

Optimized Classification and Prediction with Hybrid Machine Learning Models

¹Kalpana K Harish and ²Kuppala Saritha

¹School of Computer Science and Engineering, Presidency University, Bengaluru, Karnataka, India.

²School of Information Science, Presidency University, Bengaluru, Karnataka, India.

¹kalpanaharish1983@gmail.com, ²saritha.mphil@gmail.com

Correspondence should be addressed to Kuppala Saritha : saritha.mphil@gmail.com

Article Info

Journal of Machine and Computing (<https://anapub.co.ke/journals/jmc/jmc.html>)

Doi : <https://doi.org/10.53759/7669/jmc202606016>

Received 17 May 2025; Revised from 16 September 2025; Accepted 13 October 2025.

Available online 21 October 2025.

©2026 The Authors. Published by AnaPub Publications.

This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Abstract – In recent years, there has been an increase in thyroid disorders. Early thyroid illness identification is a crucial endeavor because of its significance in metabolism. Although there are numerous studies on the identification of whether thyroid illness leads to cancer or not, in order to properly address and handle this condition, a timely and accurate medical diagnosis is crucial. Thyroid cancer detection and diagnosis could be enhanced by algorithms that utilize machine learning, which have achieved considerable amounts of interest especially in the medical field. This paper describes how machine learning techniques are being used to diagnose thyroid cancer. K-Nearest Neighbor, Decision Tree, Random Forest, Ada-Boost with decision tree classifier, Gradient Boosting Classifier, Stochastic Gradient Boosting classifier, Extended Gradient Boosting Classifier, Extended Gradient Boosting Classifier with hyper parameter tuning, Extra Tree Classifier, Light GBM Classifier, Voting Classifier are among the machine learning approaches that are used and evaluated to assess how well they diagnose thyroid Cancer. The study evaluates the methods precision, recall, F1-score, and Accuracy. From all the above machine learning models Random Forest, Ada Boost, Extended gradient boosting classifier, Extra Tree Classifier, Light Gradient Boosting Machine outperforms the other models.

Keywords – Thyroid Detection, Machine Learning, Label Encoder, Voting Classifier.

I. INTRODUCTION

The body of an individual uses the thyroid, which is very tiny gland but essential hormone in the cervical region, to regulate activities of body such as heartbeat, digestion etc., [1]. Hormones that regulate metabolism, including body temperature along with heart rate, are released by the thyroid gland. It generates the vital hormones thyroxine(T4) as well as triiodothyronine(T3). These hormones are in charge of a number of metabolic processes, including heartbeat and weight management. Thyroid hormones, which are produced by the gland responsible for thyroid function, help to monitor several kinds of organs through the blood.

Thyroid illness symptoms may also involve higher Blood Pressure (BP), irregular pulse rates, along with high cholesterol. Thyroid disease is commonly divided into five categories: thyroid tumors, hypothyroidism, hyperthyroidism, structural abnormalities, and subclinical hypothyroidism or hyperthyroidism. Thyroid tumor is an unusual progress of cells that origins in the thyroid follicle. Symptoms of thyroid tumor cannot be recognized initially. Symptoms such as difficulty in ingesting, voice changes and bulge in the neck can be sensed in the later stages. It can be addressed by using imaging tests such as Computed Tomography (CT) and Magnetic Resonance imaging (MRI).

Hypothyroidism can be identified by medical professionals through examining the test results of body organs using various parameters. Thyroid hormone production is insufficient in hypothyroidism. It can cause a number of symptoms, including depression, constipation, increased weight, a slower heartbeat, tiredness, as well as reduced capacity to handle colds. In order to regulate thyroxine and also Thyroid Stimulating Hormone (TSH) levels, physicians adjust the medication dosages based on their patient's conditions.

Medical practitioners can diagnose hyperthyroidism by looking at the organ test results through a variety of parameters. Hyperthyroidism stimulates thyroid by generating excessive hormones causes anxiety, irritability, anger, fine crumbly hair, excessive sweating, hand tremors, rapid heartbeat, worry, sensitive skin plus muscle weakness.

The rate of development of hyperthyroidism may increase at an older age. Hyperthyroidism can be addressed by using medicines, radiation therapy, thyroid surgery etc.,

Structural abnormalities can lead to physical abnormalities in a fetus development or chromosome structure. Physical abnormalities may appear in fetus development such as heart defects, abnormal limbs, neuro defects etc., which leads to

birth inadequacy. This can be detected during the first three months of pregnancy through scanning. It can be addressed by using medicines, therapy, thyroid surgery etc., Subclinical hypothyroidism is a condition where free thyroxine levels are normal but the TSH level in the blood is elevated. It can cause a number of symptoms such as fatigue, unexpected weight gain, constipation, depression etc., which can be detected by regular blood tests. It can be treated by medication, iodine in diet.

An imbalance associated with thyroid glands causes either too much or too little hormone production, as a result thyroid illness in the body of humans. There are several thyroid-related diseases which trigger imbalances and malfunctions in the remaining organs and complications like enlarged thyroid, cardiovascular disease, pregnancy issues, and serious 219myxedema coma [2] [3]. Currently there are 200 million individuals worldwide suffering from thyroid disorder. Further forty percent of individuals worldwide are prone to insufficient TSH level and lack of iodine [4]. All the types of thyroid illnesses can be cured by early diagnosis of the appropriate thyroid illness. Further, it can be identified using various tests namely lithium, goiter, hypopituitary, Psych, TSH, T3, TT4, T4U, FTI, and TBG. These tests are essential for preventing complications and deterioration from a patient's health [5]. Similarly, thyroid illness can also be classified through numerous characteristics and symptoms namely weight gain, weight loss, unusual growth in the neck etc., [6].

The majority of existing methods employ data analysis on publicly available thyroid datasets. Further more methods like symptoms of thyroid namely voice changing, enlargement of thyroid gland, tired ness and laboratory tests like blood test, ultra-scan images, are applied to predict the thyroid cancer. These datasets are predicting inaccurate results using machine learning algorithms. In order to diagnose and forecast thyroid disease, the current research highlights the use of a variety of machine learning techniques, highlighting the significance of selecting features, data accuracy, and validation of models.

II. LITERATURE SURVEY

In [7] outlined many AI network types to identify the feature selection algorithm to classify the thyroid disease, including Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Adaptive Boosting (AB), and Gradient Boosting (GB). Each of these techniques was selected for its own unique characteristics. In comparison to Boruta, Synthetic Minority Oversampling Technique (SMOTE), and Recursive Feature Elimination (RFE), the feature selection approach Least Absolute Shrinkage and Selection Operator (LASSO) performed better with 97% Precision, 92% Recall, and 95% F1-Score. The University of California Irvine (UCI) dataset, which has 2800 occurrences and 28 categorical attributes, was employed in the study.

The [8] proposed variety of approaches, each chosen for their own qualities to predict the treatment for the thyroid disease, including AB, CatBoost (CB), Extended Gradient Boosting Classifier (XGBC), GB, Extra Tree Classifier (ETC), RF, DT, KNN, and Multilayer Perceptron Classifier (MLP). The ETC performed better, with accuracy of 84%, F1 of 84%, Precision of 85%, and recall of 84% by applying the SMOTE feature extraction technique. The primary constraint of this research pertains to the dataset's quality.

The authors [9] proposed a different methodology, each having its own quality to early predicting thyroid disease, including SVM, Neural Network (NN), KNN, DT, Logistic Regression (LR) and ensemble classifiers. The proposed model achieved 99% accuracy, 99.59% specificity and 99.39% sensitivity by using 3 feature selection methods such as SMOTE, Bayesian Optimization and Ensemble learning. The University of California Irvine (UCI) dataset, which has 3672 occurrences and 29 categorical or real attributes, was employed in the study.

In [10] proposed a variety of methods, each of the method is having its own characteristics to detect thyroid such as SVM, DT, LR, KNN, Artificial Neural Networks (ANN). They created WebApp for prediction of thyroid disease. The LR achieved a best accuracy of 96.92% compared to other methods. The University of California Irvine (UCI) dataset, which has 215 samples and 5 attributes, was employed in the study. They also suggested image dataset can be used for prediction of thyroid.

