr.org/papers/volume18/15-240/15-240.pdf>
- [25] S. Salcedo-Sanz *et al.*, “Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review,” *Theor Appl Climatol*, 2023, doi: 10.1007/S00704-023-04571-5.
- [26] F. Kamiran, T. C.-K. and information systems, and undefined 2012, “Data preprocessing techniques for classification without discrimination,” *Springer* F Kamiran, T Calders *Knowledge and information systems*, 2012•*Springer*, vol. 33, no. 1, pp. 1–33, 2011, doi: 10.1007/s10115-011-0463-8.
- [27] “‘Demystifying Naïve Bayes: Simple yet Powerful for Text Classification’ | by ajaymehta | Medium.” Accessed: Jan. 16, 2024. [Online]. Available: <https://medium.com/@dancerworld60/demystifying-na%C3%AFve-bayes-simple-yet-powerful-for-text-classification-ad92b14a5c7>
- [28] A. B. Shaik and S. Srinivasan, “A brief survey on random forest ensembles in classification model,” *Lecture Notes in Networks and Systems*, vol. 56, pp. 253–260, 2019, doi: 10.1007/978-981-13-2354-6\_27/COVER.
- [29] “Anas Brital | Random Forest Algorithm Explained .” Accessed: Jan. 16, 2024. [Online]. Available: <https://anasbrital98.github.io/blog/2021/Random-Forest/>
- [30] C. Y. Lee, J. G. Shon, and J. S. Park, “An edge detection–based eGAN model for connectivity in ambient intelligence environments,” *J Ambient Intell Humaniz Comput*, vol. 13, no. 10, pp. 4591–4600, Oct. 2022, doi: 10.1007/S12652-021-03261-2.

- [31] J. Jiang *et al.*, “Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm,” *nature.com*, Accessed: Jan. 16, 2024. [Online]. Available: <https://www.nature.com/articles/s41598-021-85223-4>
- [32] “Empirical comparison of bagging-based ensemble classifiers | IEEE Conference Publication | IEEE Xplore.” Accessed: Jan. 16, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/6289900>
- [33] S. Bera, V. S.-I. J. of R. Sensing, and undefined 2020, “Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification,” *Taylor & Francis S Bera, VK Shrivastava International Journal of Remote Sensing, 2020 • Taylor & Francis*, vol. 41, no. 7, pp. 2664–2683, Apr. 2020, doi: 10.1080/01431161.2019.1694725.
- [34] A. Gulli, A. Kapoor, and S. Pal, *Deep learning with TensorFlow 2 and Keras: regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API*. 2019. Accessed: Jan. 16, 2024. [Online]. Available: [https://books.google.com/books?hl=en&lr=&id=BVnHDwAAQBAJ&oi=fnd&pg=PP1&dq=Gulli,+A.,+Kapoor,+A.,+%26+Pal,+S.,+\(2019\).+Deep+learning+with+TensorFlow+2+and+Keras:+regression,+ConvNets,+GANs,+RNNs,+NLP,+and+more+with+TensorFlow+2+and+the+Keras+API.+Packt+Publishing+Ltd.&ots=K-r89qTs0W&sig=OUPstvlspGzudRKeTbkJ3d14kdE](https://books.google.com/books?hl=en&lr=&id=BVnHDwAAQBAJ&oi=fnd&pg=PP1&dq=Gulli,+A.,+Kapoor,+A.,+%26+Pal,+S.,+(2019).+Deep+learning+with+TensorFlow+2+and+Keras:+regression,+ConvNets,+GANs,+RNNs,+NLP,+and+more+with+TensorFlow+2+and+the+Keras+API.+Packt+Publishing+Ltd.&ots=K-r89qTs0W&sig=OUPstvlspGzudRKeTbkJ3d14kdE)
- [35] “Applied Deep Learning - Part 1: Artificial Neural Networks | by Arden Dertat | Towards Data Science.” Accessed: Jan. 16, 2024. [Online]. Available: <https://towardsdatascience.com/applied-deep-learning-part-1-artificial-neural-networks-d7834f67a4f6>
- [36] A. P.-M. and Machines and undefined 2019, “The pragmatic turn in explainable artificial intelligence (XAI),” *Springer A Páez Minds and Machines, 2019 • Springer*, vol. 29, no. 3, pp. 441–459, Sep. 2019, doi: 10.1007/s11023-019-09502-w.
- [37] M. Hossin, & M. S.-I. journal of data mining, and undefined 2015, “A review on evaluation metrics for data classification evaluations,” *academia.edu M Hossin, MN Sulaiman International journal of data mining & knowledge management process, 2015 • academia.edu*, Accessed: Jan. 16, 2024. [Online]. Available: <https://www.academia.edu/download/37219940/52151jdkp01.pdf>
- [38] “What is a confusion matrix?. Everything you Should Know about... | by Anuganti Suresh | Analytics Vidhya | Medium.” Accessed: Jan. 16, 2024. [Online]. Available: <https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5>
- [39] A. Gummadi and D. Reddy, “A Novel Machine Learning Framework for Prediction of Early-Stage Thyroid Disease Using Classification Techniques,” 2022, Accessed: Jan. 16, 2024. [Online]. Available: <https://www.academia.edu/download/90634911/9615.pdf>

- [40] P. Duggal, S. S. C. Computing, & D. S., and undefined 2020, "Prediction of thyroid disorders using advanced machine learning techniques," *ieeexplore.ieee.org* P Duggal, S Shukla 2020 10th International Conference on Cloud Computing, Data, 2020•*ieeexplore.ieee.org*, Accessed: Jan. 16, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9058102/>
- [41] J. Ahmed and M. A. R. Soomrani, "TDTD: Thyroid disease type diagnostics," *2016 International Conference on Intelligent Systems Engineering, ICISE 2016*, pp. 44–50, May 2016, doi: 10.1109/INTELSE.2016.7475160.

# A Novel Explainable AI Based Approach: Early Stage Thyroid Disease Classification

## ORIGINALITY REPORT

<b>15%</b>	<b>12%</b>	<b>4%</b>	<b>10%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

## PRIMARY SOURCES

<b>1</b>	<b>Submitted to Daffodil International University</b> Student Paper	<b>3%</b>
<b>2</b>	<b>Submitted to Huntington Beach Union High School District</b> Student Paper	<b>3%</b>
<b>3</b>	<b>dspace.daffodilvarsity.edu.bd:8080</b> Internet Source	<b>3%</b>
<b>4</b>	<b>www.researchgate.net</b> Internet Source	<b>1%</b>
<b>5</b>	<b>Sacks, D.A.. "Fasting plasma glucose test at the first prenatal visit as a screen for gestational diabetes", Obstetrics &amp; Gynecology, 200306</b> Publication	<b>1%</b>
<b>6</b>	<b>sciencescholar.us</b> Internet Source	<b>&lt;1%</b>
<b>7</b>	<b>Submitted to CSU Northridge</b> Student Paper	<b>&lt;1%</b>