A Review Paper on Utilizing Machine Learning in Predictive Modeling of Thyroid Disease

Fathma Khatun
Id:20104002
Department of Computer Science and Engineering, Bangladesh Army University of Engineering &Technology
E-mail:fathmakhatunmim@gmail. Com

Tonusri Mazumder
ID: 20104008
Department of Computer Science and Engineering, Bangladesh Army University of Engineering &Technology
E-mail:tonusrimazumder119@gmail.com

Nadira Anjum Biva
ID: 20104032
Department of Computer Science and Engineering, Bangladesh Army University of Engineering &Technology
Email: nadiraanjum080145@gmail.com

Tanjima Tabassum
ID: 201040038
Department of Computer Science and Engineering, Bangladesh Army University of Engineering &Technology
Email: cse038.tanjim@gmail.com

Sadia Binte Saleak
ID: 201040054
Department of Computer Science and Engineering, Bangladesh Army University of Engineering &Technology
Email: sadiaaumie@gmail.com

Abstract: Globally, thyroid problems provide a serious public health challenge that calls for prompt and precise diagnosis in order to provide appropriate treatment.The primary objective of this study is to develop a robust and reliable predictive model that can identify thyroid disorders at an early stage, enabling timely treatment and improving patient outcomes.Machine learning algorithms have become effective tools for diagnosing thyroid conditions more accurately, forecasting problems with the thyroid, and assisting healthcare practitioners in making decisions. The landscape of machine learning-based thyroid illness prediction is thoroughly surveyed in this review article. The study explores a number of techniques of machine learning, such as Random Forests, Decision Trees, & SVM. The focus is on how they can be applied to the analysis of a range of thyroid-related datasets, including clinical records that include hormone levels and complex image data from thyroid cyst ultrasound scans. There is a thorough examination of important issues like quality of information and ethical implications. To discover significant insights and trends about thyroid illnesses, we comprehensively analyze the datasets using state-of-the-art machine learning algorithms The review also addresses current articles that use algorithms of machine learning to detect the illness using various techniques. This review offers useful insights into the present state-of-the-art in thyroid illness prediction using machine learning through a thorough analysis of the body of existing research as well as current projects. It provides potential areas for future study while highlighting how crucial it is for data scientists and healthcare practitioners to work together. This review's synthesis contributes to the knowledge of the subject and offers practitioners and researchers a road map for utilizing machine learning for precise thyroid illness prediction and, subsequently, enhanced patient management.

Index Terms: *Thyroid Disease, Decision Tree T3, Logistic Regression, Naive Bayes, TSH, Support Vector Classifier, Random Forest, Extreme Gradient Boost, T4, K-Nearest Neighbor*

1. Introduction

Thyroid hormones are created by the thyroid gland and regulate several body processes. The recognition of thyroid dysfunction is essential to safeguarding patients from adverse effects. The diagnosis of THE disease of thyroid is made by laboratory testing for the hormone (TSH), T3 hormone, and T4 hormone. To detect thyroid function abnormalities, clinicians also employ thyroid scan or iodine uptake test findings, which take longer. In terms of time savings, computer-aided diagnostic techniques might be better for patients and clinicians. Due to demographic changes, management requirements, a lack of qualified staff, epidemics, and developing information technology, the healthcare sector is subject to substantial hazards.. These models were trained using thyroid data sets and confirmed through validation. The importance of early thyroid disease (TD) diagnosis is attributed to risk factors such as elevated cholesterol and blood pressure levels, and a rapid heart rate. Machine learning algorithms have proved effective for predicting TD, cutting costs, increasing revenue, and upholding high standards of patient care. The development of the body's baseline rate of digestion, calcium metabolism, and mental and bodily development all depend on the butterfly-shaped thyroid gland.[5]. Blood tests that measure thyroid hormone levels are frequently used to diagnose thyroid conditions. Recent technological developments have made it possible for machine learning techniques to identify thyroid diseases using medical imaging tools including an ultrasound, CT scans, even MRI scans. Around 42 million individuals worldwide suffer from thyroid disease, hormonal disorder with risk factors such as cholesterol levels, blood pressure that is elevated, and a rapid heart rate. [23] Due to the complexity of the problem and potential hazards to well-being and thyroid gland function, early diagnosis is essential the main objective of healthcare is to more precisely find ailments in their early stages Thyroid problems can lead to myxedema coma, what can prove fatal, and severe hyperthyroidism. Awareness of the condition is essential. [45] The efficacy and effectiveness of these decisions can be increased with systems. The aim of this research is to apply methods of machine learning. However it can be uncomfortable and unexpected. To aid in decision-making and categorize nodules. The best diagnostic technique, meanwhile, is FNAB, which needs knowledgeable cytopathologists and unclear cytology. For up to 80% of people with ambiguous thyroid nodules, thyroid surgery is necessary. It can be difficult to manage cytological uncertain thyroid nodules, and 50% of TN operations worldwide still involve diagnostic tests. Technology has greatly improved human lives through machine learning, especially in the field of medicine. It can use massive datasets to make decisions and forecast future data. Complex illness diagnosis is accomplished by the use of machine learning algorithms, which are frequently coupled with logic and mathematics. However, because there is a lot of data dispersed over numerous publications, determining the optimal algorithms for diagnosis might be difficult. In order to identify the best machine learning algorithms with the most consistent accuracy, this paper will do a literature review of several algorithms.

2. Literature Review

Thyroid disorders include a wide spectrum of conditions that affect the thyroid gland, an important organ for controlling metabolism and preserving homeostasis. The article examines recent developments in diagnosing, treating, and comprehending the underlying causes of thyroid disorders, shedding the spotlight on the intricate interplay of hereditary, ecological, and biological variables.

In [1] authors integrated data retrieved by the healthcare system with information utilized at third-party healthcare facilities and labs to evaluate the effectiveness of the learners on a real datasets. The goal of the research was to put thyroid disease into three separate groups: normal, hyperthyroidism, and hypothyroidism. The findings from this study, testing is generally 98.93% accurate.

In [2] the Graven Institute in Sydney, Australia provided the datasets by the authors . This paper will be used as a resource for research academics who wish to engage around thyroid condition prognosis. According to this study, testing is generally 98.93% accurate.

In [3] authors assessed the classifiers' performance using data gathered from a diagnostic facility in Dhaka, Bangladesh. In order to anticipate hypothyroidism in its early stages, this study set out to build an automatic prediction framework utilizing selection of features and classification approaches. According to this study, testing is generally 96.875% accurate.

In [4] authors assessed the classifiers' performance using data gathered by Chongqing General Hospital (Chongqing, China). t. In order to develop a new auxiliary diagnostic approach for patients with thyroid cancers, the goal is to build both malignant and benign thyroid disease diagnostic model. According to this study, testing had an overall accuracy rate of 87.4%.

In [5] authors assessed the classifiers' performance using real data collected by two hospitals in Haryana. This study paper's specific goal to develop a model using machine learning methods for evaluating euthyroidism, hyperthyroidism, and hypothyroidism in adults, adolescents, and children. This study indicates that testing is 86% accurate overall.

In [6] authors assessed the classifiers' performance using a real dataset acquired from the UC Irvine Repository. More specifically, this study stresses the benefits to society; thyroid disease's mortality as well as morbidity can be greatly reduced by recognizing the associated risk factors. According to this study, testing has an overall accuracy rate of 99.08%.

In [7] authors assessed the classifiers' performance using primary datasets made up of 1464 Indian patients. more specifically. The goal of the authors was to create and test models for predicting the risk of LLNM in these individuals based on (ML) methods. According to this study, testing accuracy is 85.03% overall.

In [8] authors assessed the classifiers' performance using data collected by Changzhou First People's Hospital. The goal of the authors was to create and test models that could forecast the possibility of LLNM in these individuals based on machine learning (ML) methods. According to this study, testing is generally 97.04% accurate.

In [9] authors analyzed a datasets that included 972 test instances and 2800 training (data) instances. The primary objective of this study is to identify patterns or correlations by the data without being aware of the anticipated results. According to this study, testing had an overall accuracy rate of 95.4%.

In [10] the Chandan Diagnosis Siba Jaunpur Center provided the datasets for the writers. In order to identify the most effective technique, most precise tool, for diagnosing illnesses of thyroid, apapers on data mining and thyroid disease diagnosis were reviewed in this work. This study indicates that testing is 87% accurate overall.

In [11] authors that used datasets from the Kaggle Machine Learning website. The steps with a great degree of accuracy. Computer learning methods are crucial in the medical field for creating a correct choice, accurate disease detection, and cost savings and the patient's time. According to the survey, overall accuracy is 96.6%.

In [12] authors assessed many studies to provide a thorough foundation on the disease classification. These are the phases of feature extraction and reducing features, classifying features, and testing GDA_WSVM for accurate thyroid illness diagnosis. The dataset was gathered by the author from the thyroid illness dataset repository at UC Irvine. With 94.8% accuracy and 91% specificity, the random forest algorithm gave us the best results.

In [13] the dataset was obtained from the Graven Institute in Sydney, Australia, and uploaded to the UC-Irvin. This work demonstrated how to use classification tools, as well as how to forecast thyroid intuitively. The overall best accuracy in this paper is 96.87%.

In [14] authors who used a dataset from the open-source scientific community in this study, we created a multiple-instance learning (MIL) integrated computer-aided diagnosis system that would concentrate on the classification of benign and malignant tissue. This study's overall accuracy rate shows 96%.

