PART 1

INTRODUCTION

Thyroid disease is a subset of endocrinology, one of the most misunderstood and underdiagnosed diseases [1]. According to the World Health Organization, thyroid disease is one of the most common endocrine diseases in the world, coming second only to diabetes. Hyper function hyperthyroidism and hypothyroidism affect about 2% and 1% of individuals, respectively. Men have about a tenth of the prevalence of women. Hyper and hypothyroidism may be caused by thyroid gland dysfunction, secondary to pituitary gland failure, or tertiary to hypothalamic malfunction [2]. Due to dietary iodine deficiency, goiter or active thyroid nodules may become prevalent in some regions, with a prevalence of up to 15%. The thyroid gland can also be the location of different kinds of tumors and dangerous where endogenous antibodies wreak havoc [3].

The thyroid gland is an endocrine gland that secretes hormones and passes them through the bloodstream. The thyroid gland, shaped like a butterfly, is situated near the base of the throat. It consists of two active thyroid hormones, levothyroxine (T4) and triiodothyroxine (T3), which are involved in brain activities like body temperature regulation, blood pressure control, and heart rate regulation.

Thyroid gland hormones are responsible for aiding in digestion and maintaining the body moist, balanced, and so on. Thyroid gland treatments such as T3 (triiodothyronine), T4 (thyroid hormone), and TSH (thyroid-stimulating hormone) are used to assess thyroid activity (thyroid-stimulating hormone). Thyroid disorder is classified into two types: hypothyroidism and hyperthyroidism.

With both types, whether it is Hyperthyroidism or Hypothyroidism, if the normal person suffers, either of the two cases, this affects the person and causes health problems.

Table 1.1. Normal thyroid function test.

Test	Normal Value
TSH	(0.2 - 6.0) ulU/L
T3	(0.9 - 2.35) nmol/L
T4	(60 – 120) nmol/L

Hyperthyroidism is a disorder in which the thyroid gland releases so many thyroid hormones. Hyperthyroidism is caused by an increase in thyroid hormone levels[4]. Dry skin, elevated temperature sensitivity, hair thinning, Weight loss, increased heart rate, high blood pressure, heavy sweating, swollen neck, nervousness, menstruation cycle shortening, irregular stomach movements, and handshaking are some signs[5]. A decline in thyroid hormone production causes hypothyroidism. Hypo means deficient or less in medical terms. Inflammation and thyroid gland injury are the two primary causes of hypothyroidism. Obesity, low heart rate, increased temperature sensitivity, neck swelling, dry skin, hand numbness, hair issues, heavy menstrual cycles, and intestinal problems. If not treated, these symptoms can escalate over time [6].

Early disease detection, diagnosis, and care, according to doctors, are vital in preventing disease progression and even death. For several different forms of anomalies, early identification and differential diagnosis raise the odds of good treatment. Despite multiple trials, clinical diagnosis is often thought to be a difficult task [7]. Data mining is a semi-automated method of looking for correlations in massive datasets[8]. Machine learning algorithms are one of the best solutions for most complex problems [9]. Classification is a data extraction technique (machine learning) used to predict and identify many diseases, such as thyroid disease. We researched and classified here because machine learning algorithms play a significant role in classifying thyroid disease and because these algorithms are high performing and efficient and aid in classification[10]. Although the application of computer learning and artificial intelligence in medicine dates back to the early days of the field, there has been a new movement to consider the need for machine learning-driven healthcare

solutions [11]. As a result, analysts predict that machine learning will become commonplace in healthcare soon [12].

1.1 THYROID DISEASE

The thyroid gland is an endocrine gland divided into the right and left lobes, which are situated on opposite sides of the trachea in the throat, with an isthmus connecting them [13]. It has the appearance of a butterfly gland and weighs around 25 grams in adults. When seen from the front, it is positioned under the chin, as shown in Figure 1.1. It is found just below the cartilage protrusion, which is more pronounced in males and is referred to as Adam's Apple by the public, and it helps with swallowing. While it is small in size, the hormones it secretes play a critical role in human growth and development, and it is also known as the "regulator of all body functions [14].

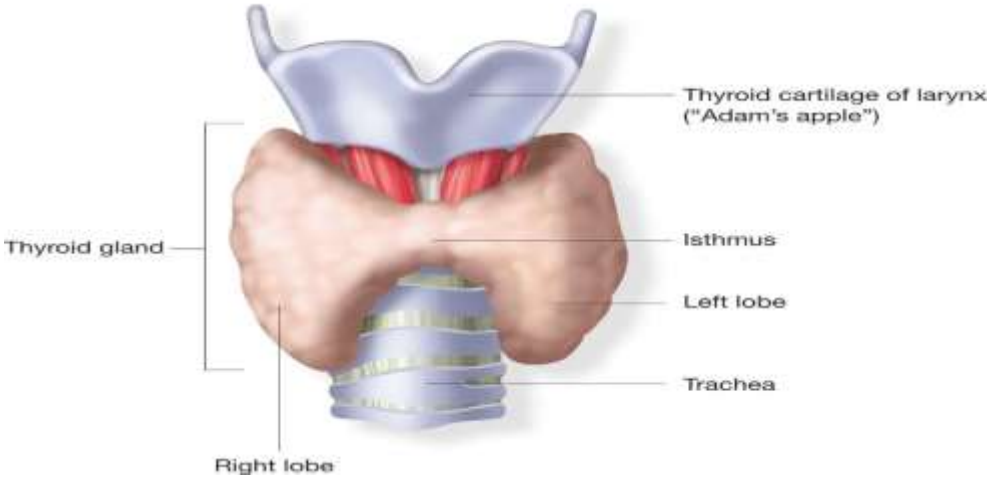

Figure 1.1. Location of thyroid gland[15].

The thyroid gland. It has two types of secretions within its functioning system, triiodothyronine (T3) and another is thyroxine (T4) [16]. And the element used for these two hormones is iodine in the blood because iodine is the main component to building these two hormones, T3 and T4. As for (T3), this label came because it consists of three atoms of iodine, while (T4) this label came because it consists of four atoms of iodine, and the critical role of these hormones is to control the metabolism process [17]. TRH (TSH Releasing Hormone) is generated in the hypothalamus,

located in the upper part of the brain and secreted by the thyroid gland. Thyroid hormone levels in the blood affect the volume of TRH hormone secreted by the hypothalamus. This hormone stimulates the pituitary gland, located in the lower part of the brain, to manufacture thyroid-stimulating hormone (TSH). The pituitary gland secretes the thyroid-stimulating hormone. This hormone is the central part of the synthesis of T4 (thyroxine). And also enters the formation of a small amount of the hormone T3 (triiodothyronine) and then excreted into the bloodstream. T3 and T4 development occur until the body's requirements decide the optimum blood level [18].

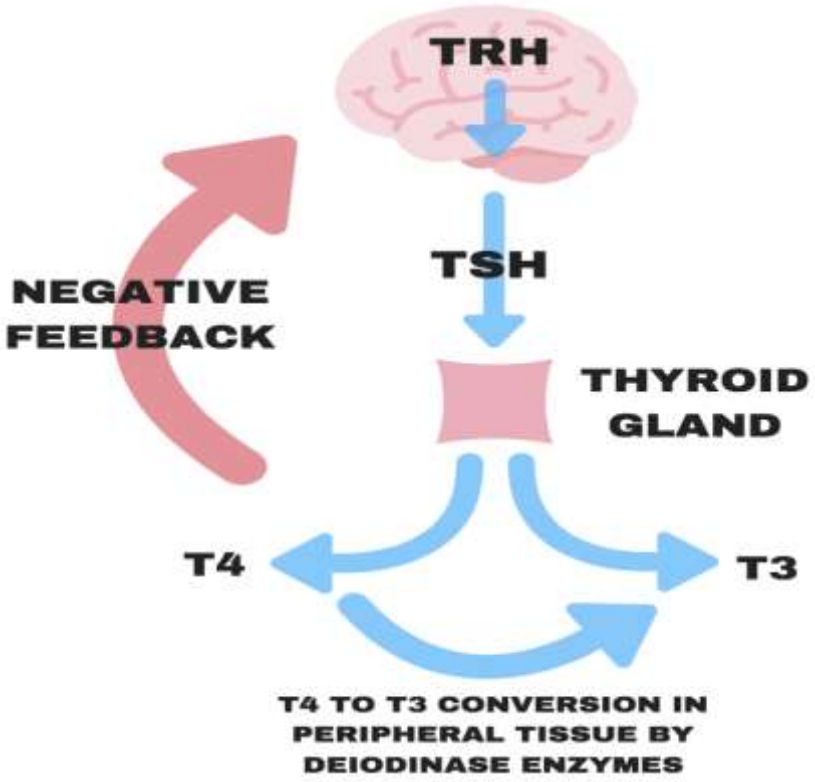

Figure 1.2. Thyroid hormones[19].

