

**DETECTION OF THYROID DISEASES USING MACHINE  
LEARNING TECHNIQUES**

By

**TEHSEEN AKHTAR**

**NUST201490166PSMME2714S**



**A thesis submitted to the Department of Robotics and Artificial Intelligence in partial  
fulfilment of the requirements for the degree of**

**DOCTOR OF PHILOSOPHY**

**IN**

**ROBOTICS AND INTELLIGENT MACHINE ENGINEERING**

**2021**

**SUPERVISED BY:**

**DR. SYED OMER GILANI**

**CO-SUPERVISOR:**

**DR. MOHSIN JAMIL**

**SCHOOL OF MECHANICAL AND MANUFACTURING ENGINEERING  
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY  
ISLAMABAD, PAKISTAN**



**In the name of ALLAH**

**The Most Beneficent and The Most Merciful**

## CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis titled “Detection of Thyroid Diseases Using Machine Learning Techniques” was conducted by Mr. Tehseen Akhtar under the supervision of Dr. Syed Omer Gilani

No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted in the School of Mechanical and Manufacturing Engineering in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the field of Robotics and Intelligent Machine Engineering at the Department of Robotics and Artificial Intelligence (R & AI), School of Mechanical and Manufacturing Engineering (SMME), National University of Science and Technology (NUST), Sector H-12, Islamabad, Pakistan.

Student Name: Tehseen Akhtar

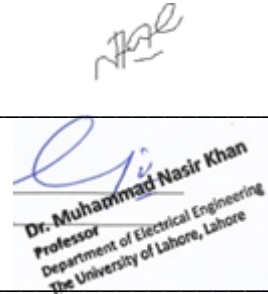
Signature: \_\_\_\_\_

### **Examination Committee:**

#### External Examiner 1:

Dr. Qasim Awais  
Associate Professor/ HOD, Electrical  
Engineering Dept., GIFT University, Pakistan

Signature: \_\_\_\_\_



Dr. Muhammad Nasir Khan  
Professor  
Department of Electrical Engineering  
The University of Lahore, Lahore

#### External Examiner 2:

Dr. Prof. Ir. Muhammad Nasir Khan  
Prof. Electrical Engineering Dept., University  
of Lahore, Pakistan

Signature: \_\_\_\_\_

#### Internal Examiner:

Dr. Shahid Ikram Ullah Butt  
Prof./ HOD, Dept. of Mechanical Engineering,  
SMME, NUST, Pakistan

Signature: \_\_\_\_\_

#### PhD Supervisor:

Dr. Syed Omer Gilani

Signature: \_\_\_\_\_

#### HoD Robotics & AI:

Dr. Hassan Sajid

Signature: \_\_\_\_\_

## **Author's Declaration**

I, **Tehseen Akhtar**, hereby state that my PhD thesis titled “**Detection of Thyroid Diseases using Machine Learning Techniques**” is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology (NUST), Islamabad, or anywhere else in the country / world.

At any time if my statement is found to be incorrect even after my Graduation, the university has the right to withdraw my PhD degree.

Dated:

Signature: \_\_\_\_\_

Name of Student: (Tehseen Akhtar)

## Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled “**Detection of Thyroid Diseases using Machine Learning Techniques**” is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and that of National University of Sciences and Technology (NUST), Islamabad, towards plagiarism. Therefore I, as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD degree, the University reserves the rights to withdraw / revoke my PhD degree and that HEC and the University has the right to publish my name on the HEC / University Website on which names of students are placed who submitted plagiarized thesis.

Date:

Student/ Author Signature: \_\_\_\_\_

Name: (Tehseen Akhtar)

## THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of PhD Thesis written by Mr. Tehseen Akhtar, (Registration No. NUST201490166PSMME2714S), of School of Mechanical and Manufacturing Engineering (SMME) has been vetted by undersigned, found complete in all respects as per NUST Statutes / Regulations / PhD Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfilment for award of PhD degree. It is further certified that necessary amendments as pointed out by GEC members and foreign/local evaluators of the scholar have also been suitably incorporated in the said thesis.

Signature: \_\_\_\_\_

PhD Supervisor: Dr. Syed Omer Gilani

Date: \_\_\_\_\_

Signature: \_\_\_\_\_



PhD Supervisor: Dr. Mohsin Jamil

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

HoD Robotics & AI: Dr. Hasan Sajid

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

Principal SMME: Prof. Dr. Javaid Iqbal

Date: \_\_\_\_\_



## **COPYRIGHT STATEMENT**

- The student author retains the copyright in the text of this thesis. Only in line with the author's instructions and housed in the library of SMME, NUST, may copies (by any method) be created in whole or excerpts. The librarian can provide further information. This page must be included in all copies created. Without the author's permission (in writing), no more copies (by any means) may be created.
  
- Subject to any prior agreement to the contrary, ownership of any intellectual property rights described in this thesis is vested in SMME, NUST, and may not be made available for use by third parties without the written permission of the SMME, which will prescribe the terms and conditions of any such agreement.
  
- The library of SMME, NUST, Islamabad, Pakistan, has further information on the situations under which disclosures and exploitations may occur.



## **ACKNOWLEDGEMENTS**

First and foremost, I want to express my gratitude to "ALLAH Almighty," who provided me with the strength and fortitude to accomplish this lengthy journey that I would not have been able to do without his help.

My family was always there for me, especially my mother, who supported me during difficult times, and my wife, who always encouraged me when I was discouraged. Their encouragement drove me to take the next step whenever I felt hopeless.

I'd want to express my gratitude to my wonderful supervisor, Dr. Syed Omer Gilani, for providing me with the necessary advice and for always pushing me in the correct direction. His guidance and inspiration are critical to the successful completion of this degree.

Prof. Dr. Yasar Ayaz and Dr. Shahid Ikram Ullah Butt, members of my guiding and assessment committee, have shown me the route to a successful conclusion. Their encouraging words and constructive criticism constantly prompted me to rethink and refocus my studies.

Special thanks to Dr. Zohaib Mushtaq, my instructor, mentor, and friend, for supplying me with the appropriate data sets and assisting me in obtaining the necessary findings for my articles.

I'd also want to thank Dr. Mohsin Jamil for giving me with up-to-date topic knowledge and for assisting me in important areas of my research, without which I would not have been able to reach my aim. The Controls Lab, the Robotics and AI department, and SMME have all provided invaluable assistance.

I'd also want to express my gratitude to the Principal of SMME, Prof. Dr. Javaid Iqbal, for all of the administrative assistance he provided me during tough times, which enabled me to fulfil my deadlines.

## ACRONYMS

<b>SFM</b>	Select from Model
<b>SKB</b>	Select K-Best
<b>RFE</b>	Recursive Feature Elimination
<b>DT</b>	Decision Tree
<b>LR</b>	Logistic Regression
<b>GB</b>	Gradient Boosting
<b>RF</b>	Random Forest
<b>TSH</b>	Thyroid Stimulating Hormone
<b>TRH</b>	Thyrotrophic-releasing hormone
<b>TBG</b>	Thymoglobulin/thyroid-binding globulin
<b>TRAb</b>	Thyroid-receptor antibodies
<b>MLP</b>	Multilayer Perceptron
<b>CHAID</b>	Chi-Squared Automatic Interaction Detector
<b>SVM</b>	Support Vector Machine
<b>BME</b>	Base Meta Estimator
<b>AB</b>	AdaBoost
<b>XGB</b>	XGBoost

## ABSTRACT

**Background:** The unusual growth of the glandular tissue on the boundary of the Thyroid gland is an indication of Thyroid disease. Thyroid disease is characterised by an unusually high or low number of hormones produced by the thyroid gland, the two most prevalent kinds are hypothyroidism (underactive thyroid gland) and hyperthyroidism (overactive thyroid gland). The main aim of this project was to introduce the concept of an efficient multi-stage ensemble i.e., the voting ensemble of the homogeneous ensemble which could be used with a variety of feature-selection algorithms for improving the diagnosis of thyroid diseases. The dataset utilised in this study was built from real-time thyroid data obtained from the teaching hospital in DG Khan at District Head Quarter (DHQ), Pakistan. Following the appropriate pre-processing processes, three kinds of attribute-selection strategies were used: The first approach used was Select from Model (SFM), the second technique was the Select K-Best (SKB), and the final methodology was the Recursive Feature Elimination (RFE). Select From Model (SFM) is a form of attribute-selection strategy that uses a model to select attributes. As potential feature estimators, the Decision Tree (DT), Logistic Regression (LR), Gradient Boosting (GB) and Random Forest denoted as the (RF) classifiers were employed in conjunction with each other. The homogeneous ensemble activated the bagging, boosting-based learners, who were then classified by the Voting ensemble, which employed both soft and hard voting to categorise the data. Other performance assessment criteria such as hamming loss, accuracy, mean square error, sensitivity and others have been implemented. The results of the experiments reveal that when the suggested approach for better thyroid sickness detection is applied in its most practicable form, it is most successful. On the dataset 1, all of the algorithms tested obtained 100 % accuracy with subset of the total no of feature in each case, however on the dataset 2, more than 98 percent accuracy was reached in every case. On the basis of accuracy and computing cost, the results given here exceeded equivalent benchmark models in their respective fields of study.

## TABLE OF CONTENTS

1. CERTIFICATE OF APPROVAL .....	iii
2. Author's Declaration .....	iv
3. Plagiarism Undertaking .....	v
4. THESIS ACCEPTANCE CERTIFICATE.....	vi
5. ACKNOWLEDGEMENTS.....	ix
6. ACRONYMS.....	x
7. ABSTRACT .....	xi
8. TABLE OF CONTENTS .....	xii
9. List of Figures.....	xv
10. List of Tables .....	xviii
11. LIST OF PUBLICATIONS.....	xix
1. CHAPTER 1 INTRODUCTION.....	1
1.1. Introduction .....	1
1.2. Background .....	1
1.3. Structure Of Thyroid Glands.....	2
1.4. Location Of Thyroid Gland.....	3
1.5. Function Of Thyroid Gland.....	4
1.5.1 Mechanism Of the Function of Thyroid Gland.....	4
1.6. Diseases of Thyroid Glands .....	6
1.6.1 Major types of diseases of thyroid gland: .....	8
1.6.2 Signs and the symptoms of diseases: .....	11
1.6.3 Methods for the detection of thyroid gland diseases: .....	11

1.7.	Research Motivation .....	15
1.8.	Detection of thyroid gland disease using machine learning:.....	15
1.9.	Report Structure .....	17
2.	CHAPTER 2 LITERATURE REVIEW .....	19
2.1.	Thyroid Disorders Prediction Using Machine Learning .....	19
2.2.	Thyroid Disease Diagnosis Using Medical Imaging .....	26
2.3.	Thyroid Disease Detection Algorithms Based on Clustering .....	31
2.4.	Methods for Detecting Thyroid Diseases that are Automated and Intelligent .....	32
3.	CHAPTER 3 METHODOLOGY .....	42
3.1.	Dataset 1 .....	42
3.2.	Dataset 2 .....	43
3.3.	Proposed Methodology .....	45
3.3.1	Data Pre-processing .....	46
3.3.2	XGBoost based Feature Significance with F-score .....	47
3.4.	Attribute Selection Approaches .....	51
3.4.1	Selection From Model (SFM).....	51
3.4.2	Recursive Feature Elimination (RFE).....	51
3.4.3	Select K-Best (SKB) based on Simple regression Extracted Features .....	52
3.5.	Isolation Forest (ISO) for Automatic Identification and Elimination of Outliers .....	52
3.6.	Homogenous Ensemble.....	54
3.6.1	Bagging.....	54

3.6.2	Boosting .....	56
3.7.	Voting Ensemble of Homogenous Ensemble.....	56
3.7.1	Soft Voting Ensemble .....	57
3.7.2	Hard Voting Ensemble.....	57
3.8.	Performance Assessment Metrics .....	58
4.	CHAPTER 4 RESULTS AND DISCUSSIONS .....	61
4.1.	Introduction .....	61
4.2.	Dataset 1 results .....	61
4.2.1	Homogenous Bagging.....	61
4.2.2	Homogenous Boosting.....	63
4.2.3	SKB feature selection on dataset 1 .....	64
4.2.4	SFM feature selection with Estimators on dataset 1.....	66
4.2.5	RFE feature selection on dataset 1 .....	70
4.3.	Dataset 2 results .....	77
4.3.1	Recursive Feature Elimination (Hard).....	80
4.3.2	Select K-Best Features (Hard) .....	84
4.3.3	Selection of Features from Model (Hard).....	86
4.3.4	Recursive Feature Elimination (soft).....	90
4.3.5	Select K-Best Features (soft) .....	94
4.3.6	Selection of Features from Model (Hard).....	96
4.4.	Comparison of proposed methodology with existing studies .....	101
5.	CHAPTER 5 CONCLUSION .....	106
6.	REFERENCES .....	108

## List of Figures

Figure 1.1. Structure of Thyroid Gland in Human Body.....	3
Figure 1.2 Position of thyroid gland in human body .....	4
Figure 1.3 Thyroid gland positive (+) and negative (-) feedback mechanism.....	5
Figure 1.4 shows the difference between the normal gland hyperthyroidisms.....	9
Figure 1.5 Effect of hyperthyroidism.....	10
Figure 3.1 Correlation heatmap between features and target variables for Dataset 1. ....	44
Figure 3.2 Correlation heatmap between features and target variables for Dataset 2. ....	44
Figure 3.3 Proposed Methodology Block Diagram. ....	46
<b>Figure 3.4 The 'Class' variable categories are described for (a) Dataset 1 (b) Dataset 2</b>	<b>48</b>
Figure 3.5 (a-d) SFM features selected using DT, GB, LR, RF, (e, f) SKB features selected using Chi2, FCI, and (g-j) RFE characteristics selected using DT, GB, LR, RF, respectively.(a-d) SFM features selected using DT, GB, LR, RF, (e, f) SKB features selected using Chi2, FCI, and (g-j) RFE characteristics selected using DT, GB, LR, RF, respectively.....	49
Figure 0.6 XGBoost assigns relevance ratings to attributes based on F-scores for features extracted using characteristics selecting algorithms (k-n) The characteristics specified for RFE using DT, GB, LR, RF, (o, p) for SKB employing Chi2, FCI, (q-t) for SFM utilizing DT, GB, LR, RF, and (q-t) for SFM utilizing DT, GB, LR, RF, respectively. ....	50
Figure 3.7 (a) Soft and (b) hard voting are two instances of ensemble concepts. ....	58
Figure 4.1 SKB with Chi (Hard Voting).....	64
Figure 4.2 SKB with FCN (Hard Voting).....	64
Figure 4.3 SKB with Chi (Soft Voting).....	65

Figure 4.4 SKB with FCN (Soft Voting) .....	66
Figure 4.5 SFM with Decision Trees (Soft Voting) .....	66
Figure 4.6 SFM with Decision Trees (Soft Voting) .....	67
Figure 4.7 SFM with GBC (Hard Voting) .....	67
Figure 4.8 SFM with GBC (Soft Voting) .....	68
Figure 4.9 SFM with LR (Hard Voting) .....	68
Figure 4.10 SFM with LR (Soft Voting) .....	69
Figure 4.11 SFM with RF (Hard Voting) .....	69
Figure 4.12 SFM with RF (Soft Voting).....	70
Figure 4.13 RFE with DT (Hard Voting).....	71
Figure 4.14 RFE with GB (Hard Voting) .....	71
Figure 4.15 RFE with LR (Hard Voting).....	72
Figure 4.16 RFE with RF (Hard Voting).....	72
Figure 4.17 RFE with DT (Soft Voting).....	73
Figure 4.18 RFE with GB (Soft Voting).....	73
Figure 4.19 RFE with LR (Soft Voting) .....	74
Figure 4.20 RFE with RF (Soft Voting) .....	74
Figure 4.21 Recursive Feature Elimination using Decision Trees (Hard Voting).....	80
Figure 4.22 Recursive Feature Elimination Based on Gradient Boosting (Hard Voting) .....	81
Figure 4.23 Recursive Feature Elimination Based on Logistic Regression (Hard Voting).....	82



Figure 4.24 Recursive Feature Elimination Based on Random Forests (Hard Voting).....	83
Figure 4.25 Chi Square Feature Selection Technique Performance .....	84
Figure 4.26 FCI Based Feature Selection Technique Performance .....	85
Figure 4.27 Selection of Features from Decision Trees .....	86
Figure 4.28 Selection of Features from Gradient Boosting .....	87
Figure 4.29 Selection of Features from Logistic Regression.....	88
Figure 4.30 Selection of Features from Random Forests .....	89
Figure 4.31 Recursive Feature Elimination using Decision Trees (soft Voting).....	90
Figure 4.32 Recursive Feature Elimination Based on Gradient Boosting (soft Voting) .....	91
Figure 4.33 Recursive Feature Elimination Based on Logistic Regression (soft Voting).....	92
Figure 4.34 Recursive Feature Elimination Based on Random Forests (soft Voting).....	93
Figure 4.35 Chi Square Feature Selection Technique Performance .....	94
Figure 4.36 FCI Based Feature Selection Technique Performance .....	95
Figure 4.37 Selection of Features from Decision Trees .....	96
Figure 4.38 Selection of Features from Gradient Boosting .....	97
Figure 4.39 Selection of Features from Logistic Regression.....	98
Figure 4.40 Selection of Features from Random Forests .....	99

## List of Tables

Table 2.1 Analysis of ML Approaches in Comparison .....	25
Table 2.2 A Study of DL Methodologies in Comparison .....	30
Table 3.1 Information about the thyroid disease dataset from the certified institution .....	43
Table 3.2 Details about the opensource KEEL dataset related to thyroid disorder. ....	45
Table 3.3 Attribute details selected by the feature selection techniques at run time. ....	53
Table 3.4 Description of outliers observed with MAE scores in the appointed features.....	55
Table 4.1 The accuracy of the bagging classifiers as a function of training and prediction time .....	62
Table 4.2 The accuracy of the boosting classifiers as a function of training and prediction time. ....	63
Table 4.3 Soft and hard voting performance metrics for homogeneous ensemble classifiers on dataset 1. ....	75
Table 4.4 . The accuracy of the bagging classifiers is described in terms of training and prediction time. ....	78
Table 4.5 The accuracy of boosting classifiers is described in terms of training and prediction time. ....	79
Table 4.6 On dataset 2, performance evaluation metrics for the homogeneous ensemble classifiers' soft and hard voting.....	102
Table 4.7 Comparison of accuracies with existing studies on dataset2.....	103
Table 4.8 Comparison of accuracies with existing studies on dataset1 .....	103

## LIST OF PUBLICATIONS

- 1. Tehseen Akhtar**, Asif Ur Rehman, Mohsin Jamil, Syed Omer Gilani.  
"Impact of an Energy Monitoring System on the Energy Efficiency of an Automobile Factory: A Case Study."  
*Energies* **2020**, 13, 2577.  
DOI: [10.3390/en13102577](https://doi.org/10.3390/en13102577)  
[IF: 3.004, HJRS Cat W]  
URL: <https://www.mdpi.com/1996-1073/13/10/2577/pdf>
- 2. Akhtar, T.;** Gilani, S.O.; Mushtaq, Z.; Arif, S.; Jamil, M.; Ayaz, Y.; Butt, S.I.; Waris, A.  
"Effective Voting Ensemble of Homogenous Ensembling with Multiple Attribute-Selection Approaches for Improved Identification of Thyroid Disorder."  
*Electronics* **2021**, 10, 3026.  
DOI: [10.3390/electronics10233026](https://doi.org/10.3390/electronics10233026)  
[IF: 2.397, HJRS Cat X]  
URL: <https://www.mdpi.com/2079-9292/10/23/3026/pdf>
- 3. Akhtar, T.;** Gilani, S.O.; Mushtaq, Z.; Arif, S.; Jamil, M.; Ayaz, Y.; Butt, S.I.  
Ensemble based effective diagnosis of thyroid disorder with various features selection techniques."  
*IEEE The Second International Conference of Smart Systems& Emerging Technologies (SMARTTECH 2022)*, pp. 1-6, (2021).  
DOI:  
[Scopus Indexed]  
URL:

## **CHAPTER 1 INTRODUCTION**

### **1.1. Introduction**

This chapter describes the introductory context & problem description, as well as why this dissertation was written to address it. According to the World Health Organization, thyroid disorders are the second most common endocrine condition worldwide. Hypothyroidism affects 1% of the population, while hyperthyroidism affects 2% of people. The prevalence of men is one-tenth that of women. Secondary to pituitary gland failure or tertiary to hypothalamic dysfunction, hyper- and hypothyroidism can be induced [1]. Goiter or active thyroid nodules may become common in some locations due to a shortage in dietary iodine. Endogenous antibodies can inflict a lot of damage in the thyroid gland, which is why it's a risky area to have them (autoantibodies). As one of the most common causes of medical diagnosis and prediction, the beginning of thyroid disease can be difficult to predict. The thyroid gland is a vital part of the human body [2]. Thyroid hormones are responsible for regulating the body's metabolic rate. There are two main types of thyroid disease, hyperthyroidism and hypothyroidism, which both affect the production of thyroid hormones, which regulate metabolic rate. In order to do analytics on the risk of thyroid disease in patients, data purification techniques were used [3]. According to this paper, machine learning plays an important role when it comes to disease prediction, and the information acquired from UCI's machine learning repository is used to create analysis and classification models for thyroid disease. It is critical to have a solid foundation of information that can be used as a hybrid model for difficult learning activities, such as medical diagnosis and prognosis tasks [4].

### **1.2. Background**

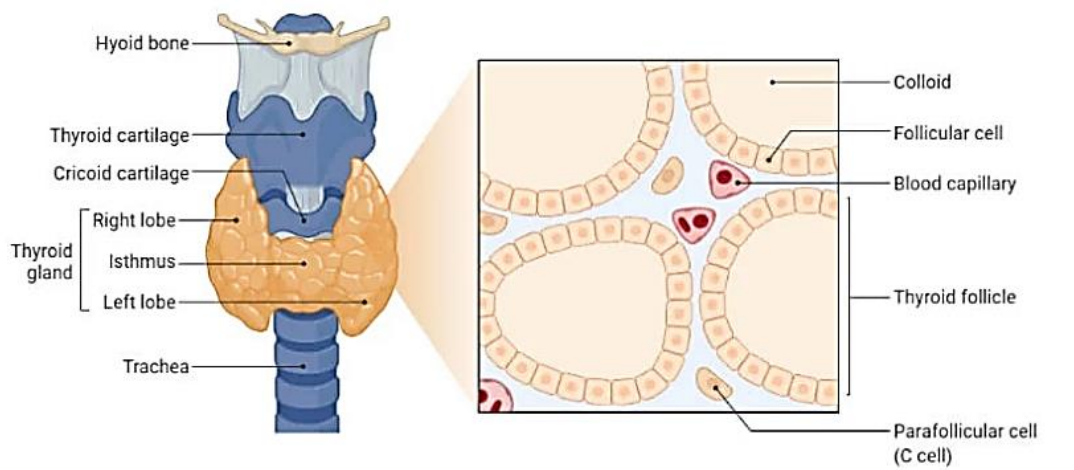
The thyroid gland produces and stores hormones that aid in the regulation of heart rate, blood pressure, body temperature, and the rate at which food is turned into energy. Thyroid hormones are required for the proper functioning of all bodily cells [5]. They aid in the regulation of body growth and the rate of chemical processes (metabolism). Thyroid hormones also aid in the growth and development of youngsters. The thyroid gland is located below the Adam's apple, wrapped around the trachea, in the lower part of the neck (windpipe). It resembles a butterfly,

with two wings (lobes) connected by an isthmus in the centre. The thyroid produces hormones with the help of Iodine; a mineral contained in some meals and iodized salt [6]. Thyroxine (T4) and triiodothyronine (T3) are the two thyroid hormones that are most significant (T3). The pituitary gland produces (TSH), which stimulates thyroid hormone production. Calcitonin is a thyroid hormone that regulates calcium levels & induces bone matrix to deposit calcium. [7].

### **1.3. Structure Of Thyroid Glands**

The thyroid gland is the first endocrine organ to emerge during foetal development. It is responsible for hormone production. In the first four weeks of pregnancy, it forms as an epithelial diverticulum, which spreads inferiorly starting at week five as the embryo develops [8]. It is formed by the foregut endoderm surrounding the base of the primitive tongue. By the seventh week of pregnancy, the embryo has developed into its final shape and size. [9] To begin with, the thyroid is made up of many follicles that are enclosed by a fibrous capsule that divides the parenchyma into multiple lobules, each of which is divided by septae. Additionally, the septae contain the nerves and blood vessels that nourish each lobule of the body. Follicles with an average diameter of 200  $\mu$ m are found in every lobule, and they are lined with simple, flat to low columnar epithelium that varies in height depending on the level of functional activity present in each lobule [10]. The higher level of functional activity present in a lobule corresponds to higher levels of follicular epithelium height in the lobule. One variant of the Hürthle cell has an abundance of granular cytoplasm, and the other has a homogeneous black nucleus that is centrally located. A Hürthle cell can be distinguished by the fact that it has a homogeneous black nucleus that is centrally located, and it can also be distinguished by the fact that it has a centrally located homogeneous black nucleus. Small follicles that extend into the core portions of bigger follicles are seen throughout the thyroid, and they should not be mistaken with papillary structures. Sanderson pollsters are found in a scattering across the thyroid, and they are not to be confused with papillary structures. In addition to colloid, a viscous fluid composed mostly of the thyroid hormone precursor protein thyroglobulin, follicles contain other substances such as keratin. It is possible to have several follicles in a single cell. It is thought that the normal thyroid gland may store thyroglobulin in its colloid for up to three months at a time. A derivation of the neural crest that develops through the ultimo branchial body, the perifollicular cell (also known as a C cell) is the thyroid's final cell type.

They are referred to as Para follicular cells because of the way they cluster together within and between follicles. They are found in the mid- and upper regions of the lobes, where they produce the greatest number of clusters of cells. The figure below shows the structure of thyroid glands in human body:



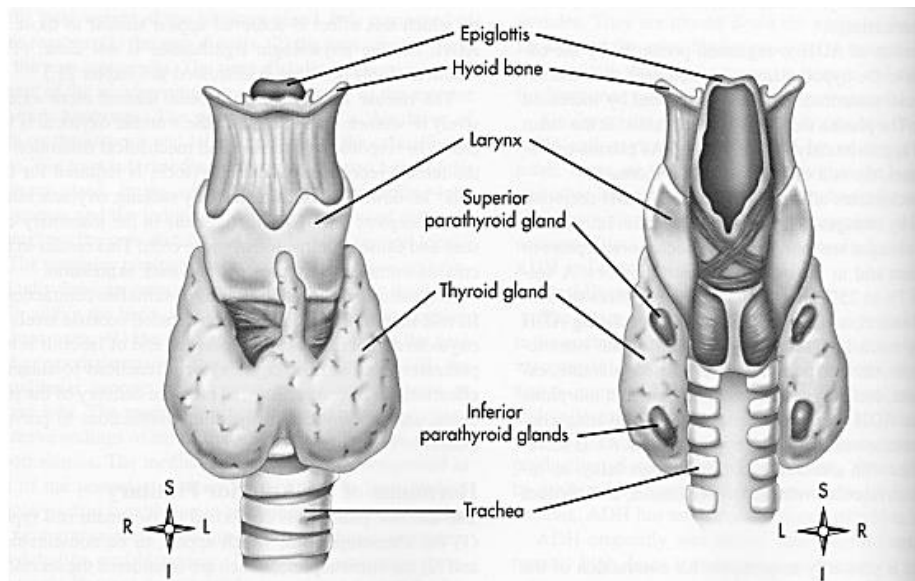
**Figure 1.1. Structure of Thyroid Gland in Human Body**

#### **1.4. Location Of Thyroid Gland**

The thyroid gland is positioned among the C5 and T1 vertebrae in the occipital region. It's between C5 and T1. It looks like a butterfly because it has two different sections (left and right). The trachea's superior rings and cricoid cartilage wrap around the thyroid lobes. This keeps the thyroid gland inside its normal place. The gland is in the visceral compartment of the neck (along with the trachea, oesophagus and pharynx). This compartment is defined by the pretracheal fascia.

Some of the important functions of thyroid gland are given below.

- They are involved in natural growth and development.
- They are also very important for myocardial contraction.



**Figure 1.2 Position of thyroid gland in human body**

- They are also responsible for the regulation of bone formation.
- They are also important for the development of white adipose tissues.

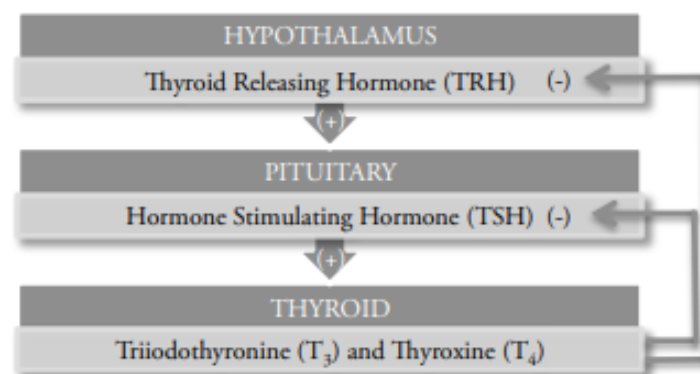
### **1.5.Function Of Thyroid Gland**

In short, they are very vital for the normal function of human body and metabolism. The fact that thyroid hormones have pleotropic effects makes them absolutely necessary for the survival and optimal functioning of the human body.

#### **1.5.1 Mechanism Of the Function of Thyroid Gland**

The thyroid gland, like other glands of the endocrine system, is controlled by a feedback loop that includes the hypothalamus, the pituitary, and the target gland (thyroid hormone) (the thyroid). As the name suggests, the HPT axis is a network of connections that connects the hypothalamus to the pituitary gland and the thyroid gland. Thyrotrophic-releasing hormone (TRH) is produced by the hypothalamus and is discharged into the venous system, which drains to the pituitary gland. The hypothalamus is also responsible for the production of insulin. Thyroid stimulating hormone (TSH), also known as thyrotropin, is produced and secreted by the pituitary gland when TRH attaches to receptors on the surface of thyrotrophic cells. Thyrotrophic stimulating hormone links to TSH receptors in the follicular cells of the thyroid gland, inducing

the production and secretion of thyroid hormones T3 and T4, as well as the production and release of T4 hormones. The thyroid gland, like all endocrine glands, has both negative and positive feedback systems; nevertheless, the negative feedback system predominates in the thyroid glands functioning. Thyroid hormones are created by the thyroid gland, and when they are released, they negatively feed back to the hypothalamus and pituitary, causing them to shut down any future thyroid hormone synthesis. This negative feedback loop is ultimately responsible for keeping hormone levels in the bloodstream generally constant.



**Figure 1.3 Thyroid gland positive (+) and negative (-) feedback mechanism.**

The thyroid gland absorbs iodine from the meals you eat to produce two primary hormones: thyroid hormone and thyroxine.

- Triiodothyronine is a hormone that is produced by the body (T<sub>3</sub>)
- Thyroxin is a hormone that regulates thyroid function (T<sub>4</sub>)

It is crucial that the levels of T3 and T4 are neither too high nor excessively low. It is necessary to maintain the balance of T3 and T4 in the body by communicating with the hypothalamus and pituitary, two glands in the brain. (TRH) is produced by the hypothalamus, which signals the pituitary, which tells the thyroid gland whether to increase or decrease the production of (TSH). The pituitary gland generates more (TSH) when the levels of T3 and T4 in the blood are low. (TSH) tells the thyroid gland to create more thyroid hormones when the levels of these hormones are low. Elevated thyroid hormone (T<sub>3</sub>) and thyroid hormone (T<sub>4</sub>) levels lead the



pituitary gland to transmit fewer (TSH) signals to the thyroid gland, causing the thyroid gland to release less of these hormones. The relationship between the pituitary gland in the brain and the generation of thyroid hormones can be better understood by looking at diagram 4, which is shown below. This demonstrates that the amount of iodine in our food activates the brain pituitary gland, which then sends signals to the thyroid gland to produce thyroid hormone.

