

أحمد عبدالله

## AI-Assisted\_Thyroid\_CDSS\_Paper\_Combined.docx

 Midocean University

### Document Details

Submission ID

trn:oid::2945:400813235

Submission Date

Jul 24, 2026, 5:05 AM GMT+3

Download Date

Jul 27, 2026, 12:58 PM GMT+3

File Name

AI-Assisted\_Thyroid\_CDSS\_Paper\_Combined.docx

File Size

3.4 MB

13 Pages

3,905 Words

23,104 Characters

# 17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Match Groups

-  **50 Not Cited or Quoted 15%**  
Matches with neither in-text citation nor quotation marks
-  **11 Missing Quotations 2%**  
Matches that are still very similar to source material
-  **0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 11%  Internet sources
- 13%  Publications
- 12%  Submitted works (Student Papers)

## Integrity Flags

### 0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Match Groups

- 50 Not Cited or Quoted 15%**  
Matches with neither in-text citation nor quotation marks
- 11 Missing Quotations 2%**  
Matches that are still very similar to source material
- 0 Missing Citation 0%**  
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**  
Matches with in-text citation present, but no quotation marks

## Top Sources

- 11% Internet sources
- 13% Publications
- 12% Submitted works (Student Papers)

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

|    |                 |                                                                                |     |
|----|-----------------|--------------------------------------------------------------------------------|-----|
| 1  | Publication     | Bhabani Prasanna Pattanaik, Dharendra Kumar Shukla, Sambit Satpathy, Kamala... | 1%  |
| 2  | Submitted works | University of Wales Swansea on 2025-07-21                                      | 1%  |
| 3  | Publication     | Emmanuel Ediri Umukoro, Godwin Kparobo Agbajor, Christian Oyinkuro Kemefa, ... | 1%  |
| 4  | Publication     | Poonam Nandal, Tapas Kumar, Meeta Singh, Mamta Dahiya. "Progressive Compu...   | <1% |
| 5  | Publication     | Liangbin Wu, Chuanqi Wang, Yong Yu. "Data-augmented machine learning for pr... | <1% |
| 6  | Internet        | mdpi-res.com                                                                   | <1% |
| 7  | Submitted works | University of Melbourne on 2025-04-13                                          | <1% |
| 8  | Submitted works | Computer Engg on 2026-07-22                                                    | <1% |
| 9  | Internet        | pubmed.ncbi.nlm.nih.gov                                                        | <1% |
| 10 | Internet        | www.ijcaonline.org                                                             | <1% |

|    |                 |                                                                                     |     |
|----|-----------------|-------------------------------------------------------------------------------------|-----|
| 11 | Submitted works | University of Northumbria at Newcastle on 2026-05-10                                | <1% |
| 12 | Submitted works | University of Sunderland on 2026-01-01                                              | <1% |
| 13 | Internet        | jisem-journal.com                                                                   | <1% |
| 14 | Publication     | V. Sekhar, P. Ajay Kumar Reddy, K. Logesh, T. Raghunadha Reddy, Dara Raju. "Ne...   | <1% |
| 15 | Publication     | Aysel Topşır, Ferdi Güler, Ecesu Çetin, Mehmet Furkan Burak, Melih Ağraz. "Thyro... | <1% |
| 16 | Publication     | Wei Feng, Jingge Gao, Zhe Zhang, Maohua Ma, Jinyu Ye, Chundi Chen. "Optimizin...    | <1% |
| 17 | Internet        | www.ijsred.com                                                                      | <1% |
| 18 | Submitted works | City University on 2024-08-09                                                       | <1% |
| 19 | Submitted works | Liberty University on 2026-07-21                                                    | <1% |
| 20 | Submitted works | London Business School on 2024-12-23                                                | <1% |
| 21 | Publication     | Shoujie Jia, Qiong Li, Wenji Zhao, Yanhui Wang. "How Do Rural Households Achiev...  | <1% |
| 22 | Internet        | bjhsjournal.com                                                                     | <1% |
| 23 | Internet        | educatorcafeteahouse.blogspot.com                                                   | <1% |
| 24 | Internet        | github.com                                                                          | <1% |

|    |                 |                                                                                 |     |
|----|-----------------|---------------------------------------------------------------------------------|-----|
| 25 | Internet        | www.pubfacts.com                                                                | <1% |
| 26 | Internet        | www.bcbsok.com                                                                  | <1% |
| 27 | Internet        | www.bio-conferences.org                                                         | <1% |
| 28 | Internet        | www.frontiersin.org                                                             | <1% |
| 29 | Submitted works | Colorado Technical University Online on 2026-07-22                              | <1% |
| 30 | Submitted works | ICSDI on 2025-12-02                                                             | <1% |
| 31 | Submitted works | University of Hertfordshire on 2025-08-25                                       | <1% |
| 32 | Submitted works | University of Stellenbosch, South Africa on 2025-03-19                          | <1% |
| 33 | Internet        | iircj.org                                                                       | <1% |
| 34 | Publication     | B. G. Nagaraja, S. Kannadhasan. "Information and Communication Systems", CRC... | <1% |
| 35 | Submitted works | COER University on 2025-08-22                                                   | <1% |
| 36 | Submitted works | Liverpool John Moores University on 2023-05-16                                  | <1% |
| 37 | Submitted works | Texas Woman's University on 2017-07-04                                          | <1% |
| 38 | Internet        | ours.ou.ac.lk                                                                   | <1% |

|    |             |                                                                                     |     |
|----|-------------|-------------------------------------------------------------------------------------|-----|
| 39 | Internet    | pjlss.edu.pk                                                                        | <1% |
| 40 | Internet    | repub.eur.nl                                                                        | <1% |
| 41 | Internet    | www.mdpi.com                                                                        | <1% |
| 42 | Publication | K. V. Sambasivarao, Anasuya Sesha Roopa Devi Bhima. "Artificial Intelligence, Co... | <1% |
| 43 | Publication | Keerthy Gopalakrishnan, jieun lee, Renisha Redij, Keirthana Aedma et al. "Tu1984... | <1% |