The [11] discussed a variety of algorithms have its own qualities how thyroid can be diagnosed, including Multi Layer Perceptron (MP), Back Propagation Neural Network (BPNN), SVM and Extreme Learning Machine (ELM). They used different types of feature selection techniques such as Principal Component Analysis (PCA), Filter Based Approach (FBA) and Wrapper Based Approach (WBP). The ELM with Recursive Feature Elimination (RFE) outperformed compared to other with 98.14% accuracy. The UCI dataset that contains 215 samples and five attributes was used in their study.

The [12] proposed an algorithm to identify whether the thyroxine level affect cardiac failure. They used limited number of patients' dataset and UCI hyperthyroidism dataset. This study highlights the cardiovascular implications on thyroid disorder. They recognized that atrial fibrillation was higher among men compared to women. In future, the model can be applied on the patients with disorders such as hypothyroidism, breathlessness, chest pain etc.,

In [13] outlined a various kinds of ML approaches have its own characteristics to predict thyroid, including SVM, KNN, ANN and DT. The SVM outperformed with an accuracy of 99.63. The UCI dataset used by them consists eight attributes. They also proposed that in future, a smaller number of features can be used for detecting thyroid disease.

Research work [14] suggested a different kind of approaches, each chosen for its unique qualities to predict the thyroid, such as SVM, Bagging, Boosting, J48 and Naïve Bayes. The SVM with pathological observation achieved better accuracy of 99.08%. Their dataset collected from 1464 patients consisting of 21 attributes. Furthermore, they also suggested that this model can be improved by adding more instances to the dataset.

The [15] proposed a variety of ML approaches, each chosen for their own qualities to classify the thyroid diseases, including SVM, RF, DT, Naïve Bayes, LR, KNN and Multi-Layer Perceptron (MLP). The RF achieved a better accuracy of 98.93%. They collected data from external hospitals of Iraq contains 1250 instances with 17 attributes.

In [16] proposed a method to detect thyroid disease using Adaptive Boosting (AB) with feature selection method Differential Evaluation based optimization Algorithm, achieved a best accuracy of 99.8%. They used Conditional Generative Adversarial Networks (CTGAN) for data augmentation. The CTGAN mitigates the probability of Overfitting, model bias problems and helps to balance the total number of samples. The author identified that the ML model shows best results using data augmentation.

The comparison details of some recent works accuracies are represented in the form table and chart is shown in **Table 1**, **Fig 1** respectively. We can learn about their finding the accuracy is accomplished by them and the algorithms they used in their research work.

III. METHODOLOGY

This paper proposes a hybrid ML approach for the diagnosis of thyroid cancer (TC). The suggested framework is depicted in **Fig 2**.

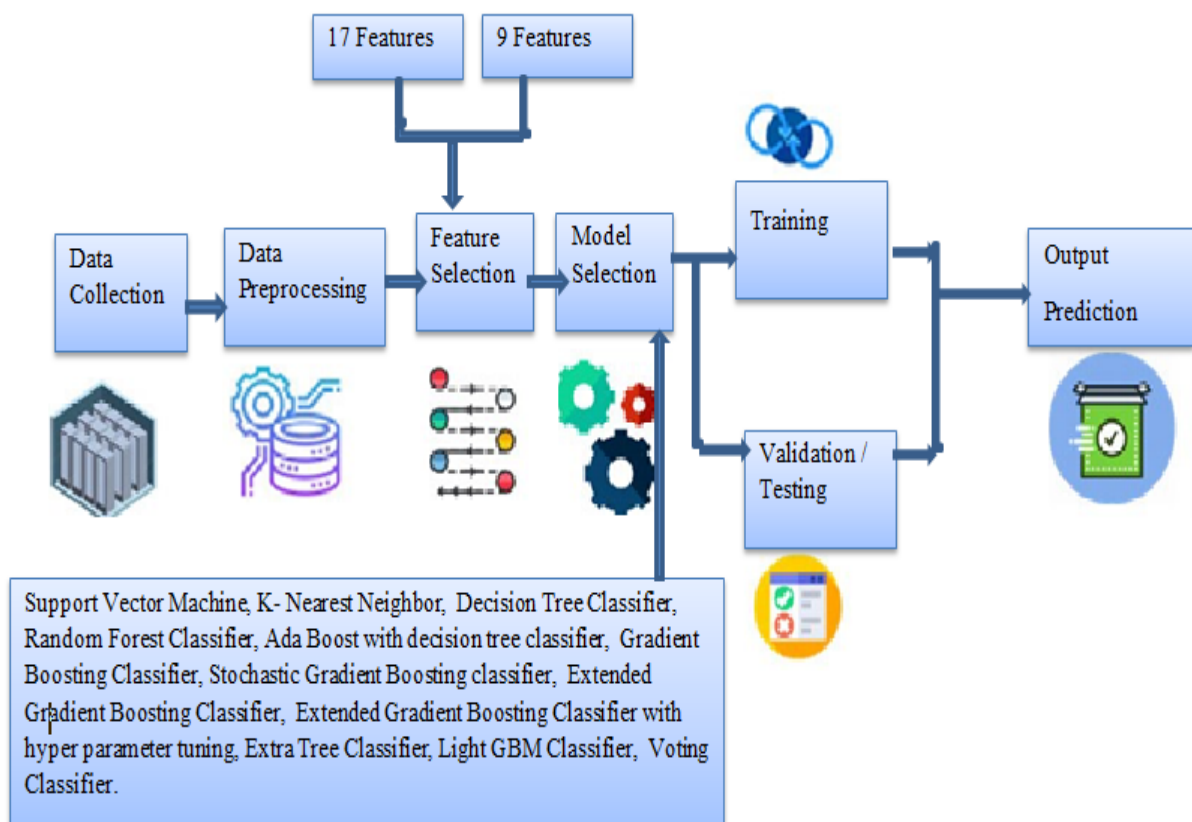


Fig 2. Proposed Architecture.

Data Collection

We obtained the dataset from Kaggle repository. Which is having 17 features, such as “Age”, “Gender”, “Smoking”, “Hx Smoking” etc., shown in **Table 2**.

Data Preprocessing

The Dataset is divided into 2 sets called test and train set with the ratio of 25% and 75% respectively used in Train Test Split model, Further, 90% data used for training, 10% data used for testing in 10-Fold Cross Validation Model and Stratified 10-Fold Cross Validation respectively.

To analyze data we used, preprocessing techniques such as feature selection, handling missing values and class encoding. We handled missing values by using the mode method. Label encoder is used to encode the data into numerical representation. This is straight forward and it turns each column value into a number. Consider a dataset of thyroid with following column “Gender”: male and female, encoded the text values by placing a running sequence for each text value: 0 and 1.

Feature Selection and Feature Importance

Generally, all the features in a dataset are not having same impact, some of them have more significance and some of them have less significance. It’s required to find the appropriate significance for the features. To find the appropriate significance

for the features from the dataset, we used method ANNOVA F-test. This test finds the relationship between the target class and conditional features. Further, this test also generates score such as high, low, which helps to find the more significance and less significance features from the given set of features.

Table 1. Comparison Table of Previous Work

Reference	[7]	[8]	[9]	[10]	[11]	[13]	[14]	[15]	[16]
RF	99%	94%	-	-	-	-	-	98.93%	98.80%
DT	99%	-	95.62%	92.30%	-	75.76%	-	98.40%	-
KNN	96%	-	91.76%	93.84%	-	98.62%	-	90.93%	-
AB	98%	93%	-	-	-	-	-	-	91%
GBM	97%	93%	-	-	-	-	-	-	98.90%
SVM	94%	90%	88.01%	95.38%	97%	99.63%	96.41%	92.27%	89.90%
LR	-	84%	87.77%	96.92%	-	-	-	91.47%	64.30%
Bagging	-	-	98.15%	-	-	-	98.29%	-	-
Boosting	-	-	99.15%	-	-	-	98.56%	-	-
ANN	-	-	94.86%	75.30%	-	97.50%	-	-	-
MLP	-	-	-	-	94%	-	-	97.60%	-
BPNN	-	-	-	-	95%	-	-	-	-
ELM	-	-	-	-	98.14%	-	-	-	-
J48	-	-	-	-	-	-	97.59%	-	-
Naïve	-	-	-	-	-	-	97.61%	81.33%	-
LDA	-	-	-	-	-	-	-	83.20%	-
LSVM	-	-	-	-	-	-	-	-	-
LGBM	-	-	-	-	-	-	-	-	-
CNN	-	-	-	-	-	-	-	-	-
LSTM	-	-	-	-	-	-	-	-	-
QSVM	-	-	-	-	-	-	-	-	-
ETC	-	-	-	-	-	-	-	-	-
XGBC	-	-	-	-	-	-	-	-	-
CB	-	-	-	-	-	-	-	-	-
ClymphN	-	-	-	-	-	-	-	-	-

Further, we used correlation method to find the association among the more significance features. The score of correlation high indicates more association between the two features and the score of less indicates less association between the features. For clear understanding we have shown the association (see **Fig 3**).

