EDRAAK

Vol. **(2024)**, 2024, **pp**. 84–100

ISSN: 3078-8412

Research Article

Thyroid Disease Prediction Using Machine Learning Model: Decision Tree

Hadeel Alkhazzar [1,*], Bassam Moslem [1]

[1] *Department of Computer and Communication, Collage of Engineering, American University of Science and Technology, Beirut, Lebanon.*

ARTICLE INFO

Article History

Received 10 Apr 2024

Revised: 5 Jun 2024

Accepted 4 Jul 2024

Published 23 Jul 2024

Keywords

Thyroid disease,

machine learning,

decision tree,

classification,

accuracy.

ABSTRACT

Thyroid illness must be correctly classified in order to have an accurate diagnosis and effective therapy. Diagnostic accuracy and efficiency have both been found to improve with the use of machine learning methods. The decision tree algorithm has been shown to be useful for healthcare classification issues. Using a dataset of 793 patients, this investigation investigates the utility of the decision tree method for classifying thyroid illnesses. The results show that the decision tree algorithm has high rates of accuracy, recall, and precision for classifying thyroid diseases. The decision tree approach is simple, can deal with missing values and noisy data, and is straightforward to comprehend when used to the categorization of thyroid diseases. This research recommends feature selection, hyperparameter adjustment, and efficient data preparation as means to enhance the efficiency of the decision tree method. The precision and dependability of machine learning models for diagnosing thyroid diseases may be further enhanced by cooperation between data scientists and healthcare practitioners. This study finds that the decision tree algorithm is a valuable resource for classifying thyroid diseases and suggests further investigation into areas including algorithm integration, bigger studies, and real-time diagnosis.

1. INTRODUCTION

Thyroid disorders have become more common in recent years. The thyroid gland is the organ that is accountable for a variety of crucial metabolic processes. Both hyperthyroidism and hypothyroidism are among the most common conditions associated with the thyroid gland, which is responsible for a number of other diseases. A sizeable number of people are identified as having thyroid conditions such as hypothyroidism and hyperthyroidism throughout the course of each year.

Both hypothyroidism and hyperthyroidism may be caused by insufficient levels of the thyroid hormones levothyroxine (T4) and triiodothyronine (T3), which are generated by the thyroid gland. There are a few different approaches proposed in the literature for identifying thyroid disease. Proactively diagnosing thyroid illness is crucial for treating patients at the right time, which in turn saves lives and lowers healthcare costs. Due to developments in data processing and computation, machine learning and deep learning algorithms can now be used to predict the early identification of thyroid disease and differentiate between different types of thyroid sickness, such as hypothyroidism and hyperthyroidism [1].

According to developments in data mining, big data, image and video processing, and parallel computing, the healthcare business is better able to employ technology in a wide variety of scenarios for the benefit of patients. Early illness detection, prediction of virus outbreaks, medication research and testing, health care data management, and patient-specific prescription recommendations are just some examples of the many applications of data mining in healthcare. For the sake of the patient, doctors work to make an accurate diagnosis as soon as possible so that treatment can begin at an earlier, more manageable stage. Thyroid disease ranks among the world's most prevalent health problems.

Twenty million Americans, according to the American Thyroid Association, the leading professional body in the field of thyroid disorders. Thyroid problems affect one in twelve people in the United States. These numbers highlight the seriousness of thyroid-related illnesses. Modern medical techniques should be used to enhance the detection and treatment of thyroid diseases [2].

1.1 Definition and overview of thyroid disease

The butterfly-shaped thyroid gland resides in the neck. The process by which food is transformed into energy is governed by hormones produced by the body. Thyroid-stimulating hormone, also referred to as TSH is a hormone released by the pituitary gland in the brain that regulates thyroid function. The thyroid gland secretes two major hormones, thyroxine (T4) and triiodothyronine (T3). T4 is primarily produced by the thyroid gland, and subsequently the liver and other organs

*Corresponding author email: h.Alkhazzar878@gmail.com

DOI: https://doi.org/10.70470/EDRAAK/2024/011

convert it into the more active hormone T3. Thyroid hormones play an important role in maintaining a healthy range of physiological functions, including core temperature, metabolic rate, and energy levels. Thyroid function may be affected by several factors, including genetics, lifestyle, and environment. The thyroid gland is a complex and vital part of the endocrine system. It produces hormones that regulate metabolism, which has far-reaching consequences for health. Thyroid illness symptoms may take numerous forms and manifest differently in different people. The thyroid secretes a hormone that regulates metabolism and has far-reaching effects on the body's every cell, tissue, and organ.
Thyroid hormones, such as triiodothyronine and levothyroxine, are produced by the thyroid gland and play an important role in protein synthesis, thermoregulation, and cellular energy production [20]. Some symptoms of low thyroid hormone production include extreme fatigue, depression, memory loss, and weight gain. This illness is known as hypothyroidism. Hyperthyroidism is characterized by the overproduction of thyroid hormone, leading to an abnormally rapid metabolic rate. This causes a variety of symptoms, such as agitation, anxiousness, muscular weakness, sudden weight loss, restless nights, and vision problems. In women who have untreated hyperthyroidism, the risk of osteoporosis (also known as brittle bones), arterial fibrillation (also known as an irregular heart rhythm), cardiomyopathy (often known as a weak heart), angina, and heart failure, in addition to other complications associated to pregnancy, is increased [3].

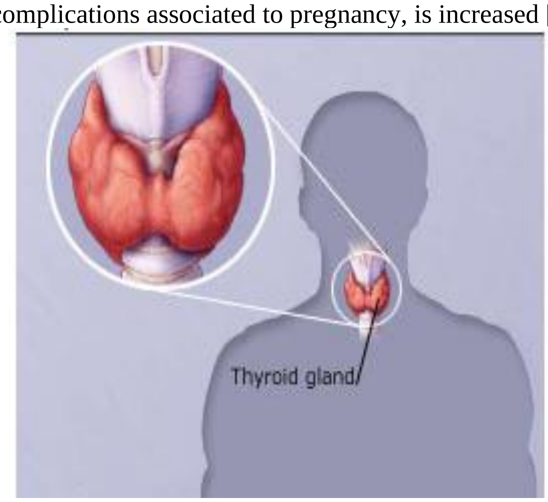

Fig 1. The thyroid in the body.

Another common inherited thyroid disorder is Graves' disease, which affects around 1% of people worldwide. Thyroid problems, if serious enough, may lead to an incurable form of cancer. Underestimated instances of thyroid diseases, such as thyroid storm and myxedema coma, are a leading cause of mortality. A case of acute hyperthyroidism is a thyroid storm, while the final step of untreated hypothyroidism is coma. Up to 60% of those who are affected by thyroid disease in the United States are unaware that they have it.
Thyroid disease affects women at an eightfold higher rate than males. Untreated hypothyroidism in pregnant women increases the risk of complications including miscarriage, premature delivery, and serious birth defects [4]. These numbers highlight the seriousness of thyroid-related illnesses but traditional way in detection has some problems like:

- There is a lack of precision, and the outcome is difficult to anticipate.
- Difficulty managing data with a high number of dimensions.
- Traditional scalability can be scalable by big data, but this may limit its applicability in real-world applications because large data may also scale traditional scalability.
- Time-consuming and labor-intensive: Traditional methods often require significant time and effort from trained professionals to manually analyze and interpret data.
- May not account for dynamic changes: Conventional approaches are often static and may not take into account changes in the patient's condition over time.

Machine learning techniques are becoming more and more important in the diagnosis of thyroid problems as a result of their unparalleled accuracy, efficacy, and ability to efficiently handle complex and comprehensive datasets. Many different machine learning approaches have been used to diagnose thyroid diseases.
A sort of method for machine learning known as a decision tree makes assumptions through the use of a model that simulates a tree. It is a hierarchical structure that allows you to visualize the possible outcomes of a decision based on a series of conditions or variables. Decision trees are constructed by iteratively dividing a training data set into subsets based on the best features that can be used to classify or predict the target variable. In other words, decision trees take a set of input data and divide it into smaller subsets based on specific features, creating a tree-like structure in the process. The tree's leaves reflect the final conclusion or categorization, while its nodes each represent a decision based on a particular aspect.

