International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

A comparative study of machine learning algorithms for thyroid disease classification

Komal *

Computer Science & Engineering, Krishna School of Technology, Drs. Kiran and Pallavi Patel Global University, India.

International Journal of Science and Research Archive, 2025, 14(02), 077-085

Publication history: Received on 19 December 2024; revised on 27 January 2025; accepted on 30 January 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.2.0305

Abstract

Thyroid disorders, affecting millions of individuals across the globe, require prompt and reliable diagnosis for optimal treatment and better patient results. On the other hand, conventional diagnostic tools are usually time-consuming and human-biased. This paper reviews an exploratory comparison of several machine learning (ML) algorithms for early diagnosis and classification of thyroid diseases based on their ability to automatize and hence the medical diagnosis. Through the comparison of the strengths and weaknesses of various ML methods, we assess them in terms of accuracy, precision, F1 score, and their applicability to clinical use. Our study utilizes datasets containing thyroid-related factors such as age, gender, TSH, T3 followed by feature selection and compares the performance of various ML techniques for thyroid disease. The purpose of this study is to contribute to the expanding literature on how machine learning can be effectively used for diagnosis enhancement of thyroid diseases and classify it into: hypothyroid, hyperthyroid, euthyroid.

Keywords: Machine Learning; Thyroid Disease; Feature Selection; Hypothyroid; Hyperthyroid; Euthyroid

1. Introduction

Thyroid is a small, butterfly-shaped gland located at the front of your neck under your skin. It's a part of your endocrine system (a network of several glands that create and release hormones). Your thyroid's main job is to control the speed of your metabolism, which is the process of how your body transforms the food you consume into energy [1]. This gland produces two active thyroid hormones which are levothyroxine (abbreviated T4) and triiodothyronine (abbreviated T3) [2], commonly known as Thyroxin (T4) and triiodothyronine(T3) [3]. Thyroid hormones control blood pressure, body temperature, and heart rate in addition to aiding in the digestion of fat, protein, and carbs. Conversely, a change in secretion can lead to a number of physical and mental health issues The most common varieties of thyroid are: Hyperthyroidism: An overactive thyroid that produces too much thyroid hormone, Hypothyroidism: An underactive thyroid that doesn't produce enough thyroid hormone, Thyroid tumors and thyroid cancers. The symptoms in Hypothyroidism may involve a person experiencing weight gain, swelling in front of neck and low pulse rate, whereas in hyperthyroidism a person may suffer from elevated blood pressure and pulse rate while having reduced body weight [4].

The majority of people worldwide suffer from thyroid diseases and abnormalities. Early detection and accurate diagnosis are critical for effective treatment and improving patient outcomes. However, conventional diagnostic methods, such as physical examinations, blood tests, and biopsies, often involve time-consuming processes and may be prone to human error or subjective interpretation. The use of machine learning (ML) algorithms to improve the precision, effectiveness, and early identification of thyroid-related disorders is therefore becoming more and more popular. Machine learning, a subset of artificial intelligence, offers significant promise for automating and improving medical diagnostics by identifying patterns in vast datasets. With the rapid advances in computational power and the availability of large-scale healthcare datasets, numerous ML algorithms have been developed and applied to predict and

* Corresponding author: Komal.

classify thyroid diseases. Machine learning algorithms have shown significant potential in improving the early diagnosis of thyroid disease. Several studies have compared the performance of various machine learning techniques for thyroid disease classification, utilizing datasets containing thyroid-related factors such as age, gender, and hormone levels [5].

This paper presents a comparative study of various machine learning algorithms used for the early diagnosis and classification of thyroid diseases. By analyzing the strengths and limitations of different ML approaches, we aim to provide insights into their effectiveness, computational efficiency, and potential for integration into clinical practice. Through this study, we seek to contribute to the growing body of research aimed at leveraging machine learning to improve thyroid disease diagnosis, thereby facilitating timely and accurate medical intervention. The organization of the paper has been done in the subsequent manner. Section 1 represents introduction, Section 2 represent the related work done by the researchers in identification of thyroid disease using ML models. Section 3 focused on the datasets used for the experimental work. Section 4 represents the methodology. Section 5 conclude the review paper.