The score of 0.9-1.0 indicates very high association between the features, similarly 0.7-0.89 score represents high association, 0.5-0.69 score represents moderate association, 0.3-0.49 score represents low association, 0.0-0.29 score negligible association and score less than zero represents negative association.

The attained scores shown in **Table 1** are obtained by using select K-best and F-statistic process. The significance of features considered in this study is having high significance. The top five attained features are: “Risk”, “Response”, “N”, “T”, “Stage”. These features play a crucial role in diagnosing thyroid cancer.

In Future, these five features can be used to diagnose thyroid cancer.

$$F_{static} = \frac{\text{Variance among batches}}{\text{Variance inside batches}} = \frac{\text{Mean Square among}}{\text{Mean Square inside}} \quad (1)$$

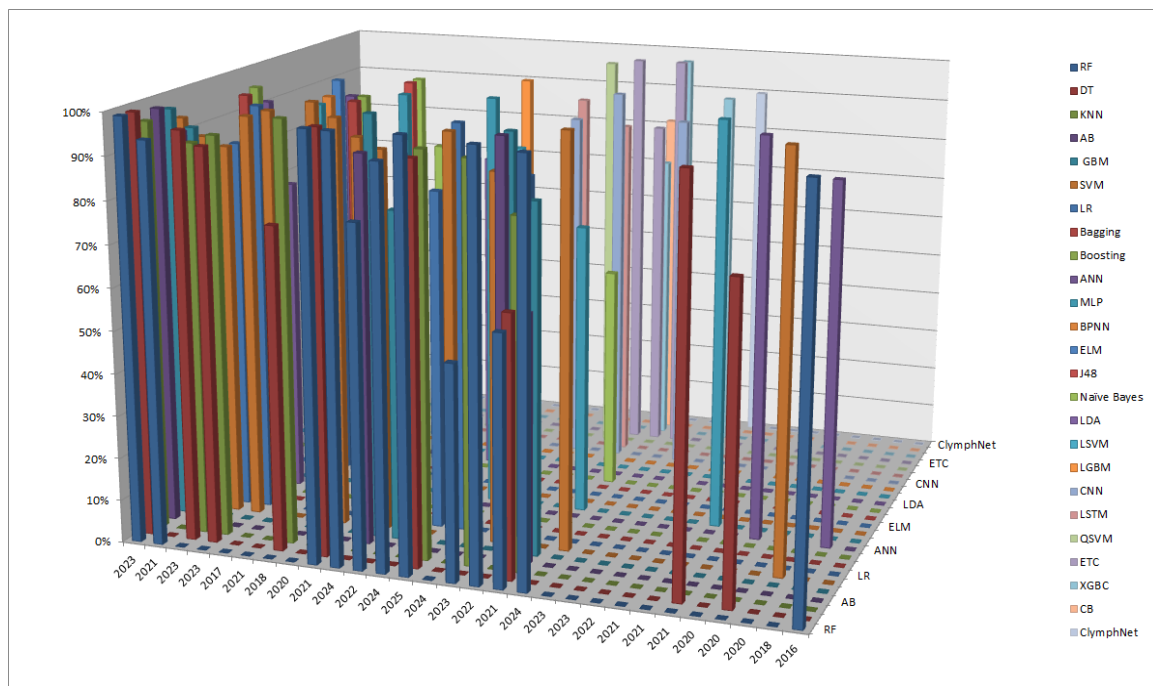


Fig 4. Comparison Chart of The Previous Work.

Table 2. Feature Importance

Sl. No.	Feature	Type	Normal range	Values	Attained
1	Risk	Categorical	-	Low (2), Intermediate (1), High (0)	372.35
2	Response	Categorical	Thyroglobulin (Tg) Tg < 1ng/ML	Excellent (1), Biochemical Incomplete (0), Structural Incomplete (3), Indeterminate (2)	264.69
3	N	Categorical	-	N0(0), N1a (1), N1b (2)	184.84
4	T	Categorical	3-40 ng/ML < 0.5mIU/L	T1a (0), T1b (1), T2(2), T3a (3), T3b (4), T4a (5), T4b (6)	171.94
5	Stage	Categorical	-	I (0), II (1), III (2), IVA (3), IVB (4)	94.61
6	Focality	Categorical	-	Uni focal (0), Multi focal (1)	60.61
7	M	Categorical	-	M0(0), M1(1)	54.33
8	Smoking	Categorical	-	Yes (1), No (0)	49.26
9	Gender	Categorical	-	F(0), M (1)	40.24
10	Age	Numerical	-	15(0),17(1).... 82(64)	25.54
11	HX Radiotherapy	Categorical	-	Yes (1), No (0)	12.11
12	Physical Examination	Categorical	-	Diffuse goiter (0), Multinodular goiter (1), Normal (2), Single nodular goiter- left (3) Single nodular goiter –right (4)	10.26
13	Adenopathy	Categorical	-	Yes (1), No (0)	9.60
14	HX smoking	Categorical	-	Yes (1), No (0)	7.76
15	Thyroid function	Categorical	3-40ng/ML 0.5-5.0mIU/L	Clinical hypothyroid (1), Clinical hyperthyroid (0), Subclinical hyperthyroid (3), Subclinical hypothyroid (4), Euthyroid (2)	1.56
16	Pathology	Categorical	-	Micropapillary (2), Follicular (0), Papillary (3), Hurthel Cell (1)	0.55
17	Recurred	Categorical	-	Yes (1), No (0)	--

In Fig 3, The “Age” has a negligible association with “Gender”, “Hx Smoking”, ”Thyroid Function”, “Response”, a negative association with “Adenopathy”, “Pathology”, ”Risk”, “Response” and a low association with “Recurred”. Similarly, “Gender” has a negligible association with “Hx Smoking”, ”Thyroid Function”, “Response”, a negative association with “Adenopathy”,