1.2 Types and diagnosis of thyroid disease

The symptoms of thyroid disease can vary from person to person and can take many different shapes. The following are some frequent forms of thyroid disease [5]:

1.2.1 Hypothyroidism

When the thyroid gland doesn't create enough hormones, this happens. Constipation, cold intolerance, weariness, and weight gain are among symptoms. the potential causes of hypothyroidism:

- Thyroiditis: thyroid gland inflammation (swelling) is the ailment. Thyroiditis might impair your thyroid's capacity to produce hormones.
- Hashimoto's thyroiditis: Hashimoto's thyroiditis is a form of autoimmune thyroiditis that does not cause discomfort and occurs when the cells of the body assault and damage the thyroid gland. This sickness is passed down through families.
- Postpartum thyroiditis: After giving birth, 5% to 9% of women develop this syndrome. In most cases, the effect is temporary.
- Iodine deficiency: Iodine is used by the thyroid to produce hormones. Many millions of individuals around the world are unable to get enough iodine.
- A non-functioning: In some cases, birth defects involve the thyroid gland. One in every four thousand newborns is affected. If the child is not addressed, he or she may develop behavioral and emotional issues. In the hospital, a screening blood test is administered to all infants to check their thyroid function.

1.2.2 Hyperthyroidism

When hormone production from the thyroid gland becomes excessive, this happens. Weight loss, anxiety, jitters, and heat sensitivity are symptoms. The following conditions can result in hyperthyroidism:

- Graves' disease: The entire thyroid gland may be hyperactive, producing an abnormally high level of hormone. This condition is also referred to as diffuse toxic goiter (thyroid enlargement).
- Nodules: Hyperthyroidism is a condition that can be caused by nodules in the thyroid that are overactive. Toxic multi-nodular thyroid nodule is the term used to describe a goiter that contains multiple nodules, while toxic autonomously functioning thyroid nodule is the term used to describe a goiter that contains only one nodule.
- Thyroiditis: This condition might cause either pain or no discomfort at all. When the thyroid is inflamed, it releases hormones it had been storing. This could last for a while, maybe a few months.
- Excessive iodine: Thyroid gland overcompensates by producing more hormones than the body needs when there is an excess of the mineral iodine in the body. A significant amount of iodine can be found in cough syrups and in some medications, such as the heart medicine amiodarone.

1.2.3 Hashimoto's thyroiditis

This particular form of hypothyroidism is brought on by an autoimmune condition. The thyroid gland is attacked by the immune system, which can result in inflammation and damage to the gland.

1.2.4 Graves' disease

This particular form of hyperthyroidism is brought on by an autoimmune condition. Antibodies made by the immune system prompt the thyroid gland to overproduce hormones.

1.2.5 Symptoms and diagnosis of thyroid disease

If you suffer from a thyroid illness, you could notice a number of different symptoms in your body. Unfortunately, the signs and symptoms of thyroid disease frequently and strikingly match those of other diseases and stages of life. Because of this, it may be difficult to establish whether the symptoms you are experiencing are the result of a problem with your thyroid or of something entirely unrelated.

The majority of symptoms linked with thyroid disease fall into one of two categories: those connected with having an excessive amount of thyroid hormone, also known as hyperthyroidism, and those connected with having an inadequate amount of thyroid hormone, also known as hypothyroidism.

Experiencing nervousness, anger, and worry may be symptoms of hyperthyroidism. Other symptoms of hyperthyroidism include being unable to fall asleep, losing weight, having a goiter or an enlarged thyroid gland, tremors, and muscle weakness. You may start experiencing irregular menstrual cycles or stop having them altogether, feeling heat-sensitive and having visual issues or irritated eyes [6].

Experiencing fatigue, gaining weight, suffering from memory loss, going through frequent, heavy menstruation periods, having coarse, dry hair, being silent or hoarse, and being sensitive to cold temperatures are some examples of the symptoms of hypothyroidism.

Diagnosing thyroid illness can be difficult at times because the symptoms of this ailment are similar to those of other conditions, which can lead to confusion in the process. You might experience symptoms that are similar to those you would have if you were suffering from a thyroid condition if you have reached an advanced age or if you are pregnant. The good news is that there are tests that can assist in evaluating whether or not the cause of your symptoms is a problem with your thyroid.

Blood tests are widely considered to be one of the most accurate ways to diagnose problems with the thyroid. Thyroid blood tests are able to establish whether or not your thyroid gland is functioning regularly by measuring the levels of thyroid hormones that are present in your blood. These tests require a blood sample, which is taken from a vein in your arm. Blood testing of the thyroid can uncover thyroid illnesses that can be caused by either hyperthyroidism or hypothyroidism, as well as indicate if you have hyperthyroidism or hypothyroidism. Thyroiditis, Graves disease, Hashimoto disease, goiter, thyroid nodule, and thyroid cancer are some of the conditions that can affect the thyroid. There are a variety of different blood tests that could be performed in order to evaluate your thyroid [7]:

- Thyroid-stimulating hormone (TSH): Hormone produced by the pituitary gland that keeps T4 and T3 levels stable in the blood. If your doctor suspects a problem with your thyroid hormones, this will likely be the first test he or she orders. Thyroid hormone excess (hyperthyroidism) is typically associated with a low TSH level, while thyroid hormone deficiency (hypothyroidism) is typically associated with an elevated TSH level. Direct measurements of thyroid hormones like thyroxine (T4) and triiodothyronine (T3) may be required if TSH is abnormal. The typical range for TSH in adults is between 0.40 to 4.50 mIU/mL (milliinternational units per liter of blood).
- T4: Thyroxine: Patients who are being treated for thyroid issues should undergo testing for both hypothyroidism and hyperthyroidism, as well as monitoring of their progress. In contrast to hyperthyroidism, which is characterized by high T4 levels, hypothyroidism is characterized by low T4 levels. Hyperthyroidism is characterized by high T4 levels. T4 levels in people should ideally fall somewhere within the range of 5.0 to 11.0 ug/dL (micrograms per deciliter of blood), as this is the optimal range that experts recommend.

2. MACHINE LEARNING ML

Machine learning, a branch of AI, is concerned with teaching computers to acquire new skills without being given any specific guidelines for doing so. It enables computers to recognize patterns, develop conclusions and predictions, and form judgments from vast amounts of data. You may find applications of machine learning in a wide variety of fields, such as the commercial world, the healthcare industry, and the financial sector, amongst others. Learning that takes place inside of a computer can be divided into three basic classifications: supervised learning, unsupervised learning, and reinforcement learning. When a machine learns from labeled datasets, this type of learning is known as supervised learning. In this type of learning, the desired output is already known. This indicates that the algorithm is able to anticipate the output for new inputs based on the patterns it has learnt from the labeled data [8]. Specifically, this means that the algorithm can predict the outcome for new inputs. When a machine uses unlabeled datasets to learn, this type of learning is known as unsupervised learning because the result that is intended is not known. In order to determine which parts of the data belong to which groups or clusters, the algorithm searches for recurring structures and connections within the information. The term "reinforcement learning" describes how a machine learns from its mistakes through the application of rewards and punishments. The algorithm learns to make decisions and carry out actions in a way that maximizes rewards while minimizing costs.

It is possible to employ supervised learning algorithms for a wide variety of tasks, some of which include classification, regression, and prediction, to name a few of these duties. The usage of these algorithms is applicable to a wide range of activities. When doing classification tasks, the goal is to assign a label to an input based on a set of classes that have been established in advance. This can be done by comparing the input to the classes. This can be accomplished by comparing the input to the classes. Performing a comparison of the input to the classes will allow you to achieve this goal. When engaging in tasks using regression, the objective is to make an accurate prediction of a continuous output value based on an input. This is something that needs to be done in order to have the finest possible results. When running prediction jobs, the objective is to make a prediction about an upcoming value based on an analysis of the data from the past. One way to accomplish this is to contrast the current value with that of the past.

