Adaptive Intelligence for Early Thyroid Disease Detection and Classification

B Sai Lokesh[1], M. Anusha[2]
[1]MCA Student, Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad
[2]Assistant Professor, Dept of MCA , Ellenki College of Engineering and Technology , Hyderabad

To Cite this Article
B Sai Lokesh, M. Anusha, "Adaptive Intelligence for Early Thyroid Disease Detection and Classification", Journal of Science Engineering Technology and Management Science, Vol. 03, Issue 07, July 2026, pp: 957-963, DOI: http://doi.org/10.64771/jsetms.2026.v03.i07.pp957-963
Submitted: 19-06-2026 *Accepted: 24-07-2026* *Published: 30-07-2026*

***Abstract*— Adaptive Intelligence for Early Thyroid Disease Detection and Classification presents an intelligent decision-support system designed to assist in the early identification of thyroid disorders using machine learning techniques. The proposed application provides separate modules for administrators and users, enabling efficient dataset management, model training, and disease prediction through a user-friendly interface. The thyroid dataset contains multiple clinical attributes associated with three diagnostic categories: Normal, Hyperthyroid, and Hypothyroid. Before model development, the data undergoes preprocessing, including cleaning, label encoding, and feature transformation to ensure compatibility with machine learning algorithms. Several classification models, including Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, and Multi-Layer Perceptron, are trained and evaluated using performance measures such as accuracy, precision, recall, and F1-score. Comparative analysis identifies Random Forest and Decision Tree as the most effective models for thyroid disease classification. The system also allows users to predict thyroid conditions by entering clinical values or uploading test data. Developed using Python, Django, and MySQL, the proposed framework simplifies disease assessment, supports early diagnosis, reduces manual effort, and provides a reliable platform for intelligent thyroid disease detection.**

***Keywords*— Thyroid Disease Detection, Machine Learning, Random Forest, Decision Tree, Thyroid Classification, Early Disease Prediction, Clinical Decision Support, Data Preprocessing, Predictive Analytics, Healthcare Informatics**

I. INTRODUCTION

Thyroid disorders are among the most common endocrine diseases and affect millions of people worldwide. The thyroid gland, located at the front of the neck, produces hormones such as triiodothyronine (T3) and thyroxine (T4), while thyroid stimulating hormone (TSH) regulates their production [3]. These hormones control important body functions, including metabolism, heart rate, body temperature, and energy balance. Any imbalance in hormone production may result in hypothyroidism or hyperthyroidism, leading to various health complications [10]. The symptoms of thyroid disease often develop gradually and resemble those of other medical conditions, making early diagnosis difficult. Delayed detection may increase the risk of severe complications and reduce treatment effectiveness. Therefore, accurate and timely identification of thyroid disorders is essential for improving patient health and supporting appropriate medical intervention [2].

Early detection of thyroid disease is important because untreated thyroid disorders can affect several body systems and reduce a person's quality of life. Hyperthyroidism commonly causes weight loss, rapid heartbeat, excessive sweating, and nervousness, whereas hypothyroidism may lead to fatigue, weight gain, dry skin, and slow heart rate [15]. Physicians generally diagnose thyroid disorders using laboratory tests that measure T3, T4, and TSH hormone levels along with clinical examination. However, interpreting these medical parameters can be difficult because many symptoms overlap with other diseases [11]. As a result, traditional diagnostic methods may require additional testing and expert evaluation. Intelligent computer-based systems can support clinicians by analysing medical data efficiently and providing reliable disease predictions for better clinical decision-making [1].

The increasing availability of electronic medical records has encouraged the use of machine learning techniques for disease prediction. Machine learning algorithms can analyse large clinical datasets, identify hidden relationships, and generate predictive models that improve diagnostic accuracy [4]. Before model training, the collected data undergo preprocessing, including data cleaning, feature selection, encoding categorical values, and normalization to improve data quality. These steps help algorithms learn meaningful patterns and reduce prediction errors. Compared with manual analysis, machine learning provides faster and more consistent classification results by processing multiple clinical attributes simultaneously [9]. Consequently, intelligent predictive models have become valuable tools for assisting healthcare professionals in diagnosing thyroid disorders more accurately and efficiently [12].