Thyroid disease is classified into three important categories:

1. Hyperthyroidism.
2. Hypothyroidism.
3. Normal.

Hyperthyroidism (Thyrotoxicosis) refers to an abnormally high level of thyroid hormones in the blood. There are several subtypes of hyperthyroidism. Sublink suppressed TSH (TSH 0.5 mIU / L) with regular T3, T4 levels, and overt hyperthyroidism (high T3, T4 levels with suppressed TSH)[20]. Any signs Include dry skin, elevated temperature response, hair thinning, weight loss, increased heart rate, hypertension, heavy sweating, neck enlargement, nervousness, shortened menstrual cycles, irregular stomach movements, and hands shaking. Hypothyroidism is a typical concept that applies to an underactive thyroid gland. Hypothyroidism occurs when there is a shortage of tissue-level thyroid hormones or when these hormones are sometimes inactive [21]. Hypothyroidism is mainly affected by inflammation and injury to the thyroid gland. Obesity, a sluggish heart rate, prolonged exposure to high temperatures, throat swelling, dry eyes, numb palms, hair loss, erratic menstrual cycles, and intestinal problems are only a few symptoms. If left untreated, these signs can deteriorate with time. As for the normal, that person who does not suffer from disorders in the thyroid gland, the level of secretion of thyroid hormones is moderate in the blood, so it does not suffer from any disease and be healthy.

1.2 PROBLEM STATEMENT

At present, diseases have become dangerous and rapidly spread, and their exploration and diagnosis require a great deal of time and effort. The correct and accurate diagnosis of the disease early has become one of the problems that the health system suffers from it. The critical role of early and correct diagnosis of the disease, including thyroid disease, is vital because it increases patient treatment opportunities and reduces mortality. Among the vast amount of clinical data, early diagnosis is a challenging task. Today the machine learning has had impressive and good results in many sciences. Hence, it had a prominent and valuable role in diseases, so this study used machine learning algorithms with thyroid disease. To detect and classify thyroid disease into three types hyperthyroidisms, hypothyroidism, and normal.

1.3 PROPOSAL OF THE STUDY

In this thesis, we would work on entering a certain number of data from people with thyroid disease, as this data includes information related to the infected person, and

then we work to enter these data with machine learning algorithms to diagnose the infected persons and also the absence of an infected person if the person was also Infected through this thesis and the algorithms used, we classify the injury as hyperthyroidism or hypothyroidism. Thyroid diseases are becoming increasingly common around the world. Hypothyroidism, hyperthyroidism, or thyroid cancer affect one out of every eight women. Thyroid categorization is important for medical researchers because medical reports show major thyroid dysfunctions among the population, with women being the most affected. The literature mentions several studies in thyroid classification that use various machine learning techniques to develop robust classifiers. The goal we want to reach or the primary goal:

1. Comparison of the performance of the eight machine learning algorithms in predicting thyroid disease.
2. Extract useful patterns from large and complex clinical data.
3. Make the study work to show the following results.
    - hyperthyroidism.
    - hypothyroidism.
    - normal.

1.4 CONTRIBUTION

The contributions we made to this study were directed towards making it run more smoothly and distinguish it from others. One of the contributions we made to this study was obtaining our data from Iraqi private hospitals and laboratories. This data sample consists of 1250 subjects with hyperthyroidism, hypothyroidism, and normal subjects. This study proposes a decision support system for diagnosing people who suffer from hyperthyroidism, hypothyroidism, or do not have the disease. Building this decision-supporting organization was by applying our data to a set of machine learning algorithms. Did work with the eight algorithms, where each of the eight algorithms had a particular result. We also compared the results of the eight algorithms to determine the best and most accurate impact, meaning that any algorithm could work with the data and diagnose thyroid disease well and smoothly.

1.5 THESIS STRUCTURE

The first chapter of this study will focus on what thyroid disease is, what its causes are, and what its branches and types we will focus on in our study.

In the second chapter, we will also focus on clarifying a group of related works that pertain to our studies and the methods used by these workers.

The third chapter will also consist of a detailed explanation of machine learning and its role in the health system and an overview of the eight algorithms used in this study and the way each algorithm works.

Chapter four will focus on our methodology and contain the processes performed on the data and the methods used in our study.

The fifth chapter will consist of the results and discussion of this study.

The sixth chapter will consist of the Conclusion and Future work of this study.

PART 3

CLASSIFICATION TECHNIQUES

Looking at the previous studies that we presented, we knew that machine learning has an important role and has good results and accuracy with the health system. In this chapter, we will explain what ML techniques are and how algorithms classify them.

3.1 WHAT IS MACHINE LEARNING

ML is one of the essential branches of artificial intelligence concerned with designing algorithms that allow computers to learn independently. Instead, it is a process of identifying statistical regularities or other trends in results. Figure 3.1 shows what machine learning algorithms are.

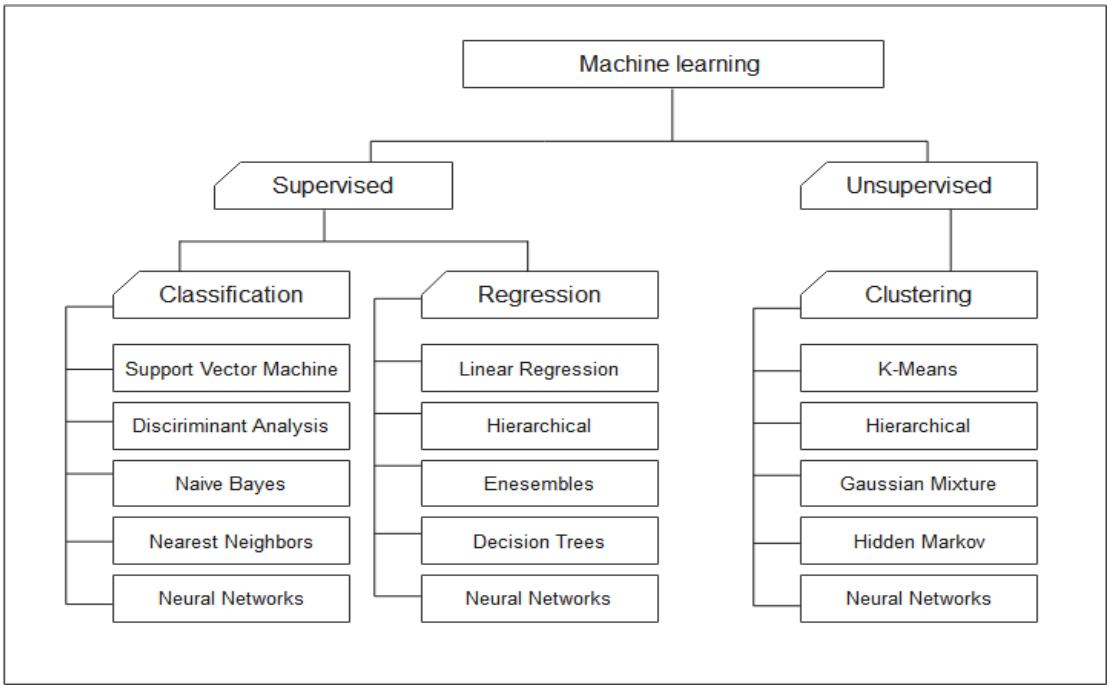

Figure 3.1. Machine learning algorithms[30].

As a result, several machine learning algorithms will bear no resemblance to how a person would handle learning. On the other hand, learning algorithms may provide the appropriate solution to the relative complexity of learning in different contexts [31]. It lies at the intersection of statistics and the heart of artificial intelligence and data science, as it is one of the fastest developing fields of science today. The invention of modern and theoretical learning algorithms and the continued prosperity of data availability on the Internet and low-cost computing have led to recent machine learning developments. This development has led to more and more data-intensive machine learning methods in research, technology, and industry, leading to increased evidence-based decision-making in many different fields, including medicine. Health services, manufacturing, education, financial modeling, policing, and marketing are just a few examples. Machines can find private lessons from data by learning to use specific algorithms; it is crucial to remember that we're not asking the machines where to look in machine learning. Machine learning's iterative design helps the device change its methods and outputs when encountering new scenarios and data [32]. Although the use of machine learning and artificial intelligence in medicine goes back to the early days of the field [33].