## **1.6. Diseases of Thyroid Glands**

In the healthcare industry, the evolution of computational biology is being employed. It enables the acquisition of previously stored patient data for the purpose of disease prediction. There are prediction algorithms that can be used to diagnose the disease at an early stage, and they are accessible. Despite the fact that medical information systems contain a large number of datasets, there are only a few intelligent systems that can easily analyse disease. Over time, machine learning algorithms have begun to play an increasingly important role in the development of new models, particularly in the resolution of complicated and non-linear issues [11]. To override the features that can be selected from multiple datasets and utilised in classification in healthy patients to make the prediction of a disease as accurate as possible, prediction models are employed in any disease. Failure to do so can result in a healthy patient receiving unneeded therapy as a result of a misclassification. In humans, the thyroid gland is an endocrine gland located in the human neck beneath the Adam's apple. It is responsible for the secretion of thyroid hormone, which has an effect on the pace of metabolism and protein synthesis. In order to determine how quickly our hearts, beat and how quickly we burn calories, the thyroid hormones are utilised. The thyroid gland secretes two types of active hormones, levothyroxine (T4) and triiodothyronine (T3), which are responsible for metabolism (T3). Several hormones are involved in the regulation of the body's temperature. These also contribute in the carrying and transmission of energy throughout the body, and they are critical in the management of protein. Iodine is thought to be the primary structural component of the thyroid gland. It has been reduced to a sham in a few specific problems [12]. Hyperthyroidism can result from a lack of these hormones in the body. There are numerous causes of hyperthyroidism and underactive thyroids, each with their own set of symptoms. There are many different types of drugs, such as thyroid surgery, which is susceptible to ionising radiation, constant discomfort of the thyroid, iodine insufficiency, and a lack of the enzyme

necessary to produce thyroid hormones. In many cases, the amount of thyroid hormone present in the bloodstream distinguishes a thyroid gland problem from a normal thyroid gland. When the thyroid gland generates excessive levels of thyroid hormones, a condition known as hyperthyroidism occurs, and when the thyroid gland produces inadequate amounts of thyroid hormones, a disease known as hypothyroidism occurs. These disorders are further differentiated based on the endocrine gland that is responsible for the problem, which can be classified as primary, secondary, or tertiary, depending on the classification. The thyroid gland, on the other hand, is responsible for the majority of ailments categorised as primary. Diseases classed as secondary or tertiary are caused by abnormalities with the pituitary or hypothalamus glands, respectively. Hypothyroidism, hyperthyroidism, thyroid sick syndrome, and drug-induced thyroid diseases are among the conditions that can affect the thyroid gland. As a result of recent improvements in highly sensitive assays for the detection of thyroid stimulating hormone, illnesses can now be classed as either overt or subclinical, depending on their clinical manifestation. Subclinical disorders are found before the patient's signs and symptoms present themselves, as well as before the patient's thyroid hormone levels become abnormally high. TSH levels are one of the few abnormalities that can be noticed in subclinical disorders, and this is one of the most common. In other words, thyroid hormone levels will be within normal ranges after treatment. The thyroid gland produces two active thyroid hormones: total serum thyroxin (T4) and total serum triiodothyronine (T3), which work together to regulate the body's metabolism. T4 and T3 are the two active thyroid hormones produced by the thyroid gland. These hormones are required for the proper functioning of each cell, each tissue, and each organ, as well as for overall energy yield and regulation, as well as for the production of proteins in the context of maintaining body temperature [13, 14]. When it comes to thyroid illness diagnosis and treatment, the functional behaviour of the thyroid disease is the central concept, and it is the determining factor in the majority of thyroid diseases. Euthyroidism, hyperthyroidism, and hypothyroidism are the terms used to classify thyroid disease. These terms refer to thyroid hormone levels that are normal, excessive, or deficient in one or more of these conditions. The state of euthyroidism is characterised by the thyroid gland's normal production of thyroid hormones as well as the maintenance of normal thyroid hormone levels at the cellular level. The clinical symptom of hyperthyroidism is caused by an excess of thyroid hormones in the circulation and intracellular thyroid hormones. Hypothyroidism is most often

due to deficiency of thyroid hormone production and poor treatment alternatives. [15], [16]. Curing disease is a constant worry for healthcare professionals and providing an accurate diagnosis at the appropriate time for a patient is critical. Recently, various enhanced diagnosis methods have been developed that may create a standard medical report as well as an additional report based on symptoms [17]. Many various questions, such as "what are the causes of thyroid disease?" and "which age group of people are afflicted by thyroid disease?" and "what is the most appropriate treatment for a disease?" may have answers when machine learning methods are used in conjunction with other techniques. Following the processing and implementation of specific approaches, health care data can be utilised to generate data for the diagnosis and treatment of diseases more efficiently and precisely, resulting in better decision-making and the reduction of the risk of death [18]. When employing machine learning techniques, it is possible to deal with a significant volume of data. Models of classification and distinction are well suited for the classification and differentiation of data classes. Classification processes are capable of dealing with both numerical and category values at the same time. There are two steps to classifying a tuple: first, the model is formed on the basis of some training data, and then the unknown tuple is presented to the model for categorization into a class label. The classification system has a significant impact on human behaviour. Contrasting various categorization algorithms is a difficult task that is highly dependent on the characteristics of the data set under consideration. When it comes to classification difficulties, logistic regression, decision trees, and k-nearest neighbour have all earned a well-deserved reputation in the statistics world [19].

### ***1.6.1 Major types of diseases of thyroid gland:***

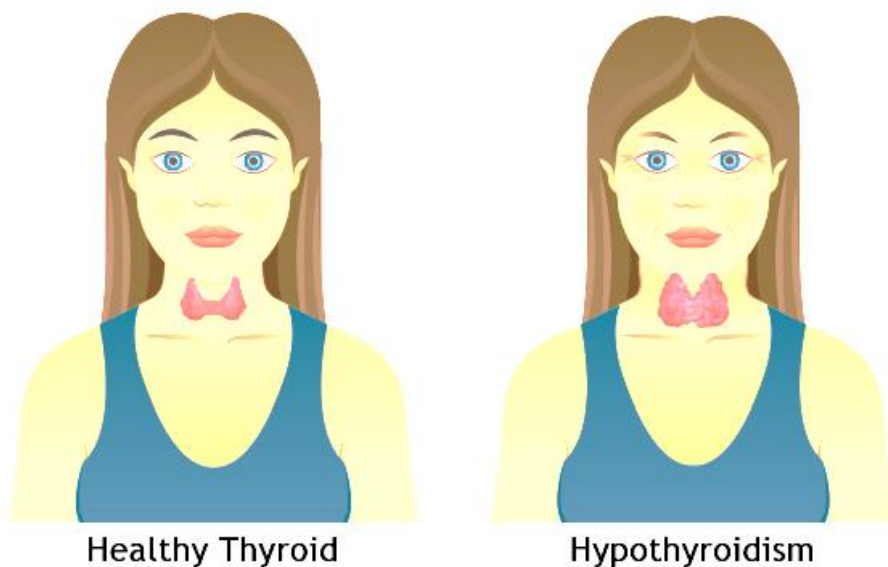
There are two most common diseases of thyroid glands which are

- a) Hypothyroidism
- b) Hyperthyroidism

#### *1) Hypothyroidism:*

It is hypothyroidism when the thyroid hormones are not available to the tissues in sufficient levels to function properly. While the vast majority of hypothyroid cases are secondary in

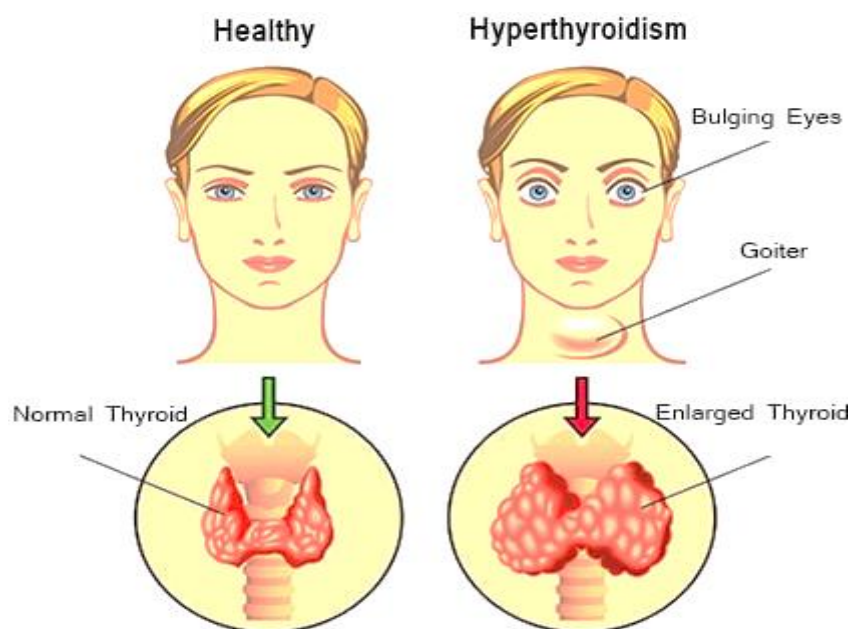
nature, the vast majority of primary hypothyroid cases occur as the result of a problem with the thyroid gland itself, such as disorders or treatments that damage the gland or interfere with the production of thyroid hormones, rather than as a result of a problem with the thyroid gland itself. To a much lesser extent, it is probable that they are caused by pituitary and/or hypothalamic disease, which leads in TSH and/or TRH shortages. In locations where there is insufficient iodine, the most prevalent cause of hypothyroidism is autoimmune thyroiditis (autoimmune thyroiditis). Hashimoto's thyroiditis is caused by the immune system's destruction of thyroid tissue, which results in thyroid gland inflammation and decreased thyroid hormone production in the individual who has been diagnosed with the condition. The fact that this condition is autoimmune in origin means that it is most found in association with other immune-based disorders. Aspects of it that are distinctive include the existence of circulating anti thyroid antibodies. In the United States, this condition affects between 2 and 15 percent of the population, and women are more likely than males to be affected by this condition, which indicates a gender bias.



**Figure 1.4 shows the difference between the normal gland hyperthyroidisms**

## 2) *Hyperthyroidism:*

As a result of a variety of various conditions, it is possible to develop hyperthyroidism. All of these disorders are associated with an excessive production of thyroid hormone. Thyrotoxicosis is a term used to describe a clinical condition characterised by increased thyroid hormone concentrations in the bloodstream. When it comes to underlying causes of hyperthyroidism (thyrotoxicosis), Grave's disease is by far the most common in the United States. Grave's disease is an autoimmune illness in which antibodies generated against thyroid-stimulating hormone receptors in the thyroid gland induce an overproduction of the thyroid hormones T3 and T4, resulting in a condition known as hyperthyroidism. Females are more susceptible to Grave's disease than males, despite the fact that the disease has a low prevalence rate in the general population (0.3 to 0.6 percent).



**Figure 1.5 Effect of hyperthyroidism**

### ***1.6.2 Signs and the symptoms of diseases:***

#### *1) Hypothyroidism:*

The presence of signs and symptoms that are associated with high TSH levels as well as low T3 and T4 concentrations in the blood is nearly certain when the disease is openly displayed. When there are no symptoms or signs of hypothyroidism, thyroid hormones are present within the normal range, and a high TSH level is present, subclinical hypothyroidism is distinguished from clinical hypothyroidism. Using the pharmaceutical levothyroxine as a treatment for hypothyroidism, patients are given oral replacement therapy. It is possible to reverse the test results as well as the clinical signs and symptoms with the injection of thyroid hormones orally.

#### *2) Hyperthyroidism*

It is likely that when hyperthyroidism is openly exhibited, there will be a drop in (TSH) and an increase in thyroid hormone (T3 and T4) concentrations in the blood. All other thyroid hormones are within normal limits in patients with subclinical hyperthyroidism, with the exception of the TSH level, which is abnormal. The treatment of hyperthyroidism can be accomplished through the use of antithyroid medicines, which work by decreasing the production of thyroid hormones, or through the use of surgical thyroidectomy or radioactive iodine, which work by reducing the amount of hyperfunctioning thyroid tissue. When treating the symptoms of hyperthyroidism, doctors frequently prescribe beta-blockers to help alleviate the symptoms while they wait for the other medications to take effect. In the case of people on antithyroid drugs, it is recommended that they have their free T4 (fT4) levels evaluated four weeks after commencing treatment and every four to eight weeks until they achieve normal thyroid function.

### ***1.6.3 Methods for the detection of thyroid gland diseases:***

There are many tests and methods which are used to diagnose the disease of thyroids. Some of them are given below:

#### a) Clinical evaluation

- b) Blood tests
- c) Thyroid neck check
- d) Imaging test
- e) Biopsy
- f) Intelligent Testing

*1) Clinical evaluation:*

When it comes to the detection and diagnosis of thyroid illness, a full clinical assessment is essential and basic to success. The involvement of a healthcare provider, such as a general practitioner or an endocrinologist, is required in order to conduct a clinical evaluation of your thyroid gland. If the patient is subjected to a comprehensive clinical evaluation, the healthcare provider would typically ask him or her the following question: The doctor used to listen for the sound of the thyroid and other organs and inspect the patient's neck for swelling and other abnormalities. Take the temperature of your body. It is possible to have a decrease in body temperature as a result of an underactive thyroid; on the other hand, hyperthyroidism can cause a minor increase in body temperature. Take a look at the general amount and quality of hair on the topic in question. Changes in the texture of one's hair, as well as hair loss and breaking, are all symptoms of hyperthyroidism or hypothyroidism, respectively.

*2) Blood test:*

A blood test to detect thyroid hormone levels will most likely be ordered by a healthcare professional if he or she suspects that someone has a thyroid disease. Your hormone levels are measured using the following test, which they do. Thyroglobulin/thyroid-binding globulin (TBG), thyroid-stimulating hormone (TSH), thyroid-receptor antibodies (TRAb). Thyroid hormones in their many forms, as well as proteins that might stimulate or reduce thyroid hormone production can all be measured using these procedures. Once the data have been compiled, the healthcare provider will examine them to determine the type and cause of the thyroid disease that the individual is suffering from.

*3) Thyroid Neck Check*

Despite the fact that it is not considered diagnostic, you can check your neck for lumps and growth on your own. An examination of the neck is not generally regarded as reliable or accurate when compared to other testing methods that are available. Although it is unlikely, it is possible to be diagnosed with thyroid illness even if your neck feels perfectly normal. A self-check, on the other hand, is not harmful and is simple and straightforward to perform. As you take a sip of water, you'll softly feel about in your neck with your fingertips to check for any lumps.

#### *4) Imaging Test*

Thyroid imaging tests can be performed to detect thyroid enlargement, atrophy, or nodules in the context of a thyroid disease diagnosis. Nodules, masses, and enlargement of the thyroid gland can all be seen with thyroid ultra-sonography. Ultrasound can be used to determine if a patient's thyroid nodule is a fluid-filled cyst or a mass of solid tissue, which might be helpful to the healthcare professional. A CT scan can provide a picture of goitre or bigger thyroid nodules, if they are present. An MRI can also be used to determine the size and form of your thyroid.

#### *5) Biopsy*

Thyroid lumps and nodules that are suspicious are evaluated with a needle biopsy, commonly known as a fine needle aspiration (FNA) biopsy. After inserting a fine needle straight into the nodule, cancerous cells are removed and tested in a laboratory to determine whether or not the nodule is malignant. Ultrasound guidance is sometimes used to guide the needle position during a biopsy by some healthcare providers. Despite the fact that 95 percent of thyroid nodules are not malignant, FNA, in conjunction with some additional tests such as the Veracyte Afirma test, can improve the accuracy of biopsy results and may prevent you from having unnecessary surgery for nodules that turn out to be harmless.

#### *6) Intelligent Systems*

There are a variety of different tests and methods that healthcare providers might use to diagnose thyroid dysfunction. These tests are considered controversial by conventional



practitioners, although some of them are acknowledged and in use by alternative, integrative, and holistic practitioners. These tests include Iodine patch tests, Saliva testing, Urinary testing, and Basal body temperature testing, among other procedures. It has not been determined whether or if these tests are reliable or valuable. Research investigations and a review of the literature reveal that there has been minimal progress in the classification methods for patients whose thyroids have been removed. Many people are familiar with the categorization methods that were employed in this study. As a result of the aforementioned concerns, this paper shows how to classify people who have been pruned by thyroid disease utilising logistic regression classification, decision tree classification, and closest neighbour classification as classification methods. The thyroid illness database categorises people who have thyroid disease into a number of different groups. Using machine learning approaches, a wide range of complex issues can be resolved [20]. We investigated and classified thyrotoxicosis here since ml algorithms perform a significant role in correctly identifying hypothyroidism, and because these techniques are high-performing & efficient (and therefore aid in classification) and so aid in classification. Computer learning and ai have been used in medicine from the very beginning of the profession. [21] Interest in machine learning-based healthcare solutions is on the rise. According to some analysts, machine learning will become a common practise in the healthcare business over the next few years [22]. Hypothyroidism and hyperthyroidism are two of the most frequent thyroid illnesses in the general population, and the most recent research focuses on the categorization of thyroid disease in these two disorders (hyperthyroidism and hypothyroidism). Naive Bayes, Decision Trees, Multilayer Perceptron, and Radial Basis Function Networks were tested and assessed in a variety of scenarios. In the study, Naive Bayes proved to be the best accurate classification model. All the models outlined above were shown to be very accurate at classifying data in this investigation, with the Decision Tree model scoring the highest classification score. It was important to use machine learning datasets from Romania and the University of California, Irvine to construct and assess our classifier (UCI). The KNIME Analytics Platform and the Weka data sets are used together. The framework for the construction and assessment of categorization models was built using data mining methods. In an effort to improve accuracy, a study of the literature found that data mining methods were utilised in much research on thyroid classification. Many models may be used to help identify thyroid dysfunctions such as hyperthyroidism or hypothyroidism following the findings of this

research. Across the board, it was shown that decision-tree models were the most effective tools for categorising data. Thyroid disease is a tough topic for medical research since it is difficult to pinpoint the source of the condition, which is a crucial aspect in medical diagnosis and prediction. Our bodies' thyroid glands are among the most important organs in our bodies' systems of metabolism. This hormone is responsible for the preservation of metabolic equilibrium. The release of thyroid hormones regulates the metabolic rate of the body. The production and release of thyroid hormones are affected by these two of the most prevalent thyroid illnesses. Thyroid hormones are essential for regulating the body's metabolism and are produced and released by both of these conditions. With the use of data purification procedures, analytical approaches may now be utilised to predict the likelihood of thyroid disease in patients. Disease prediction and classification techniques are aided greatly by machine learning.

### **1.7. Research Motivation**

As the thyroid diseases are a big problem at the present time. Also, it is not very easy to detect it in early stage, so it becomes very dangerous and painful. There are many draw backs in the detection methods of the disease of the diseases of thyroid gland. Also, it is not very cheap to detect and cure the thyroid diseases. The main objectives of our research are

- To make the detection of the diseases more accurate
- To make the detection of disease more cheap
- We can also make early detection of thyroid disease

In order to diagnose thyroid illness, we may use ml techniques. We can use machine learning to identify thyroid gland disorders.

### **1.8. Detection of thyroid gland disease using machine learning:**

There are many machine-learning techniques which can be used for the detection of the diseases of thyroid gland. In [23] author provide information about the prediction and providing medication for thyroid disease using machine learning. Machine learning is very well suited for dynamic learning tasks such as medical diagnosis, disease prediction, etc. Machine learning is essential in the disease prediction process, as well as in the investigation of classification

models for thyroid disease, both of which are now under investigation. Fundamental pattern recognition methods are used to identify and inhibit thyroid hormone production. The SVM is employed in order to forecast the approximate likelihood of a thyroid patient. Author [24] used machine learning technique for the interactive thyroid disease prediction. In order to enable the prediction, ML Algorithms, SVM, and decision tree algorithms were all employed to anticipate the estimated chance of a patient having thyroid ailment. Several machine-learning technologies and diagnoses for avoiding thyroid illness were also explored. [25] contains information regarding the author's utilisation ml algorithms to forecast thyroid disease. Researchers in Iraq utilised 8 different algorithms to diagnose thyroid problems in individuals at a public hospital, with favourable results. Random Forest, Decision Tree, Naive Bayes, Logistic Regression, K-Nearest Neighbours, Multilayer Perceptron (MLP), and Linear Discriminant Analysis were among the machine learning algorithms used in this study. Among the algorithms tested, the MLP algorithm had the highest accuracy (95.73 percent) and the lowest accuracy (98.93 percent). In [26] provide information about the use of machine learning techniques for thyroid disease diagnosis. They discuss the role played by several neural network models in the identification of thyroid dysfunctionality over the past two decades. There has been an investigation into methods and neural network models that describe the characteristics of thyroid gland non-function, its autoimmune status, and the various aspects of thyroid disease. Thyrotoxicosis (Thyroiditis) is an unending and complex infection that can occur as a result of elevated (TSH) levels or as a result of problems with the thyroid organ itself. In [27] author provide information about the identification and classification of thyroid diseases using deep learning methodology. Thyroid disorders such as hyperthyroidism and hypothyroidism are among the most frequent diseases of the body's thyroid gland. All of your metabolic functions, including digestion and energy conversion, are controlled by the thyroid, which means it is in charge of them all. It is possible to save many lives by detecting thyroid disease at an early stage. There will also be the development of a web application in which a scanned image of the inclusion will be used to eliminate both the most time-consuming thyroid type and the patient investment. According to certain theories, thyroid hormones are responsible for metabolic regulation. In [28] author use Extreme Learning Machine for the diagnosis of thyroid disease. Principal component analysis (PCA) combined with an extreme learning machine (ELM) has been developed to aid in the diagnosis of thyroid disease. The

system was developed to assist in the diagnosis of thyroid disease. A total of three steps are included in the PCA system. The first stage use PCA to create a new feature set that is the most discriminative possible. The second stage transitions to the second stage, which has as its goal the production of models. It can be regarded as a highly effective instrument for the diagnosis of thyroid illness.

## **1.9. Report Structure**

The thesis is subdivided into many chapters, each with its own set of subheadings.

**Chapter 1:** This chapter provides an outline of the thesis as well as background for the issue. This is crucial for understanding the remainder of the theory. The next section provides a summary of the numerous strategies employed in this thesis, including the many ways used to detect thyroids and to automate the detection system.

**Chapter 2:** The second chapter is a review of the literature. This chapter aids in the review of the research's fundamental to advanced principles. It examines and assesses previously published similar research initiatives. To create computer vision algorithms and generate predictions, machine learning methodologies have been applied. We examined a variety of papers in order to identify a research gap, which are mentioned in the following section.

**Chapter 3:** The third chapter of the study is allotted to the methodology. It discusses the objective of the research, as well as the methods and design used. It provides an overview of the thesis-writing process and addresses any pertinent ethical issues. This chapter describes the research strategy that was employed to execute this project. It entails the following steps: dataset collection, raw data processing, data cleaning, data pre-processing (fixing dataset imbalance and handling outliers), feature engineering, model construction (using a ml algorithm), and performance assessment of the produced models.

**Chapter 4:** Chapter 4 elaborates on results. This chapter examines the findings, contrasts them to those already published, and contributes to their assessment. This chapter illustrates the use of ml algorithms to an imbalanced thyroid disease diagnostic dataset.

**Chapter 5:** The conclusion is presented in the fifth and last chapter. Objectives, outcomes, and limitations of the study are re-stated as well as implementations, limitations, and future research ideas are examined.

## **CHAPTER 2 LITERATURE REVIEW**

### **2.1. Thyroid Disorders Prediction Using Machine Learning**

A wide range of machine learning techniques have been applied in recent studies to characterise and quantify hypothyroidism and disease diagnosis. We'll go over the methods and outcomes of previous studies that employed machine learning classification models to diagnose thyroid problems.

In article [29], it was chosen to adopt the supervised learning approach. These strategies for detecting thyroid disease type were developed using the ANACONDA software and Python programming language. In addition to SVM and KNN, we used Logistic Regression and Decision Trees as well as Nave Bayes and Random Forests. We'll plot the results to examine how accurate logistic regression is compared to random forest. Patients can obtain low-cost thyroid diagnostic reports using this strategy. Three machine learning-based strategies for identifying thyroid texture are proposed in this study [30]. ML algorithms include Random Forest Classifier, SVM and Artificial Neural System. We were able to construct 30 spectroscopic energy-based features for the 2D thyroid US pictures using autoregressive modelling, which could be utilised to discriminate among thyroid from non-thyroid textures for such classifiers during training. We employed image-based attributes rather than text-based descriptions to describe thyroid tissues. The accuracy of all three approaches combined was roughly 90%.

Two well-known heuristic feature selection procedures that might be used as examples in this case study are sequential forward and sequential backward selection. The evolutionary method, a popular strategy for nonlinear optimization problems, has made selecting features easier. The categorization of thyroid diseases is done using a support vector machine (SVM). The Intelligent System Laboratory at K.N. Toosi University of Technology of the Imam Khomeini Hospital was utilised in this [31] study to look at thyroid illness data of two kinds; one dataset was taken from the UC Irvine Machine Learning Repository, and another dataset composed of original data acquired at the Imam Khomeini Hospital's Intelligent System Laboratory.

Thanks to data mining, researchers would be able to diagnose CH quicker and choose the best treatment strategy. In this cross-sectional investigation, researchers employed the Multilayer Perceptron (MLP), Iterative Dichotomiser-3 (ID3), Support Vector Machine (SVM) and Chi-Squared Automatic Interaction Detector (CHAID) for CH diagnosis. Utilizing taxonomy models, the aforementioned classifiers, Bagging (Bootstrap aggregating) and boosting procedures, it was possible to prevent the unfavourable effects of dataset imbalance on classification results. The SVM-Bagging approach achieved the best results, with 100 percent precision and specificity, 73.33% recall, with 84.62% F-measure and 99.58 % accuracy. The investigation's findings. Article [33] discusses the usage of splitting attributes of decision trees to identify thyroid disorders. For classifying thyroid nodules, the approach described here is successful and efficient. In this study, machine learning methods such as Decision Trees, (SVM) and Nave To provide a comparison of diagnosis of thyroid ailment, Bayes was utilised. For this category, the accuracy goal is 99.89%. We previously attempted to use the Decision Tree but were unsuccessful.

KNN as well as other distance functions were employed by researchers [34] to identify thyroid disease. Three distinct approaches for feature selection will be used: chi square feature selection, KNN with feature selection and L1-based feature selection. The thyroid dataset from a recognised Pakistani hospital as well as data from the KEEL repository have both been used in this investigation. The updated dataset now includes data on blood pressure, pulse rate and BMI. It varies from earlier versions due to three new additions (BP). The KNN model was tested on these two datasets using a range of distance functions. Performance assessment measures were used to analyse the classifier's efficacy. According to the data, the ideal value range for k is 1 to 5. The correctness of the Cosine distance functions and Euclidean was tested using this new dataset, and the findings revealed that they were better. This study looked at many machine learning techniques and thyroid disease preventative diagnoses. It's very important to acquire a robust knowledge base which should be added and utilised as a hybrid model for the handling hard active learning such as prognostication and medical diagnosis [35]. Based on the medical history of a patient Decision Trees, Support Vector Machines, and K-Nearest Neighbour were utilised to determine the chance of a patient to have thyroid illness.

The research [22] employs two separate machine learning techniques to find out thyroid conditions: (SVM) and random forest. Thyroid Dataset of UC Irvine was used in the investigation. The SVM model was 91 percent accurate in this situation, whereas the Random Forests technique was 89 percent accurate. The accuracy, precision, recall, and f-score of the two ml algorithms were compared. SVM is more accurate than random forest for thyroid diagnostic testing.

Logistic regression, Nave Bayes, K-nearest neighbour (KNN), support vector machines and decision trees examples of classifiers that employ feature selection methodologies [12]. Pakistan's Dera Ghazi Khan Teaching Hospital provided the information. This thyroid dataset differed from others in that it included measures of BMI, pulse rate and blood pressure. Instead of not using any feature selection in the first trial, this study employed L1 and L2 feature selection procedures. Accuracy, precision, and area under the curve were further factors to consider. According to the findings, classifiers that used L1-based feature selection beat those that used L2-based feature selection in terms of overall accuracy (logistic regression 100%, Naive Bayes 100% and KNN 97.84%).

Thyroid disease was assessed using ml methods in [36]. To make data easier to analyse, data preparation methods were applied. These procedures make it easy to determine whether or not a patient is at risk of developing the disease. Machine learning is significantly used in disease prediction. Decision trees, KNN's, logistic regression, SVM's and other approaches are used by scientists to find whether the patient will get thyroid illness (SVMs). A website to collect user input has been established in order to make good estimations concerning illness kinds.

The inclination of ultrasonic probe while capturing picture has no influence on the CAD system's characteristics, according to [37]. The researchers employed a methodology known as Two-Threshold Binary Decomposition to retrieve multiple direction independent characteristics from 60 thyroid nodules (20 malignant, 40 normal). To assess whether nodules were malignant or not, classifiers such as Support Vector Machines (SVM) and Random Forests (RF) were integrated with nodule characteristics. A group of ten persons was used to cross-validate the categorisation. The sensitivity, overall accuracy, area under the ROC (receiver operating characteristic) curve and specificity were calculated by averaging



individual patch data. The accuracy of radio frequency (RF) is 91.6 %; the accuracy of stereo vision (SVM) is 91.6 %; and the accuracy of stereo vision (RF) is 91.6 %, and so on.

[24] classified thyroid problems into three categories based on data collected from Iraqi individuals: hypothyroid, hyperactive, and inactive. Normal thyroid function, hyperthyroidism, and hypothyroidism were all characterised. Support vector machines are used in decision trees, random forests, logistic regression, naive Bayes, linear discriminant analysis and k-neighbours, in addition to multi-layer perceptron. The most accurate classifiers were Decision Trees, Random Forests, Logistic Regression, Nave Bayes, KNN and LDA, followed by Random Forests with 88% and MLP with 89%.

The author developed an artificial neural network in [25] to construct a model for differentiating tender from toxic tissues and enhancing ultrasound diagnosis accuracy. The important sonographic indications and statistically significant changes were employed as the input layer for the ANN used to predict the malignancy of nodules. The shape, size, internal composition, echogenicity, peripheral halo on ultrasonography and calcifications had a substantial connection with malignant nodule features (absent). The training accuracy score was 82.3 %, which meant it correctly identified 82.3 percent of thyroid cancer cases (AUROC=0.818) and had an 84.5 percent sensitivity and specificity. The method's findings had an accuracy of 83.1%, sensitivity of 83.8%, and a specificity of 81.82% in the validation cohort. 82.8% is the AUROC score for this study.

On clinical datasets, Decision Tree, Support Vector Machines, and Naive Bayesian Classifiers are all investigated [3]. SVM is the most widely used algorithm in machine learning. In order to evaluate the performance of a model, many scholars combined two feature selection approaches. The wrapper approach is used to assess the classifier's performance, while the filter approach is used to pick features. For the classification of the data with binary nature Fisher Discriminant Ration is effectively used for the purpose of ranking the qualities in order of relevance. The inclusion of extra characteristics to the classification model was evaluated using three performance indicators: Formula One is all about precision and accuracy. Measuring The results of this study are significant since they were unexpected.

They used multivariate pre-processed data from the University of California, Irvine in their study [26]. Kernel Based Nave Bayes was employed in a hybrid algorithm. The Curse of Dimensionality is used to reduce the amount of 21 qualities to merely ten. This raises the Nave Bayes classification accuracy to 92 percent. Medical practitioners face a challenging task in treating these disorders. Dimensionality reduction methods are used to manage large volumes of data. Many different filter options have been investigated as suitable wrapper approaches, including the Greedy Step Wise Search (GSS) technique, Information Gain technique (IG), Linear Forward Selection (LFS) technique, Best First Search (BFS) technique, Correlation Feature Selection technique (CFS) and Chi-Square (CS) technique. The WEKA tool includes the Decision Tree approach, which was used as a classifier in this study. The characteristics of study participants with breast cancer, diabetes or heart disease were examined for statistical significance. The CFS technique outperformed the competition in terms of execution time and accuracy. The CFS approach has an 89.5% on certain datasets, 93.8% on others, and 96.8% on all of them. The consistency and accuracy of the classification were found to be Exceptionally good. A total no of accurate samples was 96.7%, specific were 96.5% and sensitive were 96.6% and remaining samples had same accuracy, specificity and sensitivity which is 96.6%.

The importance of early diagnosis in terms of saving people and increasing quality of life cannot be overstated. The healthcare business gathers massive volumes of complicated data in order to uncover underlying patterns that can be used for detection, diagnosis, and decision-making. Data mining has now become a popular method for discovering relevant and distinguishing patterns in large databases. Researchers employed a number of classification approaches to identify and treat thyroid disease, including SVM and naive Bayes classifiers that simply focused on reducing dimensionality, and K-NN [1]. In each and every parameter test K-NN beats all the other classifiers.

The study's [2] major goal was to identify hyperthyroidism and hypothyroidism using neural network models and multinomial logistic regression. To see how effectively clinical signs and laboratory tests predicted future health, researchers used a range of neural network models and multinomial logistic regression on the data. These models were assessed on average in terms of area under the curve and accuracy. The accuracy predicted was found maximum by

considering all the variables in LR (Mean Accuracy = 91.4%) and neural network models (Mean Accuracy = 91.4%) (Average Accuracy = 96.3%).

HD-classifier [13] is a ML system which employs cognitively based in-memory hyper dimensional (CMHD) data to classify tumour samples from one's with non-tumour. The validity of the technique was tested using three distinct cancer DNA methylation datasets: THCA (thyroid carcinoma), KIRP (kidney renal papillary cell carcinoma) and BRCA (breast invasive carcinoma). State-of-the-art classification algorithms, such as Decision Tree-based software and (SVM), as well as Random Forest-based software, may be used to classify massive genomics datasets in distributed computing systems. It's become worse since just a few methodologies have been using clinical data from patients who have a certain disease to discover their disease genes. The author of this paper gives a methodology named Online Mendelian Inheritance in Man (OMIN), Protein-protein interaction networks and clinical RNA sequencing data were used to identify disease genes. DgSeq calculates rewiring information from clinical RNA-Seq data and generates differential networks from it. The features gathered from the PPI and differential networks are given into the logistic regression classifiers to train them. Our dgSeq detects breast cancer, thyroid cancer, and Alzheimer's disease genes with AUC values of 0.88, 0.83, and 0.80, respectively.

