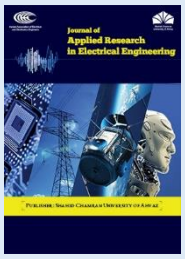
Shahid Chamran University of Ahvaz

Iranian Association of Electrical and Electronics Engineers

Journal of Applied Research in Electrical Engineering

E-ISSN: 2783-2864
P-ISSN: 2717-414X
Homepage: https://jaree.scu.ac.ir/



Research Article

A Hybrid Technique Combining Robust and DNN for Thyroid Disease Prediction on Imbalanced Dataset

Elahe Moradi [1,*]

[1] *Department of Electrical Engineering, YI.C., Islamic Azad University, Tehran, Iran*

* *Corresponding Author: Elahe.Moradi@iau.ac.ir*

Abstract: Thyroid disease poses an important public health risk, reducing the quality of life and increasing healthcare expenditures. In recent years, Artificial Intelligence (AI) algorithms have become essential tools for diagnosing and predicting thyroid disease. Although various AI-based algorithms have been applied to this task, their performance is often hindered by imbalanced data. This research proposes a hybrid approach that combines robust techniques and Deep Neural Networks (DNN) to address this challenge in thyroid disease prediction. The present paper employs the publicly available thyroid disease dataset, obtained from the University of California, Irvine machine learning repository. To manage the data imbalance, Random Under (RU) and Random Over (RO) sampling techniques are applied. A comparative analysis reveals that the hybrid approach, combining robust and DNN with RO sampling, significantly outperforms the RU sampling method on the imbalanced dataset. To assess the effectiveness of the proposed technique, the variant criteria are employed such as accuracy, F1-score, recall, precision, Area Under Curve (AUC), confusion matrix, and Receiver Operating Characteristic (ROC) curve. The hybrid approach achieves superior results, with a testing accuracy of 98.71%, recall of 100%, precision of 97.45%, and an F1-score of 98.71%. These findings demonstrate the robustness and effectiveness of the approach, underscoring its potential for real-time disease prediction.

Keywords: Deep Neural Network (DNN), Robust, Prediction, Imbalanced Dataset, Thyroid Disease, Accuracy.

Article history
Received 21 September 2024; Revised 06 June 2025; Accepted 01 September 2025; Published online November 8, 2025.
© 2025 Published by Shahid Chamran University of Ahvaz & Iranian Association of Electrical and Electronics Engineers (IAEEE)
How to cite this article
E. Moradi, "A hybrid technique combining robust and DNN for thyroid disease prediction on imbalanced dataset," *J. Appl. Res. Electr. Eng.*, vol. 4, no. 1, pp. 1-10, 2025. DOI: 10.22055/jaree.2025.48019.1138

Check for updates

1. INTRODUCTION

The thyroid gland is one of the main glands of the body, which indirectly affects all the organisms of the body, such as the heart and kidneys. Thyroid disorders can slow or speed up metabolism. When hormone levels are too low or too high, the signs of this disorder can be seen in hyperactivity and hypoactivity. The most common thyroid diseases include hypothyroidism, hyperthyroidism, thyroid gland enlargement (goiter), thyroid masses, thyroid cancer, and thyroid inflammation (thyroiditis) [1, 2]. Recently, techniques based on data mining and artificial intelligence algorithms such as machine learning, deep learning, and reinforcement learning have attracted the attention of many researchers in the diagnosis of diseases [3, 4]. These algorithms can discover patterns in medical data and thereby improve the decision-making process. Therefore, they can influence treatment costs and increase the quality of medical care [5-8].

Several data-driven algorithms for the prediction and diagnosis of thyroid disease have been discussed in the literature. In [9], the authors presented five machine learning algorithms including Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), Logistic Regression (LR), and Multilayer Perceptron (MLP) for diagnosis thyroid disease. The results displayed SVM had the best effect of diagnosing this disease with an accuracy rate of 92%. The authors of [10] studied seven AI models including Decision Tree (DT), SVM, KNN, MLP, RF, Gaussian Process, and XGBoost. Among them, RF algorithm performed the best with the accuracy 81.30%. In [11], two different classifiers including Naïve Bayes (NB), and KNN were analyzed for thyroid disease detection and the results exhibited that KNN was the most accurate classifier with 93.44% accuracy. In [12], the authors employed eleven AI algorithms to normalize which one achieved the best accuracy for diagnosing thyroid disease and the artificial neural

network classifier obtained the highest accuracy with a value of 95.87%. In [13], the authors compared two machine learning methods, decision tree and random forest, to enhance the accuracy of thyroid disease detection. A thyroid dataset from the UCI machine learning repository was utilized in the study, with the random forest model achieving an accuracy of 85%.