2. Related Work

A few studies have been performed on thyroid disease, and the authors have evaluated many of the studies to create a proper background on the disease diagnosis.

Khalid salman and Emrullah Sonuç, 2021[2] explores the application of machine learning techniques to classify thyroid conditions into hyperthyroidism, hypothyroidism, and normal categories. Using data from 1,250 Iraqi individuals aged 1 to 90 years, the study analyzed attributes such as T3, T4, and TSH levels. Data preprocessing involved cleaning, handling missing values, and normalization to enhance model performance. Various algorithms, including Support Vector Machines (SVM), Random Forest (RF), Decision Tree (DT), Naïve Bayes (NB), Logistic Regression (LR), k-Nearest Neighbors (kNN), Multi-Layer Perceptron (MLP), and Linear Discriminant Analysis (LDA), were evaluated. Random Forest achieved the highest accuracy (98.93%), demonstrating its effectiveness, while other models like Decision Tree and MLP also performed well. The study highlights the transformative role of machine learning in healthcare, emphasizing its potential to improve diagnostic accuracy and efficiency in managing thyroid diseases.

Hafiz Abbad Ur Rehman, Chyi-Yeu Lin, Zohaib Mushtaq, Shun-Feng Su, 2021[6], evaluates various machine learning classifiers to detect and diagnose thyroid disorders using a dataset from DHQ Teaching Hospital, Pakistan. The dataset includes unique features like pulse rate, blood pressure, and BMI alongside hormonal levels (T3, T4, and TSH). The study employs five algorithms—KNN, Naïve Bayes, SVM, Decision Tree, and Logistic Regression—with and without L1- and L2-based feature selection. Naïve Bayes consistently achieved 100% accuracy, while Logistic Regression and KNN also demonstrated high performance, with Logistic Regression achieving 100% accuracy under L1 feature selection. The results underscore the importance of feature selection and algorithm choice in enhancing classification accuracy. This study contributes to improving thyroid disease diagnosis by integrating advanced feature selection and machine learning techniques.

Gyanendra Chaubey, Dhananjay Bisen, Siddharth Arjaria, Vibhash Yadav, 2021[7], investigates the use of machine learning algorithms—logistic regression, decision trees, and k-nearest neighbors (kNN)—to classify and predict thyroid disease. It employs the “new-thyroid” dataset from the UC Irvine repository, consisting of 215 instances with attributes related to thyroid hormone levels. The study outlines the implementation of each algorithm and compares their performance. Results indicate kNN achieves the highest accuracy (96.88%), followed by decision trees (87.5%) and logistic regression (81.25%). The paper emphasizes the importance of algorithm choice and dataset properties in achieving effective prediction and suggests that future work could involve using larger datasets and advanced techniques like SVMs or deep learning for better accuracy.

Lerina Aversanoa, Mario Luca Bernardia, Marta Cimitileb, Martina Iammarinoa, Paolo Emidio Macchiac, Immacolata Cristina Nettorec, Chiara Verdonea, 2021[8], explores the application of machine learning to predict treatment trends for hypothyroidism patients. The study focuses on improving endocrinologists’ decision-making regarding sodium levothyroxine (LT4) dosage adjustments, which is crucial for treating hypothyroidism. Using a real dataset from Naples “AOU Federico II” hospital, the research combines patient historical and current medical data. It evaluates multiple machine learning classifiers across three preprocessed datasets, employing techniques like interpolation, balancing, discretization, and normalization. Key findings highlight the Extra-Tree Classifier as the most effective, achieving an accuracy and F1-score of 84% when predicting whether LT4 treatment should be increased, decreased, or maintained. This demonstrates the model’s potential as a reliable decision-support tool, enhancing treatment precision and patient outcomes. However, the study acknowledges limitations in dataset size and quality, emphasizing the need for broader data to generalize results further.