According to the study's [23] goal, quantitative texture analysis using machine learning would help discriminate among malignant and benign thyroid nodules. Random forest ML classifier was used to assess the sum of 306 quantitative textural parameters from 198 people for a total of 188 thyroid nodules (133 benign, 56.4 %; 102 malignant, 43.4 %). To cut down the list of acceptable features and minimise the total size, reproducibility testing and a wrapper technique were utilised. The accuracy and sensitivity of the proposed technique, as well as its area under the curve, were compared to cytopathological or histopathological findings (AUC). Results. 284 (92.2%) of the 306 texture characteristics assessed had high reproducibility (intraclass correlation was 0.80). The random forest classifier was used to successfully classify 87 malignant thyroid nodules and 117 benign thyroid nodules. The AUC of the model was 0.92. Conclusions. Using ML classification and quantitative textural analysis, nodules on the thyroid gland may be classified as benign or malignant. Our findings should be validated by a multi-

centre prospective study that uses completely independent external data. The preceding comparison study of machine learning techniques is summarised in Table 2.1.

**Table 2.1 Analysis of ML Approaches in Comparison**

<b>Reference</b>	<b>Methods</b>	<b>Data</b>	<b>Results</b>	<b>Accuracy</b>
Kour et al. [45]	SVM, RF	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 91%
Rehman et al. [12]	KNN, LR, NB	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 100%
Prochazka et al.[23]	SVM, RF	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 92%
Salman et al. [24]	Decision Tree, Multi-Layer Perceptron, Nave Bayes, Logistic Regression, K-Nearest Neighbours, And Linear Discriminant Analysis, Random Forest	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 90%
Zhu et al.[25]	ANN	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 95%
Geethaet al. [26]	Kernel Based Nave Bayes classification	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 96%
Placzek et al.[27]	Bayesian network	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 93%
Draghiciet al. [28]	Multiple-Instance Learning	Health Mining Data	Thyroid Diseases Prediction/ Thyroid Carcinoma Prediction	Highest 89%

## **2.2. Thyroid Disease Diagnosis Using Medical Imaging**

A lot of research has been done in the area of thyroid medical imaging, including a number of studies in [38]. They are employed in the diagnostic procedure. Medical imaging can be used to diagnose thyroid disease, and an up-to-date review of the studies on the subject is present in this study. Thyroid problems and thyroid diagnostics are briefly discussed here. The butterfly-shaped thyroid gland is located on the front of the neck, directly below the Adam's apple. The thyroid hormones are produced by this gland. Thyroid hormones particularly, works with the endocrine system to control metabolism. Thyroid problems include hyperthyroidism, hypothyroidism, goitre, and thyroid nodules (both benign and malignant). Ultrasound imaging is routinely used to identify and classify thyroid gland disorders. Other imaging modalities, such as CT/MRI, are also used. Radiologists and doctors' benefit from computer-assisted diagnosis (CAD) because it reduces biopsy ratios, speeds up diagnosis, and reduces effort.

Thyroid cancer is relatively common all around the globe, and it has been on the rise in Canada and United States of America in recent years. On physical examination, the majority of the patients have palpable nodules, however ultrasound investigations indicate a high number of tiny and medium-sized nodules. A biopsy sample is taken from a problematic lesion using fine needle aspiration. Because biopsies are invasive and sometimes inconclusive, many research groups have sought to create computer-aided diagnostic procedures. Radiologists manually determined clinically relevant sections in past attempts along similar lines. As a result of current artificial intelligence achievements, many novel approaches for automatically recognising certain thyroid ultrasonography characteristics are being created (AI). Our evaluation of AI's present status in thyroid cancer sonographic diagnosis is summarised in this work [39]. On the basis of approach, the numerous ways of detecting thyroid cancer are classified. We analyse how ultrasound applications can have a stronger influence on future thyroid cancer detection in this evaluation of more than 50 studies, taking into account difficulties and trends in computer-aided diagnosis and sonographic thyroid cancer diagnosis possibilities. Machine learning will be significantly used in future thyroid cancer diagnosis systems.

Thyroid ultrasonography is the subject of this study [40]. There has been the introduction of a new knowledge-based categorization system. A Densely Connected Convolutional Network is assisted throughout its learning phase by signals supplied by a group of experts in the proposed solution (Dense Net). Previously calculated feature parameters might be employed with the ensemble to construct ultrasonography domain experts through transfer learning. The number of training samples necessary would be reduced as a result. The ensemble's performance is based on a number of ImageNet-trained networks. Several experiments were conducted in order to validate the suggested strategy and give performance data for both of the experts and DenseNet depending on their expertise. Researchers discovered that a novel diagnosis method based on past information gained through consultations might be a useful tool.

Dov et al. [41] proposed MLE named approach, which presents a two-level DL methodology, as a framework to look at their influence outside of the MIL scenario. The approach assigns local malignancy ratings to informational events, which are subsequently utilised to create a global malignancy prognosis. Multiple instance labels and bags might be forecasted at the same time using single neural network's output. This algorithm exceeds all others in comparison, is as excellent as an expert at making judgements and can be used to increase or reinforce human judgments.

The influence of section-type on automated algorithms for the detection of thyroid cancer was studied by Gadermayr et al. [42]. This was accomplished by creating a two-part data collection method for every researcher. Furthermore, the examination of frozen-to-paraffin translation determine if it will aid categorization results. In order to deal with the limited number of training data while boosting classification accuracy, a specific data augmentation approach is also proposed.

Dov et al. [43] made a unique diagnostic methodology based on real-world medical practises. To begin, find and identify diagnostic picture areas containing instructive thyroid cells, which make up a tiny fraction of the whole image. Combining these regional estimations yields thyroid cancer prognoses. The deep learning approach is guided by many unique aspects of thyroid cytopathology. This approach is similar to multiple instances learning in that it uses a supervised algorithm instead of random sampling to locate diagnostically relevant areas. It is

also feasible to forecast thyroid cancer while also obtaining a human expert diagnostic score, allowing for even greater training. Experiments have proven that the proposed algorithm is as excellent as human specialists in spotting perplexing cases, and that it may be utilised for screening in the improved diagnosis of these cases.

Rivenson et al.[44] stained a microscopic picture with only a auto-fluorescence wide-field shot of a tissue sample without label, which was a time and cost-effective alternative to the traditional histochemical staining technique. Using CNN trained with generative adversarial network models, an autofluorescence picture of a tissue section that hasn't been tagged/ labbed is translated on a comparable photograph of the stained version under bright light with identical specimen. We successfully made virtual stained microscale pictures containing samples of tissue of human beings with the addition in the sections of liver tissue, thyroids, lungs, and salivary glands, and sections of thyroids and salivary glands, in the lab. This label-free virtual staining technology eliminates the need for time-consuming histochemical staining processes.

After going through a lot of research and literature we came to the conclusion that sound speed of longitudinal wave possesses diagnostic characteristics equivalent to shear wave imaging, a unique approach based on single-sided pressure-wave sound speed measurements paired with channel data was created in [15]. In this study, a fully convolutional deep neural network was used to suggest a single-sided solution for sound speed inversion. Use simulations to produce an infinite supply of real-world data, then apply what you've learned. In soft tissue, high frame rates enable longitudinal sound speed to be reversed. Prior to using real data, dummy data was utilised to evaluate the technique. With so little real-world data, several excellent findings have been made.

To discover relevant characteristics of RNA expression in this dataset, an experimental method including multiple feature selection has been applied [45]. Following that, samples were categorised using a number of machine learning algorithms that took many characteristics into consideration. The AUROC of this transcript is 0.66, indicating that it can discriminate among both early-stage and late-stage samples. This was accomplished using an AUROC-based single gene ranking method. In validation data, a panel of five protein-coding transcripts with F1

scores between 0.97 and 0.99 (confidence range: 0.91-0.99) distinguished malignant from non-cancerous samples.

This study [46] uses knowledge graph technology to join with different medical data systems to make it easier to diagnose disease. A graph of medical knowledge for thyroid disease developed in this study is used to generate an intelligent medical diagnosis. Discover the links between diverse biological elements before developing a biomedical knowledge network. The knowledge graph embedding method then converts all of the nodes and interactions in the graph into low-dimensional continuous matrices. The data from known pathological sickness relationships is utilised to train the disease diagnosis model BSTLM in a bidirectional long-short-term memory network. Experiments have demonstrated that diagnosing thyroid illness with knowledge graphs and deep learning is more accurate than using standard approaches. Medical treatment based on the knowledge graph, as depicted below, might help alleviate the country's present dearth of high-quality medical resources.

Radionics and ml are two medical diagnostics areas that have the potential to revolutionise the practise of medicine. Medical imaging algorithms based on artificial intelligence can predict tumour diagnosis and treatment response. Yes, they are. These approaches, however, still have a problem with diagnostic accuracy. Each method's data input requirements, as well as the differences and limits between them, have been carefully investigated [47]. Here's a summary of how artificial intelligence is utilised in thyroid picture analysis. A number of outstanding difficulties must be overcome before AI can be broadly deployed in healthcare. Knowledge of the risks associated with artificial intelligent based programs is essential for achieving each patient's best potential outcome.

The authors [6] propose a novel multimodal MRI-based CAD system for the identification of benign and malignant thyroid nodules. For texture learning, the suggested CAD uses convolutional neural networks (CNNs). Our system makes three significant contributions. For the first time, T2-weighted MRI and apparent diffusion coefficient (ADC) maps were combined with a CNN to model thyroid cancer. As a result, it can extract increasingly complex texture patterns from both modalities simultaneously. By adjusting diffusion gradient coefficients, a large number of scans are included into the deep learning process. Finally, for



each input, the suggested system employs a large number of channels. The suggested technique was compared to state-of-the-art CNN models available presently and different ML programs that use features which are handmade to see how accurate it is. With a diagnostic accuracy of 0.87, specificity of 0.97, and sensitivity of 0.69, the system surpassed all other techniques in the research.

Fish bones and other toxins are frequently found in the oesophagus and throat. A frequent foreign body detected in the body is a fish bone. [48] revived some old research and put up a summary. Fish bones are tough to discover while seeking for a foreign body. The danger of fish bones being stuck in the thyroid must be considered by surgeons. To prevent any misconceptions, it's advisable to keep to the patient's medical history as well as a careful fish bone. To confirm the diagnosis, computed tomography scanning (CT), ultrasound, and other diagnostic techniques might be employed.

Machine learning is becoming increasingly important to medical professions. Doctors and other healthcare practitioners need the right equipment to make accurate medical diagnosis. If we want to deliver helpful materials to medical practitioners, we will need classifiers. [9] Meta-learning techniques can be used to forecast the finest classifier for a given dataset. These techniques can be used as well. The findings show that medical diagnoses may be confidently statistically classified.

**Table 2.2 A Study of DL Methodologies in Comparison**

<b>Reference</b>	<b>Methods</b>	<b>Dataset</b>	<b>Approach</b>	<b>Results</b>	<b>Accuracy</b>
Dov et al.[41]	MLE	Malignant and Benign Thyroid Nodules	X-rays /Medical Imaging CT scan	Malignant and Benign Nodules Thyroid Carcinoma	Highest 98%
Gadermayr et al.[42]	CNN	Malignant and Benign Thyroid Nodules	X-rays /Medical Imaging CT scan	Malignant and Benign Nodules Thyroid Carcinoma	Highest 90%
Dov et al.[43]	CNN	Malignant and Benign Thyroid Nodules	X-rays /Medical Imaging CT scan	Malignant and Benign Nodules Thyroid Carcinoma	Highest 91%

Rivenson et al.[44]	Modified CNN	Malignant and Benign Thyroid Nodules	X-rays /Medical Imaging CT scan	Malignant and Benign Nodules Thyroid Carcinoma	Highest 95%
Feigin et al.[15]	CNN	Malignant and Benign Thyroid Nodules	X-rays /Medical Imaging CT scan	Malignant and Benign Nodules Thyroid Carcinoma	Highest 96%

### 2.3. Thyroid Disease Detection Algorithms Based on Clustering

Santos et al. [49] used an autoencoder with nonlinear nature to predict ageing values using the cross-sectional data collected from 1.4M individuals over 3 years. Increased morbidity has been linked to higher health-care costs across the board, In accordance with K-means clustering. Cross-sectional laboratory data, according to this study, can be used to estimate age in a variety of ways instead of just one. The pace at which individuals age and the amount of money they will have to spend in the future have a unique relationship that may be leveraged to build better approaches to avoid sickness.

The [50] employed an expert system based on fuzzy rules to diagnose hypothyroidism, the most common thyroid condition. It has been able to construct a fuzzy rule-based classifier for the identification of thyroid disease. Using a receiver operating characteristic curve, the predictive abilities of this model were compared with those of a multinomial logistic regression model (ROC). According to the data, the fuzzy rule-based technique recommended is quite successful in predicting thyroid disorders, with a 97 percent accuracy rate. Furthermore, among patients with hypothyroidism who are still in the early stages of the condition, fuzzy classification exceeds logistic regression in terms of accuracy. The use of overlapping sets in a fuzzy rule-based classifier improves the efficiency of classification and decision-making systems. Doctors who aren't familiar with modelling principles might utilise linguistic variables to help them make decisions using this method.

To make sense of the obtained data, it was pre-processed using simple NLP techniques, and chi-squared test was utilised to identify the important properties of the words [51]. After the articles had been preprocessed, iterative centroid-based unsupervised ML approach was employed to group them, this algorithm is known as K-means++. A generative probabilistic

model was utilized to discover the primary issue in every individual cluster (LDA). Medical testing, including symptoms, will be more easily recognisable, allowing patients and professionals to detect similarities and distinctions between illnesses.

#### **2.4. Methods for Detecting Thyroid Diseases that are Automated and Intelligent**

To assess the thyroid's health, its shape and size should be evaluated throughout time. Volume computation and Thyroid segmentation are two techniques that may be used to track the thyroid's health over time. There are a number of non-automated methods [52] and segmenting thyroid in the anatomy of patient is very time consuming. The accuracy and resiliency of three non-automatic segmentation algorithms were studied using freehand three-dimensional ultrasonic imaging (active contours without edges, graph cut and pixel-based classifier). These operations were found to be inefficient due to a lack of automation and machine intelligence. The quality of results of segmentation significantly improved along with automation enabled by using two ML algorithms (Convolutional Neural Network and Random Forest). The pros and cons of several algorithmic techniques are compared in this study. Finally, the thyroid volume is calculated based on the segmentation findings, and the performance of the algorithm is assessed.

Elastic light scattering (ESS) detects the spectral differences between malignant and benign nodules and can assist the surgeons' to accurately better diagnosis benign thyroid nodules before surgery. Rosen et al. [53] conducted a large prospective investigation in thyroid nodule patients and found that the ESS approach is effective. An ESS system was used to capture thyroid tissue spectra. Using spectroscopic imaging, the biopsy sample's histology was compared to that of spectroscopic imaging results. According to the manufacturer, the ESS approach has a 74 percent sensitivity, 90 percent specificity, and a 97 percent negative predictive value for distinguishing benign from malignant thyroid nodules. From the above data we can safely conclude that for the rapid diagnosis of thyroid nodules in different patients FNAB and ESS cytology can safely be utilized.

Researchers used the data contained in the Cancer Genome Atlas (TCGA) dataset to do functional analysis on differentially expressed mRNAs in order to find diagnostic lncRNAs and miRNAs that just might help explain PTC epigenetics [54]. The most effective diagnostic

lncRNA and miRNA biomarkers were discovered using Random Forest. Changes in lncRNAs and miRNAs might be used to identify PTC diagnostic biomarkers. PTC carcinogenesis is aided by epigenetic pathways, as well as before mentioned lncRNAs and miRNAs.

For the reduction of rows, The Non-Sorting Genetic Algorithm was integrated with three attribute selection data mining techniques to create a new model [55] as training and testing data. Total no of thyroid disorders under study were two having four different classes in each making total 8 classes, and total patients were 1472 out of which 500 were used for training and 972 for and with 29 different qualities and a cross validation of 5. The potential of the model obtained were evaluated using range of different benchmark like accuracy and precision. For evaluating the proposed model's applicability, a comparison with the Sequential model will be carried out, which will incorporate (traits of the given model either partially or fully).

A heated thyroid nodule is studied in [11] where its temperature distribution was determined using Finite Element Analysis. The instrumentation model was created using three-dimensional human thyroid nodule models. Visual Basic code was used to compute the temperature dispersion around a hot spot. In addition, the software includes measures for reducing thermal noise caused by temperature changes in the body. For FEA simulation, boundary values were used just like in a real life. Beginning temperatures of hotspots and its adjoint area are included. Findings of finite-element analysis assisted in the anthology of solid-state sensors for thermographic equipment. During the calibration procedure, the dynamic performance and functionality of the chosen sensors fulfilled industrial criteria. At IU Hospital, patients with Graves' disease were given the unique non-invasive diagnostic procedure, which was compared to the present method, which relies on the I Scan. The results of the new diagnostic approach were quite comparable to those acquired utilising the traditional method.

Thyroid cancer patients' treatment options are dependent on the type and stage of disease. Radio sensitivity differs across cancer cells due to their varying repair capabilities following irradiation. Thyroid cancer cells can be killed by radioactive iodine [56]. Patients' prognoses and degree of recovery after irradiation might vary greatly. To reduce unnecessary radiation exposure, predictive strategies are essential. Determining the adverse effects of radiations in those patients with thyroid cancer is very easy with our distinctive technique, in which we focus

on the amount of DNA-PK movement in cancer cells. According to our findings, the level of DNA-PK expression in thyroid cancer cell lines correlates with radio sensitivity. As a result, the level of DNA-PK's expression can be utilised to predict the outcome of thyroid cancer radiation therapy.

Thyroid glands are rarely affected by thyroid cancer. Breast cancer metastases are uncommon but renal cell carcinoma is the most common cause. If a woman has already had breast cancer, she is more likely to get thyroid nodules and thyroid cancer. When analysing a thyroid nodule, the possibility of metastatic breast cancer should be considered. A surgical extraction specimen revealed multifocal metastatic breast cancer in a 67-year-old woman who had dysphonia and dysphagia due to a multinodular goitre. Immunohistochemistry was used to rule out C cell hyperplasia and medullary thyroid carcinoma as possibilities [10]. There should be no question that a thyroid nodule in those who have had cancer in the past is metastatic.

Ectopic thyroid tissue-containing tumours can occur anywhere in the body. PTC has been discovered in ectopic tissues in a few cases, however in the normal thyroid glands of maximum no of individuals there was a primary PTC. Despite the uniqueness of isolated malignancy in other kinds of ectopic thyroid tissue with normal native tissue, PTC can be observed in an ectopic thyroglossal duct cyst. In a few rare cases of benign native thyroid glands coexisting, an unusual PTC in the midline anterior neck has been discovered, which is not consistent with thyroglossal duct cysts [57].

Thyroid hormone (TH) disruption is defined as an abnormal change in the production, function, transport, or metabolism of thyroid hormone (TH), which results in some impairment of physiological homeostasis. EDCs, such as organotin chemicals, are used in many industrial processes and ship antifouling coatings (OTCs). The consequences of over-the-counter thyroid drug usage on thyroid function, on the other hand, are still little known. If used in excessive dosages, OTCs are obesogenic and may interfere with TH metabolism. People are exposed to over-the-counter drugs that influence the thyroid axis on a daily basis. Over-the-counter drugs have been demonstrated to induce hypothyroidism in toxicology investigations. [58].

If researchers can map the brain regions that mediate sensory-perceptual processing, they will be better able to understand clinical disorders linked with abnormal processing of visceral

afferent impulses (i.e., interception). An new closed-system, electro hydrostatically driven master–slave apparatus is provided for the administration of regulated fluidic stimulations of visceral organs and interior cavities of a human body in the 3T MRI scanner. The design idea of the gadget as well as its MRI performance are discussed in depth. It was utilised to confirm the feasibility of visceral stimulation associated to detrusor distention in a functional magnetic resonance imaging (fMRI) research done on two representative patients [59]. The outcomes were positive. Because of these flaws, MR compatibility testing revealed that the device has a slight effect on imaging quality [a static SNR loss of 2.5% and a temporal SNR loss of 3.5%], as well as flow rate accuracy of 99.68 percent and delivery accuracy of 99.27 percent] and volume delivery accuracy. The recommended gadget, which has been set up to detect 5 V transistor-transistor logic (TTL) trigger signals, was used to detect MRI scanner trigger signals. This enabled for precise stimulus delivery and fMRI slice recording time. Researchers employed functional magnetic resonance imaging (fMRI) to better understand the relationship among higher visceral distension pressure and increased activity in the insula, anterior and mid-cingulate, as well as lateral prefrontal cortices, and the thalamus. Switching from manually operated to autonomously controlled MR-synchronized and MR-compatible equipment benefits clinical neuroimaging investigations of human interception.

The authors [60] introduce PheDAS (Phenome-Disease Association Study), a custom Python module that creates diagnostic EMR signatures for a disease population over time to capture system-wide co-morbidities, to improve the statistical capabilities of the PheWAS software. We're looking at the impact of merging EMR characteristics with radiological data for illnesses with dynamic and complex clinical presentations. Both of the optic nerve and diabetes have now been thoroughly investigated in these two studies, which are highlighted below. The use of EMR signature vectors in radiologically determined structural measures improves diagnostic classification for diseases of the optic nerve using elastic net regression (AUC). Despite the fact that the AUC for glaucoma has enhanced, the AUC for intrinsic optic neuropathy and optic neuropathy edoema (0.95-0.96) has reduced, while the AUC for all four disorders has increased (0.95-0.96). Although diabetes-related symptoms such as elevated blood sugar are indicated in EMR profiles, no significant variations may be identified.

Gene expression profiling is frequently employed in cancer diagnostic research, and it necessitates multiclass classification, which is a crucial bioinformatics task. Multiclass classifiers based on 1R, 1v1, or other coding schemes, as well as comparison examinations between them, have been examined in a vast number of research that aggregate binary classifiers. While the data imply that the appropriate coding depends on the context, they also suggest that this is not always the case. To answer the question, "How can we choose which coding method to perform when?" For each aggregated binary classifier in [5], the answer is to develop a multiclass classifier that employs an imaginative framework with an optimum weight value based on the observed data. Although there is no a priori solution to the optimum coding problem, our weight adjusting approach can provide a reliable outcome. The method's accuracy in different classification tasks was tested using a synthetic data set and particular gene expression profiling-based cancer detection data sets. In most instances, our technique outperforms basic voting heuristics and is on par with or better than existing multiclass predictors.

TIRADS criteria, which employ ultrasound (US) imaging to analyse visual and textural qualities, can be used to identify thyroid nodules. Composition, shape, size, and echogenicity are only a few of the factors to consider. Thyroid nodules were researched utilising geometric and morphological (G-M) criteria in order to aid doctors in making better judgments and minimise the degree of subjectivity in the latest diagnostic techniques [61]. Employing pictures from an open-access ultrasonography thyroid nodule imaging dataset, researchers discovered 27 G-M characteristics. TIRADS chose 11 standouts from this worldwide feature collection. Machine learning was used to assess the performance of each characteristic, with a score of 0 indicating benign behaviour and 1 indicating malignant activity (ML). In the taxonomy of thyroid nodules, the combination of G-M features and ML yielded excellent accuracy, specificity and sensitivity. When the outcome of the research were equated to those acquired utilising cutting-edge methodologies, it was discovered that the former performed far better. This strategy, which was employed in this study for thyroid nodule categorization in ultrasound photos, might be used to construct a CAD system for clinicians utilising simply the TIRADS system's visual qualities.

TSH must be identified in human serum samples using high-sensitivity technologies to better comprehend human physiological symptoms. Tieu et al. [62] proposed an analytical membrane-based microwave technique designed for highly responsive electrochemical immunoassay in sandwich ELISA format. It was discovered that varying dosages of TSH antigen had varied outcomes. The immunoassay identified TSH antigen concentrations as low as 0.2 mIU/L. The entire detection process took less than 60 minutes, from sample preparation to testing. The practically non-existent signal in all tested solutions proved the immunoassay's selectivity against non-specific proteins. Before a successful trial with clinical human blood samples, researchers conducted exploratory tests with typical human serum samples with TSH antigens. This membrane-based microwave-mediated analytical electrochemical immunoassay approach might be utilised to provide a low-cost, sensitive, specific, and rapid platform for TSH testing. As a result, this cutting-edge technology might serve as a potent point-of-care diagnostic tool in the future generation of biomarker detection and signalling technologies.

This study [63] describes for the first time an optofluidic lens and a microfluidic test for detecting glucose levels. A fluorescent microscope that can be carried about on a smartphone was used. The hydraulic pressure may be altered to modify the tunability of the optofluidic lens. The glucose sensor also includes a smartphone, an adjustable optofluidic lens, a microfluidic chip, a custom container, and some easily accessible optics. An enzymatic fluoresce approach with a linear detection range of 0 to 6 mM glucose was used to investigate a low detection limit of 0.33 mM glucose. These findings point to the possibility of using a small, low-cost device on-site to screen for diabetes in its early stages, as well as for other clinical diagnostic and environmental monitoring purposes. The purpose of this research [14] is to look at some of the most current results in the field of thyroid illness detection using spatiotemporal surveillance. In the relevant study, the authors addressed particular challenges and research restrictions.

Because it is inexpensive and simple to use, ultrasound electrography is becoming more popular as a cancer detection and differentiation diagnostic technique. Because conventional shear wave imaging requires a lot of power, high-end ultrasound equipment is necessary to employ contemporary shear wave imaging techniques. These techniques are similarly prone to artefacts, such as patient or sonographer movement, and operate at a low frame rate.



Longitudinal wave sound speed offers diagnostic power comparable to shear wave imaging, according to research and theory, and Feigin et al. [15] proposed an alternative approach based on single-sided pressure-wave sound speed measurements employing channel data.

Ultrafast imaging ultrasound data has recently been shown to build a vector basis that is considerably more suited for tissue and blood flow discrimination than the traditional Fourier basis, dramatically improving clutter filtering and blood flow estimate. It's still unclear if the tissue subspace/blood flow subspace border can be precisely anticipated. Using the primary aspects of the singular components, such as singular values, temporal singular vectors, and spatial singular vectors, Baranger et al. [64] created a fast estimator for automated subspace thresholding. They then compared it to a complete list of thirteen potential estimators for the project. The performance of this group in vitro was assessed on a computer-simulated patient under a range of carefully controlled conditions, including varied tissue motions and flow rates. One based on the similarity of spatial singular vectors outperformed all other estimators. This estimator's denoising skills helped to improve the Contrast to Noise ratio and address the thresholding problem by lowering the noise floor by at least 5dB. This is because successful clutter filtering in ultrafast Doppler imaging requires both temporal and spatial information. Various organs (including the carotid artery, human brain, kidney, and thyroid) were tested in vivo to see whether it worked, and it did, with outstanding clutter filtering and noise reduction results, considerably increasing the dynamic range of Superfast Power Doppler pictures.

Diseased and non-cancerous tissues have different microvascular morphology, such as malignant lesions. Quantifying tumor-specific micro vascular morphological characteristics might enhance diagnostics. Microvascular morphology is difficult to assess due to the limits of ultrasonography Doppler. Ghavami and colleagues used ultrasonic Doppler imaging. [65] The authors analysed the microvasculature's morphological features and assigned numerical values to them. To assist achieve this aim, image enhancement methods and procedures for extracting morphological characteristics that allow quantitative study of microvasculature structures are available. Vessel segments created by the skeletonization of regularised microvasculature photos address other needs such as vessel segment diameter and length. To sustain morphological traits like tortuosity, large vessel trunks are essential. New filtering processes are also available to solve this problem. The procedures were evaluated using images of breast

tumours and thyroid nodules with tiny blood arteries, and the results were published. Based on vascular morphological characteristics, breast tumours and thyroid nodules can be classified as malignant (p-value 0.005) or benign (p-value = 0.01). Researchers were able to measure the microvasculature using non-contrast ultrasound images and utilise it as a biomarker for the identification of certain illnesses.

Blood samples are taken from people all around the world to examine their glucose levels. Blood glucose monitoring and diabetes treatment options that are non-invasive, accurate, and cost-effective are in great demand. In determining blood glucose levels, capillary glucose measurement is less reliable than serum glucose testing. Currently, serum glucose levels are measured in a lab via an intrusive method. Due to the inconvenience of the intrusive operation, continuous glucose monitoring is not an option. Joshi et al. present the iGLU 2.0, a wearable, non-invasive consumer device, for accurate blood glucose monitoring on a regular basis in this study [66]. The instrument's brief near-infrared spectroscopy was created by us. The Internet-of-Medical-Things is used in smart healthcare to make patient and caregiver data available in the cloud (IoMT). To choose the best regression model from among them, the system must be calibrated and validated on healthy, prediabetic, and diabetic people. In iGLU 2.0, a powerful regression model for serum glucose levels is implemented, taking advantage of the precise measurement technique. For capillary blood glucose prediction using iGLU 2.0, AvgE and mARD are projected to be 6.09 percent and 6.07 percent, respectively, whereas for serum glucose prediction using iGLU 2.0, AvgE and mARD are expected to be 4.88 percent and 4.86 percent, respectively.

The use of tissue perfusion monitoring with Power Doppler imaging to detect inflammatory hyperaemia, diagnose deep vein thrombosis, and other clinical applications is common. Due to the instrument's lower sensitivity caused by heat and debris, detecting sluggish flow becomes more difficult. Furthermore, in order to achieve acceptable sensitivity, large ensembles are required, which decreases the frame rate and causes flash artefacts when the tissue moves considerably in [67]. The spatial coherence of backscattered ultrasound echoes can be utilised to assess flow as an alternative to traditional approaches. With this approach, the signal quality of traditional power Doppler techniques is retained or even increased, which boosts frame rate and slow flow detection. The method is possible, according to flow phantom experiments and

an in vivo thyroid study. How will it be accomplished? When compared to typical power Doppler imaging, your Doppler pictures will have an SNR increase of 15-30 dB. As a consequence, it can detect flow at 50% lower velocities or double the frame rate while maintaining the same image quality as traditional power Doppler. The results appear to be promising for clinical use of the technology.

Patients who are receiving 1-131 therapy for benign or malignant thyroid problems might benefit from knowing where the tracer will be administered. Because of septal penetration inside the collimator, using the well-known Anger scintigraphy method is challenging. The goal of the study [68] is to develop a two-dimensional 1-131 patient detector with high spatial resolution and a low acquisition time. As a result, the thickness of the septa may be altered to dramatically reduce septa penetration. Photomultipliers with excessively thick septa show "dead patches." The holes on the projection plane are filled one by one with precise displacements. Scintigrams recorded at several locations are integrated to create a single picture in order to generate the final image. Using the specified detector geometry, the precise structures of the measured phantom may be visible in 16 minutes. The newly developed method of thyroid scintigraphy is really beneficial.

This work [69] focuses on using spectral-based QUS to assess normal human thyroid function in vivo. On twenty healthy patients, we employed two ultrasonic imaging equipment and an experienced radiologist to acquire radiofrequency data spanning from 3-16 MHz. The spectral logarithmic difference technique yielded average attenuation coefficient slope (ACS) estimates of 1.69 dB/ (cm. MHz), with a standard deviation of 0.28 dB/ (cm. MHz). Using a phantom, the reference phantom approach provided BSC estimates of 0.18 sr-1.cm-1 for the required frequency range. For the entire frequency range of the study, the inter-subject variability in estimating BSCs was less than 1.5 dB. The fit of the experimentally computed BSC scattering models was further investigated using three distinct scattering models (Gaussian, fluid sphere, and exponential form factors). The exponential form factor provided the best overall goodness of fit ( $=0.917$ ) across a distance of 44-56 metres. Gaussian ( $=0.807$ ) and fluid were the next two options. Across all applications, sophisticated models of the effective dispersed diameter ( $= 0.752$ ) showed exceptional consistency. For both scanners, the estimated attenuation and backscatter coefficients for each of the scattering models used in this experiment were shown.

The hypothalamus-pituitary-thyroid system is responsible for thyrotoxicosis. This system includes the thyroid gland. In people with autoimmune thyroiditis (Hashimoto's thyroiditis), negative feedback control mechanism modelling is used to characterise the clinical course of euthyroidism, subclinical hypothyroidism, and overt hypothyroidism. Thyroxine (T4) and triiodothyronine (T3) are adversely controlled by thyroxine (T4) and triiodothyronine (T3) (TSH). Free T4 (FT4), a hormone that can be bound or unbound, can be used to identify hypothyroidism. The presence of lymphocytes that attack and finally destroy follicular cell autoantigens characterises thyroid autoimmune disease. To further understand how feedback control works, we included TSH, FT4, anti-thyroid peroxidase antibodies, and the functional thyroid gland size in our mathematical model. The associations between the previous three variables are used to create this final variable. The three variables listed above are routinely assessed. The dilemma of circulating hormones and thyroid damage occurring on two different time scales is addressed using singular perturbation theory. The mathematical model's analysis provides an effective way and circumstances, allowing the sick state to continue on its path.