In [15] authors employed a very unbalanced dataset from the NGS-based TCGA cohort. Using the decision tree method, treat thyroid illness .They used a machine learning algorithm to construct and verify an automated, highly accurate diagnosis model for PTCs. This study demonstrates a 96.6% total accuracy for testing.

In [16] authors who consulted an American Thyroid Association dataset. In terms of predicting the characteristics of Bethesda III-V TN, our ML models did well. According to this study, testing accuracy ranges from 93.6% to 96.8%.

In[17] authors who used a dataset from the Cancer Genome Atlas (TCGA) dataset can be found . In particular, they use machine learning to thyroid carcinoma. They mostly employed the random forest algorithm. This study demonstrates a testing accuracy range of 98.3%-99.5%.

In [18] authors who gathered a dataset from the American Thyroid Association. This work represents a first attempt to create a clinical tool for physicians based on interpretable machine learning. Additionally, they will entail gathering additional data and creating probabilistic projections for predictions. They mostly employed the random forest algorithm. This study demonstrates a testing accuracy range of 85%-91%.

In [19] who gathered datasets from First Affiliated Hospital of Medical University. This paper is innovative in that it investigates machine learning-based intelligent ultrasonic diagnosis of papillary thyroid cancer using diffuse image properties of the thyroid and RBM learning techniques. The accuracy of this study is the best (90.21%).

In [20] the writers gathered data from the American Thyroid Association. In this research, machine learning approaches are primarily used to predict thyroid diagnoses in their early stages and to categorize the different types of thyroid disease. This research In comparison to SVM, KNN, and DT, the Random Forest method produced better results. 99.81% accuracy rate is the highest.

In [21]We find a study of Karachi, a city in Pakistan, where author classified thyroid disease using a dataset different from those found in the UCI and KEEL repository. In order to identify and diagnose thyroid disease, this study recommends the use of effective classifiers utilizing ML algorithms, with an average precision of 97.84%.

In [22] classified thyroid illnesses using a collection of data from the University of California, Irvine Machine Learning Repository may be found in the initial group. The article explores many machine learning classification-related choice and classification of features methods for identifying illnesses of the thyroid. The research project's general accuracy is 92%.

In [23] two thyroid samples from the repository for machine learning at UCLA were utilized to train two ML methods by the authors. This study set out to investigate how successfully machine learning methods could be applied to help identify thyroid disease. The study's overall accuracy is 96.74%.

In [24] that classified thyroid illnesses using data taken from the "Gene Expression Genomics database (GEO)" may be found .The research effort under consideration set out to understand the molecular mechanism underlying "Papillary tumors of the thyroid (PTC)".This study's total accuracy rate is 98%.

In [25] the Sick-thyroid dataset used by the authors of from the UCI repository. The effectiveness of eleven machine learning algorithms is examined in this study to determine which one forecasts thyroid risk the most accurately. This study's overall accuracy is 95.90%.
In [26] a dataset on thyroid disease was received from the District Headquarters (DHQ) Hospital for this investigation by the authors. The goal of the work was to identify the optimal KNN distance function out of all the distance functions used in the paper of .The study's overall accuracy is 98.00%.
In [27] the dataset used is found by the author of in the UCI ML Portal. There are 2,800 frequencies and 28 describes in it. This study will make use of the thyroid dataset and three specific feature selections. The study's overall accuracy is 98.00%.
In [28] the authors utilized the UCI respiratory learning machine to gather information on thyroid problems. Machine learning is crucial to the disease predictions process since it allows for the investigation and classification of thyroid disease based on information from hospital databases. The study's overall accuracy is 96.98%.
In [29] author s utilized the UCI respiratory learning machine to gather information on thyroid disorders. The objective this is to apply (ANN) in systems with IoMT in order to improve thyroid disease diagnosis using results from tests and semantic explanations. This study's overall accuracy is 96.68%.
In [30] author of examined classification-based machine learning algorithms and assessed the performance of decision trees, support vector machines, and KNN a set of training data from the UCI Machine Learning repository. The goals of our study were to classify different thyroid illnesses into distinct groups and evaluate the accuracy and stability of our model Accurate thyroid illness prediction and categorization can lead to favorable patient outcomes and informed clinical decision-making. This study's overall accuracy rate is 96.85%.
In [31] the authors integrated data retrieved by the employees at the "AOU Federico II" hospital in Naples to evaluate the effectiveness of the classification methods on a real dataset. In specifically, this study suggests a method based on algorithms for machine learning that makes use of thyroid hormone parameters as well as other clinical information about the patient to determine whether the patient's treatment needs to be raised, decreased, or kept the same. .testing had an overall accuracy rate of 90.4%.
In [32] the authors employed machine learning as well as feature selection techniques to examine clinical variables from de-identified gland disorder patients in a SEER database. The 34 distinct clinical factors in the data, including the age of the patient diagnosed and info on the lymph nodes.. According to this study, testing had an overall accuracy rate of 94.5%.
In [33] the authors assessed the classifiers' performance using real data from SEER 18 Study Data + the Hurricane that Impacted Louisiana. The main goal of this work was to develop a machine learning model for forecasting that may be used to forecast bone metastases (BM) among individuals with thyroid cancer (TC) that has just been diagnosed. This study suggests that the testing is 84% accurate overall.
In [34] the authors assessed the classifiers' performance using data gathered from the Gunma-based Wakayama Medical University Hospital. This study's objective was to create a computer-based method for identifying people who could benefit from medical treatments.. According to this study, testing has an overall accuracy rate of 94%.
In[35] the authors assessed the classifiers' performance using a real dataset acquired by Shanxi Medical University. To create a machine learning model to predict medical reactions using thyroid hormones data from the world, this article specifically set out to investigate whether thyroid hormones are connected with the clinical outcomes of depression. This study indicates that testing is 86% accurate overall.
In [36] authors examined the classifiers' performance using a real dataset that was available from the UC Irvine Machine Learning Repository. The goal of this investigation is to use decision trees, regression and random forests, classification trees (CART) to analyze a datasets on thyroid disease. Once the outcomes of these classifiers were obtained, we improved the findings by utilizing a bagging ensemble technique. According to this study, testing had an overall accuracy of 99.7%.
In [37] the authors assessed the classifiers' performance using data from Turkish hospitals. more specifically with the purpose of determining the appropriate surgical intervention and lowering operation expenses by avoiding unneeded interventions and surgical procedures, various intelligent models are created in this work to support thyroid scoring systems. According to this study, testing is generally 99.79% accurate.
In [38] the authors assessed the performance of the classifiers using data that was available from the UC Irvine Machine Learning Repository. This study was done to compare how well machine learning techniques may be used to aid in the detection of thyroid illness. Two different types of ML are used in this research: Random Forest and Back-propagation using Multilayer Perceptron Neural Network. Accuracy is nearly 98% for this paper.
In [39] the authors examined the classifiers' performance using data from the UCI Repository. One of the goals of this work is to design an integrated machine learning algorithm for hypothyroid prediction using the dual filtering pre-processing techniques according to this study, testing is generally 99.74% accurate.
In [40] the New South Wales Institute and the Garvan Institute in Sydney provided the dataset to the authors . The objectives were to demonstrate how well the Feature library performed and select a standout ML model for thyroid forecasting from a variety of models. This study reveals that testing has an aggregate precision of 99.45%.
In [41] the PUMCH Ethics Committee gave its approval to the author of for the categorization techniques using real datasets. For the prediction of lymph node in thyroid, the ML algorithm may be helpful. This study found that the overall accuracy percentage for testing was 83.5%.

In [42] authors addresses the information obtained from the dataset chosen from the UCI machine learning repository and the analysis and classification models used in the thyroid disease. Numerous machine learning techniques are employed in the prognostication of common diseases. This study found that testing was general accuracy 98.62%.
In [43] the writers of evaluated the study.. The major objective of this research was to create a machine learning forecasting model that might be used to predict bone metastases in thyroid cancer patients. According to this study, the testing is generally 91.56% correct.
In [44] the Gunma-based Machine Learning Repository dataset from the UCI machine learning repository was used by the authors of to evaluate the performance of the classifiers. The purpose of the paper is to explain different data mining procedures and statistical features for interpretation of thyroid illnesses by different authors to get different prospects and for different approaches. This study found that testing has a 98.62% overall accuracy rate.
In [45] the authors of using an actual dataset that Shanxi Medical University had collected, a journal evaluated the performance of the classifiers. According to this study, testing is generally 83.24% correct.
In [46] the author of used a General Electric, USA, dataset. The impact of nodules on the spectral behavior of the US picture will be investigated. This study found that the overall accuracy percentage for testing was 96.8%.
In [47] the UCI machine learning respiratory system provided two sets of thyroid datasets that were used to conduct the study described in this publication by the author of . To deal with redundant and pointless features, these techniques mostly rely on potent classification algorithms. This study found that testing is typically 96.89% accurate.
In [48] the authors using data from the UCI Repository, the publication evaluated the performance of the classifiers. Utilizing separate convolution processes, the system combines DWI and ADC images to enable the search for deeper textural patterns in thyroid nodules. This document has a close to 88.82% accuracy rate.
In [49] the authors use information from the UCI Machine Learning Repository, the study evaluated the performance of the classifiers. This study shows how data mining and machine learning approaches can help the medical community and healthcare system. This study found that testing is typically 94.8% accurate.
In[50] the authors obtained information from the UCI) Repository, the study evaluated the performance of the classifiers. The suggested system might be expanded to incorporate Internet of Things (IoT) ideas. According to this study, testing had an overall precision of 99.46%.