# AI-Assisted Clinical Decision Support for Thyroid Risk Assessment Using an XGBoost Classifier

Ahmed Abdullah

Orchid ID:XXXX

Master of Science in Artificial Intelligence

Master of Science in Artificial Intelligence – College of Technology – Midocean University – Comoros  
[253000598@st.midocean.edu.km](mailto:253000598@st.midocean.edu.km)

## Abstract

Thyroid diseases are amongst the most common endocrine disorders around the globe; however, delays or inconsistencies in the diagnostic process may result in some health consequences. The modern diagnostic processes usually include the combination of several biomarkers measured in laboratory tests, which may be time-consuming and prone to clinician-to-clinician variability.

In this study, an AI-based CDSS is used to improve, but not replace, the existing expert systems in stratification of thyroid risks. Five widely used biomarkers including Thyroid Stimulating Hormone (TSH), Triiodothyronine (T3), Total Thyroxine (TT4), Thyroxine Uptake (T4U), and the Free Thyroxine Index (FTI) were selected from the UCI ANN Thyroid Disease dataset. An XGBoost classifier was applied to predict Normal, Hypothyroidism, and Hyperthyroidism states based on this data with an accuracy of 97.23%.

The trained classifier was then embedded into the developed prototype app using Streamlit, and allows to display a predicted classification, a level of confidence, probabilities of every class and summary information about the relevant laboratory values. As a result, the combination of a gradient-boosted classifier and the decision support interface seems to be a promising way of improving the existing system. Future work will focus on validating this approach using larger and more varied clinical datasets.

**Keywords:** *Artificial Intelligence, Clinical Decision Support System, Machine Learning, XGBoost, Thyroid Risk Assessment.*

## 1. Introduction

Thyroid disorders are considered common endocrine disorders around the world. Problems with hormone secretion may affect metabolism, the cardiovascular system, and neurology. Hyperthyroidism and hypothyroidism, which are the main types of thyroid diseases, are associated with significant morbidity if the disorder is undiagnosed.

The identification of thyroid disease usually requires the joint consideration of several laboratory indicators, among them TSH, T3, TT4, T4U, and FT. It is rare when the diagnosis is made on the basis of one test result. Such analysis may be time-consuming in medical practice. Moreover, values near the reference range boundaries will lead to inter-doctor variability in interpreting the data. Therefore, there is a possibility of using machine learning techniques for decision-making support. Several supervised learning models like decision tree, support vector machine, artificial neural networks, random forest, and gradient boosting have been tested on structured data, showing promising results (Islam et al., 2022; Alshamrani et al., 2023).

One of the prominent methods for structured healthcare data is Extreme Gradient Boosting (XGBoost), due to the efficiency of nonlinear dependency modeling and overfitting prevention. Nevertheless, the majority of the current literature is oriented toward improving the accuracy of classifications, while ignoring the actual development of the trained machine learning algorithm as the decision-making tool to be used by clinicians

(Sutton et al., 2020; Sendak et al., 2021; Rajkomar et al., 2022). Thus, the goal of the current research is to address this gap.

Specifically, based on the laboratory biomarker data from the UCI ANN Thyroid Disease Dataset, an XGBoost classifier was created to identify whether a patient has either Normal, Hypothyroidism or Hyperthyroidism conditions. The trained classifier was embedded into the Streamlit-based prototype that delivers predictions alongside with confidence scores and probabilities as well as the summary of laboratory values. It is expected that the prototype can be used as a decision support tool, which is supposed to complement, not replace, the decision-making process of the clinician. The output of the prototype is the probability, not the explanation of the diagnosis.

The model produced a classification accuracy of 97.23% on the independent test set. The rest of the paper discusses related works (Section 2), the data set, preprocessing, the model, and the prototype (Section 3), the results and discussion (Section 4), and limitations and future work (Section 5).

## 2. Previous Studies

Several machine learning algorithms have been used in diagnosing different conditions, including cardiovascular diseases, diabetes, cancer, and endocrine disorders. These models are able to find complicated patterns in structured clinical data which cannot be easily determined. There exists an opportunity to use machine learning models in diagnosing thyroid disorders since their diagnosis requires the analysis of several correlated biomarkers.

Guidelines of the American Thyroid Association recommend the diagnosis of thyroid function taking into account TSH level and thyroid hormone level in correlation with clinical findings and not any particular value (Ross et al., 2016). Normal reference ranges for TSH values usually lie between 0.4 and 4.0 mIU/L. However, some variations may occur depending on the laboratory and assay used. A euthyroid condition can be characterized by TSH in that range with hormone levels lying in the reference ranges of the laboratory.

