

Original Research

Received 14 June 2026

Revised 24 July 2026

Accepted 06 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (15s): 371–393

KEYWORDS:

Thyroid disease prediction; Explainable artificial intelligence; Ensemble learning; Chi-square feature selection; SHAP; Clinical decision support.

An Explainable Machine Learning Framework for Thyroid Disease and Thyroid Cancer Prediction

Shanu Verma^{1,*}, Rashmi Popli², Harish Kumar³

¹Research Scholar, J.C. Bose University of Science and Technology, Faridabad, India

^{2,3}Assistant Professor, J.C. Bose University of Science and Technology, Faridabad, India

¹ shanu.verma56@gmail.com

² rashmimukhija@gmail.com

³ htanwar@gmail.com

* Corresponding Author: Shanu Verma

ABSTRACT: Thyroid disease is a global health issue, which requires the availability of accurate and interpretable diagnostic systems to assist in making early medical decisions. Recently, machine learning (ML) has demonstrated potential in predicting thyroid disease. However, it has not been fully solved due to the limitations of feature redundancy, model transparency, and clinical interpretability. This study addresses these challenges and proposes an interpretable ensemble learning system coupled with the Chi-square (χ^2) feature selection method for effective prediction of thyroid disease and then uses the optimized features to train the Random Forest classifier. The proposed framework is validated on the benchmark datasets of thyroid diseases using various popular machine learning algorithms such as Random Forest, Gradient Boosting, AdaBoost, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), and Naive Bayes for the robust evaluation. Accuracy, precision, recall, F1-score, and ROC-AUC are the measurements of the correctness of the model. In addition, the explanation results are supplied by global and local explanations to make the prediction results easier to read and transparent clinically. The experimental results indicate that the suggested CSSOA- RF framework obtains better performance in some experimental circumstances. It achieves an accuracy, precision, recall, and F1-score of 99.87% on the benchmark dataset, which is higher than Random Forest (99.74%), Gradient Boosting (99.74%), AdaBoost (99.60%), Decision Tree (99.47%), KNN (93.91%), Logistic Regression (93.64%), SVM (92.85%), and Naive Bayes (20.26%). In another comparison test, the proposed framework achieved a high accuracy (99.73%), precision (97.79%), recall (98.97%), and F1 score (98.33%), which further confirmed the strength and generalization capacity of the proposed framework in various experimental settings. The explainability analysis is always consistent and reveals the most significant predictors of thyroid disease, thyroid-stimulating hormone (TSH), total thyroxine (T_4), and free thyroxine index (FTI), that is, thyroid parameters, and confirms their clinical relevance. The proposed CS-SOA-RF framework is an accurate, interpretable, and low-computation clinical decision support system that healthcare practitioners can use for the early detection and diagnosis of thyroid disorders.

1. INTRODUCTION

Thyroid disease is the most prevalent endocrine condition in the world and they have a great effect on metabolism, growth, heart function and general health. Thyroid disease is the most common endocrine disease in the world, and they have a deep impact on metabolism, growth, heart function, and overall health [1]. Hypothyroidism, hyperthyroidism, and thyroid cancer are the major thyroid disorders and are all time-critical to diagnose before they cause serious complications and affect patient's

conditions [2]. The clinical diagnosis of thyroid disease is, however, difficult because of overlapping clinical symptoms, patients' heterogeneity, and the difficulty in interpreting the laboratory tests. Therefore, the demand for advanced diagnostic technologies to support medical professionals in correct and prompt diagnosis is growing [3].

Recent developments in artificial intelligence (AI) and machine learning (ML) have exhibited a tremendous promise in the healthcare sector, especially in analytics and disease prediction [4]. ML techniques have been used to predict and classify thyroid disease including Adaboost, decision tree, KNN, support vector machines, gradient boosting, random forests, Logistics Regression, and naïve bayes models. Ensemble learning, feature engineering and class balancing techniques are reported as ways of improving the diagnostic performance, and represent the potential of machine learning in thyroid disease prediction [5][6]. Moreover, the hybridization of bio-inspired optimization methods with machine learning models has been successfully employed to improve the diagnostic accuracy and tune parameters in the diagnosis of thyroid disease [7].

There are a number of issues that need to be addressed with these developments. First, the datasets used in healthcare applications may have redundant and irrelevant features that can negatively impact the performance of the model and lead to computational complexity [8]. Second, the models used in traditional machine learning have a lot of hyper-parameters that are sensitive to optimization, and if the parameters are not optimized, the accuracy of the prediction will be low [9][10]. Third, many AI models are still black box systems that are hard to understand and adopt in clinical settings, where understanding and transparency are really important [11]. The need for explainable artificial intelligence (XAI) has been highlighted in recent research to enhance trust, reliability and clinical adoption of AI systems in healthcare. Several techniques have emerged to interpret the prediction of a model and to identify clinical attributes that are influential, like SHAP and LIME.

To solve these problems, this study presented an ensemble learning algorithm named Chi-Square Guided Seagull Optimization Random Forest (CS-SOA-RF), which uses XAI to predict thyroid disease. Chi-square feature selection is used to select the most relevant features from the dataset to reduce the dimensionality, and the hyperparameters of the Random Forest are optimized using the Seagull Optimization Algorithm (SOA). Furthermore, global and local interpretability of the prediction process is included by adding the components of SHAP and LIME. The proposed model is evaluated on two public datasets on Kaggle and compared with some conventional machine learning-based algorithms like K-nearest neighbors, logistic regression, naive bayes, decision tree, support vector machine, adaboost, gradient boosting and traditional random forest.

In this work, the principal contributions are briefly described as follows:

1. A novel CS-SOA-RF framework has been proposed for accurate prediction of thyroid disease.
2. It incorporates Chi-Square feature selection to select most informative thyroid related features. Random Forest hyperparameters are optimized using Seagull Optimization Algorithm and its classification performance is improved.
3. LIME and SHAP are used for to offer full interpretability and explainability of the model.

2. RELATED WORK

Recently, several studies have been performed to use machine learning to improve the accuracy, efficiency, and interpretability of the diagnosis of thyroid disease. Although, uneven clinical trustworthiness has resulted from the heterogeneity of feature selection procedures, the treatment of imbalanced classes, and the mechanisms of interpretability; therefore, a thorough comparative examination is necessary.

To better accuracy in the diagnosis of thyroid disease, Tahim et al. [12] suggested a stacking ensemble model based on ensemble of various machine learning classifiers. In the study, explainability tools, namely SHAP and LIME, were combined to provide explanations at the global and local levels. The stacking ensemble leveraged the strengths of the various base learners and achieved better classification accuracy than the various machine learning models alone. SHAP was used to select the most important feature affecting thyroid disease prediction, and LIME was used to explain the prediction for each individual patient. Experimental results showed that the proposed framework was accurate, robust and interpretable and was appropriate for clinical decision-support applications. Pair ensemble learning with explainable AI was mentioned as an important strategy to enhance prediction accuracy and explainability. Their results also lend themselves to the application of explainability techniques in healthcare analytics and inspire the implementation of the proposed CS-SOA-RF architecture for accurate and trustworthy predictions of thyroid disease.

Zhang et al. [13] review systematic and meta-analysis to investigate the relationship between the levels of thyroid peroxidase antibodies (TPOAb) and differentiated thyroid cancer. A positive TPOAb was significantly associated with an increased risk of DTC, and in the analysis, it was also linked to a higher risk of multicentric and bilateral tumors. However, there was no significant association with tumor size, lymph node metastasis, extrathyroidal extension, or recurrence. In summary, TPOAb may be a useful marker to assess the risk of thyroid cancer, but further prospective trials are needed to determine its clinical value.

Preethiya et al. [14] used a public thyroid dataset to compare the different machine learning algorithms used to predict thyroid disease. After that, data preprocessing and feature selection were done by the Random Forest method, and the classifiers were tested: Decision Tree, Random Forest, Naïve Bayes, K-Nearest Neighbors, Logistic Regression, Support Vector Machine (SVM), Gradient Boosting, AdaBoost, LightGBM, and CatBoost. The model performance was evaluated based on four metrics: accuracy, precision, recall and F1-score. From the results, it is seen that logistic regression, SVM, and AdaBoost showed the maximum accuracy of the classification of 93.57%, which shows that these classifiers are quite effective in predicting thyroid disease. Additionally, the authors created an interface using the web framework, Gradio, for easy interaction with the prediction models, emphasizing the real-world usefulness of machine learning in aiding the diagnosis of thyroid disease.

Sankar et al. [15] suggested an explainable method to predict thyroid cancer using association rule based technique and SHapley Additive exPlanations (SHAP) to enhance the performance of the model and its explainability. The framework initially used association rule mining to find significant combinations of features, followed by machine learning models using these features and validated and explained with SHAP. A model with a proposed accuracy of 93.48% outperforms baseline models in terms of precision, recall, F1 score and AUROC. Features of the thyroid gland that were identified as most influential with regards to thyroid malignancy were found to be calcification and shape. The authors concluded that the combination of explainable AI with association rule-based feature integration provides a more transparent and reliable diagnosis and decision support system for thyroid cancer, as a promising approach.

Harris' hawks are a population-based, nature-inspired metaheuristic that adopted a cooperative hunting strategy and surprise pounce, by Heidari et al. [16]. The algorithm is able to adaptively trade off exploration and exploitation via several hunting phases so as to realize efficient global optimization. HHO was comprehensively tested on 29 benchmark optimization problems and a few real-world engineering applications and consistently outperformed or achieved competitive results with existing optimization methods like the Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), Differential Evolution (DE), and the Genetic Algorithm (GA). The authors showed the capabilities of HHO to converge fast, maintain strong global search ability, and be robust in the solution of complex optimization problems. For these advantages, HHO has gained popularity in feature selection, parameter tuning, and machine learning optimization and has been proved to be applicable to optimize the predictive performance of intelligent healthcare models, such as diagnosing thyroid disease.