3. Methodology

a. Model/Algorithm/Technology:

i) Random Forest: A continuous data-splitting process across pre-set parameters is used in the decision tree method. Three nodes—internal nodes, leaf nodes, and root nodes—compose a decision tree. The internal node represents the test on a characteristic. [5]It is well recognized to be used in decision-making since it can be used to both constant and discrete variables. The accuracy, however, could be poor if the data set is large since the model might over fit.

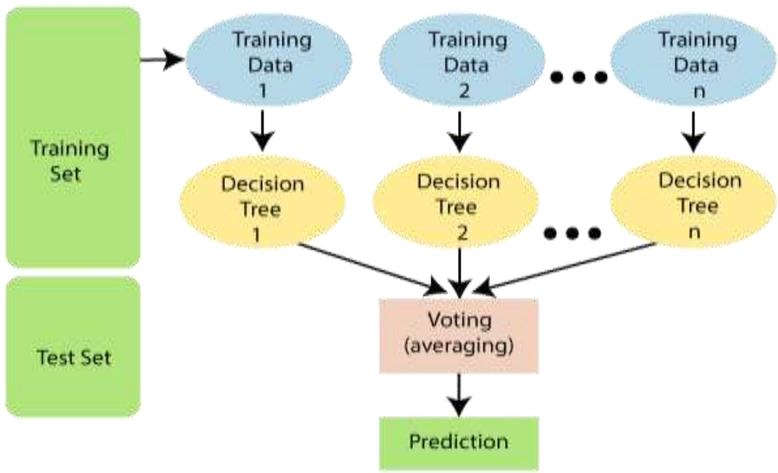

Figure 1: The algorithm for Random Forest

ii) **K-Nearest Neighbor:** The KNN model has long been known for its "laziness," as its method does not call for any assumptions about how the facts on which it depends are distributed. The KNN method does not require training before generating a prediction; therefore any fresh data can be incorporated without significantly affecting the accuracy. Additionally, KNN performs well in pattern recognition and predictive analysis, particularly when dealing with classification issues involving discrete values. [12] The K-value is going to be an integer that is odd if the classes are even-value, a measure of the number of people who are nearby, has an impact on the categorization and prediction. However, this method requires a lot of memory and takes a while to obtain the ideal K-value.

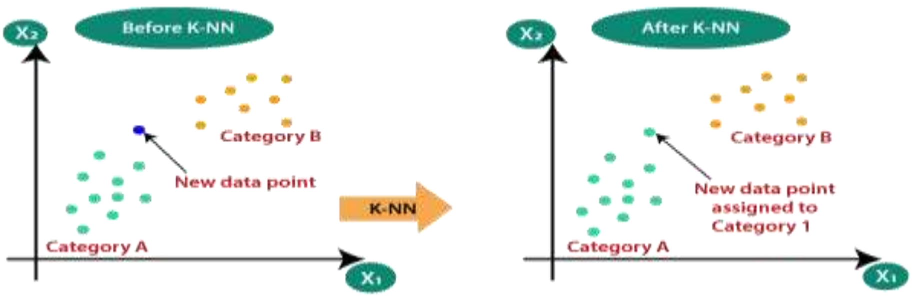

Figure 2: The algorithm for KNN

iii) **Logistic Regression:** When using the logistic regression method for classification, a linearly independent boundary is used. Linearly separable bounds between samples from various groups must first be found in order for the logistic regression technique to work. The logistic formula is then used to determine the likelihood of falling within each class described in relation to the choice limits. By providing discrete forecasts, logistic regression may analyse risk factors and anticipate the likelihood of contracting a certain disease in various analyses. When there are numerous category features, the model may encounter challenges.

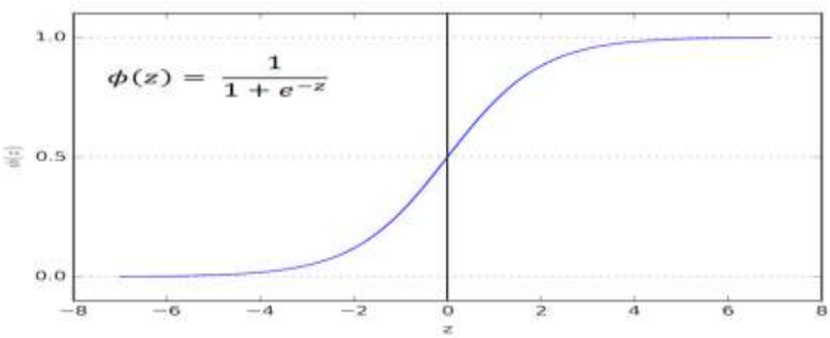

Figure 3: The algorithm for Logistic Regression

iv) **Naive Bayes:** A method for supervised machine learning called the naive Bayes classifier works under the assumption that every pair of characteristics is independent and is based on the Bayes theorem. It requires low storage capacity due to its usage as a progressive learner, adaptability to values that are lacking, and simplicity of output.

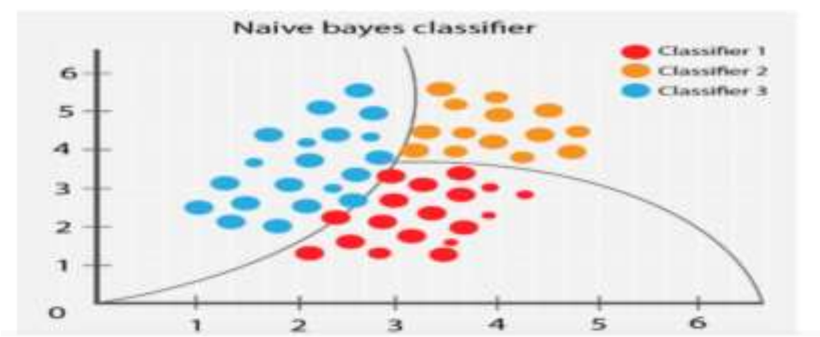

Figure 4: The algorithm for Naïve Bayes

v) **Support Vector Machine:** A hyper plane, a straight line created by splitting classes in the target with the displayed characteristics in n-dimensional space, is used to construct SVM models. The points situated in the hyperplane's side are classified as belonging to the same class, but the ones which fall in the opposing region belong to a different class. The model's main goal is to select the best line that may maximize the distance between data points. As more options for choosing the penalties and loss equations were needed and the data set was large, the model employed in this work was SVC rather than SVM.

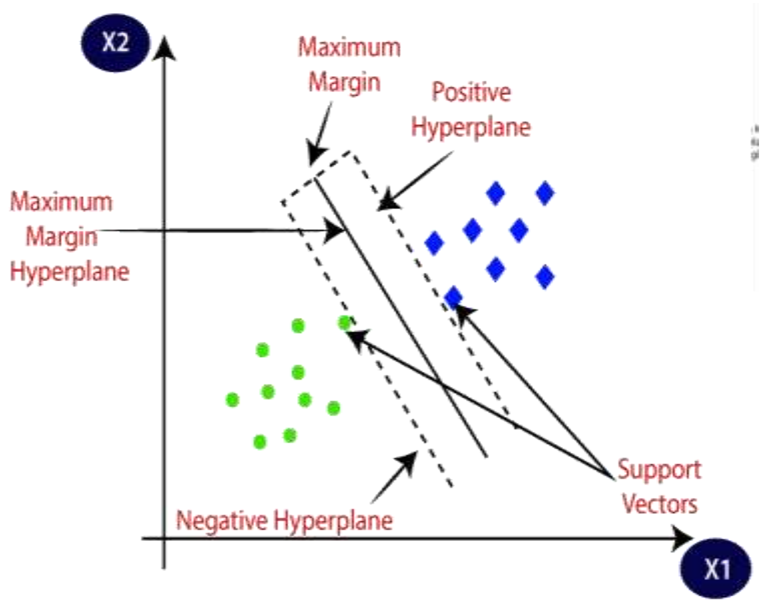

Figure 5: The algorithm for Support Vector Machine

vi) **Extreme Gradient Boost:** As tree-based ensemble machine learning technique called XGBoost has a considerable amount of predictive potential. The gradient-boosting decision tree serves as the foundational paradigm for XGBoost, which mixes different decision trees in a boosting fashion. [28] The performance of XGBoost is enhanced by employing the greatest possible tree depth, learning percentage, sub sampling ratio, and number of boosts in addition to the gradient boosting principles. More importantly, XGBoost optimizes the function's objective, the size of the tree, and the magnitude of the weights, all of which are influenced by conventional regularization settings.