Primary hypothyroidism will have elevated TSH values with low hormone levels and, conversely, primary hyperthyroidism will have low TSH and high hormone levels. It should be noted that it is just a general guideline since reference ranges may vary according to such factors as age, pregnancy, medications used, and assay itself. Table 1 outlines these common diagnostic patterns. Subclinical thyroid dysfunction, characterized by an abnormal TSH level while hormone levels remain within the reference range, is also a recognized condition that is not encompassed by the three-class classification system employed in this study.

| Thyroid Status                    | Typical TSH   | Thyroid Hormones (T3 / T4 / TT4) | Interpretive Pattern               |
|-----------------------------------|---------------|----------------------------------|------------------------------------|
| Normal (Euthyroid)                | 0.4–4.0 mIU/L | Within reference range           | Normal thyroid function            |
| Hypothyroidism (primary pattern)  | > 4.0 mIU/L   | Decreased                        | Reduced thyroid hormone production |
| Hyperthyroidism (primary pattern) | < 0.4 mIU/L   | Increased                        | Excess thyroid hormone production  |

Table (1): Typical Laboratory Patterns Associated with Thyroid Status

Only a few researches have been conducted using machine learning methods in thyroid classification in the recent past. To illustrate, in the study by Islam et al. (2022), decision trees, support vector machines, and

11 27 random forests were utilized in the classification of thyroid disease using the UCI Thyroid Disease dataset. The study indicated that there was ability of these models to classify thyroid disease from the laboratory data. Another study was carried out by Alshamrani et al. (2023) where different algorithms were used and showed that ensemble algorithms performed better than classical classifiers. The focus of the study was on the efficiency of the models rather than their application in the clinical setting.

However, on the contrary, XGBoost algorithms have become popular in application to healthcare data as this approach involves several decision trees that solve the problems of overfitting and nonlinearity.

29 Furthermore, another trend that has been observed within the research literature is that accuracy alone does not seem sufficient for the practical application in the context of healthcare. In particular, Sutton et al. (2020) stressed that clinical decision support systems should be explainable and usable while maintaining clinical judgment. Sendak et al. (2021) referred to trustworthiness, usability, and workflow integration as key barriers to the implementation of the machine learning models, whereas the same opinion was supported by Rajkomar et al. (2022). According to them, accuracy is a necessity yet not a sufficient condition to make a proper use of an algorithm. From the point of view of healthcare practice, it is important to mention that Biondi and Cappola (2021) and Chiovato et al. (2023) stressed that thyroid status could not be diagnosed solely on the basis of a laboratory test but rather required hormone profile and clinical data.

| # | Study                                                  | Year | Focus                                      | Approach                                          | Contribution                                            | Identified Limitation                                              |
|---|--------------------------------------------------------|------|--------------------------------------------|---------------------------------------------------|---------------------------------------------------------|--------------------------------------------------------------------|
| 1 | ATA Guidelines for Hyperthyroidism (Ross et al., 2016) | 2016 | Clinical diagnosis of thyroid disorders    | Clinical guideline based on laboratory biomarkers | Established standardized diagnostic protocols           | No computational or predictive modelling                           |
| 2 | High-Performance Medicine (Topol, 2019)                | 2019 | AI in healthcare                           | Conceptual framework                              | Framed integration of human expertise with AI           | No disease-specific implementation                                 |
| 3 | UCI Machine Learning Repository (Dua & Graff, 2019)    | 2019 | Benchmark dataset                          | Structured dataset for supervised learning        | Provides benchmark dataset for thyroid classification   | Limited clinical context (Shen et al., 2023)                       |
| 4 | CDSS Review (Sutton et al., 2020)                      | 2020 | Decision-support systems                   | Literature review                                 | Highlighted interpretability and usability needs        | No thyroid-specific application                                    |
| 5 | Subclinical Thyroid Disease (Biondi & Cappola, 2021)   | 2021 | Thyroid physiology and clinical evaluation | Clinical review                                   | Clarified interpretation of thyroid dysfunction         | No computational modelling                                         |
| 6 | Deep Learning in Healthcare (Esteva et al., 2021)      | 2021 | AI in healthcare (imaging)                 | Deep learning, imaging datasets                   | Demonstrated strong AI performance in disease detection | Limited applicability to structured lab data (Miotto et al., 2021) |
| 7 | ML Challenges in Healthcare (Sendak et al., 2021)      | 2021 | Real-world ML deployment                   | Analytical review                                 | Identified usability and integration barriers           | No disease-specific framework                                      |