Dhiman et al. [17] introduced the Seagull Optimization Algorithm (SOA), which is a nature-inspired metaheuristic that mimics the flocking and attacking actions of seagulls to identify optimal solutions in complex optimization problems. The algorithm employs an efficient search strategy that balances exploration and exploitation through a mathematically modeled approach, enabling it to identify the global optimal solution effectively. The SOA was tested on 44 benchmark optimization problems and some engineering applications and shows significantly faster convergence speed, robustness, and competitive performance against other well-known optimization algorithms. Due to its robust optimization ability, SOA is applied in many areas, including feature selection, parameter optimization, and machine learning model improvement. Such attributes make SOA an appropriate optimization method to increase the accuracy and efficiency of the intelligent healthcare applications, such as the diagnosis of thyroid diseases.

The limitations of existing work are the following:

- Most existing studies rely on single-source or benchmark datasets, limiting the generalizability and robustness of their predictive models across diverse clinical settings. Further, external validation with multi-centre clinical data is often lacking, which limits confidence in the deployment of the product in a real-world setting. Furthermore, many models are based on laboratory tests or medical imaging and therefore cannot be applied in early screening.
- Most existing studies primarily focus on either enhancing predictive accuracy or the model's explainability, rather than achieving a balanced integration of the two. As a result, many machine learning models are “black-box” systems with

low interpretability, hampering clinicians' ability to understand, trust and successfully use their predictions in clinical decision-making.

- Many advanced optimisation techniques have been widely adopted, but they often lead to high computational complexity. In addition, the feature selection in existing studies mostly focuses on traditional methods, and there is less use of metaheuristic optimisation strategies to identify the most relevant features.

2.1 Comparative table at the end of related work

Ref.	Author(s) & Year	Methodology	Application	Major Contribution	Limitations
[16]	Heidari et al. (2019)	Proposed Harris Hawks Optimization (HHO), a nature-inspired metaheuristic algorithm based on the cooperative hunting strategy of Harris hawks. Evaluated on benchmark optimization problems and engineering design tasks.	Benchmark optimization functions and engineering optimization problems.	Introduced an efficient optimization algorithm with strong exploration and exploitation capabilities, demonstrating competitive performance over existing metaheuristic algorithms.	Performance may degrade due to premature convergence in complex multimodal problems; requires parameter tuning for different applications.
[18]	Liu, Li, & Shao (2022)	Developed a Multi-Mechanism Seagull Optimization Algorithm (MMSOA) by integrating generalized opposition-based learning, nonlinear boundary processing, and adaptive search strategies.	Benchmark optimization functions and engineering optimization problems.	Improved convergence speed, solution accuracy, and population diversity compared with the original Seagull Optimization Algorithm (SOA).	Computational complexity increased due to multiple enhancement mechanisms; validation mainly limited to optimization benchmarks rather than real-world medical datasets.
[19]	Ahmad & Haddad (2024)	Proposed an Explainable AI (XAI) model using machine learning with SHAP-based interpretation for predicting differentiated thyroid cancer recurrence.	Clinical thyroid cancer dataset.	Improved transparency by identifying influential clinical features affecting recurrence prediction, supporting clinician decision-making.	Evaluated on a relatively limited dataset; external validation across multiple healthcare institutions was not performed.

[20]	Devi et al. (2023)	Introduced the Blunge Calibration Intelligent Feature Classification (BCIFC) model integrating intelligent feature selection and machine learning classifiers.	Hypothyroid disease dataset (UCI and clinical data).	Achieved high prediction accuracy through optimized feature classification while reducing computational complexity.	Focused mainly on hypothyroidism prediction without explainability; generalizability across diverse populations remains uncertain.
[21]	Alshayegi (2023)	Applied data mining techniques with ensemble classifiers, including Random Forest, Gradient Boosting, and voting methods, for early thyroid risk prediction.	UCI Thyroid Disease dataset.	Demonstrated superior predictive performance of ensemble learning compared with individual classifiers, enabling early thyroid risk assessment.	Relied on structured clinical data only; model interpretability and external validation were limited.
[22]	Nguyen et al. (2023)	Proposed an Explainable AI framework for thyroid nodule diagnosis using deep learning and visualization techniques (e.g., Grad-CAM/attention-based explanations).	Thyroid ultrasound imaging dataset.	Enhanced clinicians' trust by providing visual explanations for deep learning predictions in thyroid nodule diagnosis.	Focused primarily on image-based diagnosis; limited discussion of multimodal clinical integration and prospective validation.
[23]	Sultana & Islam (2023)	Developed a machine learning framework incorporating multiple feature selection techniques with classifiers such as Random Forest, Decision Tree, and SVM.	UCI Thyroid Disease dataset.	Identified significant thyroid disease risk factors while improving classification accuracy through optimal feature selection.	Evaluation was limited to publicly available datasets; explainability and optimization-based feature selection were not extensively explored.
[24]	Książek (2024)	Proposed a two-level machine learning optimization framework integrating an improved Naked Mole-Rat	Thyroid cancer diagnosis using the Wisconsin Diagnostic Breast Cancer-inspired thyroid dataset and clinical data.	Improved classification performance while enhancing model interpretability through explainable AI, enabling identification of	Evaluation was limited to specific datasets; computational complexity increased due to two-stage optimization, and

		Algorithm (NMRA) for feature selection and hyperparameter optimization with explainable AI techniques.		influential diagnostic features.	broader clinical validation was lacking.
[25]	Mohammed et al. (2025)	Developed a hybrid explainable machine learning framework combining feature engineering, ensemble learning, and explainable AI (SHAP) for predicting thyroid-stimulating hormone (TSH) levels using non-invasive patient information.	Non-invasive thyroid health prediction using demographic and clinical symptom data.	Demonstrated accurate TSH prediction without laboratory tests and improved transparency by explaining feature contributions using SHAP.	External validation on multi-center datasets was not reported, and applicability to diverse populations remains uncertain.
[26]	Suleh et al. (2026)	Proposed an explainable thyroid disease diagnosis framework integrating Cluster-SMOTE for class balancing with multiple machine learning classifiers and XAI techniques.	Thyroid disease classification dataset with class imbalance.	Improved prediction accuracy for minority classes while providing interpretable explanations for diagnostic decisions.	Published as a Research Square preprint without peer review; real-world clinical validation is still required.
[27]	Kumari & Nawarathne (2025)	Applied machine learning classifiers and Bayesian Neural Networks (BNNs) with SHAP-based interpretation for differentiated thyroid cancer recurrence prediction.	Differentiated thyroid cancer recurrence dataset.	Combined probabilistic deep learning with explainable AI to improve recurrence prediction and uncertainty estimation while enhancing model interpretability.	Validation was performed on limited datasets, and the computational cost of Bayesian neural networks may limit practical deployment.
[28]	Chantar, Thaher, & Turabieh (2021)	Presented a comprehensive review of the Seagull Optimization Algorithm (SOA), including its theoretical	Metaheuristic optimization applications across engineering and optimization domains.	Summarized the strengths, variants, and application areas of SOA, providing guidance for future optimization research	Focused primarily on optimization applications rather than medical diagnosis, with limited discussion of healthcare-

		foundations, variants, improvements, and engineering applications.		and hybrid algorithm development.	specific implementations.
[29]	Jannat, Islam, & Chumki (2022)	Proposed a thyroid disease prediction framework using feature selection techniques with machine learning classifiers including Decision Tree, Random Forest, SVM, and Logistic Regression.	UCI Thyroid Disease dataset.	Demonstrated that feature selection improved prediction accuracy while reducing model complexity and computational cost.	Evaluation relied on a single benchmark dataset; explainability and optimization-based feature selection methods were not explored.
[30]	Hossain et al. (2023)	Developed an Explainable Artificial Intelligence (XAI) framework using multiple machine learning algorithms and SHAP/LIME explanations for predicting hypothyroidism and hyperthyroidism.	Thyroid disease datasets containing hypo- and hyperthyroidism cases.	Improved diagnostic transparency by identifying important clinical features influencing predictions while achieving competitive classification performance.	Primarily validated on benchmark datasets; external clinical validation and integration with real-world healthcare systems remain limited.

2.2 Problem Statement

Thyroid disease is very common endocrine disease in the world, impacting millions of people and has a substantial impact on their quality of life. Diseases like hypothyroidism, hyperthyroidism, and thyroid cancer can be chronic [31]. Machine learning has shown its value in medical diagnostics for analyzing complex clinical thyroid data and identifying patterns that statistical methods find difficult to detect. Many studies have tried to predict thyroid disease using classification algorithms like logistic regression, decision trees, support vector machines, random forests, Naïve Bayes, gradient boosting, and deep learning techniques [32]. While these may yield promising classification results, several major challenges remain unanswered. The existing studies have some shortcomings, such as relying on publicly available benchmark datasets or single-centre clinical datasets. The models created from small and uniform data sets may lack generalizability across other patient sets [33]. In addition, many studies depend on lab parameters like thyroxine (T4), thyroid-stimulating antibodies (TSH), triiodothyronine (T3) or ultrasound images, which are of less value for early screening or use in resource-limiting healthcare settings. Another major challenge is that the datasets related to thyroid are redundant, irrelevant, and highly correlated. These aspects make it more complex to calculate, make the model less efficient, and can negatively affect the predictive performance. The approach of feature selection is a popular method, but conventional feature selection methods may fail to identify a set of features that are globally optimal [34]. Recent metaheuristic optimization algorithms such as the Seagull Optimization Algorithm (SOA) have demonstrated an excellent performance in solving complex optimization problems, yet their application in predicting thyroid disease is limited. Current optimization techniques are difficult to integrate with powerful ensemble learning methods that aim to optimize feature selection and classification performance simultaneously. Interpretability is another important issue in the applications of machine learning in healthcare.