Several machine learning algorithms have demonstrated promising performance in thyroid disease classification. Methods such as Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, Naïve Bayes, K-Nearest Neighbors, and Artificial Neural Networks have been widely used for analysing thyroid datasets [5]. Their performance is commonly evaluated using accuracy, precision, recall, F1-score, and confusion matrix to identify the most reliable classification model. Decision Tree provides simple and interpretable decision rules, while Random Forest improves prediction accuracy by combining multiple decision trees [7]. Comparative analysis of these algorithms helps researchers select the most suitable model for accurate thyroid disease prediction and supports the development of efficient clinical decision-support systems [13].

This study proposes an Adaptive Intelligence for Early Thyroid Disease Detection and Classification framework using multiple machine learning algorithms for accurate thyroid disease prediction. The system performs data preprocessing, model training, algorithm comparison, and disease classification through a structured workflow. Clinical records are categorized into Normal, Hyperthyroid, and Hypothyroid classes based on important medical attributes. The trained models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score to determine the most effective classifier [14]. The application also allows users to enter medical values or upload test datasets for disease prediction. By integrating intelligent machine learning techniques with clinical data analysis, the proposed framework supports early diagnosis, reduces manual effort, and provides an efficient decision-support tool for healthcare professionals managing thyroid disorders [16].

II. RELATED WORK

Azar et al., [2012] [1] developed an expert system for thyroid disease diagnosis by combining neural networks with fuzzy logic to improve diagnostic accuracy. Their approach was designed to handle the uncertainty that commonly exists in medical data and physician decision-making. The proposed system analyzed patient clinical parameters and generated diagnostic results using fuzzy inference rules supported by neural learning techniques. The authors demonstrated that integrating artificial intelligence with expert knowledge could improve the consistency and reliability of thyroid disease diagnosis. Their study also emphasized that intelligent systems can reduce dependence on manual interpretation while supporting physicians in making accurate clinical decisions. Experimental results indicated that the hybrid model produced encouraging prediction performance when compared with traditional diagnostic methods. This research highlights the importance of combining machine learning with expert systems to improve healthcare applications and provides a useful foundation for developing automated thyroid disease classification systems.

Shukla and Kaur, [2009] [5] presented a thyroid disease diagnosis system based on Artificial Neural Networks (ANN) to improve the identification of thyroid disorders using patient clinical data. Their study focused on training the neural network with medical attributes related to thyroid function so that it could recognize complex relationships among different health parameters. The authors explained that neural networks are capable of learning from historical medical records and generating reliable predictions without requiring manually defined rules. The experimental findings showed that the ANN model achieved good classification accuracy and successfully distinguished different thyroid conditions. The research also highlighted that proper preprocessing and training significantly influence the performance of predictive models. Their work demonstrated the practical value of neural network techniques in healthcare and encouraged the use of intelligent learning models for supporting physicians in disease diagnosis. This study serves as an important reference for machine learning applications in thyroid disease prediction.

Banu, [2016] [7] investigated the application of the Decision Tree classification algorithm for diagnosing thyroid disease through data mining techniques. The study aimed to develop an interpretable classification model capable of predicting thyroid disorders using clinical information collected from patients. The Decision Tree algorithm generated simple decision rules that made the prediction process easy to understand for healthcare professionals.

The author reported that the model achieved reliable classification performance while maintaining computational efficiency. The research emphasized that Decision Trees can effectively process both categorical and numerical medical data without requiring complicated mathematical models. In addition, the study highlighted that rule-based classification improves transparency and supports better clinical interpretation compared with many black-box techniques. The findings demonstrated that Decision Tree models provide a practical solution for thyroid disease diagnosis and can assist doctors in making faster and more accurate medical decisions while reducing manual analysis.

Chantel et al., [2016] [13] carried out a comparative study to evaluate the performance of K-Nearest Neighbor (KNN) and Naïve Bayes algorithms for thyroid disease detection. Their research analyzed clinical datasets containing various thyroid-related attributes to determine which classification technique produced better prediction results. The authors compared both algorithms using standard evaluation measures and discussed their strengths and limitations. Experimental results indicated that each algorithm offered useful classification capabilities, although prediction accuracy varied depending on the dataset characteristics and feature distribution. The study highlighted the importance of selecting appropriate machine learning algorithms based on the nature of medical data rather than relying on a single classification method. It also emphasized that comparative analysis helps researchers identify the most suitable model for disease prediction. Their work provides valuable guidance for future studies focused on developing efficient and accurate thyroid disease classification systems using machine learning.