|    |                                                               |      |                               |                                        |                                                          |                                                              |
|----|---------------------------------------------------------------|------|-------------------------------|----------------------------------------|----------------------------------------------------------|--------------------------------------------------------------|
| 8  | Deep Learning Review (Miotto et al., 2021)                    | 2021 | AI in healthcare              | Literature review                      | Emphasized integrating multiple data sources             | Interpretability remains limited                             |
| 9  | ML for Thyroid Disease Prediction (Islam et al., 2022)        | 2022 | Thyroid classification        | UCI dataset, DT/SVM/RF                 | Achieved strong classification accuracy                  | Limited interpretability (Rajkomar et al., 2022)             |
| 10 | AI in Healthcare Review (Jiang et al., 2022)                  | 2022 | AI applications in healthcare | Review                                 | Highlighted importance of clinical integration           | No disease-specific framework                                |
| 11 | ML in Medicine (Rajkomar et al., 2022)                        | 2022 | ML in clinical practice       | Review                                 | Emphasized need for interpretability and trust           | No implementation methods offered                            |
| 12 | ML Comparative Analysis for Thyroid (Alshamrani et al., 2023) | 2023 | Thyroid prediction            | Multiple ML models, structured dataset | Demonstrated ML effectiveness for thyroid classification | Limited interpretability (Shen et al., 2023)                 |
| 13 | Advances in Thyroid Diagnosis (Chiovato et al., 2023)         | 2023 | Clinical thyroid diagnosis    | Clinical review                        | Emphasized combining lab and clinical evaluation         | No ML or predictive component                                |
| 14 | ML in Endocrine Diseases (Shen et al., 2023)                  | 2023 | ML in endocrine disease       | Systematic review                      | Identified trends in ML-based endocrine prediction       | Limited interpretability                                     |
| 15 | AI in Endocrine Disease Prediction (Khan et al., 2024)        | 2024 | ML in endocrine disease       | Comparative ML models                  | Showed potential of AI for prediction performance        | Limited interpretability and usability (Sendak et al., 2021) |

**Table (2): Summary of Previous Studies**

Table 2 below highlights the reviewed AI articles in healthcare as well as others relevant to the above-mentioned studies.

From the overall discussion of the above literature, it becomes evident that there are two main gaps: firstly, for most of the studies carried out on thyroid prediction, the study stops at reporting the accuracy level and could have gone beyond just reporting; secondly, almost none of these studies incorporate GUI with a machine learning algorithm.

## 3. Methodology

### 3.1 Research Framework

It has been proposed in this research paper that CDSS assisted by artificial intelligence for three-class risk stratification in the diagnosis of thyroid diseases is needed. In this research, this process is performed in a three-step procedure, which includes data preprocessing, modeling, and designing the decision support interface. These patients will be classified into one of three classes, which are normal, hypothyroidism, and hyperthyroidism. Altogether, the entire process requires five steps, which include data collection, data preprocessing, modeling, model evaluation, and deploying the model in Streamlit.

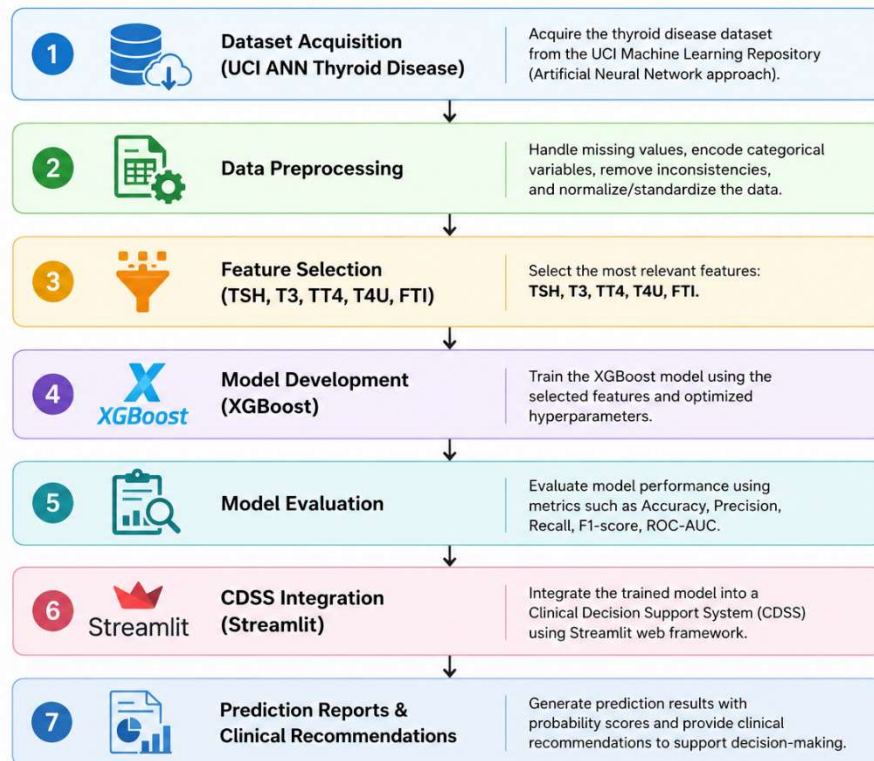


Figure (1): Overall Research Workflow

### 3.2 Dataset Description

The dataset used for the analysis is the UCI ANN Thyroid Disease Dataset, which is a benchmark dataset that is used for classifying thyroid diseases, along with their structured lab data and diagnosis labels. From the total number of features that are available in the UCI database, only five features representing lab data were selected for this analysis. These features include TSH, T3, TT4, T4U, and FTI. The diagnosis label was classified into three categories. These are Normal, Hypothyroid, and Hyperthyroid.

| Original UCI Label | Final Class | Action   |
|--------------------|-------------|----------|
| Negative           | Normal      | Retained |

|                           |                   |             |
|---------------------------|-------------------|-------------|
| Primary Hypothyroid       | Hypothyroidism    | Merged      |
| Compensated Hypothyroid   | Hypothyroidism    | Merged      |
| Secondary Hypothyroid     | Hypothyroidism    | Merged      |
| Hyperthyroid              | Hyperthyroidism   | Retained    |
| T3 Toxic                  | Hyperthyroidism   | Merged      |
| (other labels if present) | Excluded / Merged | Explain why |

### 3.3 Data Pre-processing

31 The chosen predictors from UCI ANN Thyroid Disease dataset include TSH, T3, TT4, T4U, and FTI. All the irrelevant predictors for reaching the objectives of the study were removed. Firstly, there were three categories of diagnosis: Normal, Hypothyroidism, and Hyperthyroidism. Those were encoded into numbers in order to conduct a multi-class classification problem.

