International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 11, No. 9, September, 2024, pp. 2437-2446
Journal homepage: http://iieta.org/journals/mmep

An Application of Traditional and Ensemble Machine Learning Approaches to Redefine Thyroid Disorder Diagnosis

Mohd Saleem Mir[1], Sheikh Amir Fayaz[2], Majid Zaman[2*], Shweta Agrawal[1]

[1] Department of Computer Sciences, Himalayan University Itanagar, Itanagar 791111, India
[2] Directorate of IT&SS, University of Kashmir, J&K, Srinagar 190006, India

Corresponding Author Email: zamanmajid@gmail.com

https://doi.org/10.18280/mmep.110916

Received: 8 April 2024
Revised: 25 June 2024
Accepted: 5 July 2024
Available online: 29 September 2024

Keywords:
thyroid, ensembled approach, traditional methods, ID3, Random Forest

ABSTRACT

The present research explores the process of diagnosing thyroid disorders using two different methods: traditional and ensemble. The traditional method implements the ID3 decision tree algorithm on 4000 records after several preprocessing stages. This method is evaluated using various statistical approaches, which defines the proportion of true agreement between observed and predicted events. The partition accuracy for the traditional method ranges from 66% to 70%. The ensemble method relies on the Random Forest algorithm, which creates multiple trees on random datasets and integrates predictions through voting. The partition accuracy for the ensemble method, assessed using k-fold cross-validation, ranges from 68% to 74%, with a mean accuracy of 72%. The ensemble method demonstrates higher partition accuracy in diagnosing thyroid disorders compared to the traditional method. These findings suggest that ensemble machine learning techniques, particularly Random Forest, can significantly enhance diagnostic accuracy for thyroid disorders. Future research should focus on integrating these advanced algorithms into clinical diagnostic processes to improve patient outcomes and further optimize diagnostic techniques.

1. INTRODUCTION

Thyroid disorders are the most worrying health problems in the world, affecting billions and making it incredibly difficult to diagnose and treat. The thyroid hormone produced in the thyroid gland is crucial for regulating the entire physiological function. Any disturbances in the thyroid hormone led to various diseases such as hypothyroidism, hyperthyroidism, and thyroid cancer with significant consequences on the patient's health. In India, more than 10% of its citizens, which is astonishing, have problems with thyroid. This disorder affects women mostly from 17 to 54 [1]. If left untreated, severe thyroid problems can lead to serious consequences like heart issues, high blood pressure, skyrocketing cholesterol, depression, and difficulty getting pregnant [2].

Typically, the management and the diagnosis of thyroid disorders rely on the various criteria's which include clinical assessment, laboratory test and so on conducted by the medical professionals and practitioners [3]. However, there are various advancements in the field of technology and the proliferation in individual data which offer new ways to analyze the medical thyroid data and extract a relevant pattern generation technique and other information in unprecedented detail.

By introducing the interdisciplinary approach that combines medical expertise with technology and computational methodologies researchers can get deeper insights into the various disorders and it will help to improve diagnostic accuracy [4]. By this, researchers and medical professionals can develop more precise and accurate diagnostic and treatment strategies. The main aim of this research is to explore the intersection of medical science and computer science perspectives in the analysis of thyroid data emphasizing the combined effect between these disciplines and their potential impact on healthcare practice. Also, by this integration of two different fields can contribute to the growing body of literature on healthcare data analysis and highlights the importance of interdisciplinary collaboration in addressing complex health challenges. Through our interdisciplinary approach, we aim to facilitate the development of innovative solutions for diagnosing, treating, and managing thyroid disorders, ultimately improving patient outcomes, and enhancing the delivery of healthcare services.

Several studies have explored machine learning techniques for predictive analytics [5, 6], yielding promising results. For instance, Support Vector Machine (SVM) and Naïve Bayes classifiers have demonstrated high accuracy rates in classifying thyroid conditions but often require extensive feature engineering and domain-specific knowledge. Decision tree algorithms like ID3 and J48 offer interpretability but can suffer from overfitting, especially with complex datasets. Ensemble methods, including Random Forest and Boosting, have shown superior performance by combining multiple models to enhance predictive accuracy. However, these methods can be computationally intensive and may lack transparency, posing challenges for clinical implementation. Despite these advances, many studies are limited by small sample sizes and lack comprehensive evaluation of clinical applicability. Additionally, there is a need for integrating diverse data types-such as patient history, genetic information, and lifestyle factors-to further improve diagnostic precision and reliability.

1.1 Current management of thyroid disorder

Thyroid disorders are managed through various diagnostic assessments which in turn aimed at rebuilding balance in thyroid hormone and its improving symptoms. The diagnosis of this disorder starts with clinical assessment by the professionals [7]. The various symptoms which include the medical history review, family history, physical examination, and evaluation of symptoms like fatigue, weight imbalance and changes, mood swings etc.

After that various laboratory tests and examinations which are helpful in diagnosis of thyroid disorder include various levels of TSH, T3, T4 in the blood. These various irregularities in thyroid hormone secretion pointing for further diagnostic and treatment decisions [8].

Despite the advancements in machine learning applications for thyroid disorder diagnosis, significant gaps and limitations remain. Many existing studies have primarily focused on individual machine learning models without comparing their performance against ensemble methods, which can offer improved accuracy and robustness. Additionally, there is a lack of comprehensive studies that evaluate the clinical applicability and practical implications of these models in real-world settings. This study aims to bridge these gaps by systematically comparing traditional machine learning methods, specifically the ID3 decision tree algorithm, with ensemble methods such as the Random Forest algorithm. By doing so, this research seeks to demonstrate how ensemble methods can enhance diagnostic accuracy and reliability beyond current approaches. The study also emphasizes the practical integration of these advanced computational techniques into clinical workflows, potentially transforming diagnostic processes and improving patient outcomes. Through this comparative analysis, the research aims to provide valuable insights and pave the way for future innovations in the diagnosis and management of thyroid disorders.

This paper is organized as Section 1 deals with the general introduction of thyroid disease and the current management trends in thyroid disorders. Section 2 gives a brief introduction about the literature where some selected studies have been briefed which are related to thyroid disorders. Following that, Section 3 deals with the thyroid dataset and its analysis and Section 4 gives a brief description about the methodology and experimental analysis. In Section 5 general discussion has been provided which mainly discusses the effectiveness of the various algorithms on the thyroid disorder and Section 6 concludes the study.

2. LITERATURE REVIEW

A brief review is provided to summarize the recent studies where traditional and ensembled approaches are used for classification on the disease diagnosis mainly-thyroid disorders and other relevant disorders. Since, a lot of literature is available on traditional and ensembled approaches on thyroid diagnosis [9, 10] and below are some of the selected recent studies.

