Library Progress International
Vol.44 No. 3, Jul-Dec 2024: P.9154-9168

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

An Optimized Deep Ensemble Super-Learner Model For Thyroid Disease Classification

C.Uma[1] and Dr.P.Rathiga[2]

1. Research Scholar, Department of Computer Science, Navarasam Arts and Science College for Women, Arachalur, Bharathiyar University, Erode, Tamilnadu, India.
umachinnasamy@gmail.com
2. Assistant Professor, PG and Research Department of Computer Science , Erode Arts and Science College, Bharathiyar University, Erode, Tamilnadu, India.
rathigaerodearts@gmail.com

How to cite this article: C.Uma, P.Rathiga (2024) Measuring the Success of Agile HR Initiatives in Virtual Talent Acquisition. *Library Progress International,* 44(3), 9154-9168.

ABSTRACT

Early detection of thyroid diseases is crucial to mitigate severe consequences. various Artificial Intelligence (AI) models have been developed to identify and classify thyroid diseases. But, certain models that rely on a single classifier often suffer from high prediction errors and lower generalizability and accuracy. To solve this, Optimized Deep Ensemble Super-Learner Model (ODESLM) is developed for thyroid disease prediction. In this method, Deep Ensemble Super-Learner Model (DESLM) is designed for feature selection tasks with six base learners like Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Random Forest (RF), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU).The Super Learner Model (SLM) optimally combines and weights the predictions of base models to minimize the overall prediction error. The ensemble method uses resampling or reweighting to build a stacked model to capture capturing diverse patterns for enhancing the accuracy of thyroid disease identification. The Output values of DESLM is optimized using Secretary Bird Optimization Algorithm (SBOA) which is inspired by secretary Birds' foraging behavior to avoid incorrect classifier selection in training datasets. SBOA refines training datasets, optimizes hyperparameters and identifies relevant features from base models to enhance the super-learner's performance. It adjusts model weights and parameters to balance bias and variance enhancing generalization and reducing overfitting. This optimized ensemble model improves accuracy on larger datasets with lowers predictive error in thyroid disease prediction.Experimental results show that proposed model achieves 92.73% accuracy on the thyroid disease dataset, outperforming traditional models.

Keywords: Thyroid Diseases, Ensemble Models, Super Learners, Secretary Bird Optimization Algorithm

I . INTRODUCTION

Thyroid diseases refer to a range of medical conditions affecting the thyroid gland, which is a butterfly-shaped gland located at the base of the neck [1]. The thyroid gland produces hormones that regulate many aspects of metabolism and overall bodily function. When the thyroid gland is not functioning properly, it can lead to various disorders [2].The symptoms of thyroid gland's dysfunction include hair loss, fatigue, weight gain, increased heart rate and dry skin. If left untreated, thyroid dysfunction can escalate into more severe health issues, including joint pain, infertility, obesity and heart disease [3]. Thyroid disease essential for regulating metabolism has become increasingly prevalent in recent years.

Hyperthyroidism and hypothyroidism are prevalent thyroid diseases caused by imbalances in thyroid hormone production. Hyperthyroidism results from excessive levels of thyroid hormones like T4 and T3 leading to symptoms such as rapid weight loss and increased heart rate [4]. Conversely, hypothyroidism is characterized by insufficient hormone levels, causing fatigue, weight gain, and cold sensitivity [5]. In India, approximately 42 million people are affected by thyroid diseases, a high prevalence attributed to factors such as genetic predisposition, iodine deficiency and advancements in medical awareness and diagnostic capabilities [6]. These conditions highlight the importance of early detection and effective management to mitigate their impact on health.

Traditional diagnostic methods like blood tests and physical examinations which can demand specialized expertise and might delay diagnosis [7]. The manual process is often time-consuming and prone to inaccuracies due to the extensive data analysis required [8]. To improve the effectiveness of thyroid disease detection and prevention, it is essential to integrate advanced technologies into healthcare practices. Artificial Intelligence (AI) including Machine Learning (ML) and Deep Learning (DL) have emerged as effective approaches to address these challenges and provide early warnings for disease prevention [9]. These approaches can accurately predict and classify thyroid diseases such as hypothyroidism and hyperthyroidism [10].

In recent years, numerous studies have explored thyroid diseases and their symptoms using statistical analysis, ML or DL models. However, the accuracy and robustness of these models can vary. Many studies rely on publicly available datasets, leading to issues such as model overfitting and imbalanced class problems, where the positive (disease) class has significantly fewer samples. To address this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is commonly used [11], which balances the dataset before training ML or DL models. On the other hand, it can introduce noise and compromise model efficiency.

Additionally, existing studies often focus on a limited number of thyroid diseases and binary-class classification, limiting their generalizability. Generative Adversarial Network (GAN) has great potential to address these limitations. A GAN network trains a generator to create realistic synthetic data by fooling a discriminator. The generator aims to produce samples that are indistinguishable from real data to achieve equilibrium with the discriminator. Since thyroid disease data is tabular, Gupta et al. [12] utilized the CTGAN model to generate synthetic data with a specific discrete value and balance class distribution. In this model, the generator is conditioned to adjust the tabular data input and receive additional information to generate samples according to the specified class conditions. This approach helps mitigate model bias and overfitting, but it suffers from mode collapse. This is often due to the discriminator getting stuck in a local minimum, causing the generator to repeatedly produce outputs that deceive the discriminator. This can make training CTGAN difficult, especially with large datasets.

For this reason, Hybrid Optimized GAN (HOGAN) model [13] to address imbalanced data in thyroid disorder prediction. HOGAN combines Variational Auto-Encoder (VAE) and Optimized CTGAN (OCTGAN) with Secretary Bird Optimization Algorithm (SBOA) to generate data for minority classes and prevent mode collapse by improving class-conditioning in the latent space. VAE creates meaningful latent representations from majority class data, which OCTGAN uses to generate balanced data for minority classes. OCTGAN enhances training stability and performance with adversarial training, KL divergence loss and Wasserstein gradient penalty. The balanced dataset supports the training of classifiers like SVM, Random Forest (RF), CNN and BiGRU for improved disease classification accuracy.However, the single classifier in [13] shows good results for thyroid disease classification but might struggle with generalization in large datasets, potentially losing important information. Ensemble learning enhances robustness by combining multiple models with meta-learners specifically reducing error variance and leveraging model strengths. However, effectively ensembling models is challenging task while poor selection can decrease accuracy compared to a single classifier. Successful ensemble optimization requires careful tuning and choosing complementary models to improve generalization and accuracy.