13 For the preprocessed dataset, training and validation sets were generated in an 80/20 ratio using Scikit learn's train\_test\_split function with the constant random state = 42. The stratified sampling method was used when the dataset was being divided into two subsets in order to keep the class ratio equal between both sets.

4

36 There was no need for the feature scaling or normalization of the chosen laboratory indicators before modeling as this step is not necessary when dealing with the tree-based algorithms including XGBoost algorithm in this particular study. There was no class balancing technique such as oversampling, under sampling, and class weighting used.

### 3.4 XGBoost Model Development

XGBoost was selected as a classification method due to the proved high predictability, fast performance, and reliability of this algorithm with the structured data from the medicine area. This algorithm is based on the decision tree when each following tree tries to correct mistakes of previous trees. In comparison with regular decision trees and Random Forest algorithms, the application of XGBoost allows us to provide such options as regularization to avoid overfitting and modeling of complex nonlinear dependencies between the clinical parameters.

The training of the algorithm was performed with five laboratory parameters: TSH, T3, TT4, T4U, and FTI, where encoded class diagnosis was a target parameter. Hyperparameter tuning has been performed manually in order to provide the balance between predictive ability and generalization of the model. Tuned hyperparameters are listed in Table 3.

24 The algorithm development and testing have been implemented with the use of Python 3.13.5 programming language in Anaconda environment and Jupyter Notebook. Such libraries and their versions have been used for the implementation of the algorithm: XGBoost 3.3.0, Scikit-learn 1.6.1, Pandas 2.2.3, NumPy 2.1.3.

The algorithm has been tested with the help of validation dataset and such multiclassification metrics as accuracy, precision, recall, F1-score, and confusion

| Hyperparameter                          | Value          | Description                                         |
|-----------------------------------------|----------------|-----------------------------------------------------|
| Objective                               | multi:softprob | Multi-class classification with probability outputs |
| Number of Classes (num_class)           | 3              | Normal, Hypothyroidism, Hyperthyroidism             |
| Number of Trees (n_estimators)          | 100            | Total boosting iterations                           |
| Learning Rate (learning_rate)           | 0.10           | Step size used during boosting                      |
| Maximum Tree Depth (max_depth)          | 6              | Maximum depth of each decision tree                 |
| Minimum Child Weight (min_child_weight) | 1              | Minimum sum of instance weights in a child node     |
| Subsample Ratio (subsample)             | 0.80           | Fraction of training samples used per tree          |
| Column Sampling (colsample_bytree)      | 0.80           | Fraction of features sampled per tree               |
| Gamma (gamma)                           | 0              | Minimum loss reduction required for a further split |
| Regularization (reg_lambda)             | 1.0            | L2 regularization to reduce overfitting             |
| Evaluation Metric (eval_metric)         | mlogloss       | Multi-class logarithmic loss                        |
| Random State (random_state)             | 42             | Ensures reproducibility of results                  |

Table (3): XGBoost Hyperparameters Used in the Proposed Model

With the use of the multi:softprob criterion, probabilities for each of the three diagnostic classes are obtained instead of classification only, which allows producing the confidence score in the prototype. Overfitting and underperforming on unseen data were mitigated by applying regularization and subsampling.

### 3.5 Clinical Decision Support System Implementation

To validate the applicability of the model in practice, the XGBoost model was incorporated into the interactive prototype created with the help of Streamlit library. The application takes patients' lab values as input and provides predictions on the state of their thyroid gland, confidence score, and probabilities for each of the three classes. In addition, it provides the summary of input lab values along with the messages meant to support clinicians' understanding of the input lab values.

As opposed to other research, where the metrics results are presented only, this prototype demonstrates that such classifier can be used in an interactive way.

## 4. Results and Discussion

### 4.1 Experimental Setup

Performance of XGBoost Classifier was evaluated using the test data set provided in the UCI ANN Thyroid Disease dataset, in which cases not used during the training process were employed. The performance of the model was measured using accuracy, precision, recall, F1-score, and confusion matrix.