In [14], the authors classified thyroid disease into four categories: hyperthyroid, hypothyroid, euthyroid patient and euthyroid. They did this by using a multi-class support vector machine. A high accuracy of 83.37% in a high-dimensional feature space was displayed in the results of training the proposed approach with a dataset from the UCI repository. To do a comparative analysis for improved thyroid disease prediction, several machines learning techniques, including decision trees, random forests, KNN, and artificial neural networks, were applied to the dataset in [15]. For a more comparison, classification was done on both the sample and non-sample datasets. Following the dataset's preprocessing, the random forest technique produced the best results, with an accuracy rate of 94.8%. The three machine learning methods for thyroid texture segmentation and classification that the authors compared in [16] were support vector machine, artificial neural network, and random forest classifier. Employing a feature extraction technique, they extracted features to train the classifiers. The technique achieved an accuracy of about 90% across all three approaches. The authors of [17] explored the potential predictive value of thyroid hormones in clinical outcomes using a retrospective analysis of 2,086 inpatients. They employed supervised learning techniques to create a prediction model that included sensitivity and specificity of 91%, 79%, and 0.86 on the ROC-AUC. In [18], the authors evaluated a proposed feature technique using ten machine learning classifiers. The dataset consisted of 2,211 instances from 247 patients treated at a Naples hospital. Due to class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied to oversample the minority class. Among the tested models, the EXTC model demonstrated the best performance with an F-score of 84%.

In [19], the authors developed diagnostic models for the thyroid disease via multilayer perceptron technique. The study found that the genetic algorithm with multilayer perceptron model achieved a classification accuracy of 95%, outperforming the particle swarm optimization with multilayer perceptron model, which reached 85.5%. In [20], the authors investigated a deep learning model using a modified ResNet-based Convolutional Neural Network (CNN) to detect five types of thyroid diseases. The training process was optimized using dual optimizers, Adam and Stochastic Gradient Descent (SGD), to enhance accuracy. The proposed model achieved 97% accuracy after retraining with SGD, compared to 94% with the standard ResNet. In [21], a sequential Convolutional Neural Network (CNN) method was assessed for detecting thyroid abnormalities, leveraging its ability to automate feature extraction and focus on regions-of-Interest. The proposed technique achieved an accuracy of 81.5%, with a precision of 97.4% and a sensitivity of 83.1%, demonstrating its effectiveness in classifying. Enhanced data segmentation employing K-means clustering improved scalability and efficiency with large medical datasets. The hybrid technique displayed an improvement in diagnostic accuracy and precision, with performance gains of approximately 15-20% over traditional methods.

In [22], a deep neural network (DNN) model was developed to predict thyroid disease using demographic data from an imbalanced dataset. The model incorporated various sampling techniques and activation functions, achieving a classification accuracy of 93.90% and outperforming the Random Forest (RF) baseline by 36.63% in lift performance. In [23], thyroid disease prediction was performed using both machine learning and deep learning methods. Among the evaluated models, Random Forest and Recurrent Neural Networks (RNN), combined with SMOTE for data balancing, achieved the highest classification performance with accuracies of 98.16% and 97%, respectively. In [24], the authors employed the synthetic minority oversampling technique (SMOTE-NC) technique to address data imbalance and applied a deep learning model based on Long Short-Term Memory (LSTM) for thyroid disease prediction, achieving an accuracy of 93%. Furthermore, the study reported that the Light Gradient Boosting Machine (LGBM) outperformed the LSTM model, demonstrating higher predictive accuracy in the same experimental setting. In [25], a novel hybrid model named TDIP, integrating Long Short-Term Memory (LSTM) and Support Vector Machine (SVM), was proposed for thyroid disease classification. To address class imbalance, the study applied SMOTE and employed Bidirectional Feature Elimination for optimal feature selection. The combined approach effectively enhanced model performance, achieving a classification accuracy of 97%.

Existing research shows that predictive accuracy requires further improvement when conventional techniques are applied, as prediction accuracy plays a critical role in the application of AI algorithms in medicine. Furthermore, the features of thyroid disease datasets exhibit wide ranges and the presence of outliers. Previous works have commonly utilized normalization and standardization techniques. Therefore, it is essential to explore algorithms that can ensure robustness in the presence of outliers. Additionally, considering the imbalanced nature of the dataset, appropriate balancing techniques are essential. In many previous studies, synthetic minority oversampling technique has been predominantly used.

The main objectives of this research can be summarized as follows:

- Tackling the challenges of imbalanced data: The focus is on addressing the limitations posed by imbalanced datasets, a frequent challenge in disease prediction. In this study, both random over (RO) sampling and random under (RU) sampling are used to generate synthetic samples, and balance the dataset, leading to a more robust predictive algorithm.
- Employing Robust technique in dataset preprocessing: The application of this technique on the thyroid disease dataset, in the presence of outliers, leads to improved performance and enhanced accuracy of the proposed algorithm.
- Developing an improved prediction method for thyroid disease: This research introduces a hybrid technique combining robust and deep neural networks with RO

sampling to improve thyroid disease prediction and diagnosis.