The diagnosis of thyroid disease has become increasingly reliant on machine learning techniques as a result of their unrivaled precision, effectiveness, and ability to effectively handle both intricate and extensive data sets. We will go through many different approaches to machine learning that have been used in the past for the diagnosis of thyroid diseases.

2.1Machine Learning in Healthcare

There is an extensive amount of patient health data available, as is typical in the healthcare industry. Because of this, no human being can even attempt to process it. To this end, ML provides a means of making sense of the massive amounts of data and utilizing algorithms to foretell the future consequences of patients' diseases. In the field of medicine, machine learning assists users in gaining a better grasp of the efficacy of currently available programs and in locating the treatment that will yield the greatest results for patients taking into account their specific conditions [9]. There are a lot of application in healthcare like:

- Predictions on Cardio Vascular Diseases: It is of the utmost importance to have an accurate diagnosis and prognosis of the condition in order to properly treat cardiovascular disease. Research in the field of health care places a large emphasis on diagnosing cardiac conditions, which indicates the importance of this area of study. By factoring in age, family history, diabetes, hypertension, cholesterol, smoking, alcohol use, obesity, and lack of physical activity, the two most successful tools—neural networks and genetic algorithms have introduced a method for predicting cardiovascular disease. This method is based on the idea that these main risk factors can be used to forecast heart disease. Neural networks and genetic algorithms were used to demonstrate how to use the method.
- Diabetes Predictions: Diabetes mellitus is a condition that lasts for a long time and poses a significant problem for public health all over the world. It manifests as when elevated quantities of sugar are present in the blood for an extended period of time. In recent times, it has been suggested that it may be a factor in the development of Alzheimer's disease, in addition to being a primary cause of blindness and kidney failure. There are a variety of approaches that have been developed to investigate the reasons for and treat diabetes. A person's eating habits, sleeping habits, and level of physical activity, in addition to other indicators such as BMI (Body Mass Index) and waist circumference, are some of the hypotheses that have been discussed regarding the possibility of establishing a connection between diabetes risk and the day-to-day activities that a person engages in. This possibility has been brought up in relation to the possibility of establishing a connection between diabetes risk and the activities that a person engages in. These are only some of the many different aspects that have been taken into account.
- Hepatitis Disease Prediction: Hepatitis is the medical term for a condition that causes inflammation of the liver cells in addition to damage to the liver. The primary focus was on the development of intelligent medical decision support systems to aid medical professionals in the process of medical diagnosis by using learning patterns formed from the data collected on hepatitis.
- Cancer Predictions: Several modern models for breast cancer diagnosis and recurrence prediction have been developed during the past several years using multiple risk prediction algorithms and methods [10]. The prognosis for cancer of all types and subtypes is the subject of a great deal of research. This study emphasizes the significance of determining which data sets are most helpful for a given cancer type or cancer subtype and recommends validating an approach across multiple patient cohorts. Several studies elaborated on how data mining and machine learning could be used in the real world. The survival rates and recurrence probabilities of cancer patients were calculated using these approaches.

2.2 Steps to apply machine learning

Performing a task involving machine learning can be divided into the following steps [11]:

- Data collection: Information must be collected in a machine-readable electronic format. This holds true whether the data is stored in a SQL database, a text file, a spreadsheet, or on paper. An algorithm will take these facts as the learning material that it uses in order to develop knowledge that can be acted upon, and it will use this algorithm to generate that knowledge.
- The initial phase in every project using machine learning is to investigate and prepare the data. The quality of the data used in the project will determine a substantial amount of the overall quality of the project. At this stage of the machine learning process, there is often a large degree of engagement necessary from a human. This can take the form of teaching the machine something new. According to a number that is cited very frequently, eighty percent of the effort that goes into machine learning is dedicated to data. A sizeable portion of this time is devoted to a process known as data exploration, which requires acquiring a more in-depth understanding of the data as well as the intricacies that it entails.
- Training a model by using the data: The particular goal that has to be accomplished by applying machine learning will play a role in the selection of an appropriate algorithm, and the algorithm will represent the data in the form of a model. "Training a model" is the term used to describe this procedure.
- Evaluating the performance of the model: Given that every ML model generates a solution to the learning issue that is intrinsically biased, it is vital to create an accurate evaluation of the degree to which the algorithm was able to learn from its previous experiences. An accuracy evaluation of the model can be carried out with the help of a test dataset in a variety of different ways, depending on the kind of model.
- Enhancing the efficacy of models: If improved performance is required, it is essential to make use of the sophisticated techniques that are available in order to boost the performance of the model. It is possible that at some point in time you will be obliged to switch to an altogether new kind of model.

After these procedures have been completed, the model can then be put to use for the purpose for which it was designed, provided that it looks to be functioning satisfactorily. It is possible to use the model to provide score data for predictions, to make estimates of financial data, to provide information that is appropriate for marketing or research, or to automate processes. These applications can all be accomplished through the application of the model. Both the successes and the failures of the model that has been implemented may generate more data that can be used to train the model of the following generation.

2.3 Machine Learning Classifiers

In the context of supervised learning, classification is a type of issue in which a machine learning algorithm is trained to predict the output label of new inputs based on the labeled training data. This prediction is made by comparing the new inputs to the training data. The output label is typically a categorical variable, such as determining whether or not an email is danger, whether or not an image contains a car or not , or whether or not a consumer would purchase a product. Other examples include determining whether or not an image contains a cat or a dog. The algorithm is able to acquire new knowledge from the annotated training data by recognizing recurring themes and connections between the input features and the output labels. When the algorithm has finished learning these patterns, it will be able to determine the output label for fresh inputs that it has never seen before.

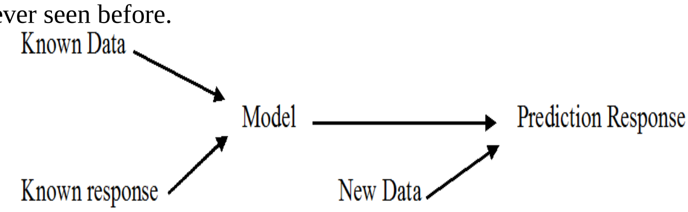

Fig 2.Supervised Learning.

There are a number of distinct classification algorithms, Decision Trees, Logistic regression, and Support Vector Machines are three of the most prevalent types of statistical models. These algorithms create predictions by employing a variety of methodologies. The nature of the problem and the kind of data that is being used should guide the selection of the appropriate algorithm.

2.3.1 Support Vector Machines SVM

SVMs search for a hyperplane that effectively splits the data in order to successfully categorize it into various categories. SVMs have been proven to be useful in the diagnosis of thyroid disease, according to a number of studies [12].
SVM offers various advantages over other machine learning algorithms, such as the capacity to easily handle high-dimensional data, which makes it well-suited for medical datasets. Another one of SVM's advantages is that it is faster than other machine learning methods. In addition to this, it is resistant to outliers, which are frequently found in medical datasets. When it is trained and calibrated appropriately, SVM is capable of achieving high accuracy in classification tasks. In general, the SVM is a potent machine learning method that has the potential to be utilized for the classification of thyroid diseases. The use of SVMs can result in more accurate diagnoses and more individualized treatment plans, which in turn leads to improved patient outcomes and an overall improvement in the patient's quality of life.

2.3.2 k-Nearest Neighbor k-NN

When it comes to machine learning, the phrase "k-Nearest Neighbor" (k-NN) refers to a straightforward non-parametric method that may be utilized for classification and regression testing. k-NN is also abbreviated as "k-NN." When k-nearest neighbors is used, the classification of a test data point is established by looking at the classifications of the k closest neighbors in the training set. This is done so that the test data point can be properly categorized. It has been demonstrated that the k-NN approach offers a high degree of accuracy when it comes to the identification of thyroid disorders.

2.3.3 Naive Bayes NB

The Naive Bayes algorithm is a straightforward approach to classification that is supported by probabilistic machine learning. The Bayes' theorem, which predicts the probability of an event occurring given the information that is easily accessible to the analyst at the moment [13], serves as the foundation for the Naive Bayes approach. Bayes' theorem is used to calculate the chance of an event occurring.
The Naive Bayes method was successfully applied as a diagnostic tool in a variety of research pertaining to the process of determining whether or not a patient has a thyroid condition.

