International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Intelligent Symptom Analysis and Report Interpretation for Thyroid Disorder Prediction Using Machine Learning

Keerti Majajan[1], Vinayak Rhatankar[2], Aniket Salokhe[3], Sohan Kurale[4]

[1]Asst. Prof / Dept. of cse(AIML), D. Y. Patil College of Engineering & Technology, Kasaba Bawada, Kolhapur, India.
[2] Dept. of cse(AIML), D. Y. Patil College of Engineering & Technology, Kasaba Bawada, Kolhapur, India
.[3]Dept. of cse(AIML), D. Y. Patil College of Engineering & Technology, Kasaba Bawada, Kolhapur, India.
.[4]Dept. of cse(AIML), D. Y. Patil College of Engineering & Technology, Kasaba Bawada, Kolhapur, India.

---***---

Abstract - The thyroid gland, situated at the front of the neck, is essential for producing hormones that regulate many of the body's critical functions. This paper presents a machine learning-based system aimed at enhancing early detection and providing personalized care for thyroid disorders. Leveraging the Random Forest algorithm chosen for its robust classification abilities the system incorporates two key models: an initial assessment tool that examines self-reported symptoms to determine if further diagnostic testing is recommended, and a diagnostic model that interprets test results to identify specific thyroid disorders. Additionally, the system provides customized diet and exercise suggestions. With an accuracy of 85%, the system effectively supports early symptom analysis and accurate diagnosis. By combining symptom evaluation with tailored health guidance, this approach enables timely detection, accurate diagnostics, and effective management of thyroid conditions, offering personalized insights for patients and support for healthcare providers.

Key Words: Machine Learning, Random Forest, Thyroid Disease , Data Preprocessing, Healthcare informatics

1.INTRODUCTION

Thyroid disorders are among the most prevalent endocrine conditions, impacting metabolism, energy regulation, and growth. The thyroid gland, located in the neck, plays a critical role in metabolic processes by producing essential hormones like thyroxine (T4) and triiodothyronine (T3) [1]. Disruptions in thyroid function can lead to hypothyroidism (underactive thyroid) or hyperthyroidism (overactive thyroid), each presenting a variety of symptoms that affect quality of life and, if untreated, can lead to serious complications, including cardiovascular disease, osteoporosis, and reproductive health issues [2][3].

Traditional diagnostic methods for thyroid disorders primarily involve blood tests that measure hormone levels, such as thyroid-stimulating hormone (TSH), T3, and T4. While these methods are accurate, they have limitations: they can be costly, require specialized equipment, and rely on patients seeking medical testing, which can delay early diagnosis [4]. Additionally, some patients may not show clear symptoms, leading to delayed disorder identification [5]. Research also indicates that demographic and lifestyle factors such as age, gender, stress levels, and dietary habits affect thyroid health, yet these are often overlooked in conventional diagnostics [6]. To overcome these limitations, researchers have explored machine learning as a tool for early and accessible detection of thyroid disorders. Aversano et al. (2021) evaluated multiple machine learning algorithms, including support vector machines (SVM), decision trees, and random forests, using clinical data on hormone levels and patient demographics [7]. Chaganti et al. (2021) investigated machine learning application using electronic health records (EHR) data, highlighting the potential for these models to detect thyroid conditions without relying solely on lab results [8]. However, these models primarily depend on clinical data rather than symptom- based screening, which limits their usefulness for early detection when patients may not seek immediate lab testing.

This research aims to improve the accessibility, efficiency, and personalization of thyroid disorder diagnosis. The primary goal is to develop a machine learning model that predicts thyroid disorders based on symptoms, facilitating early de- tection without an immediate need for lab tests. This ap- proach includes collecting and preprocessing data on common thyroid-related symptoms, with a focus on selecting relevant features such as weight changes, fatigue, and mood alterations. Various machine learning models, including decision trees and random forests, are then trained and evaluated for accuracy. Additionally, the model includes a recommendation system offering personalized lifestyle and dietary guidance based on individual symptoms. This approach promotes proactive health management, empowering patients to manage thyroid health more effectively.

- Reviewed existing machine learning methods for thyroid detection and identified gaps.
- Developed a model to predict thyroid disorders using symptom data.
- Applied Random Forest for accurate classification of thyroid stages.
- Created a system offering personalized diet and lifestyle advice

.

2. LITERATURE REVIEW

Aversano et al., "Thyroid Disease Treatment Prediction with Machine Learning Approaches," This study investigates machine learning algorithms, including support vector machines (SVM), decision trees, and random forests, to classify thyroid conditions. The researchers applied these models to a clinical dataset with hormone levels and demographic data, demonstrating their potential to improve diagnostic accuracy in thyroid disease detection.[1]

Vasant et al., "Thyroid Detection Using Machine Learning," This paper explores the use of neural networks and logistic regression for thyroid disease prediction, focusing on the accuracy of these models in detecting thyroid disorders. It also examines the impact of various clinical features on model performance, finding that neural networks achieved high accuracy but required significant computational power.[2]

Chaganti et al., "Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques," The authors analyzed the effectiveness of machine learning models on electronic health record (EHR) data to predict thyroid disorders. This study highlights the benefits of using selective features from EHRs to improve early diagnosis while addressing challenges such as data imbalance.[3]

Gupta et al., "Detecting Thyroid Disease Using Optimized Machine Learning Models Based on Different Evolution," This research reviews the role of lifestyle and dietary factors in thyroid health and suggests that integrating personalized recommendations can support thyroid management. The study also explores the role of optimized machine learning models in detecting thyroid conditions based on these lifestyle factors.[4]