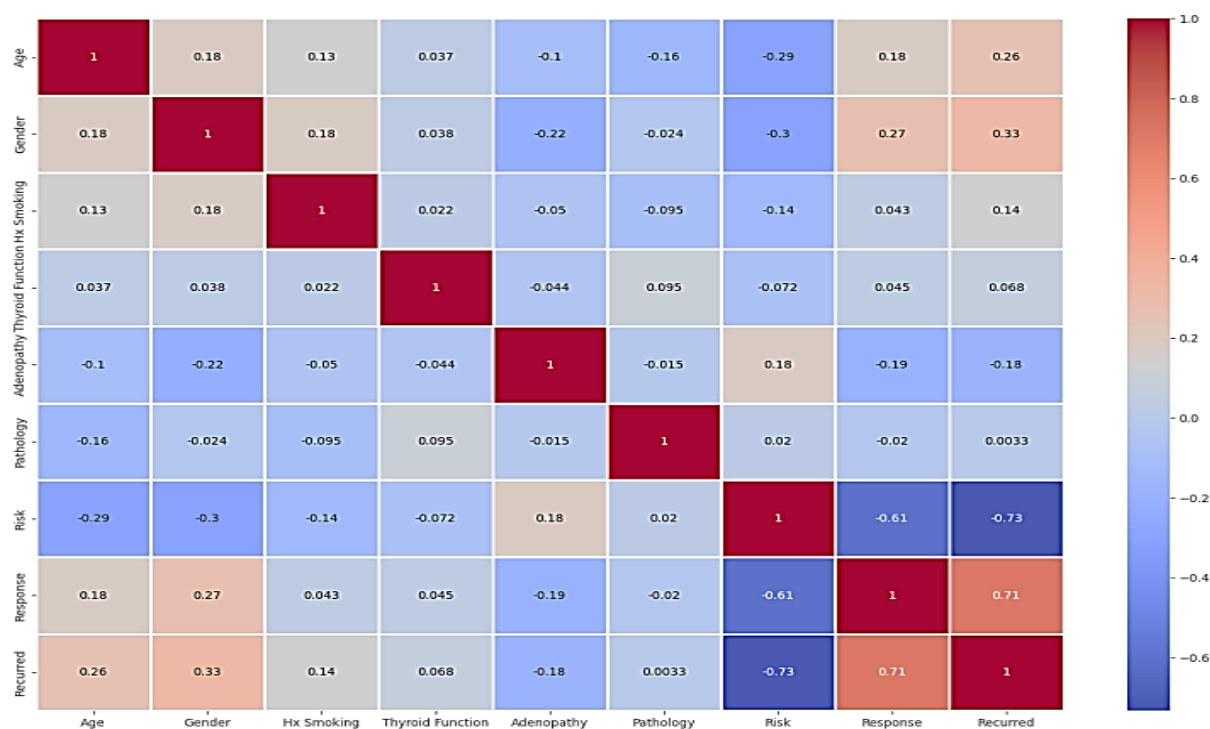


Fig 3. Correlation Heat Map of Highly Correlated Features.

“Pathology”, “Risk”, and a low association with “Recurred”. Likewise, “Hx Smoking” has a negligible association with “Thyroid Function”, “Response”, “Recurred” and a negative association with “Adenopathy”, “Pathology” and “Risk”. In the same way “Thyroid Function” has a negative association with “Adenopathy”, “Risk”, and a negligible association with “Pathology”, “Response”, “Recurred”. Further, “Adenopathy” has a negative association with “Pathology”, “Response”, “Recurred”, and a negligible association with “Risk”.

The feature “Pathology” has a negligible association with “Risk”, “Recurred”, and a negative association with “Response”, and “Risk” has a negative association with “Response”, “Recurred”. “Response” has a High association with “Recurred”.

Model Selection

We collected dataset of thyroid for recognizing cancer is a type of categorical. Generally, there are various models to process categorical data, but we don't have a knowledge which model gives good accuracy. We plan to apply various individual models namely Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF), Gradient Boosting (GB), Stochastic Gradient Boosting (SGB), Extended Gradient Boosting (XGB), Extra Tree Classifier (ETC), Light Gradient Boosting Machine (LGBM). Further we planned to use hybrid models namely Ada Boost with Decision Trees (AB) and Voting Classifier (VC) with SVM, DT, KNN, RF, GB, SGB, XGB, ETC, LGBM, AB.

Again we simulated the above model using 2 different techniques called without dimensionality reduction and with dimensionality reduction. Without dimensionality reduction, we have used 17 features such as “Age”, “Gender”, “Hx Smoking”, “Thyroid Function”, “Adenopathy”, “Pathology”, “Risk”, “Response”, “N”, “T”, “Stage”, “Focality”, “M”, “Smoking”, “Hx Radiopathy”, Physical Examination”, “Recurred”. With dimensionality reduction, we have extracted 9 distinguishing features based on the attained scores and significance such as “Age”, “Gender”, “Hx Smoking”, “Thyroid Function”, “Adenopathy”, “Pathology”, “Risk”, “Response”, “Recurred”.

Random Forest (RF)

RF is defined as an ensemble learner due to its ability to generate a significant amount of classifiers and incorporate their outputs [7]. It applies several DTs to various segments of the given dataset and determines the mean in order to improve the dataset's capacity of predicting results. When there are more trees in the forest, the overfitting problem is avoided and

higher accuracy is achieved. Some DTs may yield the right result whereas others may produce incorrect result, since the methods rely on a combination of trees to determine the class of the dataset. **Fig 4** Show the schematic structure of the Random Forest method.

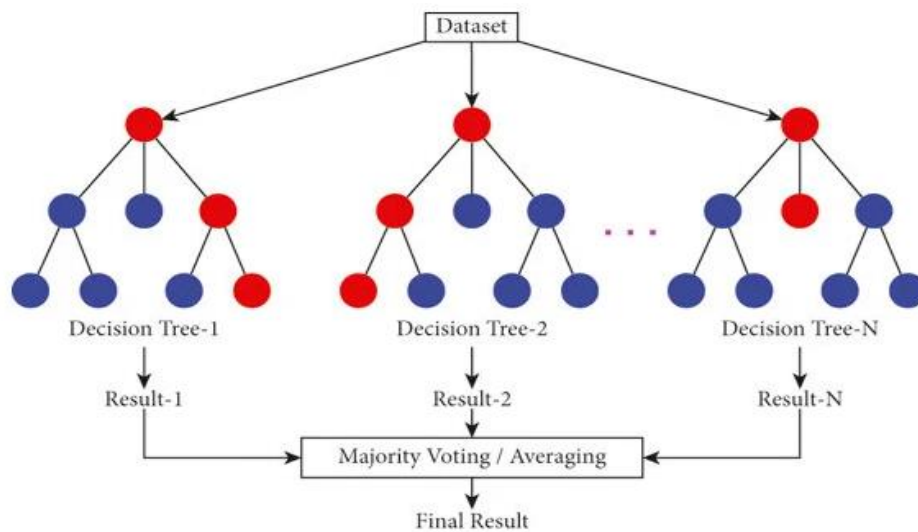


Fig 4. Random Forest Algorithm.

Decision Tree (DT)

The decision-making concept governs DT's operations. Creating hypotheses with respect to the target class by adopting a decision rule based on previous data is the main objective of applying DT [7]. It has precise and reliable tree structure. It is utilizing the multistage method; this is used to divide the difficult decision into smaller decisions and further all the smaller decisions will be combined to get the desired solution. In DT, each branch is defined by data-splitting, it will start from the root node and end by leaf node. It also has internal nodes for the dataset features, branches for decision making and leaf nodes for the desired results. **Fig 5** shows the schematic diagram of Decision Tree.

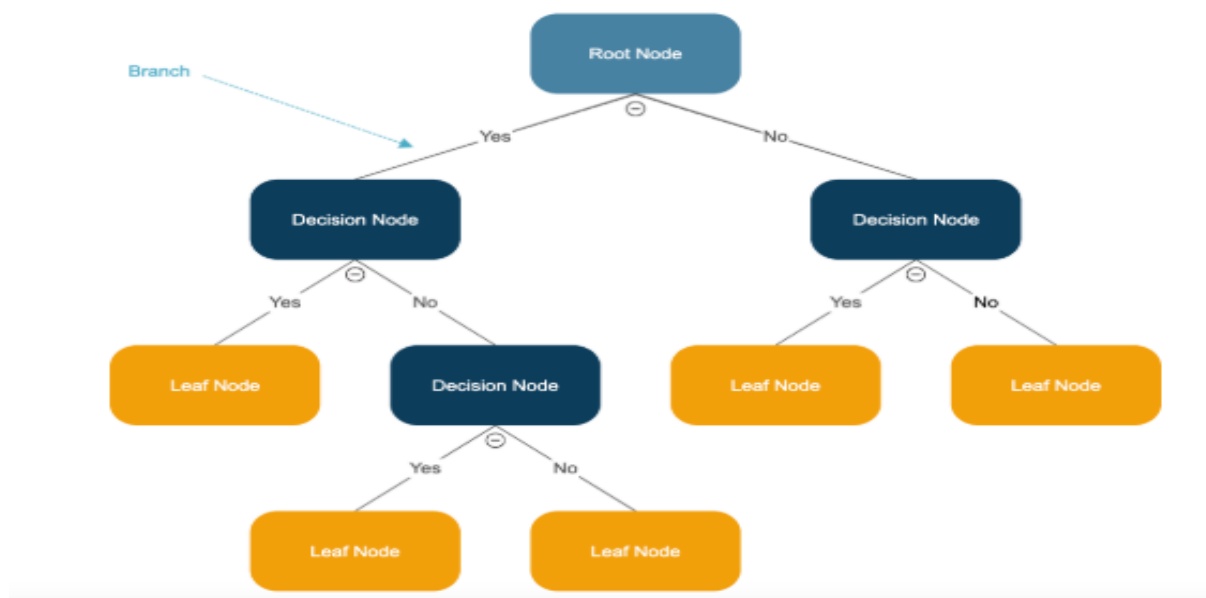


Fig 5. Decision Tree Algorithm.