Hence, in this paper, ODESLM is designed to solve above-mentioned issues forefficient thyroid disease prediction.In this method,Deep Ensemble Super-Learner Model (DESLM) is designed for feature selection tasks. The diverse base models, including Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Random Forest (RF), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) are employed. Each base model offers unique insights into the data and their predictions are combined to enhance robustness and accuracy. The Super Learner Model (SLM) optimally combines and weights the predictions of base models to minimize the overall prediction error. The ensemble method involves rearranging training datasets through resampling or reweighting and building a stacked ensemble to integrate these predictions effectively. This approach improves thyroid disease identification by capturing diverse patterns, reducing errors balancing bias and variance for more stable and reliable outcomes.

The output values of DESLM is optimized using SBOA to enhance optimal classifier selection and avoid incorrect choices while updating training datasets.SBOA is an optimization technique inspired by the foraging behavior of secretary Birds is used to fine-tune the ensemble model. In this model, SBOA refines training datasets, optimizes the hyperparameters and identifies relevant features from base models to improve the super-learner's performance. It adjusts model weights and parameters to balance bias and variance, enhancing generalization and reducing overfitting. SBOA represents model presence or absence in the bird's population, selecting those with the highest fitness and diversity. The optimized ensemble forms the final super-learner, which uses these predictions to accurately classify test data, ensuring improved accuracy for thyroid disease prediction. This model effectively enhances the generalization and accuracy for thyroid disease prediction in large datasets, while reducing prediction errors.

Below sections are prepared as follows: Section 2 covers related studies. Section 3 explains the proposed model and Section 4 showcases validation results. Section 5 précises the work and suggests potential improvements.

II. LITERATURE SURVEY

Research on predicting thyroid diseases with ML and DL is extensive. Early classification of conditions like cancer, hyperthyroidism, and hypothyroidism can greatly improve treatment and recovery. This section summarizes recent advancements in thyroid disease classification methods.

Sultana & Islam [14] conducted a study on thyroid disorder classification using ML algorithms. This model employs SMOTE for class imbalance, robust scaling for data normalization and feature selection methods like Boruta, Recursive Feature Elimination (RFE) and Least Absolute Shrinkage and Selection Operator (LASSO). The SVM, Decision Tree (DT), GBM, K-Nearest Neighbor (KNN) and RF were trained for thyroid disorder classification. They also identified key risk factors for thyroid disorder but faced limitations due to a small sample size and missing data.

Hossain et al. [15] created an interpretable AI model that utilized multiple ML algorithms to predict hypo and hyperthyroidism. It involved collecting, pre-processing and resampling a thyroid dataset. Subsequently, the most significant features were selected and classified using different ML algorithms, including DT, RF, Gradient Boosting Machine (GBM), Naive Bayes (NB), logistic regression (LR), KNN and SVM. Still, the model's accuracy suffered from neglecting overfitting issues.

Rathinakumar & Priya, [16] constructed an Optimized Multi-Layered Long Short-Term Memory (OML-LSTM) neural network with residual attention mechanism for thyroid diseases identification. Then, attention mechanism enhances feature relationship capture by focusing on key features or time steps.The LSTM with attention mechanisms was integrated to improve sequential learning and feature extraction resulting in accurate predictions. But, improperly tuned hyperparameters resulted in longer training times and reduced accuracy.

Mohan et al. [17] suggested a devised a multi-model approach for thyroid prediction. The collected images were pre-processed and segmented using bias reduction and optimized Otsu method. The hybrid Black Widow Optimization and Mayfly Optimization Approach (HBWO-MOA) was used for the feature extraction. Finally, VGG-19 with LSTM was used for thyroid disease classification. But, this model was sensitive to noise and limited to global thresholding.

Ji et al. [18] developed a Tree based self-stack ensemble model (TSSEM) for thyroid disease prediction. In this method, resampling technique was applied to address data discrepancies. Then, down-sampling was used in the majority of classes resulting in a balanced distribution of samples across each target variable. A randomized RF based stacked ensemble model was created to achieve highly accurate detection of thyroid disease. But, it leads to valuable information loss and reduced model generalization.

Obaido et al. [19] suggested aFilter-Based Feature Selection with Stacking Ensemble Model (FFS-SEM) for thyroid disease prediction. In this model, filter-based feature selection was performed to eliminate irrelevant features and improves models performance and efficiency. The stacked ensemble model was used to combine multiple model predictions, potentially identifying rare thyroid conditions that single models might miss for the disease detection. But, it results in poor performance when training on larger dataset.

Sha [20] integrated quantum computing with ML(QCML) to enhance thyroid disorder prediction. This model used a modified Quantum Particle Swarm Optimization (QPSO) for global feature selection and Quantum SVM (QSVM) for classification of thyroid disorder presence. Despite improvements, accuracy suffered due to class imbalance, and computational complexity increased with larger datasets.

Balasree et al. [21] constructed Multi-Layer Recursive Neural Network (ML-ReNN) for thyroid detection. Initially, pre-processing improves data quality through cleaning, splitting and handling missing values. Feature selection was performed using Fisher score method to choose optimal features. Data analysis was performed using Region-of-Interest (ROI). Finally, ML-ReNN was used to classify the thyroid diseases into normal, hyperthyroid and hypothyroid. But, computationally intensive and resource-demanding due to their complex architecture.

Prasad & Santhosh, [22] suggested a Multi-Stage Weight Ensemble Learning Model (MS-WELM) for Thyroid disorder prediction. Initially, the collected dataset was pre-processed and outliers were predicted. Then MS-WELM was introduced to fine-tune the weight and ensemble the models weight for efficient thyroid disorder prediction. However, training time was longer due to the computational complexity of weight fine-tuning and model stacking.

The studies on thyroid disease detection using ML and DL methods report varied results in accuracy, precision, recall and F1 score, but face key limitations. Small datasets reduce generalizability, while imbalanced datasets often lead to biased models and overfitting. Although some models achieve high accuracy, they may struggle to generalize, especially for minority classes, leading to high prediction errors. Overfitting is a recurring issue particularly for minority predictions. This study addresses these challenges by proposing an optimized ensemble model aimed at improving generalizability, reducing prediction errors and enhancing accuracy in thyroid disorder classification.

III. PROPOSED METHODOLOGY

The proposed ensemble model is detailed in this section. The key processes for predicting and classifying thyroid diseases are shown in Figure 1. These processes include data collection, preprocessing, augmentation and proposed ensembled models for thyroid diseases classification.

