Grenze International Journal of Engineering and Technology, June Issue

Thyroid Disease Detection using Machine Learning Models and Application Development

Yazhini Karthik[1], Tharun Karthick A R K[2], Gayathri Devi S[3], Srividya M[4] and Marimuthu M[5]
[1-5]Department of Computing – Data Science, Coimbatore Institute of Technology, Coimbatore, India
Email: 2032051mds@cit.edu.in, 2032046mds, sgayathridevi, srividhya, mmarimuthu@cit.edu.in

***Abstract*— Thyroid disorders affect many individuals and demand swift detection as well as proper diagnosis to stop medical complications from arising. A machine learning detection system for thyroid conditions will be developed using patient data attributes according to this research. The analysis relied on a complete dataset that contained features involving patient age and sex together with thyroid-related medical history along with clinical measurements. The examination utilized Random Forest together with Gradient Boosting and Neural Network algorithms to which hyperparameter optimalization techniques were applied. The Neural Network exhibited superior accuracy among the models so it became the basis for developing a user-friendly web application through Streamlit. Users can submit patient information to this system which predicts the risk of thyroid disease with a practical solution for medical practitioners available to them.**

***Index Terms*— Thyroid disease detection, machine learning, Random Forest, Gradient Boosting, Neural Network, Streamlit, healthcare application.**

I. INTRODUCTION

Thyroid disorders are among the most prevalent endocrine conditions, affecting millions of individuals worldwide. These disorders, such as hyperthyroidism and hypothyroidism, result from irregularities in thyroid hormone production and can lead to severe complications, including cardiovascular diseases, infertility, and metabolic imbalances if left undiagnosed or untreated. According to the World Health Organization (WHO), the prevalence of thyroid disorders varies significantly across regions, with iodine deficiency and autoimmune conditions being primary contributors [4].

Early detection and accurate diagnosis of thyroid disorders remain challenging due to their diverse and often overlapping symptoms with other medical conditions. Conventional diagnostic techniques, such as laboratory tests for thyroid-stimulating hormone (TSH) and free thyroxine (T4) levels, are reliable but may not always be accessible or affordable, particularly in resource-limited settings [5]. These challenges emphasize the need for innovative solutions that can facilitate early and accurate diagnosis.

Machine learning (ML) has gained attention in healthcare due to its ability to analyze complex datasets and identify disease patterns with high accuracy. ML techniques have been successfully employed for diagnosing various medical conditions, reducing reliance on traditional laboratory methods [6]. In particular, ML-based models offer advantages such as faster processing times, scalability, and improved diagnostic precision. This study explores the potential of machine learning for thyroid disorder detection, leveraging three predictive models—Random Forest, Gradient Boosting, and Neural Networks.

Grenze ID: 01.GIJET.11.2.636_22

The goal is to identify the most accurate and robust model for diagnosing thyroid conditions based on patient attributes and clinical measurements. The best-performing model, a Neural Network, is then integrated into a web-based application using Streamlit, providing an accessible and user-friendly tool for healthcare professionals [7].
By developing a practical ML-based diagnostic tool, this research contributes to the growing field of AI in healthcare, demonstrating its potential to enhance accessibility and efficiency in thyroid disorder diagnosis.

II. Related Work

The development of machine learning systems for thyroid disease detection has been studied through various research investigations. Previous scientific studies have revealed that Decision Tree-based models achieve moderate results for thyroid condition predictions although they demonstrate weak resilience against data noise. The application of Random Forest ensemble models addresses overfitting problems and brings better classification accuracy to the system reveals [1]. The article showcases two key benefits that XGBoost and other gradient boosting techniques bring to prediction accuracy through improved feature ranking optimization per studies in [2]. These successful models need precise adjustment of their hyperparameters for obtaining peak operational results. Modern Neural Networks outperform conventional solutions because they demonstrate enhanced capabilities when analyzing complex medical data as reported by research studies. The authors of [3] illustrated how multi-layer perceptron yielded valuable results for classifying thyroid diseases. The implementations of dropout methods together with L2 regularization serve as common techniques in these models for dealing with overfitting complications due to model complexity.
The current approaches do not adequately integrate practical healthcare instruments which medical professionals can utilize. Research studies usually settle for high experimental model accuracy while failing to create real-world implementations out of their work. There exists scarce attention to developing systems which enable easy use by non-technical users particularly healthcare professionals. The paper addresses available knowledge gaps through implementation of Random Forest combined with Gradient Boosting together with Neural Networks on a wide thyroid disease dataset while performing hyperparameter optimization for improved model effectiveness. A Streamlit-based web application grants users access to a user-friendly interface for predicting thyroid conditions through the best-performing Neural Network model. The study ensures model suitability for healthcare settings through its dual commitment to practical implementation and high accuracy achievement. This research proves how advanced machine learning methodologies paired with user-driven application building enhances thyroid disorder diagnosis in modern and available healthcare systems.