Rajasekhar Chaganti, 2022[9], addresses the increasing incidence of thyroid diseases and the need for effective prediction methods. The study critiques existing approaches that primarily focus on binary classification and often utilize small datasets without proper validation. To overcome these limitations, the authors propose a novel multi-class prediction approach that considers five different thyroid disease categories. They employ various feature engineering techniques, including forward feature selection, backward feature elimination, bidirectional feature elimination, and machine learning-based feature selection using extra tree classifiers. The results demonstrate that the extra tree classifier-based selected features yield the highest accuracy of 0.99 when combined with a random forest classifier. The study emphasizes the superiority of machine learning models over deep learning models in terms of accuracy and computational efficiency, particularly in the context of thyroid disease detection. The authors conclude that their approach significantly enhances the prediction of thyroid diseases and highlights the importance of feature selection in improving model performance.

Madhumita Pal; Smita Parija; Ganapati Panda, 2022 [3] explores the application of machine learning (ML) algorithms to predict thyroid disease efficiently and accurately. Thyroid disorders, common among women over 30, can lead to severe health complications, making early detection crucial. The authors focus on leveraging three ML classifiers—K-Nearest Neighbors (KNN), Decision Tree (DT), and Multi-Layer Perceptron (MLP)—to analyze a dataset sourced from the UCI repository with 3163 samples and 24 features. The study highlights the importance of preprocessing, including data cleaning and feature selection using correlation matrices, to minimize overfitting and optimize performance. Each algorithm's performance was evaluated in terms of accuracy and the area under the curve (AUC). Results indicate that MLP outperforms KNN and DT, achieving an accuracy of 95.72% and an AUC of 94.23%. This demonstrates the effectiveness of MLP for predicting thyroid disorders compared to other models.

Kalpna Guleria, Shagun Sharma, Sushil Kumar, Sunita Tiwari [10] investigates the application of machine learning (ML) and deep learning (DL) techniques for diagnosing hypothyroidism. With thyroid disorders affecting millions globally, early detection is crucial to prevent severe health issues such as heart disease, infertility, and obesity. The study employs a dataset with 3772 instances and 30 attributes, utilizing models like Decision Tree, Random Forest, Naive Bayes, and Artificial Neural Networks (ANN). Among the evaluated models, Decision Tree and Random Forest demonstrated the highest accuracy at 99.5758% and 99.3107%, respectively, with minimal error rates. The ANN model achieved competitive accuracy at 93.8226%, though it required more data and computational time due to its complexity. The study performed multiclass classification to identify different types of hypothyroidism (negative, compensated, primary, and secondary), finding Random Forest and Decision Tree most effective across metrics such as precision, recall, and ROC. The paper highlights the importance of preprocessing and feature selection in improving model accuracy and efficiency.

Amulya.R.Rao, B.S.Renuka, 2020[11], discusses a machine learning-based approach for early diagnosis of thyroid disease using classification algorithms like Decision Tree and Naïve Bayes. It highlights the significance of early detection to prevent complications associated with conditions like hyperthyroidism and hypothyroidism, which result from abnormal thyroid hormone levels. The dataset used includes patient attributes such as age, gender, and hormone levels (T3, T4, TSH) and is sourced from the Kaggle platform. The proposed model leverages Decision Tree algorithms to determine whether a patient has thyroid disease. If a positive diagnosis is made, the Naïve Bayes algorithm further categorizes the disease into stages (minor, major, or critical). This system aims to enhance diagnostic accuracy, reduce costs, and save time for both patients and healthcare providers. The document underscores the importance of using robust training datasets and highlights the accuracy of the proposed model, achieving approximately 95% in tests.

Priyanka Duggal, Shipra Shukla, 2020[12], discusses advanced machine learning techniques for predicting thyroid disorders. It emphasizes the importance of early and accurate diagnosis of conditions like hypothyroidism, hyperthyroidism, sick euthyroid, and euthyroid to prevent misdiagnoses and improve patient outcomes. Key machine learning methods reviewed include Naïve Bayes, Support Vector Machines (SVM), and Random Forest, with Recursive Feature Elimination (RFE) identified as the most effective feature selection technique. Among classification algorithms, SVM combined with RFE achieved the highest accuracy of 92.92%. The study highlights the critical role of feature selection and preprocessing in improving model accuracy and efficiency, enabling effective classification of thyroid disorders and supporting healthcare professionals in clinical decision-making.