This study is structured as follows: Section 2 provides an overview of the dataset for thyroid disease and explores the application of hybrid deep neural network techniques combined with data balancing methods to enhance performance. Section 3 presents the results and discusses the efficacy of the proposed methods. Finally, Section 4 concludes the research.

2. THYROID DISEASE DATASET AND PROPOSED TECHNIQUES

In this section, the thyroid disease dataset and the techniques applied in the present paper are presented.

2.1. Thyroid Disease Dataset

This study utilizes the thyroid dataset from the University of California, Irvine (UCI) Machine Learning repository [26]. The dataset includes 3772 instances with 29 attributes as input features, supports binary classification. In the thyroid disease dataset, the target variable was categorical, where 'P' denoted the presence of hypothyroidism and 'N' indicated its absence. The attributes included factors such as age, sex, thyroxine medication, query on thyroxine, sick, thyroid surgery, antithyroid medications, query hyperthyroid, treatment, pregnant, query hypothyroid, TSH measured, TSH (actual TSH value), T3, referral source, TT4 measured, T4U, T3 measured, TBG, TT4, FTI measured, TBG measured, T4U measured, FTI, lithium, tumor, goiter, psych, and hypopituitary, all assessed based on their corresponding blood levels. Hypothyroidism served as the target variable [12, 26]. The thyroid dataset consists of 3772 instances, with 92.29% (3481 samples) are classified as 'P', and only 7.71% (291 samples) are classified as 'N', indicating a significant class imbalance in the dataset. Fig. 1 depicts the distribution of the imbalanced dataset for each class in this binary classification. Fig. 2 illustrates the flowchart of the hybrid DNN techniques proposed for thyroid disease prediction. This diagram presents the overall workflow, including data preprocessing, class balancing, model training, and performance evaluation steps designed to improve predictive accuracy on an imbalanced dataset.

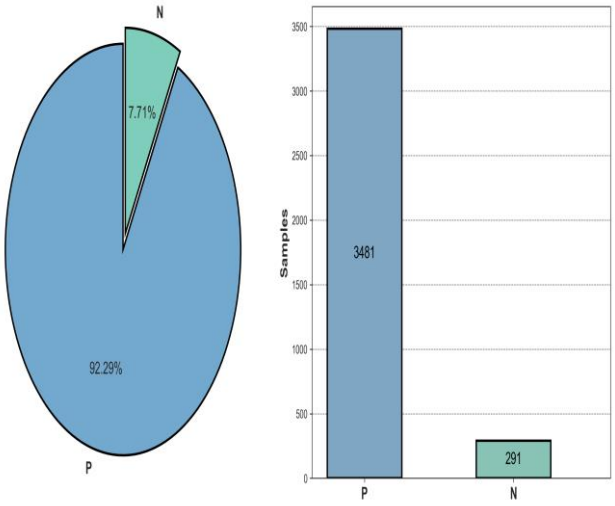


Fig. 1: Distribution of imbalanced thyroid dataset.

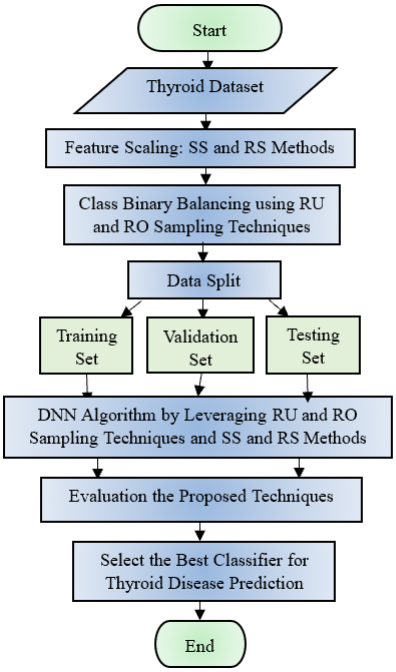


Fig. 2: Flowchart of the hybrid DNN techniques for Thyroid disease prediction.

2.2. Proposed Techniques

Outliers present in the thyroid dataset adversely influence the classifier's training, leading to longer training times and diminishing performance. Furthermore, some features have large values and a wide range, which further exacerbates the increase in training durations. As a result, the implementation of a robust feature scaling technique is required.

- **Feature Scaling**

In previous research, common feature scaling techniques such as normalization scaling and standard scaling (SS) were applied to the thyroid disease dataset. In this study, due to the presence of outliers in the public dataset, the robust scaling (RS) method is proposed, which demonstrates superior performance compared to the two conventional scaling techniques. Robust scaling approach addresses the problem of outliers by using the median and quartiles. This technique scales data based on the interquartile range (IQR), which is the difference between the third quartile (Q_3) and the first quartile (Q_1), rather than relying on the mean. The mathematical formulation of the robust scaling equation is as follows:

$$x_{rs} = \frac{x - Q_1}{Q_3 - Q_1} \quad (1)$$

Here, x denotes the main value, and x_{rs} is the scaled value [27, 28]. This paper will compare two scaling techniques, standardization and robust scaling, along with dataset balancing methods applied to the thyroid disease dataset in Section 3.