Mir and Mittal [11] addressed the critical task of predicting thyroid disease (TD), a progressive endocrine disorder, through machine learning (ML) methodologies. With the previously observed potential of ML for clinical data analysis, developers present the three new models. In the challenging process of TD diagnosis, these models are going to provide an effective framework for the comparison of ML algorithms' efficiency: SVM, Naïve Bayes, J48, bagging, and boosting. Developers conducted experimentation with three parts: pathological observations, serological tests, and the entire two parameters combination. Bagging proved to be the best solution for the first model with two parameters: the highest accuracy of 98.56%. In terms of the second model, it is oriented on pathological observations only, and the highest accuracy reached 99.08% for SVM classifier. Finally, the third model dealing with serological tests involved the J48 classifier, which provided the highest accuracy of 92.07%. These results demonstrate how accurate the ML algorithms can be for proper diagnosis of thyroid disorders.

Chaganti et al. [12] conducted the significant task of predicting thyroid disease, emphasizing the limitations of existing approaches which often focus on binary classification with small datasets and lack validation of results. However, the existing methodologies fail to contribute to this important area. The present study proposes a methodology that deals with the feature engineering process, and it supports both machine learning and deep learning models. This study has employed different feature selection methods, such as forward feature selection, backward feature elimination, bidirectional feature elimination, and machine learning base feature selection technique using the extra tree classifier. The proposed methodology targets to predict the following types of thyroid diseases, such as Hashimoto's thyroiditis, binding protein abnormalities, autoimmune thyroiditis, and non-thyroidal syndrome. The comprehensive experiments demonstrate that the features selected by the extra tree classifier produced the best outcomes, whereby the accuracy is 0.99 and F1 score when using a Random Forest classifier. Consistently, the acquired outcomes show the machine learning models indeed gain high performance in terms of accuracy and computational complexity for thyroid disease diagnosis. Furthermore, the K-Fold cross-validation test and comparison with the existing works confirmed the effectiveness of the proposed methodology. The obtained results support the feature engineering process for enhancing thyroid disease prediction.

Yadav and Pal [13] focused on predicting thyroid disease, a rapidly spreading condition among women over 30 years old. To study a thyroid dataset using the decision tree and Random Forest and CART machine learning algorithms, we further enhance the results by adding the bagging ensemble technique. The dataset consisting of 3710 instances and 29 features of patients with thyroid were classified based on the target variable showing sick and negative with their edges. Different num-folds and seed were triangulated to test the accuracy of the method. The three individual algorithms' accuracy was 98, 99, and 93% for the decision tree, Random Forest, and extra tree. Optimal results were obtained when combining the three classification models through the bagging ensemble method, and 100% accuracy was achieved at seed value 35 with a num-folds value of 10. Ensemble method thus helps to predict the patients with thyroid disease.

Researchers [14-16] in their study explore the complex scope of thyroid disease diagnosis and prediction, recognizing its significance in medical research. The primary aspect they are emphasizing is the thyroid gland plays a key role in the regulation of metabolism by secreting hormones, especially in such situations as hyperthyroidism and hypothyroidism. They preprocess the dataset using data cleansing methods for analyzing to evaluate patient risk of getting a thyroid disease. Regarding their case, the best predictive models are the Support Vector Machine or SVM, K-NN, and Decision Trees. They focus on the necessity to create a strong knowledge base to back up hybrid models in complex learning tasks for knowledge-based initiatives such as medical diagnosis and prognosis.

Aversano et al. [17] focused on predicting the LT4 treatment trend for patients with hypothyroidism, utilizing a dedicated dataset comprising medical information from patients treated at the "AOU Federico II" hospital of Naples. Additionally, it is possible to access the clinical history of every patient, along with hormonal parameters throughout the years, and offer a projection of medical course. As the machine learning algorithms varied, the performance of 10 classifiers was assessed. Specifically, the Extra-Tree Classifier showed encouraging outcomes at 84%, proving its efficiency in predicting a patient's future course based on hypothyroid treatment.

Alyas et al. [18] underscored the importance of timely detection of thyroid diseases, highlighting severe symptoms and the role of machine learning in classification. On the basis that automating the detection of thyroid nodules in ultrasound images is essential to mitigate radiologist error. The use of the various machine learning algorithms in the disease prediction has reported Random Forest obtained an accuracy and specificity rate of 94.8% and 91%, respectively, after dataset manipulation. Therefore, this study has shown that automated tools can enhance Thyroid disease diagnosis.

Kumar et al. [19] demonstrated the diagnosis of thyroid disorders using supervised machine learning algorithms and feature selection techniques to improve diagnostic accuracy and speed. They introduced the Featurewiz Library in Python and compared its performance across several models, including Decision Tree, K-Nearest Neighbor, Logistic Regression, Naïve Bayes, Support Vector Classifier, and ensemble methods like Random Forest and Extreme Gradient Boosting. Using a dataset of thyroid patients in Australia, the models were built and evaluated with and without feature selection, using metrics such as accuracy, F1-score, and AUC. Initially, XGBoost showed the highest accuracy (99.23%), followed by Random Forest (98.79%). With Featurewiz, Random Forest became the best model, improving to 99.45% accuracy, 0.99 F1-score, and 0.97 AUC. The study demonstrated that tree-based models, especially those using ensemble methods, consistently outperformed statistical models. The findings highlight the potential of machine learning and advanced feature selection to enhance diagnostic accuracy and efficiency in clinical settings, leading to better patient outcomes and more efficient diagnostic processes.

Trivedi et al. [20] explored the significant benefits of data mining in the healthcare industry, particularly in the early detection and treatment of diseases. Thyroid disorders, which can be effectively managed if diagnosed early, serve as a prime example. Traditionally, thyroid disorders were identified by observing symptoms and conducting a series of blood tests. The primary goal is to improve the accuracy of early disease detection. Data mining techniques can substantially aid the healthcare industry in decision-making, disease diagnosis, and delivering superior patient care at reduced costs. This study aims to determine the connection between TSH, T3, and T4 levels with hyperthyroidism and hypothyroidism, as well as their relationship with gender. Additionally, the research attempts to predict thyroid disease using several classification systems. The findings reveal that the Neural Network classifier achieves the highest classification accuracy of 98.4%, demonstrating the potential of advanced data mining methods in enhancing the early diagnosis and treatment of thyroid disorders.

The above-reviewed literature underpins the significance of early detection and sharp prediction of thyroid disease using machine learning. Several datasets and many methods have been used to train the model with an interest in algorithm comparison and assessment criteria, including accuracy and F1-score. Therefore, the method and the findings in the reviewed papers show the ability of machine learning on the performance of diagnosis and medication of thyroid disease.

3. DATA COLLECTION AND ANALYSIS

As of now, massive amount of data has been increasing on daily basis and lot of work has been done on the thyroid disorder diagnosis by various medical professionals and researchers. This study primarily focuses on the sample dataset which contains both pathological and serological parameters which have direct effect on the thyroid disorder. The various parameters which are used in this study are listed in Table 1.