2.3.4 Decision Trees DT

Because the information is organized into branches according to a predetermined set of guidelines, decision trees take the shape of trees. A decision tree classifier is characterized by its tree-like structure, which contributes to its high level of accuracy and consistency. The term "decision tree" is also synonymous with "CART," which stands for "classification and regression trees." In order to construct the trees, decision trees use relatively straightforward if-then logic.
When deciding how to split into two groups, the decision tree algorithm will often consider factors such as the Gini index, information gain, chi-square, and decrease in variance. A number of studies have shown that decision trees are a very effective method for the diagnosis of thyroid disease.

2.3.5 Random Forest RF

One approach that employs supervised learning is known as the Random Forest algorithm. The notion that merging numerous learning models into a single solution delivers superior outcomes overall is the core premise that drives the bagging strategy. This theory is at the heart of the bagging approach. One of the most important benefits of random forest is its ability to solve classification and regression problems. These two types of problems, when combined, account for the vast majority of challenges that contemporary machine learning systems must overcome. Random forest also has the ability to address problems involving clustering, which is another big advantage [14]. The Random Forest algorithm shares nearly identical hyper parameters with the decision tree and bagging classifier models. The researchers made the decision to implement it in the model in the hopes of enhancing both its precision and its resilience. The building of many decision trees, each of which is based on a unique subset of data and criteria, is an essential component of the Random Forest model. A number of studies have shown that the RF method is an accurate way to diagnose thyroid conditions.

2.3.6 Artificial Neural Network ANN

The acronym "ANN" refers to an artificial neural network, which is a kind of ML algorithm that models the structure and function of the human brain. Neurons, the fundamental elements of ANNs, are connected nodes that carry out a variety of data processing and analysis responsibilities.

After the individual nodes in a neural network have been activated for the first time, the connections between those nodes can then be created. Because it is now considered that these weights are having an effect on the outcomes of the prediction, it is necessary to optimize them; in order to do so, an iterative approach is utilized, which optimizes the results. However, for many years, there was no optimized strategy to select these weights. The lines that connect them each have a predetermined weight, but for many years, there was no optimized technique to decide these weights. Results from studies that employed ANNs with the intention of identifying thyroid conditions have thus far been quite favorable.

2.4 Deep Learning DL

Deep learning is a subset of machine learning that comprises educating computers to learn in the same way that people naturally do, which is through observation and the mimicking of the behavior of others. Deep learning is a term that was coined in the 1990s to describe the process of teaching computers to learn in this manner. Deep learning is one of the most important technologies that supports driverless autos since it enables autonomous vehicles to identify things like stop signs and differentiate between a pedestrian and a lamppost. In addition, deep learning is one of the most promising areas of research in computer science today.

Adding voice control into common electronic products like smartphones, tablets, TVs, and hands-free speakers requires this technology. The focus on deep learning has increased substantially in recent years, and rightly so. It means accomplishing goals that were thought to be unachievable in the past. Using "deep learning," a computer model can be trained to perform categorization tasks from raw data like images, texts, and audio. The computer's ability to take in data and transform it into useful knowledge makes this a reality. Models that have been trained using deep learning can achieve accuracy on par with or even better than that of humans. Models are trained using enormous amounts of labeled data in conjunction with intricate neural network topologies consisting of many layers. In many fields, including "image recognition," "natural language processing," "speech recognition," and so on, DL algorithms have reached state-of-the-art performance levels. When it comes to machine learning (ML), deep learning (DL) is a subfield in which neural networks play a pivotal role.

2.4.1 Multi-Layer Perceptron MLP

Recent research has shown that networks constructed using multilayer perceptron's (MLPs) are also capable of performing computer vision tasks and excel at the extraction of information on global features. In particular, a network that is purely MLP-based can accomplish a level of performance that is comparable to that of a transformer while utilizing a very straightforward architectural design. The modeling of the environment requires the utilization of a wide variety of methods, some of which may be combined. The level of intricacy of the issue that needs to be solved and the depth of one's understanding of the matter should both be taken into consideration when selecting the method that will be most effective. A full numerical model is the best method to employ if one has access to sufficient data and computer resources and an in-depth theoretical grasp of the problem [15].

The multilayer perceptron has been utilized in the solution of a wide range of problems, each of which can be placed into one of three categories: prediction, function approximation, or pattern categorization. The process of projecting future trends in a time series of data based on current and historical conditions is what we mean when we talk about prediction. The modeling of the relationship between the variables is the focus of the concept of function approximation. Pattern classification requires classifying data into discrete classes.

2.4.2 RBF Network

Input layer, Hidden layer, and Output layer make up the RBF network's three-layer structure [16]. Despite the fact that it has only a single hidden layer rather than a multilayer structure, RBF is capable of solving complicated issues in a manner that is analogous to that of a neural network that has numerous intermediary layers. The ability of the device in question to convert nonlinear input into linear output is the driving force behind the aforementioned phenomenon. The input field that is sent to each node in a hidden layer is nonlinear in nature, and it is typically given a Gaussian Function treatment when it is being processed by an activation function. Additionally, the ultimate output is the weighted summation of these nonlinear inputs to the output layer, which converts nonlinearity into linearity in the process. Although at first it was only used for the function of interpolation, it is currently utilized for a variety of large-scale prediction activities including function approximation, time series prediction, classification, and system control. Initially, it was only used for the function of interpolation.

2.5 Evaluation of Machine Learning Classifiers

In order to develop accurate models for the diagnosis of thyroid disease, it is essential to first compare and contrast available machine learning classifiers. To ensure accuracy, reliability, and resilience, it is crucial to evaluate the effectiveness of machine learning classifiers. The many measures that are used to evaluate machine learning classifiers will be discussed in the following text [17].

Examining a machine learning classifier's level of accuracy is one of the most common techniques to assess the effectiveness of the tool. The proportion of right predictions made by the classifier in relation to the total number of predictions is what is meant when we talk about accuracy.

The word "precision" refers to the ratio of correctly predicted positive occurrences to the total number of positive predictions made by the classifier.

The proportion of "actual positive cases" in a dataset that were correctly identified as "true positives" is known as the "recall" measure.

The F1 score is determined by averaging the points given for correct answers and good memories. It is helpful for non-normally distributed data sets because it strikes a balance between precision and recall.

Graphical representation of a machine learning classifier's efficacy using the receiver operating characteristic (ROC) curve. In this context, ROC refers to the "receiver operating characteristic." To do this, it compares the proportion of correct classifications (sensitivity) with the proportion of incorrect ones (1-specificity) over a range of criteria.

Area under the receiver operating characteristic curve (AUC) is a single metric that sums up a classifier's performance in machine learning. It may take on a value between 0 and 1, with 1 indicating perfect accuracy when classifying. Zero to one is the allowed range.

For a machine learning classifier, the confusion matrix is a table that shows both the actual classifications and the predicted classifications. Often referred to as a "confusion table." Accuracy, precision, and recall are just few of the evaluation metrics that may be calculated using the data presented by this method.

In addition to these metrics, cross-validation techniques like k-fold and leave-one-out cross-validation may be used to assess a machine learning classifier's efficacy. Cross-validation methods divide the dataset into training sets and testing sets, and then repeatedly test the classifier on various subsets of the data to see how well it performs overall.

When it comes to constructing accurate models for the detection of thyroid disease, one of the most crucial steps is to assess the performance of a variety of machine learning classifiers. We can ensure that machine learning classifiers are effective in detecting thyroid disorders and may assist physicians make more precise diagnoses by utilizing proper evaluation metrics and cross-validation techniques. This allows us to ensure that machine learning classifiers are effective in identifying thyroid diseases. As a consequence of this, we are able to test our machine learning classifiers to see if they can correctly identify thyroid problems.