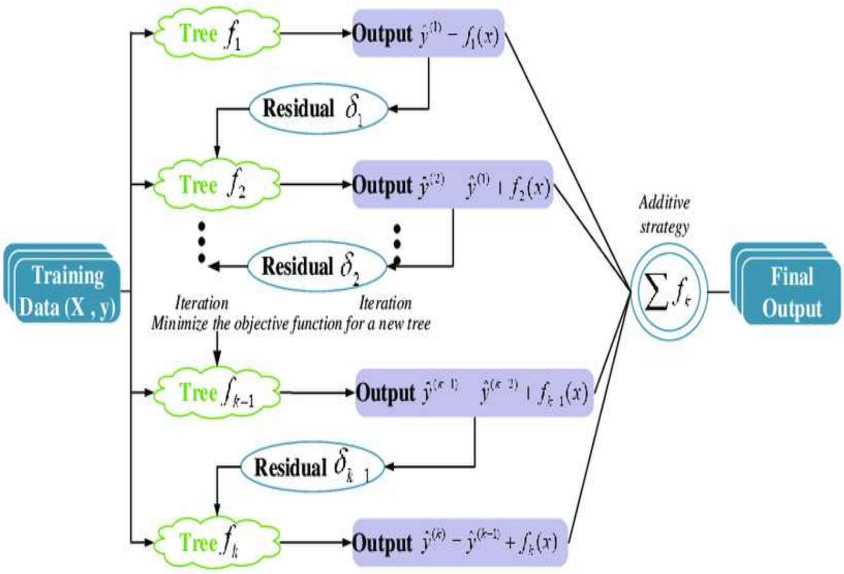

Figure 6: The algorithm for Extreme Gradient Boost

vii) **Adaptive Boosting (AdaBoost):** The framework of AdaBoost seeks to comprehend the flaws of approaches with low classification performance and to fix them [26]. Adaboost employs "sequential joins," as opposed to Random Forest, which utilizes "parallel joins." To create a classifier with a greater classification success rate. This is as a result of its flexible structure. AdaBoost is used to improve the classification effectiveness of DT based algorithms, which is greatly impacted by the amount of misfits in the evidence [27].

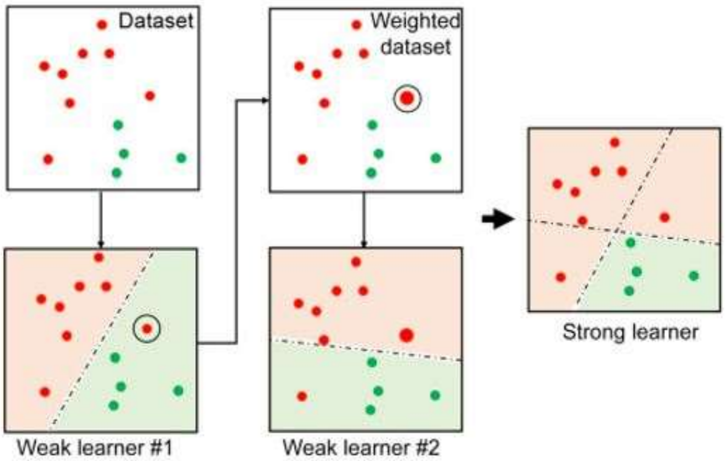

Figure 7: The algorithm for Adaptive Boosting (AdaBoost)

2. Finding/Result:

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[1]	Thyroid Disease Classification Using Machine Learning Algorithms	ICPAS 2021	External hospitals and laboratories specialized	16 input,7 algorithms	Tabular	age, gender, pregnant, surgery, TSH, T3, T4	Support Vector Machines, Random Forest, Decision Tree, Naïve Bayes, Logistic Regression, K-Nearest Neighbors, MLP	Random Forest, 98.93%
[2]	Prediction Of Thyroid Disease (Hypothyroid) In Early Stage Using Feature Selection And Classification Techniques	ICICT4SD 2021	The Graven Institute in Sydney, Australia	519 data, total of 9 attributes	Tabular	RFE, UFS, PCA	Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) & Naive Bayes (NB)	Logistic Regression, 98.93%
[3]	Thyroid Disease Prediction Using Machine Learning Approaches	Springer 2020	"Diagnostic center Dhaka, Bangladesh"	1024 patients,5 algorithm	Tabular	T3, T4, TSH	Logistic Regression, Decision Trees and K-Nearest Neighbor	KNN, 96.88%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[4]	Machine learning for identifying benign and malignant of thyroid tumors: A retrospective study of 2,423 patients	frontiers 2022	Chongqing General Hospital (Chongqing, China)	2,042 patients,4 algorithms	Image	sex, age	Gradient Boosting (XGBoost), Random Forest (RF), light Gradient Boosting Machine (LightGBM), and Adaptive Boosting	Gradient Boosting, 87.40%
[5]	"Classification of thyroid diseases using machine learning frameworks"	International Journal of Health Sciences 2022	Two hospitals in Haryana	"539 thyroid patients,5 algorithm	Image	Married, TSH, T3,T4	Logistic Regression, Naïve Bayes, Decision Tree, Random forest,	Random Forest, 93%
[6]	Hypertension and Obesity: Risk Factors for Thyroid Disease	frontiers 2022	Two hospitals in Haryana	1131 treatment, 1 algorithm	Tabular	name, age, sex, married, TSH, T3,T4,	Rule mining algorithms	Rule mining algorithms, 87%
[7]	Hypertension and Obesity: Risk Factors for Thyroid Disease	International Journal of Scientific and Technology 2020	US	I131 treatments	Tabular	name, age, sex, married, TSH, T3,T4, sick	Support Vector Machine, Bagging, Boosting	Bagging, 92%
[8]	A comparative analysis of eight machine learning models for the prediction of lateral lymph node metastasis in patients with papillary thyroid carcinoma	Frontiers 2022	"Changzhou First People's Hospital	1236 patients,8 algorithm	Tabular	"TSH, T3, T4	Logistic Regression, Gradient Boosting Machine, Extreme Gradient Boosting, Random Forest (RF)	Random Forest , 84%
[9]	Thyroid Disease Detection Using Machine Learning Approach.	Journal of Xi'an University Of Architecture & Technology 2022	"Changzhou First People's Hospital	2800 training (data) instances and 972 test instances	Image	Married, TSH, T3, T4	Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes	Random Forest, 95%
[10]	The role of data mining in diagnosing the diseases: A case study of detecting the thyroid disease	International Journal Artificial Intelligent and Informatics 2022	Chandan Diagnosis Siba Jaunpur Centre, University of California's	30 patient:33 5 data sets	Image	Skin, weight, facial swelling, heart rate, TSH, T3, and T4,age	Artificial Neural Networks (ANN), decision tree (DT), and Naïve Bayes	Support vector Machine, 87%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[11]	"A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis.	IEEE 2020	Kaggle Machine Learning Website	3000 train thyroid dataset recorded.	Tabular	T3, T4, TSH, Age, Gender, and Pregnant	"Decision Tree ID3, Native Bayes(NB), SVM, Density based clustering, Artificial Neural Networks, Radial Based	Decision Tree ID3, 96.74%
[12]	Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach.	Hindawi 2022	UCI dataset thyroid disease repository.	7200 types of records. Each record has 25 features.	Image	T3,T4, Age , Sex, medication, patient condition.	Decision tree, Random Forest algorithm, KNN, Artificial Neural Networks (ANN),	KNN , 94.8%
[13]	Thyroid Disease Prediction Using Machine Learning Approaches.	Springer 2020	“Graven Institute in Sydney, Australia, uploaded to the UC-Irvin."	5 attributes and 215 instances	Image	T3, T4, TSH, Age, Weight, Gender	logistic regression, decision trees and k-nearest neighbor (KNN) algorithms, ID3	Decision Tree , 96.87%
[14]	Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques	Medicine 2021	open-source scientific co.	99 patients and 134 ultrasound	Image	age, sex, nodule composition and shape	machine (SVM) , artificial neural network (ANN)"	Support Vector Machine, 96%
[15]	Highly accurate diagnosis of papillary thyroid carcinomas based on personalized pathways coupled with machine learning	OXFORD 2020	The NGS-based TCGA cohort contains a highly imbalanced dataset	"23 patients, 11726 merged genes from three studies as input features"	Image	age, gender, pathologic stage	Bayes algorithm, Pathifier algorithm, random forest and Support Vector Machine	Support Vector Machine , 96.6%
[16]	Application of machine learning methods to guide patient management by predicting the risk of malignancy of Bethesda III-V Thyroid nodules	EJE 2023	American Thyroid Association	7917 records reviewed, 1288 patients with 1335 TN.	Tabular	age, gender, TSH and anti-thyroid auto-antibodies levels	risk of malignancy (RM), multivariate logistic regression (LR), random forest	Random Forest, 96.8%