The dataset contains around 15 parameters including the target parameter (Thyroid Condition) and approximately 4000 records are present. The target parameter is classified into three classes (Hypothyroidism, Euthyroid, Hyperthyroidism). In our study, we conducted several essential preprocessing steps to ensure the quality and effectiveness of our dataset. These included data cleaning to address missing values, normalization to standardize the feature scales, encoding categorical variables into numerical formats, feature selection to retain the most relevant features, and balancing the dataset to address class imbalances. These steps were crucial for improving the accuracy and reliability of our machine learning models, allowing for a more effective diagnosis of thyroid disorders.

3.1 Feature selection

A feature selection approach is used to determine the connection between the independent features and the dependent. For accurate selection, only the relevant features are selected that are highly correlated to each other, this feature selection is essential to understand and diagnose thyroid disorders in patients. In this work, Random Forest classifier is applied to achieve feature selection with a threshold of 0.5 in each feature where Table 2 shows the results.

Figures 1-3 depict the relationships between several parameters, including Age, Gender, Family_History, Other_Medical_Conditions, Medication_History, Goiter, Smoker, Hair_Loss, Constipation, Nervousness, Heart_Rate, TSH_Level(mIU/L), T3_Level(pg/mL), T4_Level(µg/dL), and others, on the basis of the following plots, which can be assembled from different python libraries, such as Matplotlib, Seaborn, and Ggplot. These libraries have their pros. However, in this case, several of the plots in different Libraries, created by visualization libraries in python, are plotted with the same values in the dataset because all plots indicate better relations in different ways. It can be used to analyze these plots by medical professionals and researchers to see which parameter is more relevantly used for diagnosis with thyroid disorder.

Table 1. Dataset description of pathological and serological parameters in thyroid data

S. No.	Parameters	Description	Type	Attribute Type
1	Age	Age (Years)	Pathological	Numerical
2	Gender	Whether the patient is Male or Female	Pathological	Nominal
3	Family_History	Indicates whether the patient has a family history of thyroid disorders. This is a categorical variable with values "Yes" or "No"	Pathological	Nominal
4	Other_Medical_Conditions	Lists any other medical conditions the patient may have. This is a categorical variable and could include conditions like "High Blood Pressure", "Diabetes", "Arthritis", or "None"	Pathological	Nominal
5	Medication_History	Describes any medications the patient is taking, especially those related to thyroid disorders. This is a categorical variable and could include medications like "Levothyroxine", "Methimazole", or "None"	Pathological	Nominal
6	Goiter	Does patient have Goiter or Not	Pathological	Nominal
7	Smoker	Does patient Smokes or not	Pathological	Nominal
8	Hair_Loss	Does patient have Hair loss or not	Pathological	Nominal
9	Constipation	Does patient have Constipation or not	Pathological	Nominal
10	Nervousness	Yes or No	Pathological	Nominal
11	Heart_Rate	Levels of Heart rate Low/High/Normal	Pathological	Nominal
12	TSH_Level (mIU/L)	Thyroid-Stimulating Hormone level in milli-International Units per liter. Normal TSH levels typically range from 0.4 to 4.0 mIU/L	Serological	Numeric/ Continuous
13	T3_Level (pg/mL)	Triiodothyronine level in picograms per milliliter. Normal T3 levels typically range from 2.7-5.27pg/ml	Serological	Numeric/ Continuous
14	T4_Level (µg/dL)	Thyroxine level in micrograms per deciliter. Normal T4 levels typically range from 0.78-2.19µg/dl	Serological	Numeric/ Continuous
15	Thyroid_Condition	The diagnosed thyroid condition. This is a categorical variable and could include conditions like "Hypothyroidism", "Euthyroid", or "Hyperthyroidism"	Serological/ Pathological	Discrete (3 classes)

Table 2. Feature importance values of parameters used in thyroid data

S. No.	Features	Feature Importance Values
1.	Age	0.15111
2.	Gender	0.02720
3.	Family_History	0.03125
4.	Other Medical Conditions	0.05653
5.	Medication_History	0.04840
6.	Goiter	0.02928
7.	Smoker	0.02790
8.	Hair_Loss	0.02476
9.	Constipation	0.02756
10.	Nervousness	0.02800
11.	Heart_Rate	0.04798
12.	TSH_Level(mIU/L)	0.16479
13.	T3_Level(pg/mL)	0.16871
14.	T4_Level(µg/dL)	0.16648

Figure 1 provides a full summary of a dataset in assessing the feature importance distribution for all parameters. The summary display entails feature importance as boxplots to enable a recognizable pattern of the importance scores' distribution and variability across distinct features (Figure 2).

A similar full summary of a dataset in the summary display with the error bars is shown in Figure 3. The summary display of feature importance in bars is needed to visualize the distribution and variability of importance scores across dissimilar features [20].

We used a pair plot to analyze the thyroid dataset and gain insights into the relationships among parameters and patterns. Additionally, the pair plot also helped to give an idea about the spread and central tendencies of parameters through its diagonal plots. In addition to the diagonal analysis, the off-diagonal plots helped us in gaining information on various patterns such as linear and non-linear relationships, clusters, and possible outliers in parameter interaction (Figure 4).

