



Innovative Framework for Thyroid Disease Detection by Leveraging Hybrid AGTEO Feature Selection and GRU Classification Model

K. Hema Priya ^{a,*}, K. Valarmathi ^b

^a Department of Computer Science and Engineering, Easwari Engineering College, Chennai-600089, Tamil Nadu, India

^b Department of Computer Science and Engineering, Panimalar Engineering College, Chennai-600123, Tamil Nadu, India

* Corresponding Author Email: hemapriyaauphd@gmail.com

DOI: <https://doi.org/10.54392/irjmt2439>

Received: 28-01-2024; Revised: 07-03-2024; Accepted: 13-04-2024; Published: 30-04-2024



Abstract: Thyroid disease remains a significant health concern, necessitating advanced diagnostic tools for swift and accurate identification. The initial step involves preprocessing datasets, employing an Outlier Detection Method with Isolated Forest in conjunction with data normalization techniques to eliminate noise and standardize the data, laying a robust groundwork for subsequent analysis. Subsequently, feature extraction is conducted utilizing an Enhanced AlexNet architecture augmented by a more intricate Chameleon Swarm Algorithm (CSA) model to discern finer patterns within the data, enhancing the discriminative nature of the extracted features. Following this, a feature selection strategy employing hybrid optimization is deployed, amalgamating the strengths of Equilibrium Optimizer and Artificial Gorilla Troops Optimizer (AGTO) into a hybrid model named HAGTEO, aiming to identify the most informative features, thus reducing dimensionality and enhancing classifier efficiency. Ultimately, the Gated Recurrent Unit (GRU) classifier is employed for thyroid disease classification based on the extracted and selected features. Renowned for its capability to capture temporal dependencies, the GRU model further enhances classification accuracy. The proposed framework is subjected to testing on two distinct datasets, demonstrating its efficacy in thyroid disease detection. Experimental outcomes reveal superior performance compared to conventional methods, achieving accuracies of 98.07% and 98.00% for dataset 1 and dataset 2, respectively. As an advanced diagnostic solution for thyroid disease, it holds promising potential.

Keywords: Gated Recurrent Unit, Thyroid Disease, AlexNetwork, Artificial Gorilla Troops Optimization, Equilibrium Optimization

1. Introduction

Incidences of thyroid disease have been going up in recent years. Thyroid glands are important organs that controls metabolism. A dysfunctional thyroid gland may have any number of abnormalities. Two commonly occurring types are hyperthyroidism and hypothyroidism [1]. Thyroid disorders are diagnosed in many people, including hyperthyroidism and hypothyroidism, each year. Triiodothyronine (T3) and levothyroxine (T4) are secreted by the thyroid gland; insufficient amounts of these hormones can cause hypo- and hyperthyroidism [2]. In the literature, many methods are put forward for detecting thyroid disease diagnosis. Good treatment for the patient requires that it make thyroid disease predictions proactively, so as to save human lives and prevent unnecessary medical bills. Early thyroid diagnosis prediction is attained by utilising deep learning and machine learning methods in conjunction with the advancement of data processing and computation technologies [3]. These methods can also classify kinds of illness as hypothyroidism or hyperthyroidism etc.

In recent years, the progress of data mining, big data, image-processing technologies, and parallel computing has led to their widespread adoption across various healthcare domains, contributing significantly to improving human health and well-being. Data mining applications in healthcare could include drug discovery, early diagnosis and diagnosis, virus outbreak prediction, and patient-specific drug adaptation testing' own conditions; and management analysis of healthcare information based on statistics calculation. The health care professionals try as best they can to identify illnesses early on so that treatment may be given promptly without great expense, and the disease itself is cured quickly. One of the conditions that gravely affects a significant portion of the human population is thyroid disease [4]. On the other hand, according to the American Thyroid Association (the world's leading professional group), thyroid disease affects 20 million people in the US [5]. At least 12% of Americans will experience thyroid disorders at some point in their lives [6, 7]. According to these figures, thyroid-related

illnesses shouldn't be taken lightly. Efforts to improve health care techniques for the prevention and early thyroid disease detection should be made use of advanced technologies as much possible [8].

In agriculture, like other fields the last few years has seen a boom in deep learning (DL) methods. Computer vision and artificial intelligence have progressed so much that there are new ideas [9, 10]. These approaches are more accurate than conventional methods, providing better decision making. Because of this progress in hardware technology, DL methods are now used to solve problems that otherwise would be too complex for a reasonable amount of time [11, 12]. The findings of research in this field are no trifles. DL is already an up-to-date technique for land cover classification, and it could also be used in many other kinds of tasks. Several kinds of deep neural networks (DNNs) have led to excellent results in hyperspectral analysis [13]. As for classification, CNNs have proven to be effective.

1.1. Motivations

Given that thyroid disease diagnosis must be rapid and accurate, this study sets out to refine existing diagnostic methods. With early detection regarded as important for improving patient outcomes, the research is aimed at providing a reliable and advanced tool to the medical community. The goal is to develop appropriate tools for diagnosing and treating thyroid disease that clinicians can use without themselves becoming victims of its side effects.

1.2. Problem Statement

Existing methods of diagnosis for thyroid disease suffer from problems in terms of accuracy and speed. Misdiagnosis and delayed diagnosis can cause poor health. An accurate and modern diagnostic approach is needed to increase the precision of thyroid diseases. Many of the current techniques experience difficulties with high-dimensional and complex data, so a new approach must be able to effectively extract salient features. Facing these obstacles, this paper needs a comprehensive framework that simultaneously optimizes feature selection and classification. This allows for a more reliable and accurate detection system to help ensure timely medical interventions save lives.

1.3. Main Contributions

Data Preprocessing: By combining outlier detection with data normalization technologies, a solid foundation for analysis is laid down. The model runs on squeaky-clean and standardized data.

Advanced Feature Extraction: Using the AlexNet architecture and an optimized Chameleon Swarm Algorithm (CSA) model enabled us to capture

complicated elements in this data, yielding a robust set of features that allowed the new system to better distinguish between varied thyroid conditions.

Efficient Feature Selection: This hybrid optimization method, combining Artificial Gorilla Troops Optimizer (AGTO) with Equilibrium Optimizer (EO), results in the removal of all redundant features that add no useful information. The process not only reduces dimensionality but also dramatically improves classifier efficiency as a whole.

Classification: Leveraging The capacity of the model to represent temporal dependencies in the data, it classified thyroid disease accurately and precisely by using a specialized recurrent neural network (GRU).

1.4. Organization of the paper

The remaining sections of the study are organised in the form of shadows: A. In Section 2, the relevant works are summarized; B. The model proposed by the author is explained briefly in section 3; C: Details of results analysis as well as validation and an explanation for failure cases are found below section 4 and D: Summary and conclusion appear finally under section 5.

2. Related works

Sharma *et al.* employ three modern deep learning techniques (DeiT, Mixer-MLP, and Swin Transformer) to extract features from thyroid image datasets [14]. The feature extraction methods are built upon the MLP and Image Transformer models. The presence of numerous redundant features can lead to overfitting of classifiers, diminishing their ability to generalize effectively. Six feature transformation techniques are investigated for reducing the dimensionality of the data in order to avoid the overfitting problem: PCA, TSVD, FastICA, ISOMAP, LLE, and UMP. On the transformed dataset, the five classifiers LR, NB, SVC, KNN, and RF are assessed via the 5-fold stratified cross-validation technique. The performance is assessed using stratified cross-validation because there are significant class imbalances in both datasets. Models at various stages of analysis are ranked using the MEREC-TOPSIS MCDM technique. In the wrapper feature selection mode, the optimal strategy for dimensionality reduction is assessed in the second stage, while the optimal strategies for feature extraction and classification are chosen in the first. Applying the recently proposed FOX optimisation algorithm for meta-heuristics, the two best-ranked models are further chosen for the weighted average ensemble learning and features selection. The feature selection + random forest model based on PCA+FOX optimisation obtained the highest TOPSIS score and demonstrated exceptional performance on the ultrasound dataset, with a 99.13% accuracy, 98.82% F2-score, and 99.13% AUC-ROC

score. AUC-ROC of 95.48%, F2-score of 92.01%, and accuracy score of 90.65% were also achieved by the model on the histopathological dataset. This work takes advantage of the novel combination of various algorithms to enhance the diagnosis of thyroid cancer.

For illness diagnosis, Mohan *et al.* proposed the use of the Vgg-19 feature extractor in a novel Long Short-Term Memory-based Convolution Neural Network (LSTM-CNN), proposed by Mohan [15]. They used a novel strategy called Hybrid Black Widow Optimisation with Mayfly Optimisation Approach (HBWO-MOA) to select features. To see how well this method would classify and predict thyroid diseases, their research used ultrasound pictures from the DDTI dataset. According to recalls, F1 score, sensitivity, accuracy, and precision, the recommended Vgg-19-LSTM method beat other convolutional methods such as AlexNet-LSTM, ResNet-LSTM, and Vgg16-LSTM.