Chaubey et al., "Thyroid Disease Prediction Using Machine Learning Approaches," Focusing on the application of machine learning to symptom-based thyroid disease prediction, this paper highlights models that analyze symptoms such as weight gain and fatigue to suggest further tests if necessary. The research shows the potential for machine learning to provide early-stage warnings and reduce unnecessary lab testing.[5]

Saleh et al., "Exploring Challenges of Diagnosing Thyroid Disease with Unbalanced Data and Machine Learning," This paper examines machine learning techniques, including deep learning and ensemble methods, to address the difficulty of diagnosing thyroid disease with unbalanced data. The study highlights how these advanced techniques can handle data imbalance issues commonly found in thyroid disease datasets.[6]

Farling, "Thyroid Disease: An Overview of Thyroid Disorders," This comprehensive review discusses the symptoms, causes, and management of various thyroid disorders, including hypothyroidism, hyperthyroidism, and thyroid cancer. It provides foundational knowledge about thyroid disease, which supports the development of targeted machine learning models.[7]

Montgomery, "Thyroid Disease in Pregnancy," This research focuses on thyroid disease in pregnancy, detailing the risks that thyroid disorders pose to both maternal and fetal health. It underscores the importance of early and accurate diagnosis to manage thyroid function effectively, which aligns with the objectives of symptom-based prediction models.[8]

Peddi et al., "Awareness of Hyperthyroidism," This paper discusses the effects of hyperthyroidism on oral health and the importance of recognizing symptoms early, particularly in healthcare professions like dentistry. The study emphasizes the broader impact of thyroid disorders on overall health and the need for accessible diagnostic tools.[9]

Melish et al., "Thyroid Disease in Clinical Methods: The History, Physical, and Laboratory Examinations," This text provides a thorough overview of thyroid gland disorders, detailing the symptoms associated with functional abnormalities like hyperthyroidism and hypothyroidism, as well as structural issues such as goiters. This foundational knowledge supports the development of predictive models based on symptom analysis.[10]

3. RESEARCH METHODOLOGY

This research outlines a clear methodology to develop a machine-learning model for detecting thyroid disorders based on symptom data. The fig 1 include process like data collection, data preprocessing, feature selection, model building, and the integration of a personalized recommendation system. Each stage is tailored to ensure accurate analysis and practical insights, making the system both effective and user-friendly. This methodology aims to enhance accessibility in thyroid screening and empower individuals with personalized health guidance.

The user initiates the process by entering personal health information, including symptoms, medical history, and lifestyle factors relevant to thyroid function, into a user-friendly platform. This input undergoes careful processing to ensure data accuracy, including tasks like resolving inconsistencies, handling missing values, and standardizing information for further analysis.

The core of the system is a machine learning model, specifically a Random Forest classifier, trained on a comprehensive dataset of anonymized patient records. This dataset includes features such as symptom reports, demographic details, and relevant medical history. Through training, the model has learned to identify relationships and patterns indicative of thyroid conditions. Once the user's data is entered, it is processed by the model, which then generates a risk assessment. This risk is categorized as low, medium, or high based on the likelihood of a thyroid disorder.

Following the risk assessment, the system provides tailored recommendations based on the user's predicted risk level. For high-risk users, the system suggests immediate consultation with a healthcare professional, as timely medical intervention can be essential for effective thyroid management. For medium-risk users, it may recommend lifestyle adjustments, such as dietary modifications, stress reduction practices, and regular physical activity, which can positively impact thyroid health. For low risk users, the system might advise routine health monitoring and general wellness practices to support overall well-being.

This approach allows users to engage in proactive management of their thyroid health. By facilitating early detection of risk levels, offering personalized health guidance, and encouraging medical consultations when necessary, the system supports users in maintaining thyroid health and potentially preventing more serious complications.

The dataset was reviewed to remove any duplicate entries, irrelevant features, and records with incomplete or missing information. Missing data points were managed by either removing entries that could not be completed or imputing missing values based on statistical averages of similar cases.

Normalization and Standardization: To ensure the model's accuracy and efficiency, continuous features (e.g., age, weight) were normalized, which involves scaling numerical data to a standard range. Categorical variables (e.g., symptoms categorized by severity) were encoded to numerical values for use in machine learning models.