Table 1 Summary of literature review and the algorithms used with their result

Study	Authors	Reference	Year	Algorithms	Result-Accuracy
1	Khalid salman and Emrullah Sonuç	[6]	2021	• SVM • RF • DT • NB • LR • KNN • MLP • LDA	RF achieved the highest accuracy (98.93%)
2	Hafiz Abbad Ur Rehman, Chyi-Yeu Lin, Zohaib Mushtaq & Shun-Feng Su	[2]	2021	• KNN • NB • SVM • DT • LR	NB achieved the highest accuracy (100%)
3	Gyanendra Chaubey , Dhananjay Bisen, Siddharth Arjaria, Vibhash Yadav	[7]	2021	• LR • KNN • DT	KNN achieved the highest accuracy (96.87%)
4	Lerina Aversanoa, Mario Luca Bernardia, Marta Cimitileb, Martina Iammarinoa, Paolo Emidio Macchiac, Immacolata Cristina Nettorec, Chiara Verdonea	[8]	2021	• DT • KNN • RF • MLP • ET	EXTC achieved the highest accuracy
5	Rajasekhar Chaganti, Furqan Rustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez and Imran Ashraf	[9]	2022	• RF • LR • SVM • ADA • GBM • Deep learning algorithms	RF achieved the highest accuracy (99%)
6	Madhumita Pal; Smita Parija; Ganapati Panda	[3]	2022	• KNN • DT • MLP	MLP performs better with accuracy of 95.73%
7	Kalpna Guleria, Shagun Sharma, Sushil Kumar, Sunita Tiwari	[10]	2022	• DT • RF • NB • ANN	DT provides better results with accuracy of 99.57%
8	Amulya.R.Rao, B.S.Renuka	[11]	2020	• DT • NB	DT and NB together achieved 95% accuracy

9	Priyanka Duggal, Shipra Shukla	[12]	2020	• SVM • NB • RF	SVM provides better results with accuracy of 92.92%

SVM - Support Vector Machine	MLP - Multi-Layer Perceptron
RF - Random Forest	LDA - Linear Discriminant Analysis
DT - Decision Tree	EXTC – Extra Trees Classifier
NB - Naïve Bayes	ANN - Artificial Neural Networks
LR - Logistic Regression	KNN - k-Nearest Neighbours

2.1. Description of the dataset

The Kaggle Machine Learning Website is where the Thyroid Dataset is obtained [13]. Different attributes of patients are been used in the database to identify thyroid in patient. Age, gender, hyperthyroidism, hypothyroidism, pregnancy, T3, T4, and TSH levels are the primary characteristics taken into account.

Table 2 dataset description

Sr. No.	Attribute name	Value Type	Clarification
1	Age	number	10, 11, 23…
2	Sex	String	M=Male, F=Female
3	on_thyroxine	Boolean	t=true, f=false
4	query_on_thyroxine	Boolean	t=true, f=false
5	on_antithyroid_meds	Boolean	t=true, f=false
6	Sick	Boolean	t=true, f=false
7	Pregnant	Boolean	t=true, f=false
8	thyroid_surgery	Boolean	t=true, f=false
9	I131_treatment	Boolean	t=true, f=false
10	query_hypothyroid	Boolean	t=true, f=false
11	query_hyperthyroid	Boolean	t=true, f=false
12	Lithium	Boolean	t=true, f=false
13	Goitre	Boolean	t=true, f=false
14	Tumor	Boolean	t=true, f=false
15	hypopituitary	Boolean	t=true, f=false
16	Psych	Boolean	t=true, f=false
17	TSH_measured	Boolean	t=true, f=false
18	TSH	number	0.3, 1, 1.2…..
19	T3_measured	Boolean	t=true, f=false
20	T3	number	1.1, 2.3…….
21	TT4_measured	Boolean	t=true, f=false
22	TT4	number	76, 83, 123….
23	T4U_measured	Boolean	t=true, f=false
24	T4U	number	0.94, 1.06…..

25	FTI_measured	Boolean	t=true, f=false
26	FTI	number	45, 47, 84…..
27	TBG_measured	Boolean	t=true, f=false
28	TBG	number	11, 26, 36……
29	referral_source	String	
30	Target	String	
31	patient_id	number	840801013……..