One study, by Brindha & Muthukumaravel, aimed to assess the performance of two different classifier models in diagnosing thyroid disease [16]. After training the models using UCI repository data, these method's accuracy and precision at detecting hyperthyroidism and hypothyroidism, respectively, were examined. But their results revealed that the CNN classifier was superior to the SVM classifier, which achieved 89 % accuracy and a precision of 87 %, producing more consistent and reliable results.

Punit Gupta *et al.*, proposed an approach which uses a differential evolution (DE)-based optimization algorithm to fine-tune the parameters of machine learning models [17]. Bias field correction is combined with a hybrid optimisation technique, namely Black Widow Optimisation and Mayfly Optimisation Approach (HBWO-MOA), to select features. The LSTM and Vgg-19 of Deep Learning (DL) are presented for disease classification. Thyroid disease prediction and classification from ultrasound images in the DDTI dataset are efficient. According to this analysis, the suggested technology outperforms the convolutional methodology in terms of accuracy. In comparison to current prediction techniques Comparing the proposed Vgg-19-LSTM approach to other LSPs like AlexNet-LSTM, ResNet-LSTM, and Vgg16-LSTM, it shows better recalls, sensitivity, accuracy, precision, and F1_score.

The work by Alnaggar, *et al.* provided an improved multiclass classification model that employed XGBoost to arrange individuals into groups based on the sort of thyroid condition they had [18]. The main contributions were an enhanced feature selection accuracy of the dataset and a multiclass classification to differentiate flanked by three different types of thyroid disorders. In terms of the evaluated criteria, the extremely selective algorithm XGBoost showed the best classification. When the hyperparameters were optimised, the model best cutting-edge models with a correctness of 99%.

The use of deep learning algorithms for thyroid ultrasound image segmentation, feature extraction, and classification differentiation is covered in detail in this article by Xie, [19]. It also provides an overview of the deep learning algorithms used to process multimodal ultrasound images. Lastly, it highlights the issues with thyroid ultrasonography image diagnosis as it stands today and anticipates new avenues for research and development. This work can help advance the use of deep learning in clinical ultrasound image diagnosis of thyroid disease and serve as a resource for medical professionals diagnosing thyroid disease.

Dhamodaran, *et al.* set out to look at the feasibility of using support vector machines (SVM), (KNN), and Naive Bayes to categorise thyroid datasets into several classes [20]. By comparing several machine learning methods, it found one with the highest illness prediction accuracy. Predictions of future TD cases and estimates of the affected rate levels were shown to be more accurate using the Expert organization for TD Diagnosis (ESTDD) model. The sophisticated model achieved the specified levels of performance in terms of accuracy (98.53%), throughput (98.34%).

2.1. Research Gaps

Sharma *et al.* [14] does research on thyroid disease in which deep learning techniques and feature extraction methods are combined. But the exploration of how different feature transformation methods affect model performance is another gap. Although Mohan *et al.* Vgg-19-LSTM method appears to have potential for thyroid disease diagnosis, the performance of this technique, like all the convolutional techniques, was not tested across various datasets [15]. While Brindha *et al.* concentrate on SVM and CNN for diagnosis of thyroid, they comment on the advantages of CNN over SVM. However, there is a lack of research on evaluating these models on different datasets other than the UCI repository data [16]. However, while Gupta *et al.* Vgg-19-LSTM has the best performance of all the LSPs, there is still a gap in research concerning whether it is generalizable to different datasets, and whether it is robust to changes in the quality of images [17]. Alnaggar *et al.* present a XGBoost-based classifier with good accuracy, but the research shortfall comes in evaluating the model's performance on different datasets, as well as testing the model's scalability [18]. Xie, work on deep learning for ultrasound image diagnosis offers interesting findings [19]. However, there remains a void in terms of addressing challenges unique to thyroid ultrasound image diagnosis and putting forward avenues for future research. Although Dhamodaran *et al.* study of SVM, KNN, and Naive Bayes for thyroid disease classification is thorough, a significant gap in research remains in systematically comparing the performance of these models across different datasets while accounting for the differences in prevalence of disease [20].

3. Proposed Methodology

Figure 1 shows the proposed work flow the thyroid disease detection model.

3.1. Dataset Description

The Thyroid Disease Data Set and the Mikeizbikiv Database are the two datasets used to identify thyroid disorders.

The content shown in table 1 is gathered from the two databases that are open to the public in order to predict thyroid diseases is indicated by ID_x^{th} , where the term x is considered as $x=1,2,3...X$ and the overall quantity of data displayed is indicated by X .

3.2. Preprocessing

3.2.1. Outlier Detection Method Based on Isolated Forest

The isolated forest algorithm was employed in this study to detect outliers. The isolated forest (iForest) algorithm is an integrated learning-based unsupervised anomaly detection technique that doesn't require prior knowledge of the training set's label information [21]. The iForest divides the data space, which contains all of the samples, into two subspaces along a given dimension using a random hyperplane. Each subspace only contains a portion of the original data, and it divides the two subspaces again in the same manner, repeating the process until each subspace has just one datum left.

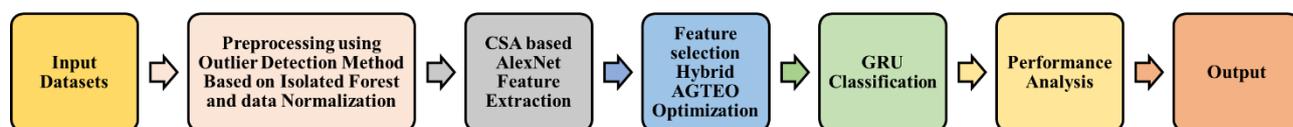


Figure 1. Block Diagram

Table 1. Description of the gathered dataset used to create the thyroid disease identification model

Datasets	Online sources	Dataset description
Dataset 1	" https://www.kaggle.com/datasets/yasserhessein/thyroid-disease-data-set?select=hypothyroid.csv : access date 2022-12-30"	The "Thyroid Disease Data Set" database 1 contains the information needed to predict thyroid diseases. The data is gathered by the Garavan Institute, and Ross Quinlan provides the documentation. There are 2800 total data instances, 972 test instances, and a large number of missing data (29). There is another database that has 9172 instances and 20 classes. The dataset is in the file format hypothyroid.csv and has a size of 276.17 KB.
Dataset 2	" https://github.com/mikeizbicki/datasets/blob/master/csv/uci/ann-train.data : access data of the link 2022-12-30"	The "mikeizbikiv database," also known as database 2, contains the information needed to predict thyroid disorders. It includes a compilation of datasets that are openly accessible. Here, basic random datasets are stored in csv files, restructured datasets are stored in graph files, an image file stores the histogram's computation of RGB signatures, a.gitignore file contains improved signature support for creating histograms, and a README.md file contains the first commit.

Only a few partitions are needed to isolate the abnormal data points because the density of the subspace containing the abnormal data is significantly lower than that of the normal data clusters.

The outlier ratio parameter was set to a low value due to the small sample size. Once the parameters were set, it looked for outliers on the three data sets. It decided to remove the outliers and use the linear interpolation method.

3.2.2. Data Normalization

In order to standardise the data because different thyroid data sets had different dimensions, this study employed maximum and minimum standardisation.

$$x^* = \frac{x - \min}{\max - \min} \tag{1}$$

where max denotes the data's maximum value and min its minimum value.

3.3. Feature Extraction using Improved AlexNet model

3.3.1. AlexNet Model Structure

In this paper, the convolution kernel, fully connected layer, convolution layer 1, convolution layer 2, and convolution layer 5 are used after the max pooling layer; the thyroid data is utilised as the model's input data [22]. 192, 3x3 convolution kernels are used in the third convolutional layer, and another 192, 3x3 convolution kernel is used in the fourth convolutional layer. 192, 3x3 convolution kernels are used in the fifth layer of convolution, 48, 11x11 convolution kernels are used in the first layer, and 128, 5x5 convolution kernels are used in the second layer; Lastly, the final classification is based on three fully connected layers.