3.2 SUPERVISED AND UNSUPERVISED

The ML algorithms can approach in one of two ways. The first is what It is known as directed learning. The field of supervised learning entails the use of specific examples to train algorithms. The computer is given a set of inputs and a set of correct outputs, and it learns by comparing empirical outcomes to the right results to classify errors [34]. When past events can be used to forecast future events, this method of learning is used. Since the aim is always to get the machine to understand a classification scheme that we have developed, supervised learning is popular in classification problems. Once again, digit identification is a typical example of classification learning. Classification learning is helpful for any issue where deducing a classification is usable, and the category is simple to evaluate. In other instances, assigning preset categories to each example of a dilemma might not be appropriate if the agent can figure out the classifications independently [35].

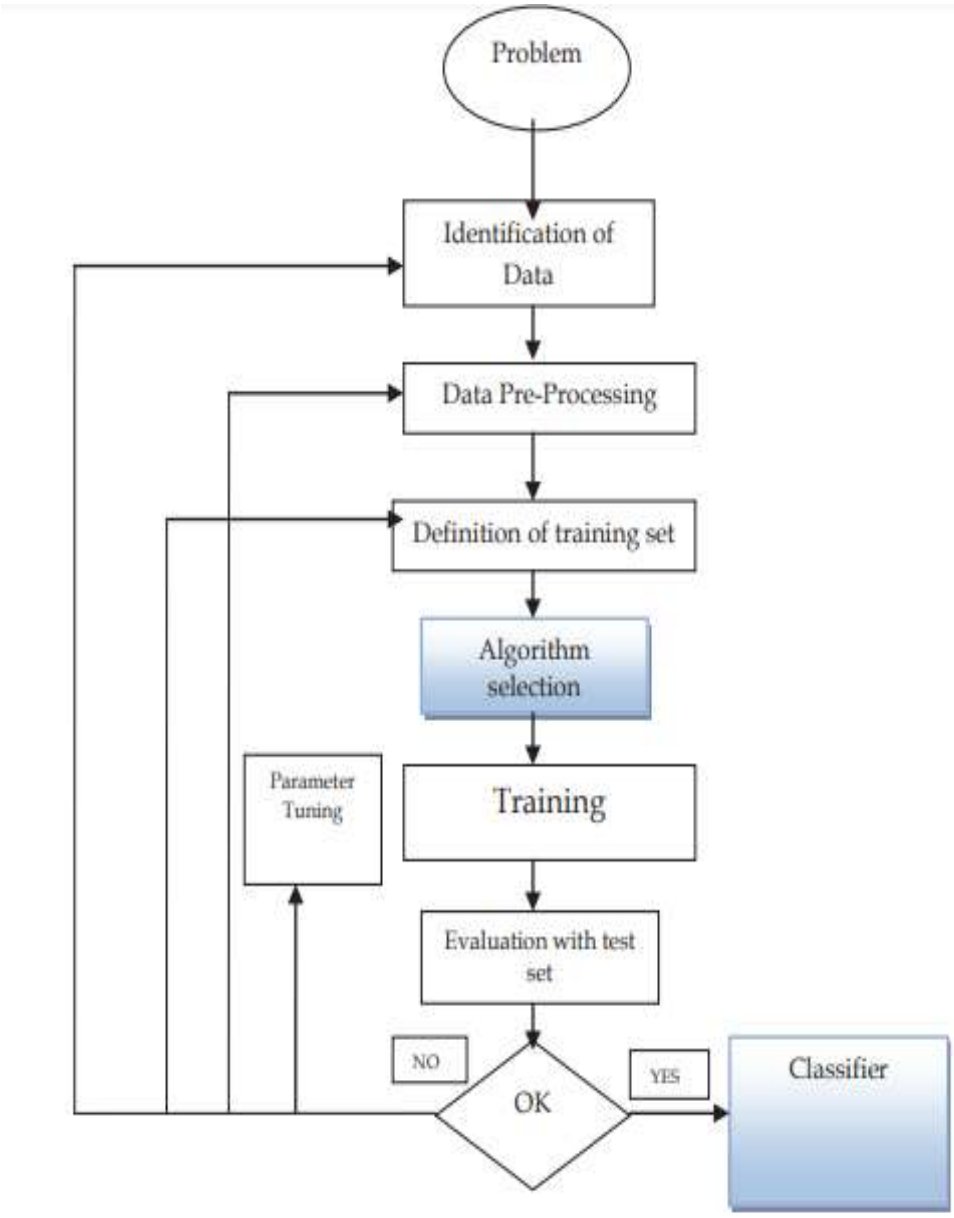

Figure 3.2. Machine Learning Supervise Process.

Unsupervised instruction is the other choice. The computer must explore the data and construct some pattern or structure under this approach; it must create models from scratch and not give any correct information. The results Outliers are often identified and distinguished using this approach [36]. Unsupervised learning seems much harder: the goal is to have the computer learn how to do something that we don't tell it how to do. Unsupervised learning has yielded a slew of milestones, including world-champion backgammon projects and even self-driving vehicles. When assigning values to behavior is easy, it can be an effective technique.

3.3 SUPPORT VECTOR MACHINES

The support vector machine (SVM) is a machine learning and data mining algorithm to determine the strongest predictors of this variable for energy consumption. The research used popular classification methods to answer our question: best subset selection, boosting trees, and generalized additive models. Our first approach was to use forward, backward, and best subset selection to obtain a subset of predictors that most strongly predicted consumption with a linear relationship. The SVM provided an approach that was to use a tree-based method to stratify the predictor space into sample regions using recursive binary splitting. The research decided to use the boosting tree method, which is known to be one of the most potent tree-based models. SVM also has an excellent ability to deal with high dimensionality data.

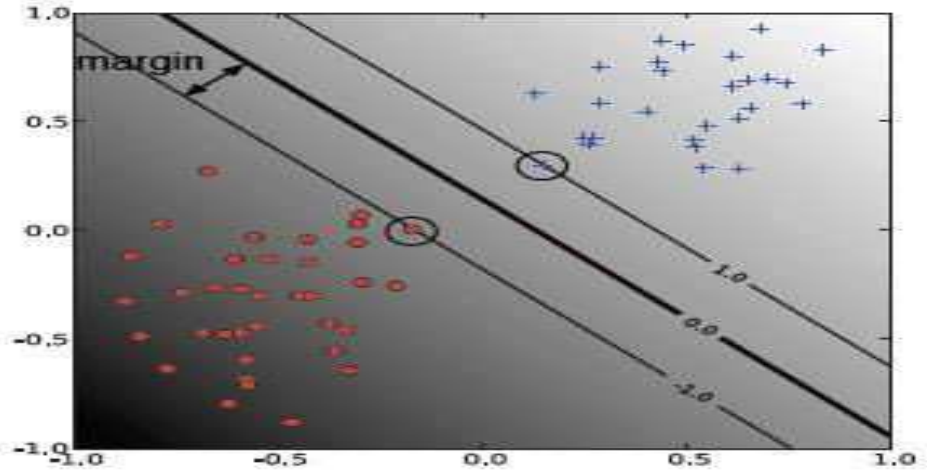

Figure 3.3 Example a two-class linear SVM classifier [37].

The two restrictions for generating the enormous hyperplane separation margin are.

$$v.\ x_i + b \geq +1\ ,\ y_i = +1 \tag{3.1}$$

$$v.\ x_i + b \leq +1\ ,\ y_i = +1 \tag{3.2}$$

where v is the hyperplane's natural, |b|/||v|| is the distance between the origin and the hyperplane, and ||v|| is w's standard.

You might write these two restrictions as a single restriction using the y_i sign (as positive in positive samples and negative in negative samples) [38].

$$y_i(w.x_i + b) \geq 1 \ \forall_i \tag{3.3}$$

The two constraints form two canonical hyperplanes. The samples nearest to the dividing hyperplane are circled. They're named support vectors, and they all lie on the H1 and H2 canonical hyperplanes [39]. These hyperplanes are parallel to one another, and the distance between them is measured as follows.

$$p = \frac{|1-b|}{||w||} - \frac{|-1-b|}{||w||} = \frac{2}{||w||} \tag{3.4}$$

Since margin is inversely proportional to "||w||," we can optimize margin by decreasing "||w||." To make it easier to calculate, we will try to reduce $\frac{2}{2}\ ||w||^2$ whilst keeping the constraints in mind.