2.2. Data pre-processing

Since the pre-processing procedure is used to analyze the data, it has a positive impact on the data and is a crucial stage in data mining. It identifies the missing data. Data preparation, cleansing, and other tasks are part of the pre-processing procedure. In this stage data was cleaned and organized. We found a set of missing data in attributes T4, TSH, T3 where we were able to process this lost data by substituting the mediator's value. After doing this, we were able to get the data in a better and more efficient manner.

3. Methodology

The study incorporates the dataset as shown in table 2. After that data is been pre-processed for a clean and organized dataset. The pre-processed data is been passed through different ML algorithms: Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), k-Nearest Neighbors (kNN) and Linear Discriminant Analysis (LDA), algorithm to generate the results.

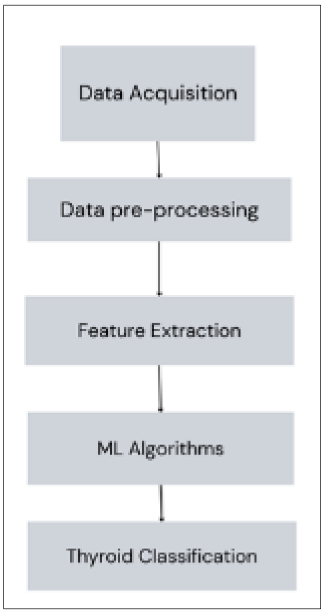

Figure 1 Thyroid Classification Methodology

3.1. Support Vector Machines

Support Vector Machines (SVMs) have emerged as a powerful tool in thyroid prediction, offering a robust approach to classify and predict thyroid disorders based on various clinical and laboratory parameters. SVMs handle high-dimensional data and can effectively separate complex decision boundaries, making them particularly suitable for analyzing the multifaceted nature of thyroid function. By mapping input features into a higher-dimensional space, SVMs can identify optimal hyperplanes that distinguish between different thyroid conditions, such as hypothyroidism, hyperthyroidism. The ability of SVMs to handle non-linear relationships and their resistance to overfitting contribute to their accuracy in thyroid prediction models. Additionally, Support Vector Machines (SVMs) can utilize kernel functions to identify complex patterns within thyroid-related datasets, allowing for more sophisticated and accurate predictions. As a result, SVM-based models have demonstrated promising results in early detection and risk assessment of thyroid disorders, potentially improving diagnostic accuracy and facilitating timely interventions in clinical practice.

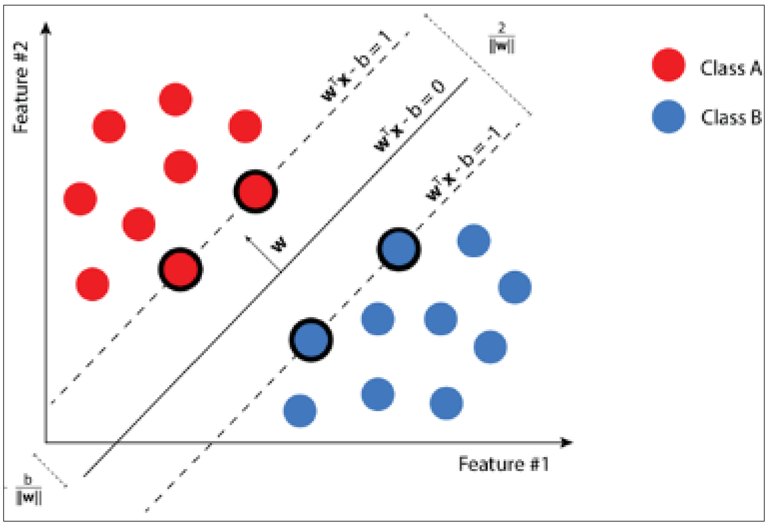

Figure 2 Support Vector Machine [14]

3.2. Random forest

The Random Forest algorithm is a robust machine learning technique widely used in predicting thyroid diseases. By building multiple decision trees and combining their outputs, it provides accurate and reliable multi-class classification, effectively distinguishing conditions like hyperthyroidism, hypothyroidism, and euthyroidism. It also highlights critical features such as TSH, T3, T4, and Free T4 levels, aiding clinicians in decision-making.