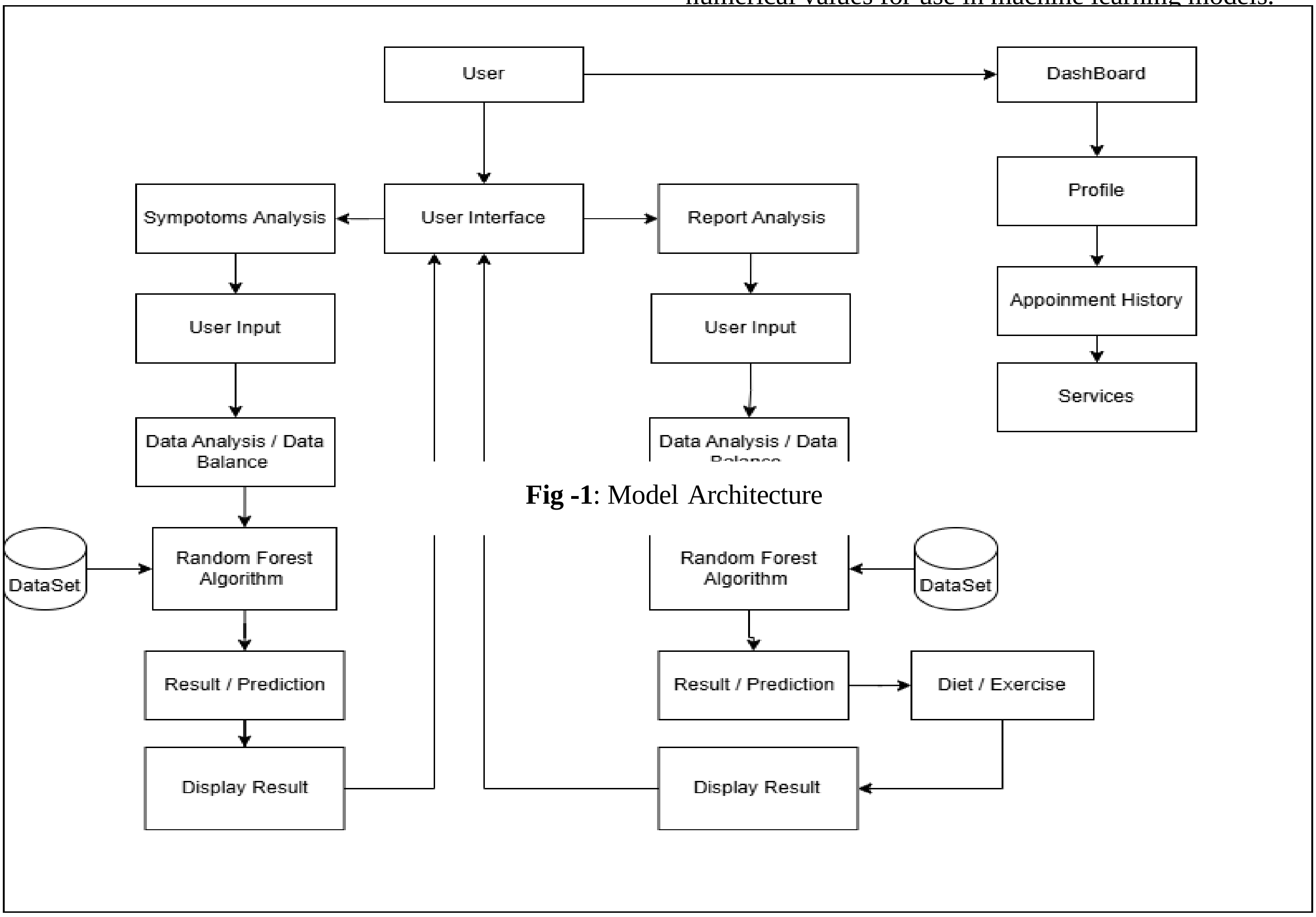


Fig -1: Model Architecture

Data Collection and Pre-processing :

Data was collected from clinical and publicly accessible health datasets that included various thyroid-related symptoms (e.g., weight fluctuations, fatigue) and corresponding diagnoses. The datasets were chosen to ensure they represent a diverse patient demographic to increase model generalization across different populations. Once the data was obtained, the pre-processing phase involved multiple steps to prepare it for analysis:

Data Cleaning :

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Additional symptoms, such as Sensitivity to Cold or

Sr No.	Attribute Name	Value Type	Clarification
1	Age	number	1, 2, 3, . . .
2	Gender	boolean	1=Male, 0=Female
3	Pregnancy	boolean	1=Yes, 0=No
4	Family_History_of_Thyroid	boolean	1=Yes, 0=No
5	Goiter	boolean	1=Yes, 0=No
6	Fatigue	boolean	1=Yes, 0=No
7	Hair_Loss	boolean	1=Yes, 0=No
8	Sensitivity_to_Cold_or_Heat	boolean	1=Yes, 0=No
9	Increased_Sweating	boolean	1=Yes, 0=No
10	Constipation_or_More_Bowel_Movements	boolean	1=Yes, 0=No
11	Depression_or_Anxiety	boolean	1=Yes, 0=No
12	Difficulty_Concentrating_or_Memory_Problems	boolean	1=Yes, 0=No
13	Dry_or_Itchy_Skin	boolean	1=Yes, 0=No
14	Weight_Change	0,1,2	0=No, 1=weight loss, 2=weight gain
15	Heart_Rate_Changes	0,1,2	0=Stable, 1=medium fast, 2=very fast
16	Muscle_Weakness	0,1,2	0=No, 1=mild, 2=Severe
17	Thyroid_Risk_Level	0,1,2	0=Low, 1=Medium, 2=High

Table -1: Attributes for Thyroid Risk Assessment

From Table. 1 we describe symptoms as follows, The dataset assigns specific numerical and binary values to its attributes for consistent analysis. Age is recorded as a straightforward numerical value, while Gender is categorized using ’1’ for male and ’0’ for female. For women, the Pregnancy attribute is marked as ’1’ if the patient is currently pregnant and ’0’ otherwise. The presence of a Family History of Thyroid issues, Goiter, Fatigue, and Hair Loss are all indicated using binary values, where ’1’ denotes the presence of the condition and ’0’ indicates its absence

women, the Pregnancy attribute is marked as ’1’ if the patient is currently pregnant and ’0’ otherwise. The presence of a Family History of Thyroid issues, Goiter, Fatigue, and Hair Loss are all indicated using binary. values, where ’1’ denotes the presence of the condition and ’0’ indicates its absence.