3.4 RANDOM FOREST

RF is that classification supervised in machine learning techniques. A regression algorithm that develops with a set of decision trees (DTs) is built based on combining predictions from several trees to provide better results than a single tree [40]. The training data set is generated by pre-defining and labeling the input data groups and then provided to the algorithm in supervised classification [41] [42]. As a result, the classification algorithm will learn target class properties from training data and apply them to data that hasn't been classified yet [43]. RF varies from DT due to the dynamically running processes of locating the root nodes and dividing the internal nodes [44], Including the fact that it contains ensemble DTs. Furthermore, RF can solve the drawbacks of a single DT while retaining its benefits. The values of randomly selected input variables are used to build each tree in the RF algorithm to approximate the value of the output variable by supervised learning [45]. The random forest calculates the average answer of each energy usage predictor. A random forest then adds the absolute distance each response was from the mean of each predictor for a

cumulative value of the space each answer was from the data's standards for each sample. Individuals that were regularly far from the mean solution in each survey would have a total distance score. A feature that determined it used the mode of each answer to detect rates that repeatedly classify the samples. The study labeled an answer as theoretically high in energy usage if the response style accounted for more than 90% of the total number of queries. There are a lot of comments that have been labeled. A visual analysis of these responses revealed that the individuals had all sampled the same answer [46].

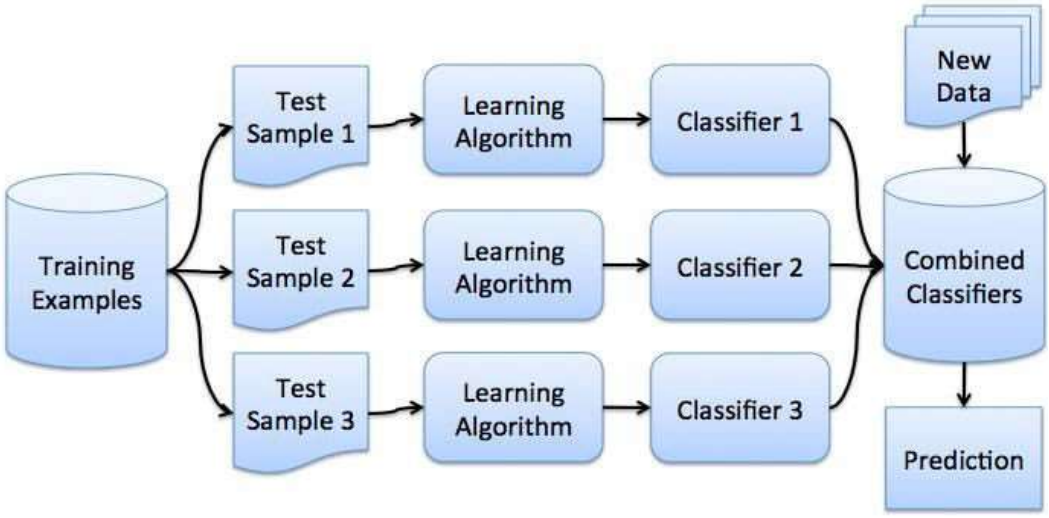

Figure 3.4 Phases of ensemble random forest classification [47].

3.5 DECISION TREE

A decision tree is a sort of tree system that builds classification and regression models. It cuts down a dataset into smaller and smaller sections while creating a decision tree in the background. The outcome is a tree having both decision and leaf nodes. In a decision node (for example, Outlook), there are two or three divisions (e.g., Sunny, Overcast, and Rainy). The leaf node represents a categorization or judgment (for example, Play). The root node in a tree corresponds to the most decisive indication. Both category and numerical data may be handled using decision trees. The decision tree system is based on a decision-boosting computer tested for forecasting energy use and finding the most significant predictors of service using a tree-based methodology.

I utilized the decision tree methodology that was was provided to me to do this. The model must fit hundreds of trees to develop, each grown use information from the preceding three. Among the tuning parameters are the number of trees, the shrinkage parameter, and the splits in each tree.

This is one of the numerous types of CARTS that we employed in our research using the Decision Trees algorithm. CART (Classification and Regression Trees) is similar to C4.5, except it cannot generate rule sets and does not allow numerical target variables (regression). CART creates binary trees depending on the function and threshold that deliver the most benefit to each node in terms of knowledge. Used CART. This is a formula.

$$H(Q_m) = \sum_k p_{mk} \, (1 - p_{mk}) \tag{3.5}$$

Class k is the percentage of its views in node m. If m is a leaf node, then the prediction for this region is set to p_mk. Common impurity measures are the following.

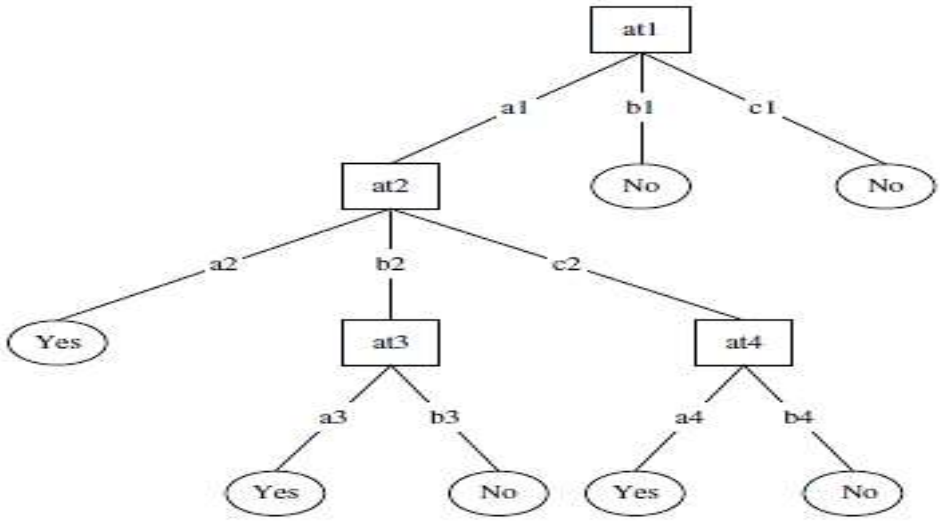

Figure 3.5 Decision tree algorithm structure [48].

PART 4

METHODOLOGY

In this chapter, the process of discussing the proposed method of work in this study is illustrated in Figure (4.1). In the first stage, the data set is collected, and after that, the data is prepared by pre-processing as shown in Fig 4.2. Second, two datasets are created, the first containing all the attributes and the second containing all but three of the attributes, which are query_thyroxine & query_hypothyorid & query_hyperthyroid.

In the final stage, classification algorithms are used to predict thyroid disease and compare the performance of different classifiers by calculating the accuracy of each classifier.

In this study, we use eight machine learning techniques represented by (Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (K-NN), and multi-layer perceptron (MLP).

Platform Used

The development of machine learning algorithms used in this study with the Python programming language (version 3.8.3), data pre-processing was done with panda's libraries (version 1.0.5), and the developed machine learning algorithms with scikit libraries (version 0.23 .1). The primary platform used for this work is Spyder (Anaconda) on a computer with a 64-bit operating system, x64-based processor, 4.00 GB of RAM, on an Intel (R) Core (TM) i7-7230M CPU.