- **RU and RO Sampling Techniques**

The imbalance refers to the unequal distribution of data across classes, which can significantly impair the performance of the classifiers, often biasing them toward the

class with more data. Therefore, balancing the dataset is crucial for improving the predictive performance in thyroid disease diagnosis. In this paper, random over and random under sampling techniques have been applied to balance the dataset.

Under sampling is a popular method for addressing class imbalance issues. It selects a subset of the majority class, resulting in high efficiency. The major limitation is that many majority class samples are discarded. By decreasing the size of the dominant class, under sampling achieves dataset balance. Retaining all instances from the minority class while randomly selecting an equal number of instances from the majority class creates a balanced dataset, suitable for further analysis [29]. Over sampling approach seeks to balance the dataset by increasing the representation of minority class samples. Instead of eliminating the abundant instances, new unique data points are generated through techniques such as resampling, bootstrapping, or the synthetic minority over-sampling [30].

The algorithms below formalize the implementation details of both class - balancing strategies employed in this study. Algorithm 1 specifies the RU Sampling procedure, which systematically reduces majority - class instances to equalize class proportions. Algorithm 2 outlines the RO Sampling procedure, which incrementally replicates minority - class instances until parity with the majority class is achieved.

Algorithm 1: Random Under (RU) Sampling Technique for Balancing Thyroid Dataset

Input:
- Original dataset $\mathcal{D} = \{(x_i, y_i)\}$ where $x_i \in \mathbb{R}^m$ represents feature vector, and $y_i \in \{0, 1\}$ represents class label.

Output:
- Balanced dataset $\mathcal{D}_{RU}$ with equal representation of minority and majority classes.

1. Identify minority and majority classes:
   - Let $\mathcal{D}_{min} = \{(x_i, y_i) \in \mathcal{D} | y_i = minority\ class\}$.
   - Let $\mathcal{D}_{maj} = \{(x_i, y_i) \in \mathcal{D} | y_i = majority\ class\}$.
2. Determine target sample count:
   - $n_{min} = |\mathcal{D}_{min}|$ (number of minority samples).
   - $n_{maj} = |\mathcal{D}_{maj}|$ (number of majority samples).
   - Set target_count $= n_{min}$.
3. Randomly under sample the majority class:
   - Shuffle $\mathcal{D}_{maj}$.
   - Select the first target_count samples from shuffled $\mathcal{D}_{maj} \rightarrow \mathcal{D}_{majSub}$.
4. Construct the balanced dataset:
   - $\mathcal{D}_{RU} = \mathcal{D}_{min} \cup \mathcal{D}_{majSub}$.
5. Return $\mathcal{D}_{RU}$.

Algorithm 2: Random Over (RO) Sampling Technique for Balancing Thyroid Dataset

Input:
- Original dataset $\mathcal{D} = \{(x_i, y_i)\}$ where $x_i \in \mathbb{R}^m$ represents feature vector, and $y_i \in \{0, 1\}$ represents class label.

Output:
- Balanced dataset $\mathcal{D}_{RO}$ with equal representation of minority and majority classes.

1. Identify minority and majority classes:
   - Let $\mathcal{D}_{min} = \{(x_i, y_i) \in \mathcal{D} | y_i = minority\ class\}$.
   - Let $\mathcal{D}_{maj} = \{(x_i, y_i) \in \mathcal{D} | y_i = majority\ class\}$.
2. Determine target sample count:
   - $n_{min} = |\mathcal{D}_{min}|$ (number of minority samples).
   - $n_{maj} = |\mathcal{D}_{maj}|$ (number of majority samples).
   - Set $target_{count} = n_{maj}$.
3. Randomly oversample the minority class:
   - Initialize $\mathcal{D}_{minRep} = \mathcal{D}_{min}$.
   - While $|\mathcal{D}_{minRep}| < target_count$:
     - Shuffle $\mathcal{D}_{min}$.
     - For each sample (x, y) in shuffled $\mathcal{D}_{min}$:
       - Append a copy of (x, y) to $\mathcal{D}_{minRep}$.
       - If $|\mathcal{D}_{minRep}| = target_count$, break.
4. Construct the balanced dataset:
   - $\mathcal{D}_{RO} = \mathcal{D}_{maj} \cup \mathcal{D}_{minRep}$.
5. Return $\mathcal{D}_{RO}$.

Table 1 presents the number of samples in the original (imbalanced) and balanced datasets through the RU and RO sampling techniques. Figs. 3 and 4 demonstration the balanced distribution of the thyroid dataset achieved through the RU sampling method and the RO sampling method.

- **Dataset Splitting for Thyroid Disease Prediction**

After applying standard scaling and robust scaling methods, as well as random under (RU) and random over (RO) sampling techniques to the thyroid dataset, the data was split into testing and training sets with an 80:20 ratio.