Heat, Increased Sweating, , Constipation or More Bowel

Movements, Depression or Anxiety, and Difficulty Concentrating or Memory Problems, are also captured using a binary scale (1 for Yes, 0 for No). For physical health changes, the dataset uses scaled values: Weight Change is classified as ’0’ for no change, ’1’ for weight loss, and ’2’ for weight gain.

Similarly, for all above symptoms this systematic assignment of numerical values ensures a structured approach to diagnosing and analyzing thyroid conditions.

No	Attribute Name	Value Type	Clarification
1	Age	Number	1, 10, 20, 50, ...
2	Gender	Boolean	1=Male, 0=Female
3	Goiter	Boolean	1=Yes, 0=No
4	Pregnancy	Boolean	1=Yes, 0=No
5	TSH	Analysis Ratio	Numeric Value
6	T3	Numeric Value	0, 1, 2, ...
7	T4	Numeric Value	0, 1, 2, ...
8	TT4	Numeric Value	0, 1, 2, ...
9	T4U	Numeric Value	0, 1, 2, ...
10	FTI	Numeric Value	0, 1, 2, ...
11	TBG	Numeric Value	0, 1, 2, ...
12	Thyroid Condition	Categorical	0=Normal 1=Hypothyroidism 2=Hyperthyroidism

Table -2: Attributes for Thyroid Report Assessment

Above Table 2 contains information collected from patients after blood tests to assess thyroid health. The first few columns include basic demographic information like Age, which is recorded numerically (e.g., 1, 10, 20, 50, etc.), and Gender, marked as ’1’ for Male and ’0’ for Female. It also includes information on whether a patient has a Goiter (1 for Yes, 0 for No) or is currently Pregnant (1 for Yes, 0 for No). These initial attributes help provide context for understanding the patient’s overall health status.
The rest of the dataset focuses on blood test results to measure thyroid hormone levels. It includes TSH (Thyroid- Stimulating Hormone) levels, which help assess how well the thyroid gland is functioning. Other important hormone indica- tors include T3, T4, TT4 (Total Thyroxine), T4U (Thyroxine Uptake), FTI (Free Thyroxine Index), and TBG (Thyroxine- Binding Globulin), all measured numerically. Finally, the Thyroid Condition attribute categorizes the patient’s thyroid health as ’0’ for Normal, ’1’ for Hypothyroidism (low thyroid activity), and ’2’ for Hyperthyroidism (high thyroid activity). This organized data helps doctors diagnose thyroid conditions based on lab results.

Data Splitting: The preprocessed data was split into training and validation sets, typically in an 80:20 ratio. The training data set was used to build and train the machine learning models, while the validation set allowed us to evaluate the model’s performance on unseen data.

A) ***Feature Selection***: Feature selection is a critical step in identifying the most relevant predictors of thyroid disorders within the dataset. We used machine learning algorithms to select features based on their importance to the model’s performance. Key features included:

Weight Fluctuations: Rapid changes in weight are a strong indicator of thyroid dysfunction, specifically weight gain for hypothyroidism and weight loss for hyperthyroidism.

Fatigue and Mood Changes: Both hypothyroidism and hyperthyroidism commonly cause significant fatigue and mood swings due to their effect on metabolic rates. vspace2cm **Temperature Sensitivity**: Individuals with thyroid conditions may report sensitivity to cold (hypothyroidism) or excessive sweating (hyperthyroidism).

Demographic Information: Factors such as age, gender, and lifestyle factors like diet and stress levels are known to influence thyroid health and were included as features.

By focusing on these critical indicators, we aimed to create a model that can accurately predict thyroid conditions with minimal input data, making the tool more accessible and user- friendly.

B) ***Model Development***: To create a reliable prediction model, various classification algorithms were tested, including decision trees and Random Forests, with the latter proving to be the most effective for our dataset.

C) ***Random Forest Algorithm for Thyroid Prediction*** : The Random Forest algorithm is a powerful ensemble learning method that operates by creating multiple decision trees during training and averaging their predictions to reach a final decision. Here’s how Random Forest works in our project:

Tree Creation: In a Random Forest, multiple decision trees are created from randomly selected subsets of the dataset. Each tree makes a prediction on whether a thyroid condition is present and, if so, at which stage (mild, moderate, or severe).

Bagging Technique: Random Forest uses a process known as ”bagging” (Bootstrap Aggregating), where each tree is trained on a randomly sampled subset of the data, which helps reduce overfitting and improves accuracy.

Feature Selection: Within each tree, only a subset of features is chosen to split nodes, ensuring that the model con- siders a variety of symptom combinations, reducing correlation between trees and further improving prediction reliability.

Voting Mechanism: After all trees have made their individual predictions, the Random Forest aggregates the results through a majority vote. For instance, if most trees indicate the presence of hypothyroidism, the final prediction will likely indicate hypothyroidism.

The Random Forest was selected due to its high accuracy, robustness to overfitting, and ability to handle large, diverse datasets. This model not only predicts the presence of a thyroid disorder but also classifies it into ...d advice.