Umar Sidiqi et al., [2019] [15] explored the use of several data mining classification techniques for diagnosing different thyroid disorders. Their study focused on comparing multiple machine learning algorithms to determine which approach could provide the highest prediction accuracy using patient clinical records. The researchers examined the influence of preprocessing, feature selection, and classifier choice on the overall performance of thyroid disease prediction models. The comparative evaluation demonstrated that machine learning algorithms can effectively recognize disease patterns and assist healthcare professionals in making timely clinical decisions. The authors also noted that accurate classification depends not only on the algorithm itself but also on the quality of the training data and preprocessing methods. Their findings confirmed that intelligent data mining techniques reduce diagnostic complexity and improve prediction reliability. This research supports the growing adoption of artificial intelligence in healthcare by demonstrating the practical benefits of automated thyroid disease diagnosis and clinical decision support systems.

III. DATASET DETAILS

The dataset used in this project consists of clinical records collected for the identification and classification of thyroid disorders. It contains multiple medical attributes related to thyroid function, including laboratory test values and patient health information, along with the corresponding disease categories Normal, Hyperthyroid, and Hypothyroid. These attributes provide important information about hormone activity and other clinical indicators required for accurate diagnosis. The data are organized in a structured tabular format, making them suitable for machine learning analysis and disease classification. Before model development, the dataset is carefully examined to identify missing values, inconsistent records, and unnecessary information that could affect prediction performance. Since the dataset contains both numerical and categorical attributes, appropriate preprocessing techniques are applied to ensure data consistency. This well-prepared dataset serves as the foundation for training intelligent machine learning models capable of accurately classifying different thyroid conditions.

To improve the quality of the dataset, several preprocessing operations are performed before model training. Categorical attributes are converted into numerical values using Label Encoder techniques so that machine learning algorithms can process them efficiently. The dataset is then cleaned, normalized, and shuffled to remove bias and improve model learning. After preprocessing, the complete dataset is divided into 70% training data and 30% testing data to evaluate model performance on unseen samples. This data preparation process helps improve prediction accuracy, reduces classification errors, and enables reliable comparison among different machine learning algorithms. Proper dataset preprocessing plays an important role in developing an efficient thyroid disease classification system.

IV. PROPOSED METHODOLOGY

The proposed system follows a structured approach to perform early thyroid disease detection and classification using machine learning algorithms. Initially, the thyroid dataset containing multiple clinical attributes and disease

categories is uploaded and examined. Data preprocessing is then carried out to improve the quality of the dataset before model training. This process includes handling missing values, converting categorical attributes into numerical values using Label Encoder, and organizing the data into a consistent format suitable for machine learning algorithms. The preprocessed dataset is shuffled to ensure balanced distribution of disease classes and then divided into training and testing datasets. These preprocessing steps help the models learn meaningful patterns from the clinical data and improve prediction reliability.

After preprocessing, several machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Naïve Bayes, and Multi-Layer Perceptron (MLP), are trained using the prepared dataset. Each model is evaluated using accuracy, precision, recall, and F1-score to compare their classification performance. The application also generates comparison graphs to visualize the effectiveness of all algorithms and identify the best-performing model. Finally, the trained model is used to predict thyroid disease by allowing users to either enter patient clinical values manually or upload a test dataset. The system classifies each record into Normal, Hyperthyroid, or Hypothyroid, providing an efficient and reliable solution for early thyroid disease diagnosis.

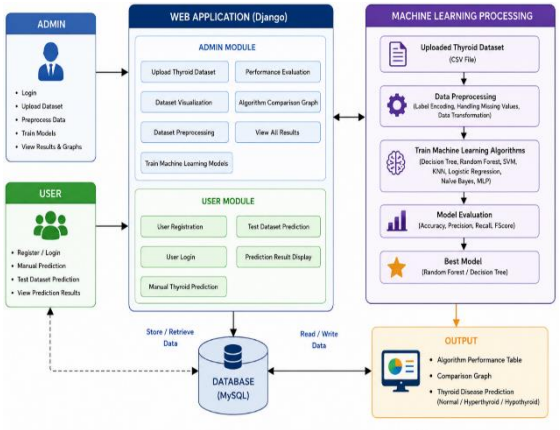

Figure [1]: Thyroid Disease Classification System

Figure [1] presents the workflow of the proposed thyroid disease classification system. The thyroid dataset is uploaded and preprocessed before training multiple machine learning algorithms. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score to identify the best-performing classifier. Finally, the system predicts whether a patient belongs to the Normal, Hyperthyroid, or Hypothyroid category and displays the prediction results along with performance comparisons.