Ref. NO	**Topic name**	**Publisher and Publication Year**	**Source of Dataset**	**No of Dataset**	**Type of Data**	**Important features**	**Algorithms**	**Best algorithm with Accuracy**
[17]	"Identification of Potential lncRNAs and miRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning"	Hindawi 2021	Cancer Genome Atlas (TCGA) dataset	506 PTC patients, lncRNA and mRNA data of 502 PTC patients, and data of 506 PTC patients were recorded.	Image	lncRNA, mRNA, , miRNA, , family history, personal history, data	Random forest (RF), decision tree and support vector machine (SVM)"	Random Forest , 99.5%
[18]	"Predicting Malignancy in Pediatric Thyroid Nodules: Early Experience With Machine Learning for Clinical Decision Support"	OXFORD 2021	American Thyroid Association	198 pediatric patients, 140 potentially eligible patients, 53 had benign	Tabular	sex, age, number of years on dialysis, cytometry and total number of human leukocyte	Random forest (RF), decision tree, false positive (FP), false negative (FN)	Random Forest , 91%
[19]	Ultrasonic Intelligent Diagnosis of Papillary Thyroid Carcinoma Based on Machine Learning	Hindawi 2022	First Affiliated Hospital of Medical University.	a total of 70 patients	Image	Male patients, Female patients, Medical History	Genetic algorithm, ID3 algorithm of decision tree.	Decision Tree, 90.21%
[20]	Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques	MDPI 2022	American thyroid association.	The dataset contains 7200 samples, and each sample has 21 attributes	Tabular	treatment condition, health condition, general health	Random Forest, Random Forest (RF), Logistic Regression, Support Vector Machine (SVM), AdaBoost (ADA), Gradient Boosting Machine Decision tree (DT), Logistic Regression, (DT), Gradient Descent (GD),	Random Forest , 99%
[21]	Performance Analysis of Machine Learning Algorithms for Thyroid Disease	Spinger 2021	District Headquarter s (DHQ) Teaching Hospital, Dear Ghazi Khan, Pakistan	three classes and 309 patient entry	Tabular	blood pressure, pulse rate and BMI , T3,T4,TSH, Gender,Age, Pregnant	KNN,SVM, Naive Bayes, Decision Tree, Logistic Regression	Random Forest , 97.84%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[22]	Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques	IEEE Xplore 2020	UCI Machine Learning Repository	7200 instances and 27 attributes.	Tabular	TSH, T4U, TT4, T3, and FTI	Naive Bayes Algorithm, Random Forest Algorithm	Naïve Bayes, 92%
[23]	Decision Support for Diagnosing Thyroid Diseases using Machine Learning	Ijsrp 2021	UCI machine learning repository	2 dataset, One set-215 instances, the other set-1621 instances, 9 attributes	Tabular	(T3) and Thyroxin (T4),TSH	neural network with back propagations, random forest classifier,ID3 algorithm	Random Forest , 96.74%
[24]	Comparative analysis of thyroid disease and predict them using machine learning techniques	Journal Of Health Science 2022	Gene Expression Omnibus database (GEO)’	one hundred and sixty-four tissue samples	Image	T3, T4, TSH, TRH	Random Forest (RF), K Nearest Neighbors (KNN), and Support Vector Machine (SVM)	Random Forest , 98%
[25]	Application of machine learning algorithms to predict the thyroid disease risk: An experimental comparative study	PeerJ Computer Science 2022	UCI machine learning repository	3,162 rows and 25 data columns, 2,800 instances and 972 cases.	Tabular	age, gender, thyroid_, TSH , T3,TT4, T4U, FTI, TBG, Pregnant,	Decision tree, Random forest, Light- GBM, XGBoost, KNN, ANN, SVC, CatBoost, Extra Trees	Random Forest , 95.90%
[26]	Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease	Journal of Chinese Institute of Engineers 2021	District Headquarters (DHQ) Teaching Hospital, Dear Ghazi Khan, Pakistan	Three classes and 309 patient entry	Tabular	blood pressure, pulse rate and BMI ,T3,T4,TSH, Gender, Age, Pregnant	KNN,SVM, Naive Bayes, Decision Tree, Logistic Regression	ANN, 98.00%
[27]	Machine learning framework with feature selection approaches for thyroid disease classification and associated risk factors identification	Journal of Electrical Systems and Information Technology 2023	UCI ML Repository	2,800 instances	Tabular	Age, TSH, T3,TT4, T4U, FTI	Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), (KNN), (AB), Gradient boosting (GB)	KNN, 98%
[28]	PREDICTION AND PROVIDING MEDICATION FOR THYROID DISEASE USING MACHINE LEARNING TECHNIQUE (SVM)	Journal of Computer and Mathematics Education 2020	UCI respiratory learning machine	3152 records, contains total 23 columns and using 2521	Image	T3, T4, ANN,	Decision Tree (DT), Support Vector Machine (SVM), (KNN), PCA	Random Forest , 96.98%

Ref. NO	**Topic name**	**Publisher and Publication Year**	**Source of Dataset**	**No of Dataset**	**Type of Data**	**Important features**	**Algorithms**	**Best algorithm with Accuracy**
[29]	"A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things"	Journal of Supercompu ting 2020	UCI machine learning repository	consists of 21 input variables of 7200 patients, with 15 binary and 6 continuou s attributes	Tabular	age, gender, thyroid, TSH , T3,TT4, T4U, FTI, TBG, Pregnant	k-nearest neighbor (KNN), ANN, SVM, MMLP, Native Bayes	SVM, 98.68%
[30]	UTILIZING MACHINE LEARNING FOR THE DETECTION OFTHYROID DISORDERS	Industrial Engineering Journal 2023	UCI Machine Learning repository	3,772 patients, including 3,212 negative cases and 560 positive cases	Image	T3, T4, and TSH and	"artificial neural network (ANN), Random Forest, SVM,"	SVM, 98.85%
[31]	Thyroid Disease Treatment prediction with machine learning approaches	Elsevier B.V 2021	AOU Federico II" Naples hospital	2211 instances referring to a total of 247 patients.	Tabular	Height, Body mass index, Age, Pathology Weight, Gender	AdaBoost, Gradient Boosting , XGBC, and CatBoost., Bayesian algorithm and the K-Nearest Neighbors,	ANN, 0.904
[32]	Machine Learning and Feature Selection Applied to SEER Data to Reliably Assess Thyroid Cancer Prognosis	Scientific Reports 2020	Surveillance Epidemiolog y and End Results (SEER) database	In total, 25,063 thyroid cancer entries	Tabular	age, location of nodal disease, and primary disease	namely, Fisher's discriminant ratio, Kruskal-Wallis' analysis, and Relief-F	Extra-Tree Classifier9 4.5%
[33]	Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer	WILEY 2021	SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana	A total of 17,138 patients.	Tabular	T stage, histology, race, sex, age, and N stage.	Decision tree (DT)	Multilayer Perception, 84%
[34]	Development and preliminary validation of a machine learning system for thyroid dysfunctiondiagnosis based on routine laboratory tests	Communicat ion Medicine 2022	Wakayama Medical University Hospital, Gunma	176,727 subjects	Tabular	"sex, aspartate amino red blood cell count (RBC), serum, (ALP), uric acid (UA), (LDH), albumin, (A/G), FT3, FT4 and TSH "	K-nearest neighbor (KNN), CATBoost	Random Forest , 94%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[35]	Exploring the potential of thyroid hormones to predict clinical improvements in depressive patients: A machine-learning analysis of the real-world based study	Elseiver 2022	Shanxi Medical University	2086 individuals	Tabular	FT3, FT4 and TSH	SVM with Gaussian kernel function.	Gradient-Boosted Decision Tree, 86%
[36]	Prediction of thyroid disease using decision tree ensemble method	Springer Link 2020	UC Irvine Machine Learning Repository	3710 instances and 29 features	Tabular	Age, Sick, T3,TT4, TSH ,	decision tree, random forest tree, and extra tree	HAMD + HAMA + FT3+FT4, 99.69%
[37]	Reducing Operation Costs of Thyroid Nodules Using Machine Learning Algorithms with Thyroid Nodules Scoring Systems	Springer Link 2022	Hospitals in Turkey	2609	Tabular	T3, T4, TSH	J48 Decision Tree (J48 DT),	Decision tree, 99.79%
[38]	Decision Support for Diagnosing Thyroid Diseases using Machine Learning	International Journal of Scientific and Research Publications, 2021	UCI machine learning repository	1621 instances	Tabular	Age, Gender and Pregnancy	Multilayer perceptron Neural Network with and Random Forest (RF)	Random Forest , 98.49%
[39]	HYBRIDIZATION OF MACHINE LEARNING ALGORITHM FOR THE PREDICTION OF HYPOTHYROID	Science World Journal 2022	"University of California Irvine (UCI) "	1221 instances	Tabular	FT3, FT4, and TSH	. SVM and Adaboost M1 Random Forest	Random Forest , 99.80%
[40]	Prediction of Thyroid Disease using Machine Learning Approaches and Featurewiz Selection	JTEC 2023	Sydney-based New South Wales Institute and the Garavan Institute	126 feature	Tabular	T4, T3, and TSH ,T4U, TT4, and FTI.	Ensembled machine learning algorithms (Random Forest and Extreme Gradient Boost).	Logistic Regression 99.45%
[41]	Machine Learning Algorithms for Prediction of Central Lymph Node Metasis in Patient with Papillary Thyroid Cancer	frontiesin 2020	Peking Union Medical College Hospital's (PUMCH)	1103 people , Male is 297 and Female is 806.	Tabular	age, sex, FBG, FT3, FT4, TPO-Ab	RFC,ANN, DT,GBDT, XGBoost ,AdaBoost	Random Forest , 90.00%
[42]	Interactive Thyroid Disease Prediction System Using Machine Learning Technique	IEEE 2022	University of California Irvine (UCI) Machine Learning Repository"	29 attributes of dataset	Tabular	(T3 and T4), iodine	ANN,SVM, DT,KNN	Ultra-sound , 98.62%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[43]	Thyroid Classification and Segmentation in Ultrasound Images Using Machine Learning Algorithms	IEEE 2022	The Institutional Review Board approved the study. Health Insurance Portability and Accountability	50 images, 30 thyroid nodul ,20 normal	Tabular	Age,T3, T4, iodine	ANN,RFC, DT, XGBoost, AdaBoost	KNN, 96.50%
[44]	Thyroid Disease Diagnosis Based on Genetic Algorithms using PNN and SVM	IEEE 2019	University of California Irvine (UCI) Machine Learning Repository	53 and 189 2D thyroid US images	Tabular	FBG,FT4, FT3	Support Vector Machine SVM, ANN , RFC	ELM, 98.62%
[45]	Diagnosis method of thyroid disease combining knowledge graph and deep learning	IEEE 2020	Shanxi Medical University	Positive 1200 and negative 3600. Total 4800.	Image	T4,T3, Appetite, TSH	BPNN,SVM,RNN,LSTM	Random Forest Classifier , 83%
[46]	Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches	frontiesin 2022	General Electric, USA	7200 samples and 21 features	Image	Gender, Familiar anamnesis Menarche	Proposed Method	knowledge Graph ,99%
[47]	Machine Learning Algorithms in Healthcare	IEEE 2020	University of California Irvine (UCI) Machine Learning Repository	174 patient	Tabular	Height, Body mass index, Age, Weight, Gender	ANN, Random forest, LR, SVM, Decision tree	Proposed method , 99%
[48]	THYROID CANCER COMPUTER-AIDED DIAGNOSIS SYSTEM USING MRI-BASED MULTI-INPUT CNN MODEL	IEEE 2022	University of California Irvine (UCI) Machine Learning Repository	40 benign nodules in 32 patients	Tabular	DWI and ADC images	CNN	Naive Bayes, 88.00%
[49]	Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach	Hindawi 2022	University of California Irvine (UCI) Machine Learning Repository	7200 samples and 21 features	Tabular	Age , thyroxin Male (M), female(F), Sick , Pregnant , Thyroid surgery, T131 treatment	KNN,ANN, Naive Bayes, Random Forest	Proposed Method , 94.80%