Table 1: Imbalanced and balanced instances of thyroid disease.

	Total number of instances	Minority of instances	Majority of instances	Ratio of majority to minority instances
Imbalanced samples	3772	291	3481	11.96
RU sampling method	582	291	291	1.00
RO sampling method	6962	3481	3481	1.00

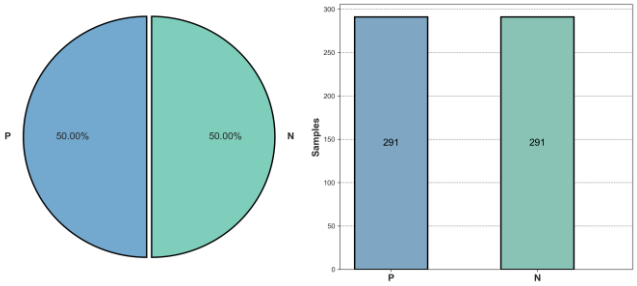


Fig. 3: Distribution of balanced thyroid dataset using RU sampling method.

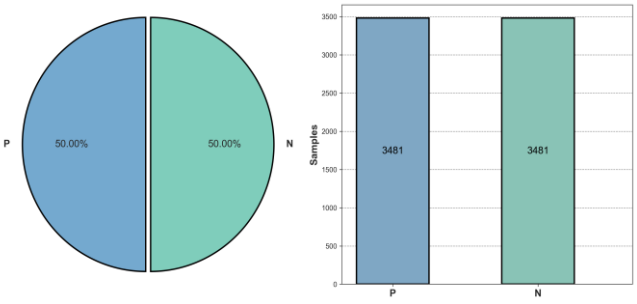


Fig. 4: Distribution of balanced thyroid dataset using RO sampling method.

The training set was further divided into training and validation subsets in an 85:15 ratio. All experiments, aimed at comparing deep neural network algorithms, were conducted using Jupyter Notebook and the Python programming language. Pre-processing was carried out using Python libraries, such as NumPy, Pandas, Matplotlib, and Seaborn. The Scikit-learn library, along with Keras and TensorFlow, was used for training of the DNN algorithm.

- **Deep Neural Network algorithm**

Most prior studies on thyroid disease prediction have predominantly utilized classical machine learning algorithms. This paper, leverages deep neural networks (DNN), a subset of deep learning, to classify the dataset and detect thyroid disease. Fig. 5 illustrates the overall architecture of the DNN used in this study, consists of one input layer, two hidden layers, and one output layer. A sequential approach was used to initialize the model and build the layers. In the proposed algorithm, the number of neurons in the input layer was aligned with the features of the dataset. The first hidden layer comprised 256 neurons, the second hidden layer contained 64 neurons, and the output layer, designed for the classification of the thyroid dataset, featured a single neuron. The activation of the input and hidden layers was facilitated by a Rectified Linear Unit (ReLU) function, which processes inputs by converting any negative values to zero. The ReLU function is mathematically defined as:

$$ReLu(x) = \max(0, x) \quad (2)$$

The output layer utilized a sigmoid activation function, defined as follows:

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Here, x is equal value of the sum of weighted, and e is equal Euler's number [16].

To prevent overfitting, a dropout layer was added after each layer. For optimization, the Adam optimizer was applied for training the deep learning algorithm. The loss was calculated using binary cross-entropy, while accuracy served as the metric for evaluating the algorithm. The model was trained for 300 epochs, with a batch size of 32, a learning rate of 0.001, and a validation split of 0.15. Deep neural network classification was performed by Python 3.9.12, leveraging libraries such as Scikit-learn, Keras, and TensorFlow for algorithm development and evaluation. To provide a concise overview of the experimental setup, all key parameters used in the proposed deep neural network algorithm are summarized in Table 2.

- **Performance Assessment Metrics**

This study utilizes various measures, including accuracy, precision, F1-score, recall, receiver operating characteristic (ROC) curve, and area under the curve (AUC) to assess the performance of robust artificial intelligence methods applied to the thyroid dataset. Also, the performance of the proposed algorithms is evaluated using the confusion matrix, as shown in Table 3. To define the metric formulas, it is essential to first outline the main components of the confusion matrix. TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

Table 2: Parameters for the proposed deep neural network algorithm.

No.	Parameter	Value / Description
1	Activation Function (Hidden)	ReLU
2	Activation Function (Output)	Sigmoid
3	Optimizer	Adam
4	Loss Function	Binary Cross-Entropy
5	Learning Rate	0.001
6	Batch Size	32
7	Epochs	300
8	Hidden Layer 1 Neurons	256
9	Hidden Layer 2 Neurons	64
10	Output Layer Neurons	1

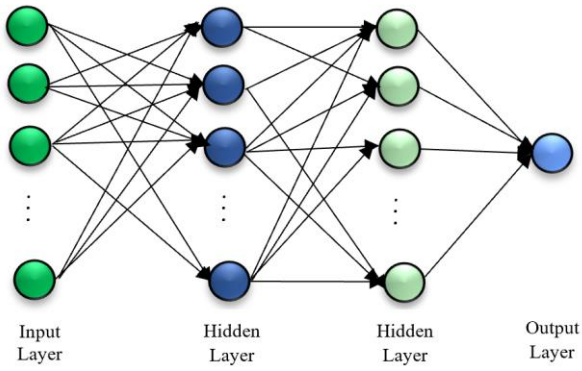


Fig. 5: Deep neural network for thyroid disease prediction.