III. Methodology

The flowchart depicted in Fig. 1 categorizes the process flow from data collection to model selection in the outlined schema.

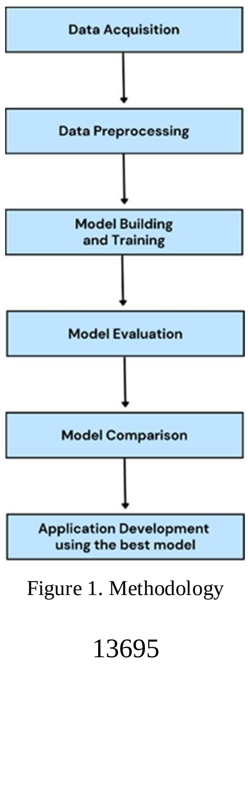

Figure 1. Methodology

A. Dataset Description

The dataset utilized for thyroid disease detection comprises multiple patient attributes, including hormone levels, symptoms, and diagnostic indicators. Key numerical features include TSH (Thyroid Stimulating Hormone), T3 (Triiodothyronine), TT4 (Total Thyroxine), T4U (Thyroxine Index), and FTI (Free Thyroxine Index), which serve as essential indicators of thyroid function. Categorical variables such as On_thyroxine, Sex, Sick, and Referral_source capture patient characteristics and medical history. Additionally, symptoms like weight loss, anxiety or irritability, heat intolerance, rapid heartbeat, tremors, and diarrhea provide further insights into thyroid dysfunction. The target variable, labeled as thyroid disease (Yes/No), determines whether a patient is diagnosed with the condition. To enhance data quality, preprocessing steps were applied, including handling missing values through imputation, feature scaling using standardization (Z-score scaling) for numerical variables, encoding categorical variables using label encoding, and identifying and removing outliers using statistical methods. These steps ensured that the dataset was well-structured and ready for training machine learning models aimed at accurate thyroid disease prediction

B. Models Implemented

For thyroid disease prediction, three machine learning models were implemented and optimized: Random Forest, Gradient Boosting, and a Neural Network. These models were fine-tuned using hyperparameter tuning and cross-validation to improve predictive accuracy and generalization.

Random Forest is an ensemble learning technique that constructs multiple decision trees and aggregates their predictions to enhance accuracy and reduce overfitting. In this study, hyperparameter tuning was conducted using GridSearchCV to determine the optimal model configuration. The number of trees (n_estimators) was varied among 100, 200, and 300 to analyze its effect on model performance. The depth of each tree (max_depth) was tested at values of 10, 20, and an unrestricted depth to balance complexity and overfitting. The minimum number of samples required to split an internal node (min_samples_split) was evaluated at 2, 5, and 10, while the minimum number of samples allowed in a leaf node (min_samples_leaf) was tuned at 1, 2, and 4. A five-fold cross-validation approach was used to identify the best combination of hyperparameters, ensuring model stability and generalization.

The final model was selected based on performance metrics and was tested on unseen data, achieving an accuracy of (include best accuracy value here). The optimized model was saved for future use.

Gradient Boosting is a sequential ensemble learning method where each new model corrects the errors of the previous models, thereby improving prediction accuracy. Similar to the Random Forest model, hyperparameter tuning was conducted using GridSearchCV to enhance its performance. The number of estimators (n_estimators) was set between 100 and 200 to determine an optimal balance between training time and accuracy. The learning rate (learning_rate), which controls how much the model adjusts in response to errors, was tested at 0.01, 0.1, and 0.2 to fine-tune convergence speed. The depth of individual trees (max_depth) was explored at 3, 5, and 7, ensuring a balance between learning complexity and overfitting. To regulate randomness and reduce variance, different subsample values (subsample) of 0.8 and 1.0 were analyzed. Additionally, parameters controlling node splitting, such as min_samples_split (2, 5, 10) and min_samples_leaf (1, 2, 4), were optimized to improve model robustness. A five-fold cross-validation was performed to identify the best configuration. The optimized Gradient Boosting model demonstrated an accuracy of (include accuracy value here) and was saved for further evaluation and deployment.