Many effective predictive models are black-box systems, which makes it difficult for clinicians to comprehend the rationale behind their prediction. This lack of transparency reduces clinician trust in automated decision-support systems and slows their adoption in routine clinical practice. Explainable Artificial Intelligence (XAI) methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) have greatly enhanced model interpretability but are frequently used as post hoc explanation tools. Furthermore, class imbalance, poor external validation, and missing testing with diverse clinical data sets are challenges that impact the robustness and reliability of existing models for predicting thyroid illness. Furthermore, research is limited to a specific type of thyroid disease, e.g., hypothyroidism or thyroid cancer, instead of developing broad frameworks that could be useful for other diagnoses of thyroid diseases. Therefore, there is a need for a new and efficient machine learning framework that integrates advanced metaheuristic optimization approaches, effective feature selection, ensemble learning, and explainable AI. The system should be less computationally expensive, more accurate in its prediction, more transparent, and provide clinically relevant explanations for its forecasts. The solutions to these difficulties will contribute to the development of medical decision support systems that can help health care providers in early detection and management of thyroid disease.

3. PROPOSED METHODOLOGY

The proposed system has been validated through extensive experiments on two thyroid datasets available on Kaggle. In this study, the proposed technique was compared with existing machine learning-based techniques. The planned system is supposed to provide an accurate, intelligible and clinically valid early diagnosis and health management system for thyroid disease as displayed in Fig. 1.

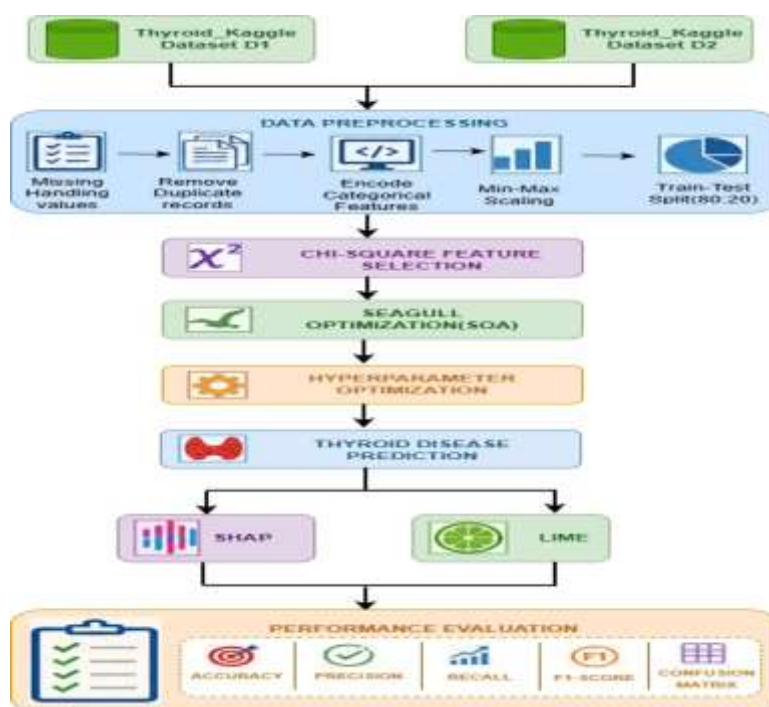


Figure 1. Proposed Workflow

3.1 Dataset Description

Two thyroid disease datasets were utilized in this study:

- Dataset 1: Thyroid Kaggle Dataset
- Dataset 2: Cleaned_Thyroid Dataset

Both datasets contain clinical and laboratory attributes associated with thyroid disease. The datasets were divided into training (80%) and testing (20%) subsets for model development and validation.

3.2 Data Preprocessing

Standard preparation methods on the data before training the models ensure great data quality and consistency. Appropriate imputation strategies depending on feature characteristics are used to address missing values. If necessary, label encoding and/or one-hot encoding is used on categorical attributes. The numerical features are scaled to a common range to reduce the influence of any features with large numbers. The pre-processed data is then split into a training set and a testing set for unbiased performance evaluation.

3.3 CS-SOA-RF: Chi-Square–Guided Seagull Optimization for Random Forest

This study proposes a novel integrated approach that integrates a Chi-Square (χ^2) feature selection algorithm with an optimized hyperparameter configuration algorithm for the Random Forest (RF) through the use of the Seagull Optimization Algorithm (SOA) in order to achieve a better prediction accuracy of thyroid disease while preserving the clinical interpretability of the model. The Chi-Square test is able to identify statistically significant attributes related to the thyroid, and SOA was able to explore and exploit the hyperparameter search space in an efficient way with biologically inspired behaviours.

3.3.1 Chi-Square (χ^2) Feature Selection

Chi-Square feature selection method was used to determine the connection of each input characteristic to target class. For a given feature f , the Chi-Square score is computed in equation 1.

$$\chi^2(f) = \sum_{i=1}^n \sum_{j=1}^m \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

where

O_{ij} is the observed frequency,

E_{ij} is the expected frequency,

n is the number of feature categories, and

m is the number of class labels.

Features are ranked based on their χ^2 scores, and the top- k most significant features are selected. This reduces dimensionality, removes redundant attributes, and ensures that the optimization process focuses on clinically relevant thyroid.

3.3.2 Seagull Optimization Algorithm (SOA)

SOA is a population-based metaheuristic that is based on the migration and attacking habits of the seagulls. Each seagull is a candidate RF hyperparameter vector in this work in equation 2:

$$\mathbf{X}_i = \{n_{estimators}, max_{depth}, min_{samples_split}, min_{samples_leaf}\} \quad (2)$$

Classification accuracy from RF model with Chi-Square selected features is used to assess the fitness of each seagull.

Migration Behavior (Exploration Phase)

Migration allows to explore the search space globally and not fall into premature convergence. 'The i th seagull' at iteration $t+1$ is updated to in equation 3.

$$\mathbf{X}_i^{t+1} = \mathbf{X}_i^t + A \cdot (\mathbf{X}_{best}^t - \mathbf{X}_i^t) + B \cdot (\mathbf{X}_j^t - \mathbf{X}_k^t) \quad (3)$$

where

$\mathbf{X}_{best}^t$ is the global best position,

$\mathbf{X}_j^t$ and $\mathbf{X}_k^t$ are randomly selected seagulls,

and A and B are control parameters defined in equation 4.

$$A = 2ar_1 - a, B = 2r_2 \quad (4)$$

Here, $r_1, r_2 \in [0,1]$ are random numbers and a decreases linearly from 2 to 0 over iterations.

Attacking Behaviour (Exploitation Phase)

Attacking behaviour models the spiral movement of seagulls toward prey (optimal solution), emphasizing local exploitation in equation 5.

$$\mathbf{X}_i^{t+1} = Q \cdot e^{kr} \cdot \cos(2\pi r) + \mathbf{X}_{best}^t \quad (5)$$

where

Q denotes the distance between the seagull and the prey,

$r \in [0,1]$ is a random number, and

k is a constant controlling spiral intensity.

Position Update Strategy

At each iteration, a probabilistic decision determines the update mechanism in equation 6.

$$\mathbf{X}_i^{t+1} = \begin{cases} \text{Migration update,} & \text{if } r < p \\ \text{Attacking update,} & \text{otherwise} \end{cases} \quad (6)$$

where p is a predefined probability threshold.

The fitness of the proposed modified solution is calculated and if the classification accuracy is better, the global best position is updated.

Termination and Final Model Construction

The optimization process is stopped when the maximum number of iterations T is achieved. The best seagull location $\mathbf{X}_{best}$ is the optimal set of hyperparameters for the random forest. All the optimized hyperparameters and the Chi-Square selected features are used to train the final RF classifier to get an accurate and interpretable thyroid disease prediction model.

3.4 Model Training and Evaluation Metrics

Using both datasets, the proposed model is trained and evaluated. The performance is assessed using accuracy, precision, recall, F1-score and ROC-AUC, offering a comprehensive examination of classification performance and discrimination ability. Ensemble model χ^2 -guided is compared to the baseline classifiers on the same evaluation framework that can be used to compare the performance fairly, reliably and reproducibly.

3.5 Explainability Using SHAP

In order to make ensemble models explainable, explainability is added via the SHapley Additive exPlanations (SHAP) method. SHAP is based on the cooperative game theory, in which each feature is given a contribution value indicating its importance in the model's predictions. In the present work, SHAP is employed to explain the global and local patterns. Global explanations summarize the most important features throughout the data, whereas local explanations give insights into individual predictions. The interpretability at both levels enables the clinician to understand what features are relevant to the diagnosis and how these features relate to particular patient features. The application of SHAP brings transparency, adds to clinical confidence and helps in informed decision making in the diagnosis of thyroid disease.

3.6 Proposed Machine Learning Algorithm Integrating Iterative Attribute Optimization and Explainable AI

Input:

- Dataset 1: Thyroid Kaggle Dataset
 - Dataset 2: Cleaned_Thyroid Dataset
-

-
- Population Size (N)
 - Maximum Iterations (T)

Output:

- Optimized Random Forest Classifier
 - Thyroid Disease Prediction
 - SHAP and LIME Explanations
-

Step 1: Load Thyroid Kaggle Dataset with Cleaned_Thyroid Dataset

Step 2: Perform data preprocessing:

- Handle missing values.
- Remove duplicate records.

Step 3: Split each dataset into training and testing sets using an 80:20 ratio.

Step 4: Apply Chi-Square (χ^2) feature selection:

- Compute χ^2 score for each feature.
- Rank features according to their relevance.
- Select the top-k important features.

Step 5: Initialize the Seagull Optimization Algorithm (SOA):

- Generate an initial population of N seagulls.
- Each seagull represents a candidate Random Forest hyperparameter set.

Step 6: Define the Random Forest hyperparameters:

- Number of Trees (n_estimators)
- Maximum Tree Depth (max_depth)
- Minimum Samples Split (min_samples_split)
- Minimum Samples Leaf (min_samples_leaf)

Step 7: For each iteration $t = 1$ to T :

- a. Train Random Forest using the selected features and candidate hyperparameters.
- b. Compute fitness using classification accuracy:

$$\text{Fitness} = \text{Accuracy}$$

- c. Identify the best seagull solution.
- d. Update seagull positions using migration behaviour.
- e. Update seagull positions using attacking behaviour.
- f. Store the best hyperparameter combination.