3.3.2. Learning Phase

The data are sampled to create the training dataset during the learning phase:

$$(U_{R_x}, Y) = \{(R_x^{(1)}(N), y^{(1)}), (R_x^{(2)}(N), y^{(2)}) \dots, (R_x^{(K)}(N), y^{(K)})\} \tag{2}$$

where U_{R_x} symbolises the collection of $R_x(N)$, Y symbolises the group of labels y , and $(R_x^{(k)}(N), y^{(k)})$ represents the training dataset's k -th data ($k = 1, 2, \dots, K$). Furthermore, one-hot encoding is employed to encrypt the labels on the data sets used for training and validation:

$$y^{(k)} = \{[1, 0]^T, H_1 [0, 1]^T, H_0\} \tag{3}$$

where H_1 represents a PU. Similarly, a 2x1 class score vector represents the output of the last fully connected layer of the AlexNet model:

$$f_{\theta}(R_x^{(K)}(N)) = [f_{\theta|H_1}(R_x^{(K)}(N)) f_{\theta|H_0}(R_x^{(K)}(N))] \tag{4}$$

where $f_{\theta}(\cdot)$ is the expression that has the model parameter for the AlexNet θ and $f_{\theta|H_i}(\cdot)$ is the phrase that goes with H_i . In this case, $f_{\theta|H_i}(R_x^{(K)}(N))$ shows the rating for classification of H_i .

As a result, the two fictitious probability expressions that follow:

$$P(y^{(k)} = 1 | R_x^{(K)}(N); \theta) = f_{\theta|H_1}(R_x^{(K)}(N)) H_1 P(y^{(k)} = 0 | R_x^{(K)}(N); \theta) = f_{\theta|H_0}(R_x^{(K)}(N)) H_0 \tag{5}$$

where $P(y^{(k)} = i | R_x^{(K)}(N); \theta)$ is the likelihood conditional on $P(y^{(k)} = i | R_x^{(K)}(N))$ under θ when $i = 1$ or 0.

Consequently, maximizing the conditional probability is the goal of the AlexNet model's training, specific

$$L(\theta) = P(Y | U_{R_x}; \theta) = \prod_{k=1}^K (f_{\theta|H_1}(R_x^{(K)}(N)))^{y^{(k)}} (f_{\theta|H_0}(R_x^{(K)}(N)))^{1-y^{(k)}} \tag{6}$$

To facilitate computation, it present the logarithmic function:

$$l(\theta) = \log L(\theta) = \sum_{k=1}^K y^{(k)} \log f_{\theta}(R_x^{(K)}(N)) + (1 - y^{(k)}) \log (1 - f_{\theta}(R_x^{(K)}(N))) \tag{7}$$

It want $l(\theta)$ to be as large as possible, that is $-l(\theta)$ to be as tiny as feasible, and the loss function can be obtained:

$$Loss_{JCM-AlexNet}(\theta) = -\frac{1}{K} \sum_{k=1}^K y^{(k)} \log (\hat{f}_{\theta|H_1}(R_x^{(k)}(N))) + (1 - y^{(k)}) \log (\hat{f}_{\theta|H_0}(R_x^{(k)}(N))) \tag{8}$$

The aim of training an AlexNet is to determine which θ would maximise the MAP $P(Y | U_{R_x})$, namely:

$$\theta^* = \arg \max_{\theta} P(Y | R_x^{(K)}(N); \theta) \tag{9}$$

where θ^* symbolises the optimal θ under the MAP.

Using Equation (8) as the basis for the loss function, it employed the CSA optimisation algorithm to gradually adjust the model's parameter θ . This allowed the training process to converge, resulting in the optimal parameter θ being obtained by the model, and ultimately, the trained model, which can be represented as:

$$f_{\theta^*}(R_x) = [f_{\theta^*|H_1}(R_x) f_{\theta^*|H_0}(R_x)] \tag{10}$$

where $f_{\theta^*}(R_x)$ expresses how the AlexNet model was trained following input R_x and $f_{\theta^*|H_i}(R_x)$ shows the rating for classification of H_i .

3.3.3. Test Statistic Design

Two MAPs can be produced in accordance with the training process outcomes from the previous section, as shown by Equations (9) and (10):

$$P(H_1 | R_x) = f_{\theta^*|H_1}(R_x) H_1 P(H_0 | R_x) = f_{\theta^*|H_0}(R_x) H_0 \tag{11}$$

The Bayesian theorem allows us to obtain:

$$P(R_x | H_1) = \frac{P(H_1|R_x) \cdot P(R_x)}{P(H_1)} = \frac{f_{\theta^*|H_1}(R_x) \cdot P(R_x)}{P(H_1)} \tag{12}$$

$$P(R_x | H_0) = \frac{P(H_0|R_x) \cdot P(R_x)}{P(H_0)} = \frac{f_{\theta^*|H_0}(R_x) \cdot P(R_x)}{P(H_0)} \tag{13}$$

where $P(R_x | H_i)$ symbolises the conditional likelihood of a specific H_i , $P(R_x)$ the marginal probability, and $P(H_i)$ represents the training process's prior probability of H_i . When $P(R_x | H_1)$ and $P(R_x | H_0)$ are established, the likelihood ratio (LR) functions superbly as a test statistic, as the NP theorem shows.

In light of the false alarm probability, the Neyman-Pearson (NP) theorem states that $P_f = P(H_1 | H_0)$, to optimise the likelihood of detection $P_d = P(H_1 | H_1)$, there are:

$$L(R_x) = \frac{P(R_x|H_1)}{P(R_x|H_0)} > \lambda \tag{14}$$

where $L(R_x)$ is both the detection threshold and the likelihood ratio.

Equation (14), when Equations (12) and (13) are replaced, yields the AlexNet LRT based on the likelihood ratio test (LRT):

$$L_{AlexNet}(R_x) = \frac{f_{\theta^*|H_1}(R_x)}{f_{\theta^*|H_0}(R_x)} \cdot \frac{P(H_0)}{P(H_1)} = \frac{f_{\theta^*|H_1}(R_x)}{f_{\theta^*|H_0}(R_x)} \tag{15}$$

$$T_{JCM-AlexNet} > \lambda$$

where the false alarm probability constraint can be used to determine the threshold λ . To facilitate analysis, it creates the training datasets H_1 and H_0 possess an equal quantity of samples, meaning that $P(H_1) = P(H_0) = 0.5$; in addition:

$$f_{\theta^*|H_i}(R_x) = e_{H_i} \cdot f_{\theta^*}(R_x) \tag{16}$$

$$e_{H_i} = \{[1,0], i = 1 [0,1], i = 0 \tag{17}$$

3.3.4. CSA based Hyper parameter Tuning

One of the newest metaheuristics is CSA, which Braik introduced in 2021. The hunting and food-finding processes of chameleons, a highly specialised class of animals with the capacity to adjust their colour in order to better fit their environment, are modelled by this algorithm [23]. Chameleons are omnivores, meaning they eat insects and can live in semi-desert regions, lowlands, mountains, and deserts [23]. They follow the prey with their eyes, track them, and then attack them as part of a multi-step food hunting process, as Figure 2 illustrates. The following subsections provide an

explanation of this algorithm's mathematical models and steps.

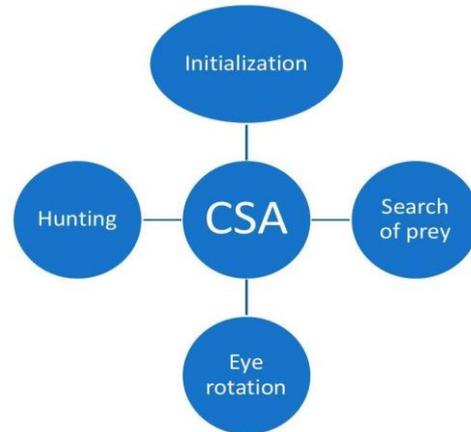


Figure 2. Procedures for CSA

3.3.4.1 Initialization and Function Evaluation

By creating an initial population at random, the population-based metaheuristic known as CSA initiates the optimisation process. A d-dimensional search area is used to generate the n-sized chameleon population, which is made up of all possible solutions to the optimisation problem. Equation (18) describes the chameleon's location in the search area at any given iteration:

$$y_t^i = [y_{t,1}^i, y_{t,2}^i, \dots, y_{t,d}^i] \tag{18}$$

where $i = 1, 2 \dots t$ symbolises the number of iterations, $y_{t,d}^i$ symbolises the chameleon's position. Equation (19) demonstrates the process of creating the initial population according to the size of the issue as well as how many chameleons are in the search region:

$$y^i = l_j + r(u_j - l_j) \tag{19}$$

where y^i is the i th chameleon's initial vector, u_j and l_j refer to the search space's upper and lower bounds, accordingly, and r is a number from zero to one that is uniformly random. Based on the evaluation of the objective function, the quality of each step's solution is evaluated for each new position.

3.3.4.2 Search of Prey

Equation (20) establishes a foundation for describing the chameleons' movement patterns during their search based on how they update their position:

$$y_{t+1}^{i,j} = \{y_t^{i,j} + P_1(P_t^{i,j} - G_t^j)r_2 + P_2(G_t^j - y_t^{i,j})r_1 y_t^{i,j} + \mu(u^j - l^j)r_3 + l_b^j \text{sgn}(\text{rand} - 0.5)r_1 < P_p r_1 \geq P_p \tag{20}$$

wherein t and $(t + 1)$ denotes the t th and $(t + 1)$ th iteration step, accordingly. i and j represent the i th chameleon in the j th dimension. $y_t^{i,j}$ and $y_{t+1}^{i,j}$ are the chameleon's present and new positions, respectively.