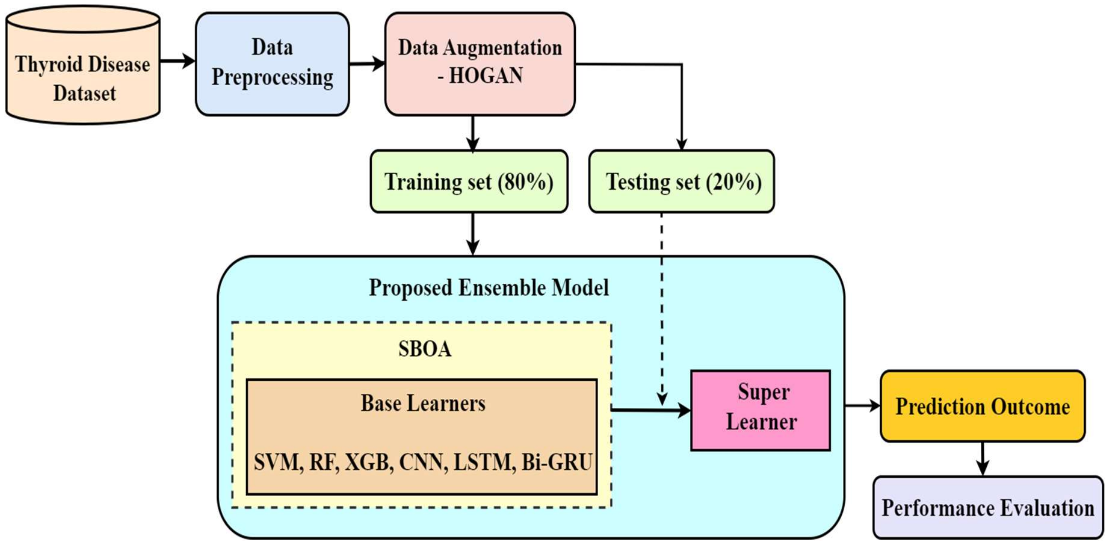

Figure 1. Design of Proposed Methodology

3.1 Dataset Description

In this study, the thyroid disease dataset is obtained from the UCI Machine Learning Repository [23]. It contains 9172 samples, each with 31 features. It comprises 6771 samples of normal individuals and the remaining samples have hypothyroidism and hyperthyroidism. Table 1 provides detailed descriptions of the features in this dataset.

Table 1. Detailed Description of the Features in the Thyroid Disease Dataset

Features	Descriptions
Age	Age of the patient in years
Sex	Gender of the patient
On thyroxine	Having triiodothyronine or not
Query on thyroxine	Type of thyroid hormone
On antithyroid meds	Prevent the thyroid form
Sick	About a person's sickness
Pregnant	Is patient pregnant
Thyroid surgery	Remove a part or all of the thyroid gland
I131 treatment	Radioactive iodine therapy
Query hypothyroid	Patient's query hypothyroid
Query hyperthyroid	Makes more hyperthyroid hormones
Lithium	Used to treat certain psychiatric diseases
Goitre	Swelling in front of the neck
Tumor	An abnormal mass of tissue
Hypopituitary	Fails to produce more hormones

Psych	Combining forms denoting the mind
TSH measured	Measure hormone is in patient's blood
TSH	Thyroid-stimulating hormone
T3 measured	Involves having patient blood drawn
TT4 measured	Measures the label of total T4
TT4	Total thyroxine
T4U	Thyroxine utilization rate
T4U measured	Measures or not
FTI	Thyroxine use rate
FTI measured	Measures or not
TBG	Thyroxine-binding globulin
TBG measured	Thyroxine-binding globulin measured or not
T3	Lab report for triiodothyronine
Referral source	Referencing source
Class	Normal (no thyroid conditions), hypothyroidism (underactive thyroid) and hyperthyroidism (overactive thyroid)

3.2 Data Pre-processing and Data Augmentation

Data pre-processing is vital for all classification models as it helps extract insights from the dataset. This model uses a label encoder to convert categorical features into numeric vectors, features with over 50% missing values, including T3, TSH, Age, T4U, TT4, TBG, FTI, and Sex, are excluded excludes and applies mean interpolation to impute missing values in the remaining features. For data augmentation, HOGAN model is proposed by combining VAE and OCTGAN models to address data imbalance in thyroid diseases prediction. The data pre-processing and augmentation are briefly illustrated in [13]. The proposed ensemble model for thyroid diseasesprediction is illustrated in below sections.

3.3 Proposed Ensemble Model for Thyroid disease prediction

The complete framework of the proposed ensemble model for thyroid diseases classification are illustrated in below sections

3.3.1 Considered Models for Ensemble Learning

In this method, SVM, XGB, RF, CNN, LSTM and Bi-GRU are considered as base learners which are briefly illustrated below.

SVM: It is a supervised learning algorithm that finds the hyperplane that best separates data into different classes. It uses kernel functions (e.g., linear, polynomial, Radial Basis Function (RBF)) to transform data into higher dimensions for better separation. It is effective in high-dimensional spaces and with a clear margin of separation. It's robust against overfitting, especially in high-dimensional datasets.

XGB: This model is based on gradient boosting, where weak learners (typically decision trees) are combined to create a strong predictive model. It optimizes a loss function and uses regularization to prevent overfitting. It effectively handles missing data well, and is highly efficient in terms of speed and performance. It provides feature importance scores and is effective with both large datasets and complex relationships.

RF: RF is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It uses bagging (bootstrap aggregating) to improve robustness. It eliminates overfitting compared to individual decision trees handles large datasets well and provides feature importance scores.

CNN: It is one sort of DL model designed for grid-like data, such as images. They consist of convolutional layers (for feature extraction), pooling layers (for down-sampling) and fully connected layers (for classification) It captures spatial hierarchies in data in which the images or similar data are used. It automatically learns relevant features reducing the need for manual feature extraction.

LSTM: It is types of Recurrent Neural Network (RNN) designed to learn and remember long-term dependencies in sequential data. They have memory cells and gates that regulate the flow of information. It is well-suited for time-series or sequential data, as they can capture temporal dependencies and patterns that other models might miss.

Bi-GRU: It is the variation of the (Gated Recurrent Unit) which processes data in both forward and backward directions. This bidirectional approach helps capture information from past and future contexts. It is useful for capturing dependencies in both directions in sequential data, which can be beneficial for understanding complex patterns in disease progression. Hence, combining these models in an ensemble can leverage their individual strengths to improve the accuracy and robustness of thyroid disease prediction.To evaluate these models, repeated Stratified K-Fold Cross-Validation is employed, ensuring each fold retains the class distribution of the entire dataset and providing a reliable performance estimate. This process involves averaging the accuracy across multiple folds and iterations to assess each model's generalization ability accurately. The mean accuracy obtained from these evaluations is then used as the fitness function in the SBOA [13]. This approach allows for the optimal selection and combination of base learners by incorporating their individual performances into the optimization process, ultimately aiming to enhance the overall predictive accuracy and robustness of the ensemble model.

3.3.2 Utilizing SBOA for Selecting Base Classifier Subsets

In this model, SBOA is used to select the base classifiers from a pool of candidates. By following below steps, SBOA can effectively selects the subset of base classifiers for arobust ensemble model.

Step 1: Initialization and Candidate Evaluation

The total number of base classifiers is analyzed and train each classifier on the same validation dataset to ensure comparability. Evaluate their performance using metrics such as accuracy, precision, recall, and F-measure. Record these metrics in a structured format, such as a list or matrix, to facilitate comparison. This setup allows for an informed selection of the best-performing classifiers.