V. RESULT AND DISCUSSION

The experimental evaluation demonstrates that the proposed thyroid disease classification system can accurately identify different thyroid conditions using machine learning techniques. After completing data preprocessing, including cleaning, label encoding, and feature preparation, the dataset was divided into training and testing sets for model development. Several classification algorithms, namely Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, Naïve Bayes, and Multi-Layer Perceptron (MLP), were trained and compared. Their performance was assessed using accuracy, precision, recall, and F1-score. Among the evaluated models, Random Forest and Decision Tree achieved the highest prediction performance and produced more reliable classification results than the remaining algorithms. The confusion matrices indicated that most patient records were classified correctly with only a few prediction errors. Performance comparison graphs further confirmed the consistency of these models across different evaluation metrics. The prediction module also successfully classified new patient records into Normal, Hyperthyroid, and Hypothyroid categories. These findings demonstrate that the developed system provides accurate, consistent, and efficient support for thyroid disease diagnosis using clinical data.

Figure [2]: Machine Learning Model Performance

Figure [2] presents the performance evaluation of the trained machine learning algorithms used for thyroid disease classification. The table compares Decision Tree, KNN, Logistic Regression, MLP, Naïve Bayes, Random Forest, and SVM using accuracy, precision, recall, and F1-score. The results show that Decision Tree and Random Forest achieved the highest overall performance, while the remaining algorithms also produced reliable classification results. This comparison helps identify the most suitable model for accurate thyroid disease prediction.

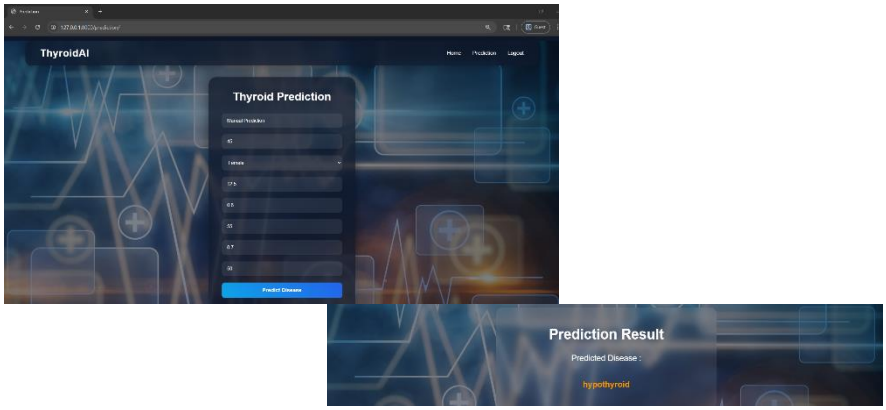

Figure [3]: Thyroid Disease Prediction Results

Figure [3] shows the thyroid disease prediction process using manually entered patient clinical values. The user enters the required medical parameters through the prediction form, and the trained machine learning model analyzes the input data. The system then displays the predicted thyroid condition, which in this case is Hypothyroid. This module provides a quick and reliable way to classify thyroid disorders and supports early clinical decision-making.

Figure [4]: Thyroid Dataset Prediction Results

Figure [4] displays the prediction results generated after testing the uploaded thyroid dataset. The system analyzes clinical attributes such as Age, Gender, TSH, T3, TT4, T4U, and FTI, and classifies each record as Normal, Hyperthyroid, or Hypothyroid. The results demonstrate the ability of the trained machine learning model to accurately predict thyroid conditions for multiple patient records, supporting efficient large-scale disease screening and clinical analysis.

DISCUSSION

The outcomes of this study highlight the effectiveness of machine learning in supporting the early detection of thyroid disorders. Proper dataset preparation, including preprocessing, label encoding, and feature transformation, played a significant role in improving the quality of the training data and enhancing model performance. The

comparative analysis revealed that Random Forest and Decision Tree outperformed the other classification algorithms because they effectively captured the relationships among clinical attributes while maintaining stable prediction accuracy. Although algorithms such as SVM, KNN, Logistic Regression, Naïve Bayes, and MLP also produced satisfactory results, their performance was comparatively lower for this dataset. The use of evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices provided a comprehensive understanding of each model's strengths and limitations. In addition, graphical comparisons made it easier to identify the most suitable algorithm for thyroid disease classification. Overall, the proposed system demonstrates that intelligent machine learning techniques can reduce manual diagnostic effort, improve prediction reliability, and provide valuable decision support for healthcare professionals in the early diagnosis and management of thyroid disorders.