## **CHAPTER 3 METHODOLOGY**

In this chapter on methodology, we concentrated on and detailed our chosen methodologies and strategies for diagnosing thyroid ailment, which we had previously discussed. A full overview of the approaches that have been applied is provided. To begin, this chapter describes the original thyroid disease datasets, various feature selection techniques, and various machine learning-based ensemble algorithms, including bagging and boosting classifiers, before describing the ensemble of ensemble voting approaches (both soft and hard) on thyroid disease datasets at the conclusion. The following is a breakdown of the information included in this chapter in depth. The introduction of the original datasets, the pre-processing section, details of the machine learning ensemble classifiers used, implemented features selection techniques, proposed ensemble of ensemble approach with voting technique, and at the end, the performance evaluation measurements that will be used to check the efficiency of models and approaches will all be covered in the following section.

### **3.1. Dataset 1**

In this experimental study project, the data set relating to thyroid illness was the primary emphasis. The information in this dataset was taken from a well-known district headquarter hospital in the Pakistani city of Dera Ghazi Khan, which is located in the Punjab region. This information was obtained from the Dera Ghazi Khan headquarter hospital in the famous district of Dera Ghazi Khan in the province of Punjab, Pakistan. Endocrinologists in Karachi, Pakistan, extensively evaluated and verified the dataset in order to ensure its integrity and credibility [12]. 309 entities are included inside the dataset, each of which is directly related with the total number of subjects. Each individual is subjected to 10 distinct screening tests, which are further subdivided into characteristics, as well as one goal variable, which is denoted as 'Class.' This dependent variable is further subdivided into three separate groups marked by the letters 'Hypo' for Hypothyroidism, 'Normal' for Normal Thyroids, and 'Hyper' for Hyperthyroidism. There are a maximum of 13 missing data in a 'T3' feature, each of which is represented by the letter '?'. The details of this dataset are shown in Table 3.1. whereas the Figure 3.1 depicts a correlation of the output variable with other features.

### 3.2. Dataset 2

The dataset in [71] was taken from KEEL repository named as “Thyroid 0387” and also available at University of California Irvine (UCI) repository containing thousands of datasets related to multiple domains. This dataset is focused on the classification task based on the 21 attributes with 15 having integer value and 6 having real values. There is no missing value present in this dataset. There is total 7200 entries in the dataset and one target variable ‘Class’ that further categorized into three domains, ‘Normal’, ‘Hyperthyroidism’ and ‘Hypothyroidism’. The details about the dataset (D) are demonstrated in Table 3.2.

**Table 3.1 Information about the thyroid disease dataset from the certified institution**

<b>First Dataset of Thyroid</b>	
<b>Characteristics Labels</b>	<b>Variety of features</b>
Sr. No.	1 to 309
Institution IDs	Distinct Number
Pregnancy	No, Yes
Body Mass Index (BMI)	Under weight Balanced Overweight
Blood Pressure (BP)	High Normal Low
Pulse Rate (PR)	50 to 110
Thyroid Test (T3)	0.15 TO 3.7 (Missing values = 13 denoted by ‘?’)
Thyroid Test (TSH)	0.05 to 100
Thyroid Test (T4)	0.015 to 30
Gender	Male Female
Age	6 to 62
Class	‘0’ as Hypo ‘1’ as Hyper ‘2’ as Normal

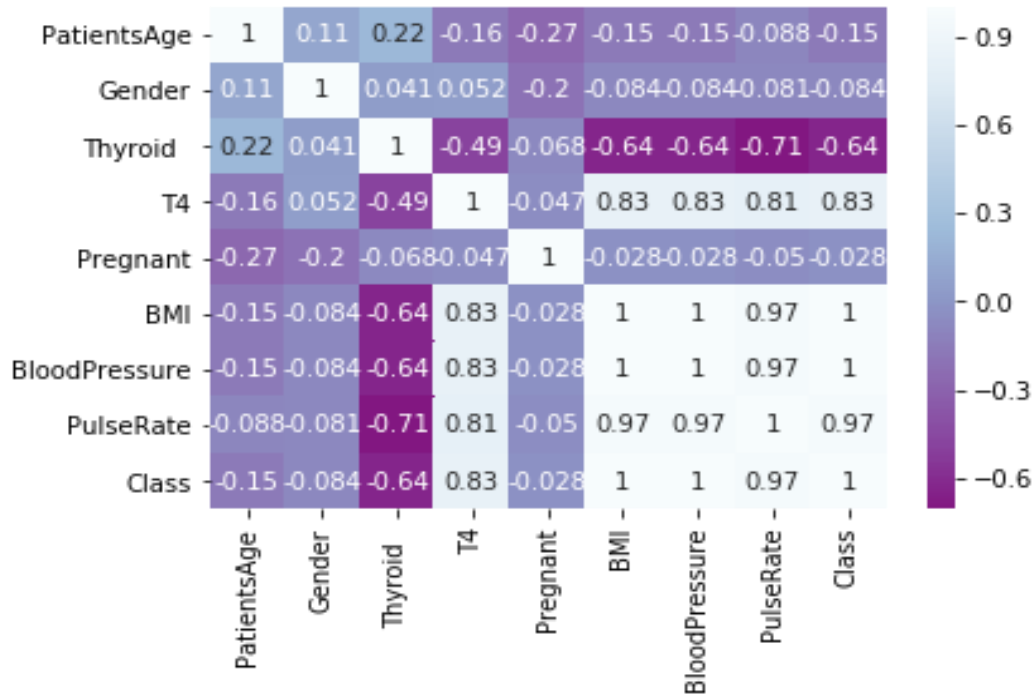


Figure 3.1 Correlation heatmap between features and target variables for Dataset 1.

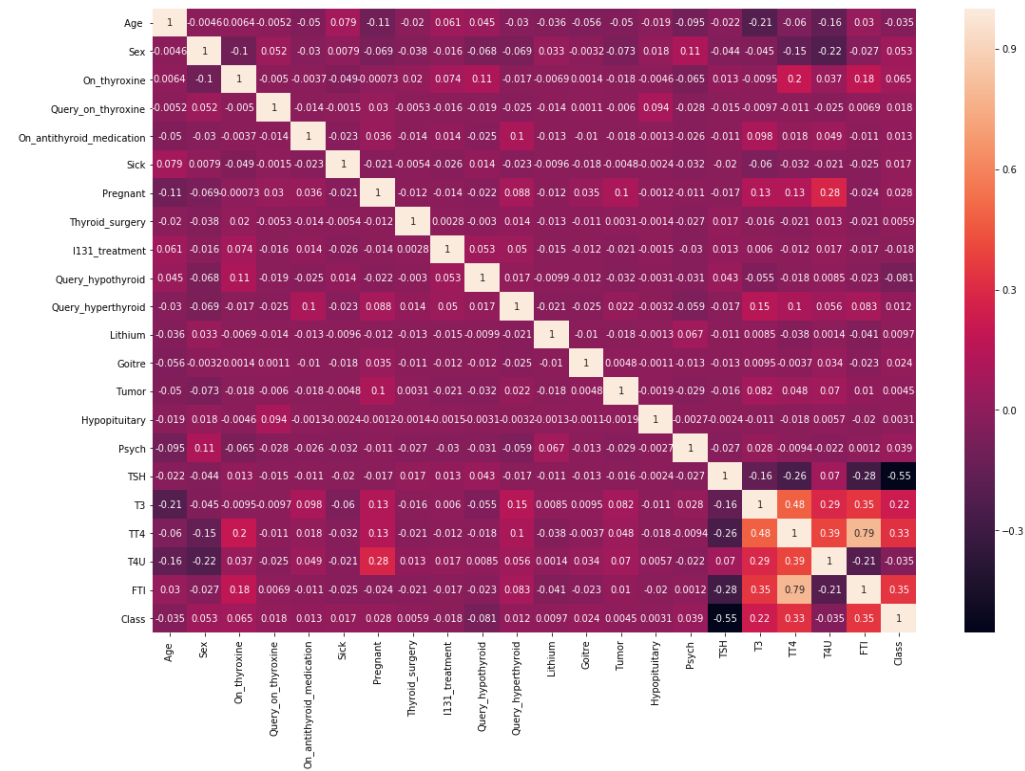


Figure 3.2 Correlation heatmap between features and target variables for Dataset 2.

**Table 3.2 Details about the opensource KEEL dataset related to thyroid disorder.**

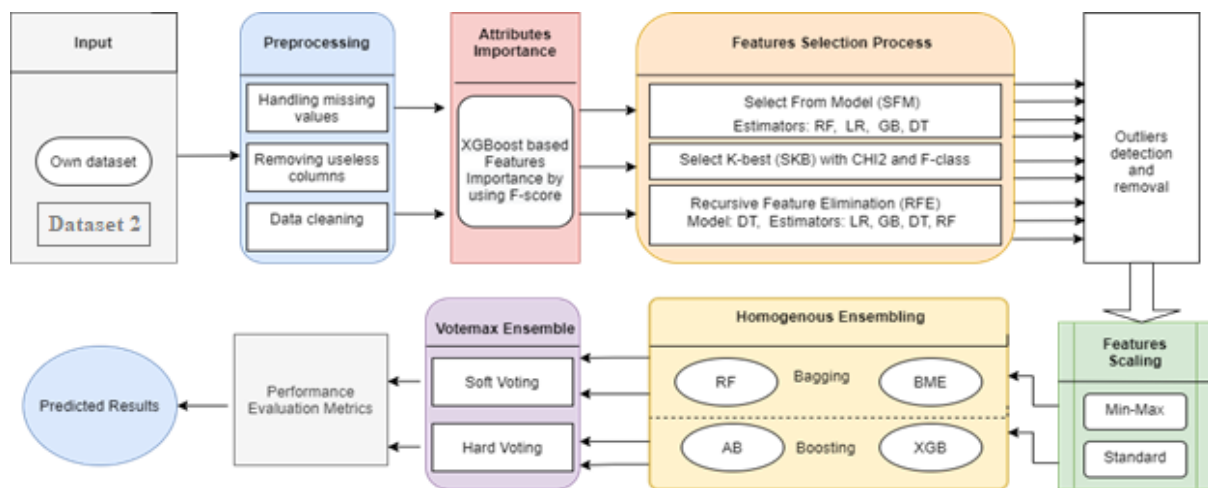
<b>KEEL Repository thyroid0387 dataset</b>		
<b>S. No</b>	<b>Features Names</b>	<b>Features Domain</b>
1	Age	0.01 to 0.97
2	Sex	0 to 1
3	On_thyroxine	0 to 1
4	Query_on_thyroxine	0 to 1
5	On_antithyroid_medication	0 to 1
6	Sick	0 to 1
7	Pregnant	0 to 1
8	Thyroid_surgery	0 to 1
9	I131_treatment	0 to 1
10	Query_hypothyroid	0 to 1
11	Query_hyperthyroid	0 to 1
12	Lithium	0 to 1
13	Goitre	0 to 1
14	Tumour	0 to 1
15	Hypopituitary	0 to 1
16	Psych	0 to 1
17	TSH	0.0 to 0.53
18	T3	0.0005 to 0.18
19	TT4	0.0020 to 0.6
20	T4U	0.017 to 0.233
21	FTI	0.0020 to 0.642
Target variable	Class	'1' as Normal '2' as hyperthyroidism '3' as hypothyroidism

### **3.3. Proposed Methodology**

Figure 3.3 depicts the research approach used in this study. The ability to present and visualize the data is critical prior to commencing the analytical process. But when comes to artificial intelligence, the possibilities are endless (AI), By using data pre-processing, it is possible to enhance both the description of the information and the reliability of the model by removing or cleaning out worthless data and missing values, and by reducing the number of missing values. The XGBoost classifier was then used to graphically evaluate the significance of characteristics based on the F-score [72] to identify their relative significance. Moreover, the three attribute selection methods employed, SFM, SKB, and RFE, are shown in Figure 1



together with their respective estimators. The next phase involved identifying and removing anomalies; it is crucial to identify and detect anomalies after attribute selection; as a consequence, the absence and presence of anomalies are depending on the total number of features chosen. The scaling technique is then used to normalize the data derived from the chosen characteristics, which is the following stage. This is accomplished through the use of both traditional and minimal-maximal scaling, as well as feature scaling, which makes the data more regular. Finally, homogeneous ensemble bagging and boosting methods such as RF and Base Meta Estimator (BME), AdaBoost (AB), and XGBoost (XGB) are used. Next, the predictions are sent to a second voting method that includes both hard and soft voting. This is another reason why the performance evaluation metrics are used. They can show how well procedures used in the final product are shown.



**Figure 3.3 Proposed Methodology Block Diagram.**

### 3.3.1 Data Pre-processing

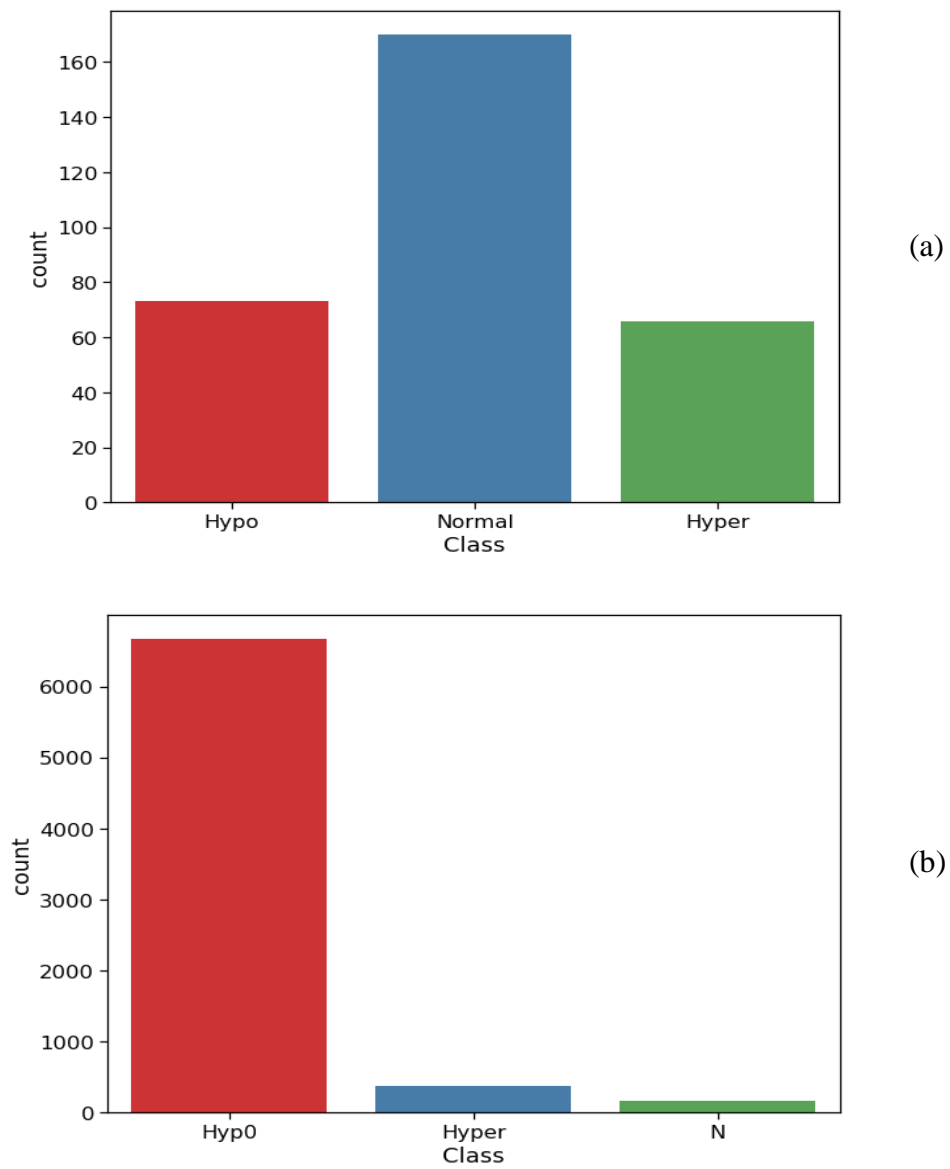
An open-source file format called CSV was used to hold all the data used in this study. These features should be excluded since they have no direct influence on the outcomes "Class" and have a detrimental effect on the effectiveness of the models, respectively, in order to minimise the risk of missing data. The vast bulk of the attribute data in this collection is represented as text or characters, not integers or reals. Libraries have a hard time implementing these values since they can't be used to perform operations on them directly. For example, the phrase

"Pregnancy" indicates "Yes" with a 1 and "No" with a 0 when text or strings are converted to real numbers or integer values. In a similar vein, all of the other characteristics have been modified. The mean values were used to impute the 13 missing values that were contained in the 'T3' and symbolized by the letter '?' in order to improve performance. The pre-processing stage has completed all of the data cleaning procedures required. For the dataset 2, it is targeted on the classification job based on the 21 characteristics, 15 of which have integer values and 6 of which have real values. The attributes in this dataset are as follows: There are no missing values in this dataset, which is a good thing. The Fig. 3.4. explains the ratio of the target variable classes distribution on both datasets.

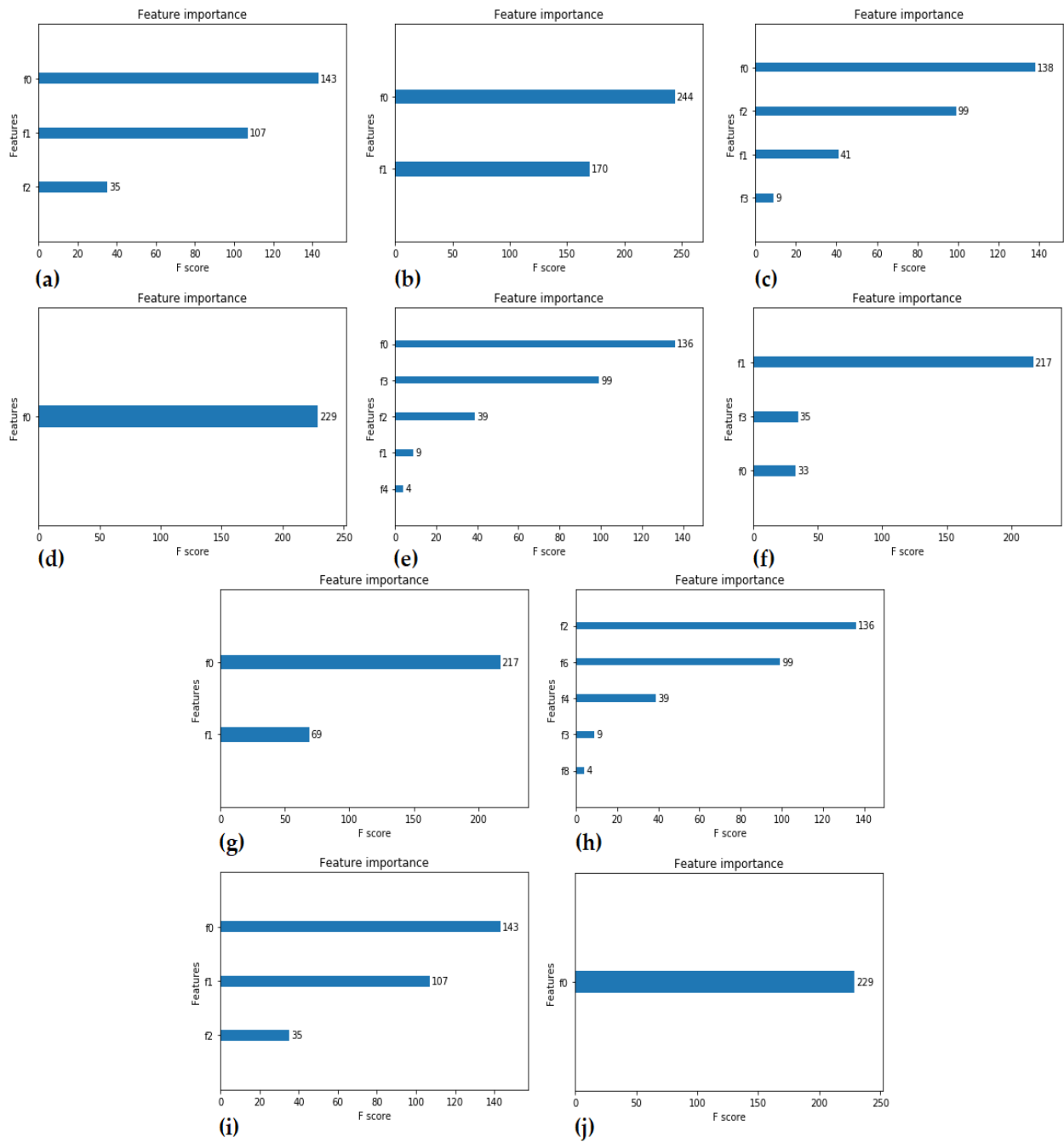
### ***3.3.2 XGBoost based Feature Significance with F-score***

An output variable's value is determined by how important an input feature is in predicting it. Scores for feature relevance play a critical role in the development of a predictive modelling project because they provide information about features that can be used to improve model efficiency and effectiveness, as well as insight into the model itself and the basis for reducing spatial dimensions for high-dimensional data and attribute selection. There are a variety of ways to determine feature significance scores. To name just a few, the F-score may be used to identify the relative relevance of several attributes, including RF and DT focused attributes [12]. Models may efficiently be transformed into groups by using the SFM class, which uses XGBoost-based features importance [73]. A classifier that is already being trained on the entire training sample may be used instead of starting from scratch. As soon as a specific threshold has been crossed, it may decide which traits to use. These attributes are selected for training and testing in SFM's convert () function based on this threshold. An XGBoost example demonstrates how to use the XGBoost method to develop and evaluate a model on an existing dataset. SFM instances are determined by the importance of the training data's features, which are then encased in this model. In order to choose features, the training dataset is employed, and the classifier is trained using the set of attributes that have been picked from the dataset. Last but not least, the model is assessed on the test dataset, which is selected using the same feature selection technique as before. This procedure is quite beneficial in obtaining a more accurate diagnosis of thyroid problem. For each attribute selection strategy, Figures 3.5 and 3.6

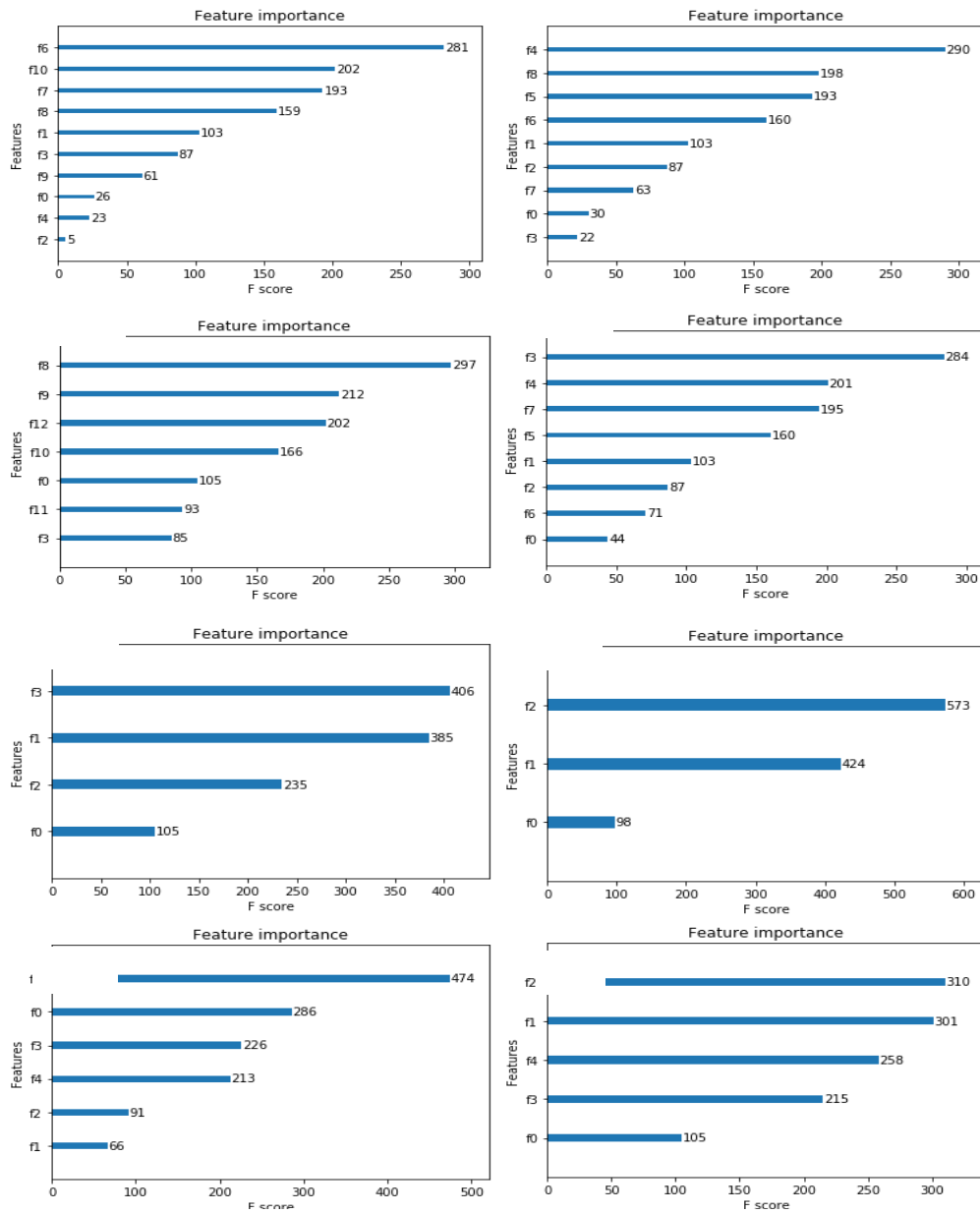
illustrate that the most relevant characteristics are those with the highest F-scores for the parameter.



**Figure 3.4** The 'Class' variable categories are described for (a) Dataset 1 (b) Dataset 2



**Figure 3.5 (a-d) SFM features selected using DT, GB, LR, RF, (e, f) SKB features selected using Chi2, FCI, and (g-j) RFE characteristics selected using DT, GB, LR, RF, respectively.(a-d) SFM features selected using DT, GB, LR, RF, (e, f) SKB features selected using Chi2, FCI, and (g-j) RFE characteristics selected using DT, GB, LR, RF, respectively.**



**Figure 3.6 XGBoost assigns relevance ratings to attributes based on F-scores for features extracted using characteristics selecting algorithms (k-n) The characteristics specified for RFE using DT, GB, LR, RF, (o, p) for SKB employing Chi2, FCI, (q-t) for SFM utilizing DT, GB, LR, RF, and (q-t) for SFM utilizing DT, GB, LR, RF, respectively.**

### **3.4. Attribute Selection Approaches**

Feature selection in model creation refers to the process of picking the most reliable, quasi, and determining factors for use in model production. A rigorous approach is needed to reduce the size of datasets as the quantity and variety of them grows. Selecting relevant features is a fundamental goal of feature selection since it improves predictive power while decreasing model processing costs. This information is provided in Table 2 for each feature-selection approach.

#### ***3.4.1 Selection From Model (SFM)***

It is possible to utilise the SFM in combination with any estimator that allocates priority to each attribute based on a specified characteristic (such as the coef function) or an important getter. It is considered irrelevant if the matching relevance of attribute values falls below a lower limit that has been specified. There are known methods for calculating a threshold using text input as well as numerical information. Thresholds may be calculated using text input and heuristics such as "mean," "median," and distinct pairs of these, such as "0.1\*mean."; for example, you may use the max features option in combination with them to restrict the number of characteristics that can be selected in connection with eligibility. The sklearn package [74] was used in the development of the SFM implementation. These are the estimators that were utilized in this approach: LR [75], RF [76], DT [77], and GB [78], among others.

#### ***3.4.2 Recursive Feature Elimination (RFE)***

To choose features, RFE employs wrapper structure to select attributes. Taking this into account, the method's core employs an RFE-wrapped classifier that aids in the feature set of the classifier. It differs from the filtration feature selection approach, which picks features based on the highest and lowest scores acquired for each feature. This method Wrapper-based RFE employs a filter-based selection of features internally to choose which characteristics to use. Beginning with all characteristics in a training instance, the RFE technique is used to identify a set of attributes. The entire attributes selection process may be completed by fitting the specified Learning algorithm into the model's base, assessing features based on their significance, deleting the least critical variables, and re - training the model. When just a few qualities are left, the procedure is iterated. Machine learning models (for example, decision

trees) or a statistical technique may be used to assess the significance of a given attribute. RFE is implemented as an algorithm in the scikit-learn machine learning package [74]. RFE transformations require that you construct your class first using the "estimator" argument and the "n features to select" function, which let you choose the number of attributes from which to choose. When analysing data for this research, we used the similar estimators from section 2.4.1 as in section 2.4.1 as the foundation for our analysis: DT, GB, LR, and RF.

### ***3.4.3 Select K-Best (SKB) based on Simple regression Extracted Features***

When using this method, statistical measurements may be utilised to identify the traits that have the greatest correlation with the output variable. To choose a certain number of features, the scikit - learn program's Select K-Best module is used in conjunction with various data sets. Table 3.3. provides a detailed breakdown of the most important components of our investigation. There are two functions in SKB: chi2 and FCI. chi2 is used for non-negative features, while FCI is used to calculate the ANOVA, F-value for a particular cohort.

### **3.5. Isolation Forest (ISO) for Automatic Identification and Elimination of Outliers**

Preparation is crucial to ensuring that data scientists' models correctly reflect reality. Outliers may be present in a dataset, which are data points that fall outside of the predicted range. The term "outlier" refers to these extraordinary independent values that are out of the ordinary. It is a unique observation that distinguishes itself from the others. Understanding and, in certain cases, reducing these outlier values may aid in the improvement of ML modelling and model capabilities more broadly. Because each dataset has its own set of properties, there is no one approach that can be used to define and identify outliers in general. The review of raw data and

**Table 3.3 Attribute details selected by the feature selection techniques at run time.**

<b>DATASET 1</b>				
<b>Techniques for Feature Selection</b>	<b>Functions/ Estimators</b>	<b>Gross Features used</b>	<b>Features Selected</b>	<b>Feature Selection Time (s)</b>
Select From Model (SFM)	GB	09	04	0.154
	DT	09	01	0.032
	RF	09	03	0.135
	LR	09	02	0.010
Select K Best (SKB)	FCI	09	03	0.006
	Chi2	09	05	0.014
Recursive Feature Elimination (RFE)	GB	09	03	1.172
	DT	09	01	0.010
	RF	09	02	0.235
	LR	09	05	0.092
<b>DATASET 2</b>				
<b>Techniques for Feature Selection</b>	<b>Functions/ Estimators</b>	<b>Gross Features used</b>	<b>Features Selected</b>	<b>Feature Selection Time (s)</b>
Select From Model (SFM)	GB	21	03	1.239
	DT	21	04	0.013
	RF	21	05	0.050
	LR	21	06	0.079
Select K Best (SKB)	FCI	21	08	0.009
	Chi2	21	11	0.012
Recursive Feature Elimination (RFE)	GB	21	09	14.930
	DT	21	11	0.082
	RF	21	08	1.294
	LR	21	13	7.570

the determination of whether or not a certain outcome is an abnormality are all part of the standard procedure. Based on the data accessible to us, statistical approaches may be used to discover events that seem to be odd or improbable basis of information. Based on the data from the training sample, the fit model will decide which samples are outliers and which are inliers in the training cases. Outliers from the training dataset will be deleted before the model is updated to the remaining cases and tested on the whole testing dataset. To find outliers in this research study, we employed the Isolation Forest, a tree-based outlier identification technique.



According to [79], anomalies which are both small and distinct may be found by modelling typical information in a way that allows for the discovery of anomalies. Table 3.4 demonstrates the mean absolute error (MAE) and the identification of outliers in the dataset using each feature selection strategy.

### **3.6. Homogenous Ensemble**

In machine learning and artificial intelligence, ensemble methods are used to combine many learning algorithms into a single predictive model that is more accurate than any one algorithm could be on its own. Homogenous ensemble, on the other hand, is a cluster of classification models of various types, each constructed on a distinct data sample [80]. In this research work, the two most important forms of homogeneous ensembles have been deployed as beginning ensembles, and the results have been evaluated. Bagging and boosting are two different things.