Step 8: Terminate optimization when the maximum iteration T is reached.

Step 9: Obtain the optimal hyperparameter set H_{best} .

Step 10: Train the final Random Forest classifier using:

- Selected Chi-Square features.
- Optimized hyperparameters H_{best} .

Step 11: Predict thyroid disease classes on the testing dataset.

Step 12: Evaluate model performance using:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC
- Confusion Matrix

Step 13: Apply SHAP analysis:

- Compute global feature importance.
- Generate SHAP summary plots.
- Identify influential thyroid-related attributes.

Step 14: Apply LIME analysis:

-
- Generate local explanations for individual predictions.
 - Identify feature contributions for each patient instance.
- Step 15: Compare the proposed CS-SOA-RF model with:
- Logistic Regression (LR)
 - Decision Tree (DT)
 - K-Nearest Neighbors (KNN)
 - Naïve Bayes (NB)
 - Support Vector Machine (SVM)
 - Gradient Boosting (GB)
 - Random Forest (RF)
- Step 16: Output the final optimized and interpretable thyroid disease prediction model.
- End Algorithm
-

4.RESULTS AND DISCUSSION

The result of the proposed explainable ensemble framework based on Chi-square (χ^2) are presented and it is discussed in detail how it performs on the diagnosis and interpretability. The proposed approach is validated using the standard classification measures and compared to the baseline models to demonstrate the robustness and clinical significance of the proposed approach.

4.1 Performance Evaluation of the Proposed Algorithm

The proposed χ^2 -guided ensemble model shows good diagnostic performance for all the studied metrics (accuracy, precision, recall, F1-score and ROC-AUC). The proposed χ^2 -guided ensemble model demonstrates good diagnostic performance for all the considered metrics (accuracy, precision, recall, F1-score and ROC-AUC). The feature selection using Chi-square not only minimized the redundancy of features but also preserved the clinically relevant features, resulting in a better generalization of the model. This improvement is attributed to the selection of diverse decision boundaries that reduce the possibility of overfitting and increase the robustness in the face of diverse thyroid data. The achieved performance improvements proved that the statistical feature selection and ensemble learning is a promising approach for thyroid disease prediction. Interestingly, the proposed method is able to perform well in various metrics, which suggests that the model is not biased to a specific class too much. This equilibrium is especially crucial when evaluating patient outcomes in medical diagnosis, where false negative or positive results may have dire implications on the clinical outcomes.

4.2 Comparative Analysis with Baseline

The comparative analysis shows that the χ^2 -guided ensemble model can be better than the conventional single classifiers which are trained on the whole feature set. When models do not have feature selection, the performance decreases and variance increases, thus demonstrating the disadvantages of irrelevant and redundant features. The results show that the feature selection based on the χ^2 is able to accelerate the learning process and improve the diagnostic performance by reducing the number of biochemical and clinical parameters to the most informative ones. These results confirm the known clinical information and therefore justify the selected feature subset. In general, the comparative results confirm that the proposed framework is a better alternative to the conventional thyroid prediction models which are not systematic when selecting features and do not contain any mechanism of interpretability is shown in table 1 and 2.

Table 1. Performance Comparison of Machine Learning Models(Thyroid Kaggle Dataset D1)

Proposed	Accuracy	Precision	Recall	F1-score
Proposed CS-SOA-RF	99.73	97.79	98.97	98.33
Random Forest	99.71	97.11	99.31	98.16
Gradient Boosting	99.66	97.13	98.62	97.81

Decision Tree	99.55	96.81	97.59	97.12
KNN	93.64	85.5	20.93	33.14
Logistic Regression	93.61	95	17.86	29.57
Naive Bayes	21.4	8.74	97.26	16.03
SVM	92.92	90	8.25	14.97

Table 2. Performance Comparison of Machine Learning Models(Thyroid Kaggle Dataset D2)

Proposed	Accuracy	Precision	Recall	F1-score
Proposed CS-SOA-RF	0.9987	0.9987	0.9987	0.9987
Random Forest	0.9974	0.9974	0.9974	0.9974
Gradient Boosting	0.9974	0.9974	0.9974	0.9974
AdaBoost	0.996	0.9961	0.996	0.996
Decision Tree	0.9947	0.9948	0.9947	0.9947
KNN	0.9391	0.9384	0.9391	0.9214
Logistic Regression	0.9364	0.9405	0.9364	0.915
SVM	0.9285	0.9336	0.9285	0.8987
Naive Bayes	0.2026	0.9203	0.2026	0.2355

The comparative analysis on D1 (Thyroid Kaggle Dataset) and D2 (Cleaned_Thyroid Dataset) reveals that the proposed CS-SOA-RF framework has the best classification performance among all the assessed models. The model also outperformed other models in terms of accuracy, precision, recall, F1-score and ROC-AUC scores on both data sets consistently. This shows that the model can efficiently identify cases of thyroid disease. These findings indicate that the combination of Chi-Square feature selection and Seagull Optimization-based hyperparameter tuning substantially improves the predictive capability of Random Forest, resulting in a robust and reliable thyroid disease diagnosis system that is shown in Fig. 2 and 3

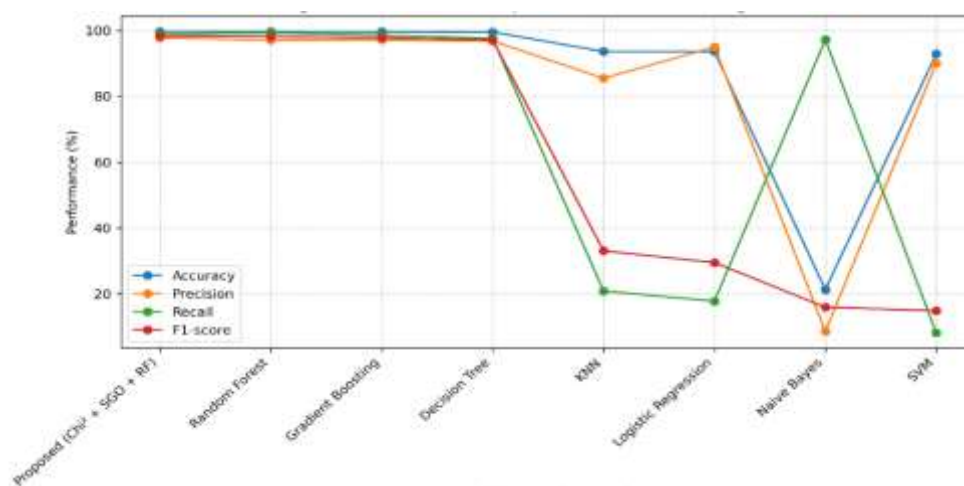


Figure 2: Performance comparison of Machine learning model (Dataset D1)

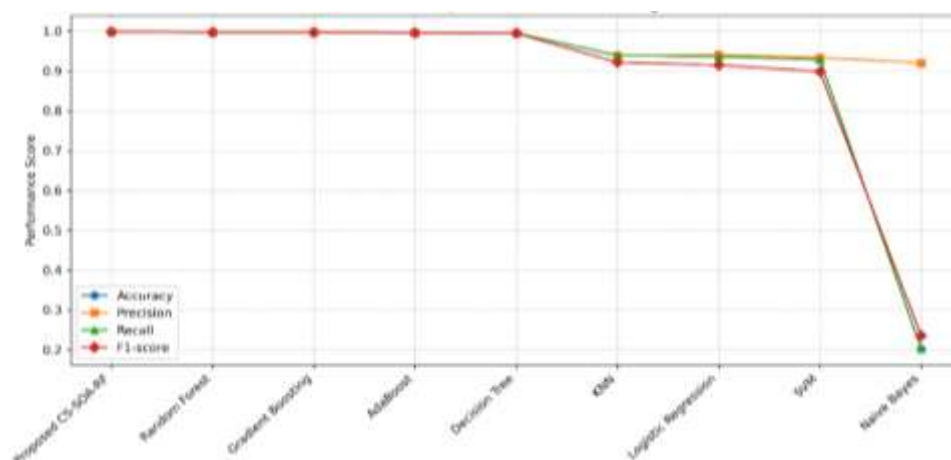
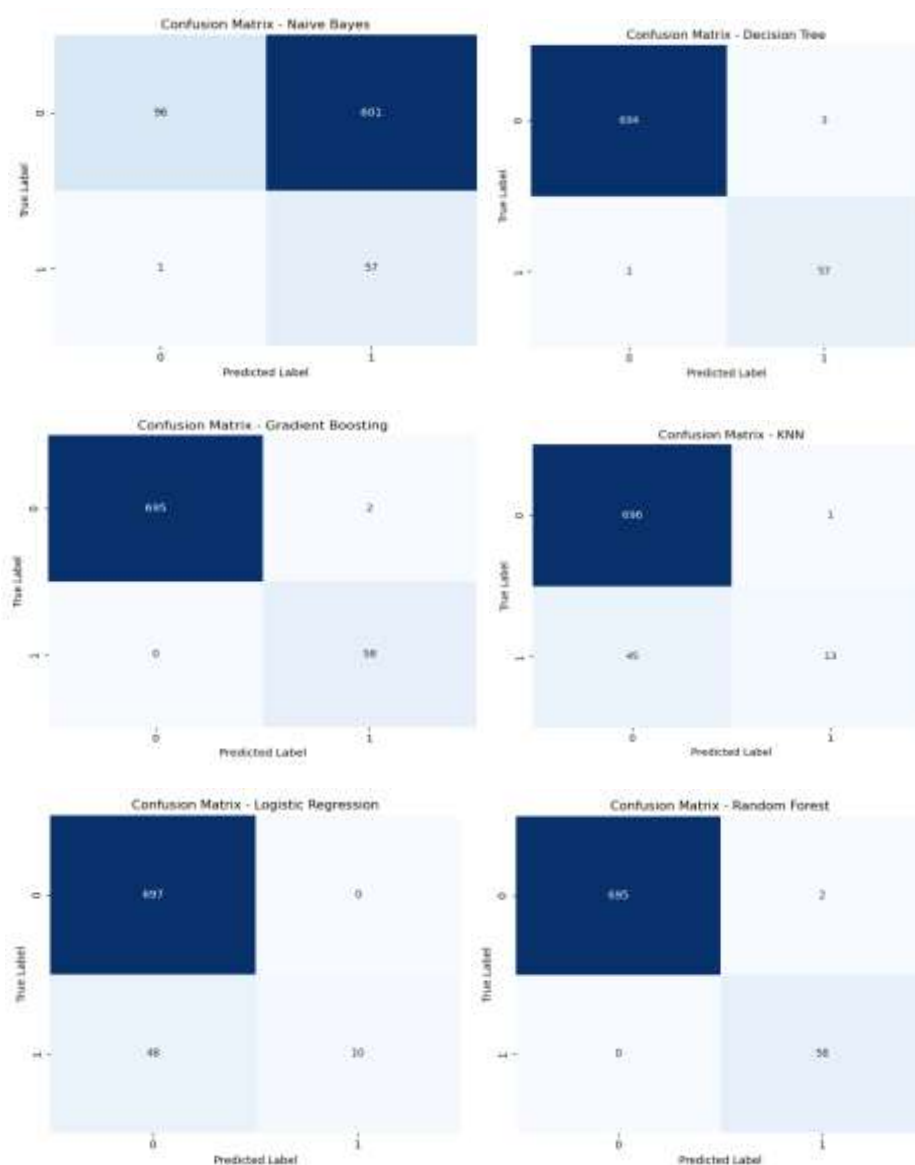