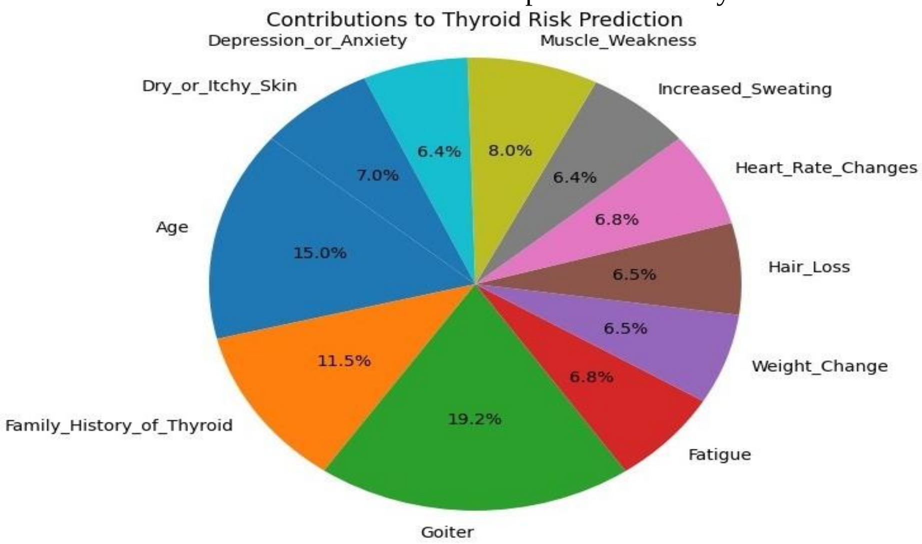


Fig. 2: Overall Symptoms Rate

Above pie ... a role in pr... section of the chart represents a specific characteristic or symptom, such as Age Group, Heart Rate Changes, Weight Change, Muscle Weakness, and Family History of Thyroid. The size of each section reflects how common or influential that factor is in predicting thyroid risk within the dataset.

Fig -2 : symptom based report

The chart helps us identify which factors have a greater impact on thyroid health. For example, some characteristics are more prevalent in the dataset and, therefore, have larger sections in the chart. This gives us valuable insights into which factors should be given more attention when evaluating thyroid health risks.

By analyzing these factors, doctors and healthcare providers can improve their approach to diagnosing thyroid problems. The information allows for more focused attention on the symptoms and characteristics that matter most in predicting thyroid dysfunction. This targeted approach can lead to more accurate diagnoses and help healthcare professionals better assess a patient's thyroid health risk.

Mathematical Expressions

In this research, the mathematical modeling primarily re- volves around decision-tree-based algorithms, specifically the Random Forest model. This approach handles the contribution of symptoms to thyroid risk by creating an ensemble of decision trees, each contributing to the final prediction.The following is a mathematics and mechanics of how this model evaluates symptoms and delivers an overall thyroid risk pre- diction.

A Random Forest is a collection of individual decision trees, each built from a random subset of the data. For the research:

Input Data: Each input represents patient data with features (symptoms) such as fatigue, weight change, and others, along with patient-specific factors like age, gender, and pregnancy status.

Output: The model assigns a Thyroid Risk Level (0 = low, 1 = medium, 2 = high) by combining the outputs from all the individual decision trees.

Each decision tree independently predicts a thyroid risk level based on the presence or absence of specific symptoms, and the Random Forest aggregates these predictions to produce a more accurate and stable outcome.

Mathematical Formulation in Random Forest

The main mathematical operations in the Random Forest involve:

a) **Decision Tree Building:** Each decision tree is built using a subset of features and samples, which is mathematically represented by the following process:

1) **Feature Selection:** For each node in a tree, a random subset of features F is selected. The number of features chosen (usually N for classification problems, where N is the total number of features) controls the diversity among trees.

2) **Node Splitting:** At each node, a feature f_i from F and a threshold value t are selected to split the data. The goal is to find the feature-threshold pair (f_i, t) that

maximizes information gain (or minimizes impurity). The impurity is often calculated using **Gini impurity** for classification:

$$Gini = 1 - \sum_{i=1}^{n} p_i^2$$

where p_i is the probability of each class at a node, and c is the number of classes (in this case, thyroid risk levels 0, 1, or 2).

3) **Recursive Partitioning:** The process of selecting features and splitting continues recursively until each node contains data that corresponds predominantly to one risk class or meets a predefined stopping criterion (like maximum depth).

b)Aggregation of Trees (Ensemble Voting): Once the forest of trees is built, each tree independently classifies a sample and votes on the thyroid risk level. The final prediction is obtained by majority voting (for classification), where the thyroid risk level with the highest number of votes across all trees is selected.

$$\text{Prediction} = arg\ max \sum_{i=1}^{T} I(T_i = c)$$

here: - T_i represents the prediction of the i-th tree, - T is the total number of trees, and - $I(T_i = c)$ is an indicator function that equals 1 if the i-th tree votes for class c and 0 otherwise.

3.Feature Importance and Contribution of Symptoms

To assess how each symptom contributes to thyroid risk, the Random Forest model calculates feature importance, which tells us the relative significance of each feature (symptom) in predicting the outcome.

Feature importance for a feature f can be calculated by measuring the average decrease in impurity across all trees when splitting on f:

$$\text{Importance}(f) = \frac{1}{N} \sum \underset{T}{\Delta}\text{Gini}(f)$$

where: - $\Delta Gini_i(f)$ is the reduction in Gini impurity for feature f in the i-th tree, - T is the total number of trees.

This calculation gives a numerical value that shows how much each symptom influences the thyroid risk prediction. Features with higher values have a stronger impact on the outcome.

4.How It Works in Model:

Symptom Evaluation: Based on the symptoms present (marked 'yes' in your dataset), the Random Forest model evaluates the pattern and predicts the risk of thyroid disorder by analyzing the importance and interactions of each symptom. Prediction and Explanation Generation: Since each tree makes an independent prediction, you gain a robust, averaged result that reduces the risk of overfitting, providing a stable prediction. The feature importance scores also help create personalized explanations by highlighting the symptoms that contributed most to the predicted risk level.