Ref. NO	Topic name	Publisher and Publication Year	Source of Dataset	No of Dataset	Type of Data	Important features	Algorithms	Best algorithm with Accuracy
[50]	Thyroid Prediction Using Machine Learning Techniques	SPRINGER 2019	University of California Irvine (UCI) Machine Learning Repository"	174 patient	Tabular	Age , Male, female, Sick , Pregnant , Thyroid surgery, T131 treatment , Query	SVM, Decision Tree, Logistic Regression, Random Forest, Feed Forward image	Random Forest, 98%

4. Results and Discussion

The studies employ Random Forest Logistic Regression, gradient boosting, Rule mining methods, Support vector machines, Decision Trees, KNN, Artificial Neural Networks, Extra-Tree Classifiers, Multi-layer Perceptron, and Naive Bayes and may algorithms.

Paper	Algorithm	Accuracy
[1]	Logistic Regression	98.93%
[2]	KNN	96.88%
[3]	Gradient Boosting	87.8%
[4]	Bagging	92%
[5]	Decision Tree	96.87%
[6]	Support Vector Machine	96.6%
[7]	Random Forest	99.5%
[8]	Naïve Bayes	92%
[9]	ANN	98%
[10]	Extra Tree Classifier	94.5%

Table No-01: Finding accuracy of all algorithm

The studies employ many algorithms. The greatest accuracy methods utilized in the publications we analyzed are displayed in this graph. In this case, Gradient Boosting has the lowest accuracy (86%) while Random Forest has the highest accuracy (99.80%).

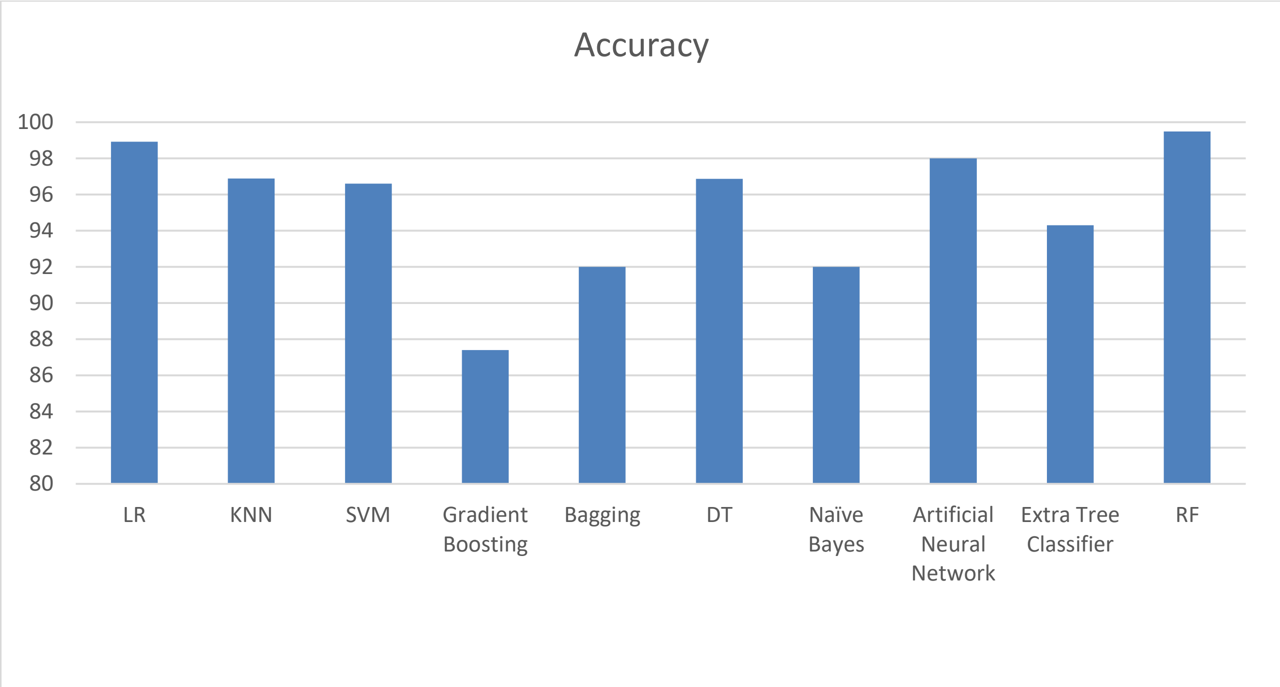

Figure-8: Highest accuracy of the algorithms

A table listing the specific algorithms utilized in the research article:

Algorithms	Number of Uses Algorithm of Papers	Count
Random Forest (RF)	[1][2][4][5][8][9][18][17][16][15][12][24][23][22][20][30][27][25][45][44][42][41][39][38][37] [31][49][47][50]	29
Support Vector Machine (SVM)	[1][2][7][8][9][14][12][11][18][17][15][24][21][20][30][29][28][27]	18
K-Nearest Neighbors (KNN)	[1][3][13][12][21][30][29][28][27][26][25][24][23][32][35][30]	16
Artificial Neural-Network (ANN)	[8] [9] [10][15][14][13][23][25][30][29][31][45][43][49][47]	15
Logistic-Regression	[1] [2] [3][5][8][13][21][20] [16][26][50][47][34]	13
Decision Tree	[1][2][3][5][8][24][23][22][21][20][19][18][17][16][15][14][13][12][11] [10][28][27][26][25][44][43][42][39][38][37][50][47]	32
Naive Bayes	[2][5] [7] [9] [10][11][17][15][22][21][29][26][49]	13
Extra-Tree Classifier	[32]	1
Gradient-Boost	[4] [7] [8][20][27][20]	6
Bagging	[7]	1

Table No-02: Number of algorithms that utilized in the papers

Table No-02 shows us which algorithms are employed in which papers. Additionally, this algorithm's reference paper number is assigned. Here several algorithms are utilized in various publications. Additionally, this approach has a reference paper number associated with it. We observe here that numerous papers were given the same approach since in this situation; one study occasionally employed different algorithms to display the algorithms utilized.

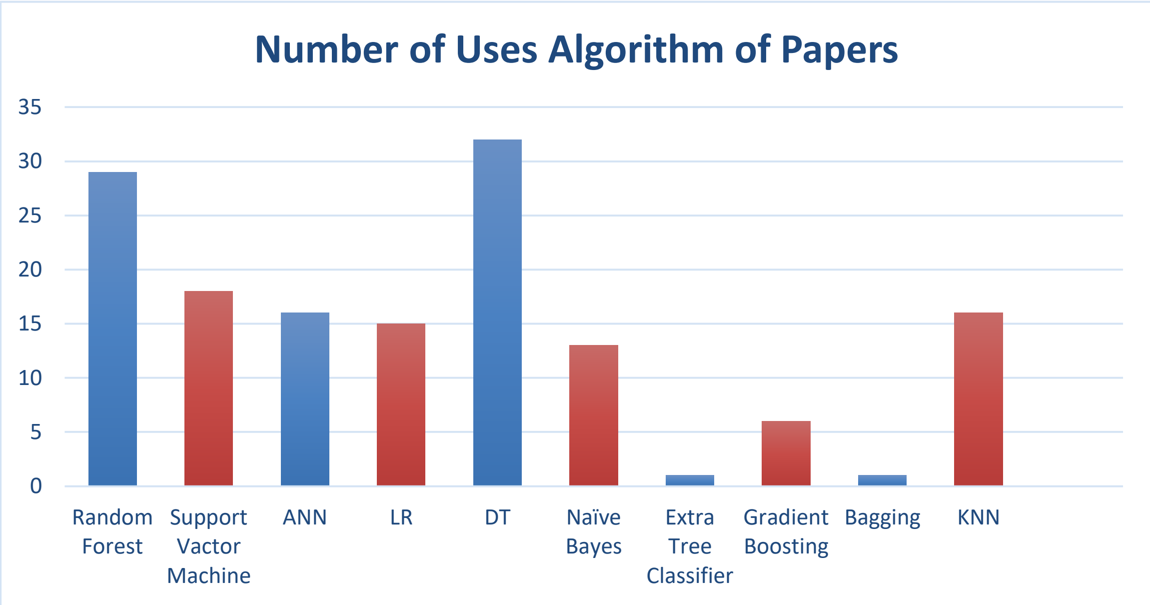

Figure-9: Portrayal of the algorithms utilized

In total, 10 Algorithms are shown in Fig-9, with Decision Tree being the most frequently used algorithm and Extra Tree appearing in just one of them. We effectively illustrated this with a graph chart that we created using them.