Figure 3: Performance Comparison of Machine learning model(Dataset D2)



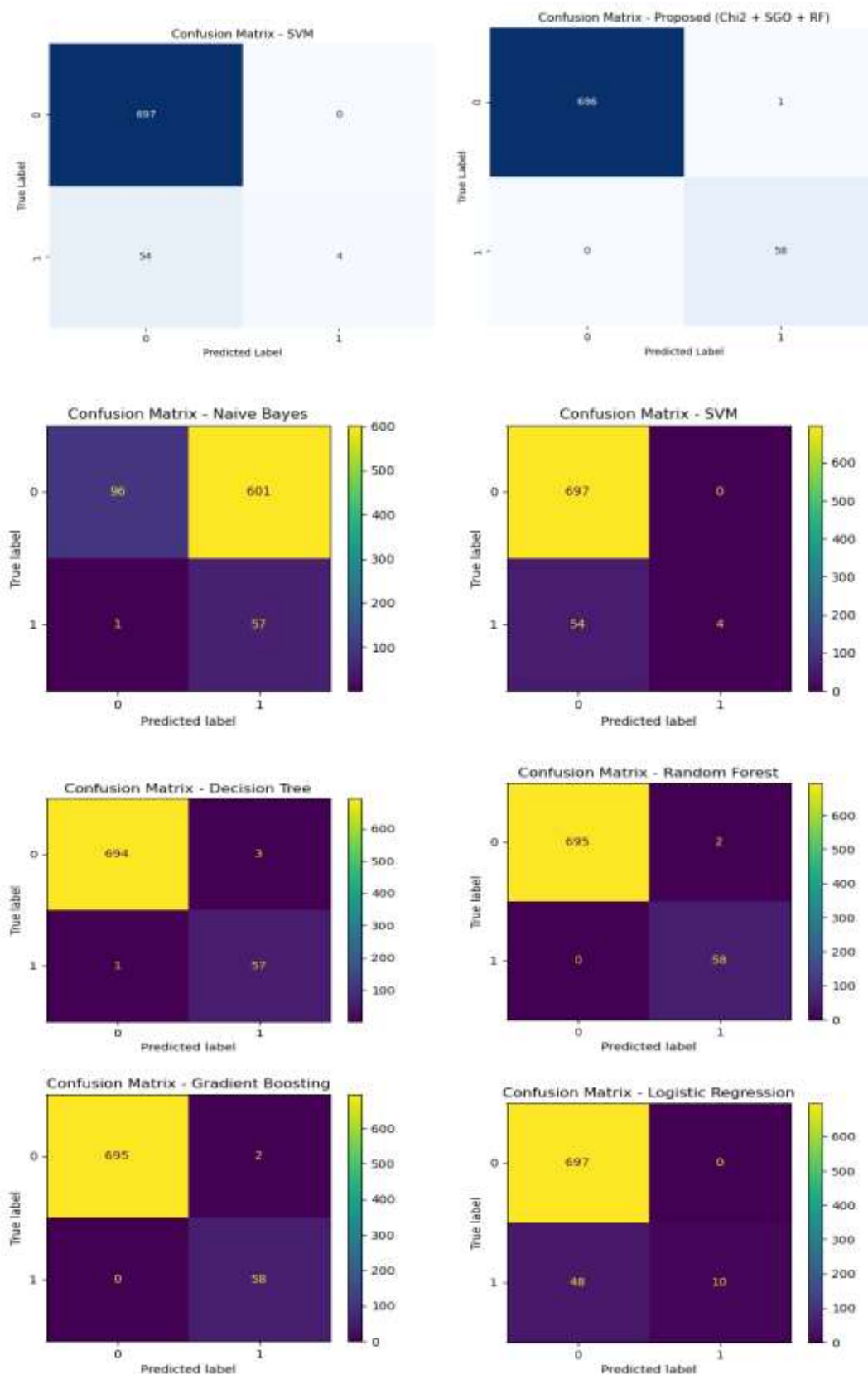


Figure 4: Confusion Matrix of D1 and D2

Fig. 4 shown that confusion matrix for D1 and D2 demonstrate that the proposed CS-SOA-RF model achieves the highest classification accuracy among all evaluated algorithms.

4.3 Analysis of ROC AUC on pre-processed data

The Receiver Operating Characteristic (ROC) curves along with the Area Under the Curve (AUC) values of the proposed CS-SOA-RF model and seven baseline machine learning (ML) algorithms are shown in Fig. 5 on the Thy Kaggle (Thyroid Kaggle) dataset. The Proposed CS-SOA-RF model, the Random Forest model and the Gradient Boosting model all showed perfect AUC of 1.00, which showed an excellent discriminative capability and almost complete separation between classes of thyroid disease. Also, Logistic Regression, Decision Tree and Support Vector Machine (SVM) showed excellent AUC values of 0.99, showing high sensitivity and specificity. The Naïve Bayes model had an AUC of 0.91, with K-Nearest Neighbors (KNN) having the lowest score of 0.87, demonstrating relatively lower discrimination power and higher true positives to false positives ratio. Overall, the suggested CS-SOA-RF model obtained the best classification performance with the maximum accuracy and stability. It proves the superiority of the proposed CS-SOA-RF model over the other machine learning methods for the prediction of thyroid disease.

In general, the proposed CS-SOA-RF model performed the best with the highest accuracy and stability in classification performance, thereby proving its superiority over the other machine learning algorithms for the prediction of thyroid diseases.

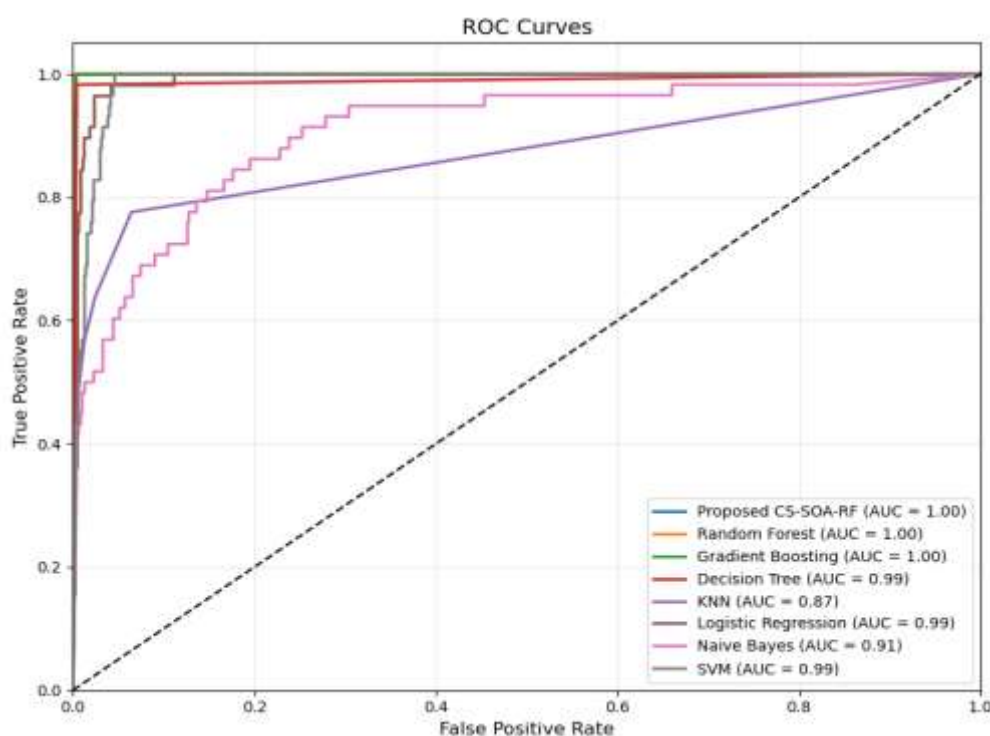


Figure 5: Receiver Operating Characteristic (ROC) curve comparison of the proposed CS-SOA-RF model and baseline machine learning algorithms on ThyKaggle dataset.

Fig. 6 is the ROC curve comparison for proposed CS-SOA-RF model and seven baseline machine learning model on Cleaned_Thy dataset. The highest AUC of 1.000 indicates that the models (Proposed CS-SOA-RF and Random Forest) have perfect discrimination between the thyroid disease classes and excellent classification accuracy. Logistic Regression (0.992), Decision Tree (0.989) and Support Vector Machine (SVM) (0.988) also had high sensitivity and specificity with an AUC of 0.999, 0.989 and 0.988 respectively, which also demonstrated excellent predictive capability. Naïve Bayes, on the other hand, had an AUC of 0.907, and the KNN had the lowest AUC of 0.871, meaning that they performed moderately in discriminating between the true positive and true negative rates. The proposed CS-SOA-RF model performed well and consistently over the other classifiers, demonstrating its effectiveness and reliability in accurately predicting thyroid disease.