Support Vector Machine (SVM)

SVM is a method that analyses the data and groups into two categories. It creates an arrangement of data with the largest possible margins among the two. Data points that fall in one side of the imaginary line shall be placed in one group, and those that fall on opposite side will be placed into other group [7]. The maximum distance between the two different points is referred to as hyperplane, this is the decision boundary used to classify the data points. The vectors near to hyperplane that determines its location. **Fig 6** shows the schematic diagram of SVM.

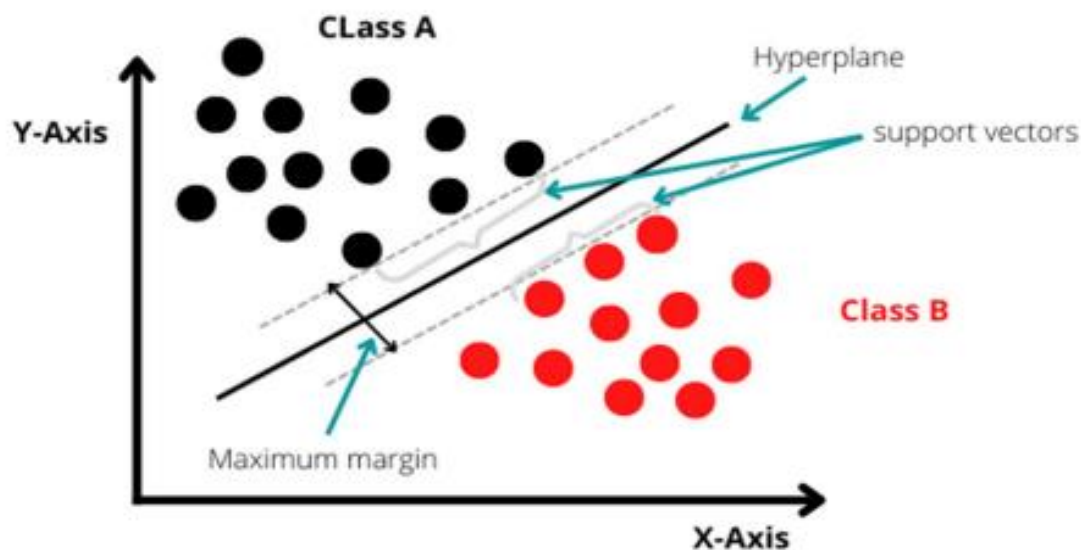


Fig 6. SVM Algorithm.

K-Nearest Neighbors (KNN)

By calculating the likelihood of the degree to which every category fits with the training data, the model using KNN identifies which of "K" possible groups that the testing data significantly matches [7]. In order to forecast a suitable class for the testing data, KNN looks at the difference between the test data as well as training records. K is the amount of neighbors [7]. To find the best K there is no pre-defined method, when you select small value for K it results in producing unstable decisions. In order to avoid ties in categorization, often recommended to select an odd number for K. **Fig 7** shows the schematic diagram of KNN. Here, we can notice class B and class A instances are represented by two colors. There is one instance in orange color, which is new data; it should be classified by KNN.

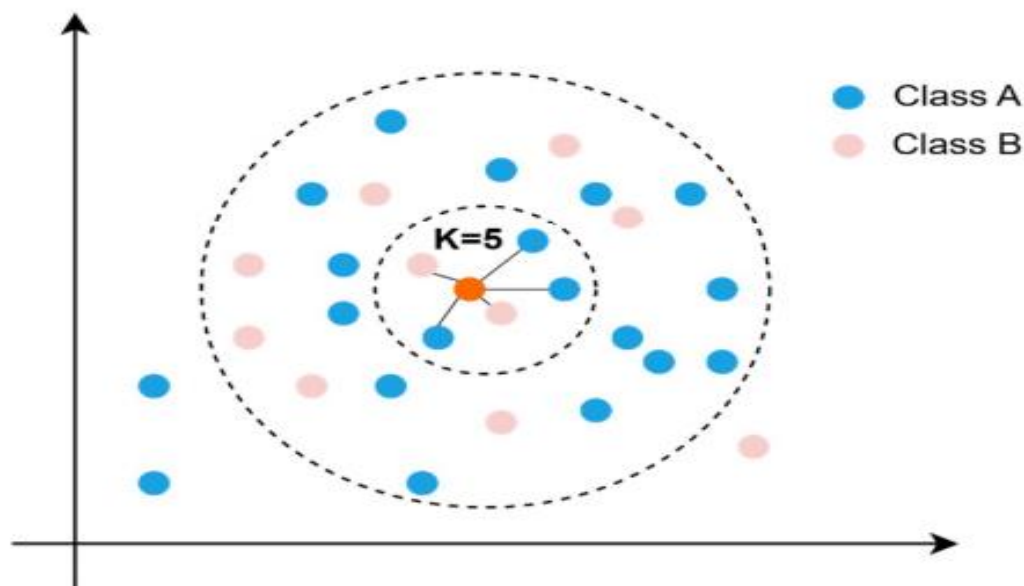


Fig 7. KNN Algorithm.

Adaptive Boosting (AB)

Any machine learning (ML) algorithm can be made more effective by using AB. It generates good results for slow learners. DT is probably the most commonly used approach for AB, since it is considered to be the most efficient for a single level. Initially, this algorithm uses the original dataset for classification; gradually it utilizes all instances, in order to fix the flaw in the previous instance [7]. The classification model is replicated, and each record or instance is trained on distinctive dataset. By providing weights to discrete data points, some subsets of dataset could be generated. Then, strong classifier is produced by joining weak classifiers. Finally, classifiers obtained with the more accuracy are assigned with higher weight.

Our frameworks training parameters are as follows: n-estimators of 200, 0.001 learning rate, SAMME method, and decision- tree classifier in place of base estimator.

Gradient Boosting (GB)

GB is an excellent method for ML which represents decision tree for complicated datasets. To produce strong model, it combines some weakly predicted models. The GB contains 3 main components such as utilization of poor predictors, loss function and additive model, that adds trees by utilizing gradient descent function. It finds the best model by combining preceding model with the next model. GB method used for improving accuracy by removing overfitting. It prevents degeneration by performing proper fitting process. The GB rounds have positive correlation and reduction of an error. The best learning rate of our model is 0.1.

Stochastic Gradient Boosting (SGB)

A machine learning optimization method called stochastic gradient descent (SGD) is used to reduce mistakes in prediction models. SGD changes the parameters utilizing one point of data at a time, in contrast to traditional gradient descent; this will compute the gradient & modifies the model attributes by using the complete dataset. SGD assists in determining the unique model parameters to increase prediction accuracy. Initially, it starts with poor learner, which forecasts an initial output depends on the train dataset. Later subsampling can be done by creating subset for training data and features. Errors can be generated by comparing with predicted results and actual values and then find train the new learner on this subsampled data by focusing on the previous iteration. Lastly combine all the predictions by using all the base learners to make latest prediction. These steps should be repeated till you get stopping criteria.

Extreme Gradient Boosting (XGB)

XGB is a machine learning approach that uses gradient developed decision trees. This is one of the more recent community learning methods. XGB solves the problem of overfitting during the training. Hence the accuracy improved. It consists of two objective functions namely, loss function and regularization. Loss function determines the discrepancy between actual value and each predicted value. The regularization eliminates overfitting and maintains the complexity of the model. It uses L1, L2 regularization that helps a model to avoid overfitting. It uses second partial derivative in loss function. It is faster compared to other gradient boosting because it constructs the trees parallel. It also handles the missing values.