$P_t^{i,j}$ and G_t^j represent the chameleon's optimal and worldwide positions, respectively.

Where, P_1 and P_2 two positive values that govern the capacity for exploration. r_1, r_2 , and r_3 are generated using random uniform numbers that range from 0 to 1. r_i is a randomly generated integer, uniformly generated at index i , in the interval $0-1$. P_p shows the likelihood that the chameleon will perceive prey. $sgn(\text{rand} - 0.5)$ influences how exploration and exploitation are carried out, and can be either -1 or 1. μ is a function that grows less as the iteration count increases and depends on the iteration's parameter.

3.3.4.3 Chameleon's Eyes Rotation

Through the use of their eyes, chameleons can determine the location of their prey. They can see the prey 360 degrees thanks to this rotating feature [24]. The subsequent actions take place as follows:

- The chameleon's initial position serves as its centre of gravity (i.e., the beginning);
- The rotation matrix is located, revealing the location of the prey;
- The rotation matrix at the centre of gravity is used to update the chameleon's position;
- At last, the chameleons are put back in the starting position.

3.3.4.4 Hunting Prey

When their prey gets too close, chameleons attack. The ideal chameleon is the one that is closest to the prey and is thought to yield the best results. This chameleon attacks its prey with its tongue. Its situation is made better by the chameleon's ability to spread its tongue to twice its original length. The chameleon is able to successfully snatch prey as a result and makes use of the pursuit space [24]. Equation (21) provides a numerical representation of the chameleon's tongue's speed as it extends towards prey:

$$v_{t+1}^{i,j} = wv_t^{i,j} + c_1(G_t^j - y_t^{i,j}) + c_2(P_t^{i,j} - y_t^{i,j})r_2 \quad (21)$$

where $v_{t+1}^{i,j}$ shows the i th chameleon's new velocity in the j th dimension of the iteration $t + 1$, and $v_t^{i,j}$ shows the i th chameleon's current velocity in the j th dimension.

3.4. Feature Selection using Hybrid AGTEO Optimization

3.4.1. Optimizer for Artificial Gorilla Troops

The AGTO emulates the way of life of a group of gorillas in the wild. The two processes that make up the GTO are the exploration and exploitation processes, which are similar to other metaheuristic optimisation

techniques [25, 26]. The best solution is represented by the silverback gorilla, and the gorillas' and candidates' locations are denoted by the letters X and GX, respectively. The following is a description of the GTO's phases [27–29].

3.4.1.1 Exploration Phase

Three mechanisms underpin the GTO's exploitation phase: gorillas moving into new areas, gorillas moving to familiar locations, and gorillas moving towards one another. The transitions between these motions are adjusted using an adjustable operator (P) in the manner described below:

$$GX(t + 1) = (UB - LB) \times r_1 + LB \text{ If } rand < p \quad (22)$$

$$GX(t + 1) = (r_2 - C) \times X_r(t) + L \times H \text{ If } rand \geq 0.5 \quad (23)$$

$$X(i) - L \times (L \times (X(t) - GX_r(t)) + r_3 \times (X(t) - GX_r(t))) \text{ If } rand < 0.5 \quad (24)$$

Where;

$$C = F \times \left(1 - \frac{t}{T_{Max}}\right) \quad (25)$$

$$F = \cos(2 \times r_4) + 1 \quad (26)$$

$$L = C \times l \quad (27)$$

$$H = Z \times X(t) \quad (28)$$

$$Z = [-C, C] \quad (29)$$

where UB is a control variable's upper limit, while LB is the lower boundary. r_1, r_2, r_3 , and r_4 are arbitrary numbers in $[0 - 1]$. The value of l varies between -1 and 1.

3.4.1.2. Exploitation Phase

The silverback gorilla, the swarm's leader, is followed by the male and female gorillas. However, backback or young male gorillas replace the silverback when it ages or dies begin to struggle for control of the females and the leadership position. Two elements are used to modify the transition during the exploitation phase. The following are the updated gorilla positions if $C \geq W$:

$$GX(t + 1) = L \times M \times (X(t) - X_{silverback}) + X(t) \quad (30)$$

$$M = \left(\left|\frac{1}{N} \sum_{i=1}^N GX_i(t)\right|^g\right)^{\frac{1}{8}} \quad (31)$$

$$g = 2^L \quad (32)$$

If $C < W$, the updated gorilla positions are as follows:

$$GX(i) = X_{silverback} - (X_{silverback} \times Q - X(t) \times Q) \times A \quad (33)$$

$$Q = 2 \times r_5 - 1 \quad (34)$$

$$A = \beta \times E \quad (35)$$

$$E = \{N_1, rand \geq 0.5 N_2, rand < 0.5\} \quad (36)$$

r_5 is an arbitrary figure inside $[0 - 1]$. β is set in advance. E is an arbitrary number.

3.4.2. The Equilibrium Optimizer

An effective optimizer that mimics the control volume's balance is the EO. The search agents are represented by the focus while in the dynamic equilibrium state. The mass-balanced equation is represented by the following equation:

$$V \frac{dc}{dt} = QX_{eq} - QX + G \quad (37)$$

where V, Q , and X are the following, in that order: volume, flow rate, and concentration.

$$X = X_{eq} + (C_0 - C_{eq}) \exp[-\lambda(t - e_0)] + \frac{G}{\lambda V} (1 - \exp[-\lambda(t - e_0)]) \quad (38)$$

where $\lambda = \left(\frac{Q}{V}\right)$. X_0 and e_0 are the starting time and the initial focus. A vector pool (X_{pool}) is built using the EO technique, comprising the average solution and the top four solutions, as shown below:

$$X_{avg} = \frac{X_1 + X_2 + X_3 + X_4}{4} \quad (39)$$

$$X_{pool} = \{X_1, X_2, X_3, X_4, X_{avg}\} \quad (40)$$

The following is the formulation of the EO's main equation:

$$X = X_{pool} + (X - X_{pool}) \cdot F + \frac{G}{\lambda V} (1 - F) \quad (41)$$

Where;

$$F = a_1 \text{sign}(r - 0.5) [e^{-\lambda e} - 1] \quad (42)$$

$$e_0 = \left(1 - \frac{T}{T_{Max}}\right)^{\left(\frac{a_2 T}{T_{Max}}\right)} \quad (43)$$

$$G = G_0 e^{-k(e - e_0)} \quad (44)$$

$$G_0 = GCP(X_{pool} - \lambda X) \quad (45)$$

$$GCP = \{0.5r_1 r_2 \geq GP \ 0 r_2 < GP\} \quad (46)$$

wherein r and λ are two randomly generated vectors. a_1 and a_2 are two values for constants that were chosen to be, respectively, 2 and 1. r_1 and r_2 are arbitrary parameters between 0 and 1. The value of the constant GP was chosen to be 0.5. When comparing the generated solution to the previous solution in the memory-saving step of the EO, if the new solution proves to be superior, it is refreshed.

The tendency of the GTO towards local optima and stagnation are its main drawbacks. In this way, the phases of the conventional GTO are integrated with the EO technique's exploitation and exploration stages to enhance the latter's search capabilities. The AGTEO

that is being presented aims to integrate the EO and GTO. The three exploration operators of the GTO (motion to a known location, motion to an unknown location, and the exploration technique of the EO, which is a particle's memory-saving approach), when combined, and motion to other gorillas), an excellent hybrid AGTEO is suggested. It can also search effectively. Moreover, the suggested hybrid algorithm combines the GTO and EO's respective exploitation techniques. This includes the GTO's concentration updating and the EO's particle motion concerning the best solution, or silverback. Notably, the current optimisation process iteration is represented by the symbol t . Until the stopping criteria are met that is, repeat this process until the number of iterations in the current iteration equals the maximum number.

3.5.GRU Classification

A significant problem with fully connected neural networks has been effectively resolved by Recurrent Neural Network (RNN) development, particularly with regard to LSTM networks. The problem is that there is a lot of data loss over time or in space in fully connected networks, leading to problems with vanishing and exploding gradients [30]. The introduction of gates into the LSTM network effectively addressed these two problems, which are referred to as vanishing and exploding gradients. Information flow in an LSTM network is regulated and controlled by input, output, and forgetting gates. By this breakthrough, the problems caused by data loss are addressed by the capacity of LSTM to identify and acquire knowledge from sequences with extended dependencies. The GRU, which is essentially an improved LSTM model, is another significant development in this regard. Simpler network topology and fewer training parameters are achieved by the GRU, while maintaining the training efficacy of LSTM. The GRU is therefore better suited to handle sequential data, such as LIB charging and discharging profiles. One of the GRU update equations' steps, the update gate, is shown in Figure 3. Z_t and the reset gate r_t are the two key gates at the center of this architecture. These gates are crucial for regulating data flow and locating data temporal dependencies.