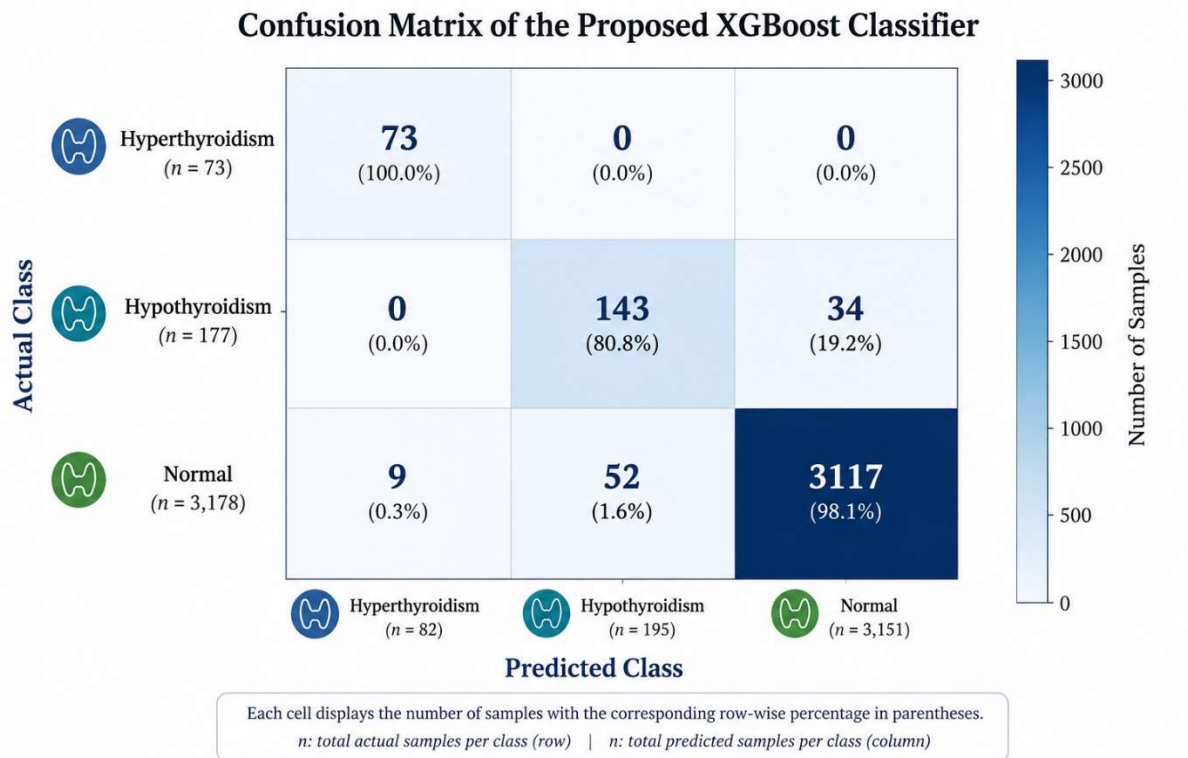
### 4.2 Model Performance Evaluation

| Metric    | Value  |
|-----------|--------|
| Accuracy  | 97.23% |
| Precision | 97.00% |
| Recall    | 97.00% |
| F1-score  | 97.00% |

**Table (4): Performance Evaluation of the Proposed XGBoost Model**

The model showed an accuracy of 97.23% in the independent test set, which implies that the vast majority of the samples were correctly classified. This finding corroborates what is known from the literature about the performance of assembling algorithms on structured thyroid lab data sets (Islam et al., 2022; Alshamrani et al., 2023); nonetheless, a direct comparison between the results is possible only after verifying the reported values of precision, recall, and F1-score, as well as the size of the data set (see Sections 3.2-3.3).