3. THE PROPOSED APPROACH

The decision tree algorithm is a well-known machine learning technique that employs a tree-like representation of decisions and their potential outcomes. The decision tree algorithm may be trained on a dataset of Thyroid Disease for patient information, such as thyroid hormone levels, clinical symptoms, and other relevant criteria, in the case of thyroid illness categorization. The algorithm then utilizes this information to build a tree-like model of decision rules that may be used to categorize new patient cases as having a certain type of thyroid illness or being healthy. The decision tree technique is advantageous for thyroid illness categorization because of its capacity to handle both numerical and categorical data, transparency, and ease of use. Physicians can easily comprehend the generated decision tree model, which can assist guide therapeutic decision-making. Moreover, decision tree algorithms may be adjusted for high accuracy rates and are simply updated as new data becomes available. The process is following these steps:

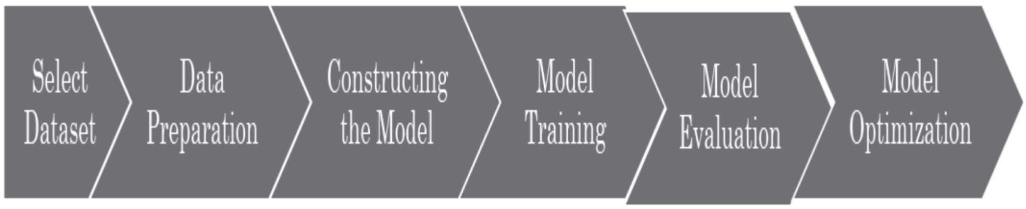

Fig 3. The process steps.

3.1 Dataset

The "hypothyroid.csv" file contains clinical information required to diagnose thyroid illness. It includes 3772 observations and 30 variables, such as patient age, gender, and thyroid hormone levels, among others. Each finding is categorized as "hypothyroid" or "negative." Figure 4 shows the dataset. The attributes are:

TABLE I. The attributes

Attribute Name	Possible Values
“age”:	“continuous”
“sex”:	“M, F”
“on thyroxine”:	“f, t”
“query on thyroxine”:	“f, t”
“on antithyroid medication”:	“f, t”
“sick”:	“f, t”
“pregnant”:	“f, t”
“thyroid surgery”:	“f, t”
“I131 treatment”:	“f, t”
“query hypothyroid”:	“f, t”
“query hyperthyroid”:	“f, t”
“lithium”:	“f, t”
“goitre”:	“f, t”
“tumor”:	“f, t”
“hypopituitary”:	“f, t”
“psych”:	“f, t”
“TSH measured”:	‘f, t”
“TSH”:	“continuous”
“T3 measured”:	“f, t”
“T3”:	“continuous”
“TT4 measured”:	“f, t”
“TT4”:	“continuous”
“T4U measured”:	“f, t”
“T4U”:	“continuous”
“FTI measured”:	“f, t”
“FTI”:	“continuous”
“TBG measured”:	“f, t”
“TBG”:	“continuous”
“Referral source”:	“WEST, STMW, SVHC, SVI, SVHD, other”
“binary Class”:	“P, N”

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	query hypothyroid	...	TT4 measured	TT4	T4U measured	T4U	FTI measured	FTI
0	41	F	f	f	f	f	f	f	f	f	...	t	125	t	1.14	t	109
1	23	F	f	f	f	f	f	f	f	f	...	t	102	f	?	f	?
2	46	M	f	f	f	f	f	f	f	f	...	t	109	t	0.91	t	120
3	70	F	t	f	f	f	f	f	f	f	...	t	175	f	?	f	?
4	70	F	f	f	f	f	f	f	f	f	...	t	61	t	0.87	t	70
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3767	30	F	f	f	f	f	f	f	f	f	...	f	?	f	?	f	?
3768	68	F	f	f	f	f	f	f	f	f	...	t	124	t	1.08	t	114
3769	74	F	f	f	f	f	f	f	f	f	...	t	112	t	1.07	t	105
3770	72	M	f	f	f	f	f	f	f	f	...	t	82	t	0.94	t	87
3771	64	F	f	f	f	f	f	f	f	f	...	t	99	t	1.07	t	92

3772 rows × 30 columns

Fig 4. Dataset.

3.2 Preprocessing

Before starting the training process, the necessary data processing operations must be carried out. When we see the data, we notice that there are “?” values. We consider these values as missing values, so they must be removed. We remove it by converting those values to null values. When calculating the number of null values for each column, we notice that the entire column "TBG" contains null values, which means it must be removed. For the remaining null values in the remaining columns, we remove the lines containing null values. All previous removal operations are carried out using functions provided by the sklearn library. We check the unique values for each column. We notice that there are columns that contain only one value, and therefore they are not useful in the training process. We remove those columns. Figure 2 shows the dataset after removing the null values.

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	query hypothyroid	...	tumor	hypopituitary	psych	TSH	T3	TT4	T4U	FTI
0	41	F	f	f	f	f	f	f	f	f	...	f	f	f	1.3	2.5	125	1.14	109
4	70	F	f	f	f	f	f	f	f	f	...	f	f	f	0.72	1.2	61	0.87	70
7	80	F	f	f	f	f	f	f	f	f	...	f	f	f	2.2	0.6	80	0.7	115
8	66	F	f	f	f	f	f	f	f	f	...	t	f	f	0.6	2.2	123	0.93	132
9	68	M	f	f	f	f	f	f	f	f	...	f	f	f	2.4	1.6	83	0.89	93
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3766	19	F	f	f	f	f	f	f	f	f	...	f	f	f	8.8	2.7	108	1.11	97
3768	68	F	f	f	f	f	f	f	f	f	...	f	f	f	1	2.1	124	1.08	114
3769	74	F	f	f	f	f	f	f	f	f	...	f	f	f	5.1	1.8	112	1.07	105
3770	72	M	f	f	f	f	f	f	f	f	...	f	f	f	0.7	2	82	0.94	87
3771	64	F	f	f	f	f	f	f	f	f	...	f	f	f	1	2.2	99	1.07	92

2643 rows × 23 columns

Fig 5.Dataset after removing null values.

We separate the target column from the rest of the columns. Below we will specify which columns to encode or which to scale using the number of unique values previously calculated. Columns whose number of unique values is greater than or equal to 10 are specified to be scaled and columns whose number of unique values is less than 10 are specified to be encoded. The columns that need to be scaled are scaled by the StandardScaler method. The columns that need to be encoded are encoded with the get dummies method. Figure 3 shows the columns after scaling and encoding. In the Target column there are two values "p" and "N" converted into 1 and 0 respectively.

0	-0.593285	-0.155859	0.605651	0.483484	0.734949	-0.013427	0	0	0	0	...	0	0	0	0	0	0	1	0	0	0
4	0.830789	-0.180055	-0.972673	-1.321686	-0.639743	-1.214683	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
7	1.321849	-0.118312	-1.701131	-0.785776	-1.505291	0.171382	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
8	0.634365	-0.185061	0.241423	0.427073	-0.334256	0.695007	0	0	0	0	...	0	0	0	1	0	0	0	0	1	0
9	0.732577	-0.109968	-0.487035	-0.701159	-0.537914	-0.506250	1	0	0	0	...	0	0	0	0	0	0	0	0	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3766	-1.673617	0.157028	0.848471	0.003986	0.582206	-0.383044	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1
3768	0.732577	-0.168374	0.120013	0.455278	0.429462	0.140581	0	0	0	0	...	0	0	0	0	0	0	0	0	1	0
3769	1.027213	0.002671	-0.244216	0.116809	0.378548	-0.136632	0	0	0	0	...	1	0	0	0	0	0	0	0	0	1
3770	0.929001	-0.180889	-0.001396	-0.729364	-0.283341	-0.691058	1	0	0	0	...	0	0	0	0	0	0	0	0	1	0
3771	0.536153	-0.168374	0.241423	-0.249866	0.378548	-0.537051	0	0	0	0	...	0	0	0	0	0	0	0	0	0	1

2643 rows × 25 columns

Fig 6. Dataset columns after preprocessing.

3.3 Methodology

In this section, we will talk about the problem facing our research. We will review some information about thyroid disease. We will then move on to presenting the proposed model and what it includes and explaining its operation.