Essentially, an update gate is a traditional recurrent network's combination of input and forgetting gates. It chooses which fresh data to include and which information from the prior state to keep. The following equation (47) displays the Z_t formula:

$$Z_t = \sigma(W_z \times [h_{t-1}, X_t] + c_z) \quad (47)$$

where Z_t is a GRU model's update gate at time step t that determines the amount of newly added data and the amount of the previous hidden state to preserve σ . Information flow is regulated by Using the activation function, values between 0 and 1 are squeezed.

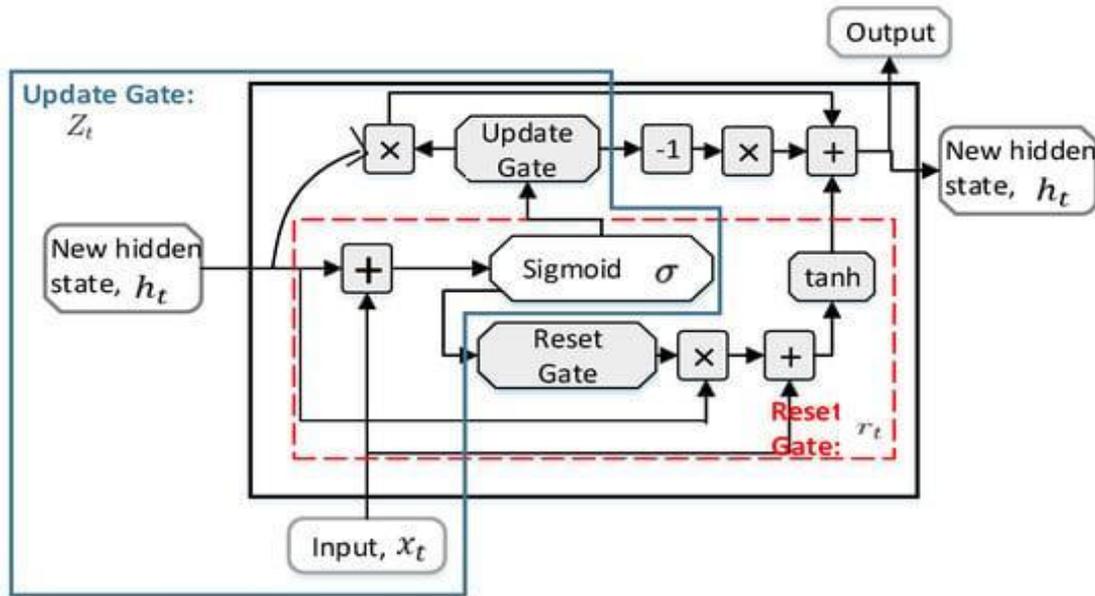


Figure 3. The basic structure of a GRU

The update gate's weight matrix, W_z , determines the relative importance of the previous hidden state h_{t-1} and input X_t . $[h_{t-1}, X_t]$ is the current input concatenated X_t and earlier concealed condition h_{t-1} to create the input of the update gate. c_z is a term for bias that alters the decision boundary of the sigmoid to adjust the gate's behavior in equation (47).

The reset gate is symbolized by r_t , is yet another essential part of the design. The amount of the previous condition that needs to be reset or erased is set when calculating the current state. Through adjusting the memory reset, r_t enables the network to interpret data and identify pertinent patterns and dependencies while keeping or discarding specific details from earlier time steps. The equation for r_t is displayed as follows in equation (48):

$$r_t = \sigma(W_r \times [X_t, h_{t-1}] + c_r) \tag{48}$$

wherein r_t The amount of the previous hidden state that (the reset gate) defines h_{t-1} to reset or neglect when utilizing the current input to compute the fresh concealed state at each time step t, X_t . By squeezing the weighted sum of the input along with the earlier concealed state, the sigmoid activation function σ ensures that r_t accepts values in the range of 0 to 1, regulating the quantity of reset necessary in Equation (48).

4. Results and Discussions

4.1. Experimental Setup

The efficacy of the DL models was evaluated using the WEKA 3.8.6 environment. A data mining programme called WEKA is under the GNU General Public Licence. Data preparation, visualisation, and more are just some of the many features it offers alongside its extensive model library.

The subsequent parameter of a individual system was used to behaviour the assessment:

Processor: Intel(R) Core (TM) i7-97250H CPU @ 2.60 GHz;

Memory: 16 GB;

OS: Windows 10 Home, 64-bit

CPU based on the x64 instruction set

4.2. Performance Metrics

The performance metrics used for comparing the outcomes of the recommended strategy. They include the F-score, Precision, Sensitivity, Accuracy, Specificity, and Negative Predictive Value (NPV). These criteria are used to measure the recommended model's categorization performance. The letters "FN", "FP", "TN", and "TP" in the table stand for "false negative", "false positive", "true negative", and "true positive", respectively.

$$Specificity (SPEC) = \frac{TN}{TN+FP} \tag{49}$$

$$Sensitivity (SEN) = \frac{TP}{TP+FN} \tag{50}$$

$$Accuracy (ACC) = \frac{TP+TN}{TP+TN+FP+FN} \tag{51}$$

$$Precision (PR) = \frac{TP}{TP+FP} \tag{52}$$

$$NPV = \frac{TN}{TN+FN} \tag{53}$$

In the validation results for feature selection in table 2 and figure 4, the HAGTEO-GRU model consistently outperforms other methods in terms of a variety of evaluation metrics on both Dataset 1 and Dataset 2. More importantly, HAGTEO-GRU reaches the highest accuracy percentages 98.0681 % and 98.00424 %, respectively of correct classifications on

Dataset 1 and Dataset 2 respectively, proving its ability to correctly classify cases of thyroid disease. In addition, the model performs well in terms of sensitivity, specificity, precision and F1-Score, two measures of correctly picking positive cases, and two of avoiding false positives, respectively. It is remarkable that HAGTEO-GRU always outperforms alternative models like SMO-GRU, SLO-GRU, IAOA-GRU and HAGTSO-GRU, in terms of all performance metrics. Additionally, the low FPR and FNR values affirm the model's proficiency in minimizing misclassifications. Taken as a whole, these results point up the effectiveness of HAGTEO as a feature selector, making it a promising and reliable choice for thyroid disease detection in clinical practice.

When all the classification results from 2 separate datasets are compiled in a comprehensive analysis as shown in table 3 and figure 5, the proposed model, HAGTEO-GRU, demonstrates a clear performance advantage in thyroid disease detection. On

Dataset 1, HAGTEO-GRU gives an outstanding accuracy of 98.0681 %, beating offerings such as LSTM, 1D-CNN, IAOA-EL, and HAGTSO-HDL. In particular, HAGTEO-GRU proves to be highly sensitive, specific and precise, being able to accurately isolate positive samples while at the same time reducing both false positives and false negatives. Its F1-Score and MCC values show the ability of the model to select the features well and perform feature classification, which further highlight the model's balanced performance. On Dataset 2, HAGTEO-GRU maintains its high standard of excellence with an accuracy of 98.00424%, beating other models in all of the various evaluation metrics. These consistently high precision, recall, and F1-Score demonstrate HAGTEO-GRU's stability in dealing with various datasets. These results all attest to the utility of the proposed model, HAGTEO-GRU, as a promising and reliable instrument for thyroid disease detection. Its efficacy in contributing to higher accuracy in clinical diagnosis is of considerable significance.