Step 2: Application of SBOA

- **Exploration Phase:** SBOA explores a subset of classifiers to get an estimate of their performance. This is analogous to the "observation" phase in the secretary problem where the goal is to gather information about the candidates. For instances, define a threshold or proportion p of classifiers to evaluate before making a selection. For example, if $p = 0.5$, then evaluate the first 3 classifiers to get a sense of their performance.
- **Selection Phase**:After the exploration phase, use the performance metrics from the observed classifiers to set a benchmark.In the selection phase, choose classifiers that have performance metrics better than the best observed in the exploration phase.

Step 3: Subset Formation

Formulate the final subset of classifiers based on the selection criteria established in the previous phase.The aim is to select a subset of classifiers that, when combined, optimize performance metrics and improve the overall model.

Step 4: Ensemble Formation

Combine the selected classifiers into your Ensemble Super-Learner Model (ESLM) using methods such as stacking, bagging, or boosting to leverage their strengths and improve prediction accuracy.

Step 5: Evaluation and Tuning:

Evaluate the performance of the ESLM with the selected classifiers on a test dataset.Fine-tune the parameters and hyperparameters of the selected classifiers to further optimize the ensemble performance.

3.3.4 Illustration of Super Learner Model

A SLM [24] in ensemble learning is a sophisticated method designed to combine different base models to produce a more accurate and robust predictive model.SLM extends the idea of meta-learning by focusing on how to effectively combine and weigh predictions from various base models to minimize prediction errors. The key feature of SLM is the inclusion of a meta-learner, which is responsible for integrating the predictions from the base models. This meta-learner is trained using the base models' predictions along with the actual target labels. Its role is to determine the optimal weights for each base model's prediction, ultimately generating a final, improved prediction through this weighted combination.The below steps provides the construction of super learner model.

Step 1: The base learners or candidate learner S are the algorithms selected to construct the super learners.

Step 2: The candidate learner are trained and validated until all the k parts of the blocks hve been reversed as the validation set.

Step 3: As the reference to figure 2, a matrix is constructed with the R outputs of the candidate learners in the validation block. The matrix becomes the input of the validation block. The predictions are stored with the real value x of the determined samples.

Step 4: Themodel construction (z) and fitting for the regression of observed results (β) onto outcomes predicted by candidate learnings as

$$E(y|R) = m(z;\beta) \tag{1}$$

Where, E denotes expected value of the predicted outcome$y$ with original input values$R$, $y$will be the target variable or the observed resultand Rrepresents the predictors or features used for the prediction.

Step 5: The super learner model is constructed by integrating the prediction from each candidate with learner with $m(z;\beta)$ from step 1-4.

Step 6: The candidate algorithm is completely trained with the entire dataset

Step 7: The final output values are determined from the training stage for the prediction tasks

3.3.5 Design of ODESLM

In the ODESLM for thyroid disease classification, a diverse set of base classifiers including SVM, XGB, RF, CNN, LSTM and Bi-GRU, is employed. The model leverages the SBOA to optimally select a subset of these classifiers. Figure 2 depicts the complete framework for the proposed ODESLM model.

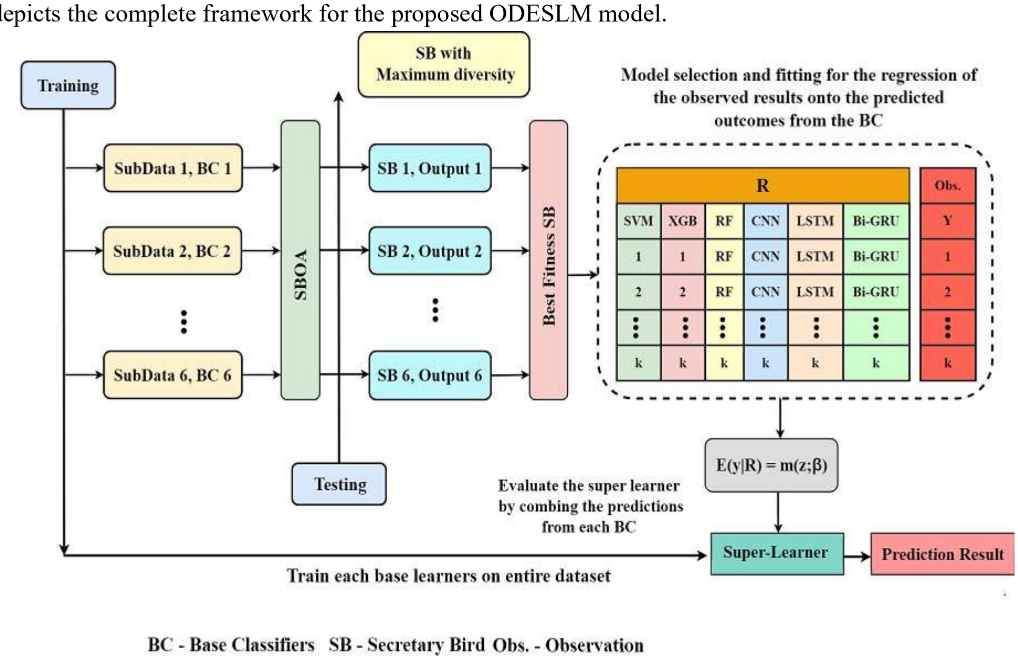

Figure 2 Working Module of the Proposed ODESLM Model

In this process, the presence or absence of each base classifier in the ensemble is encoded as asecretary bird within the SBOA. This algorithm iteratively evaluates various combinations of classifiers referred to as secretary birds, by assessing their fitness values based on performance criteria such as accuracy and diversity. After each iteration, the fitness and diversity metrics are calculated for each secretary bird, and the one exhibiting the highest diversity score, along with a strong fitness value, is selected for inclusion in the ensemble.Once the optimal set of base classifiers is selected, they form the backbone of the SLM. The predictions from these classifiers are passed to the SLM, which learns the optimal weights to combine them, minimizing prediction errors. This combination of diverse classifiers enhances accuracy and generalization by leveraging the strengths of each model.

During the optimization process, SBOA generates a population of secretary birds, where each bird represents a unique combination of base classifiers. The fitness of each bird is evaluated based on its classification performance, and its diversity is measured by examining the differences in the learning characteristics of the classifiers it comprises. The top-performing secretary bird demonstrates the highest diversity and fitness which is selected for the final model. This selected subset of base classifiers is applied to the test dataset, where the SLM combines their predictions to predict the thyroid disease classification labels.This iterative selection process coupled with SBOA's ability to explore a diverse range of classifier combinations, ensures that the ODESLM explores diverse classifier combinations. This approach improves classification performance, boosts model robustness and reduces predictive errors by choosing the most complementary classifiers.The model’s final evaluation is conducted using k-fold cross-validation (k = 5), ensuring that the chosen ensemble performs consistently across multiple folds of the data, reducing the risk of overfitting and improving its generalization capabilities.