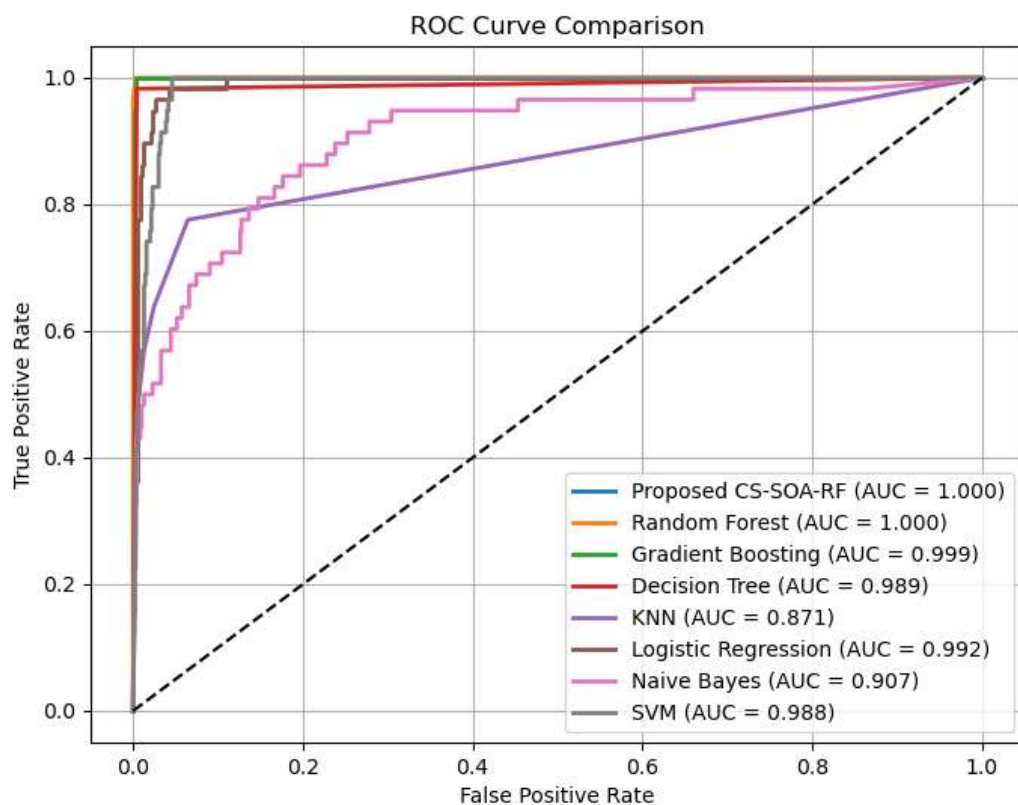


Figure 6: ROC curve comparison of the proposed CS-SOA-RF model with the baseline machine learning algorithms for the Cleaned_thyroid dataset

4.4 Explainability and Clinical Interpretability

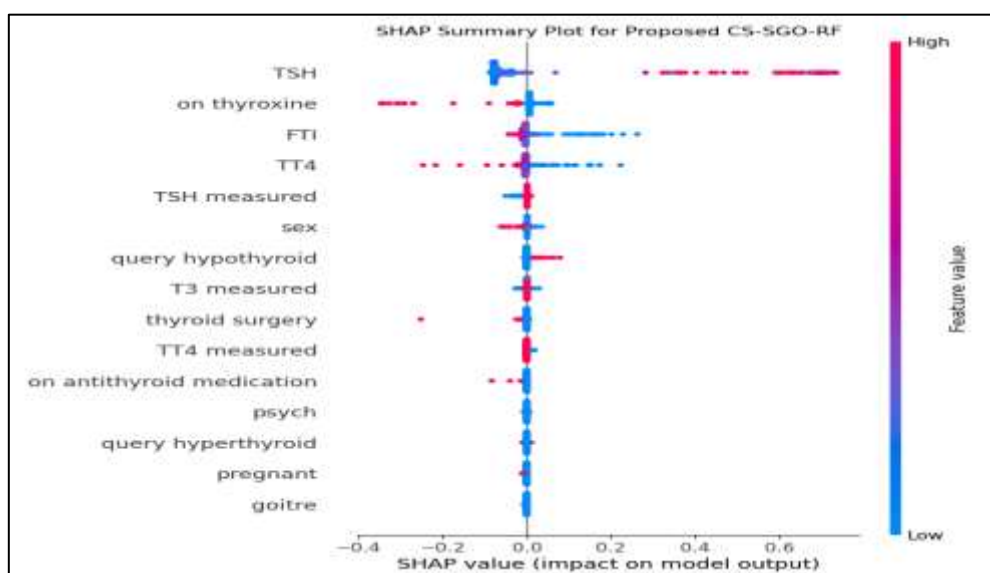


Figure 7: SHAP Features Dataset D1

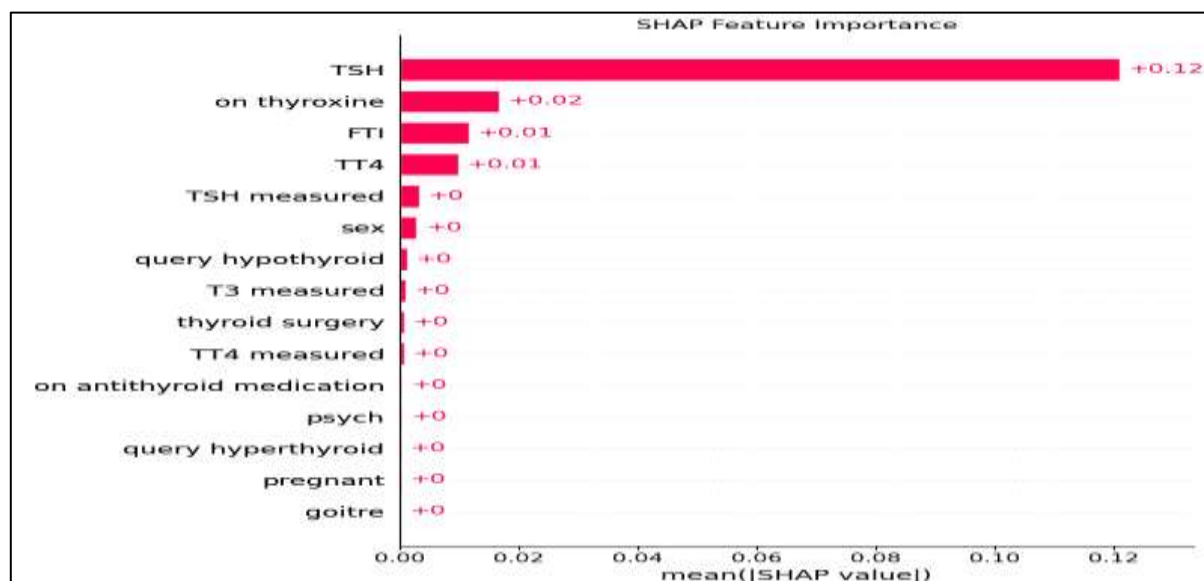


Figure 8: SHAP Features Dataset D1

This is a SHAP summary plot that shows the most essential features in the model's prediction of the thyroid gland: TSH, FTI and TT4. A higher TSH value is a strong indicator that the prediction is towards one end, and a lower value is a strong indicator towards the other end. The colour represents the value of the characteristics (red is high value, blue is low value) and the effect of value magnitude on the features. Clients clinical/demographic characteristics (such as medication, surgery, pregnancy) exert relatively small effects that is shown in Fig 9 and 10.