#### ***3.6.1 Bagging***

Bagging is a technique for improving accuracy since it reduces variance to a significant degree. The conclusion is that overfitting, which was a key issue with many estimation techniques in the past, is no longer a concern. During the learning process, a homogeneous weak classifier learns items in parallel, independently of one another, and then combines them by taking the average of the findings. It is more effective than utilizing single models since the weak basis classifiers are integrated to form a single, strong classification model rather than using several weak base classifiers. The most significant disadvantage of this strategy is that it is computationally costly. Even if we are dealing with regression or classification, we get a function that iterates over the training dataset, delivers an output, and is explained in terms of that dataset. Some variances may be expected due to the training dataset's theoretical variation.

At its core, bagging is an attempt to produce a model with less variation via "averaged" forecasts from many different models. A lack of independent models is caused by the enormous amount of data required. Due to the excellent "approximate properties" of boot-samples, we can fit almost independent models. In order to do so, it is necessary to create numerous bootstrap samples, each of which serves

**Table 3.4 Description of outliers observed with MAE scores in the appointed features**

<b>Dataset 1</b>							
<b>Techniques for Feature Selection</b>	<b>Functions/ Estimators</b>	<b>Gross Features used</b>	<b>Gross ingress available in the Dataset</b>	<b>Training data i.e., 75% of Total</b>	<b>Features Selected</b>	<b>Observed Outliers in Elected Features</b>	<b>MAE estimates</b>
Select From Model (SFM)	GB	09	309	231	04	23	0.0001
	DT	09	309	231	01	0	0.0001
	RF	09	309	231	03	23	0.0001
	LR	09	309	231	02	23	0.0001
Select K Best (SKB)	FCI	09	309	231	03	22	0.026
	Chi2	09	309	231	05	23	0.0001
Recursive Feature Elimination (RFE)	GB	09	309	231	03	23	0.0001
	DT	09	309	231	01	0	0.0001
	RF	09	309	231	02	19	0.295
	LR	09	309	231	05	23	0.0001
<b>DATASET 2</b>							
<b>Techniques for Feature Selection</b>	<b>Functions/ Estimators</b>	<b>Gross Features used</b>	<b>Gross ingress available in the Dataset</b>	<b>Training data i.e., 75% of Total</b>	<b>Features Selected</b>	<b>Observed Outliers in Elected Features</b>	<b>MAE estimates</b>
Select From Model (SFM)	GB	21	7200	5400	03	540	0.107
	DT	21	7200	5400	04	540	0.107
	RF	21	7200	5400	05	540	0.107
	LR	21	7200	5400	06	540	0.107
Select K Best (SKB)	FCI	21	7200	5400	08	540	0.107
	Chi2	21	7200	5400	11	540	0.107
Recursive Feature Elimination (RFE)	GB	21	7200	5400	09	540	0.107
	DT	21	7200	5400	11	540	0.107
	RF	21	7200	5400	08	540	0.107
	LR	21	7200	5400	13	540	0.107

as a distinct (almost) independent dataset drawn from the true distribution. If we develop a weak classifier for each of these data points, we can then merge them to produce ensemble approaches that have a reduced level of variance. Bootstrap samples are characterized by approximate independence and same distribution, which holds true for both trained base models and their samples. The following are the bagging classifiers that were employed in this

study. Random Forest (RF) [81] and Base Meta Estimator (BME) [82] are two popular *ensemble* learning algorithms.

### **3.6.2 Boosting**

This ensemble method is the most widely utilised and most potent of the group techniques. This was originally developed to deal with categorization concerns, but it was subsequently expanded to include comments relating to regression problems. Models are no longer fitted individually, but as a group, and must be fitted frequently in order to be trained in the same way. Training a model requires a series of iterative steps, each of which is reliant upon that models that have come before it. If these strategies are combined, the result is an ensemble of classification models less biased than the individual weak classifiers that make up the ensemble. Just a few of the newer, more generally accessible algorithms for speed optimization include Gradient Boosting Machine (GBM), AdaBoost, and Light GBM. Take a look at this scenario: Both the first and second models were wrong in their predictions. In the next step, we'll integrate the data and utilise them to build a more accurate forecast. Nothing more than a bid to enhance the long-term results. For example, boosting shows how to change a bad classifier into a better one utilising the core principles of transformation. In this investigation, we used the following boosting models: In addition to AdaBoost [83], there is XGBoost (XGB) [84].

### **3.7. Voting Ensemble of Homogenous Ensemble**

Often referred to as a "majority voting ensemble," a voting ensemble is a learning algorithm ensemble model that integrates predictions from a variety of different models. In theory, it is possible to employ this method to improve model performance, with the goal of outperforming any one model in the ensemble. In a voting ensemble, the forecasts from several models are merged to form a final prediction. This technique may be used to characterise and predict data, among other things. Within context of regression, this means calculating the mean of the models' predictions, as explained above. The amount of votes a label receives from the general audience determines its popularity.

In certain ways, a voting ensemble may be seen as a meta-model, or as a model of models. It can be used as a meta-model with any aggregate of machine learning models that have

previously been trained, and the existing models are completely unaware that they are being used as part of the ensemble. A voting ensemble is the ideal solution when you already have two or more models that perform well on a predictive modelling assignment. The ensemble models must be in wide agreement on their predictions [85]. Hard voting and soft voting are the two strategies for forecasting majority votes for categorization [86]. Soft voting and hard voting are the two methods of voting. Below are the details for both voting systems.

### 3.7.1 Soft Voting Ensemble

Soft voting in action is seen in Figure 3.6(a). Soft voting is the process of calculating the predicted probability or score for each target class in order to decide which class label has the best chance of being chosen. Based on the models, it also forecasts which class has the greatest total likelihood of being selected for the test. Assume the classifiers from  $C_1, C_2, \dots, C_n$  and the probability distributions for each classifier are represented by the variables  $Prob_{max}^n$  and  $Prob_{min}^n$ . Consider the following scenario: Class1 = 0 and Class2 = 1 are the class descriptions assuming there are two classes in total. The weights assigned to each classifier are denoted by the letters  $W_1, W_2, \dots, W_n$ . The following is the formula for calculating the probability for the target class:

$$Prob(Class_1) = W_1 * Prob_{min}^1 + W_2 * Prob_{min}^2 + \dots + W_n * Prob_{min}^n \quad (1)$$

$$Prob(Class_2) = W_1 * Prob_{max}^1 + W_2 * Prob_{max}^2 + \dots + W_n * Prob_{max}^n \quad (2)$$

The target variable classes' averages are determined as follows:

$$Avg(Class_1) = \frac{(Prob_{min}^1 + Prob_{min}^2 + \dots + Prob_{min}^n)}{n} \quad (3)$$

$$Avg(Class_2) = \frac{(Prob_{max}^1 + Prob_{max}^2 + \dots + Prob_{max}^n)}{n} \quad (4)$$

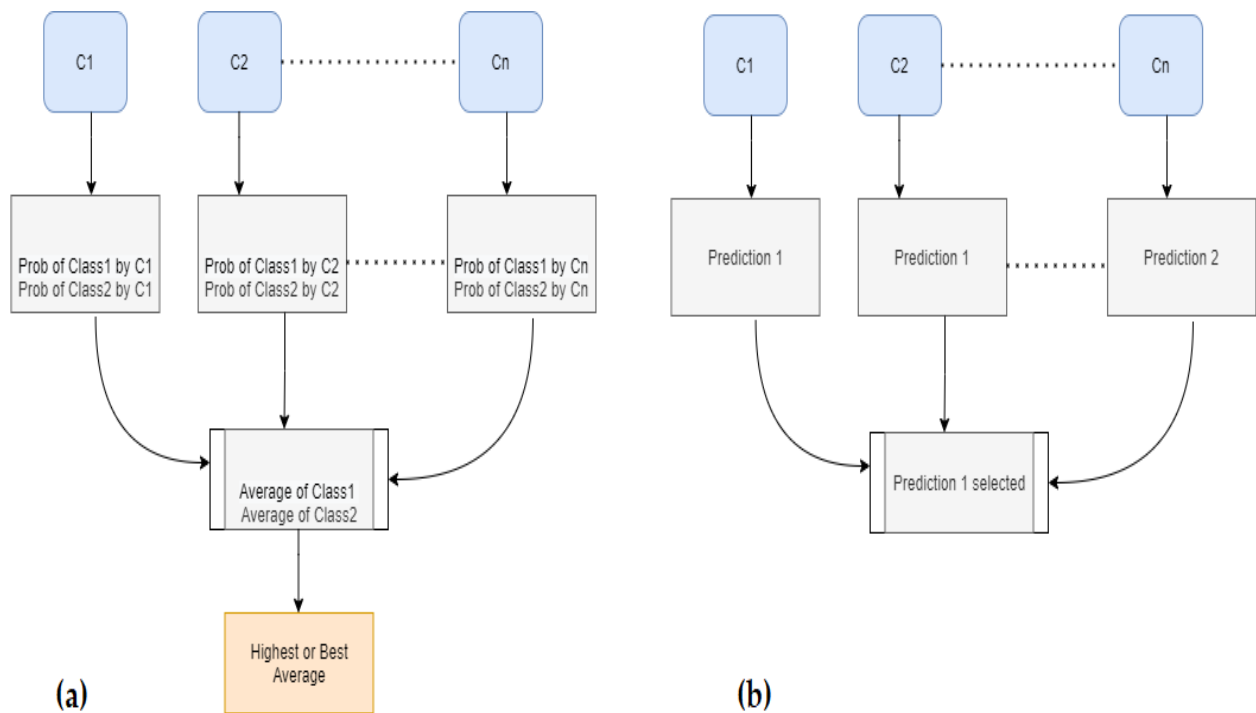
### 3.7.2 Hard Voting Ensemble

Votes for each class label must be added up before determining the target class that receives highest votes, which is known as hard voting. Using models, we can guess which class will get

the most votes. For input data, the hard voting classifier uses the mode of all predictions given by different forecasters. This mode is then utilised for classification by the hard voting model. However, even if each algorithm's weight is equal, it is regarded differently than when weights vary. Consider the following scenario: the total number of classifiers is  $n$ , and they are designated by the letters  $C_1, C_2, \dots, C_n$ , whilst their predictions are given by the letters  $P_0$  and  $P_1$ . The events of hard voting categorization were described mathematically in the equation below and in Figure 3.6 (b).

$$|C_1, C_2, \dots, C_n| \quad (5)$$

$$|P_0, P_0, \dots, P_1| \quad (6)$$



**Figure 3.7 (a) Soft and (b) hard voting are two instances of ensemble concepts.**

### 3.8. Performance Assessment Metrics

There are a variety of ways in which classification algorithms may be assessed. When comparing and contrasting different learning approaches, it is important to interpret metrics

analysis correctly. Diagnostic tests for breast cancer and human physiological conditions are evaluated using ML classifiers [87–88, 89–90]. " A = True positive (TP), B = True negative (TN), C = False positive (FP), and D = False negative are all included in the confusion matrix (FN). False positives are defined as systems that fail to accurately forecast an event's result. False positives occur when the given modal anticipates a true value however the actual answer was false. A wrong value is expected when the system's output is also incorrect, implying that algorithm is anticipating an incorrect value. If a false positive occurs, it means that the system predicted a genuine value, but the actual result was incorrect. An example of a false negative is when a system predicts that an outcome would be untrue, but the actual outcome is one with a genuine value.

The accuracy measure (AC) is the most commonly used to quantify a classifier's performance. It is determined by dividing the number of successful predictions by the total no of variables predicted.

$$Accuracy, ACU = \frac{A+B}{A+B+C+D} \times 100\% \quad (7)$$

True positive ratio (TPR) or recall is the ratio of real predicted positive samples to true positive samples in a given test.

$$Sensitivity, Recall, TPR = \frac{A}{A+D} \times 100\% \quad (8)$$

An F1 Score also sometimes called the F-measure calculates and recalculates the harmonized mean of accuracy and memory which is presented in this study. However, the worst-performing models are those that have an error rate of more than one and a score of zero. The F1-score equation is as follows:

$$F1\ Score = \frac{2A}{2A+C+D} \times 100\% \quad (9)$$

Brain W. Matthews invented the Matthews correlation coefficient (MCC) in 1975, and it is still in use today. The relationship here between observed and projected classes is represented by this correlation coefficient. This is obtained by calculating the MCC with the help of the confusion matrix. A positive number represents perfect prediction, while the negative number

shows a discrepancy between predicting and actual values. The term MCC is defined further down.

$$MCC = \frac{A \times B - C \times D}{\sqrt{(A+C)(A+D)(B+C)(B+D)}} \times 100\% \quad (10)$$

As PPV stands for Positive Predictive Value, it refers to the proportion of relevant occurrences that are found in a database while searching.

$$Precision, PPV = \frac{A}{A+C} \times 100\% \quad (11)$$

The Cohen kappa statistic is a statistical tool for estimating the degree of agreement between two assessors. This approach may also be used to assess the performance of a classification model in the real world.

$$Cohen\ kappa = \frac{p_0 - p_e}{1 - p_e} \times 100\% \quad (12)$$

where  $p_0$  denotes the model accuracy, and  $p_e$  reflects the degree to which the projected quantities of the classes agree with the actual values of the classes.

As a statistical metric, the Mean Absolute Error (MAE) is among the most frequently used statistics in statistical analysis.

$$MAE = \frac{1}{N} (\sum_{i=1}^N |Y_i - Y'|) \times 100\% \quad (13)$$

Squaring the total difference across the whole dataset yields the mean square error (MSE) statistic, which is used to compare actual and anticipated values.

$$MSE = \frac{1}{N} (\sum_{i=1}^N (Y_i - Y')^2) \times 100\% \quad (14)$$

Hamming loss (HL): It is the proportion of labels that were incorrectly anticipated.

$$Hamming\ Loss = \frac{1}{|N| \cdot |L|} (\sum_{j=1}^{|L|} \sum_{i=1}^{|N|} (Y_{i,j} \oplus Z_{i,j})) \times 100\% \quad (15)$$

where  $Y_{i,j}$  is the original target, and  $Z_{i,j}$  represents the predicted value?

## **CHAPTER 4 RESULTS AND DISCUSSIONS**

### **4.1. Introduction**

A homogeneous ensemble hybrid with several feature sets and a large number of estimators is described here. Machine learning-related tools and packages were used in conjunction with the Jupyter notebook with Python environment. Classifier hyper parameters have been tuned to a 75/25 split for training and testing, using a ratio of 75/25. It's here that the outcomes of the suggested strategy of an ensemble hybrid with three classification strategies and numerous estimators are given.

### **4.2. Dataset 1 results**

#### ***4.2.1 Homogenous Bagging***

Methodology for a homogeneous hybrid ensemble with three feature selection techniques and a large number of estimations is presented in this section. Programming on a Python platform with several machine learning modules and packages was used for the experiment. Classifiers have been tweaked using the splitting strategy, with a 75 percent training-to-testing ratio. For every attribute selection approach and estimator/function used, Table 1 shows the RF and BME's performance, learning and prediction durations. All of the classifiers, with the exception of the LR estimator from SFM attribute selection, achieved 100% accuracy on the applicable feature selection processes using the estimators. By selecting just one feature, the RFE feature selection using DT as an estimator, and utilising the forecast bagging model (BME), we were able to achieve the fastest training and prediction times with 100% accuracy.



**Table 4.1 The accuracy of the bagging classifiers as a function of training and prediction time**

<b>Homogenous ensemble (Bagging)</b>						
<b>Features selection techniques</b>	<b>Estimators or functions used</b>	<b>Selected features</b>	<b>Bagging Classifiers</b>	<b>Accuracy In %</b>	<b>Training Time (s)</b>	<b>Prediction Time (s)</b>
Select From Model (SFM)	LR	02	RF	98.71	0.2869	0.0029
			BME	100.0	0.0149	0.0020
	RF	03	RF	100.0	0.0861	0.0079
			BME	100.0	0.0139	0.0009
	DT	01	RF	100.0	0.3040	0.0029
			BME	100.0	0.0269	0.0019
	GB	04	RF	100.0	0.0873	0.0079
			BME	100.0	0.2844	0.0009
Select K Best (SKB)	Chi2	05	RF	100.0	0.0643	0.0039
			BME	100.0	0.0108	0.0019
	FCI	03	RF	100.0	0.0743	0.0049
			BME	100.0	0.0329	0.0050
Recursive Feature Elimination (RFE)	LR	05	RF	100.0	0.0289	0.0030
			BME	100.0	0.0089	0.0009
	RF	02	RF	100.0	0.0259	0.0029
			BME	100.0	0.0129	0.0019
	DT	01	RF	100.0	0.0320	0.0039
			BME	100.0	0.0129	0.0009
	GB	03	RF	100.0	0.0329	0.0049
			BME	100.0	0.0129	0.0020

#### 4.2.2 Homogenous Boosting

There are two boosting predictors that perform well in Table 2. Except for the FCI function with SKB methodology through XGB classifier implication, all estimators achieve 100 percent accuracy with their feature selection techniques.

**Table 4.2 The accuracy of the boosting classifiers as a function of training and prediction time.**

Homogenous ensemble (Boosting)						
Features selection techniques	Estimators or functions used	Selected features	Boosting Classifiers	Accuracy In %	Training Time (s)	Prediction Time (s)
Select From Model (SFM)	LR	02	AB	100.0	0.1037	0.0049
			XGB	100.0	0.9898	0.0009
	RF	03	AB	100.0	0.1047	0.0050
			XGB	100.0	1.3160	0.0019
	DT	01	AB	100.0	0.0490	0.0059
			XGB	100.0	1.3354	0.0009
	GB	04	AB	100.0	0.1187	0.0049
			XGB	100.0	0.9752	0.0008
Select K Best (SKB)	Chi2	05	AB	100.0	0.0757	0.0049
			XGB	100.0	1.0682	0.0216
	FCI	03	AB	100.0	0.0678	0.0069
			XGB	97.43	1.0034	0.0009
Recursive Feature Elimination (RFE)	LR	05	AB	100.0	0.0594	0.0059
			XGB	100.0	1.0484	0.0009
	RF	02	AB	100.0	0.0628	0.0059
			XGB	100.0	1.0614	0.0009
	DT	01	AB	100.0	0.0927	0.0049
			XGB	100.0	1.1596	0.0009
	GB	03	AB	100.0	0.0638	0.0059
			XGB	100.0	1.0472	0.0019

Figures following demonstrate the confusion matrixes for the RF and AB soft voting ensembling of the predictors. The soft voting method is characterized using equal weights for each classifier. The confusion matrices with SFM feature selection are represented below:

### 4.2.3 SKB feature selection on dataset 1

#### 1) Hard Voting

Figure 4.1 below shows the SKB with chi square feature selection technique performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism.

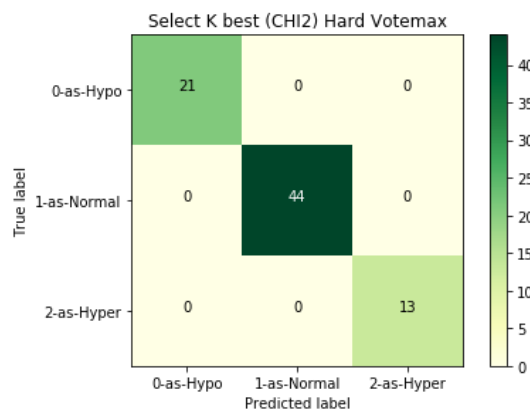


Figure 4.1 SKB with Chi (Hard Voting)

Figure 4.2 below shows the SKB with FCN technique performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism.

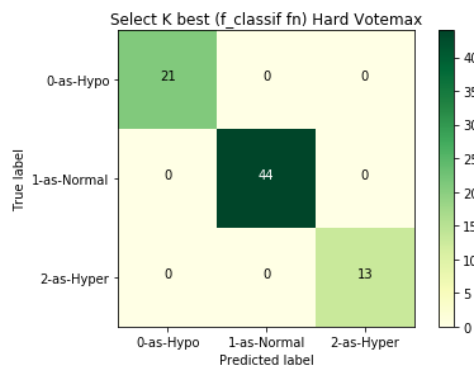
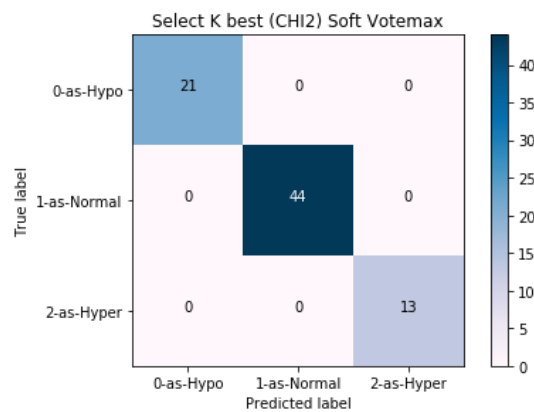


Figure 4.2 SKB with FCN (Hard Voting)

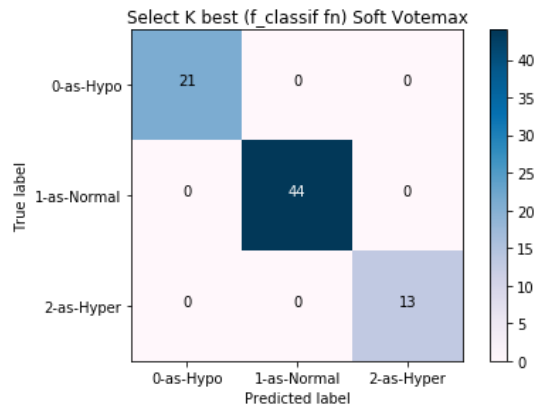
## 2) Soft Voting

Figure 4.3 below shows the SKB with Chi Square technique performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism.



**Figure 4.3 SKB with Chi (Soft Voting)**

Figure 4.4 below shows the SKB with Chi Square technique performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism.

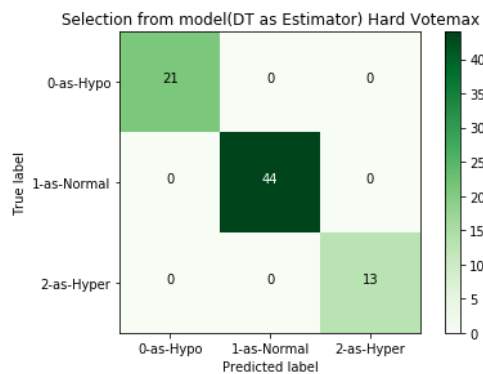


**Figure 4.4 SKB with FCN (Soft Voting)**

#### 4.2.4 SFM feature selection with Estimators on dataset 1.

##### 1) SFM with Decision Trees

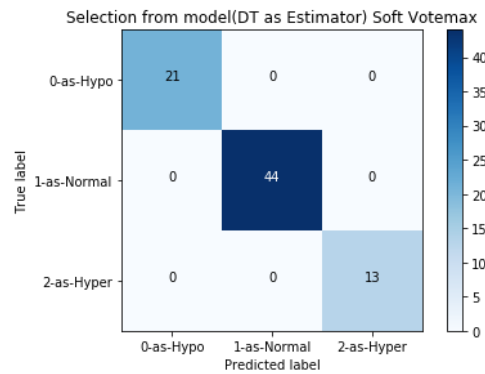
Figure 4.5 below shows the SFM with estimator of decision trees performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.5 SFM with Decision Trees (Soft Voting)**

Figure 4.6 below shows the SFM with estimator of decision trees performance in the form of confusion matrix, with soft voting 21 data points has been truly classified as Hypothyroidism,

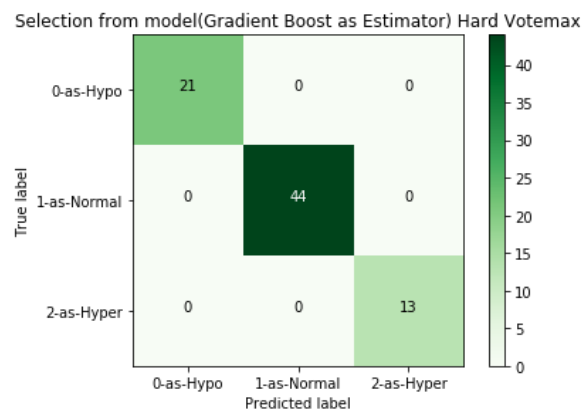
while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.6 SFM with Decision Trees (Soft Voting)**

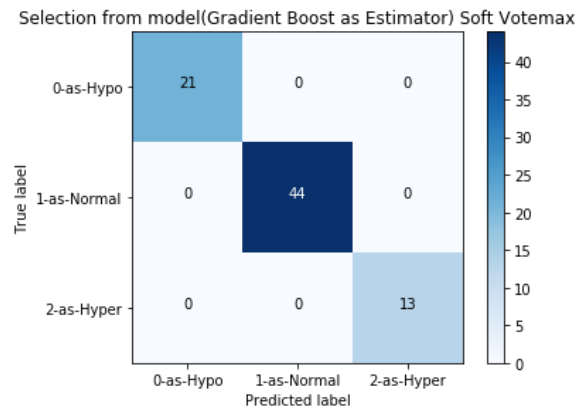
## 2) SFM with GBC

Figure 4.7 below shows the SFM with estimator of GBC performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.7 SFM with GBC (Hard Voting)**

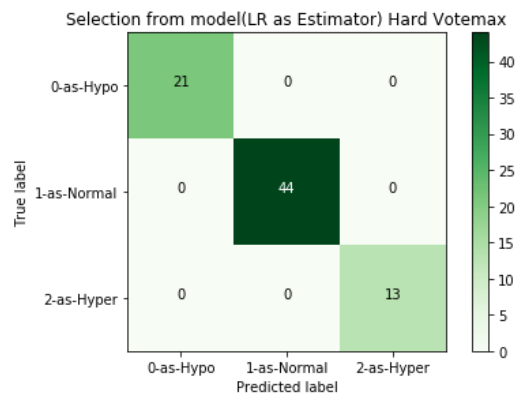
Figure 4.8 below shows the SFM with estimator of GBC performance in the form of confusion matrix, with soft voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.8 SFM with GBC (Soft Voting)**

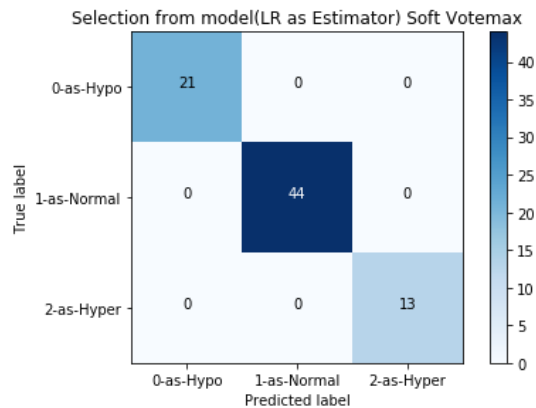
### 3) SFM with Logistic Regression

Figure 4.9 below shows the SFM with estimator of LR performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.9 SFM with LR (Hard Voting)**

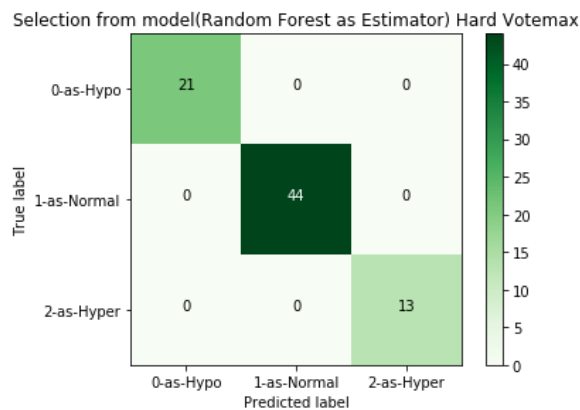
Figure 4.10 below shows the SFM with estimator of GBC performance in the form of confusion matrix, with soft voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



**Figure 4.10 SFM with LR (Soft Voting)**

4) *SFM with RF*

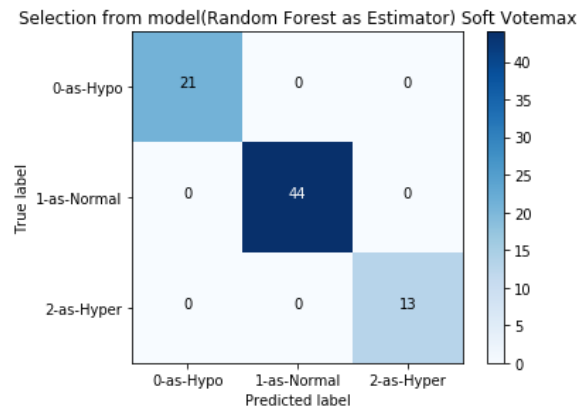
Figure 4.11 below shows the SFM with estimator of RF performance in the form of confusion matrix, with hard voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism.



**Figure 4.11 SFM with RF (Hard Voting)**

Figure 4.12 below shows the SFM with estimator of RF performance in the form of confusion matrix, with soft voting 21 data points has been truly classified as Hypothyroidism, while 44 data points truly classified as Normal, and 13 data points classified as Hyperthyroidism



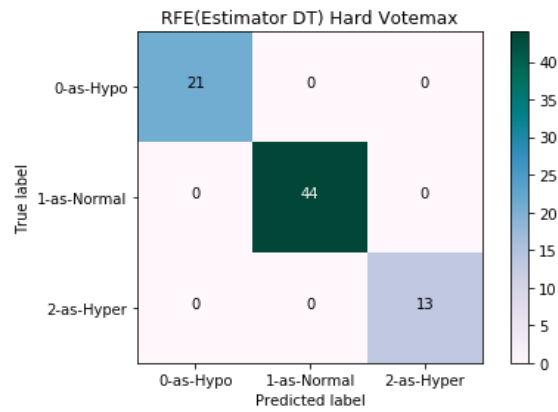


**Figure 4.12 SFM with RF (Soft Voting)**

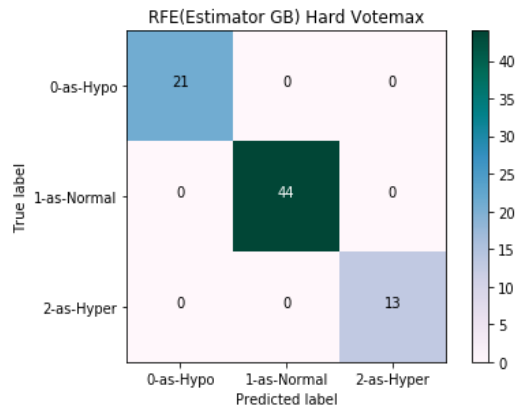
#### 4.2.5 RFE feature selection on dataset 1

##### 1) Hard Voting

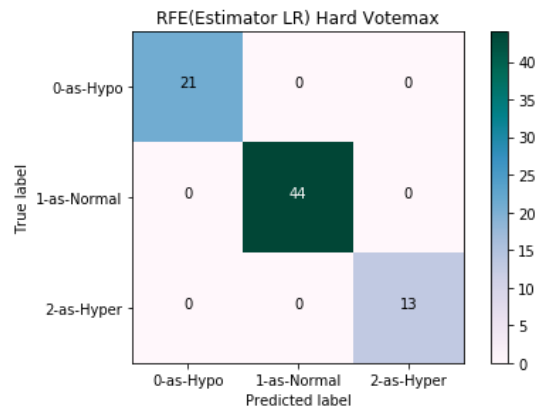
The RFE attribute selection based on the multiple estimators have been used. In Fig 4.13 the DT estimator utilized and selected only a single feature. Whereas for the GB estimator 03 features, for LR 05 and for RF 02 features have been selected for better performance. The Figures 4.14, 4.15 and 16 have been illustrated the confusion matrices results for hard voting ensemble technique with RFE attribute selection. while after the homogenous ensembling the voting technique has been used and attained the 100% prediction accuracy for all used estimators.