Algorithm: ODESLM for Thyroid Disease Detection

Input: Collected Thyroid disease dataset with features and labels

Output: Accurate classification of thyroid disease

1. Start
2. Load the input data
3. \\ **Data Pre-processing**
4. Apply Label Encoding to convert categorical features into numeric vectors
5. Eliminate features with more than 50% missing values, including T3, TSH, Age, T4U, TT4, TBG, FTI, and Sex.
6. Apply Mean Interpolation to impute missing values in the remaining features.
7. **Data Augmentation**
8. Implement HOGAN for data augmentation to address class imbalance
9. Apply VAE to learn latent representations from majority class data
10. Utilize OCTGAN to generate synthetic data for minority classes
11. SBOA optimizes VAE and OCTGAN to prevent mode collapse and enhance class-conditioning in the latent space.

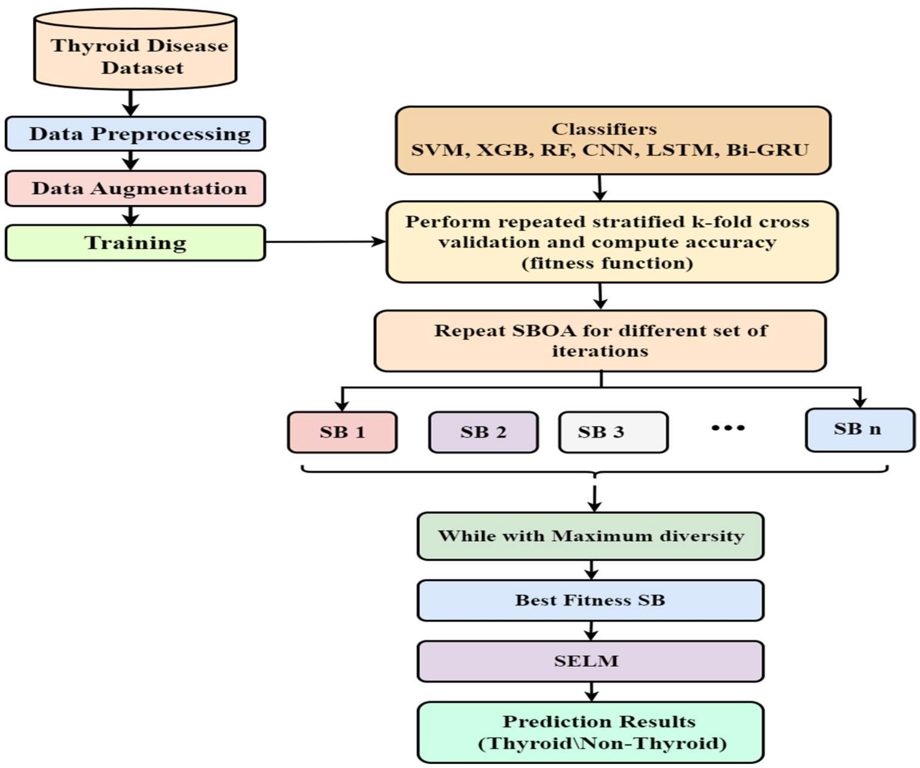

Figure 3 Flow work of the proposed model

12. **Base Classifier Training**
13. Train the following base classifiers (SVM, XGB, RF, CNN, LSTM and Bi-GRU) on the balanced dataset from HOGAN
14. \\ **Constructing DESLM**
15. Combine the predictions of base models using SLM
16. Rearrange training datasets through resampling or reweighting.
17. Build a stacked ensemble model by integrating predictions from base models
18. Weight and combine predictions to minimize overall prediction error
19. **Optimization with DESLM Output with SBOA**
20. Use SBOA for model optimization
21. Fine-tune hyperparameters and adjust weights of base models to balance bias and variance
22. Optimize feature selection by identifying relevant features
23. Ensure the final SLM generalizes well to unseen data and reduces overfitting.

24. \\ **Prediction and Evaluation**
25. Devise the optimized ensemble model to classify thyroid disease in test data.
26. Evaluate the model's performance based on accuracy, generalization, and error minimization.
27. End

Hence, the proposed model effectively enhances generalization and accuracy for thyroid disease prediction in large datasets by leveraging proposed ensemble learning. Also, it combines diverse classifiers to minimize prediction errors, reduce overfitting, and improve reliability.Figure 3 depicts the overall flow work of the proposed ODESLM model for thyroid diseases prediction.

4. RESULTS AND DISCUSSION

This section outlines the experiments conducted for thyroid disease prediction and classification using both machine learning (ML) and deep learning (DL) algorithms. The experiments were run on a laptop featuring an Intel® Core™ i5-4210 CPU @ 3GHz, 4GB RAM, and 1TB HDD, operating on Windows 10 64-bit with MATLAB 2019b. The dataset used is described in Section 3.1, and the analysis was performed on a balanced dataset generated by the HOGAN model as referenced in [13]. This model was used to augment the thyroid dataset by generating synthetic instances, which helped minimize bias and improve experimental reliability. Table 2 presents the sample counts before and after augmentation. The augmented dataset was split into 80% for training and 20% for testing, with the training set used to train ML/DL classifiers, and the test set for evaluating and comparing classification results.

Table 2. Statistics of Balanced Thyroid Disease Dataset

Classes	# of samples before augmentation	# of samples after augmentation	# of training samples	of test samples
Healthy	6771	6771	5417	1354
Hyperthyroid	182	6682	5346	1336
Hypothyroid	601	6601	5281	1320

4.1 Parameter Settings

Table 3provides a list of parameters used to simulate both the existing and proposed models for performance evaluation.

Table 3 Hyperparameter Settings for Various Models

Models	Hyperparameters	Range
SVM	Kernel	Linear
	Regularization variable (C)	5.0
XGB	Minimum child weight	1
	Maximum depth	8
	Learning rate	0.1
	# of predictors	120
	Minimum split loss	0
	Subsample	0.8
	L_1 regularization	0.1
	L_2 regularization	10
RF	# of predictors	250
	Maximum depth	20
CNN	Filters in convolution layer	32
LSTM	# of LSTM units	100
	# of LSTM layers	3
BiGRU	Embedding dimension	18
	# of GRU layers	3
	# of hidden units	256
ODESLM	Learning rate	0.01
	Optimizer	Adam
	Loss function	Cross-entropy loss

	Batch size	96
	Epochs	50
	Dropout rate	0.5
	Momentum	0.7

4.2 Performance Evaluation Measures

The model's ability to detect thyroid disease is evaluated using the following performance metrics:

- Accuracy is the percentage of correctly detected samples out of the total samples tested.

$$Accuracy = \frac{True\ Positive\ (TP)+True\ Negative\ (TN)}{TP+TN+Fals\quad Positive\ (FP)+Fa\quad Negative\ (FN)} \qquad (2)$$

In Eq. (31), TP defines the situation where the model accurately categorized that a subject has thyroid when they have thyroid. TN represents the situation where the model accurately categorized that a subject does not have thyroid when they truly do not have thyroid. FP denotes the situations where the model inaccurately categorized that a subject has thyroid when they do not have thyroid. FN indicates the situation where the model inaccurately categorized that a subject does not have thyroid when they do have thyroid.