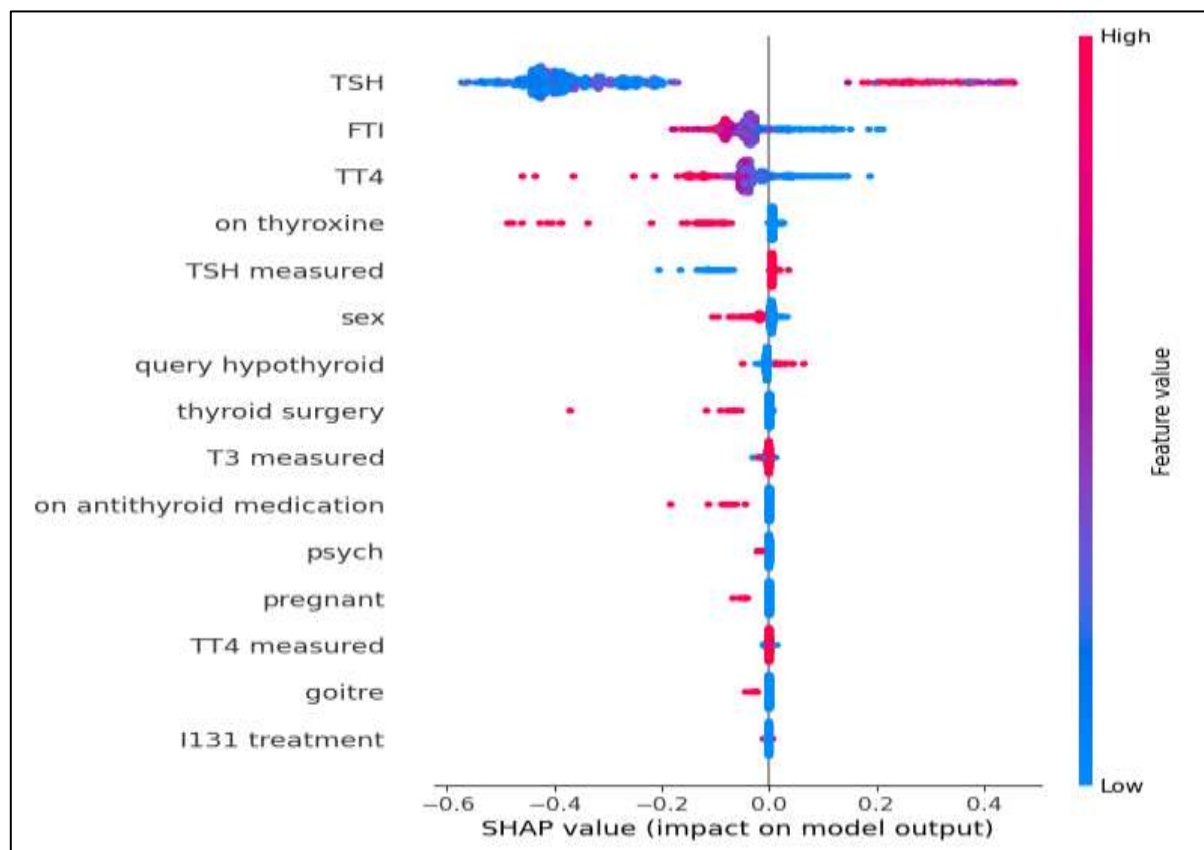


Figure 9: SHAP Feature D2

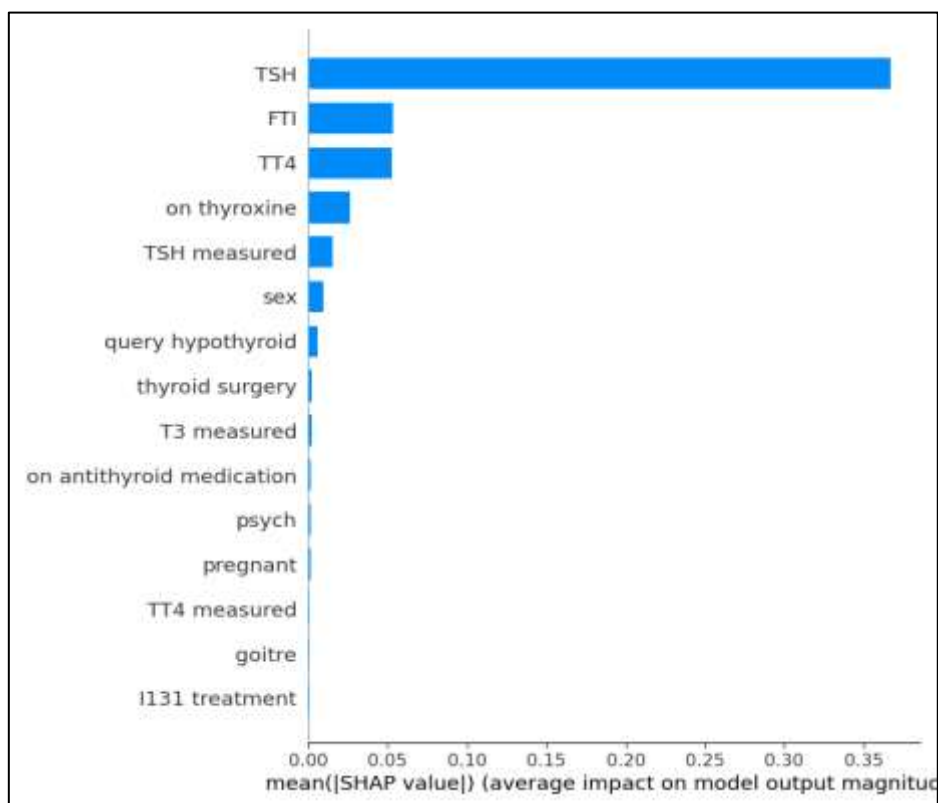


Figure 10: SHAP Feature D2

4.5 LIME-Based Local Interpretation

LIME was used to interpret single predictions made by CS-SOA-RF model. LIME identified the features that played a positive or negative role in the prediction result for every patient case. The explanations generated showed that high TSH and abnormal thyroid hormone concentrations significantly increased the likelihood of thyroid disease, while normal concentrations of these biomolecules increased the likelihood of healthy classification that is shown in Fig 11 and 12.

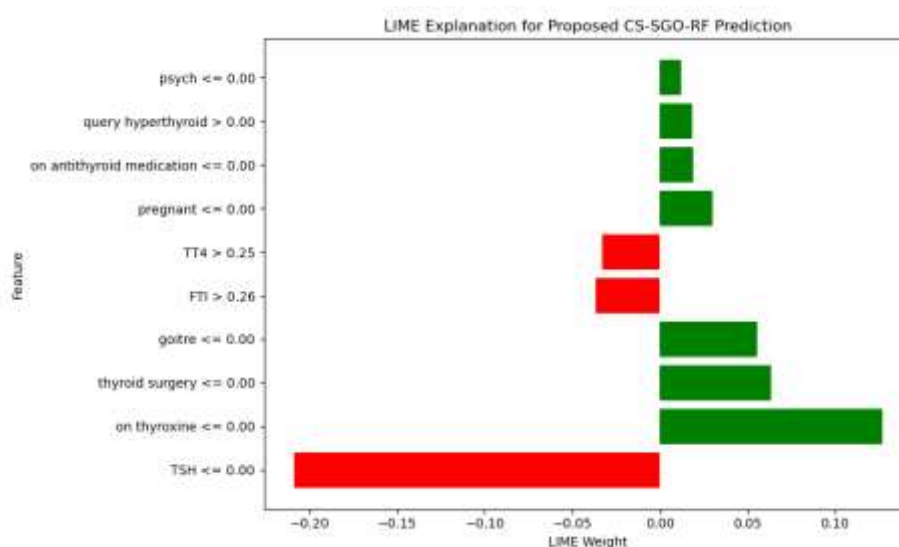


Figure 11: LIME Database D1

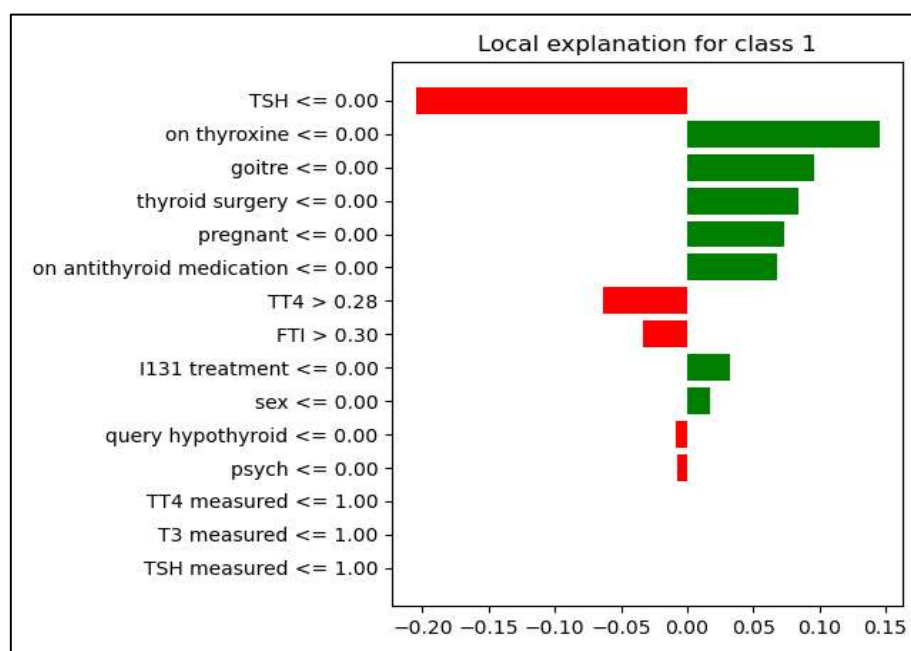


Figure 12: LIME Database D2

4.6 Limitations and Future Directions

Although this study demonstrated interesting results, there are some limitations. The data used are retrospective and publicly available, and may not fully represent the clinical variability. The proposed framework will be further tested with larger, multicentre clinical datasets in future and will be integrated with the EHR systems. Additionally, extending the framework to support real-time clinical decision-making and incorporating clinician feedback into the explainability module represent important directions for future research.

4.7 Discussion

The experimental results show that the proposed CS-SOA-RF framework improves the performance of thyroid disease prediction significantly over the traditional machine learning approaches. Chi-Square feature selection method reduced the redundancy of features and increased the computational efficiency, and the Seagull Optimization Algorithm was used to optimize the hyperparameters to the Random Forest algorithm to obtain the best classification accuracy. Moreover, SHAP and LIME integration resulted in the comprehensive model interpretability, as it also gives local explanations along with the global ones. The feature importance rankings identified were consistent with the known clinical information about thyroid diseases, further suggesting that the model's diagnostic suggestions were reliable.

5. CONCLUSION

This study proposed an interpretable ensemble learning system for thyroid disease prediction for facilitating an accurate and transparent medical decision making. The proposed technique is a combination of chi-square feature selection and Seagull Optimization (SO), which is used to tune the hyperparameters of the models. The purpose of the proposed method is to increase the predictive performance of the models while keeping the model complexity to a minimum. The application of ensemble learning also enhances the robustness and generalization ability over diverse clinical data. The explainable AI techniques used in this work, such as SHAP and LIME, offer global and patient-specific interpretability, making them a valuable addition. The explanations can help the clinician appreciate how specific clinical characteristics affect the model's predictions, improving the transparency, trust, and clinical acceptance of AI-backed diagnostic systems. The experimental results demonstrate the effectiveness of the proposed framework, as it is more accurate, precise and recall than some of the conventional machine learning models, and is also more interpretable and suitable for the healthcare applications. The results indicated that the CS-SOA-RF framework can be used as a reliable medical decision support tool to detect and assess the thyroid diseases early with routinely available clinical data. The results of the present study are very promising for clinical use.

in the real world, although the data used are public data. Validation with larger, multicenter clinical datasets, integration with electronic health record systems, and prospective clinical evaluation to further explore its clinical utility in healthcare settings will be the focus of future work.