- **Performance Assessment Metrics**

This study utilizes various measures, including accuracy, precision, F1-score, recall, receiver operating characteristic (ROC) curve, and area under the curve (AUC) to assess the performance of robust artificial intelligence methods applied to the thyroid dataset. Also, the performance of the proposed algorithms is evaluated using the confusion matrix, as shown in Table 3. To define the metric formulas, it is essential to first outline the main components of the confusion matrix. TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

Accuracy serves as a measure of the model's overall proficiency, signifying the percentage of the test dataset that the model accurately classifies. Correctly identified instances are termed TP and TN, whereas misclassified instances are FP and FN. Accuracy is determined as follows:

$$Accuracy\ (\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (4)$$

Table 3: The confusion matrix.

		Actual Class	
		+	-
Predicted Class	+	TP	FP
	-	FN	TN

Conversely, precision and recall seek to capture the proportion of TP and TN. Precision, in particular, assesses the classifier's ability to refrain from incorrectly labeling negative instances as positive, and it is defined as follows:

$$Precision\ (\%) = \frac{TP}{TP + FP} \times 100 \quad (5)$$

Alternatively, recall quantifies the sensitivity of the model. It represents the proportion of correctly predicted instances for a given class relative to the total occurrences of that class, and is expressed as follows:

$$Recall\ (\%) = \frac{TP}{TP + FN} \times 100 \quad (6)$$

The F1-score, primarily used for comparing classifier models rather than global precision, reflects the weighted harmonic average of the Precision and Recall metrics. By giving more importance to lower values, the F1-score ensures that a classifier achieves a high score only when both precision and recall are strong. The formula is defined as follows [18, 20]:

$$F1 - measure\ (\%) = 2 \times \frac{1}{(1/Precision) + (1/Recall)} \times 100 \quad (7)$$

Furthermore, a thorough evaluation of the hybrid algorithms is conducted using the AUC and ROC curve, providing a comprehensive analysis of their performance.

3. Results and Discussion

After preprocessing the thyroid disease dataset, which includes examining missing values and outliers, feature scaling is performed using both standard scaling and robust scaling methods. Due to the imbalance in the binary classes, the dataset is balanced utilizing the RU and RO sampling techniques. Subsequently, the data is partitioned into three segments including training, validation, and testing, employing the train-test split function from the Scikit-learn framework in Python. Finally, the deep neural network algorithm is applied in four scenarios to evaluate the prediction of thyroid disease using various classifier metrics. The following are the four hybrid approaches used for the classification of thyroid disease in this study:

- Hybrid technique deep neural network and RU sampling, by leveraging standard scaling,
- Hybrid technique deep neural network and RU sampling, by leveraging robust scaling,
- Hybrid technique deep neural network and RO sampling, by leveraging standard scaling,
- Hybrid technique deep neural network and RO sampling, by leveraging robust scaling.

Table 4 presents a comparison and evaluation of the deep neural network algorithm across the four specified scenarios, utilizing metrics including accuracy, recall, F1-score, and precision. In Table 5, the performance of the hybrid deep neural network techniques is assessed using two criteria: AUC and computational time, providing a more comprehensive comparison. Figs. 6 and 7 illustrate the comparison and evaluation of the performance of four hybrid deep neural network techniques based on classifier metrics. It is evident that the hybrid deep neural network approach utilizing RO sampling and robust scaling achieves superior performance among the proposed hybrid techniques.

Table 4: Assessment of the proposed techniques.

Techniques	Assessment criteria (%)			
	Accuracy	Precision	Recall	F1-score
RU_SS	94.87	96.23	92.73	94.44
RU_RS	95.73	94.64	96.36	95.50
RO_SS	98.64	97.31	100.00	98.64
RO_RS	98.71	97.45	100.00	98.71

Table 5: Comparison of AUC and computational time for the proposed techniques.