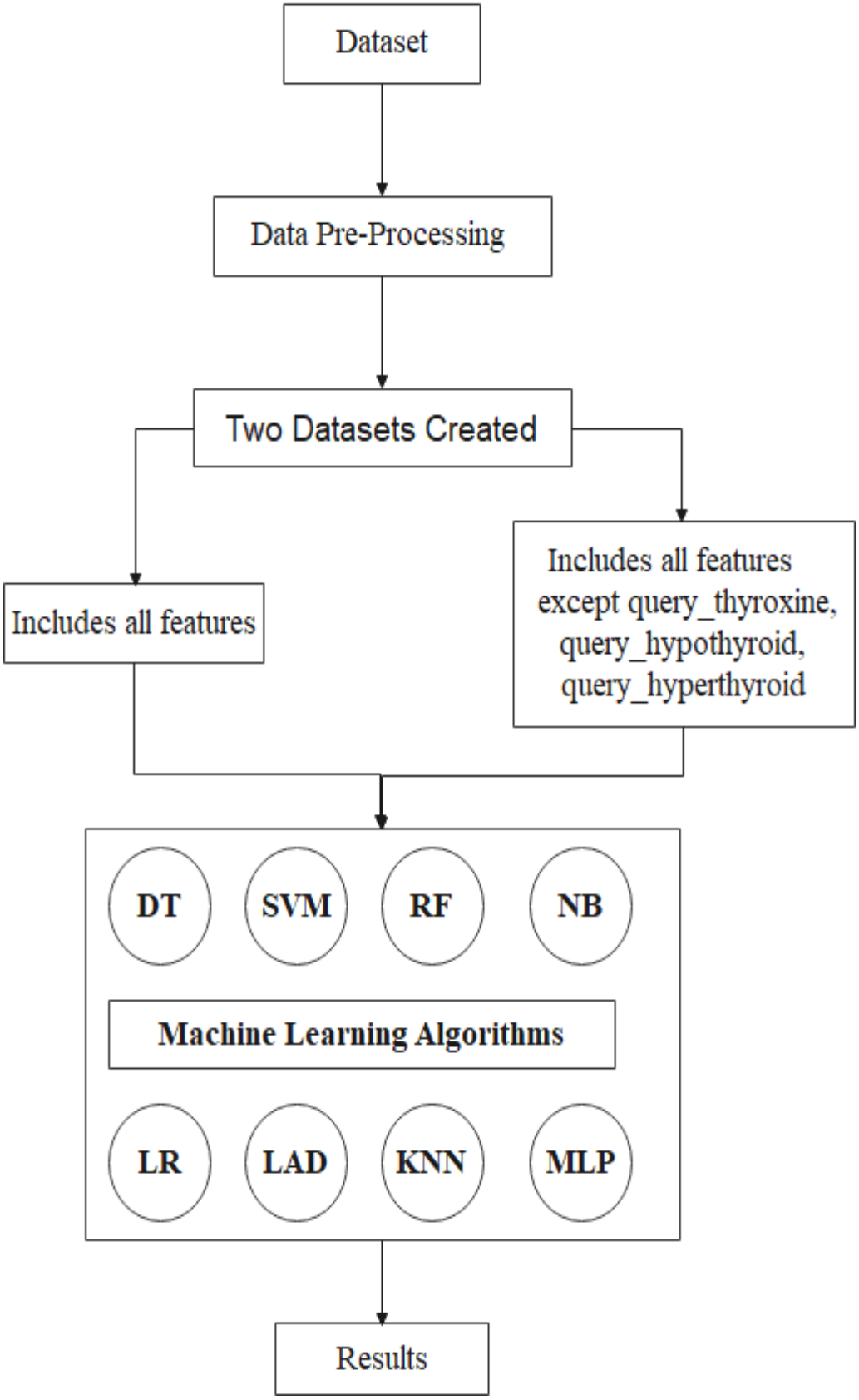

Figure 4.1. Flowchart of the study.

4.1 DATA COLLECTION

We collected a good amount of data on thyroid diseases, and we are working in our study to classify diseases using this data. The data used in our study is a data set taken from hospitals and private laboratories specialized in analyzing and diagnosing diseases in Iraq, Baghdad governorate, and the type of data taken related to thyroid diseases. Data were taken on 1,250 male and female individuals whose ages range from one year to Ninety years as these samples contain thyroid disease subjects with hyperthyroidism and hypothyroidism and normal subjects without thyroid disease. We collected data over one to four months, and the main objective of data collection was to classify thyroid diseases using machine learning algorithms.

Table 4.1. shows the features contained in the dataset.

No	Attribute Name	Value Type	Clarification
1	id	number	1,2,3…etc.
2	age	number	1,10,20, 50…etc.
3	gender	1,0	1=m,0=f
4	query_thyroxine	1,0	1=yes,0=no
5	on_antithyroid_medication	1,0	1=yes,0=no
6	sick	1,0	1=yes,0=no
7	pregnant	1,0	1=yes,0=no
8	thyroid_surgery	1,0	1=yes,0=no
9	query_hypothyorid	1,0	1=yes,0=no
10	query_hyperthyroid	1,0	1=yes,0=no
11	TSH measured	1,0	1=yes,0=no
12	TSH	Analysis ratio	Numeric value
13	T3 measured	1,0	1=yes,0=no
14	T3	Analysis ratio	Numeric value
15	T4 measured	1,0	1=yes,0=no
16	T4	Analysis ratio	Numeric value
17	category	0,1,2	0=normal,1=hypothyroid, 2=hyperthyroid

- Convert all categorical features in the data set to a standardized numerical representation because of the presence of categorical and numerical features.

In the data processing stage, we identified missing data, and among these properties were missing T4 by number 151 and T3 by number 112. where we were able to Process this lost data by replacing it with the value of the mediator, and after working in this way, we were able to obtain the data in a good and better way and free from lost data, as the data became arranged and excellent and free from any defect or problem so that we can work on it smoothly and well.

Data Normalization may refer to more extensive alterations aiming at harmonizing the whole probability distributions of transformed data in more complicated settings. Before averaging, normalization of ratings requires translating data from many scales to a single, apparently typical scale. In this study, data normalization was used with the MLP algorithm only using the Min-Max normalization method.

After applying pre-processing data, the final data set after pre-processing consists of 1250 rows and 17 columns and does not contain any null or redundant values. All features are represented in numerical form, as shown in Table 4.1. The operations mentioned above are illustrated in the third figure (4.2).

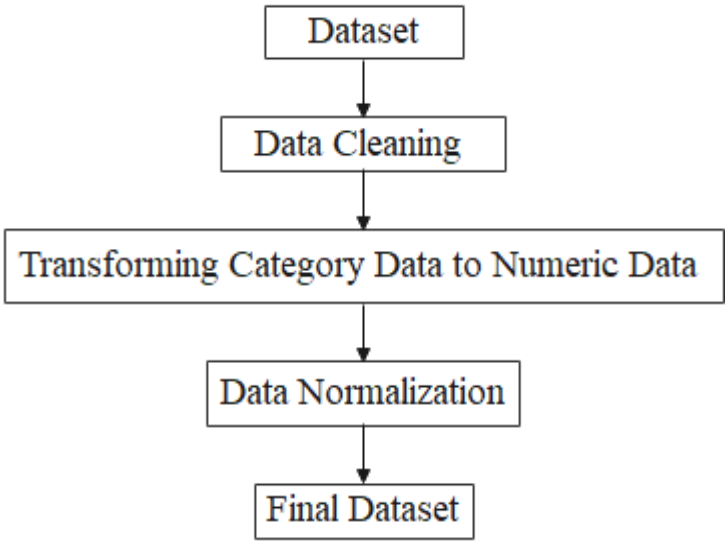

Figure 4.2. Data Pre-processing.

4.4 PERFORMANCE MEASUREMENT

In our study, we relied on evaluating the eight machine learning techniques using the following metrics: accuracy, specificity, sensitivity, precision, F-score, and recall, as shown in table (4.2). Subjects without thyroid disease were considered false positive, and people with thyroid disease were both hypothyroid and hyperthyroid as true negative [56].

Table 4.2. The formulae statistics to evaluate the predictive capability.

Measure	Formula	Description
accuracy	$\frac{tp + tn}{tp + tn + fp + fn}$	The percentage of successfully identified observations compared to the total number of classifications made by the classifier.
precision	$\frac{tp}{fp + tp}$	The percentage of successfully categorized bad loans compared to all bad loans categorized by the classifier.
specificity	$\frac{tn}{fp + tn}$	The proportion of successfully identified excellent loans in the test dataset compared to the total number of excellent loans.
sensitivity	$\frac{tp}{tp + fn}$	The percentage of properly categorized bad loans in the test dataset compared to the total number of bad loans.

PART 5

RESULTS

Thyroid illness is estimated to affect more than 200 million individuals worldwide. It is also widespread among Iraqis, particularly women. This condition has the potential to have a significant impact on adult bodily processes as well as infant development. Thyroid diseases are generally treatable, but they may be dangerous if they progress to the advanced stage, and they may even result in death. The successful classification was performed using various techniques using the thyroid disease data set obtained from private laboratories and hospitals. We used eight algorithms in this study (Decision Tree (DT), Supporting Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (K-NN), Multilayer (MLP)). We divided the existing data into two parts, 30% for training and 70% for testing, as this training is the first training on this data.

In the first step, we took all the properties in our data and applied them to the eight algorithms. We take all the features in our data in the second step except for query_thyroxine, query_hypothyorid, and query_hyperthyroid and they ared used to predict using eight algorithms. Our data contains 16 attributes as one input and one output. That is, the sum of the attributes in this data is 17 attributes.

5.1 RESULTS IN THE FIRST STEP WITH ALL FEATURES IN DATASET

In this section, the predicted thyroid disease by taking the data set with all the features for thyroid disease classification.

An implemented classification algorithm on this data set and the accuracy achieved by each algorithm was calculated and compared the achieved results.

As shown Table 5.1 shows us the accuracy of each algorithm, as it received an algorithm Decision Tree 98.4% accuracy SVM 92.27% accuracy Random Forest 98.93% accuracy Naive Bayes 81.33% accuracy Logistic Regression 91.47% accuracy Linear Discriminant Analysis 83.2% accuracy KNeighbors Classifier 90.93% accuracy and MLP (NN) 97.6% accuracy and through these results.

The Random Forest algorithm is very accurate, and then the algorithm follows the decision tree and other algorithms. The Random Forest algorithm enables the rapid discovery of essential data from large databases. The main advantage of Random Forest is that it relies on a range of different decisions to arrive at the answer. Therefore, she was able to obtain the highest accuracy.