- Precision: It is computed as:

$$Precision = \frac{TP}{TP+F} \qquad (3)$$

- Recall: It is measured by

$$Recall = \frac{TP}{TP+FN} \qquad (4)$$

- F1-score: It is determined by

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recal} \qquad (5)$$

4.3 Experimental Results with Single Classifier with Proposed Ensemble Classifier

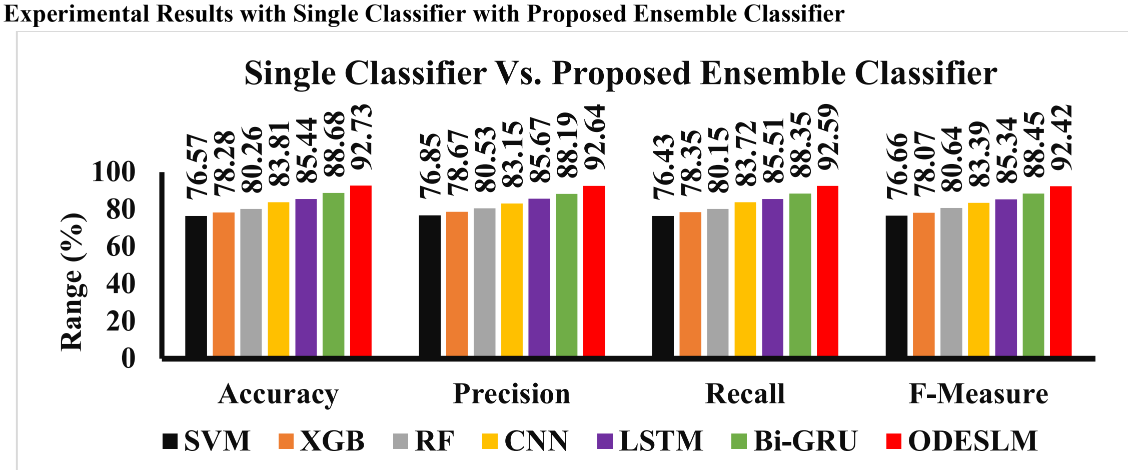

Figure 4 Performance Analysis of Considered Single Classifier vs. Proposed Ensemble Classifier

Figure4 illustrates a comparative analysis of the considered single classifier (SVM, XGB, RF, CNN, LSTM and Bi-GRU) with the proposed ensemble method (ODESLM). The performance results indicate that the proposed ODESLM model consistently outperforms individual classifiers. In terms of accuracy, ODESLM shows an improvement of 21.10%, 18.46%, 15.54%, 10.64%, 8.53% and 4.57% over SVM, XGB, RF, CNN, LSTM, and Bi-GRU, respectively. The precision of ODESLM also exceeds these classifiers by 20.55%, 17.76%, 15.04%, 11.41%, 8.14% and 5.05%. Additionally, ODESLM achieves higher recall, surpassing SVM, XGB, RF, CNN, LSTM, and Bi-GRU by 21.14%, 18.17%, 15.52%, 10.59%, 8.28% and 4.79%. Finally, F-Measure of ODESLM is greater by 20.56%, 18.38%, 14.61%, 10.83%, 8.29% and 4.49% compared to SVM, XGB, RF, CNN, LSTM and Bi-GRU.

This is due to the ODESLM which enhances performance by integrating the strengths of individual classifiers while mitigating their weaknesses. Through a super-learning approach, it captures diverse patterns in the dataset, leading to better generalization and accuracy. This ensemble method balances the bias-variance tradeoff, improving key metrics like precision, recall, and F-measure.

4.4 Experimental Results with Other Thyroid Classification Model and Proposed Classification Model

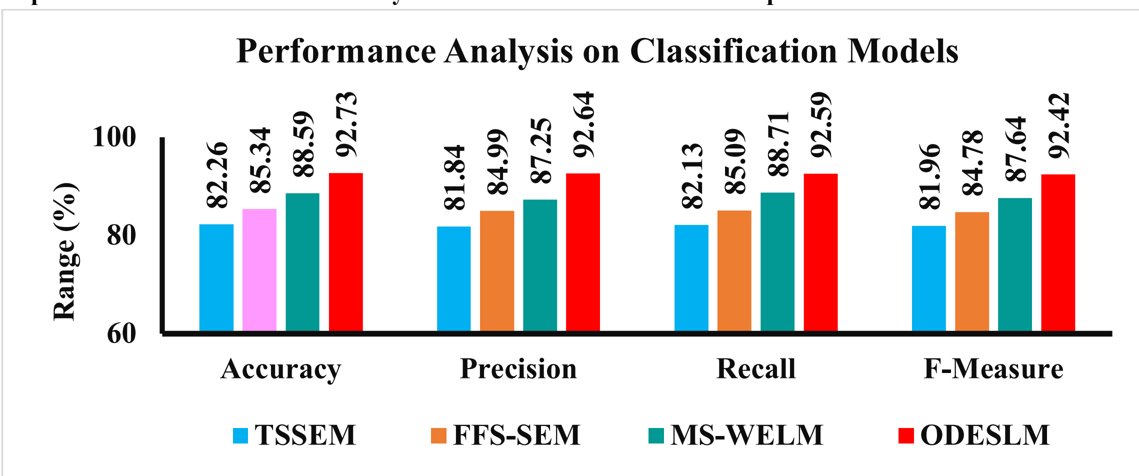

Figure 5 Performance Analysis of Existing and Proposed Ensemble Classification Models

The figure 5 presents a comparative analysis between the existing thyroid classification models, including TSSEM [18], FFS-SEM [19], MS-WELM [22] and the proposed ODESLM model. Both the proposed and existing models were evaluated on the same thyroid dataset. The analysis reveals that ODESLM consistently outperforms these models. In terms of accuracy, ODESLM demonstrates an improvement of 12.73%, 8.66% and 4.67% overTSSEM, FFS-SEM and MS-WELMrespectively. Regarding precision, ODESLM surpasses the competing models by 13.19%, 9% and 6.18%. Additionally, ODESLM achieves higher recall, with increases of 12.74%, 8.81% and 4.37% compared to TSSEM, FFS-SEM and MS-WELM. Finally, ODESLM's F-Measure shows improvements of 12.76%, 9.01% and 5.45% over these models, confirming its superior performance in thyroid classification tasks.