3.3.1 Task definition

Thyroid illness is a common medical issue that impairs the thyroid gland's ability to produce hormones that govern numerous internal processes. Early identification and categorization of thyroid illness can result in more effective therapy and better patient outcomes. ML techniques, such as DT algorithms, are increasingly employed in medical diagnostics and can help with thyroid illness categorization.

3.3.2 Typical Architecture

The decision tree classifier provided a good ability in classification operations because of its method that simulates the human decision-making method. Bagging Classifier appeared to guide the work of classifiers and raise the level of their training by strengthening the work of classifiers that are considered weak in the classification as it works to train more than one classifier and get the best results from them. Figure 4 shows the proposed model.

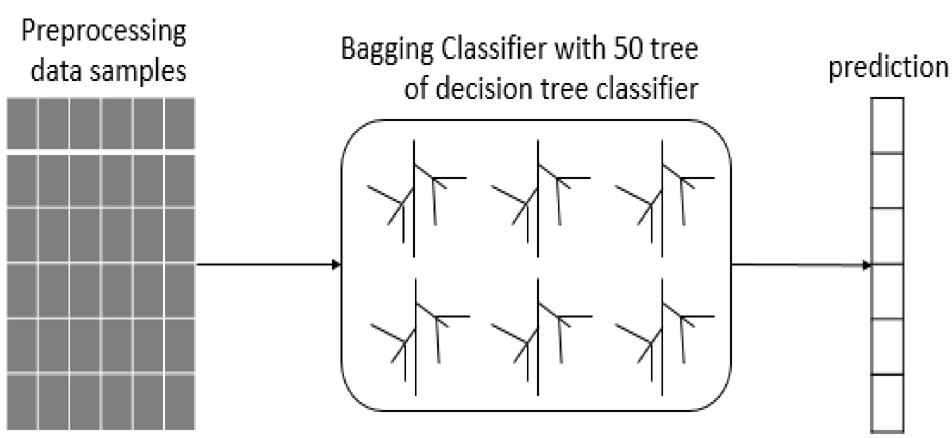

Fig 7. Proposed model.

3.3.2.1 Decision Tree Classifier

One common supervised ML approach for solving both classification and regression issues is the decision tree classifier. In order to produce a group of similar instances, it recursively divides the data into subsets according to the values of the input characteristics. A series of binary judgments is used to construct the tree structure, with each node in the tree indicating a rule for making a decision based on a feature value. The target variable's class labels or numerical values are shown as leaves in the tree diagram. A decision tree has three different types of nodes: the root, the intermediate, and the leaf.
The root node is the node at the very top of the decision tree and it stands in for the whole dataset. It is in charge of creating the first branching choice in the tree, and it does so by selecting the one characteristic that most effectively divides the dataset into two or more groups. The first node receives no new connections.

In a decision tree, internal nodes are any nodes other than the root and leaf nodes. They have one incoming edge and two or more outgoing edges and reflect choices depending on the values of one or more characteristics. Each branching edge denotes a potential value for the feature under evaluation. The dataset is partitioned into smaller subsets by internal nodes according to the feature values until a stopping condition is reached.
The ultimate results of a model are represented by the nodes at the very tip of the decision tree. Class labels or numerical values are assigned to each leaf node in a tree structure according to whether or not the job at hand is classification or regression. Each leaf node receives one incoming edge but sends no edges out.
A decision tree is trained by an algorithm that iteratively divides the dataset into smaller and smaller subsets, depending on the values of the characteristics, all the way down to the leaf nodes. In order to maximize information gain or reduce impurity, the algorithm seeks the optimal split at each internal node. Metrics like information gain and impurity are used to assess how well a decision tree split is doing.
Once the decision tree has been trained, it may be used to make predictions based on previously unknown data. Using the values of the new data point's properties, the algorithm determines the best route from the root node to the relevant leaf node. The projected value or class label for the new data point is then taken from the output of the leaf node. Figure 5 shows the types of nodes in a decision tree.
Some benefits of decision tree classifiers:
DT are widely used in data analysis and decision-making because of their intuitive structure.

- They are able to process numeric, category, and missing value data.
- Both classification and regression problems are amenable to decision trees.
- They are able to deal with feature-target connections that are not linear.
- The most crucial characteristics for accurately forecasting the target variable may be isolated with the use of decision trees.

Some drawbacks of decision tree classifiers:

- If the tree is too big or complicated, decision trees might fall victim to overfitting. As a result, the model's ability to generalize to novel data may suffer.
- As a result, they may create different trees for various subsets of the data depending on even subtle changes in the training data.
- For problems with complicated or non-linear interactions between features and the goal variable, decision trees may not be the best model.
- The effectiveness of decision trees may be affected by the selection of hyperparameters like the tree's maximum depth and the splitting criteria.

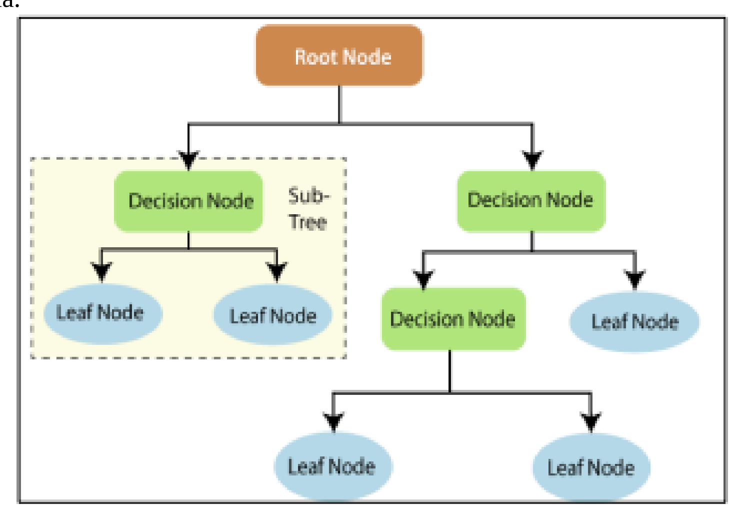

Fig 8. Decision Tree nodes.

3.3.2.2 Bagging Classifier

Bootstrap sampling is the foundation of Bagging Classifier; it involves generating several random samples of the training data with replacement, and then using each sample to train its own independent instance of the base classifier. Once enough classifiers have given their predictions, a final prediction is established based on a majority vote.
Because each base classifier is trained with a unique subset of features based on random sampling of the training data, Bagging Classifier may also assist in mitigating the impact of noisy or irrelevant features in the data. This may make the model more resistant to the effects of irrelevant or noisy data.

Random Forest is a form of Bagging Classifier that employs a bagging algorithm built on top of decision trees. Each DT that makes up a Random Forest is given its training using a unique combination of a distinct subset of the data and a unique collection of characteristics. The ultimate forecast is arrived at by tallying up the votes from all of the different trees.
Since Bagging Classifier necessitates training several instances of the underlying classifier, it may raise the computational cost of training and predicting. Parallel processing methods or utilizing smaller sections of the training data may help with this.
By decreasing overfitting and increasing resilience to noisy or irrelevant information, Bagging Classifier is a strong strategy for enhancing the performance of classification models. However, it may not always be the ideal option, and when opting to utilize it, it is important to weigh the pros and downsides, including the trade-offs between performance and computational cost.

3.4 Model Interpretation

The methodology can be divided into several steps:

3.4.1 Data Preparation

Preprocessing the data is the initial stage in employing machine learning algorithms for thyroid illness categorization. In this scenario, the "hypothyroid.csv" file contains a dataset of thyroid illness patients. Thyroid hormone levels, clinical symptoms, and other pertinent characteristics are all included in the dataset. To import and preprocess the data, we will utilize Python's Pandas module, which will include:

- removing unnecessary columns.
- Missing data handling.
- Encoding categorical variables.
- Data is separated into a testing set and a training set.

3.4.2 Choose a Feature

The following step is to choose the appropriate characteristics from the dataset to be utilized in the classification model. To discover the most significant characteristics, we may utilize several feature selection approaches such as correlation analysis and feature importance.