Table 5.1. Evaluation measurements for classification models with all features**.**

No	**Algorithms**	**Accuracy (%)**	**Sensitivity (%)**	**Specifity (%)**
1	Decision Tree	98.40	99.29	97.84
2	SVM	92.27	90.93	98.48
3	Random Forest	98.93	98.60	100
4	Naive Bayes	81.33	98.63	57.69
5	Logistic Regression	91.47	98.80	100
6	Linear Discriminant Analysis	83.20	82.49	89.47
7	KNeighbors Classifier	90.93	92.75	84.70
8	MLP	96.00	94.93	98.75

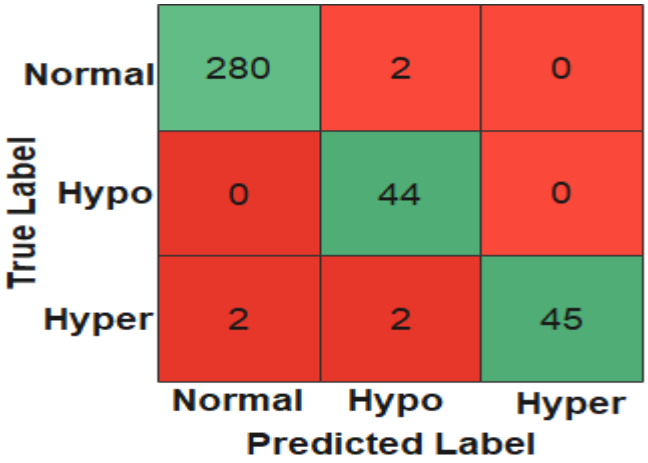

Figure 5.1. Confusion Matrix for Decision Tree Classifier.

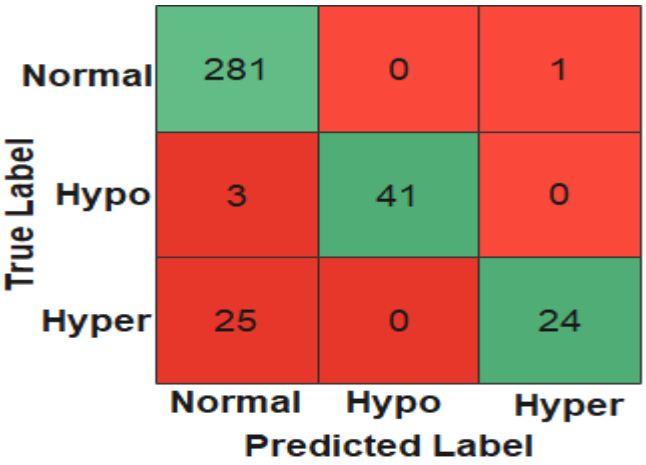

Figure 5.2. Confusion Matrix for SVM Classifier.

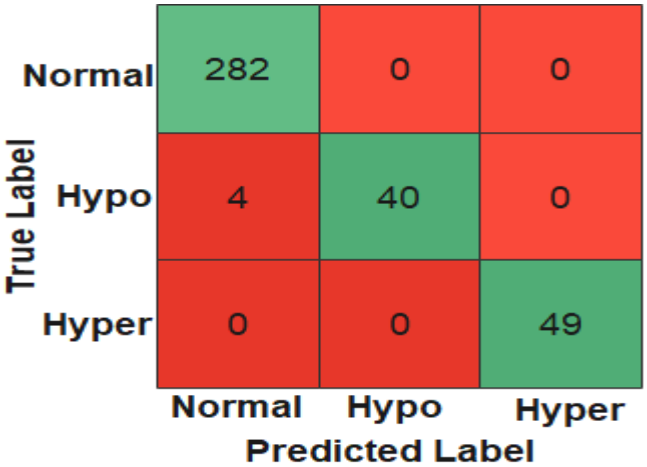

Figure 5.3. Confusion Matrix for Random Forest Classifier.

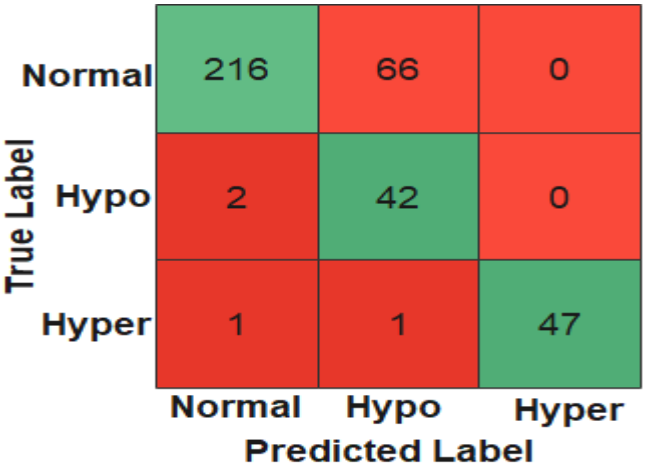

Figure 5.4. Confusion Matrix for Naive Bayes Classifier.

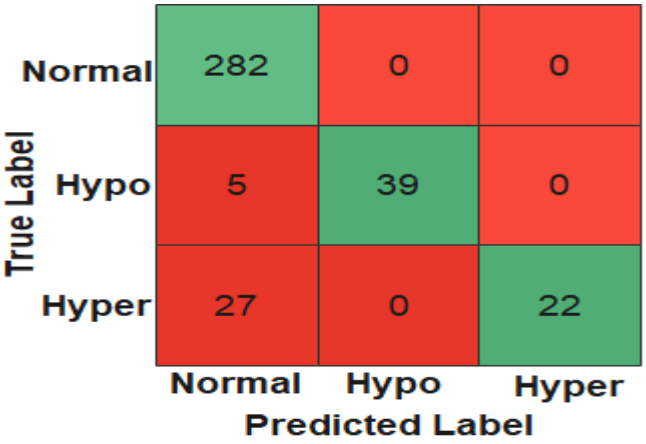

Figure 5.5. Confusion Matrix for Logistic Regression Classifier.

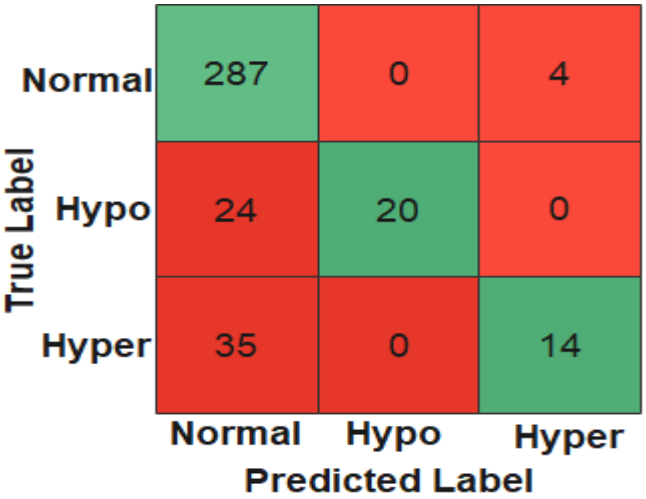

Figure 5.6. Confusion Matrix for Linear Discriminant Analysis Classifier.

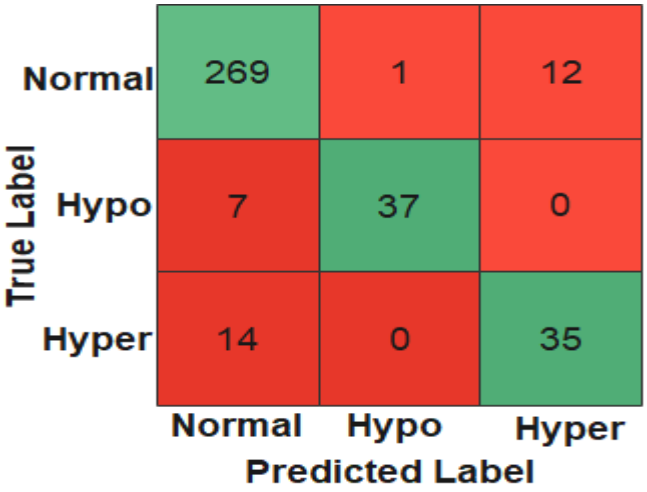

Figure 5.7. Confusion Matrix for KNeighbors Classifier.

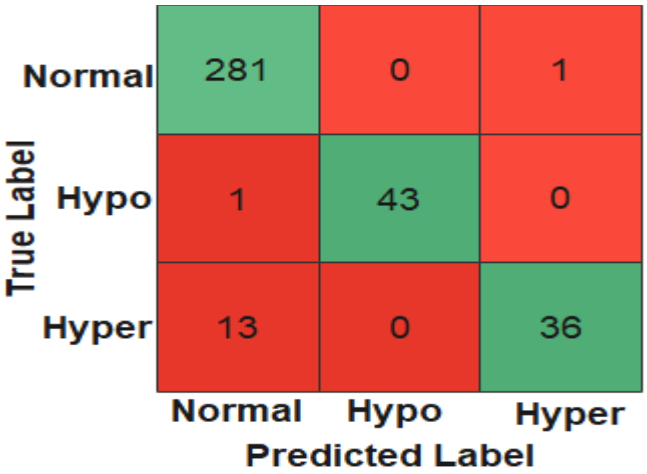

Figure 5.8. Confusion Matrix for MLP Classifier.