Table 2. Effectiveness evaluation of the developed GRU-based thyroid disease prediction model among various Meta heuristic algorithms

Performance analysis on dataset 1					
Validation measures	SMO-GRU	SLO-GRU	IAOA-GRU	HAGTSO-GRU	HAGTEO-GRU
Accuracy	92.4369	93.68152	96.30613	97.1762	98.0681
Sensitivity	92.51644	93.71687	96.28845	97.1369	98.1456
Specificity	92.39714	93.66385	96.31496	97.03076	98.1234
Precision	85.17459	88.08872	92.89004	94.23869	95.1432
FPR	8.502863	6.336162	3.585049	2.869249	1.987622
FNR	8.383563	6.283138	3.611558	2.763204	1.896574
NPV	92.39714	93.66385	97.31496	98.03076	98.86721
FDR	15.72541	11.91128	7.009975	5.661318	4.547643
F1-Score	87.69211	90.81566	94.55872	95.6659	96.1287
MCC	0.81311	0.861028	0.91799	0.934737	0.94342
Performance analysis on dataset 2					
Accuracy	91.46341	92.68293	93.69035	94.48567	98.00424
Sensitivity	91.43924	92.70325	93.67998	94.48433	98.01235
Specificity	91.75258	92.43986	93.81443	94.50171	97.90722
Precision	99.25164	99.32287	99.45105	99.51588	99.8342
FPR	8.247423	7.560137	6.185567	5.498283	3.072783
FNR	8.560758	7.296754	6.320023	5.515657	2.787646
NPV	91.75258	92.43986	93.81443	94.50173	96.90721
FDR	0.748363	0.677131	0.548948	0.484112	0.2657
F1-Score	95.18541	95.89896	96.47929	96.93488	98.35443
MCC	0.622022	0.657389	0.692603	0.720902	0.826832

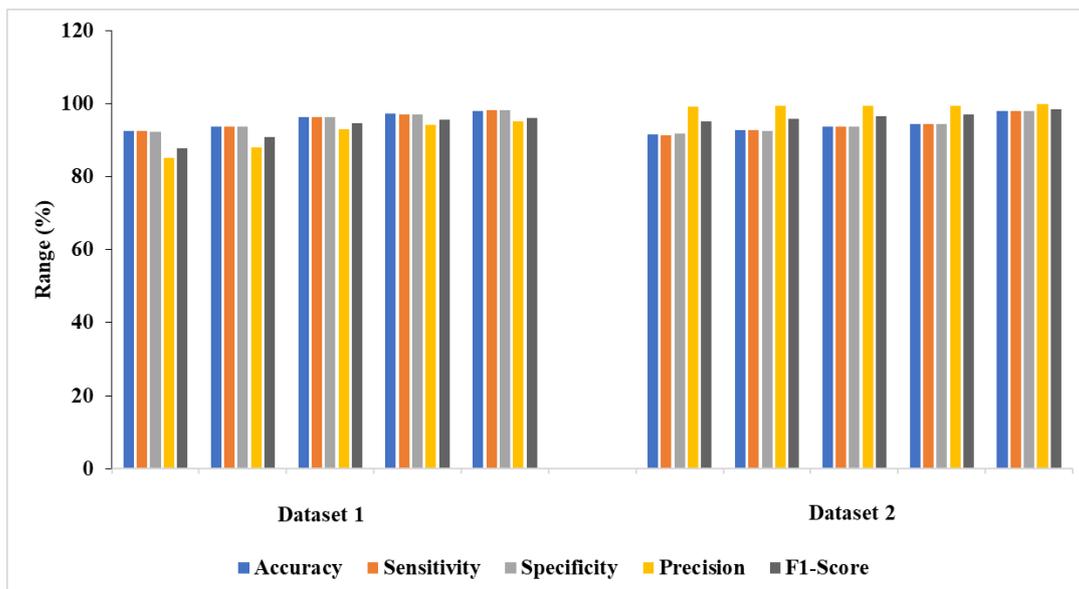


Figure 4. Feature selection evaluation

Table 3. Performance estimation of the developed GRU-based thyroid disease detection technique among distinct conventional thyroid disease detection algorithms

Dataset 1					
Validation measures	LSTM	1D-CNN	IAOA-EL	HAGTSO-HDL	HAGTEO-GRU
Accuracy	91.40156	92.00247	96.30612	97.0661	98.0681
Sensitivity	91.4369	92.1527	96.28844	97.1368	98.1456
Specificity	91.38388	91.92736	96.31495	97.03075	98.1234
Precision	84.14247	85.0918	92.89003	94.23868	95.1432
FPR	8.615119	8.072641	3.685048	2.969247	1.987622
FNR	8.563297	7.847296	3.711559	2.863203	1.896574
NPV	91.38388	91.92736	96.31495	97.03075	98.86721
FDR	15.85753	14.9082	7.109974	5.761317	4.547643
F1-Score	87.63817	88.48161	94.55871	95.6658	96.1287
MCC	0.812286	0.825248	0.91798	0.934736	0.94342
Dataset 2					
Accuracy	91.59597	92.02235	91.9141	97.00424	98.00424
Sensitivity	91.53415	92.04432	91.87015	97.01235	98.01235
Specificity	92.09522	91.75658	92.43986	96.90722	97.90722
Precision	99.28649	99.27551	99.31677	99.7342	99.8342
FPR	7.90368	8.247223	7.560137	3.092784	3.072783
FNR	8.445649	7.957583	8.129848	2.987647	2.787646
NPV	92.09522	91.75458	92.43986	96.90722	96.90721
FDR	0.716311	0.743394	0.68323	0.2658	0.2657
F1-Score	95.26129	95.51249	95.44844	98.35445	98.35443
MCC	0.625424	0.63453	0.636473	0.826832	0.826832