This is because ODESLM improves generalization by capturing diverse data patterns through its ensemble approach. It balances the bias-variance tradeoff, reducing overfitting and enhancing models performance. This makes it particularly effective for handling the complexity and noise in medical datasets like thyroid disease where accurate and timely diagnosis is essential.

4.5 Error Rate Analysis

The error rate is defined as the proportion of incorrect predictions made by the model out of the total number of predictions. It is calculated by

$$Error\ rate = 1 - Accuracy \tag{6}$$

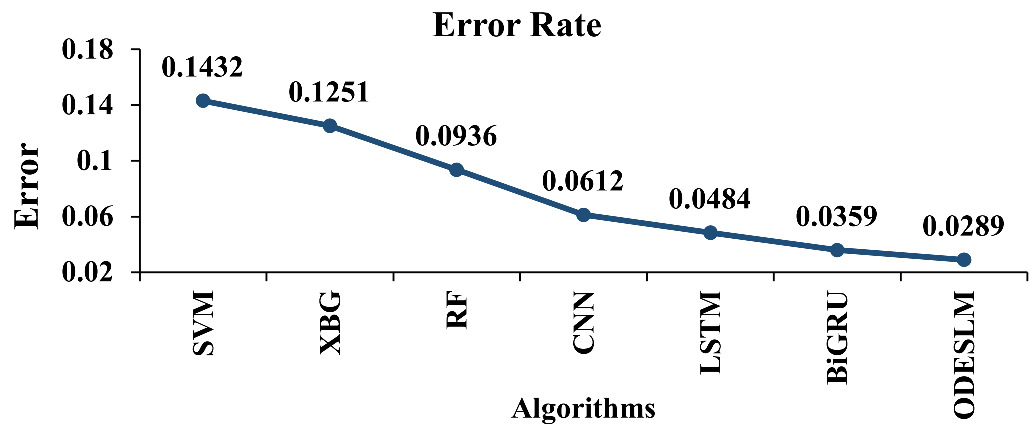

Figure 6 Error Rates ofConsidered Single Classifier with Proposed Ensemble Classifier on Collected Thyroid Disease Datasets

Figure 6 displays the error rates of single classifier with proposed ensemble classifier on collected thyroid disease datasets. The comparison between single classifiers and the proposed ODESLM model reveals significant differences in error rates for thyroid disease prediction. The ODESLM model shows a marked reduction in error rates compared to individual classifiers such as SVM, XGB, RF, CNN, LSTM, and BiGRU, with improvements ranging from 79.82%, 76.89%, 69.12%, 52.78%, 40.29% and 19.49%. This is because the ability of ODESLM modelODESLM model significantly reduces error rates and enhances accuracy by integrating and optimizing multiple classifiers, addressing the complexities and limitations faced by single models.

The figure 7 compares error rates between single classifiers and the proposed ODESLM model on thyroid disease datasets. The ODESLM model demonstrates a substantial reduction in error rates compared to individual classifiers such as TSSEM, FFS-SEM, and MS-WELM, with improvements of 52.07%, 30.36%, and 19.27%, respectively. As ODESLM leverages advanced feature extraction and optimization techniques using SBOA which capture complex data patterns better than the more limited feature handling of single classifiers.

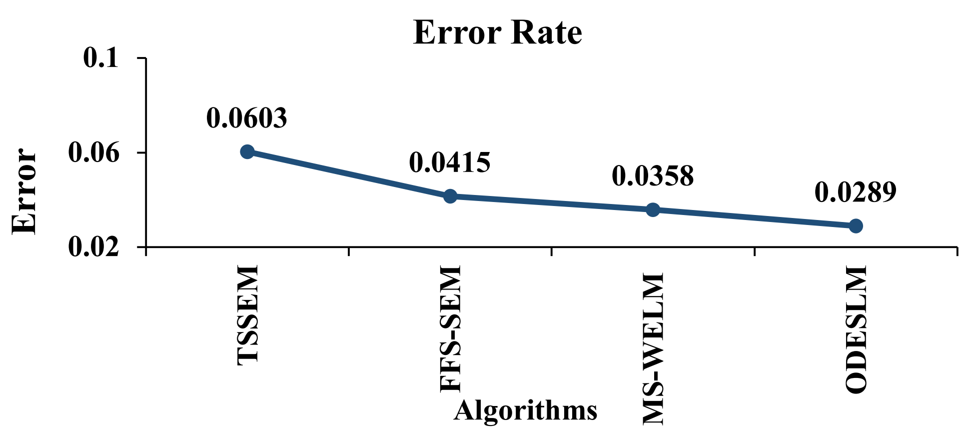

Figure 7 Error Rates of Existing and Proposed Ensemble Classification Models on Collected Thyroid Disease Datasets

4.6 Comparison of Time Complexity with Existing Studies

The computational time complexity is evaluated based on the training and prediction times. Training time is defined as the amount of time needed to train ML and DL algorithms on a dataset. Prediction time is defined as the amount of time taken by trained algorithms to make predictions on new, unseen data.

Table 4 Comparison of Single Classifier vs. Proposed Ensemble Classifier for Time Analysis

Models	Training time (s)	Prediction time (s)
SVM	254	0.1401
XGB	221	0.1275
RF	203	0.1123

CNN	179	0.1011
LSTM	156	0.090
Bi-GRU	121	0.083
ODESLM	**94**	**0.071**

The table 4 compares time complexity between single classifiers and the proposed ensemble classifier for thyroid disease prediction. The proposed model shows a training time of 94 seconds which is 62.99%, 57.46%, 53.69%, 47.49%, 39.74% and 22.31% lesser than SVM, XGB, RF, CNN, LSTM and Bi-GRU. While, the prediction time ofODESLM is 0.071 seconds, which is 49.32%, 44.31%, 36.78%, 29.77% and 21.11% lesser than SVM, XGB, RF, CNN, LSTM and Bi-GRUwhich is significantly lower than the training and prediction times of the single models.These reduced times highlight the proposed model's efficient use of computational resources, achieving a balance between performance and efficiency with significantly lower training and prediction durations compared to other models.

Table 5 Comparison of Existing vs. Proposed Ensemble Classifier for Time Analysis

Ref. No.	Models	Training time (s)	Prediction time (s)
[18]	TSSEM	159	0.111
[19]	FFS-SEM	134	0.103
[22]	MS-WELM	107	0.092
Proposed Model	**ODESLM**	**94**	**0.071**

The table 5 analyzes theprediction time between existing and the proposed ensemble classifier for thyroid disease prediction. The proposed model shows a training time of 94 seconds which is 40.88%, 29.85% and 12.15% lesser than TSSEM, FFS-SEM and MS-WELM. While, the prediction time ofODESLM is 0.071 seconds, which is 36.04%, 31.07% and 22.83% lower than TSSEM, FFS-SEM and MS-WELM. This analysis demonstrates that the proposed model's fast prediction capability is crucial for real-time applications, highlighting its efficiency in delivering quick results for thyroid disease prediction.