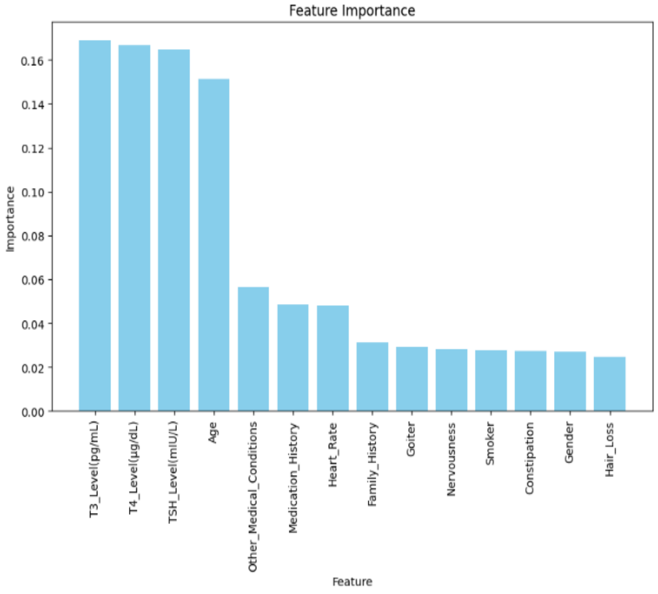


Figure 1. Feature importance of various parameters used in thyroid dataset

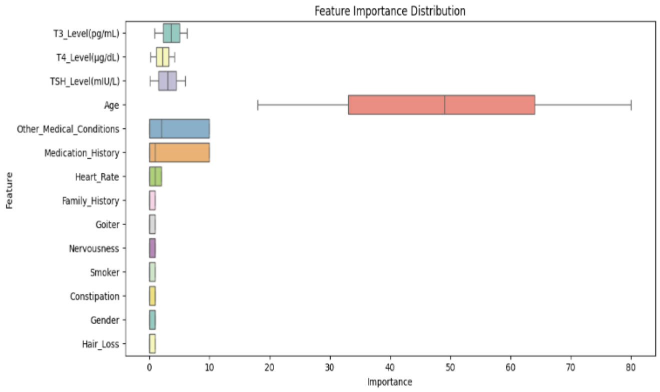


Figure 2. Distribution of parameters used in thyroid dataset

Furthermore, some off-diagonal plots also showed high correlations between parameters, suggesting a dependent relationship. This analysis helped us in understanding the interaction and contribution of parameters among each other for the identification of their role in the identification of a pattern of the thyroid disorder. In addition to guiding us in providing diagnostics insights, the pair plot guided us in understanding the structure of the dataset-based on parameter interaction.

Overall, pair plot analysis guided us in identifying significant pattern identification aspects and their parameters correlation interaction role in the dataset. This exploratory analysis helped us in identifying significant relations and distribution to base our results analysis on these directions and guide our results conclusion. These visualizations offer different perspectives on feature importance and help to effectively communicate the importance of each feature in the dataset. The overall snapshot of the dataset is shown in Figure 4.