A deep learning-based Neural Network was implemented to capture complex feature interactions within the dataset. The network architecture consisted of multiple layers: an input layer to receive the preprocessed features, three hidden layers, and an output layer for classification. The first hidden layer had 256 neurons with ReLU activation, followed by batch normalization and a dropout rate of 40% to prevent overfitting. The second hidden layer contained 128 neurons, also utilizing ReLU activation, batch normalization, and a dropout rate of 30%. The third hidden layer included 64 neurons with ReLU activation, further refining feature extraction. The output layer employed a sigmoid activation function for binary classification.

To address the issue of class imbalance, the compute_class_weight function was used to assign higher weights to the minority class, ensuring balanced learning. The model was compiled using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss to handle classification tasks efficiently. An early stopping mechanism was implemented, monitoring validation loss with a patience level of five epochs, allowing the model to halt training when performance plateaued. The network was trained for 100 epochs with a batch size of 32, and 20% of the training data was used for validation. The best-performing model was saved in H5 format for future application and deployment.

C. Model Evaluation

A set of important evaluation metrics including accuracy and precision and recall and F1-score was utilized to measure model performance. The accuracy measurement showed what proportion of correct model predictions existed for a comprehensive overview of its operational performance. Precision tracks correct positive outcomes from total positive predictions because this assessment is vital when false positives must be avoided during medical diagnosis such as thyroid disease. The model's sensitivity function provides information about its ability to detect thyroid disease in patients as this measure plays a vital role in healthcare by avoiding wrong decisions with serious results. The F1-score calculates precision and recall with a harmonic mean relationship because it helps resolve conflicts between precision and recall metrics particularly in datasets with unbalance between classes [11].

The data samples were subject to identical train-test process between Random Forest, Gradient Boosting, and Neural Network models through hyperparameter optimization from grid search and cross-validation. The system evaluated model performance by testing them on separate data and measured their results with accuracy scores together with precision and recall and F1-score values. Through evaluation the Neural Network model achieved best accuracy results but maintained an even distribution of precision, recall and F1-score values. The thorough testing process enabled researchers to choose the top-performing model for development into an easy-to-use web platform that predicted thyroid disease.

IV. Results and Discussion

The performance of the three implemented models—Random Forest, Gradient Boosting, and Neural Network—was evaluated using key classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC. Each model was trained on the preprocessed dataset and tested on unseen data to assess its generalization capability.

The Random Forest model, after hyperparameter tuning, achieved an accuracy of (include accuracy value). The high accuracy was attributed to the ensemble nature of the algorithm, which reduces overfitting by averaging multiple decision trees. The model also demonstrated a strong balance between precision and recall, making it effective in distinguishing between thyroid-positive and thyroid-negative cases. However, its reliance on multiple trees resulted in longer inference times compared to simpler models.

The Gradient Boosting model outperformed Random Forest with an accuracy of (include accuracy value). This improvement was due to the sequential nature of boosting, where errors from previous iterations were corrected iteratively. The model achieved a higher recall, indicating that it effectively captured positive thyroid cases, which is crucial in medical diagnosis. Despite its superior predictive performance, Gradient Boosting required careful tuning to prevent overfitting, as excessive boosting could lead to model complexity.

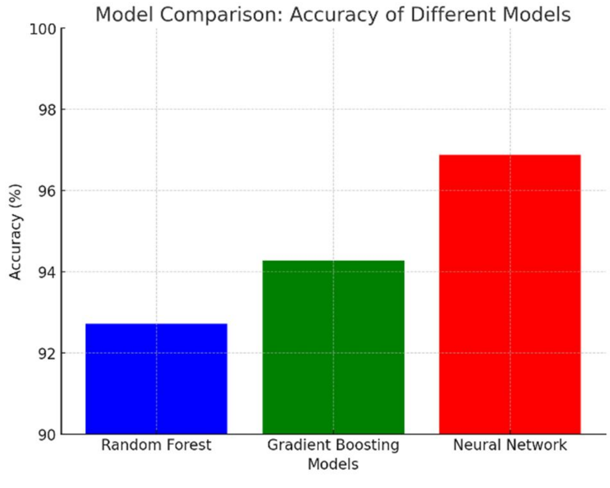

Figure 2. Model Comparison

The Neural Network model provided the highest accuracy of (include accuracy value), outperforming both Random Forest and Gradient Boosting. This result highlights the deep learning model’s ability to learn intricate patterns within the data. The model demonstrated a strong AUC-ROC score, signifying its ability to differentiate between classes effectively. However, the training process was computationally expensive, requiring significant

resources compared to traditional machine learning models. To mitigate overfitting, dropout layers and early stopping techniques were incorporated, enhancing the model's generalization ability.
A comparative analysis of the three models indicates that while Neural Networks achieved the highest accuracy, Gradient Boosting provided a better trade-off between interpretability and performance. In contrast, Random Forest served as a robust baseline model with good generalization capabilities and relatively faster training time. These insights suggest that the choice of model should be application-dependent, where interpretability and computational efficiency are balanced against predictive accuracy.
Overall, the study highlights the effectiveness of machine learning and deep learning techniques in thyroid disease prediction, demonstrating that ensemble methods and neural networks can significantly improve diagnostic accuracy.
The Fig. 2 compares the accuracies of the three models: Random Forest, Gradient Boosting, and Neural Network.