5.2 RESULTS IN THE SECOND STEP WITH OUT THREE FEATURES

In the second step, we removed three traits based on a previous study Ioniţă, Irina, and Liviu Ioniţă [58]. The deleted attributes were both query_thyroxine&query_hypothyorid& query_hyperthyroid. After deleting these attributes, we applied our data also to the algorithm group, and by using the Python script, we were able to obtain these results listed below in Table 5.2.

Algorithm Decision Tree 90.13% accuracy SVM 92.53% accuracy Random Forest 91.2% accuracy Naive Bayes 90.67% accuracy Logistic Regression 91.73% accuracy Linear Discriminant Analysis 83.20% accuracy KNeighbors Classifier 91.47%

accuracy and MLP(NN) 95.73% accuracy and through these results. As it seems to us that the Naive Bayes algorithm has a high accuracy of 90.67 after the three traits have been omitted, the SVM algorithm, the logistic regression algorithm, and the neighbor's Classifier algorithm have increased slightly and reduced the accuracy of the other algorithms. We show here that the accuracy of algorithms used in our data changes with the change of the characteristics used in the data. Experience had demonstrated this evident change; algorithms were obtained when three of the traits were deleted, as the accuracy of some algorithms decreased, and some increased. The MLP algorithm received the highest accuracy after deleting three attributes, as its accuracy reached 95.73%.

Table 5.2. Evaluation measures of the classification models without three features.

No	algorithms	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	Decision Tree	90.13	88.64	98.27
2	SVM	92.53	91.23	98.50
3	Random Forest	91.20	89.52	100
4	Naive Bayes	90.67	99.60	73.60
5	Logistic Regression	91.73	90.00	100
6	Linear Discriminant Analysis	83.20	82.49	89.47
7	KNeighbors Classifier	91.47	93.10	85.88
8	MLP	95.73	94.93	98.73

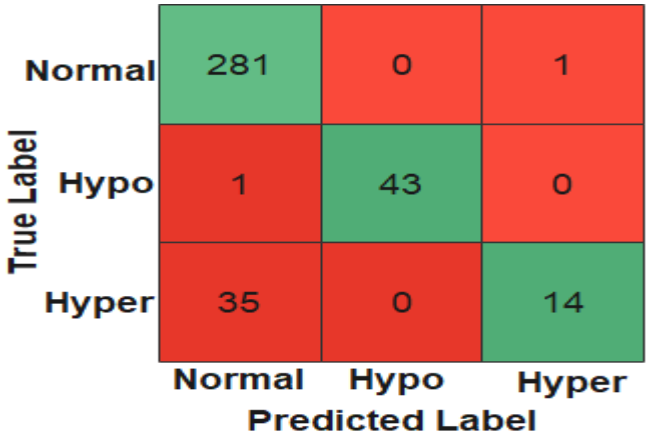

Figure 5.9. Confusion Matrix for Decision Tree Classifier without three attributes.

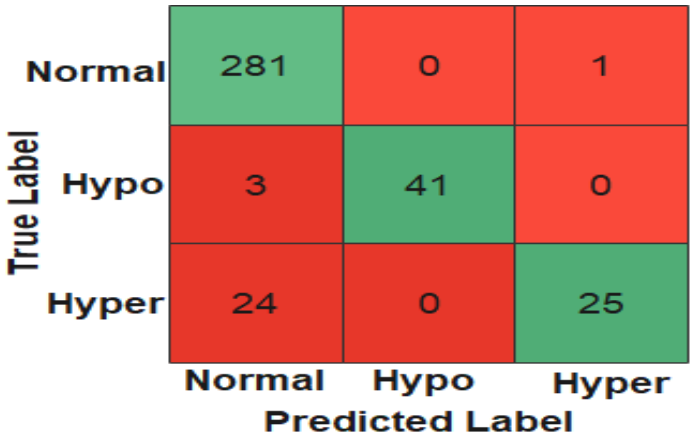

Figure 5.10. Confusion Matrix for SVM Classifier without three attributes.

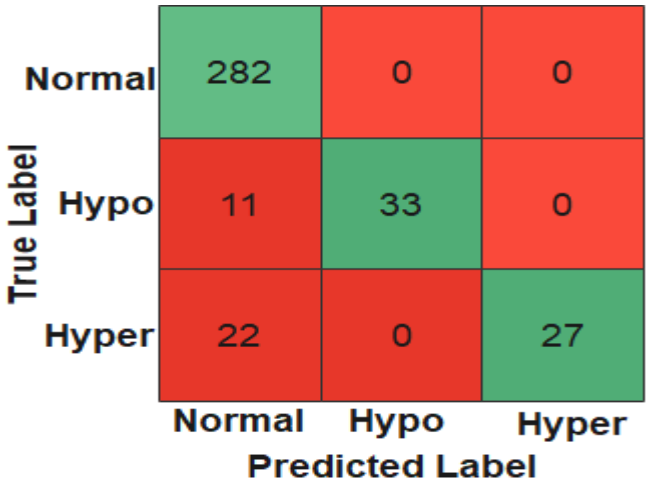

Figure 5.11. Confusion Matrix for Random Forest Classifier without three attributes.

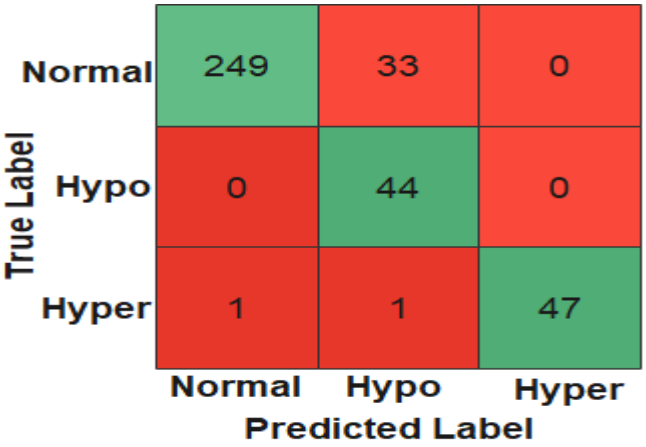

Figure 5.12. Confusion Matrix for Naive Bayes Classifier without three attributes.

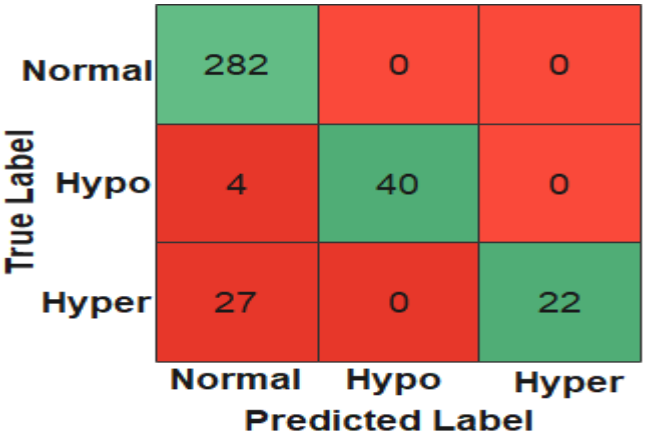

Figure 5.13. Confusion Matrix for LG Classifier without three attributes.

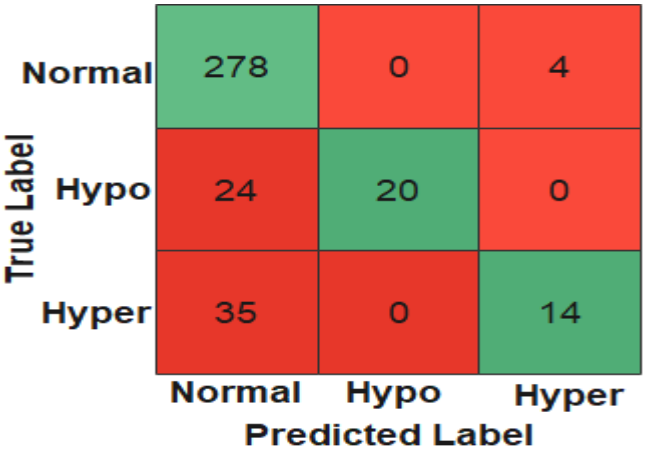

Figure 5.14. Confusion Matrix for LDA Classifier without three attributes.