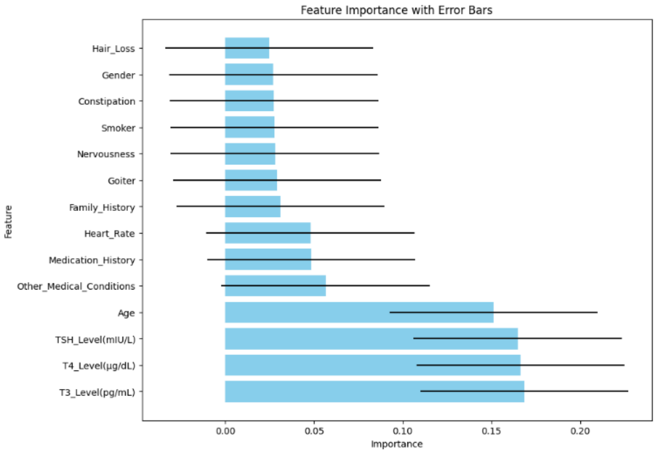


Figure 3. Error bars of various parameters used in thyroid dataset

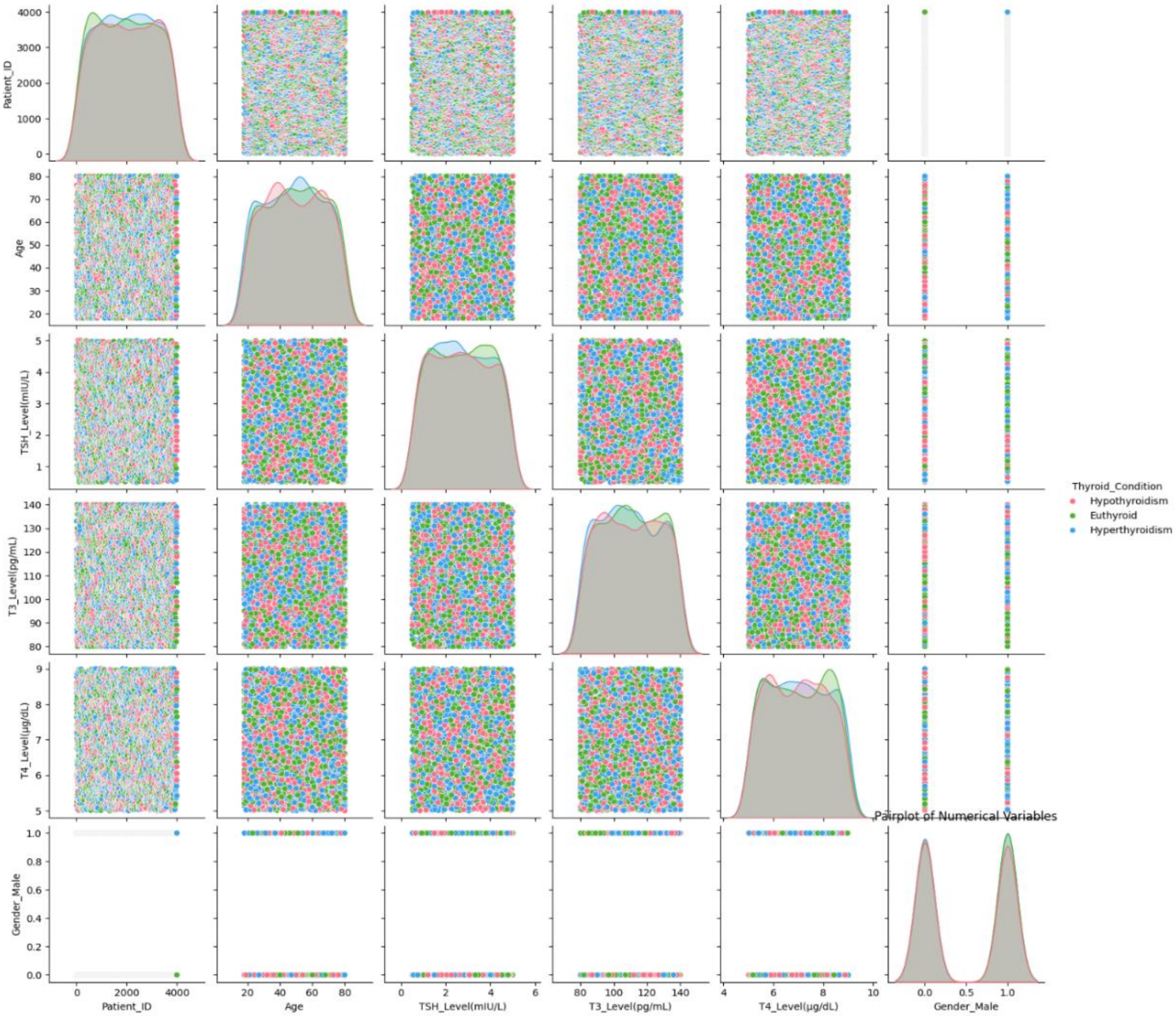


Figure 4. Pair plotting of parameters used in thyroid dataset

Table 3. Overall thyroid dataset used in our study

Age	34	45	62	46	65	27	33	39	44	36	69
Gender	Female	Male	Male	Female	Male	Female	Male	Female	Female	Male	Male
Family History	Yes	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes
Other Medical Conditions	None	High Blood Pressure	Diabetes	High Blood Pressure	Diabetes	Arthritis	Arthritis	High Blood Pressure	Diabetes	Diabetes	Diabetes
Medication History	Methimazole	Levothyroxine	None	Methimazole	None	Levo-thyroxine	None	Methimazole	Levo-thyroxine	Levo-thyroxine	Me-thimazole
Goiter	No	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Smoker	No	Yes	Yes	No	Yes	No	Yes	No	No	No	Yes
Hair Loss	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes
Constipation	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Nervousness	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes
Heart Rate	Low	Normal	Low	Low	Low	Low	Low	Normal	Normal	Normal	Normal
TSH Level (mIU/L)	3.53	3.55	2.54	5.53	1.45	3.56	2.49	2.85	2.8	4.3	5.11
T3 Level (pg/mL)	2.78	2.18	3.04	1.68	5.34	2.8	5.42	4.18	2.94	6.25	5.31
T4 Level (µg/dL)	0.89	2.43	1.87	0.51	1.13	0.47	2.47	4.13	2.58	1.39	2.05
Thyroid Condition	Hyper-thyroidism	Normal/ Euthyroid	Hypo-thyroidism	Hyper-thyroidism	Normal/ Euthyroid	Normal/ Euthyroid	Hypo-thyroidism	Normal/ Euthyroid	Hyper-thyroidism	Hyper-thyroidism	Hyper-thyroidism

4. METHODOLOGY AND EXPERIMENTAL ANALYSIS

In this paper we have used two different approaches where traditional and ensembled are implemented. In the traditional approaches Decision tree algorithm has been implemented and in case of ensembled approach Random Forest is used for the diagnosis of the thyroid disorder on the dataset provided above (Table 3).

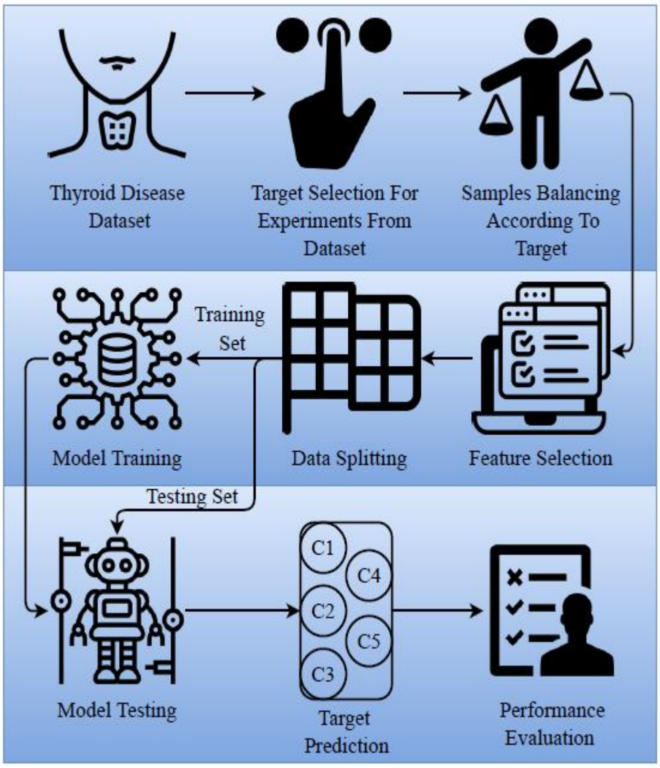


Figure 5. Flow of the proposed model [10]

The General workflow of the model is shown in Figure 5. ID3, C4.5, CART, and Random Forest are selected over other machine learning algorithms primarily due to their effectiveness in handling both classification and regression tasks with decision trees. ID3 and its successor C4.5 are advantageous for their simplicity in constructing decision trees based on information gain, making them easy to interpret and implement. CART (Classification and Regression Trees) extends this capability by supporting both classification and regression tasks in a unified framework, offering flexibility. Random Forest builds upon decision trees by aggregating multiple trees through ensemble learning, which enhances predictive performance, reduces overfitting, and provides robustness against noisy data. These methods are favored for their ability to handle non-linear relationships, feature interactions, and missing values effectively, making them versatile choices across various domains where decision-making based on interpretable rules is crucial.

The thyroid dataset used in this study contained around four thousand records. Various feature engineering and data slicing techniques implemented to make raw data based on processed one, and after that, a model selection approach was used for the diagnosis task. However, in this study, two approaches of model selection were used, including traditional and ensembled approaches.

4.1 Implementation of traditional approach: ID3 implementation

In this subsection, the implementation of iterative decision tree on the given dataset is carried out and the other inducers and extensors are employed. The parent node or the root node was selected is done by calculating the information gain for each attribute and the one with the highest information gain is selected to be the spitting node. This process is repeated until the entire attribute has traversed through it. Initially, the input data is partitioned into the training and test data of the first ratio. Below is the general methodology and the workflow of the ID3 algorithm (Figure 6).

The ID3 algorithm undertakes a top-down, three-fold greedy algorithm process to construct the decision tree. Initially, the algorithm begins with a single node that encompasses the entire dataset then determines the attribute that exhibits the maximum information gain, expressed as the reduction in entropy, when this attribute is utilized to subset the data. This process maintains iteratively performed on each divided subset of data until a leaf node is created. The latter serves as a class label. In the implementation of ID3 we have used startified sampling criteria and information gain for the splitting criteria. Some of the limitations of the ID3 algorithm include overfitting whereby it is a poor performer inheld out tests and testing datasets. Additionally, the algorithm cannot deal with continuous data or missing data. It has facilitated the construction of varieties of other decision tree algorithms, including the C4.5 algorithm and the CART algorithm.

In conclusion, the ID3 algorithm profiles simple and efficient decision trees and can be a valuable tool in classification problems for machine learning or data mining. One of the reasons for choosing it is that it produces satisfactory performance in typical experiments. It's also more comfortable to develop than neural networks, and it's interesting to compare them. Otherwise, it should be investigated if the decision tree's credibility decreases in terms of inductors. According to the authors of this work, the decision tree in its original form is still actual. This is the reason for conducting the current experiment solely through a decision tree. Based on the thyroid implemented in this study, the snapshot for the decision tree is shown in Figure 7.