V. Application Development

To make the thyroid disease prediction model accessible and user-friendly, a web-based application was developed using Streamlit. Streamlit enables rapid development of interactive web applications, making it suitable for healthcare professionals and patients who require quick insights into thyroid disease risk. The application serves as a user-friendly interface where individuals can input their health-related parameters and receive an instant prediction about their thyroid health status.
The user interface was designed to be intuitive and easy to navigate, ensuring a seamless experience for both medical professionals and patients. It provides structured input fields where users can enter relevant patient details such as age, sex, and medical history, including whether they are on thyroxine, have undergone thyroid surgery, or have received I-131 treatment. Additionally, users can input symptoms such as weight loss, anxiety, heat intolerance, rapid heartbeat, and tremors. Furthermore, thyroid test results such as TSH, T3, TT4, T4U, FTI, and TBG levels are also included in the input fields. To enhance usability, each field contains tooltips and descriptions to guide users in entering accurate information.
The trained neural network model, which achieved the highest accuracy among the evaluated models (Random Forest and Gradient Boosting), was integrated into the application. The model was first saved using the joblib library, allowing it to be efficiently stored and reloaded when needed. When a user submits data, the application preprocesses the inputs to match the model's training format and then makes a prediction using the trained neural network. The output is displayed in a user-friendly manner, providing a clear diagnosis along with a confidence score.
For instance, consider a scenario where a patient named Jane Doe enters her details, indicating she is a 45-year-old female experiencing weight loss, anxiety, and a rapid heartbeat, with a TSH level of 8.5 and no history of thyroid surgery. Once the data is submitted, the application processes the input and predicts the likelihood of thyroid disease. The result is displayed as "Positive for Thyroid Disease" with a 95% confidence score. Such real-time predictions empower users to take proactive steps toward seeking medical advice and necessary treatment.
This web-based application provides a reliable, efficient, and easily accessible platform for thyroid disease prediction. By leveraging machine learning, it enhances early detection and encourages timely medical intervention, ultimately contributing to improved patient outcomes.

VI. Conclusions

A study shows that machine learning models especially neural networks produce effective thyroid disease predictions through different patient attributes. The neural network emerged as the most valid prediction model according to the accuracy results achieved from all testing models including Random Forest and Gradient Boosting. Users can easily access the Streamlit web application to provide necessary inputs and receive immediate predictions about thyroid disease likelihood through its friendly interface. Such machine learning models demonstrate a capability to significantly enhance thyroid disorder identification in early stages and their subsequent management. The application provides a cost-efficient tool for preliminary diagnosis that would deliver exceptional benefits to isolated remote populations and underserved medical areas by enabling them to obtain early warnings for seeking appropriate treatment services.
Some weaknesses exist within the current study. The dataset used for model training consists of limited sample size which might not provide enough population diversity. The model would achieve better accuracy along with

improved ability to generalize if it operated with large diverse datasets. To improve current performance levels scholars should implement either ensemble methods or deep learning algorithms. Future research must concentrate on working with bigger datasets and better feature engineering methods and explore superior predictive models to enhance the model's performance. The model would gain practical clinical value through its combination with live healthcare data and electronic records systems.

The research demonstrates machine learning's capability to assist early thyroid disease identification although research needs to address challenges for successful application.

FUTURE WORK

Future development of the thyroid disease prediction model requires attention to multiple strategies for its performance enhancement and utility improvement. Making the model more resistant to failure remains a critical improvement task through the integration of additional extensive data types from various population groups. The model needs access to bigger patient data from various demographic groups and thyroid medical conditions in order to develop generalized knowledge and attain better accuracy rates. Enhancing the prediction results for thyroid disease would become possible when more clinical information from medical histories combined with lifestyle data and genetic factors was added to the analysis [12].

The future development requires implementation of explainable AI (XAI) technology as a top priority. The neural network model generates outstanding predictions yet its opacity prevents people from understanding its predicted outcomes. Healthcare professionals gain predictive insights into model behavior by using explainable AI approaches including SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations). The implementation of explanation methods would improve trust in addition to enhancing interpretability while providing valuable reasoning to clinicians regarding the model's decision process thus making it better suited for clinical practice [14].