3.4.3 Constructing the Decision Tree Model

After the data has been preprocessed and the important characteristics picked, we can start creating the decision tree model. To create the model, we will use Python's Scikit-learn module. We will also increase the estimators parameter to 50 to guarantee that the model is resilient to noise and outliers

3.4.4 Model Training

To train the decision tree model, we will utilize the preprocessed and chosen feature data. The training set is used to fit the decision tree algorithm to the data.

3.4.5 Model Evaluation

The decision tree model's performance will then be evaluated using the testing set. To assess the model's performance, we will employ the following metrics:

- Accuracy: gauges the model's overall performance.
- Recall is the percentage of true positives that were successfully categorized.
- Precision is the percentage of true positives in the anticipated positive class.
- Confusion matrix displays the true positive, false positive, true negative, and false negative rates.

3.4.6 Model Optimization

Model optimization entails making slight adjustments to the model in order to boost its efficiency. Changing the model's hyperparameters and picking the most relevant characteristics to include are common approaches.
Changing the model's hyperparameters
The hyperparameters of a model are the parameters that are set before the model is trained, and they are responsible for determining the general behavior of the model. Examples of hyperparameters include the number of layers in a neural network, the level of regularization, the learning rate, and the type of kernel that is utilized in a support vector machine.
Experimenting with a model's hyperparameters can help us determine the best possible configuration for the model. It is possible for us to accomplish this objective if we first generate a grid of potential hyperparameters and then train the model using every possible combination of those hyperparameters. Then, we will need to identify which combination results in the best overall performance for the model, and we will use that combination going forward.
Feature selection

The process of deciding which features will be incorporated into the construction of a model is referred to as "feature selection." When seeking to identify thyroid disease, this may involve giving particular findings on blood tests, imaging features, or patient characteristics higher priority than others. The selection of features manually by subject matter experts is one approach, while the selection of features automatically by employing algorithms to determine the most helpful qualities is another approach.
In order to fine-tune the model, we might experiment with a number of different feature sets to see which one is the most successful. One way to accomplish this is by the use of forward selection, which entails beginning with a small collection of features and progressively extending it while keeping an eye on the correctness of the model. Alternately, we may use the process of backward elimination, in which we start with all of the features and then delete them one by one while keeping track of how accurately the model predicts things after each deletion.
The model's performance and accuracy in predicting thyroid disease may be enhanced by hyperparameter adjustment and feature selection. Important clinical implications include the potential for earlier and more accurate diagnosis, both of which improve patient outcomes.

4. METRICS

In order to evaluate the efficacy of the model, we make use of these distinct criteria. We will provide you with an overall summary of them. Once we have knowledge of the Confusion matrix that they are based on, we may compute the metrics that underpin them. In the field of machine learning, the effectiveness of classification models can be evaluated through the utilization of a confusion matrix, which can provide a summary of the results.
For the purpose of determining whether or not a classification model is effective, it is possible to compute metrics such as recall, accuracy, and precision.

4.1 Confusion matrix

The confusion matrix is a thorough description of the prediction results in the classification problem. It is one of the most important metrics, and it is also one of the most important metrics. Accuracy, Precision, and Recall are some of the other metrics that may be achieved by its use. The confusion matrix has four primary components. The term "true negative" refers to the number of cases that are, in fact, negative and that the model accurately identified as negative. FP stands for "false positive," which refers to the number of events that should have been classed as negative but were instead wrongly identified as positive by the model. The number of cases that are truly positive but were wrongly labeled as negative by the model is represented by the FN, which stands for false negative. The term "true positive" refers to the number of cases that are, in fact, positive and that the model accurately identified as positive. The following is the formula to determine these values:

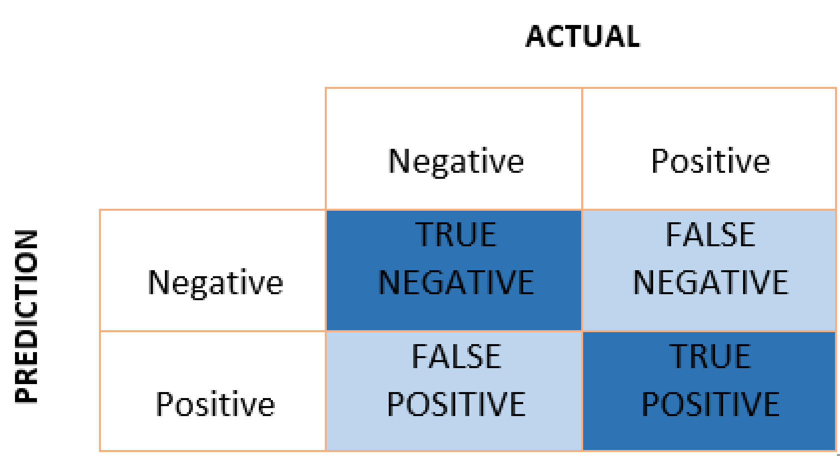

Fig 9. Confusion matrix.

- TN: The percentage of cases that the model correctly recognized as positive instances even when they weren't actually positive.
- FP: The percentage of examples that were incorrectly labeled as positive by the model when they should have been negative instead.
- FN: The percentage of occurrences that the model incorrectly classified as having a negative outcome, even if in reality they had a good outcome.
- TP: The percentage of cases in which the model correctly recognized them as positive instances even when they did not include those characteristics.

4.2 Recall

It is the percentage of times that our model came up with accurate predictions out of the total number of times that it got anything right. To determine the ratio, divide the total number of true positives and false negatives by the number of true positives only. This will give you the ratio. The retrieval's objective is to guarantee that the level of precision with which the machine learning model can recognize true positives will improve in a manner that is proportional to the level's value, as that is the desired outcome. It is essential to keep in mind that the fact that the percentage is large does not necessarily mean that the model is accurate merely because of that fact alone. This is due to the fact that a high percentage of recovery is indicative of the fact that the model will not neglect any instances of positivity when the percentage is high. Recall is defined as the ratio of positive results to all positive results plus all false negatives. Following is the formula used to determine recall:

Recall= (TP) / (TP + FN)

4. 3 Precision

The frequency with which desirable results were forecasted and then achieved. To calculate it, take the total number of genuine positives and divide it by the combined total of positive and false positive results. When it comes to machine learning models, higher precision equals fewer instances of false positives. This advantage increases in direct proportion to the market price of precision. In order for a model to have a high score for Precision, the majority of the model's correct positive predictions need to correspond to reasonable expectations.

A classification model's precision is the percentage of positive cases (true positives) it properly identifies, relative to the total number of positive instances (including false positives) categorized by the model. In other words, precision measures how well the model predicts the outcome of a given event. Precision is determined by the following formula:

Precision = (TP) / (TP + FP)

4.4 Accuracy

The term "accuracy" refers to a method that can be used to evaluate an algorithm in terms of how frequently it correctly identifies a given data point. The term "accuracy" refers to the degree to which the actual quantity of data points matches the fraction of those that were correctly expected. In order to arrive at an answer, you must first perform the computation of dividing the total number of true positives and true negatives by the total number of true positives, true negatives, false positives, and false negatives. If the algorithm correctly identifies a data point as either true or false, then that point is said to have a true positive or true negative value, respectively.

One popular way of measuring a classification model's efficacy is its accuracy. It is a metric that quantifies how many occurrences in a dataset were properly categorized. Accuracy measures how frequently a model correctly identifies a given input's class label. Following is the formula used to determine accuracy:

Accuracy = (TP + TN) / (TP + FP + TN + FN)

5. RESULT

The decision tree technique with n estimators = 50 was used to identify thyroid illness using the "hypothyroid.csv" dataset, and the resulting model obtained a high degree of accuracy with a 99.75% accuracy score. With a recall score of 0.998 and a precision score of 0.998, the model likewise displayed excellent levels of recall and precision. Additionally, the confusion matrix indicates that just two out of 793 total instances were misclassified, yielding a 99.75% accuracy rate. According to the matrix, the model properly identified 739 true positive instances and 52 true negative cases, while misclassifying one false positive case and one false negative case. Overall, these findings indicate that the decision tree method is extremely successful in properly classifying thyroid illness and may be used to help physicians in making appropriate diagnoses and treatment recommendations for patients

5.1 Comparison

Our proposed model was compared with two research papers in terms of results.