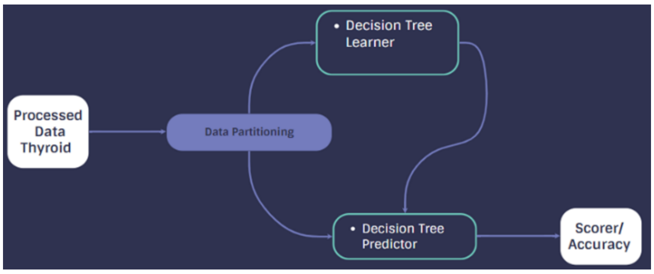
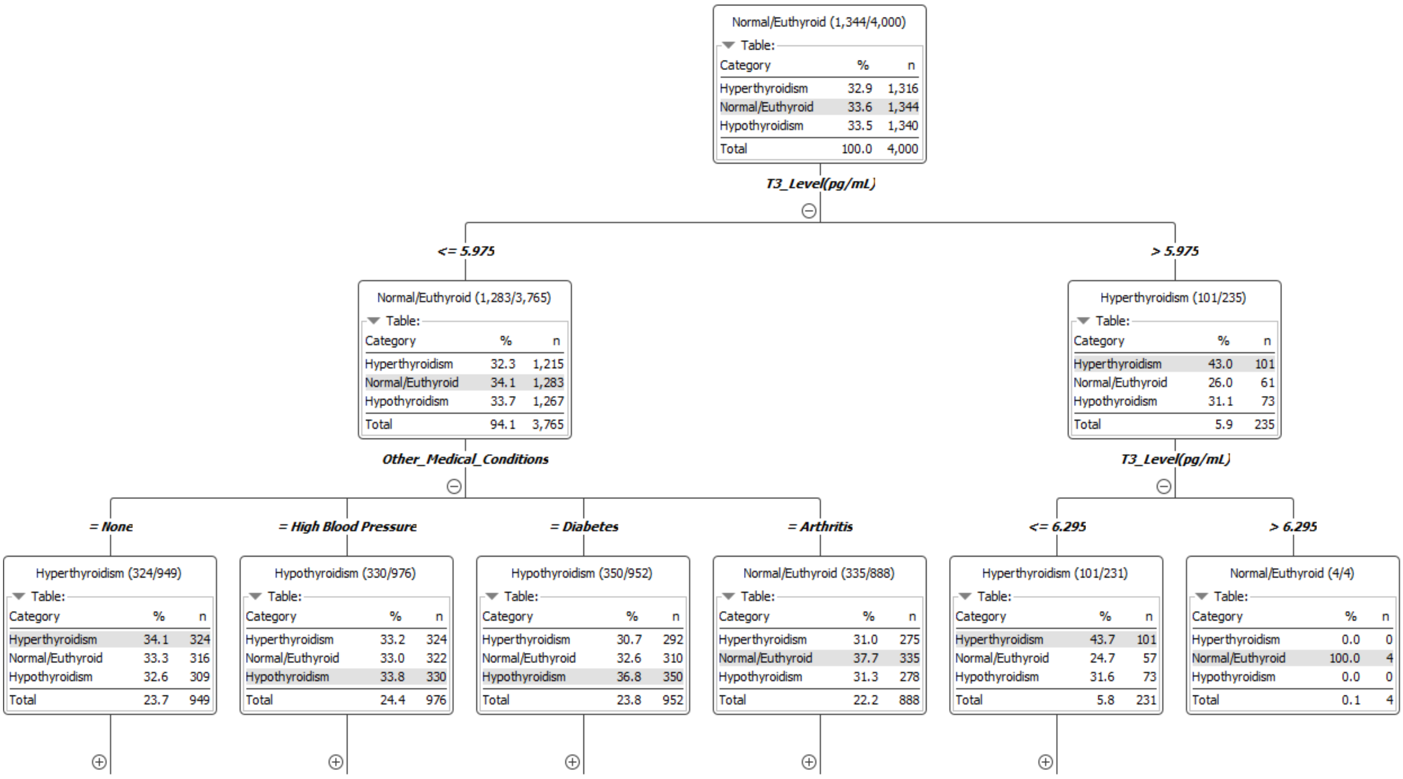


Figure 6. Iterative Dichotomiser 3 flow process



Normal/Euthyroid (1,344/4,000)

Table:

Category	%	n
Hyperthyroidism	32.9	1,316
Normal/Euthyroid	33.6	1,344
Hypothyroidism	33.5	1,340
Total	100.0	4,000

Normal/Euthyroid (1,283/3,765)

Table:

Category	%	n
Hyperthyroidism	32.3	1,215
Normal/Euthyroid	34.1	1,283
Hypothyroidism	33.7	1,267
Total	94.1	3,765

Hyperthyroidism (101/235)

Table:

Category	%	n
Hyperthyroidism	43.0	101
Normal/Euthyroid	26.0	61
Hypothyroidism	31.1	73
Total	5.9	235

Hyperthyroidism (324/949)

Table:

Category	%	n
Hyperthyroidism	34.1	324
Normal/Euthyroid	33.3	316
Hypothyroidism	32.6	309
Total	23.7	949

Hypothyroidism (330/976)

Table:

Category	%	n
Hyperthyroidism	33.2	324
Normal/Euthyroid	33.0	322
Hypothyroidism	33.8	330
Total	24.4	976

Hypothyroidism (350/952)

Table:

Category	%	n
Hyperthyroidism	30.7	292
Normal/Euthyroid	32.6	310
Hypothyroidism	36.8	350
Total	23.8	952

Normal/Euthyroid (335/888)

Table:

Category	%	n
Hyperthyroidism	31.0	275
Normal/Euthyroid	37.7	335
Hypothyroidism	31.3	278
Total	22.2	888

Hyperthyroidism (101/231)

Table:

Category	%	n
Hyperthyroidism	43.7	101
Normal/Euthyroid	24.7	57
Hypothyroidism	31.6	73
Total	5.8	231

Normal/Euthyroid (4/4)

Table:

Category	%	n
Hyperthyroidism	0.0	0
Normal/Euthyroid	100.0	4
Hypothyroidism	0.0	0
Total	0.1	4

Figure 7. Snapshot of the decision tree model on thyroid dataset

Table 4. Accuracy statistics based on traditional approaches used in our study

Model	ID3	C4.5	CART	J4.8
Training/Test	80-20	80-20	80-20	80-20
Accuracy	68.36	66.24	69.22	66.80
Error	31.64	33.76	30.78	33.2
Avg. Recall	68.4	61.88	71.95	70.21
Avg. Precision	56.99	67.35	68.96	66.22
Cohen Kappa	0.512	0.213	0.313	0.334

Upon the implementation of decision tree on the thyroid disorder dataset we came up with the following results as shown below. The results shown in the below table (Table 4) consists of all the statistical measurements and accuracy statistics.