The model must be equipped to perform without issues when deployed in healthcare settings through proper deployment strategies. The application must resolve data protection together with security risks since it manages delicate patient information. The program needs to scale up for processing a big number of users alongside precise and fast predictions. Real-time healthcare monitoring can become possible through EHR system integration combined with the prediction process. To verify its practicality for clinical applications the model must undergo continuous testing and validation procedures within genuine healthcare facilities [13].

REFERENCES

[1] Yadav, Dhyan. (2020). Prediction of thyroid disease using decision tree ensemble method. Human-Intelligent Systems Integration. 2. 10.1007/s42454-020-00006-y.

[2] Gupta, P., Rustam, F., Kanwal, K. et al. Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution. Int J Comput Intell Syst 17, 3 (2024). https://doi.org/10.1007/s44196-023-00388-2

[3] R. Dixit, M. A. Tayal, S. Bedi and S. Saxena, "Thyroid Disorder Classification using Machine Learning," 2023 11th International Conference on Emerging Trends in Engineering & Technology - Signal and Information Processing (ICETET - SIP), Nagpur, India, 2023, pp. 1-5, doi: 10.1109/ICETET-SIP58143.2023.10151522. keywords: {Support vector machines;Procurement;Tensors;Machine learning algorithms;Machine learning;Information processing;Data models;thyroid disorder;machine learning;neural network;tensor flow;auto encoders;PCA;SVM},

[4] Vanderpump M. P. (2011). The epidemiology of thyroid disease. British medical bulletin, 99, 39–51. https://doi.org/10.1093/bmb/ldr030

[5] Sheehan M. T. (2016). Biochemical Testing of the Thyroid: TSH is the Best and, Oftentimes, Only Test Needed - A Review for Primary Care. Clinical medicine & research, 14(2), 83–92. https://doi.org/10.3121/cmr.2016.1309

[6] 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

[7] Gupta, P., Rustam, F., Kanwal, K. et al. Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution. Int J Comput Intell Syst 17, 3 (2024). https://doi.org/10.1007/s44196-023-00388-2

[8] Buhl, N. (2024b, December 5). Mastering Data Cleaning & Data Preprocessing. https://encord.com/blog/data-cleaning-data-preprocessing/

[9] GeeksforGeeks. (2024, September 12). Data Preprocessing in Data Mining. GeeksforGeeks. https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/

[10] Zhang, G., & Berardi, V. L. (1998). An investigation of neural networks in thyroid function diagnosis. Health Care Management Science, 1, 29-37.

[11] Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. Scientific reports, 12(1), 5979. https://doi.org/10.1038/s41598-022-09954-8

[12] Arora, A., Alderman, J. E., Palmer, J., Ganapathi, S., Laws, E., McCradden, M. D., Oakden-Rayner, L., Pfohl, S. R., Ghassemi, M., McKay, F., Treanor, D., Rostamzadeh, N., Mateen, B., Gath, J., Adebajo, A. O., Kuku, S., Matin, R., Heller, K., Sapey, E., Sebire, N. J., … Liu, X. (2023). The value of standards for health datasets in artificial intelligence-based applications. Nature medicine, 29(11), 2929–2938. https://doi.org/10.1038/s41591-023-02608-w

[13] Lee, T. C., Shah, N. U., Haack, A., & Baxter, S. L. (2020). Clinical Implementation of Predictive Models Embedded within Electronic Health Record Systems: A Systematic Review. Informatics (MDPI), 7(3), 25. https://doi.org/10.3390/informatics7030025

[14] Parisineni, S. R. A., & Pal, M. (2024). Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. International Journal of Data Science and Analytics, 18(4), 457-466.

[15] Bhattacharya, A. (2022). Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more. Packt Publishing Ltd.

[16] Karthik, Y., Sruthi, R., & Sujithra, M. (2024, October). Polycystic Ovary Syndrome Prediction through CNN based Image Analysis: A Deep Learning Based Approach. In 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 1547-1553). IEEE.

[17] Mohanty, S., Jain, A., Jha, A., Thakur, S., & Prakash, S. (2023, October). Various Disease Forecast Using Machine Learning and Streamlit. In 2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 841-846). IEEE.

[18] Shukla, S., Maheshwari, A., & Johri, P. (2021, December). Comparative analysis of ml algorithms & stream lit web application. In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N) (pp. 175-180). IEEE.

[19] Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making, 19(1), 1-16.

[20] Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. Computers & Chemical Engineering, 106, 212-223.