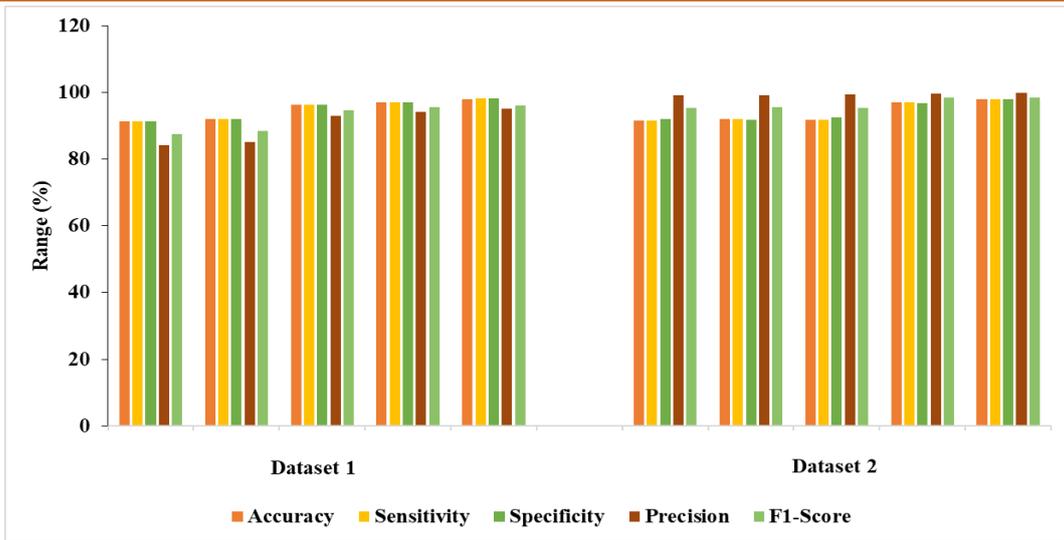


Figure 5. Classification validation

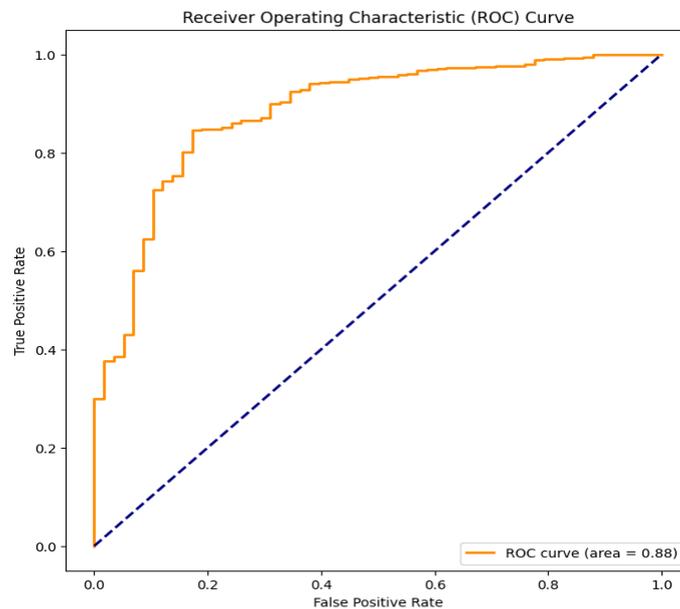


Figure 6. ROC of RNN

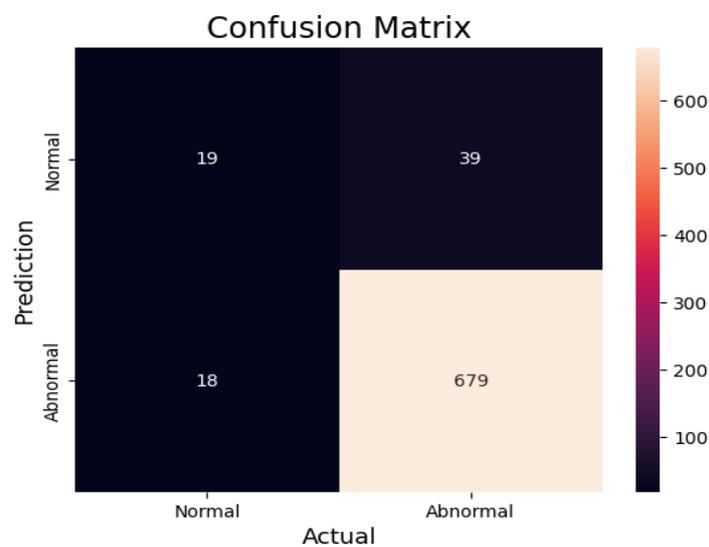


Figure 7. Confusion matrix of RNN

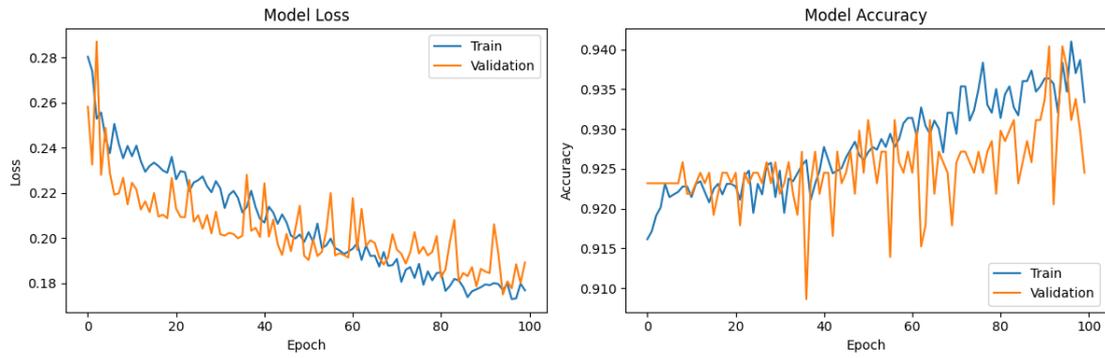


Figure 8. Model loss and model accuracy of RNN

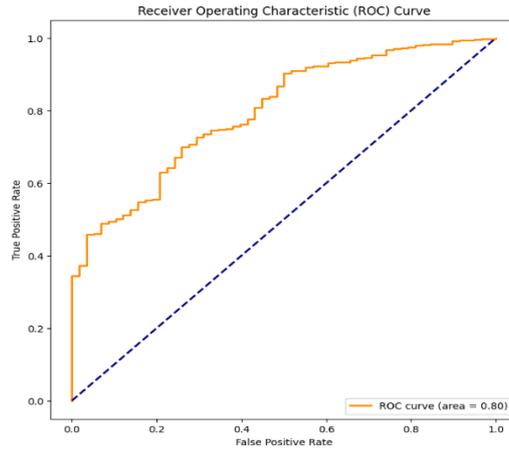


Figure 9. ROC of Proposed GRU model

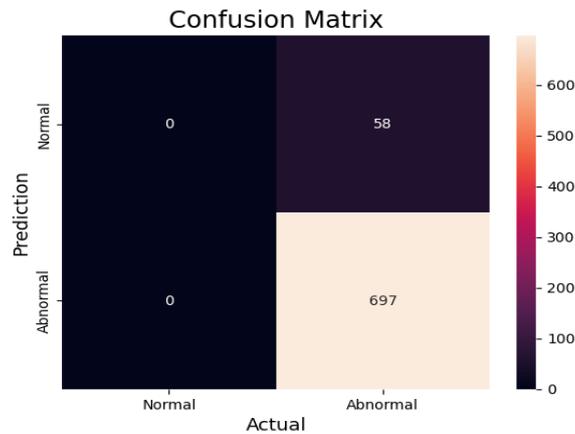


Figure 10. Confusion matrix of proposed GRU model

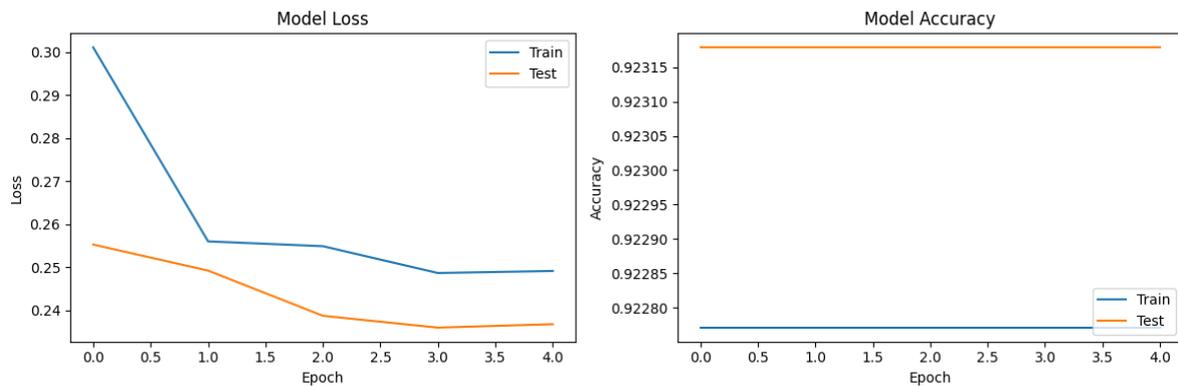


Figure 11. Model loss and model accuracy of proposed GRU model

The Scope of proposed GRU model versus current RNN model in comparative analysis this paper examines the most illuminating performance metrics through ROC curves, confusion matrices, and model loss-accuracy plots. The ability of the RNN to distinguish true positive from false positive rates is depicted in Figure 6, which shows the corresponding ROC curve. At the same time, you can compare the proposed GRU model directly with Figure 9: its ROC curve. More importantly, the proposed GRU model has a better ROC curve, which is not only more sensitive but also more specific than the RNN. Confusion matrices for the RNN and the proposed GRU model are given in Figures 7 and 10, respectively. These matrices give a global perspective of the models' performance across all the classes. The proposed GRU model performs consistently better than the RNN at classifying examples. This is apparent in the higher values along the diagonal, which represent correct predictions. Training results over epochs for the RNN and the proposed GRU model are presented in figures 8 and 11, respectively. The model suggested in the results section converges to lower loss values and higher accuracy faster, thus proving its efficiency in learning and generalizing patterns from the dataset. In sum, these visualizations emphasize the superior performance of the proposed GRU model relative to the standard RNN model. The ROC analysis input, confusion matrices, and training plots, altogether point to the increased discriminative power, accuracy, and efficiency of the proposed GRU model in the face of thyroid disease detection. These findings highlight the promise of GRU for being a high-level and powerful classifier that can help improve diagnostic accuracy in clinical situations.

5. Conclusion

This paper presents a comprehensive model for thyroid disease detection, encompassing state-of-the-art techniques from data preprocessing to feature extraction, selection, and classification. The methodology addresses diagnostic challenges in thyroid disease with a focus on enhancing efficiency and accuracy. Initial preprocessing involves outlier detection using Isolated Forest and normalization to ensure a clean, standardized dataset. Integration of these steps ensures the reliability of subsequent stages. Feature extraction employs the powerful AlexNet architecture augmented with an improved Chameleon Swarm Algorithm (CSA) to identify subtle data structures, enhancing feature discrimination. Deep learning is utilized for its efficacy in handling complex, high-dimensional data. Feature selection employs a HAGTEO optimization approach, combining Artificial Gorilla Troops Optimizer (AGTO) and Equilibrium Optimizer (EO) to reduce dimensionality and enhance classification effectiveness. The Gated Recurrent Unit (GRU) classifier leverages temporal relationships for precise disease classification. Comprehensive testing

on two datasets demonstrates high accuracy (98.0681% and 98.00424% for dataset 1 and dataset 2 respectively), outperforming traditional methods. However, limitations include potential biases in the datasets and generalization ability. Future research directions could involve exploring additional advanced deep learning architectures, incorporating diverse datasets, and conducting real-world clinical trials to further refine and expand the proposed framework.