Light Gradient Boosting Machine (LGBM)

LGBM is an open source framework, developed by Microsoft. It is an ensemble learning model. By adding poor learners it constructs strong learner in descent manner. It is scalable, efficient and produces high accuracy. It will use DT for minimizing the usage of memory and improve the training time. It is fastest compared to other gradient boosters. It uses L1 and L2 regularization for solving overfitting problem.

Extra Tree Classifier

ETC is an ensemble learning function. It will generate large number of trees, but each tree is generated randomly and is diverse in nature. It merges the predictions of all these trees to get the better accuracy and resolve the overfitting problem. It selects the split point randomly unlike random forest. This can be used for finding the most influential attributes in the dataset. It will choose the split point randomly. Final prediction can be done by combining the predictions of all the individual trees. It consumes less time for training and is used for extracting important features.

Voting Classifier

A VC is a ML model that learns from several models and finds the output based on the class with highest likelihood. It displays the output based on the majority votes; It performs average of each classifier given as an input to VC. It is used to build only one model from several models and predicts output by using majority votes. There are two types of voting classifiers namely soft voting and hard voting. In hard voting the prediction can be done based on the highest majority votes. Similarly, in soft voting highest average voting probability can be considered for predicting the final prediction.

IV. RESULTS AND DISCUSSION

The evaluations were carried out on a computer using Windows 10 & including a Core i5 and 64 GB of RAM. The coding environment is Python 3, and packages including scikit-learn, pandas, NumPy, matplotlib, Collections, seaborn, and time were utilized. Without the usage of GPU acceleration, Google Colab was used for both code execution and model training.

We have compared our study with some available studies using similar datasets. Our model outperformed compared to the existing works (see **Table 1, Fig 1**). In future, we will try to explore the performance of the methods for additional datasets. On the other hand, to get better accuracy, we have applied Train-test-split, 10 fold cross validation, Stratified 10 Fold cross validation and stratified 10 Fold with hyper parameter tuning. The results of without dimensionality reduction and with dimensionality reduction are shown in the **Table 3, Fig 8** and **Table 4, Fig 9** respectively.

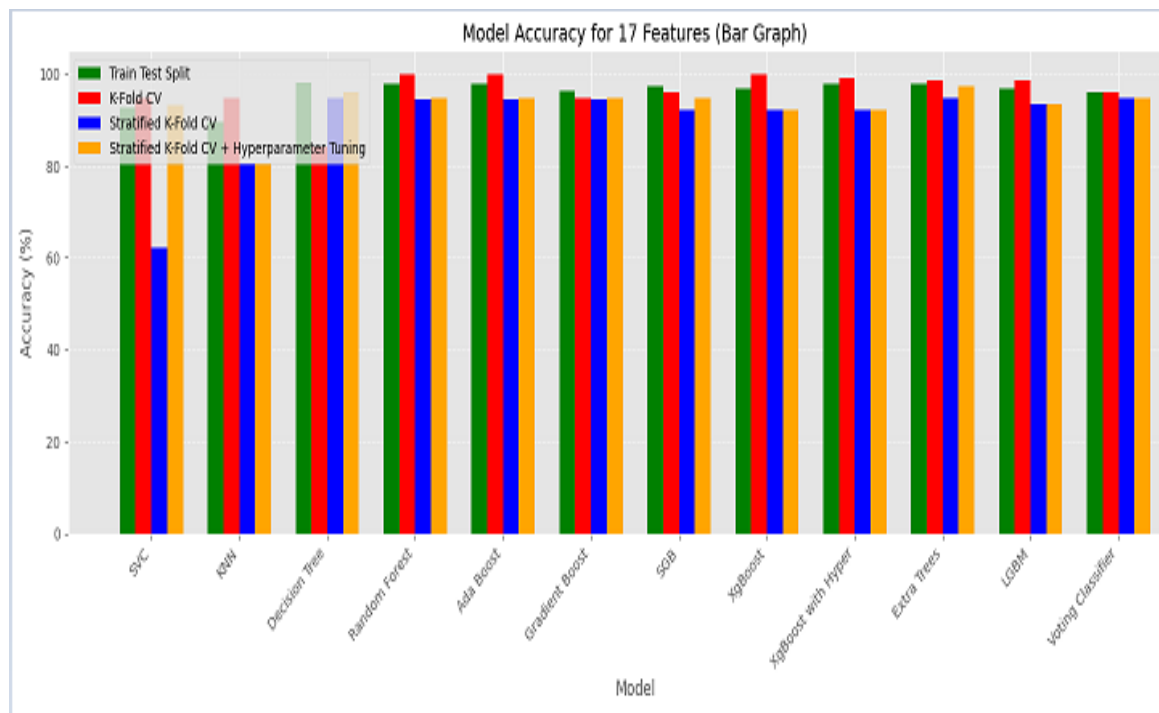


Fig 8. Chart for without Dimensionality Reduction.

In Without dimensionality Reduction, we have applied 12 different hybrid models to get best accuracy, out of 12 models four models produced best results namely RF, AB, XGB and XGB with hyper parameter tuning and L GBM with 10 Fold Cross validation are outperformed with an accuracy of 100%,100%,100% and 99.02% respectively.

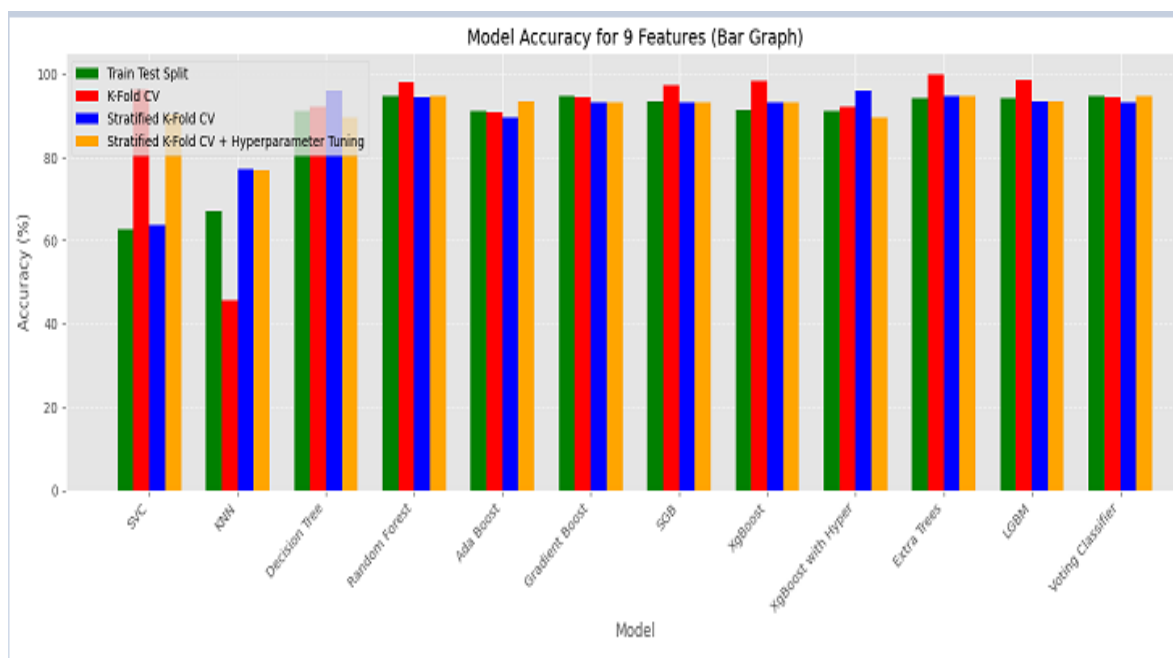


Fig 9. Chart for with Dimensionality Reduction.

Similarly, in case of with dimensionality reduction technique, again 12 models applied on the dataset for checking the accuracy, the best four models which produced the best accuracy are RF, XGB, ETC and LGBM with 10 Fold Cross Validation outperformed with an accuracy of 98.11%, 98.39%, 100% and 98.68% respectively.

Time Taken to Execute Each Algorithm

Execution time for each algorithm for without dimensionality reduction and with dimensionality reduction is shown in **Fig 6** and **Fig 7** respectively.