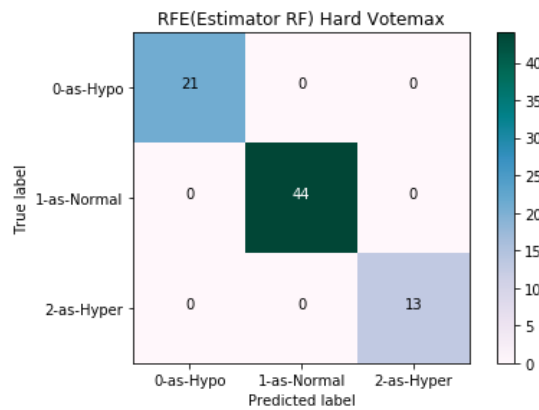
**Figure 4.13 RFE with DT (Hard Voting)**



**Figure 4.14 RFE with GB (Hard Voting)**



**Figure 4.15 RFE with LR (Hard Voting)**

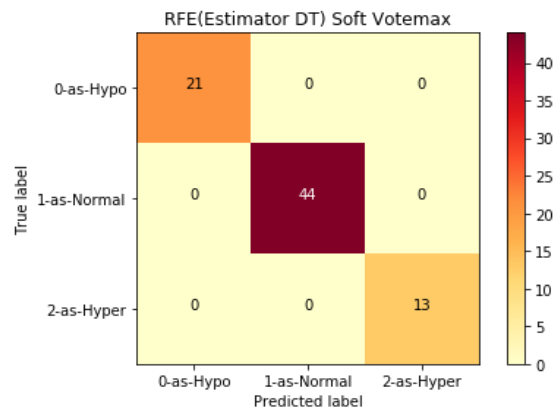


**Figure 4.16 RFE with RF (Hard Voting)**

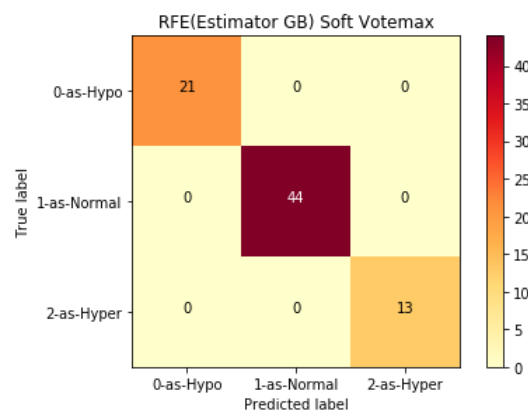
## 2) *Soft Voting*

We employed the RFE attribute selection based on several estimators. The DT estimator confusion matrix has shown in Figure 4.17 that only used and picked one feature. While the GB estimator 03 features were chosen for greater performance, the LR with 05 and RF with 02 features were chosen for better performance. The findings of the confusion matrices for the soft voting ensemble approach with RFE attribute selection are shown in Figures 4.18, 4.19, and 20. while following the homogeneous assembly, the soft voting procedure was employed to achieve the highest 100% prediction accuracy for all estimators by using RFE feature selection. All the figures shown below have 100 % true predictions and 0% false estimations by the proposed approaches. The Table 3. Illustrates the overall performance of all ensemble

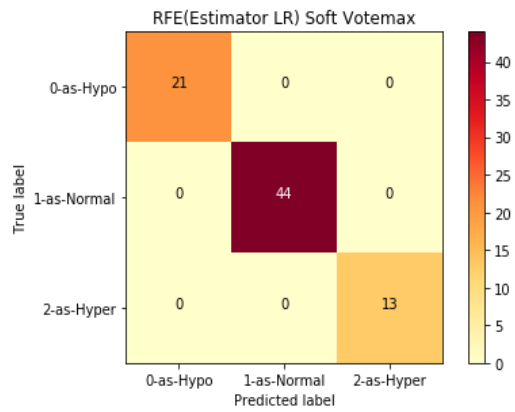
classifiers after the bagging and boosting models' predictions with a final voting of hard and soft voting ensemble techniques. This approach shows state of the art performance not only in terms of accuracy and other performance measurements but also have very less computational cost.



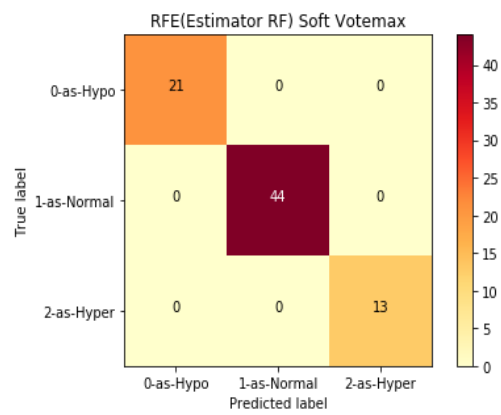
**Figure 4.17 RFE with DT (Soft Voting)**



**Figure 4.18 RFE with GB (Soft Voting)**



**Figure 4.19 RFE with LR (Soft Voting)**



**Figure 4.20 RFE with RF (Soft Voting)**

**Table 4.3 Soft and hard voting performance metrics for homogeneous ensemble classifiers on dataset 1.**

Ensemble of homogeneous groups (Voting classifier)															
Features selection techniques	Estimators or functions used	Selected features	Voting Classifiers	Accuracy In %	Training Time (s)	Prediction Time (s)	Recall %	Precision %	F1-score %	MCC %	Cohen Kappa %	MSE %	MAE %	Hamming Loss %	
Select From Model (SFM)	LR	02	Soft	100.0	1.062	0.007	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Hard	100.0	1.103	0.007	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	RF	03	Soft	100.0	0.176	0.010	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
			Hard	100.0	0.152	0.009	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	DT	01	Soft	100.0	1.415	0.010	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
			Hard	100.0	1.216	0.014	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	GB	04	Soft	100.0	0.190	0.009	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
			Hard	100.0	1.060	0.009	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
		Chi2	05	Soft	100.0	0.208	0.015	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0

Select K Best (SKB)	FCI	03	Hard	100.0	0.375	0.008	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Soft	100.0	0.159	0.013	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
			Hard	100.0	0.175	0.022	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
Recursive Feature Elimination (RFE)	LR	05	Soft	100.0	1.095	0.031	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Hard	100.0	1.138	0.009	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	RF	02	Soft	100.0	1.125	0.009	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Hard	100.0	1.363	0.011	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	DT	01	Soft	100.0	1.102	0.011	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Hard	100.0	1.230	0.010	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0
	GB	03	Soft	100.0	1.087	0.016	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0	
			Hard	100.0	1.073	0.011	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0	0.0

Cross-examination of the approach used in this study with other investigations on the same data is shown in Table 3. Findings were achieved by combining multiple homogeneous ensembles with attribute selecting strategies. Researchers set out to improve performance while minimizing the amount of time required for learning and forecasting in this analysis. A second step of the ensemble process, Voting (soft and hard) voting, was used to build the hybrid implementation of multiple feature selection, identify outliers and anomalies, and make final predictions using first ensemble classifiers from bagging and boosting approaches. Not only did this approach provide better outcomes, but it also took less time to train, and forecast compared to earlier hybrid models that used advanced algorithms with a wide range of techniques, such as neural networks. Existing procedures are more costly to implement and take longer to train and verify results, as shown by this comparison.

### **4.3. Dataset 2 results**

Similarly, to dataset 1 the second dataset that is open source has been handled identically. The first step of the methodology is the implementation of the homogenous ensemble techniques both bagging and boosting in a hybrid combination with estimators-based features selection approaches.

The Table 4.4 and 4.5 illustrates the results of both methodologies with their training and testing time.



**Table 4.4 . The accuracy of the bagging classifiers is described in terms of training and prediction time.**

Homogenous ensemble (Boosting)						
Features selection techniques	Estimators or functions used	Selected features	Boosting Classifiers	Accuracy In %	Training Time (s)	Prediction Time (s)
Select From Model (SFM)	LR	06	AB	92.66	0.1431	0.0119
			XGB	92.50	1.557	0.0119
	RF	05	AB	98.55	0.1969	0.0139
			XGB	99.00	0.4819	0.0080
	DT	04	AB	97.77	0.1779	0.0169
			XGB	98.88	0.5832	0.0085
	GB	04	AB	97.77	0.0538	0.0069
			XGB	97.22	0.5529	0.0109
Select K Best (SKB)	Chi2	11	AB	98.72	0.1464	0.0160
			XGB	99.00	0.8659	0.0087
	FCI	08	AB	98.55	0.1870	0.0124
			XGB	99.11	0.7289	0.0039
Recursive Feature Elimination (RFE)	LR	13	AB	98.72	0.0887	0.0069
			XGB	99.00	0.8776	0.0069
	RF	08	AB	98.55	0.1236	0.0109
			XGB	99.05	0.6382	0.0059
	DT	11	AB	98.55	0.1186	0.0119
			XGB	99.00	0.7789	0.0099
GB	09	AB	98.61	0.0568	0.0049	
		XGB	99.16	0.7373	0.0059	

**Table 4.5 The accuracy of boosting classifiers is described in terms of training and prediction time.**

Homogenous ensemble (Bagging)						
Features selection techniques	Estimators or functions used	Selected features	Bagging Classifiers	Accuracy In %	Training Time (s)	Prediction Time (s)
Select From Model (SFM)	LR	06	RF	92.11	0.3667	0.0293
			BME	92.00	0.1414	0.0039
	RF	05	RF	99.11	0.0488	0.0049
			BME	99.16	0.0498	0.0039
	DT	04	RF	98.94	0.0997	0.0080
			BME	99.00	0.0449	0.0047
	GB	03	RF	98.00	0.0962	0.0089
			BME	97.55	0.0201	0.0030
Select K Best (SKB)	Chi2	11	RF	99.05	0.0590	0.0079
			BME	99.00	0.0578	0.0086
	FCI	08	RF	99.00	0.1328	0.0087
			BME	98.94	0.0574	0.0048
Recursive Feature Elimination (RFE)	LR	13	RF	99.00	0.0698	0.0059
			BME	99.16	0.0408	0.0030
	RF	08	RF	98.77	0.1406	0.0089
			BME	98.72	0.0279	0.0030
	DT	11	RF	99.05	0.1246	0.0059
			BME	99.00	0.0304	0.0029
	GB	09	RF	99.05	0.0548	0.0049
			BME	99.05	0.0453	0.0030

### 4.3.1 Recursive Feature Elimination (Hard)

RFE is a feature selection algorithm that is a wrapper type. Fitting the specified machine learning algorithm used in the model's core, sorting features based on relevance, eliminating the least important features, and re-fitting the model after each step of the process is how it is performed. Each of these steps is continued until there are no longer any distinguishing characteristics. Figure 4.21 below shows the Recursive Feature Elimination based on decision trees with hard voting 41 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1642 data points classified as Hypothyroidism.

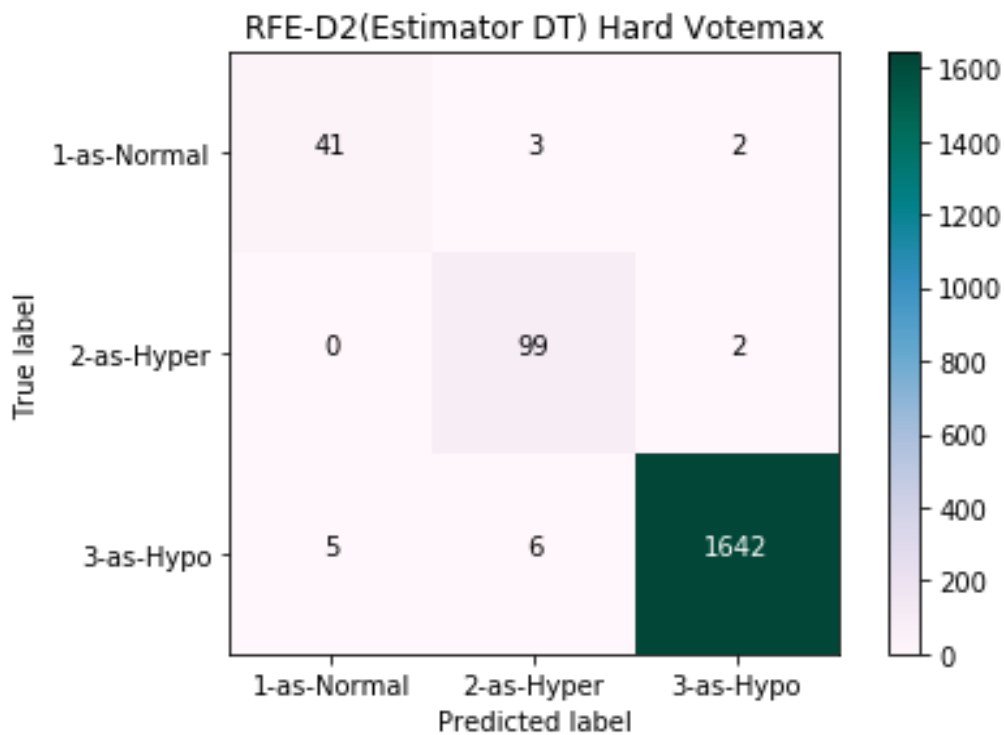
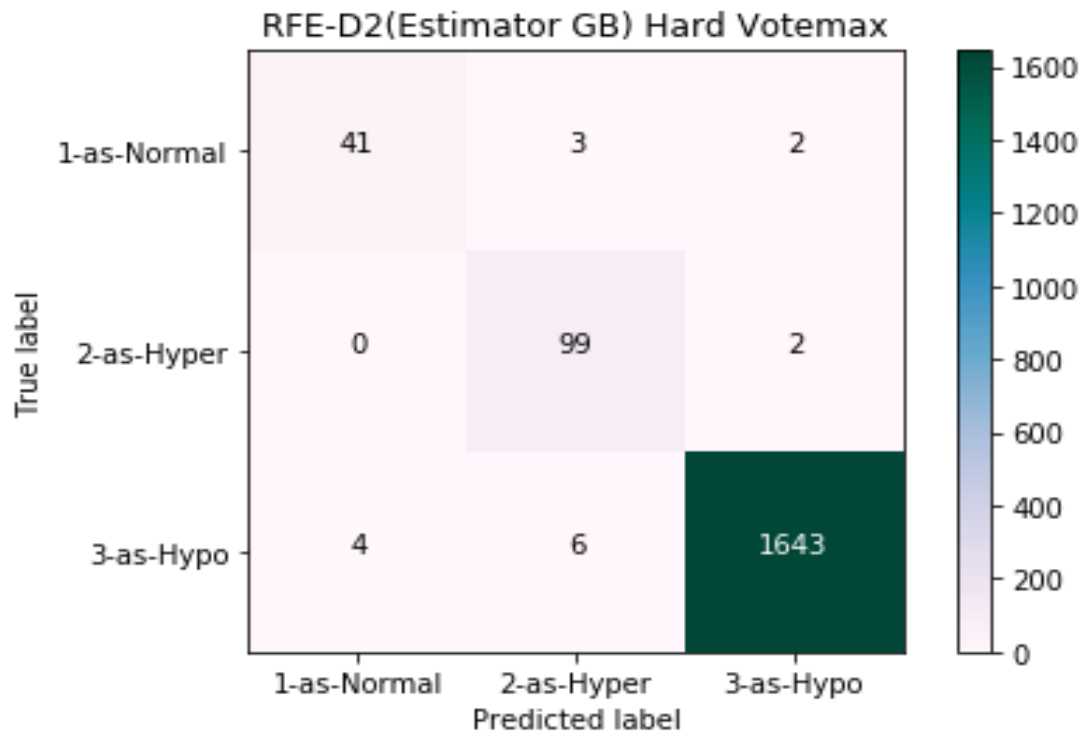


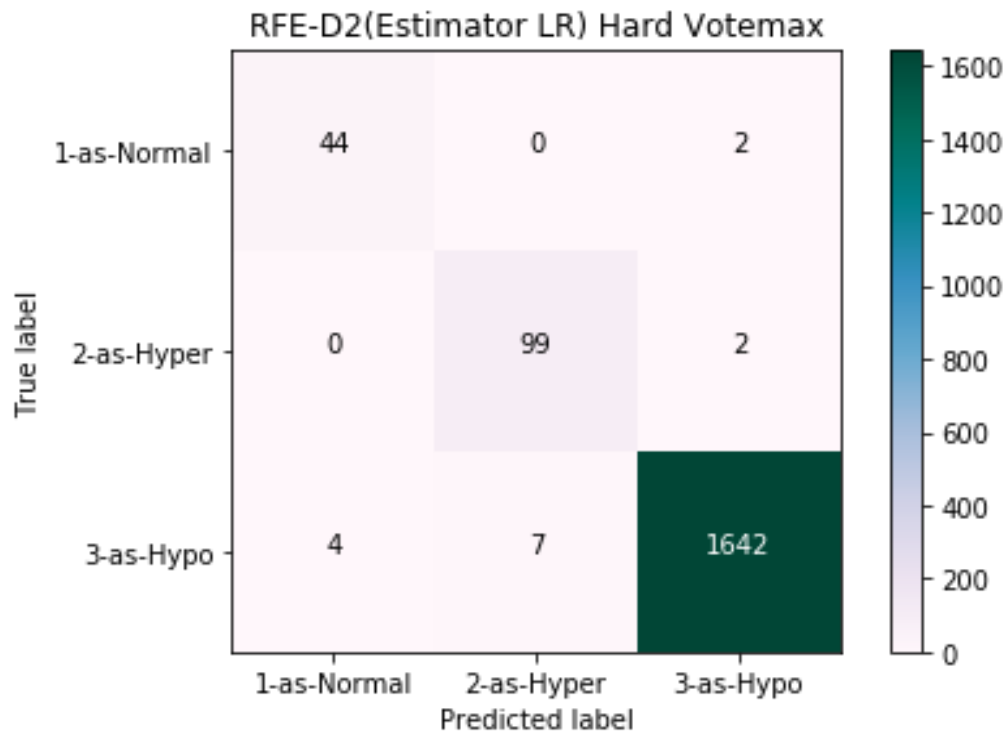
Figure 4.21 Recursive Feature Elimination using Decision Trees (Hard Voting)

Figure 4.22 below shows the Recursive Feature Elimination based on gradient boosting with hard voting 41 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1643 data points classified as Hypothyroidism.



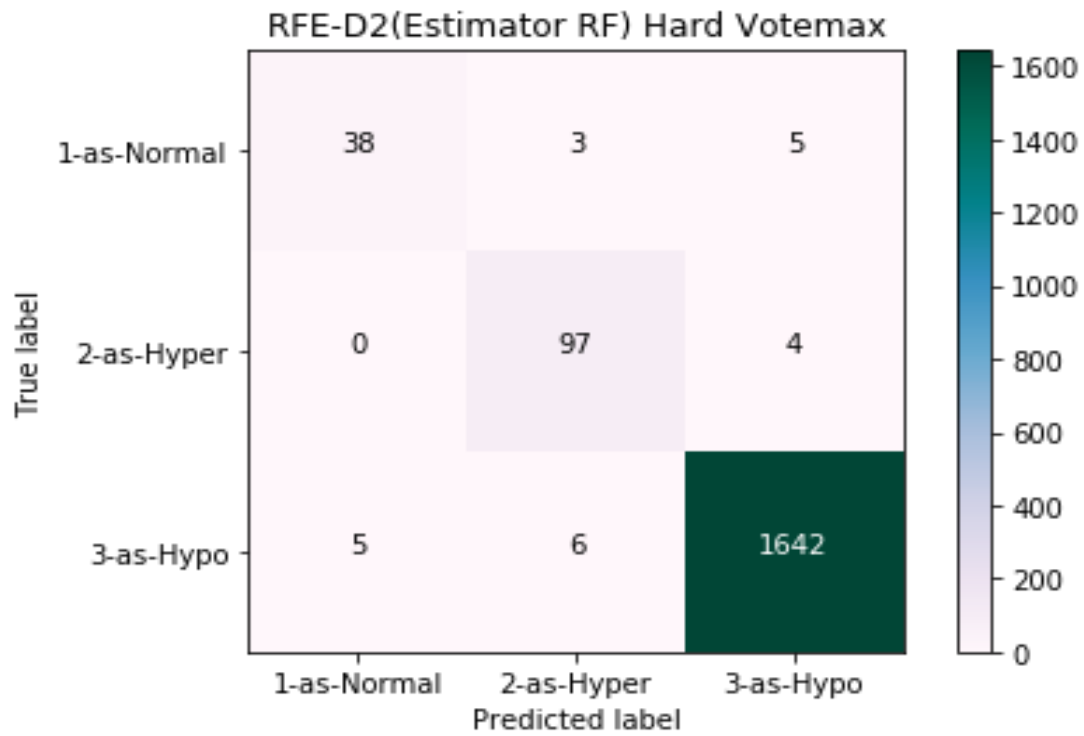
**Figure 4.22 Recursive Feature Elimination Based on Gradient Boosting (Hard Voting)**

Figure 4.23 below shows the Recursive Feature Elimination based on logistic regression with hard voting 44 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1642 data points classified as Hypothyroidism.



**Figure 4.23 Recursive Feature Elimination Based on Logistic Regression (Hard Voting)**

The below figure 4.24 shows the Recursive Feature Elimination based on random forests with hard voting 38 data points has been truly classified as Normal Patients, while 97 data points truly classified as Hyperthyroidism and 1642 data points classified as Hypothyroidism.



**Figure 4.24 Recursive Feature Elimination Based on Random Forests (Hard Voting)**

### 4.3.2 Select K-Best Features (Hard)

We use the feature selection approach to pick out the characteristics of our data that have the greatest impact on the target variable. In other words, we select the predictors that are most likely to be associated with the target variable. The Select Best technique selects the features based on which feature has received the highest score out of k. By altering the value of the 'score func' argument, we may use the classification and regression data analysis procedure. When preparing a large dataset for training, selecting the optimal features is a critical step in the process. Figure 4.25 below shows the chi square feature selection technique performance in the form of confusion matrix, with hard voting 42 data points has been truly classified as Normal Patients, while 100 data points truly classified as Hyperthyroidism and 1640 data points classified as Hypothyroidism.

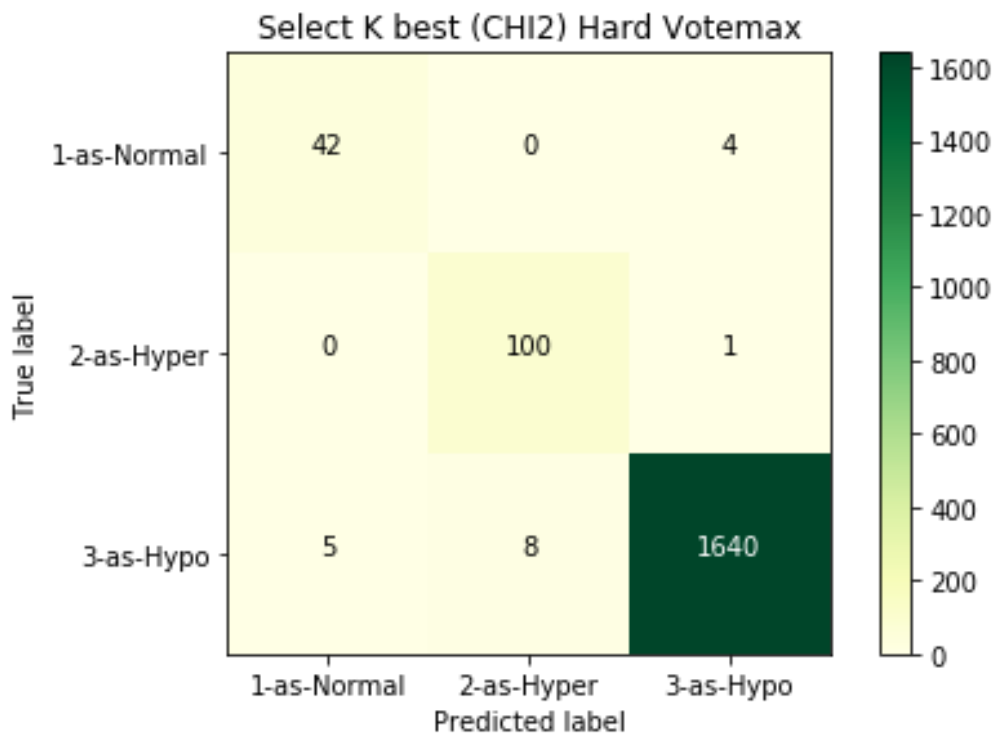
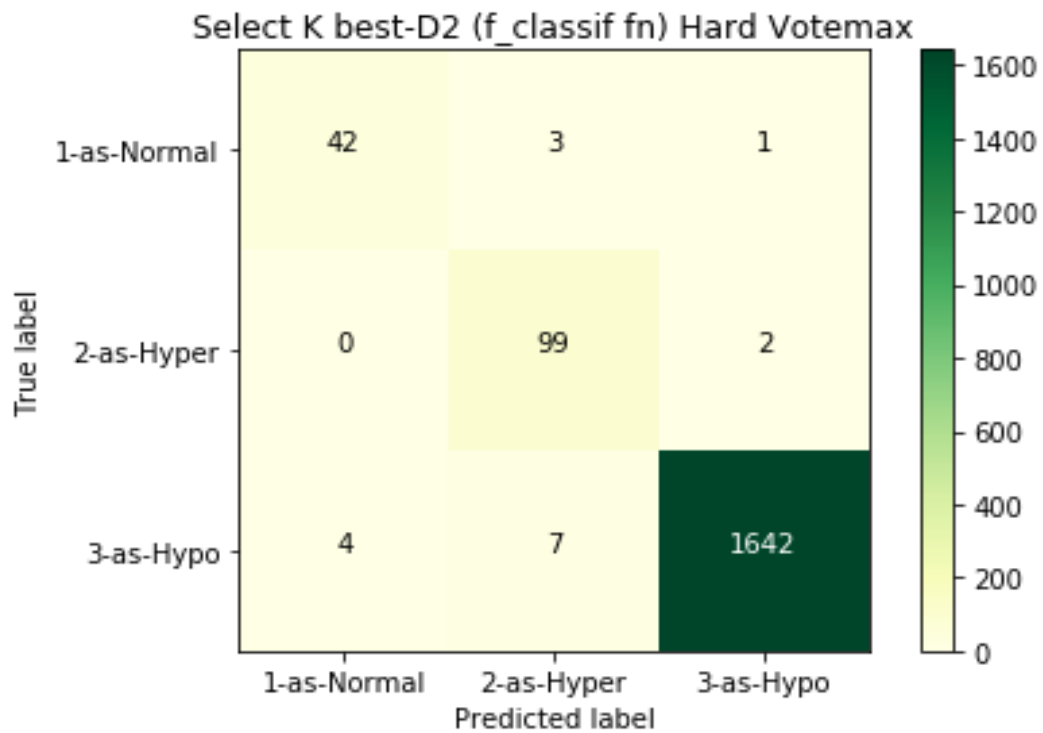


Figure 4.25 Chi Square Feature Selection Technique Performance

Figure 4.26 below shows the FCI feature selection technique performance in the form of confusion matrix, with hard voting 42 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1640 data points classified as Hypothyroidism.



**Figure 4.26 FCI Based Feature Selection Technique Performance**



### 4.3.3 Selection of Features from Model (Hard)

By using machine learning Estimators, k-best features are selected. Figure 4.27 below shows the selection of features from decision trees in the form of correctly classified data with hard voting 46 data points has been truly classified as Normal Patients, while 101 data points truly classified as Hyperthyroidism and 1639 data points classified as Hypothyroidism.

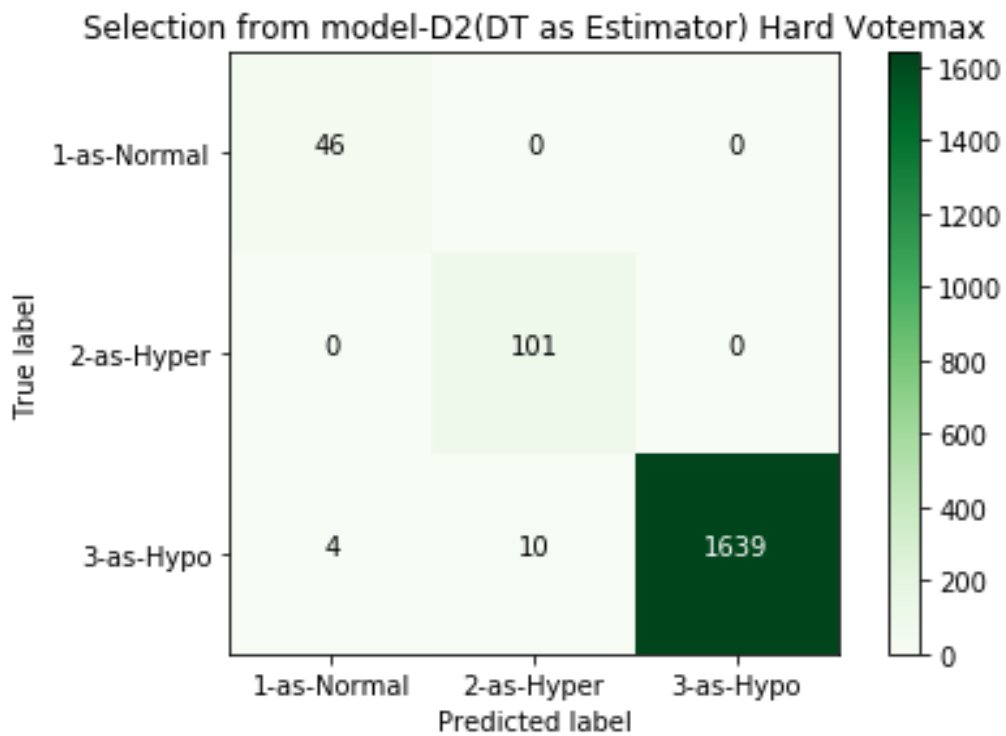
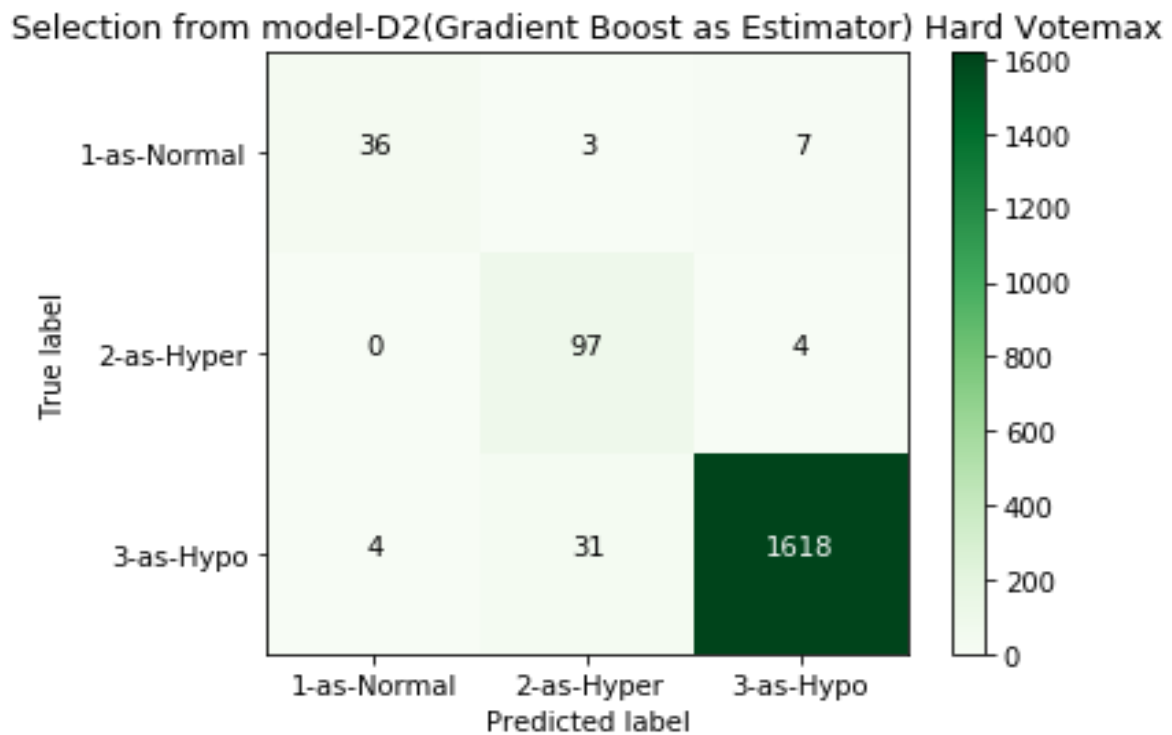


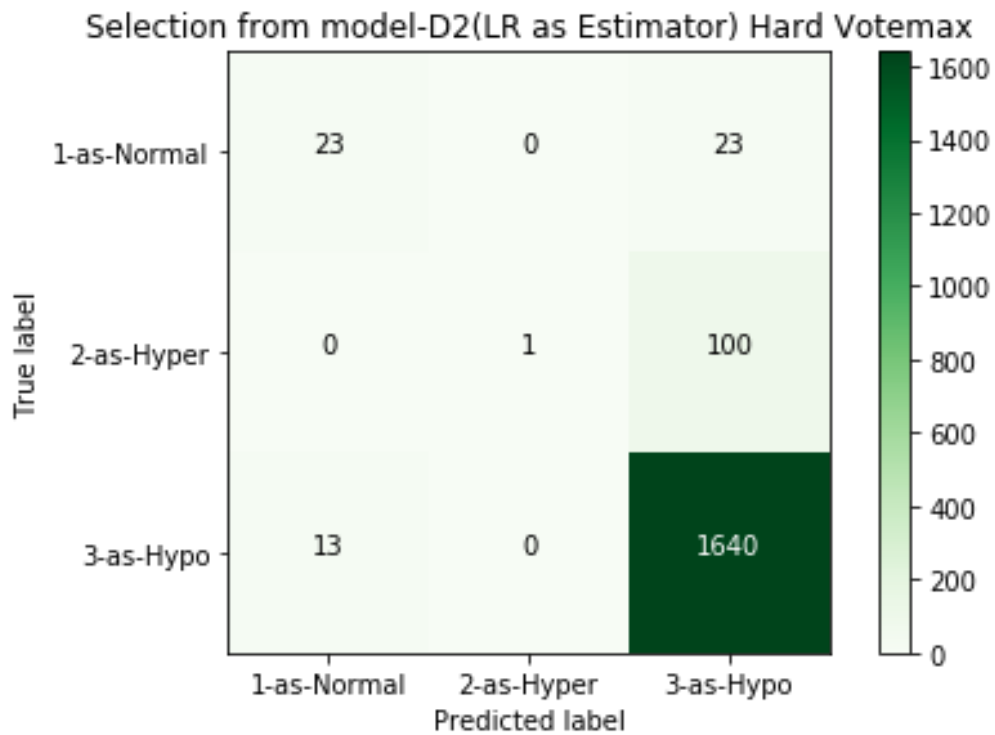
Figure 4.27 Selection of Features from Decision Trees

Figure 4.28 below shows the selection of features from GBC in the form of correctly classified data with hard voting 36 data points has been truly classified as Normal Patients, while 97 data points truly classified as Hyperthyroidism and 1618 data points classified as Hypothyroidism.