Criteria	Techniques			
	RUS_SS	RUS_RS	ROS_SS	ROS_RS
AUC (%)	97.62	98.21	99.08	99.31
Computational Time (s)	21.11	22.42	143.18	141.72

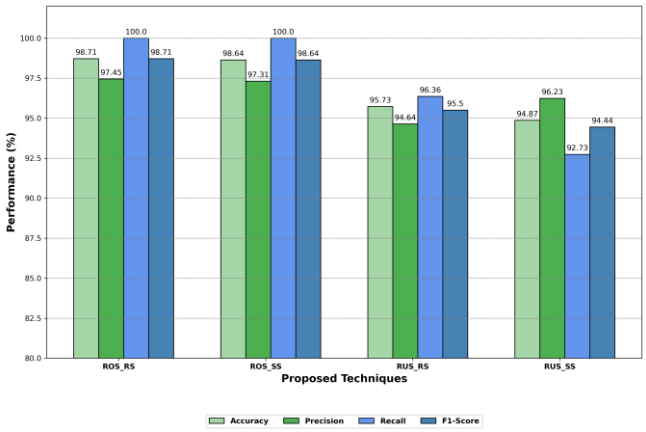


Fig. 6: Performance evaluation of the proposed techniques on thyroid dataset.

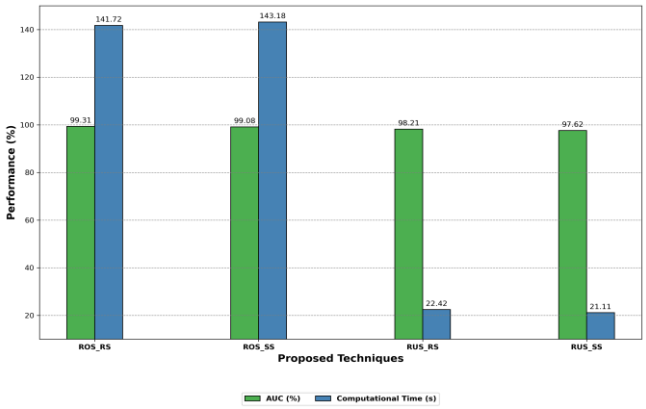


Fig. 7: Comparison AUC (%), and computational time of the proposed techniques on thyroid dataset.

It is also worth mentioning that the two hybrid deep neural network techniques employing RO sampling, in conjunction with both standard scaling and robust scaling methods, outperform the hybrid techniques using RU sampling in classifying the thyroid disease dataset. Nevertheless, the RU sampling hybrid models show computational times approximately one-seventh of their RO sampling counterparts. For a more comprehensive evaluation, Figs. 8, 9, 10, and 11 present the ROC curves and confusion matrices of the proposed techniques, demonstrating the advantage of the hybrid deep neural network technique utilizing RO sampling and robust scaling.

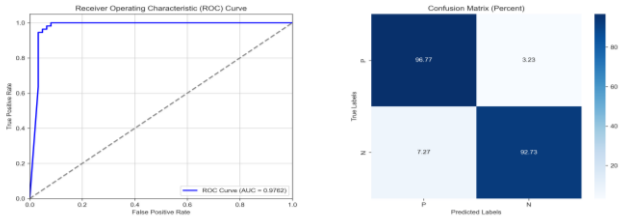


Fig. 8: ROC curve and confusion matrix of hybrid DNN technique utilizing RU sampling and SS.

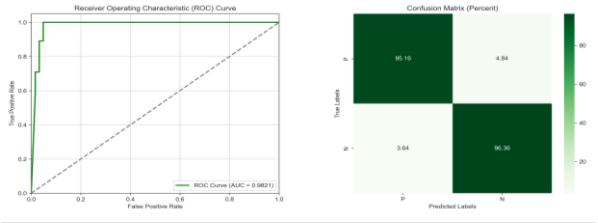


Fig. 9: ROC curve and confusion matrix of hybrid DNN technique utilizing RU sampling and RS.

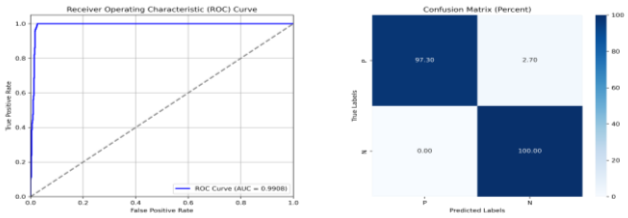


Fig. 10: ROC curve and confusion matrix of hybrid DNN technique utilizing RO sampling and SS.

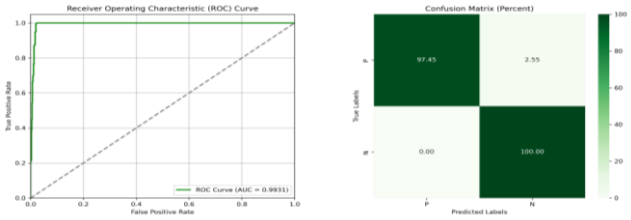


Fig. 11: ROC curve and confusion matrix of hybrid DNN technique utilizing RO sampling and RS.

Figs. 12, 13, 14, and 15 illustrate the accuracy curves for the training and testing techniques of deep neural networks utilizing RU sampling with standard scaling (SS), RU sampling with robust scaling (RS), RO sampling with standard scaling (SS), and RO sampling with robust scaling (RS), respectively. These findings indicate that the deep neural networks using RO sampling with RS achieve the best performance after 300 epochs.