Example case: 1. The model computes the probability of each thyroid risk level by assessing the presence and contribution of symptoms across the forest.

2. Based on the model's predicted risk and the feature importance values, the explanations for each prediction can focus on the most impactful symptoms—tailored to each patient's specific data.

This method provides a predictive and interpretive model that is both accurate and useful for personalized health advice in thyroid risk assessment.

c) **Recommendation System**

Once the model detects a thyroid disorder and determines its stage, it provides personalized lifestyle and dietary recommendations to the user. This recommendation system was developed with a focus on providing practical, accessible health advice that supports patients in managing their thy- roid condition. Key features of the recommendation system include:

Dietary Recommendations: Based on the specific thyroid condition, the system suggests nutrient-rich foods that support thyroid health. For hypothyroidism, the model may suggest iodine-rich foods like fish, while for hyperthyroidism, it could advise a balanced intake of protein and fats to maintain energy levels.

Lifestyle Adjustments: The system provides advice on daily routines and practices that help alleviate symptoms. For example, it may recommend regular physical activity and stress management techniques, both of which can positively impact thyroid function

Testing and Monitoring: If symptoms or model predictions indicate a severe or rapidly progressing condition, the recommendation system advises the user to seek additional blood tests (e.g., TSH, T3, T4) to confirm diagnosis and assess the need for medication or other medical interventions

4. RESULT AND DISCUSSION

The results of this study highlight the performance of two machine learning models developed for thyroid disease diagnosis. The first model, the **Symptoms Analysis Model**, achieved an accuracy of 85%, while the second model, the **Report Analysis Model** , slightly outperformed it with an ac- curacy of 86%. Both models demonstrated effective classification of thyroid conditions, with distinct strengths in identifying specific classes. However, both models encountered challenges in accurately detecting Hypothyroid and Hyperthyroid conditions, as indicated by lower recall values for these categories. These results underscore the potential of machine learning techniques in the diagnostic process for thyroid disorders, while also revealing areas for improvement in capturing all possible cases of Hypothyroid and Hyperthyroid conditions.

	Precision	Recall	F1-score	Support
Low	0.82	1.00	0.90	382
Medium	0.80	0.90	0.85	414
High	0.90	0.70	0.78	404
Accuracy			0.85	1200
Macro avg	0.84	0.87	0.84	1200
Weighted avg	0.85	0.85	0.85	1200

Fig -3: Symptoms Analysis Model performance

From fig.3 the Symptoms Analysis Model aimed to classify thyroid conditions into three categories: Low, Medium, and High. It achieved an accuracy of around 85%. The model performed very well in detecting Low thyroid conditions, with perfect recall (1.00), though its precision was slightly lower at

0.82. For Medium thyroid conditions, the precision was 0.80, and recall was 0.90, showing balanced performance. However, for High thyroid conditions, the model had high precision (0.90) but lower recall (0.70), meaning it missed some true cases. Despite this, the overall performance across all classes was solid, with an average precision, recall, and F1-score of around 0.85, indicating good model reliability

	Precision	Recall	F1-score	Support
Normal	0.89	0.95	0.92	330
Hypothyroid	0.83	0.73	0.78	335
Hyperthyroid	0.86	0.78	0.82	335
Accuracy			0.86	1000
Macro avg	0.86	0.82	0.84	1000
Weighted avg	0.86	0.85	0.86	1000

Fig -4: Report analysis model performance

From fig.4 Report analysis model was used to classify thyroid conditions into Normal, Hypothyroid, and Hyperthyroid. It achieved a slightly better accuracy of 86%. The model performed excellently for Normal conditions, with high precision (0.89) and recall (0.95), resulting in an F1-score of

0.92. For Hypothyroid and Hyperthyroid conditions, the model had reasonable precision (0.83 and 0.86) but lower recall (0.73 and 0.78), meaning it missed some true cases. Overall, Model 2 showed better accuracy and generalization compared to Model 1, especially for Normal thyroid conditions, though both models had similar challenges in detecting Hypothyroid and Hyperthyroid conditions.