### 4.3 Confusion Matrix Analysis

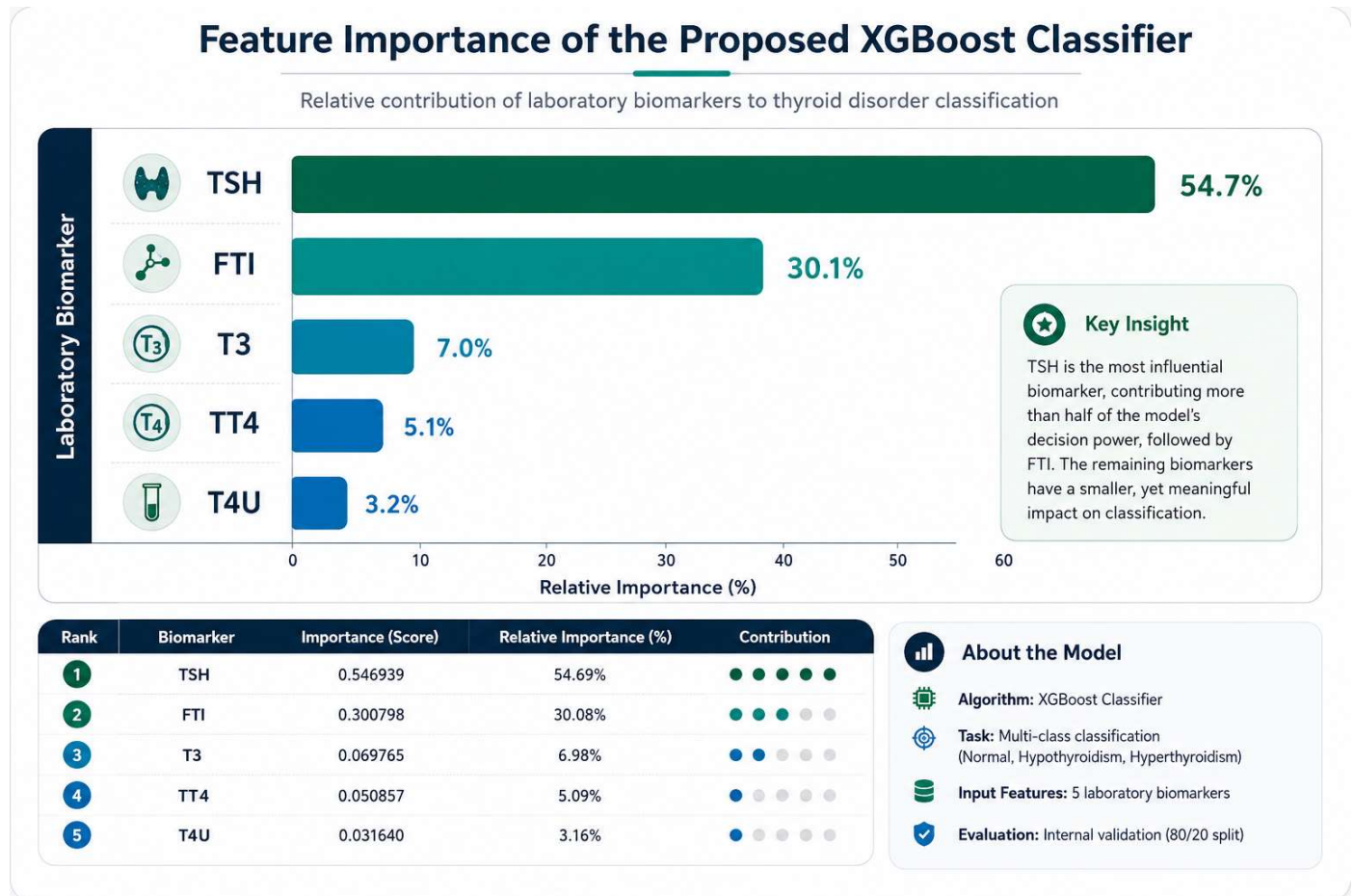


**Figure (2): Confusion Matrix of the Proposed XGBoost Model**

The confusion matrix for XGBoost classifier is provided in figure 2. The XGBoost model had 97.23% accuracy because it correctly classified the samples in all three categories. The recall for the hyperthyroid category was 100% because all 73 cases were accurately classified. For the hypothyroid category, there were 143 out of 177 samples that were classified correctly, while 34 samples were misclassified as normal. There were 3,117 out of 3,178 samples in the normal category that were correctly classified, while 61 samples were misclassified.

The main error that occurred was linked to the classification of the Hypothyroidism and Normal classes, but there were no errors in the classification of the Hyperthyroidism class. It seems like the classifier performs good on average, but not good enough concerning each particular diagnostic class. More precisely, the precision of the classifier for the Hypothyroidism class was lower, meaning that there was a certain percentage of cases when the Hypothyroidism class turned out to be Normal. Even though it may lead to more wrong diagnoses of Hypothyroidism, these will probably be easily clarified during the clinical examination.

## 4.4 Feature Importance Analysis



**Figure (3): Feature Importance of the Selected Laboratory Biomarkers**

Figure (3) demonstrates the weight of the five biomarkers used in the proposed XGBoost classifier. It can be seen that TSH is the most important biomarker and its importance is 54.69%, whereas the second most important feature is FTI and the importance of the same is 30.08%. Other biomarkers possess comparatively lower levels of importance; however, T3 (6.98%), TT4 (5.09%), and T4U (3.16%) made up the list.

It is evident from the above result that TSH and FTI biomarkers are the dominant features in making the classification decision and they account for over 84% of the importance of total features. Although T3, TT4, and T4U were comparatively less important than other biomarkers, these additional pieces of information helped in making the classification more accurate for the diagnosis of Normal, Hypothyroidism, and Hyperthyroidism diseases. The dominance of TSH is due to the reason that it is the most important laboratory biomarker for diagnosing thyroid dysfunction.

9

## 4.5 Clinical Decision Support System Evaluation

**AI-based Thyroid Risk Assessment**  
Machine Learning Assisted Decision Support for Thyroid Disorders

**Navigation**  
Dashboard  
New Assessment  
Assessment History  
Model Information  
About System

**Model Summary**  
Algorithm: XGBoost - Clinical Validation  
Accuracy: 97.23%  
Dataset: USQ Thyroid Dataset  
Features: 11 Laboratory Values  
Version: v2.0  
Research prototype only. Not a medical diagnosis tool.

**Patient Information**  
Patient ID: P-001  
Age (Years): 42  
Gender: ☒ Male ☐ Female

**Laboratory Values**  
TSH: 2.50000  
T3: 1.20000  
TT4: 0.50000  
T4U: 1.00000  
FTI: 0.00000  
Detected Input Type: Clinical Laboratory Values

**Thyroid Risk Assessment Report**  
Enter laboratory values and click Run AI Assessment.

Figure 4. AI-Assisted Clinical Decision Support System Interface Before Prediction

**AI-based Thyroid Risk Assessment**  
Machine Learning Assisted Decision Support for Thyroid Disorders

**Patient Information**  
Patient ID: P-001  
Age (Years): 42  
Gender: ☒ Male ☐ Female

**Laboratory Values**  
TSH: 2.50000  
T3: 1.20000  
TT4: 0.50000  
T4U: 1.00000  
FTI: 0.00000  
Detected Input Type: Clinical Laboratory Values