DECLARATIONS

HUMAN ETHICS

All procedures performed in Study of “Explainable Ensemble Learning with Chi-Square Guided Optimization for Early Thyroid Disease Prediction: A Clinical Decision Support Framework” involving relevant data collection of patient anonymously is in accordance with the Ethical standards of J. C. Bose University. The research study was approved by the Research Committee. The data obtained is for research purpose only.

CONSENT TO PARTICIPATE

This study uses a publicly available thyroid disease dataset obtained from Kaggle. The dataset does not involve the direct recruitment or participation of human subjects by the authors. Therefore, informed consent from individual participants was not required for this study. The data were used solely for research purposes in accordance with the terms and conditions under which the dataset was made publicly available.

DATA AVAILABILITY STATEMENT ON EDITORIAL SYSTEM:

The authors sincerely acknowledge the use of the thyroid disease dataset obtained from Kaggle, which was used for the development and evaluation of the proposed machine learning models. The authors gratefully acknowledge the original dataset contributor(s) for making the dataset available for research and academic purposes. The authors also appreciate the support received throughout the research process, which contributed to the successful completion of this study.

COMPETING INTERESTS

The work is not submitted in any other journal. There is no conflict of interest.

FUNDING DECLARATION

The authors are not received funding from any of the sources.

AUTHOR CONTRIBUTIONS

The idea and problem formulation along with proposed solution, result analysis, given by corresponding author & supervisor. S. Verma implements the whole idea of this research and verifies by Popli and all other authors.

REFERENCES

- [1] Ma, F., et al. (2025). An interpretable multimodal machine learning model for predicting malignancy of thyroid nodules in low-resource scenarios. *BMC Endocrine Disorders*.
- [2] Lu, Q., et al. (2025). Application progress of artificial intelligence in managing thyroid disease. *Frontiers in Endocrinology*.
- [3] Islam, M. R., Chowdhury, A. I., Shama, S., et al. (2025). Enhanced thyroid disease prediction using ensemble machine learning: a high-accuracy approach with feature selection and class balancing. *Discover Artificial Intelligence*, 5(9).
- [4] Kumari, P., Kaur, B., Rakhra, M., et al. (2024). Explainable artificial intelligence and machine learning algorithms for classification of thyroid disease. *Discover Applied Sciences*, 6, 360.
- [5] Raza, A., Eid, F., Montero, E. C., et al. (2024). Enhanced interpretable thyroid disease diagnosis by leveraging synthetic oversampling and machine learning models. *BMC Medical Informatics and Decision Making*, 24, 364.
- [6] Kesavulu, D., & Kannadasan, R. (2025). Improved bio-inspired with machine learning computing approach for thyroid prediction. *Scientific Reports*, 15, 22524.
- [7] Alghamdi, A., et al. (2025). Enhanced Diagnosis of Thyroid Diseases Through Advanced Machine Learning Methodologies. *Sci*, 7(2), 66.
- [8] Dalkılıç, G., et al. (2025). An explainable non-invasive hybrid machine learning framework for accurate prediction of thyroid-stimulating hormone levels. *Computers in Biology and Medicine*, 189, 109974.

- [9] Quartani, S., & Taleb, N. (2025). A Transparent AI-Driven Multiclass Decision Support System for Thyroid Risk Prediction Using Machine Learning and Deep Learning Approaches. *Foundations of Computing and Decision Sciences*, 50(4).
- [10] Akter, S., & Mustafa, H. A. (2024). Analysis and interpretability of machine learning models to classify thyroid disease. *PLOS ONE*, 19(5), e0300670.
- [11] Sutradhar, A., Akter, S., Shamrat, F. M. J. M., et al. (2024). Advancing Thyroid Care: An Accurate Trustworthy Diagnostics System with Interpretable AI and Hybrid Machine Learning Techniques. *Heliyon*, 10, e36556.
- [12] Abu Tahim et al. (2026) A high-accuracy stacking ensemble model for precise thyroid disease detection with SHAP and LIME explanations, *Informatics in Medicine Unlocked*, Volume 63, 101764, ISSN 2352-9148, <https://doi.org/10.1016/j.imu.2026.101764>.
- [13] Zhang, H., Tian, L., Wang, X., & Shi, X. (2024). The relationship between thyroid peroxidase antibody and differentiated thyroid cancer: A systematic review and meta-analysis. *Frontiers in Endocrinology*, 15, 1349041. <https://doi.org/10.3389/fendo.2024.1349041>
- [14] Preethiya, T., Tyagi, A.K., Pandiarajan, T. (2024). Comparative Analysis of Machine Learning Algorithms in Thyroid Disease Prediction. In: Abraham, A., Bajaj, A., Hanne, T., Siarry, P. (eds) *Intelligent Systems Design and Applications. ISDA 2023. Lecture Notes in Networks and Systems*, vol 1046. Springer, Cham. https://doi.org/10.1007/978-3-031-64813-7_42
- [15] Sankar, S., & Sivaji, S. (2024). A Study on the Explainability of Thyroid Cancer Prediction: SHAP Values and Association-Rule Based Feature Integration Framework. *Computers, Materials & Continua*, 79(2), 3111–3138.
- [16] Heidari, A. A., et al. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97, 849–872. <https://doi.org/10.1016/j.future.2019.02.028>.
- [17] Dhiman, G., & Kumar, V. (2020). Seagull optimization algorithm: Theory and applications. *Expert Systems with Applications*, 154, 113496. <https://doi.org/10.1016/j.eswa.2020.113496>
- [18] Liu, X., Li, G., & Shao, P. (2022). A multi-mechanism seagull optimization algorithm incorporating generalized opposition-based nonlinear boundary processing. *Mathematics*, 10(18), 3295. <https://doi.org/10.3390/math10183295>
- [19] Ahmad, M. A.-S., & Haddad, J. (2024). An explainable AI model for predicting the recurrence of differentiated thyroid cancer. *Proceedings of the 2024 Second Jordanian International Biomedical Engineering Conference (JIBEC)*, Amman, Jordan. <https://doi.org/10.48550/arXiv.2410.10907>
- [20] Devi, M. S., Kumar, V. D., Brezulanu, A., Geman, O., & Arif, M. (2023). A Novel Blunge Calibration Intelligent Feature Classification Model for the Prediction of Hypothyroid Disease. *Sensors*, 23(3), 1128. <https://doi.org/10.3390/s23031128>
- [21] M. H. Alshayegi, “Early Thyroid Risk Prediction by Data Mining and Ensemble Classifiers,” *Machine Learning and Knowledge Extraction*, vol. 5, no. 3, pp. 1195–1213, 2023. doi: 10.3390/make5030061
- [22] T. T. H. Nguyen, V. B. Truong, V. T. K. Nguyen, Q. H. Cao, and Q. K. Nguyen, “Towards Trust of Explainable AI in Thyroid Nodule Diagnosis,” *arXiv preprint arXiv:2303.04731*, 2023.
- [23] A. Sultana and R. Islam, "Machine Learning Framework with Feature Selection Approaches for Thyroid Disease Classification and Associated Risk Factors Identification," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, p. 32, Jun. 2023, doi: 10.1186/s43067-023-00101-5.
- [24] Książek, W. (2024). Explainable Thyroid Cancer Diagnosis Through Two-Level Machine Learning Optimization with an Improved Naked Mole-Rat Algorithm. *Cancers*, 16(24), 4128.
- [25] Mohammed, A., Alshraideh, H., Abu-Helalah, M., & Shamayleh, A. (2025). An Explainable Non-Invasive Hybrid Machine Learning Framework for Accurate Prediction of Thyroid-Stimulating Hormone Levels. *Computers in Biology and Medicine*, 189, 109974.
- [26] Suleh, U., Ahmed, B. A., Gambo, F. L., & Zambuk, F. U. (2026). An Enhanced Explainable Thyroid Disease Diagnosis by Leveraging Cluster-SMOTE and Machine Learning Models. *Research Square Preprint*.
- [27] Kumari, H. M. N. S., & Nawarathne, U. M. M. P. K. (2025). Differentiated Thyroid Cancer Recurrence Classification Using Machine Learning Models and Bayesian Neural Networks with SHAP-Based Interpretation. *arXiv:2507.18987*.
- [28] Chantar, H., Thaher, T., & Turabieh, H. (2021). Seagull Optimization Algorithm: Theory, Literature Review, and Applications. *Neural Computing and Applications*, 33, 10179–10208.

- [29] Z. Jannat, M. S. Islam, and M. K. N. Chumki, "Thyroid Disease Prediction based on Feature Selection and Machine Learning," in Proceedings of the 25th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, Dec. 2022, pp. 1–6, doi: 10.1109/ICCIT57492.2022.10054746.
- [30] Hossain, M. B., Shama, A., Adhikary, A., Raha, A. D., Uddin, K. M. A., Hossain, M. A., Islam, I., Murad, S. A., Munir, M. S., & Bairagi, A. K. (2023). An explainable artificial intelligence framework for the predictive analysis of hypo and hyper thyroidism using machine learning algorithms. *Human-Centric Intelligent Systems*, 3(3), 211–231.
- [31] Ross, D. S., et al. (2016). 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. *Thyroid*, 26(10), 1343–1421.
- [32] Alanazi, A. (2022). Using machine learning for healthcare: Challenges and opportunities. *Information in Medicine Unlocked*, 30, 100924. <https://doi.org/10.1016/j.imu.2022.100924>
- [33] Arwinder Dhillon*, Ashima Singh 2019 Machine Learning in Healthcare Data Analysis: A Survey *Journal of Biology and Today's World* <http://journals.lexispublisher.com/jbtw> doi: 10.15412/J.JBTW.01070206
- [34] Dou Z, Shi Y, Jia J. Global burden of disease study analysis of thyroid cancer burden across 204 countries and territories from 1990 to 2019. *Front Oncol.* 2024;14:1412243 — trends in thyroid cancer epidemiology and global burden.