5.USER INTERFACE

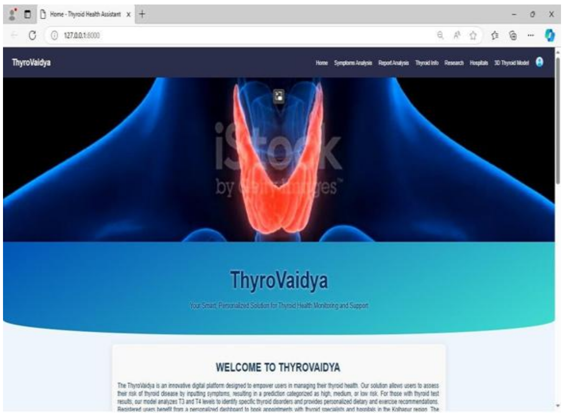


Fig -5 : User Interface Overview

Fig -6: Symptoms Analysis Form

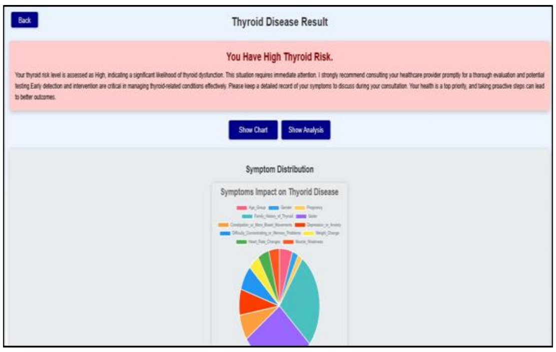
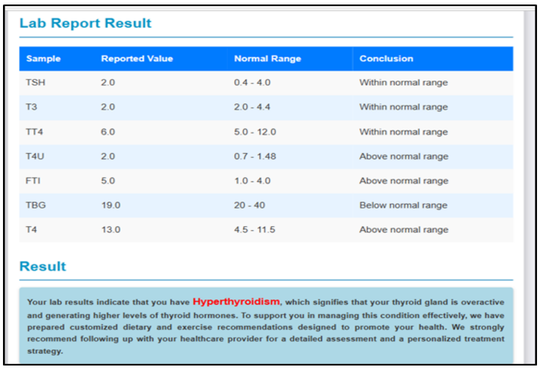


Fig -7: Symptoms Analysis Results

Fig -8: Report Analysis Form


Lab Report Result

Sample	Reported Value	Normal Range	Conclusion
TSH	2.0	0.4 - 4.0	Within normal range
T3	2.0	2.0 - 4.4	Within normal range
TT4	6.0	5.0 - 12.0	Within normal range
T4U	2.0	0.7 - 1.48	Above normal range
FTI	5.0	1.0 - 4.0	Above normal range
TBG	19.0	20 - 40	Below normal range
T4	13.0	4.5 - 11.5	Above normal range

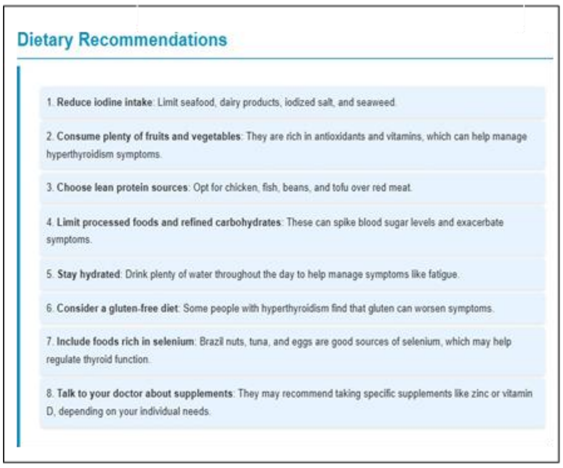
Result

Your lab results indicate that you have Hyperthyroidism, which signifies that your thyroid gland is overactive and generating higher levels of thyroid hormones. To support you in managing this condition effectively, we have prepared customized dietary and exercise recommendations designed to promote your health. We strongly recommend following up with your healthcare provider for a detailed assessment and a personalized treatment strategy.

Fig -9: Report Analysis Result

Fig -10: Recommendation

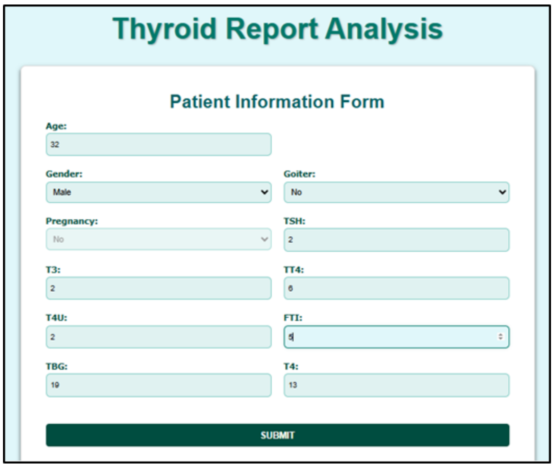
Diet Recommendation



6. TEST SAMPLE

From Table 3 machine learning model analyzed following outputs: Person 1 (25-year-old male, Hypothyroidism): This individual exhibits signs of hypothyroidism, with a normal TSH level of 1.5 mIU/L, but lower-than-expected T3 and T4 levels (1.2 ng/dL and 8.0 µg/dL), suggesting insufficient thyroid hormone production. While TSH is within the reference range, the reduced T3 and T4 indicate underactive thyroid function. Deficiencies in Vitamin D and Selenium have been identified, both of which are vital for maintaining proper thyroid activity and supporting immune function. A high-protein, low-sugar diet is recommended to help regulate metabolism and provide steady energy. Incorporating moderate cardio and yoga is suggested to enhance overall health, reduce stress, and maintain well-being.

Person 2 (45-year-old female, Euthyroid): This individual has normal thyroid function, with TSH (0.8 mIU/L), T3 (1.1 ng/dL), and T4 (6.5 µg/dL) levels all falling



within the normal range, indicating a healthy thyroid status. However, there are identified deficiencies in iron and iodine, which can impact overall health and lead to symptoms like fatigue.