**Prediction Results**  
Predicted Class: **Normal**  
Confidence: 90.00%  
Risk Level: Low

**Class Probability Distribution**  
Normal: 90.00%  
Hypothyroid: 10.00%

**Laboratory Reference Ranges**

| Lab Test | Unit   | Normal Range | Current Value | Status |
|----------|--------|--------------|---------------|--------|
| TSH      | mIU/L  | 0.1 - 0.4    | 2.5           | High   |
| T3       | pmol/L | 1.0 - 1.8    | 1.2           | Normal |
| TT4      | nmol/L | 0.8 - 1.2    | 0.5           | Low    |
| T4U      | nmol/L | 0.8 - 1.2    | 1.0           | Normal |
| FTI      | pmol/L | 0.1 - 0.2    | 0.0           | Low    |

**Recommendation**  
Based on the prediction, the patient's thyroid condition is classified as Normal.

Figure 5. Sample Prediction Result Generated by the Proposed AI-Assisted Clinical Decision Support System

Figure 4 and Figure 5 illustrate the GUI of the proposed AI-powered CDSS, created using the Streamlit library. In Figure 4, there is shown the screen where the data about the patient along with five laboratory markers (TSH, T3, TT4, T4U, and FTI) are inputted. Further, the data is used as an input to the XGBoost classifier model, and a report is generated.

Figure 5 illustrates the sample of the prediction generated by the proposed approach. The prediction screen shows the patient's thyroid condition, the probability score, the risk level, the distribution of the class probability scores, the reference values of the laboratory test, and the clinical recommendation, which is a result of the algorithm's prediction. It can be seen that the integration of the machine learning model into the CDSS has been successfully performed.

42

## 4.6 Discussion

Such results prove that XGBoost is highly discriminative in distinguishing Normal, Hypothyroidism, and Hyperthyroidism classes using five popular lab indicators for the same benchmarking dataset. The mentioned fact is in agreement with the already existing literature according to which ensembles give superior results compared to individual classifiers when dealing with structured medical data. Nevertheless, most of the previously performed studies limited their investigations only to providing performance metrics, and the current one proposes a developed Streamlit prototype allowing predicting classes with confidence scores and summarizing values of lab indicators.

The assessment is conducted on the basis of a single public benchmarking dataset without any involvement of clinicians or patients' records, let alone any clinical implementation. Therefore, the results presented here should not be regarded as any evidence of clinical relevance. Further research is required in order to test the proposed method on more heterogeneous clinical data and explore its generalizability and possible limitations. Also, it would be useful to utilize some methods of explaining ML-based predictions, for example SHAP and LIME, in order to provide more detailed information for clinicians about individual predictions (the methods were not used in this study).

## 5. Conclusion

39

The AI-powered Clinical Decision Support System (for classification of the thyroid disease into three classes) was constructed on the basis of XGBoost classifier with five laboratory markers (TSH, T3, TT4, T4U, and FTI) of UCI ANN Thyroid Disease dataset. The constructed system demonstrated the accuracy of 97.23% on the test set and therefore proved the capability of the gradient boosting technique for precise classification of Normal, Hypothyroidism, and Hyperthyroidism.

1

Nevertheless, apart from the evaluation of the model performance, the classifier was integrated in the Streamlit-powered prototype to provide predictions and probabilities together with lab values summary for the user as an example of the use of such models in decision support system.

1

1

There are some limitations of the approach. First of all, it relies on one publicly available benchmark dataset instead of using several clinical ones, it is not clinically validated, and there are some facts regarding the experiment which should be checked by the author of the paper (the size of the dataset, its classes distribution, train/test splitting, and precision/recall/F1-score of each model). Future research would consist of using this approach to other clinical and larger datasets, usability validation of the prototype by clinicians, and using explainability methods like SHAP or LIME to interpret individual results. Thus, this feasibility approach could be clinically assessed.

## References

- 14 Alshamrani, M., Alqahtani, A., & Alzahrani, F. (2023). Machine learning approaches for thyroid disease prediction: A comparative analysis. *IEEE Access*, 11.
- Biondi, B., & Cappola, A. (2021). Subclinical thyroid disease: Clinical relevance and management. *The Lancet Diabetes & Endocrinology*, 9.
- Chiovato, L., Antonelli, F., & Vitti, P. (2023). Thyroid disorders and their management: Recent advances in diagnosis and treatment. *Endocrine Reviews*.
- 10 Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. University of California, Irvine. <http://archive.ics.uci.edu/ml>
- 6 33 Islam, M. M., Poly, T. N., Alsinglawi, B., Lin, L.-F., Chien, S.-C., Liu, J.-C., & Li, Y.-C. (2022). Machine learning approaches for thyroid disease prediction: A comparative study. *Healthcare*, 10. <https://www.mdpi.com/2227-9032/10/1/59>
- 41
- 2 22 Rajkomar, A., Dean, J., & Kohane, I. (2022). Machine learning in medicine. *New England Journal of Medicine*, 380.
- Ross, D. S., Burch, H. B., Cooper, D. S., Greenlee, M. C., Laurberg, P., Maia, A. L., Rivkees, S. A., Samuels, M., Sosa, J. A., Stan, M. N., & Walter, M. A. (2016). 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid*, 26. <https://doi.org/10.1089/thy.2016.0229>
- 7 Sendak, M., Gao, W., Brajer, N., & Balu, S. (2021). Machine learning in health care: A critical appraisal of challenges and opportunities. *NPJ Digital Medicine*, 4.
- 3 Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3. <https://www.nature.com/articles/s41746-020-0221-y>
- 19 Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25.