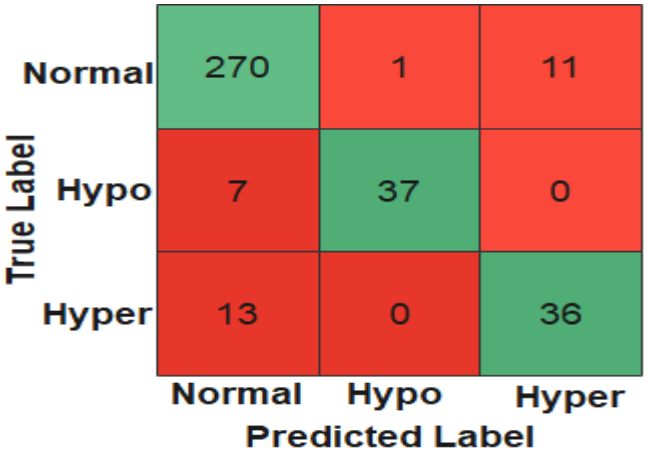

Figure 5.15. Confusion Matrix for KNN Classifier without three attributes.

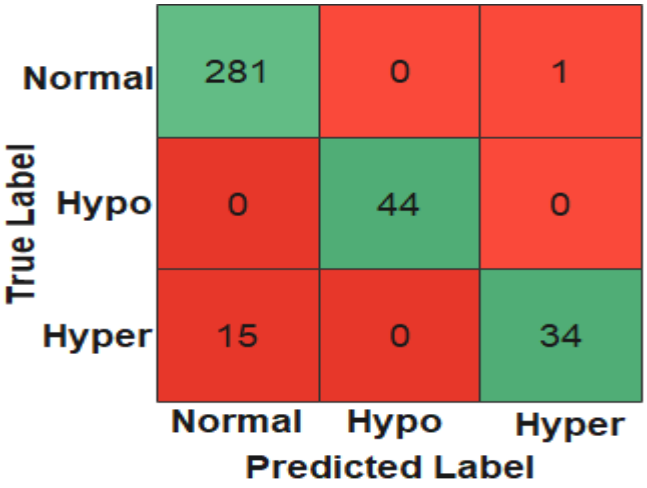

Figure 5.16. Confusion Matrix for MLP Classifier without three attributes.

5.3. DISCUSSION

Today the health field needs solid support to help overcome many obstacles. Among the vast amount of clinical data, it becomes tough to work with. We note the excellent and effective progress of artificial intelligence techniques in the medical field, especially machine learning algorithms, and their valuable and prominent role in classifying, predicting, or diagnosing diseases. Our study deals with machine learning algorithms to classify thyroid disease. These algorithms are (Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic

Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (K-NN), and multi-layer perceptron (MLP), as well as comparison of algorithm results.

After arranging and cleaning the dataset, we created two datasets with all the attributes, and the second, we erased three of them, which are (query_thyroxine&query_hypothyorid& query_hyperthyroid). With the first step, the Random Forest algorithm obtained 98.93% accuracy, followed by other algorithms. In the second step, the MLP algorithm received 95.73% accuracy, followed by the different algorithms. With this step also, the accuracy of the Naive Bayes algorithm increased and obtained 90.67% after its accuracy in the first step was 81.33%. The KNN algorithm increased to 91.47% from 90.93%, the accuracy of some other algorithms increased slightly, the accuracy of some of the algorithms decreased.

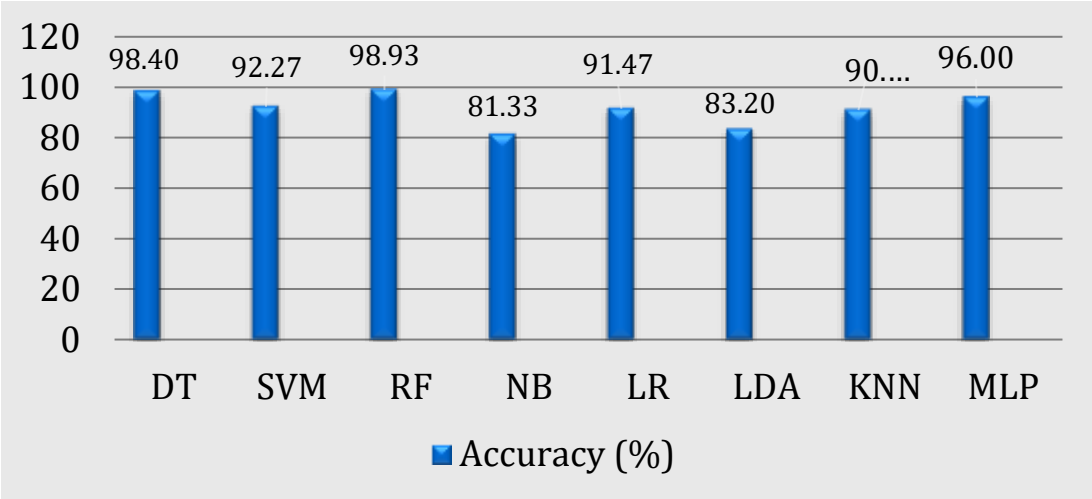

Figure 5.3.1 Evaluation measurements for classification models with all features.

It indicates that the three characteristics had a negative performance on the results of algorithms whose accuracy increased after deletion, while they had a good effect with algorithms whose accuracy decreased after deletion; this indicates that these properties had an essential and valuable role with these algorithms for diagnosing disease. In each of the two steps, we reached excellent and effective results in analyzing and classifying thyroid disease to contribute to helping the health system and its workers.

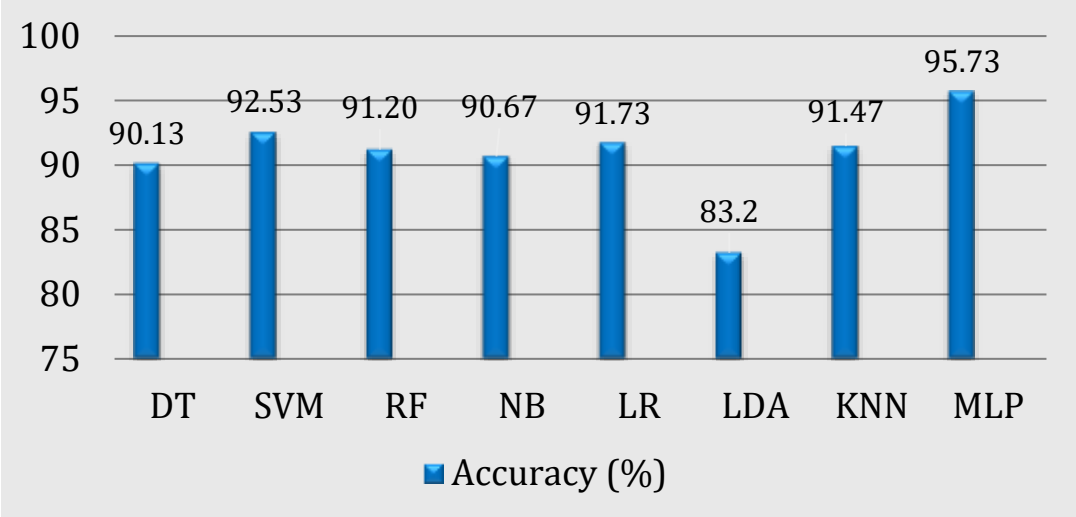

Figure 5.3.2 Evaluation measures of the classification models without three features.

PART 6

CONCLUSION AND FUTURE WORK

The study showed that the could diagnose thyroid functional disease with we use eight machine learning techniques. We were represented by (Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Naive Bayes (NB), Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (K-NN), and multi-layer perceptron (MLP)). This study assists to doctors and medical staff in the healthcare field. We were also able to compare the eight algorithms and which one could reach the best accuracy.

We predicted and classified thyroid disease by applying machine learning techniques to a data set consisting of 1250 actual samples. We divided the dataset as follows: 30% of the data were used for training, and 70% were used for testing. After applying these techniques to dataset one that consists of all the characteristics, the random forest algorithm obtained an accuracy rate of 98.93%. In the second step, and based on a previous study, we deleted a set of features which are 1- query_thyroxine 2- query_hypothyorid 3-query_hyperthyroid. We applied machine learning techniques to this data, and the MLP algorithm got the highest accuracy of 95.73%. The results obtained in this study help us in the rapid prediction of thyroid disease. And the classification of the disease (Hyperthyroidism or Hypothyroidism). Future work should focus on improving the performance of classification algorithms and using different approaches from feature selection methods to obtain better results. The work can use deep learning algorithms for improving accuracy and other metrics.