VI. CONCLUSION

The proposed Adaptive Intelligence for Early Thyroid Disease Detection and Classification system demonstrates that machine learning can provide an effective solution for identifying thyroid disorders accurately and efficiently. The system successfully integrates data preprocessing, model training, algorithm evaluation, and disease prediction into a single web-based platform. Several machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Naïve Bayes, and Multi-Layer Perceptron, were trained and compared using clinical thyroid data. Performance evaluation based on accuracy, precision, recall, and F1-score showed that Random Forest and Decision Tree achieved the most reliable classification results among the evaluated models. The application also supports both manual data entry and dataset-based prediction, making it suitable for different clinical scenarios. Proper preprocessing and feature preparation contributed significantly to improving prediction consistency and reducing classification errors. The developed framework simplifies the diagnosis process, minimizes manual effort, and provides timely prediction results that can assist healthcare professionals in clinical decision-making. Overall, the study demonstrates the potential of intelligent machine learning techniques to support early thyroid disease detection and improve the quality and efficiency of healthcare services.

REFERENCES

[1] Azar, at, Hasaean, A.E. and Kim, T. Expert system based on neural fuzzy rules for thyroid diseases diagnosis, Computer Science, Artificial Intelligence, arXiv:1403.0522, Pp. 1-12,2012.
[2] Kales, A. ESTDD: Expert system for thyroid diseases diagnosis, Expert Syst Appl., Vol. 34, No.1, Pp.242–246,2008.
[3] a. Schuck, "World Health Organization," 2000. [Online]. Available: https://www.who.int/. 2nd International Conference on Physics and Applied Sciences (ICPAS 2021) Journal of Physics: Conference Series 1963 (2021) 012140 IOP Publishing doi:10.1088/1742-6596/1963/1/012140 11
[4] Koura, K., Exarches, T.P. Exarches, K.P., Karamus, M.V. and Fotiadis, D.I. (2015) Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal, Vol. 13, Pp.8–17.
[5] Shukla, A. & Kaur, P. (2009). Diagnosis of thyroid disorders using artificial neural networks, IEEE International Advance computing Conference (IACC 2009)– Patiala, India, pp 1016-1020.
[6] Aswad, Salma Abdullah, and Eulah Sinus. "Classification of VPN Network Traffic Flow Using Time Related Features on Apache Spark." 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE, 2020.
[7] Banu, G. Rasta. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
[8] Chindi, Jamil Ahmed, et al. "TDV: Intelligent system for thyroid disease visualization." 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). IEEE, 2016.
[9] Travis B Murdoch and Allan S Ditsy. The inevitable application of big data to health care. Jama, 309(13):1351–1352, 2013.
[10] Dr. Srinivasan B, Pavia K "Diagnosis of Thyroid Disease: A Study" International Research Journal of Engineering and Technology Volume: 03 Issue:
11 | Nov – 2016 [11] Aytürk Kelsea and kales, Ali. "ESTDD: Expert system for thyroid diseases diagnosis." International Research Journal of Engineering and Technology (IRJET) Volume: 03 Issue: 11 | Nov -2017 34.1 (2017): 242- 246
[12] Khushboo Taneja, Parveen Sehgal, Preanal "Predictive Data Mining for Diagnosis of Thyroid Disease using Neural Network" International Journal of Research in Management, Science & Technology (E-ISSN: 2321- 3264) Vol. 3, No. 2, April 2016

[13] Chantel, Khushboo, et al. "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques." CSI transactions on ICT 4.2-4 (2016): 313-319.
[14] Banu, G. Resita. "A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease." International Journal of Computer Sciences and Engineering 4.11 (2016): 64-70.
[15] Umar Sidiqi, Dr, Syed Mutahar Saqib, and Rafi Ahmad Khan. "Diagnosis of various thyroid ailments using data mining classification techniques." Int J Sci Res Copout Sci Inf Technol 5 (2019): 131-6.
[16] Sandhya, Mrs. K. "EFFECTIVE PREDICTION OF HYPOTHYROID USING VARIOUS DATA MINING TECHNIQUES."