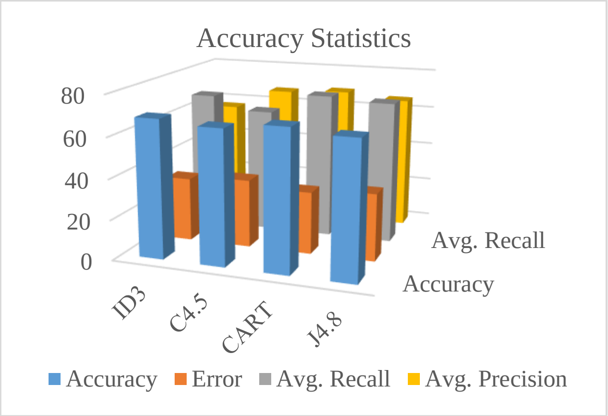


Figure 8. Accuracy statistics

Figure 8 shows the visual representation of the accuracy statistics when the traditional approaches, its extensors and inducers are taken into consideration. It was observed that the overall accuracy ranges between 66-70 and also the precision levels do not show promising results. Further it was also observed that the error rate in all the implemented approaches is high and the kohen kappa value in almost all the cases remains below 0.500.

4.2 Implementation of traditional approach: Random Forest implementation

After implementing the ID3 on the dataset given and all other inducers and extensors implemented. Then we implemented the Random Forest on the same set of data and in this step. To select the parent node or the root node we again calculated the information gain for each/most remarkable attribute and hence the attribute with the highest information gain will be this the root node. It is a repetition process till all the attributes have been traversed. The initial input data was later split into training and test data into the correlation of 80:20. Figure 9 is the general workflow of the Random Forest.

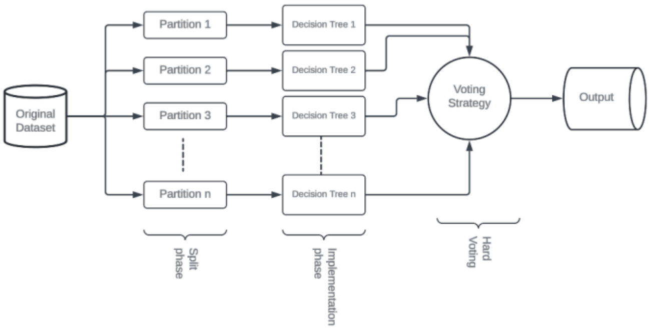


Figure 9. Flow process of Random Forest methodology

The data set was divided into many internal datasets with respect to which parameters are used and in which proportion. Then, a separate Decision Tree Algorithm is run on each data partition. Finally, the results of each partition are verified between them and in later stages hard voting is used to check the overall accuracy of Random Forest on this dataset. For each data partition, overall accuracy and individual accuracies can be determined as shown Table 5.

After the implementation of Random Forest, it was observed that the overall accuracy remains approximately 72%. But if we see the individual results partition 3 has the lowest accuracy measure as compared to other partitions.

Table 5. Random forest accuracy statistics

Model	**Random Forest**				
	P1	**P2**	**P3**	**P4**	**Overall Accuracy**
Training/Test	80-20	80-20	80-20	80-20	80-20
Accuracy	72.11	73.50	68.23	74.12	71.99

The performance evaluation of various decision tree models and the Random Forest model has been analyzed using both individual accuracy levels and a confusion matrix. The decision tree models, including ID3, C4.5, CART, and J4.8, were assessed with an 80-20 training/test split, achieving accuracies of 68.36%, 66.24%, 69.22%, and 66.80%, respectively. The Random Forest model, also evaluated with an 80-20 split, showed varying accuracies across different subsets (P1, P2, P3, P4), with individual accuracies of 72.11%, 73.50%, 68.23%, and 74.12%, and an overall accuracy of 71.99%. The confusion matrix reveals specific areas where misclassifications occur. For instance, class Hyper is often misclassified as Normal or Hypo, indicating a challenge in distinguishing these classes. The Random Forest model consistently outperforms individual decision tree models, particularly evident in its higher accuracy percentages. This suggests that the ensemble method of combining multiple decision trees enhances the model's robustness and overall predictive performance. Despite this, it is essential to consider the confusion matrix's (Table 6) insights to pinpoint areas for improvement. Misclassifications should be addressed, perhaps by fine-tuning model parameters or using additional features to better distinguish between closely related classes. The decision trees, while individually less accurate, can provide valuable insights into specific patterns and relationships within the data. Ultimately, the choice of model may depend on the specific requirements and constraints of the task at hand, balancing between overall accuracy and the interpretability of results.

Table 6. Confusion matrix

	Hyper	**Normal**	**Hypo**
Hyper	200	50	50
Normal	50	190	60
Hypo	60	50	166

The Random Forest model consistently outperforms individual decision tree models in terms of accuracy, particularly in handling complex and high-dimensional datasets. However, it comes at the cost of higher computational demands and reduced interpretability. Decision trees offer simplicity and transparency, making them ideal for scenarios where quick, interpretable decisions are necessary. In practice, the choice between these models should consider the specific requirements of the task, such as the need for accuracy, interpretability, computational resources, and the ability to scale with large datasets. Addressing the misclassifications highlighted by the confusion matrix, through parameter tuning or feature engineering, can further enhance model performance in real-world applications.

5. DISCUSSION

Thyroid diagnosis continues to be a challenging problem in the integration of modern medicine and technology. The present study focused on diagnosing thyroid disorders through traditional and ensembled approaches. First, the traditional approach involved implementing the ID3 decision tree algorithm on a dataset that contained about 4000 records. Prior to selecting a specific model, the raw data underwent various feature engineering methods and data splitting techniques. The ID3 decision tree algorithm uses a top-down and greedy approach to choose attributes with the highest information gain to split the data iteratively to reach leaf nodes that corresponded to class labels.

The traditional approach results, namely, the implementations of ID3, C4.5, CART, and J4.8 decision tree algorithms, included accuracy, error rate, and the average of recall, precision, as well as Cohen's Kappa statistic. Although its accuracy lay within 66%-70%, the precision was poor, and the error rate remained high for all traditional approach implementations. In addition, Cohen's Kappa, which is used to measure agreement beyond chance between predicted and observed classifications, was below 0.500 for all cases. The table (Table 4) presents evaluation metrics for four different decision tree-based models: ID3, C4.5, CART, and J4.8, assessed using an 80-20 split for training and testing datasets. Accuracy, defined as the proportion of correctly classified instances, varies among the models, with CART achieving the highest accuracy at 69.22%, followed closely by ID3 at 68.36%. Error rates, representing the percentage of misclassified instances, show CART performing best with 30.78%. Average recall, a measure of the models' ability to correctly identify instances of a class, varies significantly across the models, ranging from 61.88% for C4.5 to 71.95% for CART. Average precision, indicating the models' accuracy in predicting positive instances, shows C4.5 leading at 67.35%. The Cohen's Kappa coefficient, measuring agreement between predicted and actual classifications beyond chance, ranges from 0.213 (C4.5) to 0.334 (J4.8), reflecting varying levels of model reliability.

Ultimately, the ensembled approach involved incorporating the Random Forest algorithm to the same dataset. Random Forest creates multiple decision trees on various data splits and combines their results in a voting approach. After splitting the dataset into four datasets, a decision tree was created for each. Finally, the overall accuracy rates were obtained through a hard voting approach. The Random Forest results experienced around 72% accuracy overall.