5. CONCLUSION

In this paper, ODESLM is designed for thyroid disease prediction. It employs six base learners like SVM, XGBoost, Random Forest, CNN, LSTM, and Bi-GRU which is integrated through a SLM which optimally combines and weights their predictions to minimize error. The ensemble approach uses resampling and reweighting techniques to capture diverse data patterns, enhancing prediction accuracy. The model is further optimized using the Secretary Bird Optimization Algorithm (SBOA), inspired by the foraging behavior of secretary birds, which refines training datasets, optimizes hyperparameters, and identifies relevant features. This balance of bias and variance improves generalization and reduces overfitting, particularly for larger datasets in thyroid disease prediction. The proposed model achieves an accuracy of 92.74%, an error rate of 0.289, training time of 94s and prediction time of 0.071s.

REFERENCES

1. Das, D., Banerjee, A., Jena, A. B., Duttaroy, A. K., & Pathak, S. (2022). Essentiality, Relevance, and Efficacy of Adjuvant/Combinational Therapy in the Management of Thyroid Dysfunctions. Biomedicine & Pharmacotherapy, 146, 112613
2. Ignjatović, M. (2010). The thyroid gland in works of famous old anatomists and great artists. Langenbeck's archives of surgery, 395, 973-985.
3. Hedman, C., Djärv, T., Strang, P., & Lundgren, C. I. (2017). Effect of thyroid-related symptoms on long-term quality of life in patients with differentiated thyroid carcinoma: a population-based study in Sweden. Thyroid, 27(8), 1034-1042.
4. Bode, H., Ivens, B., Bschor, T., Schwarzer, G., Henssler, J., & Baethge, C. (2022). Hyperthyroidism and clinical depression: a systematic review and meta-analysis. *Translational Psychiatry*, *12*(1), 362.
5. Hegedüs, L., Bianco, A. C., Jonklaas, J., Pearce, S. H., Weetman, A. P., & Perros, P. (2022). Primary hypothyroidism and quality of life. *Nature Reviews Endocrinology*, *18*(4), 230-242.
6. Nafisa, A., Ikram, N., Khursheed, S., Anjum, R., & Akhtar, N. (2021). Epidemiologic profile of thyroid disorders in a Tertiary Care Hospital, a Five Years Analysis. *Journal of Rawalpindi Medical College*, *25*(4).
7. Islam, S. S., Haque, M. S., Miah, M. S. U., Sarwar, T. B., & Nugraha, R. (2022). Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study. *PeerJ Computer Science*, *8*, e898.
8. Abbad Ur Rehman, H., Lin, C. Y., & Mushtaq, Z. (2021). Effective K-nearest neighbor algorithms performance analysis of thyroid disease. *Journal of the Chinese Institute of Engineers*, *44*(1), 77-87.

9. Aversano, L., Bernardi, M. L., Cimitile, M., Maiellaro, A., & Pecori, R. (2023). A systematic review on artificial intelligence techniques for detecting thyroid diseases. *PeerJ Computer Science*, *9*, e1394.
10. Jha, R., Bhattacharjee, V., & Mustafi, A. (2022). Increasing the prediction accuracy for thyroid disease: a step towards better health for society. *Wireless Personal Communications*, *122*(2), 1921-1938.
11. Elreedy, D., Atiya, A. F., & Kamalov, F. (2023). A Theoretical Distribution Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Learning. Machine Learning, 1-21.
12. Gupta, P., Rustam, F., Kanwal, K., Aljedaani, W., Alfarhood, S., Safran, M., & Ashraf, I. (2024). Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution. International Journal of Computational Intelligence Systems, 17(1), 3.
13. Uma (2024) Addressing Imbalanced Thyroid Disease Data Classification Using Hybrid Optimized Generative Adversarial Network
14. Sultana, A., & Islam, R. (2023). Machine Learning Framework with Feature Selection Approaches for Thyroid Disease Classification and Associated Risk Factors Identification. Journal of Electrical Systems and Information Technology, 10(1), 32.
15. Hossain, M. B., Shama, A., Adhikary, A., Raha, A. D., Uddin, K. A., Hossain, M. A., ... & 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.
16. Rathinakumar, N.,& Priya, R., (2023). Predicting Thyroid Disease using Optimized Multi-Layered Long Short-Term Memory (LSTM) Neural Network with Residual Attention Mechanism. International Journal of Engineering Trends and Technology. 71. 75-84.
17. Mohan, E., Saravanan, P., Natarajan, B., Kumer, S. A., Sambasivam, G., Kanna, G. P., & Tyagi, V. B. (2023). Thyroid detection and classification using DNN based on hybrid meta-heuristic and LSTM technique. *IEEE Access*, *11*, 68127-68138.
18. Ji, S. (2024). SSC: The novel self-stack ensemble model for thyroid disease prediction. Plos one, 19(1), e0295501
19. Obaido, G., Achilonu, O., Ogbuokiri, B., Amadi, C. S., Habeebullahi, L., Ohalloran, T., ... & Aruleba, K. (2024). An Improved Framework for Detecting Thyroid Disease Using Filter-Based Feature Selection and Stacking Ensemble. IEEE Access.
20. Sha, M. (2024). Quantum Intelligence in Medicine: Empowering Thyroid Disease Prediction through Advanced Machine Learning. IET Quantum Communication, 5, 123-139.
21. Balasree, K., & Dharmarajan, K. (2024, March). Thyroid Detection and Recognition Based on Multi-Layer Recursive Neural Network (ML-RNN) Using in Deep Learning. In 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 190-195). IEEE.
22. Prasad, M. L., & Santhosh, R. (2024). Design of a Multi-Stage Ensemble Model for Thyroid Prediction Using Learning Approaches. Intelligent Automation & Soft Computing, 39(1).
23. https://archive.ics.uci.edu/dataset/102/thyroid+disease.
24. Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical applications in genetics and molecular biology*, *6*(1).

Authors

Ms.C.Uma, is Research Scholar, Navarasam Arts and Science College for women, Arachalur. She is working as Assistant Professor in Department of Computer Technology and Information Technology, Kongu Arts and Science College, Erode. My area of research is Data Mining and my teaching experience 15 years. I published review papers in various National and International Journals and presented 2 papers in National Conferences. I has attending more than 20 workshop.

Dr. P. Rathiga, is working as Assistant Professor, PG and Research Department of Computer Science, Erode Arts and Science College, Erode – 9. She did Ph.D. from Bharathiar University, Coimbatore. Her area of research is Network Security and Wireless Network. She has a teaching experience of 19 years. She has guided 21 M.Phil. Scholars and guiding 4 Ph. D, Scholars. She has published 16 research papers in National and International Journals & presented 18 papers in National and International seminars and Conferences. She has attended more than 25 workshop and Faculty Development programme. She has delivered expert lectures on Cryptography and Network Security. She has Organized 2 National Level Seminar and one International

Seminar.