5. Conclusion

In an effort to make the data from the full study more accessible to professionals and machine learning researchers, we have organized it. Recent advances in modeling prediction for thyroid disease are largely attributable to the accessibility of extensive healthcare knowledge and the use of machine learning tools. It might be a thrilling and worthwhile undertaking to write a review paper on the potential future application of thyroid disease prediction. There are many fascinating directions to explore as machine learning algorithms advance.[1]The creation of extremely precise, individualized diagnostic models, which lessen both false positive and false negative detection of thyroid disease, is one of the most important topics. This review article will provide an in-depth analysis of thyroid illness prediction as it is today, setting the groundwork for additional research and progress in this important area of healthcare. It is still very important to focus on early thyroid prediction, where predictive modeling can be quite useful in spotting minor signals of malignancies in thyroid nodules. The future for studying in the field of thyroid illness prediction is tremendously bright, with new directions for investigation and innovation being made possible by developments in science and information analysis. Additionally, telemedicine and machine learning integration have tremendous possibilities for monitoring and diagnosis, guaranteeing that people in impoverished areas have access to top-notch medical treatment.[32] To evaluate the long-term impacts of artificial intelligence (AI) upon patient results and healthcare effectiveness, future studies are crucial. Consider how emerging technologies, such as quantum technology for sophisticated simulations or advances in bioinformatics for in-depth molecular analysis, may affect the prediction of thyroid disease. The need to create the highest standards for privacy, prejudice, as well as AI use in medicine makes the ethical component equally important. The detection, therapies, and control of thyroid problems could be profoundly changed by upcoming studies and articles in this topic, transforming healthcare in the process. The approaches used to forecast thyroid disease are thoroughly reviewed in this review paper, which also explores prospective directions for further study.[2] This review attempts to highlight the future extent of thyroid illness prediction, focusing light on creative tactics and cross-disciplinary strategies that could transform diagnosis and treatment. It does this by assessing recent accomplishments, obstacles, and new technology. This dataset contains ML techniques that is useful in healthcare industry. A list of the best algorithms from all the articles we looked at has been created. The research-reviewing literature does not favor the methodology. This is commended. We only focused on the algorithms used and how precise they were. The results of this study should make it easier for practitioners for forecasting the illnesses we have discussed here. One common finding is that the machine must first gather the information before creating a prognosis. Usually, it depends on how the computer was programmed. We were able to exhibit the best diagnosis accuracy using a data set from another research study because we did not employ any applications in our investigation. We didn't always receive accurate diagnoses for our patients. This is one of the problems with our work. Another limitation is the choice of database. By choosing a few more databases, it is possible to get the variety of results that we presented here.

References

[1]. Salman, K., & Sonuc, E. (2021). Thyroid Disease Classification Using Machine Learning Algorithms. *Journal of Physics: Conference Series*, *1963*(1). https://doi.org/10.1088/1742-6596/1963/1/012140

[2]. Chaubey, G., Bisen, D., Arjaria, S., & Yadav, V. (2021). Thyroid Disease Prediction Using Machine Learning Approaches. *National Academy Science Letters*, *44*(3), 233–238. https://doi.org/10.1007/s40009-020-00979-z

[3]. Riajuliislam, M., Rahim, K. Z., & Mahmud, A. (2021). Prediction of Thyroid Disease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques. *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021 - Proceedings*, 60–64. https://doi.org/10.1109/ICICT4SD50815.2021.9397052

[4]. Tseng, Y.-J., Sen, J., Business School, P., Tae Keun Yoo, I., Y-y, G., Z-j, L., Guo, Y., Li, Z., Du, C., Gong, J., Liao, P., Zhang, J., & Shao, C. (n.d.). *Machine learning for identifying benign and malignant of thyroid tumors:* A retrospective study of 2,423 patients. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.960740

[5]. Verma, S., Popli, R., Kumar, H., & Srivastava, A. (2022). Classification of thyroid diseases using machine learning frameworks. *International Journal of Health Sciences*, 7552–7566. **https://doi.org/10.53730/ijhs.v6nS1.6603**

[6]. Liu, F., & Zhang, X. (2022). Hypertension and Obesity: Risk Factors for Thyroid Disease. *Frontiers in Endocrinology*, *13*. https://doi.org/10.3389/fendo.2022.939367

[7]. Mir, Y. I., & Mittal, S. (2020). Thyroid Disease Prediction Using Hybrid Machine Learning Techniques: An Effective Framework. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, *9*, 2. www.ijstr.org

[8]. Feng, J. W., Ye, J., Qi, G. F., Hong, L. Z., Wang, F., Liu, S. Y., & Jiang, Y. (2022). A comparative analysis of eight machine learning models for the prediction of lateral lymph node metastasis in patients with papillary thyroid carcinoma. *Frontiers in Endocrinology*, *13*. https://doi.org/10.3389/fendo.2022.1004913

[9]. Sankar, R. R., & Sanjay Mahulikar, C. (n.d.). *Thyroid Disease Detection Learning Disease Detection Using Machine Learning Approach*. https://doi.org/**10.33564/IJEAST.2021.v05i09.028**

[10]. Hambeishi, N. K., Hussein, S. A., & AlAzzawi, W. K. (2022). The role of data mining in diagnosing the diseases: A case study of detecting the thyroid disease. *International Journal Artificial Intelligent and Informatics*, *3*(2), 92–103. https://doi.org/10.33292/ijarlit.v3i2.51

[11]. Rao, A. R., & Renuka, B. S. (2020, November 6). A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis. *2020 IEEE International Conference for Innovation in Technology, INOCON 2020*. https://doi.org/10.1109/INOCON50539.2020.9298252

[12]. Alyas, T., Hamid, M., Alissa, K., Faiz, T., Tabassum, N., & Ahmad, A. (2022). Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach. *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/9809932

[13]. Chaubey, G., Bisen, D., Arjaria, S., & Yadav, V. (2021). Thyroid Disease Prediction Using Machine Learning Approaches. *National Academy Science Letters*, *44*(3), 233–238. https://doi.org/10.1007/s40009-020-00979-z

[14]. Zhou, H., Liu, B., Liu, Y., Huang, Q., & Yan, W. (2022). Ultrasonic Intelligent Diagnosis of Papillary Thyroid Carcinoma Based on Machine Learning. *Journal of Healthcare Engineering*, *2022*. https://doi.org/10.1155/2022/6428796

[15]. Park, K. S., Kim, S. H., Oh, J. H., & Kim, S. Y. (2021). Highly accurate diagnosis of papillary thyroid carcinomas based on personalized pathways coupled with machine learning. *Briefings in Bioinformatics*, *22*(4). https://doi.org/10.1093/bib/bbaa336

[16]. D'Andréa, G., Gal, J., Mandine, L., Dassonville, O., Vandersteen, C., Guevara, N., Castillo, L., Poissonnet, G., Culié, D., Elaldi, R., Sarini, J., Decotte, A., Renaud, C., Vergez, S., Schiappa, R., Chamorey, E., Château, Y., & Bozec, A. (2023). Application of machine learning methods to guide patient management by predicting the risk of malignancy of Bethesda III-V thyroid nodules. *European Journal of Endocrinology*, *188*(3), 249–257. https://doi.org/10.1093/ejendo/lvad017

[17]. Yang, F., Zhang, J., Li, B., Zhao, Z., Liu, Y., Zhao, Z., Jing, S., & Wang, G. (2021). Identification of Potential lncRNAs and miRNAs as Diagnostic Biomarkers for Papillary Thyroid Carcinoma Based on Machine Learning. *International Journal of Endocrinology*, *2021*. https://doi.org/10.1155/2021/3984463

[18]. Radebe, L., van der Kaay, D. C. M., Wasserman, J. D., & Goldenberg, A. (2021). Predicting Malignancy in Pediatric Thyroid Nodules: Early Experience with Machine Learning for Clinical Decision Support. *Journal of Clinical Endocrinology and Metabolism*, *106*(12), E5236–E5246. https://doi.org/10.1210/clinem/dgab435

[19]. Zhou, H., Liu, B., Liu, Y., Huang, Q., & Yan, W. (2022). Ultrasonic Intelligent Diagnosis of Papillary Thyroid Carcinoma Based on Machine Learning. *Journal of Healthcare Engineering*, *2022*. https://doi.org/10.1155/2022/6428796

[20]. Chaganti, R., Rustam, F., de La Torre Díez, I., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques. *Cancers*, *14*(16). https://doi.org/10.3390/cancers14163914

[21]. Abbad Ur Rehman, H., Lin, C. Y., Mushtaq, Z., & Su, S. F. (2021). Performance Analysis of Machine Learning Algorithms for Thyroid Disease. *Arabian Journal for Science and Engineering*, *46*(10), 9437–9449. https://doi.org/10.1007/s13369-020-05206-x

[22]. Duggal, P., & hukla, S. S. (n.d.). *Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques*. https://doi.org/10.1038/s41598-020-62023-w

[23]. Jayamini, W. K. D., & Weerasinghe, K. G. H. D. (2021). Decision Support for Diagnosing Thyroid Diseases using Machine Learning. *International Journal of Scientific and Research Publications (IJSRP)*, *11*(6), 1–5. https://doi.org/10.29322/ijsrp.11.06.2021.p11402