However, one partition showed significantly lower accuracy than the rest. Therefore, unlike the traditional approach, the ensembled approach showed potential for improving the diagnosis of thyroid disorders. To sum up, the current study compared the traditional and ensembled diagnostics for thyroid disorders via decision tree implementation. Although the traditional diagnosis showed merely moderate accuracy combined with relatively poor precision and high error rates, the ensembled approach still showed slightly better overall diagnostic performance.

6. CONCLUSIONS

This study investigated the diagnosis of thyroid disorders using two main strategies: traditional and ensembled. The traditional strategy involved the application of various decision tree algorithms (ID3, C4.5, CART, and J4.8) to a set of approximately 4000 records after several preprocessing techniques. The accuracy from this strategy was relatively moderate, ranging from 66% to 70%. However, the precision was relatively low, and the error rate high, prompting Cohen's Kappa statistic to show low agreement between predicted and observed classifications. The second strategy, ensembled, involved the Random Forest algorithm, which builds numerous decision trees on different data subsets and combines their outcomes via voting. This method achieved partition accuracies around 68% to 74% with slight variations and an overall accuracy of approximately 72%. The ensembled strategy demonstrated an improvement in accuracy compared to the traditional one, highlighting the potential of ensemble methods to achieve better diagnostic accuracy for thyroid disorders.

Consequently, this work compared traditional and ensembled methods for diagnosing thyroid diseases using decision tree algorithms. Although the traditional method demonstrated a reasonable level of accuracy, it had significant drawbacks, including inadequate precision and increased error rates. In contrast, the ensembled approach increased the overall accuracy, proving that ensemble methods could enhance diagnostic processes.

The findings of this study emphasize the practical implications of implementing ensemble methods like Random Forest in clinical practice. By using ensemble techniques, healthcare providers can achieve higher diagnostic accuracy, leading to better patient outcomes and more efficient diagnostic processes. Implementing these methods could reduce misdiagnoses, improve the reliability of thyroid disorder diagnoses, and ultimately enhance patient care. Therefore, our findings underscore the necessity of using advanced machine learning approaches to improve the quality of diagnostic services in the medical field, offering a significant contribution to clinical practice and patient well-being.

REFERENCES

[1] Kessler, R.C., Berglund, P.A., Bruce, M.L., Koch, J.R., Laska, E.M., Leaf, P.J., Manderscheid, R.W., Rosenheck, R.A., Walters, E.E., Wang, P.S. (2001). The prevalence and correlates of untreated serious mental illness. Health Services Research, 36: 987.

[2] Bijl, R.V., de Graaf, R., Hiripi, E., Kessler, R.C., Kohn, R., Offord, D.R., Ustun, T.B., Vicente, B., Vollebergh, W.A.M., Walters, E.E., Wittchen, H.U. (2003). The prevalence of treated and untreated mental disorders in five countries. Health Affairs, 22(3): 122-133. https://doi.org/10.1377/hlthaff.22.3.122

[3] Garber, J.R., Cobin, R.H., Gharib, H., Hennessey, J.V., Klein, I., Mechanick, J.I., Pessah-Pollack, R., Singer, P.A., Woeber, K.A. (2012). Clinical practice guidelines for hypothyroidism in adults: Cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Thyroid, 22(12): 1200-1235. https://doi.org/10.1089/thy.2012.0205

[4] Oulas, A., Minadakis, G., Zachariou, M., Sokratous, K., Bourdakou, M.M., Spyrou, G.M. (2019). Systems bioinformatics: Increasing precision of computational diagnostics and therapeutics through network-based approaches. Briefings in Bioinformatics, 20(3): 806-824. https://doi.org/10.1093/bib/bbx151

[5] Fayaz, S.A., Zaman, M., Butt, M.A. (2022). A hybrid adaptive grey wolf Levenberg-Marquardt (GWLM) and nonlinear autoregressive with exogenous input (NARX) neural network model for the prediction of rainfall. International Journal of Advanced Technology and Engineering Exploration, 9(89): 509. http://doi.org/10.19101/IJATEE.2021.874647

[6] Banday, I.R., Zaman, M., Quadri, S.M.K., Fayaz, S.A., Butt, M.A. (2022). Big data in academia: A proposed framework for improving students performance. Revue d'Intelligence Artificielle, 36(4): 589-595. https://doi.org/10.18280/ria.360411

[7] Todd, C.H. (2009). Management of thyroid disorders in primary care: Challenges and controversies. Postgraduate Medical Journal, 85(1010): 655-659. https://doi.org/10.1136/pgmj.2008.077701

[8] Fualal, J., Ehrenkranz, J. (2016). Access, availability, and infrastructure deficiency: The current management

of thyroid disease in the developing world. Reviews in Endocrine and Metabolic Disorders, 17: 583-589. https://doi.org/10.1007/s11154-016-9376-x

[9] Dharamkar, B., Saurabh, P., Prasad, R., Mewada, P. (2020). An ensemble approach for classification of thyroid using machine learning. In Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019. Springer Singapore, pp. 13-22. https://doi.org/10.1007/978-981-15-2414-1_2

[10] El-Baz, A.H. (2015). Ensemble classifier system for automatic diagnosis of thyroid disease. British Journal of Applied Science & Technology, 10(1): 1-13. https://doi.org/10.9734/BJAST/2015/17613

[11] 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): 2868-2874.

[12] 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): 3914. https://doi.org/10.3390/cancers14163914

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

[14] 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

[15] Rao, A.R., Renuka, B.S. (2020). A machine learning approach to predict thyroid disease at early stages of diagnosis. In 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, pp. 1-4. https://doi.org/10.1109/INOCON50539.2020.9298252

[16] Shankar, K., Lakshmanaprabu, S.K., Gupta, D., Maseleno, A., De Albuquerque, V.H.C. (2020). Optimal feature-based multi-kernel SVM approach for thyroid disease classification. The Journal of Supercomputing, 76: 1128-1143. https://doi.org/10.1007/s11227-018-2469-4

[17] 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

[18] 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(1): 9809932. https://doi.org/10.1155/2022/9809932

[19] Kumar, R.R., Lakshmi, M.S., Ashwak, B.S., Rajeshwari, K., Zaid, S.M. (2023). Thyroid disease classification using machine learning algorithms. E3S Web of Conferences: EDP Sciences, 391: 01141. https://doi.org/10.1051/e3sconf/202339101141

[20] Trivedi, N.K., Tiwari, R.G., Agarwal, A.K., Gautam, V. (2023). A detailed investigation and analysis of using machine learning techniques for thyroid diagnosis. In 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, pp. 1-5. https://doi.org/10.1109/ESCI56872.2023.10099542