Time Taken in Without Dimensionality Reduction

Fig 10 illustrates the time complexities among the different models. KNN consumed less execution time and Voting Classifier consumed more execution time with 0.0023 seconds and 0.6459 seconds respectively compared to all other models by using Train test split. Similarly, KNN consumed less execution time and Random Forest consumed more execution time with 0.0143 seconds and 1.6457 seconds respectively compared to all other models by using K-Fold CV. Likewise, KNN consumed less execution time and Voting Classifier consumed more execution time with 0.0067 seconds and 1.3523 seconds respectively compared to all other models by using Stratified K-Fold CV. Similarly, KNN consumed less execution time and Voting Classifier consumed more execution time with 0.0052 seconds and 1.2029 seconds respectively compared to all other models by using Stratified K-Fold CV with dimensionality reduction.

Time Taken in with Dimensionality Reduction

Fig 11 illustrates the time complexities among the different models. The Ada Boost consumed less execution time and Voting Classifier consumed more execution time with 0.0110 seconds and 0.71 seconds respectively, compared to all other models. Similarly, DT consumed less execution time and Random Forest consumed more execution time with 0.0037 seconds and 1.6415 seconds respectively compared to all other models by using K-Fold CV. Likewise, DT consumed less execution time and Voting Classifier consumed more execution time with 0.0015 seconds and 1.4366 seconds respectively compared to all other models by using Stratified K-Fold CV. Similarly, the KNN consumed less execution time and Voting Classifier consumed more execution time with 0.0038 seconds and 1.5482 seconds respectively compared to all other models by using Stratified K-Fold CV with dimensionality reduction.

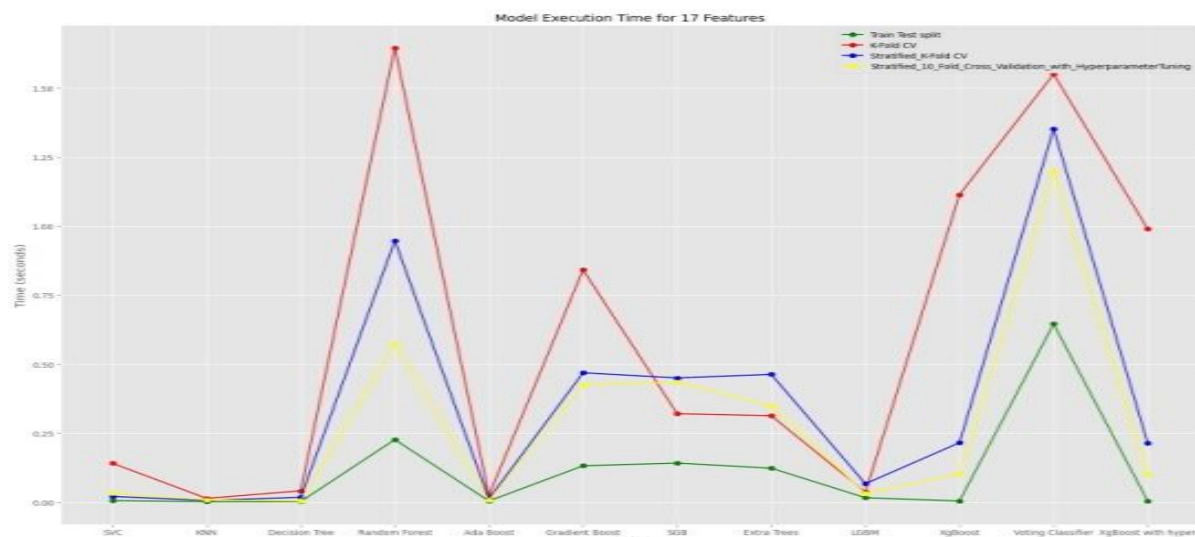


Fig 10. Time Complexity Graph for 17 Features.

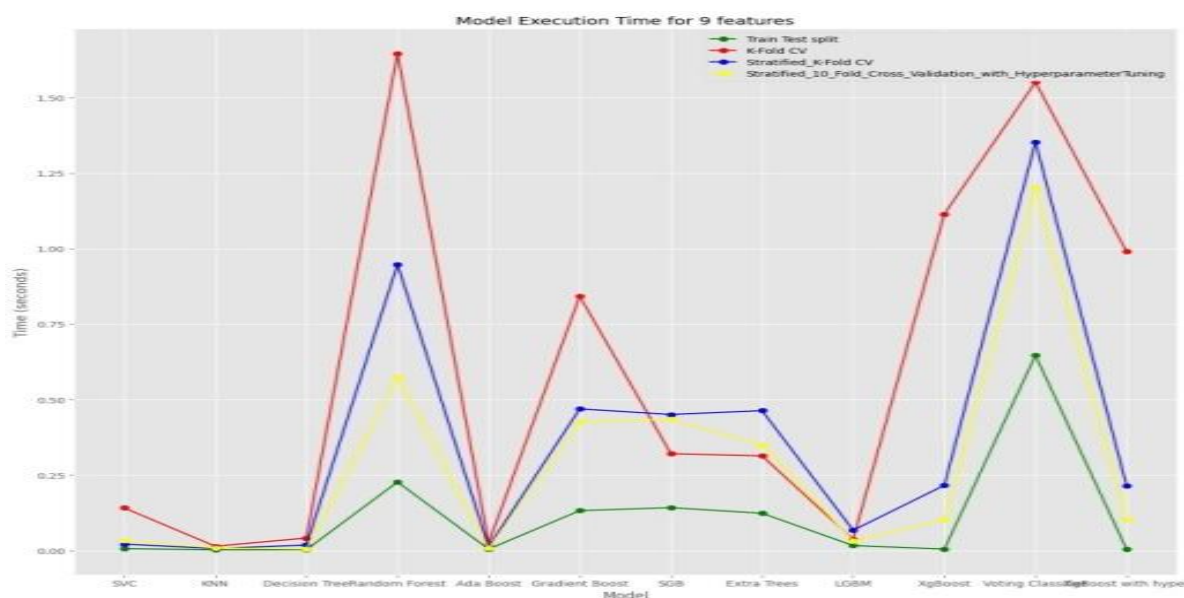


Fig 11. Time Complexity Graph for 9 Features.

Evaluation Metrics

All of the research findings are evaluated using several metrics, each one having its own relevance. A true positive (TP) occurs when an algorithm for prediction produces an outcome that proves out to be accurate. The result of a predictive system that is truly incorrect is called True Negative (TN). The result of model used for prediction that seems true but is actually false is called false positive (FP). False negative (FN) seems an outcome of model for prediction which is false however actually accurate.

Table 3. Results Table for Without Dimensionality Reduction

Model	Without dimensionality reduction						
	Precision	Recall	F1 Score	Train test split Accuracy	10 Fold CV Accuracy	Stratified 10 Fold CV Accuracy	Stratified 10 Fold CV with hyper parameter tuning
Support Vector Machine	0.28	1	0.44	92.70%	94.73%	62.15%	93.30%
K- Nearest Neighbour	0.72	0.76	0.73	89.58%	94.74%	80.50%	80.50%
Decision Tree Classifier	0.96	0.96	0.96	98%	84.21%	94.74%	96.05%
Random Forest Classifier	0.93	0.93	0.93	97.92%	100%	94.59%	94.74%
AdaBoost with decision tree classifier	0.94	0.96	0.95	97.92%	100%	94.47%	94.74%
Gradient Boosting Classifier	0.9	0.93	0.92	96.35%	94.74%	94.47%	94.74%
Stochastic Gradient Boosting classifier	0.92	0.97	0.94	97.39%	96.02%	92.11%	94.74%
Extended Gradient Boosting Classifier	0.9	0.95	0.92	96.87	100%	92.11%	92.11%
Extended Gradient Boosting Classifier with hyper parameter tuning	0.94	0.96	0.97	97.92%	99.02%	92.11%	92.11%
Extra Tree Classifier	0.95	0.95	0.95	97.91%	98.69%	94.74%	97.37%
Light GBM Classifier	0.91	0.96	0.93	96.87%	98.68%	93.42%	93.42%
Voting Classifier	0.63	0.96	0.68	96%	96%	94.74%	94.74%

Accuracy

It shows the percentage within the input data that produced predictions that were correct.

