IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 7, July 2025

DOI: 10.17148/IJARCCE.2025.14733

A Comparative Analysis of Machine Learning for the Classification of Thyroid Dysfunction

Anup Kumar[1], Suryakant Pathak[2], Varun Bansal[3]

Research Scholar, Department of Computer Science & Engineering, Shobhit University, Gangoh, Saharanpur India[1]

Professor, Department of Computer Science & Engineering, Shobhit University, Gangoh, Saharanpur India[2]

Professor, Department of Computer Science & Engineering, Shobhit University, Gangoh, Saharanpur, India[3]

Abstract: The thyroid gland, one of the body's key endocrine organs, produces two essential hormones that regulate metabolic activity. Abnormal functioning of this gland can lead to disorders such as hypothyroidism and hyperthyroidism, both of which significantly disrupt the body's normal physiological processes. Although thyroid disorders are generally diagnosed through blood tests, these tests often yield ambiguous or noisy results, making accurate diagnosis difficult. To address this challenge, the present study incorporates data cleaning methods and machine learning techniques to enhance the accuracy of thyroid disease detection and prediction. Clean and structured data improved the reliability of the analysis. Various machine learning algorithms, including logistic regression, decision trees, k-nearest neighbors (KNN), support vector machines (SVM), XG Boost, and artificial neural networks (ANN), were employed to model and predict.

Keywords: hypothyroidism, hyperthyroidism, diagnosed, machine learning algorithms decision tree, KNN, SVM, ANN

I. INTRODUCTION

Approximately four out of ten people are shrinking from thyroid disorders, which are becoming increasingly prevalent among women. A vital endocrine gland, the thyroid coordinates several kinds of metabolic activities, which includes the converting of food into energy. Hypothyroidism and hyperthyroidism are the two main classes into which thyroid dysfunction usually looks classified.

Symptoms of hypothyroidism include low blood pressure, decreased sweating, hair loss, fatigue, and even mental health distrub like depression. On the other hand, people who have hyperthyroidism frequently complain of increased sweating, irritability, blood pressure, muscle weakness, and general discomfort.

A few dietary and lifestyle reforms can help manage thyroid-related problems, even though clinical treatment is frequently required. These include making sure the results are enough protein and fibre, as well as increasing your consumption of green leafy vegetables.

II. REVIEW OF LITERATURE

Ref. No.	Methodology / Tools	Objective	Research Gap	Publisher
1.	XG Boost Algorithm and hyper-parameter optimization techniques	An Intelligent Machine Learning Model for Thyroid Disease Early Detection	To further enhance prediction capabilities, more varied datasets and machine learning techniques are attached.	SEEJPH
2.	Differential evolution (DE)-based optimization algorithm	A Better Machine Learning Model Based on Differential Evolution for Thyroid Disease Detection	There is still much to learn about the problem of class imbalance in datasets. Few studies combine sophisticated methods like conditional GANs for data augmentation or differential evolution for model optimization. This leaves room for more thorough models that successfully handle data imbalance and multi class	Springer

DOI: 10.17148/IJARCCE.2025.14733

			classification.	
3.	Forward feature selection, backward feature elimination, bidirectional feature elimination, and machine learning-based feature selection using extra tree classifiers are adopted.	Non-thyroidal syndrome (NTIS), autoimmune thyroiditis (compensated hypothyroid), binding protein (increased binding protein), and primary Hashimoto's thyroiditis can all be predicted using the suggested method. hypothyroid).	Because we relate to the 5-class classification problem and feature reduction as the research findings, we intend to expand the number of classes in our subsequent work.	Scopus
4.	Machine Learning algorithms, SVM - support vector machine, decision tree, logistic regression, KNN - K-nearest neighbours, ANN- Artificial Neural Network	To develop a machine learning-based system that uses cleaned blood test data to predict thyroid disorders (hypothyroidism and hyperthyroidism) and to build a web application for user input and real-time disease prediction.	Thyroid nodules and cancer that cannot be detected in blood test results can be predicted with image processing of ultrasonic scanning of thyroid images. All thyroid-related diseases can be predicted by combining the two results.	IJEAST
5.	Support vector machines, random forest, decision tree, naive bayes, logistic regression, k-nearest neighbors, multilayer perceptron (MLP), linear discriminant analysis.	This study aims to distinguish between three types of thyroid disease: hyperthyroidism, hypothyroidism, and normal.	To improve prediction accuracy and practical applicability in thyroid disease detection, more research is required on robust feature engineering, model comparison, and validation techniques.	IOP
6.	Support Vector Machine, Decision tree, Logistic Regression, K-nearest neighbor, Naive Bayes.	To investigate diseases based on pertinent disease parameters and efficiently classify them using sophisticated classification algorithms like Naive Bayes, K-Nearest Neighbour (KNN), Decision Tree, Support Vector Machine (SVM), and Logistic Regression. By training and testing these models on disease-specific datasets, the goal is to achieve accurate disease prediction.	Future worlds for improved disease prediction analysis are being built by fields like machine learning, artificial intelligence, and deep learning.	IJET