References

- [1] G. Chaubey, D. Bisen, S. Arjaria, V. Yadav, Thyroid disease prediction using machine learning approaches. *National Academy Science Letters*, 44, (2021) 233–238. <https://doi.org/10.1007/s40009-020-00979-z>
- [2] M. Garcia de Lomana, A.G. Weber, B. Birk, R. Landsiedel, J. Achenbach, K.J. Schleifer, M. Mathea, J. Kirchmair, in silico models to predict the perturbation of molecular initiating events related to thyroid hormone homeostasis. *Chemical Research in Toxicology*, 34(2), (2020) 396–411. <https://doi.org/10.1021/acs.chemrestox.0c00304>
- [3] A. Webster, S. Wyatt, (2020) *Health, Technology and Society*. Palgrave Macmillan Singapore. <https://doi.org/10.1007/978-981-15-4354-8>
- [4] K.H. Priya, K. Valarmathi, (2022) *Big Data Analytics on Thyroid: A Review on Algorithms, Attributes and Future*. 1st International Conference on Computational Science and Technology (ICCST), IEEE, Chennai, India. <https://doi.org/10.1109/ICCST55948.2022.10040437>
- [5] American Thyroid Association, General Information/Press Room. American Thyroid Association. <https://www.thyroid.org/media-main/press-room/>
- [6] D. Chen, J. Hu, M. Zhu, N. Tang, Y. Yang, Y. Feng, Diagnosis of thyroid nodules for ultrasonographic characteristics indicative of malignancy using random forest. *BioData Mining*, 13, (2020) 1-21. <https://doi.org/10.1186/s13040-020-00223-w>
- [7] M.R. Kwon, J.H. Shin, H. Park, H. Cho, S.Y. Hahn, K.W. Park, (Radiomics study of thyroid ultrasound for predicting BRAF mutation in papillary thyroid carcinoma: preliminary results. *American Journal of Neuroradiology*, 41(4), (2020) 700-705. <https://doi.org/10.3174/ajnr.A6505>
- [8] A.J. Idarraga, G. Luong, V. Hsiao, D.F. Schneider, False Negative Rates in Benign Thyroid Nodule Diagnosis: Machine Learning for

- Detecting Malignancy. *Journal of Surgical Research*, 268, (2021) 562–569. <https://doi.org/10.1016/j.jss.2021.06.076>
- [9] X. Chai, Diagnosis method of thyroid disease combining knowledge graph and deep learning. *IEEE Access*, 8, (2020) 149787-149795. <https://doi.org/10.1109/ACCESS.2020.3016676>
- [10] M.A.A.R. Asif, M.M. Nishat, F. Faisal, M.F. Shikder, M.H. Udoy, R.R. Dip, R. Ahsan, Computer aided diagnosis of thyroid disease using machine learning algorithms. *International Conference on Electrical and Computer Engineering (ICECE)*, IEEE, Bangladesh. <https://doi.org/10.1109/ICECE51571.2020.9393054>
- [11] E. Sonuç, (2021). Thyroid disease classification using machine learning algorithms. In *Journal of Physics: Conference Series*, IOP Publishing, 1963(1), 012140. <https://doi.org/10.1088/1742-6596/1963/1/012140>
- [12] Y. Habchi, Y. Himeur, H. Kheddar, A. Boukabou, S. Atalla, A. Chouchane, A. Ouamane & W. Mansoor. AI in Thyroid Cancer Diagnosis: Techniques, Trends, and Future Directions. *Systems*, 11(10), (2023) 519. <https://doi.org/10.3390/systems11100519>
- [13] X. Zhang, V.C. Lee, J. Rong, J.C. Lee, F. Liu, Deep convolutional neural networks in thyroid disease detection: a multi-classification comparison by ultrasonography and computed tomography. *Computer Methods and Programs in Biomedicine*, 220, (2022) 106823. <https://doi.org/10.1016/j.cmpb.2022.106823>
- [14] R. Sharma, G.K. Mahanti, G. Panda, A. Rath, S. Dash, S. Mallik, R. Hu, A framework for detecting thyroid cancer from ultrasound and histopathological images using deep learning, meta-heuristics, and MCDM algorithms. *Journal of Imaging*, 9(9), (2023) 173. <https://doi.org/10.3390/jimaging9090173>
- [15] E. Mohan, P. Saravanan, B. Natarajan, S.A. Kumer, G. Sambasivam, G.P. Kanna, V.B. Tyagi, Thyroid Detection and Classification Using Dnn Based on Hybrid Meta-Heuristic and Lstm Technique. *IEEE Access*, 11, (2023) 68127 – 68138. <https://doi.org/10.1109/ACCESS.2023.3289511>
- [16] V. Brindha, A. Muthukumaravel, (2023) Efficient Method for Predicting Thyroid Disease Classification using Convolutional Neural Network with Support Vector Machine. In *Computational Intelligence for Clinical Diagnosis*, EAI/Springer Innovations in Communication and Computing. Springer, Cham. <https://doi.org/10.1109/ACCESS.2023.3289511>
- [17] P. Gupta, F. Rustam, K. Kanwal, W. Aljedaani, S. Alfarhood, M. Safran & I. Ashraf. Detecting Thyroid Disease Using Optimized Machine Learning Model Based on Differential Evolution. *International Journal of Computational Intelligence Systems*, 17(3), (2024) 1-19. <https://doi.org/10.1007/s44196-023-00388-2>
- [18] M. Alnaggar, M. Handosa, T. Medhat, M.Z. Rashad, Thyroid disease multi-class classification based on optimized gradient boosting model. *Egyptian Journal of Artificial Intelligence*, 2(1), (2023) 1-14. <https://doi.org/10.21608/ejai.2023.205554.1008>
- [19] Y. Xie, W. Yu, S. Song, W. Wang, W. Gao, Y. Jia, S. Wen, C. Wang, Wang, S. (2023) Thyroid disease diagnosis based on feature interpolation and dynamic weighting ensemble model. *Research Square*. <https://doi.org/10.21203/rs.3.rs-2851005/v1>
- [20] S. Dhamodaran, B.B. Shankar, B. Balachander, D. Saravanan, D.S. Kharate, (2023) Estimation of Thyroid by Means of Machine Learning and Feature Selection Methods. In *Artificial Intelligence for Smart Healthcare*, EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-031-23602-0_19
- [21] S. Zhou, C. Song, J. Zhang, W. Chang, W. Hou, L. Yang, A hybrid prediction framework for water quality with integrated W-ARIMA-GRU and LightGBM methods. *Water*, 14(9), (2022) 1322. <https://doi.org/10.3390/w14091322>
- [22] S. Ren, H. Wu, W. Chen, D. Li, Polarization domain spectrum sensing algorithm based on AlexNet. *Sensors*, 22(22), (2022) 8946. <https://doi.org/10.3390/s22228946>
- [23] M.S. Braik, Chameleon Swarm Algorithm: A bio-inspired optimizer for solving engineering design problems. *Expert Systems with Applications*, 174, (2021) 114685. <https://doi.org/10.1016/j.eswa.2021.114685>
- [24] P.J. Vasovala, M.I. Mirchiwala, V. Mayank, V.H. Ghanchi, Application of ant colony optimization technique in economic load dispatch of IEEE-26 bus system with valve point loading. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 9, (2021) 51–58. <https://doi.org/10.22214/ijraset.2021.32713>
- [25] A. Karimipouya, S. Karimi, H. Abdi, Microgrid frequency control using the virtual inertia and ANFIS-based Controller. *International Journal of Industrial Electronics Control and Optimization*, 2(2), (2019) 145-154. <https://doi.org/10.22111/ieco.2018.26293.1071>

- [26] S.K. Akula, H. Salehfar, (2019) Frequency Control in Microgrid Communities Using Neural Networks. 2019 North American Power Symposium (NAPS), IEEE, USA. <https://doi.org/10.1109/NAPS46351.2019.9000219>
- [27] G.I. Sayed, A.E. Hassanien, (2021) A Novel Chaotic Artificial Gorilla Troops Optimizer and Its Application for Fundus Images Segmentation. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics, Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham. https://doi.org/10.1007/978-3-030-89701-7_28
- [28] A. Ginidi, S.M. Ghoneim, A. Elsayed, R. El-Sehiemy, A. Shaheen, A. El-Fergany, Gorilla Troops Optimizer for Electrically Based Single and Double-Diode Models of Solar Photovoltaic Systems. Sustainability, 13, (2021) 9459. <https://doi.org/10.3390/su13169459>
- [29] B. Abdollahzadeh, F.S. Gharehchopogh, S. Mirjalili, Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems. International Journal of Intelligent Systems, 36(10), (2021) 5887–5958. <https://doi.org/10.1002/int.22535>
- [30] S. Jafari, Y.C. Byun, A CNN-GRU Approach to the Accurate Prediction of Batteries' Remaining Useful Life from Charging Profiles. Computers, 12(11), (2023) 219. <https://doi.org/10.3390/computers12110219>

Has this article screened for similarity?

Yes

About the License

© The Author(s) 2024. The text of this article is open access and licensed under a Creative Commons Attribution 4.0 International License.

Acknowledgement

We would like to thank our organization and colleagues who helped out in publishing this research manuscript.

Authors Contribution Statement

K. Hema Priya - Conceptualization, data collection, investigation, analysis, validation, writing original draft; K. Valarmathi - investigation, analysis, validation, writing original draft. Both the authors read and approved the final version of the manuscript.

Funding

The Authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

The Authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Data Availability

Data will be provided based on request if required.