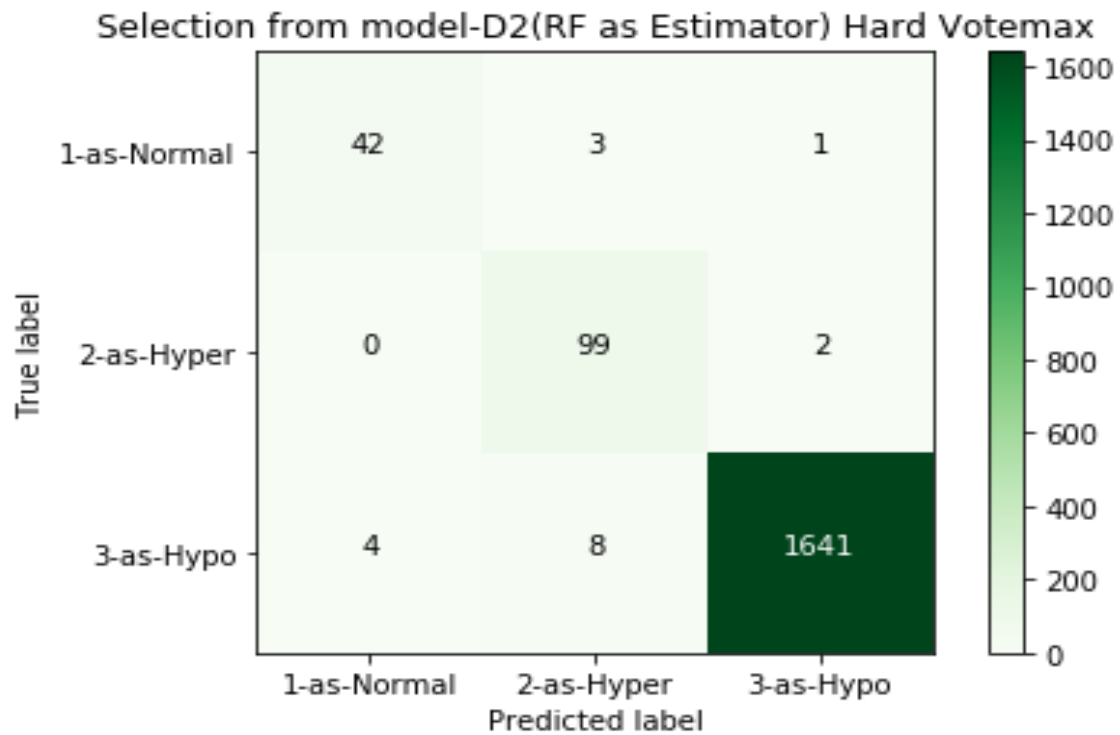
**Figure 4.28 Selection of Features from Gradient Boosting**

Figure 4.29 below shows the selection of features from LR in the form of correctly classified data with hard voting 23 data points has been truly classified as Normal Patients, while 1 data points truly classified as Hyperthyroidism and 1640 data points truly classified as Hypothyroidism.



**Figure 4.29 Selection of Features from Logistic Regression**

Figure 4.30 below shows the selection of features from RF in the form of correctly classified data with hard voting 42 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1641 data points truly classified as Hypothyroidism.



**Figure 4.30 Selection of Features from Random Forests**

#### 4.3.4 Recursive Feature Elimination (soft)

RFE is a feature selection algorithm that is a wrapper type. This is accomplished by fitting the specified machine learning algorithm that was used in the core of the model, ranking features according to relevance, deleting the least important features, and re-fitting the model after each step of the process. Each of these steps is continued until there are no longer any distinguishing characteristics. Figure 4.31 below shows the Recursive Feature Elimination based on decision trees with soft voting 43 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1643 data points truly classified as Hypothyroidism.

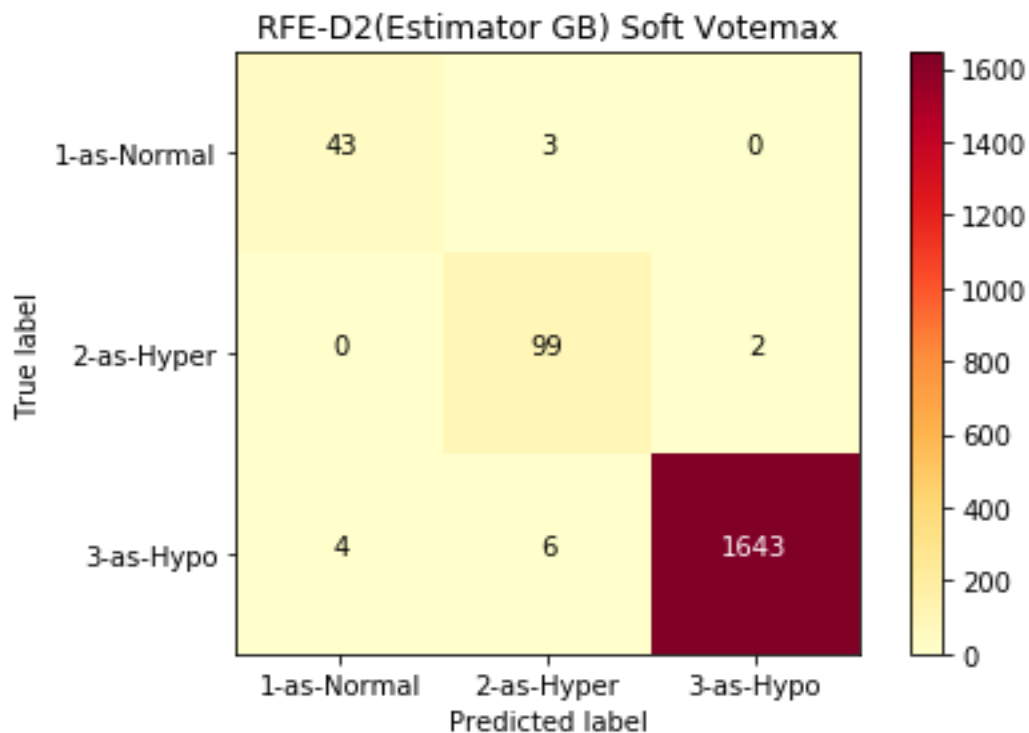


Figure 4.31 Recursive Feature Elimination using Decision Trees (soft Voting)

Figure 4.32 below shows the Recursive Feature Elimination based on gradient boosting with soft voting 45 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1643 data points truly classified as Hypothyroidism.

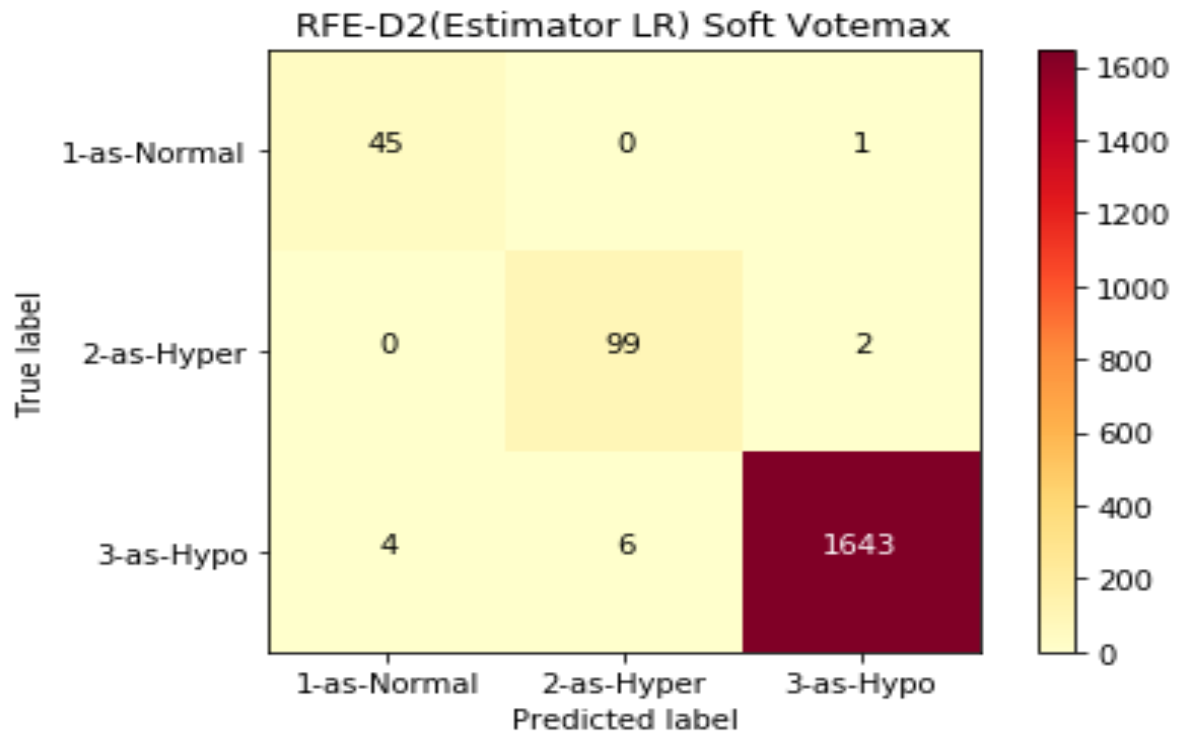
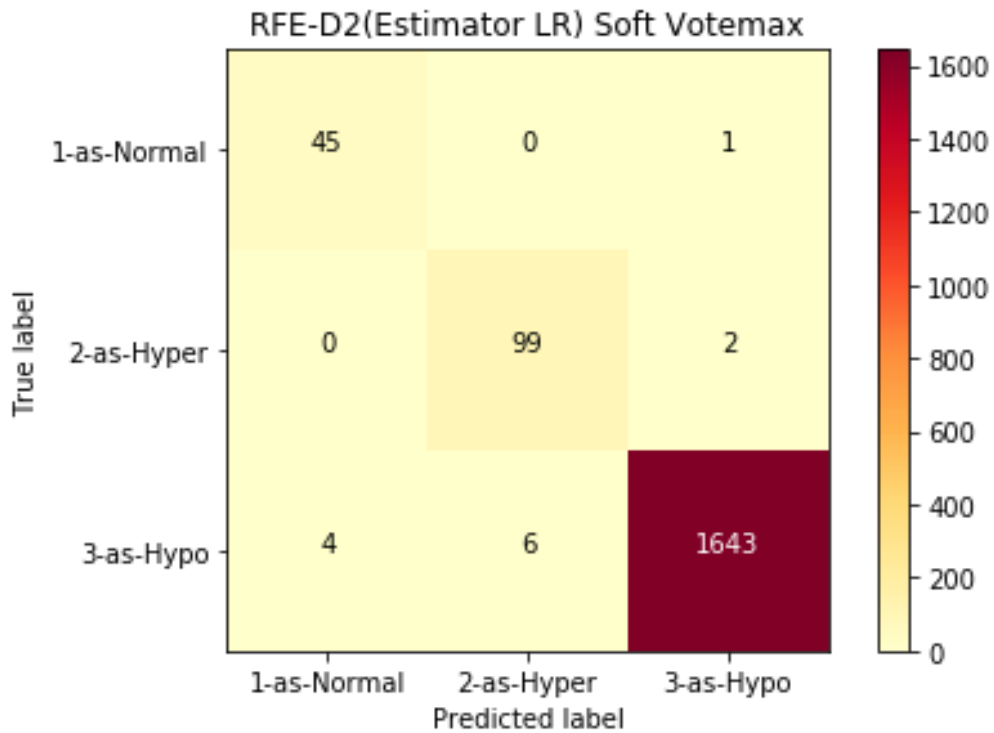


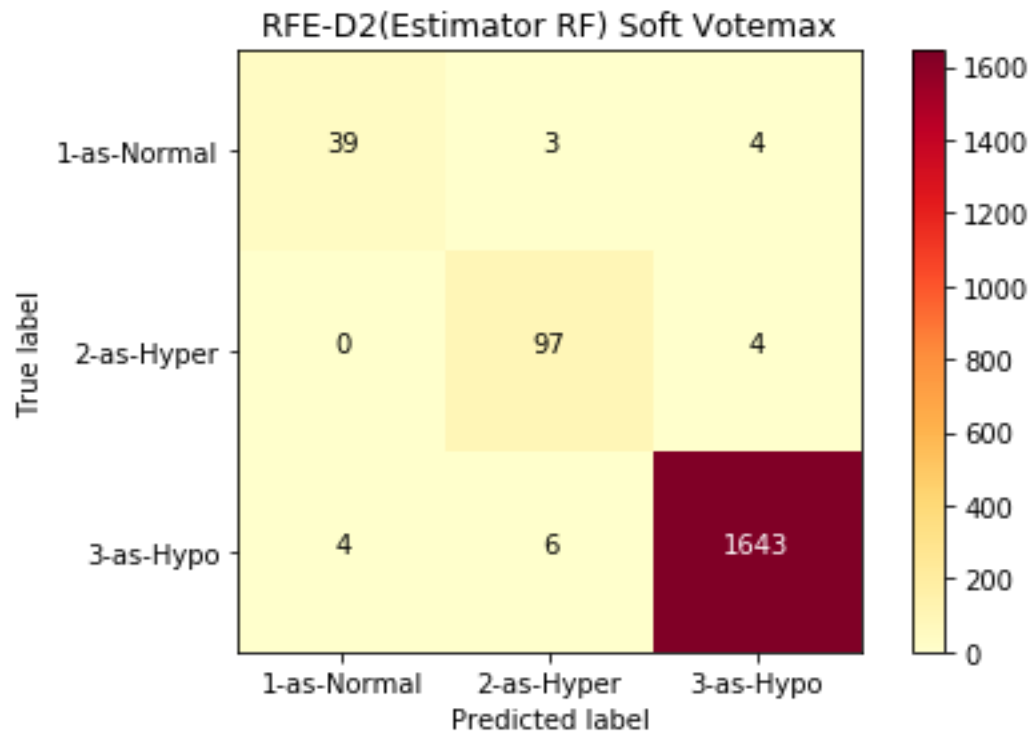
Figure 4.32 Recursive Feature Elimination Based on Gradient Boosting (soft Voting)

The confusion matrix shows the Recursive Feature Elimination based on logistic regression with soft voting 45 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1642 data points truly classified as Hypothyroidism. The figure 4.33 represented as follows:



**Figure 4.33 Recursive Feature Elimination Based on Logistic Regression (soft Voting)**

Figure 4.34 below shows the Recursive Feature Elimination based on random forests with soft voting 39 data points has been truly classified as Normal Patients, while 97 data points truly classified as Hyperthyroidism and 1643 data points truly classified as Hypothyroidism.



**Figure 4.34 Recursive Feature Elimination Based on Random Forests (soft Voting)**



### 4.3.5 Select K-Best Features (soft)

We use the approach of feature selection to pick out the characteristics of our data that have the greatest impact on the target variable. In other words, we select the predictors that are most likely to be associated with the target variable. The Select Best technique selects the features based on which feature has received the highest score out of k. By altering the value of the 'score func' argument, we may use the procedure for both classification and regression data analysis. When preparing a large dataset for training, selecting the optimal features is a critical step in the process. Figure 4.35 below shows the chi square feature selection technique performance in the form of confusion matrix, with soft voting 42 data points has been truly classified as Normal Patients, while 100 data points truly classified as Hyperthyroidism and 1640 data points truly classified as Hypothyroidism.

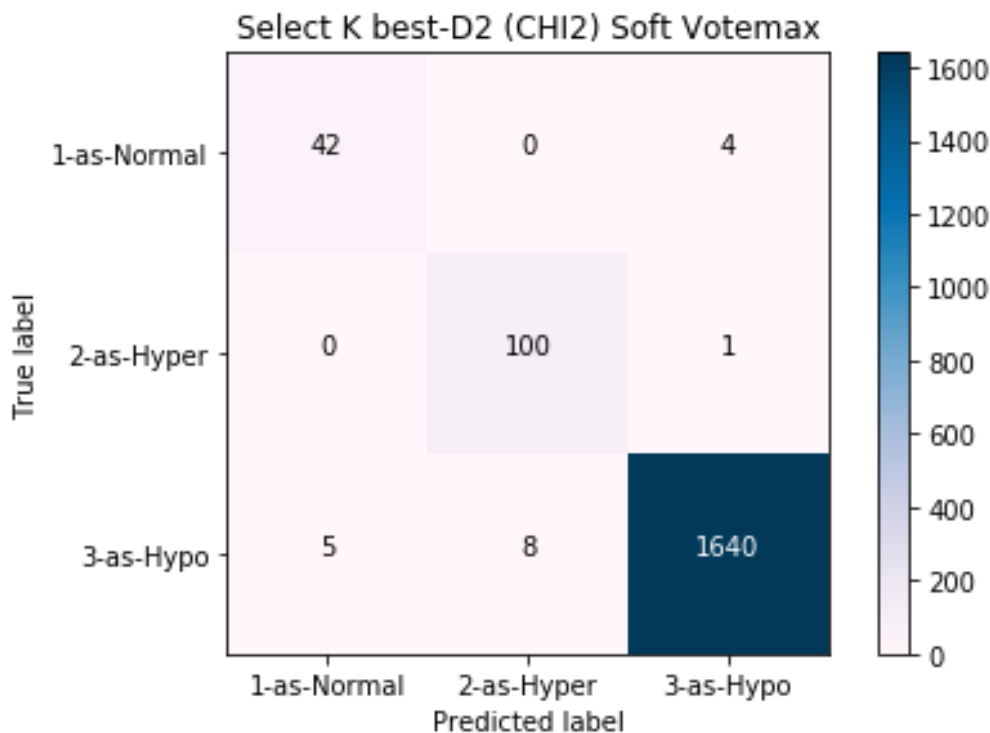
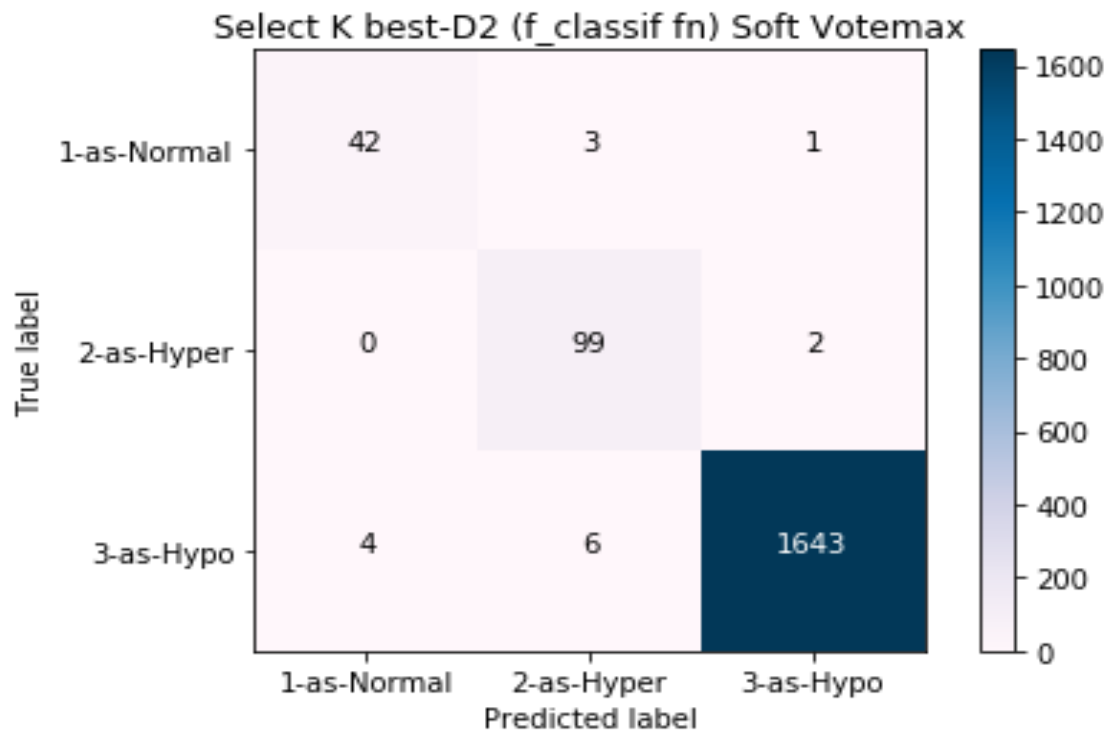


Figure 4.35 Chi Square Feature Selection Technique Performance

Figure 4.36 below shows the FCI feature selection technique performance in the form of confusion matrix, with soft voting 42 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1643 data points truly classified as Hypothyroidism.



**Figure 4.36 FCI Based Feature Selection Technique Performance**

### 4.3.6 Selection of Features from Model (Hard)

By using machine learning Estimators, k-best features are selected. Figure 4.37 below shows the selection of features from decision trees in the form of correctly classified data with soft voting 44 data points has been truly classified as Normal Patients, while 101 data points truly classified as Hyperthyroidism and 1639 data points truly classified as Hypothyroidism.

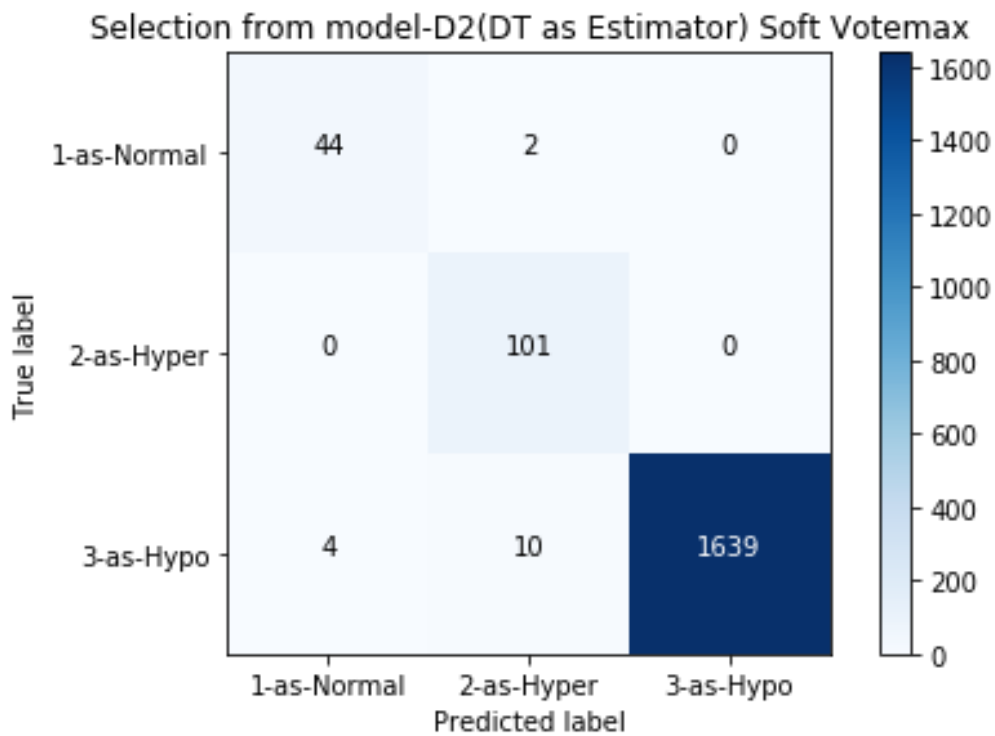
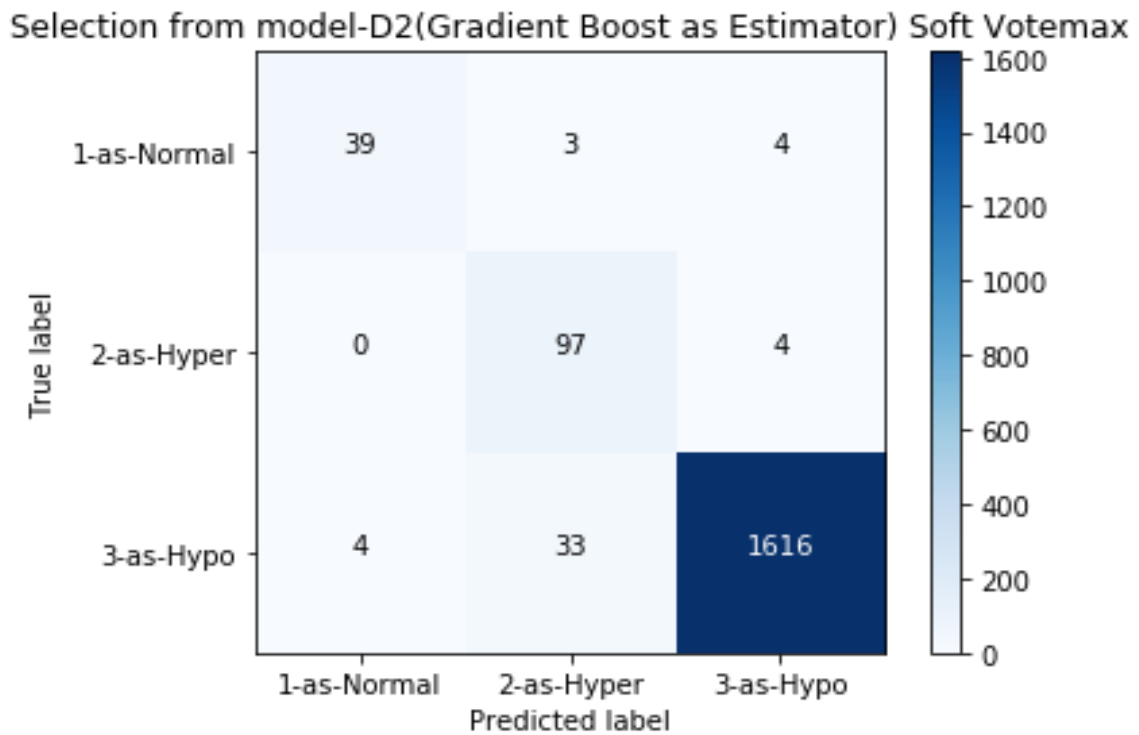


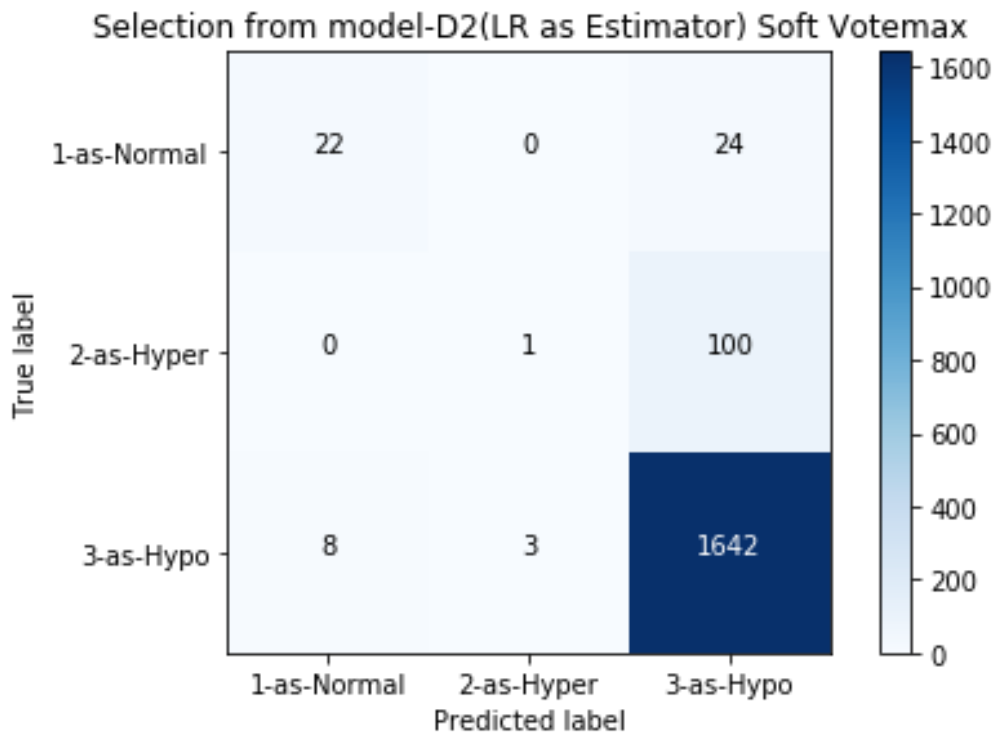
Figure 4.37 Selection of Features from Decision Trees

Figure 4.38 below shows the selection of features from GBC in the form of correctly classified data with soft voting 36 data points has been truly classified as Normal Patients, while 97 data points truly classified as Hyperthyroidism and 1616 data points truly classified as Hypothyroidism.



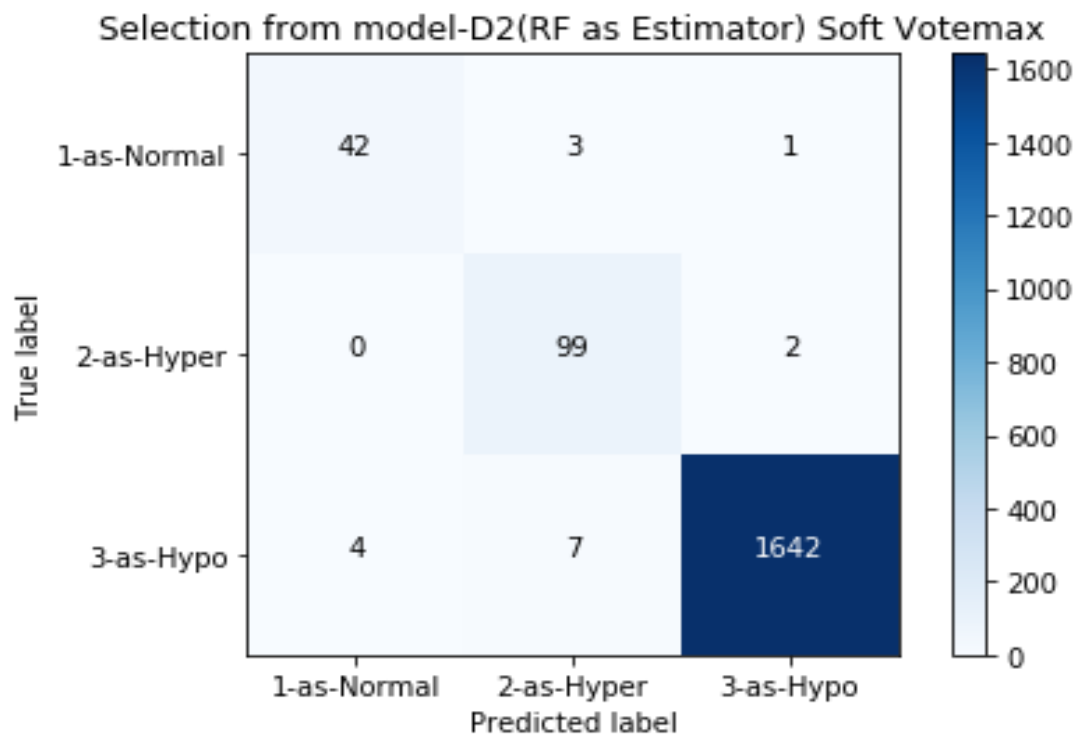
**Figure 4.38 Selection of Features from Gradient Boosting**

Figure 4.39 below shows the selection of features from LR in the form of correctly classified data with soft voting 22 data points has been truly classified as Normal Patients, while 1 data points truly classified as Hyperthyroidism and 1642 data points truly classified as Hypothyroidism.



**Figure 4.39 Selection of Features from Logistic Regression**

Figure 4.40 below shows the selection of features from RF in the form of correctly classified data with soft voting 42 data points has been truly classified as Normal Patients, while 99 data points truly classified as Hyperthyroidism and 1642 data points truly classified as Hypothyroidism.



**Figure 4.40 Selection of Features from Random Forests**

Using the Voting ensemble process, which contains both hard and soft voting procedures, Table 6 displays the last step of the experiment, also known as the Voting ensemble stage. This stage of the assembling process has a larger computing cost than the first, but the proposed method has delivered cutting-edge outcomes with accurate results. The highest accuracy achieved was 99.27 percent, and the other metrics (precision, recall, MCC, kappa, and so on) have also produced the highest results by using this proposed methodology. While using both hard and soft voting techniques to implement the Voting Ensemble of Homogeneous Ensemble, the results are computed with a little delay.