TABLE III: Thyroid Parameters, Normal Ranges, and Suggested Diet/Exercise

Parameter	**Normal Range**	**Person 1**	**Person 2**	**Person 3**
Age	-	25	45	38
Gender	-	Male	Female	Female
Goiter	-	Yes	No	No
TSH	0.4 - 4.0 mIU/L	1.5	0.8	0.2
T3	80 - 200 ng/dL	1.2	1.1	2.5
T4	5.0 - 12.0 µg/dL	8.0	6.5	13.5
FTI	1.0 - 4.3	3.4	2.9	4.8
TT4	5.5 - 12.5 µg/dL	120	110	180
T4U	0.8 - 1.8	0.9	1.1	1.3
Thyroid Condition	-	Hypothyroidism	Euthyroid (Normal)	Hyperthyroidism
Nutrient Deficiency	-	Vitamin D, Selenium	Iron, Iodine	Magnesium, Calcium
Suggested Diet	-	High-protein, low-sugar	Iron-rich foods	Low-caffeine, balanced diet
Suggested Exercises	-	Moderate cardio, yoga	Strength training, stretching	Low-impact exercises, meditation

To correct these deficiencies, a diet rich in iron sources such as leafy vegetables, lean meats, and beans is recommended. Engaging in strength training exercises +and regular stretching will promote better muscle tone and flexibility. While her thyroid function is normal, addressing these nutrient gaps is crucial for maintaining optimal health.

Person 3 (38-year-old female, Hyperthyroidism): This sample displays hyperthyroidism, with low TSH (0.2 mIU/L) and high levels of T3 and T4 (2.5 ng/dL and 13.5 µg/dL), indicating an overactive thyroid. Hyperthyroidism can result in symptoms such as rapid weight loss, a racing heartbeat, and increased anxiety due to the elevated metabolic rate. The individual also shows magnesium and calcium deficiencies, which are common in those with hyperthyroidism due to the body's heightened metabolic demands. A balanced, low- caffeine diet is advised to avoid overstimulation of the thyroid and reduce stress on the body. For physical health and mental balance, low-impact exercises such as yoga and meditation are beneficial to alleviate stress and promote relaxation.

7. Conclusion

In this research a machine learning model aimed at predicting thyroid disorders by analyzing symptom-based data, offering a practical and accessible alternative to conventional lab-based tests. Designed to support early detection, the model is particularly useful in settings with limited resources, where access to laboratory facilities may be restricted. Achieving an accuracy rate of up to 86 percent, the model demonstrates a strong capability to differentiate between various thyroid conditions, allowing for early intervention when needed. Using the Random Forest algorithm, known for its robustness and accuracy, the model effectively addresses certain limitations of traditional diagnostic methods, such as high costs and the necessity for specialized equipment. This innovation promises to improve access to thyroid disorder screening and enhance patient outcomes through timely diagnosis.

Beyond diagnosis, this model includes a recommendation system that provides personalized lifestyle and dietary advice, helping individuals take proactive steps in managing their thyroid health. By delivering insights that are tailored to each individual's symptoms and risk level, the model enriches traditional healthcare approaches by connecting diagnosis with actionable, individualized support. Future work will explore broadening the dataset to include diverse populations and improving the model's accuracy to ensure it meets the needs of a wider audience. As the field of machine learning in healthcare advances, incorporating additional data types, such as demographic and genetic information, could enhance the model's predictive accuracy, contributing to the growth of personalized healthcare solutions for thyroid and other endocrine disorders.

References

1. National Institutes of Health, "Thyroid Gland and Hormone Function," NIH, accessed 2024.
2. American Thyroid Association, "Thyroid Health," ATA, 2024.
3. Mayo Clinic, "Thyroid Disease Complications," Mayo Clinic Health Information, 2024.
4. Smith, T. et al., "Thyroid function tests: an overview," Journal of Clinical Endocrinology, 2023.
5. Jones, M. et al., "The Hidden Symptoms of Thyroid Disease," Endocrinology Journal, 2023.
6. Gupta, P. et al., "Detecting thyroid disease using optimized machine learning models," Springer, 2024.
7. Aversano, L. et al., "Thyroid Disease Prediction using Machine Learning," Elsevier, 2021.
8. Chaganti, R. et al., "Machine Learning in Thyroid Disease Prediction," National Library of Medicine, 2021.
9. Vasant, C. et al., "Symptom-Based Detection of Thyroid Disease," IJEAST, 2021.
10. Saleh, D. et al., "Challenges in Diagnosing Thyroid Disease Using Machine Learning," Baghdad Science Journal, 2022.
11. R. et al., "Thyroid Disease Diagnosis Using Artificial Intelligence," Journal of Endocrine Disorders, 2023.
12. Singh, V. et al., "Predicting Thyroid Disorders with Deep Learning Models," Journal of Medical Informatics, 2024.
13. Brown, L. et al., "A Comprehensive Survey on Machine Learning Approaches for Thyroid Disease Diagnosis," Journal of Computational Medicine, 2023.
14. Chakraborty, S. et al., "Machine Learning Techniques for Predicting Thyroid Function Abnormalities," IEEE Transactions on Bioinformatics, 2024.
15. Zhang, Y. et al., "Automated Thyroid Disease Diagnosis Using Symptom-Based Data," International Journal of Artificial Intelligence in Medicine, 2023.
16. Kumar, D. et al., "Feature Selection in Thyroid Disease Prediction Models," Computers in Biology and Medicine, 2024.
17. Wang, Y. et al., "An Ensemble Learning Approach for Thyroid Disease Classification," Journal of Machine Learning Research, 2023.
18. Hassan, M. et al., "Using Machine Learning to Improve Thyroid Disorder Diagnosis: A Review," Journal of Health Informatics, 2024.

19. Sharma, N. et al., "Predicting Thyroid Disease from Clinical Data Using Support Vector Machines," Journal of Clinical Bioinformatics, 2023.
20. Liu, X. et al., "Deep Neural Networks for Thyroid Disease Prediction Based on Clinical Symptoms," Artificial Intelligence in Medicine, 2024.