[24]. Veeranki, S. R., & Varshney, M. (2022). Comparative analysis of thyroid disease and predict them using machine learning techniques. *International Journal of Health Sciences*, 11005–11014. https://doi.org/10.53730/ijhs.v6ns3.8459

[25]. Islam, S. S., Haque, M. S., Miah, M. S. U., Sarwar, T. bin, & Nugraha, R. (2022). Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study. *PeerJ Computer Science*, *8*. https://doi.org/10.7717/PEERJ-CS.898

[26]. Abbad Ur Rehman, H., Lin, C. Y., & Mushtaq, Z. (2021). Effective K-Nearest Neighbor Algorithms Performance Analysis of Thyroid Disease. *Journal of the Chinese Institute of Engineers, Transactions of the Chinese Institute of Engineers,Series A*, *44*(1), 77–87. https://doi.org/10.1080/02533839.2020.1831967

[27] Sultana, A., & Islam, R. (2023). Machine learning framework with feature selection approaches for thyroid disease classification and associated risk factors identification. *Journal of Electrical Systems and Information Technology*, *10*(1). https://doi.org/10.1186/s43067-023-00101-5

[28]. Ramya, M., & Kumar, S. (n.d.). PREDICTION AND PROVIDING MEDICATION FOR THYROID DISEASE USING MACHINE LEARNING TECHNIQUE (SVM). In *Turkish Journal of Computer and Mathematics Education* (Vol. 11, Issue 03). **https://10.31838/ecb/2023.12.s3.277**

[29]. Hosseinzadeh, M., Ahmed, O. H., Ghafour, M. Y., Safara, F., hama, H. kamaran, Ali, S., Vo, B., & Chiang, H. sen. (2021). A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things. *Journal of Supercomputing*, *77*(4), 3616–3637. https://doi.org/10.1007/s11227-020-03404-w

[30]. Jayanth Bharadwaj, K., Reddy, N. D., & Srinivas, P. J. (n.d.). *UTILIZING MACHINE LEARNING FOR THE DETECTION OF THYROID DISORDERS*.

[31].. Aversano, L., Bernardi, M. L., Cimitile, M., Iammarino, M., Macchia, P. E., Nettore, I. C., & Verdone, C. (2021). Thyroid disease treatment prediction with machine learning approaches. *Procedia Computer Science*, *192*, 1031–1040. https://doi.org/10.1016/j.procs.2021.08.106

[32]. Mourad, M., Moubayed, S., Dezube, A., Mourad, Y., Park, K., Torreblanca-Zanca, A., Torrecilla, J. S., Cancilla, J. C., & Wang, J. (2020). Machine Learning and Feature Selection Applied to SEER Data to Reliably Assess Thyroid Cancer Prognosis. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-62023-w

[33]. Liu, W. C., Li, Z. Q., Luo, Z. W., Liao, W. J., Liu, Z. L., & Liu, J. M. (2021). Machine learning for the prediction of bone metastasis in patients with newly diagnosed thyroid cancer. *Cancer Medicine*, *10*(8), 2802–2811. https://doi.org/10.1002/cam4.3776

[34]. Hu, M., Asami, C., Iwakura, H., Nakajima, Y., Sema, R., Kikuchi, T., Miyata, T., Sakamaki, K., Kudo, T., Yamada, M., Akamizu, T., & Sakakibara, Y. (2022). Development and preliminary validation of a machine learning system for thyroid dysfunction diagnosis based on routine laboratory tests. *Communications Medicine*, *2*(1). https://doi.org/10.1038/s43856-022-00071-1

[35]. Qiao, D., Liu, H., Zhang, X., Lei, L., Sun, N., Yang, C., Li, G., Guo, M., Zhang, Y., Zhang, K., & Liu, Z. (2022). Exploring the potential of thyroid hormones to predict clinical improvements in depressive patients: A machine learning analysis of the real-world based study. *Journal of Affective Disorders*, *299*, 159–165. https://doi.org/10.1016/j.jad.2021.11.055

[36]. Yadav, D. C., & Pal, S. (2020). Prediction of thyroid disease using decision tree ensemble method. *Human-Intelligent Systems Integration*, *2*(1–4), 89–95. https://doi.org/10.1007/s42454-020-00006-y

[37]. Ayvaz, E., Kaplan, K., Kuncan, F., Ayvaz, E., & Türkoğlu, H. (2022). Reducing Operation Costs of Thyroid Nodules Using Machine Learning Algorithms with Thyroid Nodules Scoring Systems. *Applied Sciences (Switzerland)*, *12*(22). https://doi.org/10.3390/app122211559

[38]. Jayamini, W. K. D., & Weerasinghe, K. G. H. D. (2021). Decision Support for Diagnosing Thyroid Diseases using Machine Learning. *International Journal of Scientific and Research Publications (IJSRP)*, *11*(6), 1–5. https://doi.org/10.29322/ijsrp.11.06.2021.p11402

[39]. Khadijat, *, Abdulwahab, L., & Mustapha, R. (n.d.). HYBRIDIZATION OF MACHINE LEARNING ALGORITHM FOR THE PREDICTION OF HYPOTHYROID. *Science World Journal*, *18*(1), 2023. www.scienceworldjournal.org

[40]. Lim, S. T., Khaw, K. W., Chew, X., & Yeong, W. C. (n.d.). Prediction of Thyroid Disease using Machine Learning Approaches and Featurewiz Selection. *Journal of Telecommunication*.

[41]. Wu, Y., Rao, K., Liu, J., Han, C., Gong, L., Chong, Y., Liu, Z., & Xu, X. (2020). Machine Learning Algorithms for the Prediction of Central Lymph Node Metastasis in Patients With Papillary Thyroid Cancer. *Frontiers in Endocrinology*, *11*. https://doi.org/10.3389/fendo.2020.577537

[42]. Tyagi, A., Mehra, R. and Saxena, A., 2018, December. Interactive thyroid disease prediction system using machine learning technique. In 2018 Fifth international conference on parallel, distributed and grid computing (PDGC) (pp. 689-693). IEEE. https://doi.org/10.1109/PDGC.2018

[43]. Selvathi, D. and Sharnitha, V.S., 2011, July. Thyroid classification and segmentation in ultrasound images using machine learning algorithms. In 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies (pp. 836-841). IEEE.2022,. **10.1109/ICBBE.2009.5163689**

[44]. Saiti, F., Naini, A.A., Shoorehdeli, M.A. and Teshnehlab, M., 2009, June. Thyroid disease diagnosis based on genetic algorithms using PNN and SVM. In 2009 3rd International Conference on Bioinformatics and Biomedical Engineering (pp. 1-4). IEEE,2022, DOI: https://doi.org/10.1109/ICSCCN.2011.6024666

[45]. Chai, X., 2020. Diagnosis method of thyroid disease combining knowledge graph and deep learning. IEEE Access, 8, pp.149787-149795.

[46]. Poudel, P., Illanes, A., Ataide, E. J. G., Esmaeili, N., Balakrishnan, S., & Friebe, M. (2019). Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches. *IEEE Access*, *7*, 79354–79365. https://doi.org/10.1109/ACCESS.2019.2923547

[47]. Ferdous, M., Debnath, J., & Chakraborty, N. R. (n.d.). *Machine Learning Algorithms in Healthcare: A Literature Survey*.

[48]. Naglah, A., Khalifa, F., Khaled, R., Razek, A. A. K. A., & El-Baz, A. (2021). Thyroid cancer computer-aided diagnosis system using mri-based multi-input CNN model. *Proceedings - International Symposium on Biomedical Imaging*, *2021-April*, 1691–1694. https://doi.org/10.1109/ISBI48211.2021.9433841

[49]. Alyas, T., Hamid, M., Alissa, K., Faiz, T., Tabassum, N., & Ahmad, A. (2022). Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach. *BioMed Research International*, *2022*. https://doi.org/10.1155/2022/9809932

[50]. Raisinghani, S., Shamdasani, R., Motwani, M., Bahreja, A., & Raghavan Nair Lalitha, P. (2019). Thyroid prediction using machine learning techniques. *Communications in Computer and Information Science*, *1045*, 140–150. https://doi.org/10.1007/978-981-13-9939-8_13

Authors' Profiles

Fathma Khatun is a committed student now enrolled in the esteemed Bangladesh Army University of Engineering & Technology to pursue a Bachelor of Science in Computer Science & Engineering. She has an immense passion in technology, networking and computer science. Web development is one of her interest areas. Software development, machine learning, and software design

Tonusri Mazumder is a dedicated student currently pursuing her Bachelor of Science in Computer Science &Engineering at the prestigious Bangladesh Army University of Engineering & Technology. She possesses a keen interest in subjects related to computer science, software design, and technology. Her areas of interest are Web Development. Networking, Machine learning and software development

Nadira Anjum Biva is currently a committed student at the esteemed Bangladesh Army University of Engineering & Technology, where she is studying a Bachelor of Science in Computer Science & Engineering. Her interests lie in the fields of technology, software design, and computer science. One of her interests is web development. Software development, networking and machine learning

Tanjima Tabassum is a committed student who is presently enrolled in the esteemed Bangladesh Army University of Engineering & Technology to pursue a Bachelor of Science in Computer Science & Engineering. She has an intense interest in technology, software design, and computer science topics. Web development is one of her interests. Software development, machine learning and networking

Sadia Binte Saleak is a committed student pursuing a Bachelor of Science in Computer Science & Engineering in Bangladesh Army University of Engineering & Technology . She has an intense interest in technology, software design, and computer science topics. Machine Learning is one of her interests. Software development, networking and machine learning