This algorithm excels due to its high accuracy, resistance to overfitting, and ability to handle noisy and large datasets. It can also manage missing data effectively through imputation and offers interpretability by ranking feature importance. In practice, thyroid disease prediction involves collecting patient data, preprocessing it to handle missing values and normalize inputs, and training the Random Forest model. The model's performance is validated using metrics like accuracy and sensitivity before deployment for analyzing new patient data.

3.3. Decision tree

The Decision Tree algorithm is a simple yet effective machine learning method for identifying thyroid disorders like hypothyroidism, hyperthyroidism, and euthyroidism. It works by creating a flowchart-like structure, where decisions are made based on thresholds or conditions related to features such as TSH, T3, T4, and Free T4 levels. This approach makes Decision Trees highly interpretable, enabling clinicians to visualize the decision-making process and understand the factors influencing diagnoses. Decision Trees handle non-linear relationships and mixed data types well, making them suitable for thyroid identification. They require minimal preprocessing and work effectively with small to medium-sized datasets. In practice, the model splits data into branches based on thresholds that maximize information gain or minimize entropy, classifying conditions accordingly.

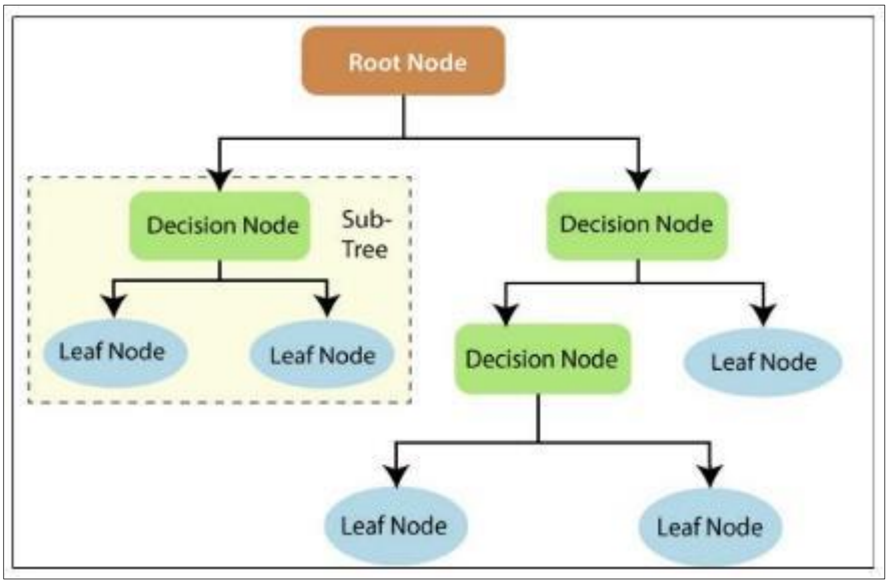

Figure 3 Decision tree [15]

3.4. k-Nearest Neighbors

The k-Nearest Neighbors (kNN) algorithm is a simple yet effective machine learning method used in thyroid diagnosis. It classifies data points based on the majority class among the k closest data points in the feature space, making it well-suited for identifying thyroid disorders like hypothyroidism, hyperthyroidism, and euthyroidism. The algorithm relies on distance metrics (e.g., Euclidean, Manhattan) to measure similarity between instances. In thyroid diagnosis, kNN leverages patient data, such as TSH, T3, T4 levels, and clinical symptoms, to predict the likelihood of a thyroid condition. It is particularly effective for datasets where class separability is evident and works well with both binary and multi-class classification tasks. One of its strengths is its simplicity and minimal assumption about data distribution, making it highly adaptable. However, kNN is sensitive to the choice of k (the number of neighbors) and the scaling of features, which requires preprocessing steps like normalization. Additionally, its computational cost increases with large datasets.

3.5. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a statistical method used in thyroid disease diagnosis for dimensionality reduction and classification. It projects data onto a lower-dimensional space while maximizing class separation, making it effective for identifying conditions like hypothyroidism, hyperthyroidism, and euthyroidism. LDA leverages features such as TSH, T3, and T4 levels to distinguish patterns among thyroid conditions. Its strengths include simplicity, interpretability, and efficiency, especially with linearly separable classes. However, it assumes Gaussian data distribution and struggles with non-linear relationships or imbalanced datasets, requiring preprocessing like feature scaling and data balancing. By reducing noise and highlighting relevant features, LDA improves diagnostic accuracy and is often combined with other techniques for enhanced thyroid disorder classification and risk assessment.