Using the F1 score of the features, the table below shows the relative relevance of the selected features in each attribute selection technique. The F1 score of the features is based on their importance in the selected features. In the first step, we employed Logistic Regression, Gradient Boosting, Random Forests, and Decision Trees to select the feature from model (SFM) that would be used in the second phase of the experiment. In the study, six features were extracted through the use of Logistic Regression, and the accuracy of LR was 92.50 percent with soft voting and 92.44 percent with hard voting, according to the findings. In the case of using RF, the features that were retrieved amounted to five, and the accuracy of RF was 99.05 percent with soft voting and 99.00 percent with hard voting. It was discovered that while using DT, four features could be recovered, and that the accuracy of DT with soft voting was 99.11 percent, while the accuracy of DT with hard voting was 99.22%. Using the GBC, three features were extracted, and the accuracy of the GBC was found to be 97.33 percent when using soft voting and 97.27 percent when using hard voting when soft voting was utilized.

Chi Square and FCI tests were used to determine which K parameters were among the best. The results of this step are shown in the following table. The Chi square was used to determine the best parameters, and the accuracy for hard voting was 99.00 percent, while for soft voting, the accuracy was 99.05 percent, resulting in an accuracy of 99.00 percent. Using hard voting, FCI found the best parameters from a pool of 8 K candidates with an accuracy of 99.05 percent and soft voting with an accuracy of 99.11 percent, respectively.

Implementing Recursive Feature Elimination in the third phase, which is now underway, has been made possible through the application of machine learning Estimators. Hard voting results

in an accuracy of 99.16 percent for hard voting and 99.27 percent for soft voting, with LR extracting 13 features at random with an accuracy of 99.16 percent for hard voting and 99.27 percent for soft voting. If you vote in hard voting, the RF extracts 8 features at random and with 98.7 percent accuracy, however if you vote in soft voting, the accuracy is 98.8 percent. While utilizing hard voting, DT extracts 11 features with a 99 percent accuracy rate, and when using soft voting, the accuracy rate is similarly 99 percent, according to the company. Using a random number generator, GBC extracts nine features with an accuracy of 99.0 percent for hard voting and 99.1 percent for soft voting, respectively. The greatest MCC score was achieved by DT at four features that were selected following a rigorous voting process, earning a score of 95.4 percent on a scale of 100.

#### **4.4. Comparison of proposed methodology with existing studies**

There are numerous strategies that have already been developed that specifically narrate the identification of thyroid ailment by utilizing various ML and DL approaches. Table 7. contains a full comparison of the accuracies of the different methods. Our model achieved an accuracy of 99.27 percent, which is higher than the accuracy of previously reported existing work. Similarly for the Table 8, our proposed methodologies attained the highest accuracy of 100%



**Table 4.6 On dataset 2, performance evaluation metrics for the homogeneous ensemble classifiers' soft and hard voting.**

Features selection techniques	Estimators or functions used	Selected features	Voting Classifiers	Accuracy In %	Training Time (s)	Prediction Time (s)	Recall %	Precision %	F1-score %	MCC %	Cohen Kappa %
Select From Model (SFM)	LR	06	Soft	92.50	1.158	0.053	89.00	93.00	90.00	23.82	31.13
			Hard	92.44	1.973	0.053	93.00	92.00	90.00	24.44	31.15
	RF	05	Soft	99.05	0.6453	0.024	99.00	99.00	99.00	94.01	93.97
			Hard	99.00	1.610	0.025	99.00	99.00	99.00	93.68	93.63
	DT	04	Soft	99.11	1.794	0.033	99.00	99.00	99.00	94.43	94.54
			Hard	99.22	1.795	0.029	99.00	99.00	99.00	95.24	95.12
	GB	03	Soft	97.33	1.597	0.024	98.00	97.00	97.00	84.42	84.00
			Hard	97.27	0.663	0.030	98.00	97.00	97.00	83.73	83.42
Select K Best (SKB)	Chi2	11	Soft	99.05	1.957	0.025	99.00	99.00	99.00	94.06	94.00
			Hard	99.00	1.095	0.021	99.00	99.00	99.00	93.65	93.61
	FCI	08	Soft	99.11	0.820	0.033	99.00	99.00	99.00	94.30	94.34
			Hard	99.05	2.082	0.070	99.00	99.00	99.00	94.01	93.97
Recursive Feature Elimination (RFE)	LR	13	Soft	99.27	0.961	0.018	99.00	99.00	99.00	95.40	95.37
			Hard	99.16	2.496	0.023	99.00	99.00	99.00	94.69	94.66
	RF	08	Soft	98.83	0.863	0.020	99.00	99.00	99.00	92.41	92.41
			Hard	98.72	1.706	0.026	99.00	99.00	99.00	91.69	91.69
	DT	11	Soft	99.00	0.940	0.024	99.00	99.00	99.00	93.62	93.59
			Hard	99.00	1.955	0.023	99.00	99.00	99.00	93.62	93.59
	GB	09	Soft	99.16	0.924	0.023	99.00	99.00	99.00	94.72	94.68
			Hard	99.05	1.911	0.035	99.00	99.00	99.00	93.95	93.93

**Table 4.7 Comparison of accuracies with existing studies on dataset2.**

Datasets	Methodologies	Accuracy	References
KEEL repository thyroid 0387	KNN with Euclidean distance function	96.90	[1]
KEEL repository thyroid 0387	KNN with Minkowski distance function	93.44	[87]
KEEL repository thyroid 0387	(Soft voting)	99.27	This study dataset 2

**Table 4.8 Comparison of accuracies with existing studies on dataset1**

References	Methodology	Accuracy	Recall	F1-score	Training time (s)	Prediction time (s)	Dataset
[12]	WLSVC(L1) + KNN	97.8	96	97	0.53	0.361	DHQ, DG Khan Pakistan
	WLSVC(L2) + DT	76.9	67	61	0.681	0.372	
	WLSVC(L2) + SVM	86.0	79	85	0.511	0.361	
[34]	WCHI+ KNN(Euclidean)	100	100	100	1.032	0.806	DHQ, DG Khan Pakistan
	WCHI+ KNN(Minkowski)	99.3	99	99	1.18	0.827	
	WCHI+ KNN(Chebyshev)	98.7	97	98	1.11	0.808	
	WCHI+ KNN(Manhattan)	99.3	99	99	1.01	0.749	
	WCHI+ KNN(Correlation)	77.3	76	76	0.899	0.655	
This study Dataset 1	Homogenous ensemble + Voting(hard) +SFM(RF)	100	100	100	0.152	0.009	DHQ, DG Khan Pakistan
	Homogenous ensemble + Voting(soft) + SKB(FCI)	100	100	100	0.159	0.013	
	Bagging (BME)+RFE(DT)	100	100	100	0.0129	0.0009	
	Homogenous ensemble + Voting(hard) + RFE(GB)	100	100	100	1.073	0.011	

The boosting classification of AdaBoost (AB) and XGBoost (XGB) in comparison to other Estimators is shown in the table 4.4. In SFM, the accuracy of LR with AB and XGB was 92.66 percent and 92.50 percent, respectively, according to the results. While in RF, the accuracy of the AB ensemble with RF and XGB was 98.55 percent and 99.00 percent, respectively, according to the results. The accuracy of DT with AB was 97.77 percent, and the accuracy of DT with XGB was 98.88 percent. The accuracy of GB with AB was 97.77 percent, and the accuracy of GB with XGB was 97.22 percent.

The two boosting estimators named XGB, and AB are summarised in the table 4.4. As long as you don't use an XGB classifier, all the forecasting models and their associated feature selection procedures are accurate to within a margin of error. A comparison of this study's methodology with other research on the same data set may be found in the preceding table, as shown. Findings were achieved by integrating multiple homogeneous ensemble algorithms (bagging, boosting) with several attribute selection processes in proposed research. During this project, the researchers aimed to increase accuracy while simultaneously reducing training and prediction time. An additional step of ensemble included voting (soft and hard) in order to accomplish the hybrid implementation of different selecting features, outliers' and anomalies detection, as well as the overall estimate. Our proposed hybrid Estimator technique outperformed earlier hybrid Estimators, with improved F1-scores and recall as well as accuracy, and of 100%, for the same degree of accuracy, but needing less learning and predicting time. To train and verify results, present techniques need substantially more time and resources, as well as being significantly more costly to implement.

Human life depends on the capacity to recognise and identify illness at an early stage. Machine learning methods have made identification more exact and accurate. Because of the similar symptoms, it may be difficult to tell the difference between thyroid disease and other conditions. Data from the thyroid dataset has shown a positive impact on classifier performance that beats earlier studies in terms of accuracy. Multi-stage ensemble i.e., Votmax ensemble of homogenous ensemble along with three different feature selection techniques were the goals of this research effort. Pre-processing and outlier detection from the provided attributes must be completed before the classification procedure can begin. In the first assembly phase, each of the ensembles contributed two algorithms. As part of the classifying procedure, an ensemble

of both voting techniques has been used. The maximum feasible accuracy of 100 percent was achieved by combining all feature engineering strategies with multiple estimators and ensemble procedures while using the least computer resources, according to the findings. It was also found to be perfect in terms of the other performance assessment metrics that had been used. According to our findings, our method was the most effective in analysing the thyroid illness dataset.

## **CHAPTER 5 CONCLUSION**

This dissertation is based on the implementation of different popular ensemble algorithms from bagging and boosting techniques. Previously the researchers have had a great deal of success in identifying thyroid diseases; nonetheless, it is recommended that patients take advantage of the large number of criteria available for diagnosing thyroid disorders. The addition of further criteria would involve the administration of more clinical tests on individuals, which would be both expensive and time-consuming. As a consequence, such forecasting must be developed in such a manner that they require the fewest number of parameters possible in order to diagnose diseases while also conserving both time and money for the patients involved. When compared to prior studies, this research study's dataset offers less, however it contains traits that are extremely essential and valuable in improving illness detection. The two datasets have been used in this dissertation one is open source and other is the local hospital from DG Khan. Cleaning the data samples before modelling is crucial for ensuring that the data is representative of the scenario as accurately as possible. It is fairly uncommon for a dataset to contain missing eigenvalues that do not fit in with the overall data set. To improve the accuracy of the developed models, it has been proven that outlier values may be identified and deleted from the model's dataset. Finding and restoring incomplete data and outliers with average values of the attributes that were utilised are early pre-processing operations in this research. With the help of characteristic relevance ratings, as well as the projected feature significance derived from a dataset, the feature selection method may be made better. Attributes are selected from the training dataset, the model is trained using features selected, and the model is evaluated using test data in this stage. It is possible to automatically choose the qualities in given dataset that have the greatest impact on your output variables. The XGBoost feature significance test was performed on both datasets prior to selecting the most important features. Many models may result in poor performances due to the obvious presence of irrelevant characteristics in your data. In addition to improving accuracy, feature extraction method prior to modelling saves training and the risk of overfitting. Recursive feature elimination (RFE) was used to choose from model and select K-best (SKB) in this study.

Bagging and Boosting Homogeneous ensembles are introduced in this experimental work by proposing classifiers for further processing by a voting ensemble that includes both hard and soft votes. Using both intended and open-source datasets, this method yielded discoveries that were at the cutting edge. Recall, hamming loss and accuracy are only a few of the performance parameters used to evaluate overall performance.

After the comparison of deployed methodologies in this experimental work our proposed approach attains the best performance on both datasets. By comparing the proposed methodology results with already available existing work, we realized our proposed augmentation approach achieved an improvement in accuracy, sensitivity, AUC, lower error and miss-classification rate. A state-of-the-art result can be obtained through proposed augmentation approaches.

In future we will try to introduce the augmentation that not only capable to improve the accuracies of the ML models but also increase the total number of instances or entries, that will rapidly improve the performance of the proposed algorithms or methodologies. Such type of approach is also very helpful where the datasets are very small and in a limited number of entries.

## REFERENCES

- [1] R. Pal, T. Anand, and S. K. Dubey, "Evaluation and performance analysis of classification techniques for thyroid detection," *Int. J. Bus. Inf. Syst.*, vol. 28, no. 2, pp. 163–177, 2018.
- [2] S. Chowdhury and P. pratim Chakraborty, "Universal health coverage - There is more to it than meets the eye," *J. Fam. Med. Prim. Care*, vol. 6, no. 2, pp. 169–170, 2017.
- [3] A. K. Singh, "A Comparative Study on Disease Classification using Machine Learning Algorithms," *SSRN Electron. J.*, 2019.
- [4] S. Park, B. Shin, W. Sang Shim, Y. Choi, K. Kang, and K. Kang, "Wx: a neural network-based feature selection algorithm for transcriptomic data," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, 2019.
- [5] N. Yukinawa, S. Oba, K. Kato, and S. Ishii, "Optimal aggregation of binary classifiers for multiclass cancer diagnosis using gene expression profiles," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 6, no. 2, pp. 333–343, 2009.
- [6] M. Ghazal, G. Giridharan, and A. El-baz, "Cancer Using Multi-Input CNN," pp. 1–13, 2021.
- [7] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. P. L. F. De Carvalho, "Effectiveness of Random Search in SVM hyper-parameter tuning," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2015-Sept, 2015.
- [8] F. Ladich and H. Winkler, "Acoustic communication in terrestrial and aquatic vertebrates," *J. Exp. Biol.*, vol. 220, no. 13, pp. 2306–2317, 2017.
- [9] M. A. Moreno-Ibarra, Y. Villuendas-Rey, M. D. Lytras, C. Yáñez-Márquez, and J. C. Salgado-Ramírez, "Classification of diseases using machine learning algorithms: A comparative study," *Mathematics*, vol. 9, no. 15, pp. 1–21, 2021.

- [10] A. F. Ghias, G. Epps, E. Cottrill, and S. K. Mardekian, "Multifocal Metastatic Breast Carcinoma to the Thyroid Gland Histologically Mimicking C Cell Lesions," *Case Rep. Pathol.*, vol. 2019, pp. 1–5, 2019.
- [11] A. Helmy, M. Holdmann, and M. Rizkalla, "Application of thermography for non-invasive diagnosis of thyroid gland disease," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 1168–1175, 2008.
- [12] H. Abbad Ur Rehman, C. Y. Lin, Z. Mushtaq, and S. F. Su, "Performance Analysis of Machine Learning Algorithms for Thyroid Disease," *Arab. J. Sci. Eng.*, vol. 46, no. 10, pp. 9437–9449, 2021, doi: 10.1007/s13369-020-05206-x
- [13] F. Cumbo, E. Cappelli, and E. Weitschek, "A brain-inspired hyperdimensional computing approach for classifying massive DNA methylation data of cancer," *Algorithms*, vol. 13, no. 9, pp. 1–13, 2020.
- [14] K. L. Tsui, S. Y. Wong, W. Jiang, and C. J. Lin, "Recent research and developments in temporal and spatiotemporal surveillance for public health," *IEEE Trans. Reliab.*, vol. 60, no. 1, pp. 49–58, 2011.
- [15] M. Feigin, D. Freedman, and B. W. Anthony, "A Deep Learning Framework for Single-Sided Sound Speed Inversion in Medical Ultrasound," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 4, pp. 1142–1151, 2020.
- [16] M. Education and M. Ramya, "Prediction and Providing Medication for Thyroid Disease Using Machine Learning," vol. 11, pp. 1099–1107, 2020.
- [17] A. Tyagi, R. Mehra, and A. Saxena, "Interactive thyroid disease prediction system using machine learning technique," *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 689–693, 2018.
- [18] S. Razia and M. R. Narasinga Rao, "Machine learning techniques for thyroid disease diagnosis - A review," *Indian J. Sci. Technol.*, vol. 9, no. 28, 2016.



- [19] C. S. Nageswari, M. N. V. Kumar, C. Raveena, J. S. Sharma, and M. Y. Devi, “An Identification and Classification of Thyroid Diseases Using Deep Learning Methodology,” *Rev. Gestão Inovação e Tecnol.*, vol. 11, no. 2, pp. 2004–2015, 2021.
- [20] L. N. Li, J. H. Ouyang, H. L. Chen, and D. Y. Liu, “A computer aided diagnosis system for thyroid disease using extreme learning machine,” *J. Med. Syst.*, vol. 36, no. 5, pp. 3327–3337, 2012.
- [21] C. R, C. Vasan, C. MS, and D. H S, “Thyroid Detection Using Machine Learning,” *Int. J. Eng. Appl. Sci. Technol.*, vol. 5, no. 9, pp. 173–177, 2021.
- [22] H. Kour and V. Sharma, “Performance Evaluation of SVM and Random Forest for the Diagnosis of Thyroid Disorder,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. 5, pp. 945–947, 2021.
- [23] A. Prochazka, S. Gulati, S. Holinka, and D. Smutek, “Patch-based classification of thyroid nodules in ultrasound images using direction independent features extracted by two-threshold binary decomposition,” *Comput. Med. Imaging Graph.*, vol. 71, pp. 9–18, 2019.
- [24] K. Salman and E. Sonuc, “Thyroid Disease Classification Using Machine Learning Algorithms,” *J. Phys. Conf. Ser.*, vol. 1963, no. 1, 2021.
- [25] L. C. Zhu *et al.*, “A model to discriminate malignant from benign thyroid nodules using artificial neural network,” *PLoS One*, vol. 8, no. 12, pp. 1–6, 2013.
- [26] K. Geetha and C. S. S. Baboo, “An Empirical Model for Thyroid Disease,” *Glob. J. Comput. Sci. Technol. E Network, Web Secur.*, vol. 16, no. 1, pp. 242–250, 2016.
- [27] A. Placzek, A. Pluciennik, A. Kotecka-Blicharz, M. Jarzab, and D. Mrozek, “Bayesian Assessment of Diagnostic Strategy for a Thyroid Nodule Involving a Combination of Clinical Synthetic Features and Molecular Data,” *IEEE Access*, vol. 8, pp. 175125–175139, 2020.
- [28] S. Draghici, “Machine learning techniques,” *Stat. Data Anal. Microarrays Using R Bioconductor*, pp. 999–1024, 2020.

- [29] A. Professor, G. Singh, U. Anil Kumar, P. Ramesh, and A. Kumar Dubey, “Comparative Analysis of Thyroid Disease based on Hormone Level using Data Mining Techniques,” vol. 8, no. 14, pp. 30–34, 2020, [Online]. Available: [www.ijert.org](http://www.ijert.org).
- [30] P. Poudel, A. Illanes, E. J. G. Ataide, N. Esmaceli, S. Balakrishnan, and M. Friebe, “Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches,” *IEEE Access*, vol. 7, no. MI, pp. 79354–79365, 2019, doi: 10.1109/ACCESS.2019.2923547.
- [31] M. R. N. Kousarrizi, F. Seiti, and M. Teshnehlal, “An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification,” *Int. J. Electr. Comput. Sci.*, vol. 12, no. 01, pp. 13–19, 2012, [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.655.363&rep=rep1&type=pdf>.
- [32] S. S. Zarin Mousavi, M. Mohammadi Zanjireh, and M. Oghbaie, “Applying computational classification methods to diagnose Congenital Hypothyroidism: A comparative study,” *Informatics Med. Unlocked*, vol. 18, no. October 2019, p. 100281, 2020, doi: 10.1016/j.imu.2019.100281.
- [33] K. Dharmarajan, K. Balasree, A. S. Arunachalam, and K. Abirmai, “Thyroid Disease Classification Using Decision Tree and SVM,” *Indian J. Public Heal. Res. Dev.*, vol. 11, no. 03, pp. 224–229, 2020, doi: 10.37506/ijphrd.v11i3.822.
- [34] H. Abbad Ur Rehman, C. Y. Lin, and Z. Mushtaq, “Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease,” *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A*, vol. 44, no. 1, pp. 77–87, 2021, doi: 10.1080/02533839.2020.1831967.
- [35] S. Mishra, P. K. Mallick, H. K. Tripathy, A. K. Bhoi, and A. González-Briones, “Performance evaluation of a proposed machine learning model for chronic disease datasets using an integrated attribute evaluator and an improved decision tree classifier,” *Appl. Sci.*, vol. 10, no. 22, pp. 1–35, 2020, doi: 10.3390/app10228137.

- [36] P. Luo, L. P. Tian, J. Ruan, and F. X. Wu, "Disease Gene Prediction by Integrating PPI Networks, Clinical RNA-Seq Data and OMIM Data," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 16, no. 1, pp. 222–232, 2019, doi: 10.1109/TCBB.2017.2770120.
- [37] B. Colakoglu, D. Alis, and M. Yergin, "Diagnostic Value of Machine Learning-Based Quantitative Texture Analysis in Differentiating Benign and Malignant Thyroid Nodules," *J. Oncol.*, vol. 2019, 2019, doi: 10.1155/2019/6328329.
- [38] A. A. Sheeja and S. Suresh Babu, "A Review of Thyroid Disorder Detection, Segmentation and Classification on Medical Images," *Int. J. Eng. Adv. Technol.*, no. 2, p. 88, 2013.
- [39] F. Abdolali *et al.*, "A systematic review on the role of artificial intelligence in sonographic diagnosis of thyroid cancer: Past, present and future," *Front. Biomed. Technol.*, vol. 7, no. 4, pp. 266–280, 2020, doi: 10.18502/fbt.v7i4.5324.
- [40] D. Avola, L. Cinque, A. Fagioli, S. Filetti, G. Grani, and E. Rodola, "Multimodal Feature Fusion and Knowledge-Driven Learning via Experts Consult for Thyroid Nodule Classification," *IEEE Trans. Circuits Syst. Video Technol.*, 2021, doi: 10.1109/TCSVT.2021.3074414.
- [41] D. Dov *et al.*, "Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images," *Med. Image Anal.*, vol. 67, 2021, doi: 10.1016/j.media.2020.101814.
- [42] M. Gadermayr *et al.*, "Frozen-to-Paraffin: Categorization of Histological Frozen Sections by the Aid of Paraffin Sections and Generative Adversarial Networks," 2020, doi: 10.1007/978-3-030-87592-3\_10.
- [43] D. Dov, S. Kovalsky, J. Cohen, D. Range, R. Henao, and L. Carin, "Thyroid Cancer Malignancy Prediction From Whole Slide Cytopathology Images," pp. 1–14, 2019, [Online]. Available: <http://arxiv.org/abs/1904.00839>.

- [44] Y. Rivenson, H. Wang, Z. Wei, Y. Zhang, H. Gunaydin, and A. Ozcan, “Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue,” 2018, [Online]. Available: <http://arxiv.org/abs/1803.11293>.
- [45] S. Bhalla, H. Kaur, R. Kaur, S. Sharma, and G. P. S. Raghava, “Expression based biomarkers and models to classify early and late-stage samples of Papillary Thyroid Carcinoma,” *PLoS One*, vol. 15, no. 4, pp. 1–24, 2020, doi: 10.1371/journal.pone.0231629.
- [46] X. Chai, “Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning,” *IEEE Access*, vol. 8, pp. 149787–149795, 2020, doi: 10.1109/ACCESS.2020.3016676.
- [47] F. Marinozzi and P. Trimboli, “Artificial Intelligence in Thyroid Field — A Comprehensive Review,” pp. 1–18, 2021.
- [48] E. H. Wu, L. Huang, Y. Zhou, and X. Zhu, “Migratory Fish Bone in the Thyroid Gland: Case Report and Literature Review,” *Case Rep. Med.*, vol. 2018, 2018, doi: 10.1155/2018/7345723.
- [49] A. J. Santos *et al.*, “Modeling differential rates of aging using routine laboratory data; Implications for morbidity and health care expenditure,” 2021, [Online]. Available: <http://arxiv.org/abs/2103.09574>.
- [50] N. Asaad Sajadi, S. Borzouei, H. Mahjub, and M. Farhadian, “Diagnosis of hypothyroidism using a fuzzy rule-based expert system,” *Clin. Epidemiol. Glob. Heal.*, vol. 7, no. 4, pp. 519–524, 2019, doi: 10.1016/j.cegh.2018.11.007.
- [51] U. Machine and L. Technique, “IoT Based Health — Related Topic Recognition from Emerging Online Health Community ( Med Help ),” pp. 1–15.
- [52] P. Poudel, A. Illanes, D. Sheet, and M. Friebe, “Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches,” *J. Healthc. Eng.*, vol. 2018, 2018, doi: 10.1155/2018/8087624.

- [53] J. E. Rosen *et al.*, “Preoperative discrimination of benign from malignant disease in thyroid nodules with indeterminate cytology using elastic light-scattering spectroscopy,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 8, pp. 2336–2340, 2014, doi: 10.1109/TBME.2013.2267452.
- [54] F. Yang *et al.*, “Identification of Potential lncRNAs and miRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning,” *Int. J. Endocrinol.*, vol. 2021, 2021, doi: 10.1155/2021/3984463.
- [55] S. Kurnaz, M. S. Mohammed, and S. J. Mohammed, “A high efficiency thyroid disorders prediction system with non-dominated sorting genetic algorithm NSGA-II as a feature selection algorithm,” *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–6, 2020, doi: 10.1109/INCET49848.2020.9154189.
- [56] M. Ihara, K. Ashizawa, K. Shichijo, and T. Kudo, “Expression of the DNA-dependent protein kinase catalytic subunit is associated with the radiosensitivity of human thyroid cancer cell lines,” *J. Radiat. Res.*, vol. 60, no. 2, pp. 171–177, 2019, doi: 10.1093/jrr/rry097.
- [57] L. Evans, S. Park, C. Elliott, and C. Garrett, “Ectopic Papillary Carcinoma in the Midline Neck Accompanied by a Benign Thyroid Gland,” *Case Rep. Otolaryngol.*, vol. 2019, pp. 1–3, 2019, doi: 10.1155/2019/9172942.
- [58] M. De Oliveira *et al.*, “Disruptive effect of organotin on thyroid gland function might contribute to hypothyroidism,” *Int. J. Endocrinol.*, vol. 2019, 2019, doi: 10.1155/2019/7396716.
- [59] B. Jarrahi, R. Gassert, J. Wanek, L. Michels, U. Mehnert, and S. S. Kollias, “Design and Application of a New Automated Fluidic Visceral Stimulation Device for Human fMRI Studies of Interoception,” *IEEE J. Transl. Eng. Heal. Med.*, vol. 4, no. February, pp. 1–12, 2016, doi: 10.1109/JTEHM.2016.2538239.
- [60] S. Chaganti *et al.*, “Improve Diagnostic Classification using Medical Image Computing,” 2017.

- [61] E. J. Gomes Ataide, N. Ponugoti, A. Illanes, S. Schenke, M. Kreissl, and M. Friebe, “Thyroid nodule classification for physician decision support using machine learning-evaluated geometric and morphological features,” *Sensors (Switzerland)*, vol. 20, no. 21, pp. 1–14, 2020, doi: 10.3390/s20216110.
- [62] M. Van Tieu, A. Go, Y. J. Park, H. V. Nguyen, S. Y. Hwang, and M. H. Lee, “Highly sensitive ELISA using membrane-based microwave-mediated electrochemical immunoassay for thyroid-stimulating hormone detection,” *IEEE Sens. J.*, vol. 19, no. 21, pp. 9826–9831, 2019, doi: 10.1109/JSEN.2019.2925020.
- [63] C. Song, Y. Yang, X. Tu, Z. Chen, J. Gong, and C. Lin, “A Smartphone-Based Fluorescence Microscope with Hydraulically Driven Optofluidic Lens for Quantification of Glucose,” *IEEE Sens. J.*, vol. 21, no. 2, pp. 1229–1235, 2021, doi: 10.1109/JSEN.2020.3019889.
- [64] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demene, “Adaptive Spatiotemporal SVD Clutter Filtering for Ultrafast Doppler Imaging Using Similarity of Spatial Singular Vectors,” *IEEE Trans. Med. Imaging*, vol. 37, no. 7, pp. 1574–1586, 2018, doi: 10.1109/TMI.2018.2789499.
- [65] S. Ghavami, M. Bayat, M. Fatemi, and A. Alizad, “Quantification of morphological features in non-contrast-enhanced ultrasound microvasculature imaging,” *IEEE Access*, vol. 8, pp. 18925–18937, 2020, doi: 10.1109/ACCESS.2020.2968292.
- [66] A. M. Joshi, P. Jain, S. P. Mohanty, and N. Agrawal, “IGLU 2.0: A New Wearable for Accurate Non-Invasive Continuous Serum Glucose Measurement in IoMT Framework,” *IEEE Trans. Consum. Electron.*, vol. 66, no. 4, pp. 327–335, 2020, doi: 10.1109/TCE.2020.3011966.
- [67] Y. L. Li and J. J. Dahl, “Coherent flow power doppler (CFPD): Flow detection using spatial coherence beamforming,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 6, pp. 1022–1035, 2015, doi: 10.1109/TUFFC.2014.006793.

- [68] N. Bussmann *et al.*, “Development of a high resolution, 2D detector system for iodine-131-scintigraphy,” *IEEE Trans. Nucl. Sci.*, vol. 47, no. 3 PART 3, pp. 1085–1087, 2000, doi: 10.1109/23.856551.
- [69] J. Rouyer, T. Cueva, T. Yamamoto, A. Portal, and R. J. Lavarello, “In Vivo Estimation of Attenuation and Backscatter Coefficients from Human Thyroids,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 63, no. 9, pp. 1253–1261, 2016, doi: 10.1109/TUFFC.2016.2532932.
- [70] B. Pandiyan, S. J. Merrill, and S. Benvenga, “A patient-specific model of the negative-feedback control of the hypothalamus-pituitary-thyroid (HPT) axis in autoimmune (Hashimoto’s) thyroiditis,” *Math. Med. Biol.*, vol. 31, no. 3, pp. 226–258, 2014, doi: 10.1093/imammb/dqt005.
- [71] Alcalá-Fdez, J.; Fernández, A.; Luengo, J.; Derrac, J.; García, S.; Sánchez, L.; Herrera, F. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *J. Mult.-Valued Log. Soft Comput.* 2011, 17, 255–287.
- [72] J. Brownlee, “Feature Importance and Feature Selection With XGBoost in Python,” 2016. [Online]. Available: <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/>. [Accessed: 15-Oct-2021].
- [73] F. Pedregosa, R. Weiss, and M. Brucher, “Scikit-learn : Machine Learning in Python,” *J. of Machine Learn. Res.* 12, vol. 12, pp. 2825–2830, 2011.
- [74] A. F. Seddik and D. Shawky, Muhammad, “Logistic Regression Model for Breast Cancer Automatic Diagnosis,” in *SAI, Intelligent Systems Conference*, 2015, pp. 150–154.
- [75] L. Universit, M. Curie, P. V. I. Bo, P. Cedex, and B. Yu, “Analysis of a Random Forests Model,” *J. Mach. Learn. Res.*, vol. 13, pp. 1063–1095, 2012.
- [76] P. Sathiyarayanan, P. S, S. Saranya, and M. M, “Identification of Breast Cancer Using The Decision Tree Algorithm,” in *IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1–6.

- [77] A. Natekin and A. Knoll, "Gradient boosting machines , a tutorial," *Front. Neurorobot.*, vol. 7, no. December, 2013.
- [78] J. Brownlee, "4 Automatic Outlier Detection Algorithms in Python," 08-Jul-2020. [Online]. Available: <https://machinelearningmastery.com/model-based-outlier-detection-and-removal-in-python/>. [Accessed: 16-Oct-2021].
- [79] A. Lal and B. Datta, "Performance Evaluation of Homogeneous and Heterogeneous Ensemble Models for Groundwater Salinity Predictions : a Regional-Scale Comparison Study," *Water Air Soil Pollut.*, vol. 231, no. 320, 2020.
- [80] D. Opitz and R. Maclin, "Popular Ensemble Methods : An Empirical Study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, 1999.
- [81] Z. Alam, M. S. Rahman, and M. S. Rahman, "Informatics in Medicine Unlocked A Random Forest based predictor for medical data classification using feature ranking," *Informatics Med. Unlocked*, vol. 15, no. April, p. 100180, 2019.
- [82] H. AB, "Meta-analysis in medical research," *Hippokratia*, vol. 14, pp. 29–37, 2010.
- [83] K. Srimani, P and S. Koti, "Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Approach," *J. Adv. Comput.*, no. July, 2016.
- [84] X. Y. Liew, N. Hameed, and J. Clos, "Machine Learning with Applications An investigation of XGBoost-based algorithm for breast cancer classification," *Mach. Learn. with Appl.*, vol. 6, no. April, p. 100154, 2021.
- [85] J. Brownlee, "How to Develop Voting Ensembles With Python," 17-Apr-2021. [Online]. Available:<https://machinelearningmastery.com/voting-ensembles-with-python/>. [Accessed: 15-Oct-2021].
- [86] Z. Mushtaq, A. Yaqub, A. Hassan, and S. F. Su, "Performance analysis of supervised classifiers using PCA based techniques on breast cancer," in *International Conference on Engineering and Emerging Technologies, ICEET*, 2019.



[87] Chandel, K.; Kunwar, V.; Sabitha, S.; Choudhury, T.; Mukherjee, S.: A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques CSI Trans. 4(2–4), 313–319 (2016)