- In [18] several classification models such as KNN, Random Forest, Naive Bayes and ANN have been used. KNN showed 70% as a metric of Sensitivity, while Random Forest showed 94.8% as a metric of Sensitivity, Naive Bayes showed 98%, and ANN showed 77.4%. It can be noted that Naive Bayes is the best classification model among the previous classifiers in terms of Sensitivity.
- In [19] use LSTMBAM (Longa Short Term Memory Bidirectional Associative Memory) and show an accuracy of 98.94%. By comparison between the results of our proposed model and the results of the two research papers, we find that it has shown higher results in terms of Sensitivity and accuracy. Table II shows the Comparison of our proposed model with [18] and [19].

TABLE II: COMPARISON WITH OTHER WORKS.

Paper	Model	Dataset	Results
[[18]	KNN, ANN, Random Forest, Naive Bayes	Diagnose thyroid illness	KNN 70% ANN 77.4% Random Forest 94.8% Naive Bayes 98%
[[19]	LSTMBAM (Longa Short Term Memory Bidirectional Associative Memory)	Diagnose thyroid illness	Accuracy: 98.94%
Our model	Bagging Decision Trees Classifier	Diagnose thyroid illness	Recall (Sensitivity): 99.86% Accuracy: 99.75 %

6.CONCLUSION & RECOMMENDATION

6.1Recommendation

Based on the study's findings, the following suggestions for categorizing thyroid illness using the decision tree algorithm may be made:

- Feature Selection: Identifying the most significant traits for accurate predictions is a critical phase in the classification process. As a result, it is advised to employ sophisticated feature selection approaches to determine the most important features for categorization.
- Tuning Hyperparameters: Hyperparameters are important in the performance of machine learning algorithms. Consequently, it is vital to fine-tune the hyperparameters of the decision tree algorithm to improve its performance and produce superior outcomes.
- Data Preprocessing: Effective data preprocessing is essential for reducing noise and inconsistencies from data, which can increase the decision tree algorithm's accuracy. As a result, it is advised that the data be thoroughly preprocessed before being fed into the algorithm.
- Although the decision tree method displayed good rates of accuracy, recall, and precision in detecting thyroid illness, it is advised that the model's performance be evaluated on diverse datasets to assess its resilience and generalizability.
- Collaboration: Working together, data scientists and healthcare practitioners can design more accurate and reliable machine learning models for identifying thyroid illnesses. As a result, it is suggested that healthcare professionals be included in the model creation process to ensure that the models are matched with clinical needs.

6.2 Conclusion

To summarize, thyroid illness categorization utilizing machine learning algorithms, notably the decision tree approach, has shown significant potential in properly detecting thyroid disorders. The decision tree technique was used to the "hypothyroid.csv" dataset with good accuracy, recall, and precision, with just two cases misclassified out of 793 total cases. The decision tree algorithm can help healthcare practitioners make accurate diagnosis and treatment recommendations for patients, which can lead to better patient outcomes and quality of life. However, further study is needed to improve the decision tree algorithm's effectiveness by combining different algorithms, producing more interpretable and transparent models, and investigating real-time and automated diagnostics. Overall, the decision tree algorithm is a helpful tool for thyroid illness categorization, and its continuing development and optimization can considerably benefit both patients and healthcare providers.

6.3 Future work

Despite the fact that the decision tree algorithm has shown good accuracy rates in identifying thyroid illness, there is still considerable opportunity for future research in this subject. Some potential future research areas include:

- Integration of Several Algorithms: While the decision tree method is quite good in classification, combining different algorithms can result in more accurate predictions. As a result, future research might concentrate on merging several machine learning techniques to enhance overall thyroid illness categorization performance.
- Large-Scale Research: While the current study obtained great accuracy rates, it was based on a limited dataset. Future research might concentrate on larger datasets to increase the model's generalizability.

- Explain ability is critical for establishing confidence in machine learning models, particularly in medical situations where decisions have serious repercussions. As a result, future research might concentrate on designing decision tree algorithms that are easier to read, transparent, and explain to healthcare practitioners.
- Real-Time Diagnosis: The development of machine learning models capable of reliably diagnosing thyroid illnesses in real-time has the potential to significantly improve patient outcomes. As a result, future research might concentrate on constructing decision tree algorithms capable of performing real-time thyroid illness detection based on patient symptoms and other clinical data.
- Automated Diagnosis: Automated diagnosis can assist healthcare providers in more efficiently and accurately diagnosing thyroid problems. As a result, future research might concentrate on constructing decision tree algorithms that can autonomously detect thyroid illnesses based on clinical data without the need for human intervention.

Funding:

The authors confirm that no funding was acquired from any organization, grant agency, or institution. This research was undertaken without any external financial contributions.

Conflicts of Interest:

The authors declare no competing financial interests in this study.

Acknowledgment:

The authors would like to thank their institutions for providing the necessary facilities and guidance, which proved vital in achieving the study's objectives.

References:

[1] M. Mohamedali, S. R. Maddika, A. Vyas, V. Iyer, and P. Cheriyath, "Thyroid disorders and chronic kidney disease," *International Journal of Nephrology*, vol. 2014, 2014.

[2] K. AlexanderErik et al., "2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum," *Thyroid*, 2017.

[3] D. Gnocchi, K. R. Steffensen, G. Bruscalupi, and P. Parini, "Emerging role of thyroid hormone metabolites," *Acta Physiologica*, vol. 217, no. 3, pp. 184-216, 2016.

[4] A. Stagnaro-Green and E. Pearce, "Thyroid disorders in pregnancy," *Nature Reviews Endocrinology*, vol. 8, no. 11, pp. 650-658, 2012.

[5] A. Smith, J. Eccles-Smith, M. d'Emden, and K. Lust, "Thyroid disorders in pregnancy and postpartum," *Australian Prescriber*, vol. 40, no. 6, p. 214, 2017.

[6] J. P. Walsh, "Managing thyroid disease in general practice," *Medical Journal of Australia*, vol. 205, no. 4, pp. 179-184, 2016.

[7] A. J. Wassner, "Congenital hypothyroidism," *Clinics in Perinatology*, vol. 45, no. 1, pp. 1-18, 2018.

[8] A. Dhillon and A. Singh, "Machine learning in healthcare data analysis: a survey," *Journal of Biology and Today's World*, vol. 8, no. 6, pp. 1-10, 2019.

[9] B. Nithya and V. Ilango, "Predictive analytics in health care using machine learning tools and techniques," in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2017, pp. 492-499, doi: 10.1109/ICCONS.2017.8250771.

[10] J. Das, K. M. Gayvert, and H. Yu, "Predicting cancer prognosis using functional genomics data sets," *Cancer Informatics*, vol. 13, Suppl. 5, pp. 85, 2014.

[11] B. Nithya, "Study on predictive analytics practices in health care system," *IJETTCS*, vol. 5, 2016.

[12] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, p. 160, 2021.

[13] S. Umadevi and K. S. JeenMarseline, "Applying classification algorithms to predict thyroid disease," *International Journal of Engineering Science and Computing*, vol. 7, no. 10, pp. 15118-15120, 2017.

[14] K. Ammulu and T. Venugopal, "Thyroid data prediction using data classification algorithm," *International Journal of Innovative Research in Science and Technology*, vol. 4, no. 2, pp. 208-212, 2017.

[15] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14-15, pp. 2627-2636, 1998.

[16] "Although at first it was only used for the function of interpolation, it is currently utilized for a variety of large-scale prediction activities including function approximation, time series prediction, classification, and system control," 2016.

[17] P. Flach, "Performance evaluation in machine learning: the good, the bad, the ugly, and the way forward," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 9808-9814, Jul. 2019.

[18] T. Alyas et al., "Empirical method for thyroid disease classification using a machine learning approach," *BioMed Research International*, vol. 2022, 2022.

[19] D. Priyadharsini and S. Sasikala, "Novel hybrid LSTBAM - bidirectional associative memory deep learning based thyroid disease prediction," *Journal of Pharmaceutical Negative Results*, pp. 1179-1185, 2022.