$$\text{Accuracy} = \frac{\text{Number of Correct Prediction}}{\text{Total Number of Predictions}} \quad (2)$$

Precision

What percentage that favorable predictions are valid depends upon the precision. It is computed as the ratio of correctly predicted positive results compared to outcomes that the classifier expected to be positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3)$$

Table 4. Results Table for Without Dimensionality Reduction

Model	With dimensionality reduction						
	Precision	Recall	F1 Score	Train test split Accuracy	10 Fold CV Accuracy	Stratified 10 Fold CV Accuracy	Stratified 10 Fold CV with hyper parameter tuning
Support Vector Machine	0.92	0.78	0.82	62.66%	96.22%	63.61%	90.23%
K- Nearest Neighbour	0.68	0.72	0.97	66.99%	45.52%	77.08%	77%

Decision Tree Classifier	0.77	0.83	0.77	91.14%	92.11%	96.05%	89.47%
Random Forest Classifier	0.87	0.9	0.88	94.62%	98.11%	94.59%	94.74%
AdaBoost with decision tree classifier	0.85	0.83	0.84	91.14%	90.79%	89.47%	93.42%
Gradient Boosting Classifier	0.84	0.88	0.85	94.79%	94.59%	93.28%	93.28%
Stochastic Gradient Boosting classifier	0.84	0.88	0.85	93.57%	97.22%	93.28%	93.28%
Extended Gradient Boosting Classifier	0.84	0.88	0.85	91.32%	98.39%	93.13%	93.13%
Extended Gradient Boosting Classifier with hyper parameter tuning	0.77	0.83	0.77	91.14%	92.11%	96.05%	89.47%
Extra Tree Classifier	0.88	0.88	0.88	94.27%	100%	94.74%	94.74%
Light GBM Classifier	0.84	0.88	0.85	94.27%	98.68%	93.42%	93.42%
Voting Classifier	0.87	0.9	0.88	94.79%	94.59%	93.27%	94.74%

Recall

It usually calculates the percentage of True Positive which has been incorrectly classified. The proportion of True Positive towards the sum of True Positive followed by False Negative.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

F1-Score

It comprises assessing a model that uses binary classification by looking at its favorable class predictions. Recall as well as precision is used to calculate it. The F1-score is in the range of 0 and 1.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

V. CONCLUSION

A prediction of thyroid cancer has become a significant medical issue with an alarming rise in recent years, necessitating automated prediction algorithms. Various existing models have been proposed for prediction of thyroid cancer. But still, those models will take more time. To overcome, the drawbacks in existing models. We applied machine learning techniques using dimensionality reduction and dimensionality complete. The models of machine learning techniques using dimensionality reduction such as KNN, DT, RF, AB, GB, SGB, XGB, XGB with hyper parameter tuning, ETC, LGBM and VC. Further, these models trained using four different techniques such as train test split, 10-fold cross validation, stratified 10-fold cross validation and 10-fold cross validation with hyper parameter tuning. The model ETC produced best accuracy of 100% using 10 Fold cross validation with 0.2030 seconds. Likewise, the KNN produced the lowest accuracy of 45.52% using 10 fold cross validation with 0.0039 seconds. Also, these models applied on dimensionality complete trained with four different techniques. The Models RF, AB and XGB generated the best accuracy of 100% using 10 fold cross validation with 1.6457, 0.0128 and 1.1142 seconds respectively. In the same way the KNN generated the least accuracy of 80.50% with 0.0143 seconds.

Although, the proposed model produced satisfactory results. There are several areas where future work could be focused. The dataset used in this study is structured. We have planned for unstructured datasets using deep learning models for the same in future.

CRedit Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Kalpana K Harish and Kuppala Saritha; **Methodology:** Kalpana K Harish; **Software:** Kuppala Saritha; **Data Curation:** Kalpana K Harish; **Writing- Original Draft Preparation:** Kalpana K Harish and Kuppala Saritha; **Visualization:** Kalpana K Harish; **Investigation:** Kuppala Saritha; **Supervision:** Kalpana K Harish; **Validation:** Kuppala Saritha; **Writing- Reviewing and Editing:** Kalpana K Harish and Kuppala Saritha; All authors reviewed the results and approved the final version of the manuscript.

Acknowledgment

I sincerely thank Dr. Syed Siraj Ahmed for his valuable support and generous contribution of time in completing this paper. His expert assistance in executing the programs was instrumental to this work. I deeply appreciate his dedication and encouragement throughout the process.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

- [1]. A. K. Sharma, R. Arya, R. Mehta, R. Sharma, and A. K. Sharma, "Hypothyroidism and Cardiovascular Disease: Factors, Mechanism and Future Perspectives," *Current Medicinal Chemistry*, vol. 20, no. 35, pp. 4411–4418, Oct. 2013, doi: 10.2174/09298673113206660255.
- [2]. healthline: The 6 Common Thyroid Problems & Diseases. <https://www.healthline.com/health/common-thyroid-disorders> (2018)
- [3]. UK, N.H.S.: Underactive thyroid (hypothyroidism). <https://www.nhs.uk/conditions/underactive-thyroid-hypothyroidism/> (2021)
- [4]. "The untapped potential of the thyroid axis," *The Lancet Diabetes & Endocrinology*, vol. 1, no. 3, p. 163, Nov. 2013, doi: 10.1016/s2213-8587(13)70166-9.
- [5]. B. Zhang., "Machine Learning–Assisted System for Thyroid Nodule Diagnosis," *Thyroid*, vol. 29, no. 6, pp. 858–867, Jun. 2019, doi: 10.1089/thy.2018.0380.
- [6]. P. Gupta., "Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Jan. 2024, doi: 10.1007/s44196-023-00388-2.
- [7]. A. Sultana and R. Islam, "Machine learning framework with feature selection approaches for thyroid disease classification and associated risk factors identification," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, Jun. 2023, doi: 10.1186/s43067-023-00101-5.
- [8]. L. Aversano., "Thyroid Disease Treatment prediction with machine learning approaches," *Procedia Computer Science*, vol. 192, pp. 1031–1040, 2021, doi: 10.1016/j.procs.2021.08.106.
- [9]. M. H. Alshayegi, "Early Thyroid Risk Prediction by Data Mining and Ensemble Classifiers," *Machine Learning and Knowledge Extraction*, vol. 5, no. 3, pp. 1195–1213, Sep. 2023, doi: 10.3390/make5030061.
- [10]. K. Pavya and B. Srinivasan, "Feature selection algorithms to improve thyroid disease diagnosis," *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, pp. 1–5, Mar. 2017, doi: 10.1109/igeht.2017.8094070.
- [11]. M. Deepika, "A Prognostic Thyroid Disorder analysis based on Thyroid Test Measure with ECG," *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, pp. 1–8, Feb. 2021, doi: 10.1109/icaect49130.2021.9392450.
- [12]. A. Tyagi, R. Mehra, and A. Saxena, "Interactive Thyroid Disease Prediction System Using Machine Learning Technique," *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 689–693, Dec. 2018, doi: 10.1109/pdgc.2018.8745910.
- [13]. K. salman and E. Sonuç, "Thyroid Disease Classification Using Machine Learning Algorithms," *Journal of Physics: Conference Series*, vol. 1963, no. 1, p. 012140, Jul. 2021, doi: 10.1088/1742-6596/1963/1/012140.
- [14]. K. Sangaiah, A. Javadpour, C.-C. Hsu, A. Haldorai, and A. Zeynivand, "Investigating Routing in the VANET Network: Review and Classification of Approaches," *Algorithms*, vol. 16, no. 8, p. 381, Aug. 2023, doi: 10.3390/a16080381.
- [15]. N. M. Xi, L. Wang, and C. Yang, "Improving the diagnosis of thyroid cancer by machine learning and clinical data," *Scientific Reports*, vol. 12, no. 1, Jul. 2022, doi: 10.1038/s41598-022-15342-z.
- [16]. A. Raza, F. Eid, E. C. Montero, I. D. Noya, and I. Ashraf, "Enhanced interpretable thyroid disease diagnosis by leveraging synthetic oversampling and machine learning models," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Nov. 2024, doi: 10.1186/s12911-024-02780-0.

Publisher's note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.