4. Conclusion

Thyroid disease is one of the illnesses that affect people all over the world, and its prevalence is rising. The medical reports indicate significant imbalances in thyroid disorders, so our study focuses on the identification and differentiation of thyroid in categories: hyperthyroid, hypothyroid and euthyroid. Machine learning showed us good results using several algorithms, it got observed that using feature selection the accuracy of algorithms increased. Also a few datasets included some unique features: BMI, blood pressure, pulse rate, diabetes which enhanced the performance of the algorithms. Using feature selection, it got identified that TSH, TT4, T3, T4U, FTI, age and sex have been considered as important features for Thyroid disorder identification. Feature selection not only increases accuracy of algorithms but also reduces overfitting, training time.

References

[1] Preethiya, T., Tyagi, A.K., Pandiarajan, T. (2024). Comparative Analysis of Machine Learning Algorithms in Thyroid Disease Prediction. In: Abraham, A., Bajaj, A., Hanne, T., Siarry, P. (eds) Intelligent Systems Design and Applications. ISDA 2023. Lecture Notes in Networks and Systems, vol 1046. Springer, Cham.

[2] Abbad Ur Rehman, H., Lin, CY., Mushtaq, Z. et al. Performance Analysis of Machine Learning Algorithms for Thyroid Disease. Arab J Sci Eng 46, 9437–9449 (2021).

[3] M. Pal, S. Parija and G. Panda, "Enhanced Prediction of Thyroid Disease Using Machine Learning Method," 2022 IEEE VLSI Device Circuit and System (VLSI DCS), Kolkata, India, 2022, pp. 199-204

[4] Pal, R.; Anand, T.; Dubey, S.K.: Evaluation and performance analysis of classification techniques for thyroid detection. Int. J. Bus. Inf. Syst. 28(2), 163–177 (2018)

[5] P. Kumari, B. Kaur, A. K. Rawat and M. Rakhra, "Role of Machine Learning in the Prediction of Thyroid Disease," 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), Greater Noida, India, 2024, pp. 1-5

[6] Salman, Khalid & Sonuç, Emrullah. (2021). Thyroid Disease Classification Using Machine Learning Algorithms. Journal of Physics: Conference Series. 1963. 012140. 10.1088/1742-6596/1963/1/012140.

[7] Chaubey, G., Bisen, D., Arjaria, S. et al. Thyroid Disease Prediction Using Machine Learning Approaches. Natl. Acad. Sci. Lett. 44, 233–238 (2021).

[8] Lerina Aversano, Mario Luca Bernardi, Marta Cimitile, Martina Iammarino, Paolo Emidio Macchia, Immacolata Cristina Nettore, Chiara Verdone, Thyroid Disease Treatment prediction with machine learning approaches, Procedia Computer Science, Volume 192, 2021, Pages 1031-1040, ISSN 1877-0509

[9] Chaganti, R.; Rustam, F.; De La Torre Díez, I.; Mazón, J.L.V.; Rodríguez, C.L.; Ashraf, I. Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques. Cancers 2022, 14, 3914.

[10] Kalpna Guleria, Shagun Sharma, Sushil Kumar, Sunita Tiwari, Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning, Measurement: Sensors, Volume 24, 2022, 100482, ISSN 2665-9174

[11] A. R. Rao and B. S. Renuka, "A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4

[12] P. Duggal and S. Shukla, "Prediction Of Thyroid Disorders Using Advanced Machine Learning Techniques," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 670-675

[13] https://www.kaggle.com/datasets/amirmohammadparvizi/thyroid-disease-detection-dataset

[14] Pecha, Marek & Horák, David. (2020). Analyzing l1-loss and l2-loss Support Vector Machines Implemented in PERMON Toolbox. 10.1007/978-3-030-14907-9_2.

[15] Classification of Adult Income Using Decision Tree - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Structure-of-Decision-Tree-Algorithm_fig3_370770672