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 7, July 2025

DOI: 10.17148/IJARCCE.2025.14733

7.	Sampling Methodology , SVM Model, Development of website using PYTHON FLASK given an interactive user interface	1) To comprehend the health concerns of the patient in light of thyroid issues 2) To determine the expert system's parameters for thyroid disease prediction. 3) To research the thyroid disease prediction expert systems currently in use. 4) To forecast thyroid conditions based on patient blood tests and symptoms. 5) To create and implement an expert system for thyroid prediction disease.	1. Absence of lifestyle-based or region-specific data 2. Strategies for Early Screening and Awareness 3. Prediction models with limited personalisation 4. Monitoring and Longitudinal Tracking 5. Evaluation of ML Algorithms in Practical Clinical Contexts 6. Health Information System Integration	ABCD Index
8.	Ensemble ML classifier (hard voting), XGBost, SelectKBest	Since thyroid conditions are becoming more common worldwide, early detection is essential to lowering complications and death rates. In order to accurately predict thyroid disease, this study applies standard machine learning models enhanced with ensemble learning techniques to analyse clinical symptoms and address important diagnostic challenges.	To determine the model's fairness, accuracy, transparency, and results, this study can be expanded to the classification of thyroid disease using an interpretable machine-learning approach. Additionally, this research can be expanded using a self-developed dataset.	Scopus
9.	Evaluated seventeen machine learning models and implemented an Ensemble ML classifier using a hard voting strategy.	By combining clinical feature analysis with an ensemble learning approach, this study seeks to enhance the early detection of thyroid disorders through machine learning. The suggested model outperforms current approaches with 100% sensitivity and 97.27% accuracy after evaluating seventeen models and utilising strategies like random oversampling, XGBoost, and SelectKBest.	Research will concentrate on creating more ensemble-based classifiers for AI applications and investigate sophisticated class balancing strategies from an AI standpoint.	Scopus
10.	Semi-supervised learning Methods, namely FixMatch, Co-training, and self-training, in conjunction with supervised learning algorithms, specifically Naive Bayes and logistic regression.	The objective was to get around the drawbacks of traditional diagnostic methods, which might not be able to identify minute thyroid hormone imbalances.	The purpose of this research is to use publicly accessible datasets to create a model for the classification of thyroid diseases. In the future, wearable technology will be used to gather thyroid-related data, such as pictures, and sophisticated deep learning methods (such as CNNs, LSTM, and RNNs) will be used to improve diagnosis. To find the best method for thyroid monitoring, a comparison of deep learning and semi-supervised learning (SSL) will be done.	scopus

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 7, July 2025
DOI: 10.17148/IJARCCE.2025.14733

III. OBJECTIVE

1. To comprehend the health concerns of the patient in light of thyroid issues
2. To determine the expert system's parameters for thyroid disease prediction.
3. To research the thyroid disease prediction expert systems currently in use.
4. To forecast thyroid conditions based on patient blood tests and symptoms.
5. To create and implement an expert system for thyroid disease prediction.
6. To assess how well different machine learning algorithms detect thyroid conditions.
7. Determine which biomarkers—such as TSH levels and antibodies—are most important for predicting thyroid disease.

IV. CONCLUSION

The research revealed that, the thyroid disease is mostly affected in females than males. Additionally, studies showed that thyroid function is not inherited. Thyroid disorders are more common in people aged 15 to 25. The researcher added that complications can be reduced with the aid of a healthy diet and regular exercise. A simple way to predict thyroid disease and symptoms of thyroid disease detection using intelligent machine learning Scientists in their research have used different algorithms to help machine learn to predict thyroid. What is the diagnosis of thyroid disease, but we have seen that all the scientists checked the data and picked up the data set from one city or one place and did it which is not enough. So after reading the research paper we came to know that the data used is not enough. Now our goal is to make such a model or album which can work in the setting of thyroid disease and also diagnose thyroid disease.

REFERENCES

[1]. M.V. Sheela Devi , Dr. Parveen Kumar (2025): An Intelligent Machine Learning Model for Early Detection of Thyroid Disease
[2]. Punit Gupta, Furqan Rustam, Khadija Kanwal, Wajdi Aljedaani, Sultan Alfarhood, Mejdl Safran, Imran Ashraf (2024): Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution
[3]. Rajasekhar Chaganti, Furqan Rustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez, and Imran Ashraf (2022): Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques
[4]. Chandan R, Chetan Vasan, Chethan MS, Devikarani H S (2021): Thyroid Detection Using Machine Learning
[5]. Khalid salman, Emrullah Sonuc (2021): Thyroid Disease Classification Using Machine Learning Algorithms
[6]. E. Laxmi Lydia, N. Sharmil, K. Shankar and Andino Maseleno (2019): Analysing the Performance of Classification Algorithms on Diseases Datasets
[7]. M.V. Sheela Devi, Dr. Parveen Kumar (2024): An Innovative Machine Learning Approach for Early Detection of Thyroid Disorders
[8]. Khandaker Mohammad Mohi Uddin, Abdullah Al Mamunb, Anamika Chakrabarti, Rafid Mostafiz (2024): An ensemble machine learning-based approach to predict thyroid disease using hybrid feature selection
[9]. Md. Rezaul Islam, Aniruddha Islam Chowdhury, Sharmin Shama, Md. Masudul Hasan Lamyea (2025): Enhanced thyroid disease prediction using ensemble machine learning: a high-accuracy approach with feature selection and class balancing
[10]. Melih Agraz (2024): Semi-Supervised Machine Learning Approaches for Thyroid Disease Prediction and its Integration Withthe Internet of Everything