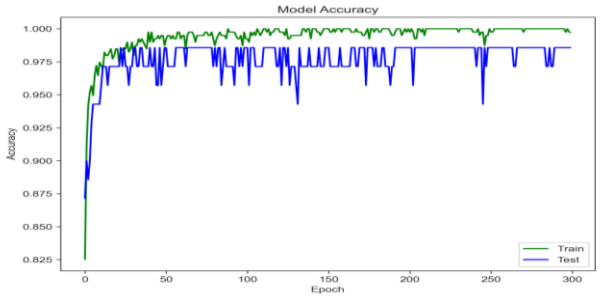


Fig. 12: The training and testing accuracy for the DNN employing RU sampling with SS.

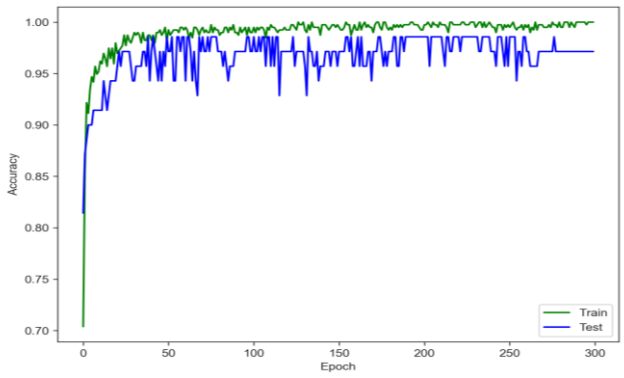


Fig. 13: The training and testing accuracy for the DNN employing RU sampling with RS.

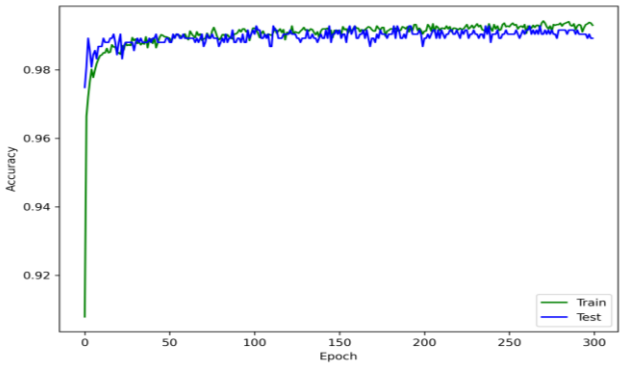


Fig. 14: The training and testing accuracy for the DNN employing RO sampling with SS.

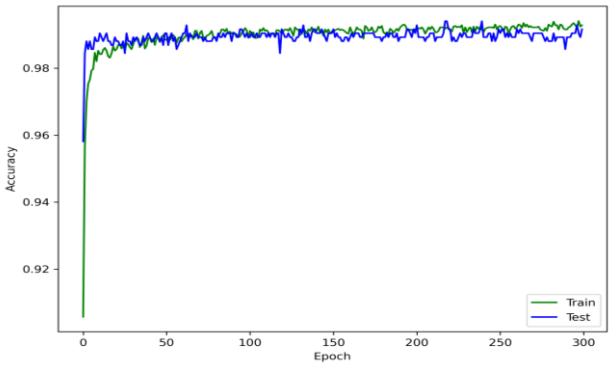


Fig. 15: The training and testing accuracy for the DNN employing RO sampling with RS.

In Table 6, the accuracy and loss values of the proposed techniques are compared. The hybrid deep neural network technique with RO sampling and robust scaling outperforms the other techniques. Figs. 16, 17, 18, and 19 depict the loss curves for the training and testing techniques of hybrid DNN employing RU sampling with SS, RU sampling with RS, RO sampling with SS, and RO sampling with RS, respectively. These results illustration that the hybrid DNN using RO sampling with RS achieve the best performance after 300 epochs.

Table 6: Accuracy and loss comparison of the proposed techniques.

Criteria	Techniques			
	RUS_SS	RUS_RS	ROS_SS	ROS_RS
Accuracy (%)	94.87	95.73	98.64	98.71
Loss	0.52	0.44	0.09	0.07

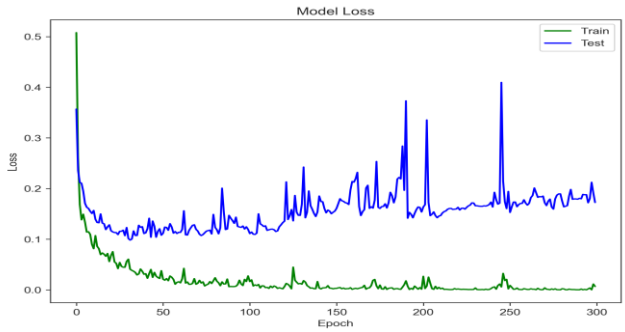


Fig. 16: The training and testing loss for the DNN employing RU sampling with SS.

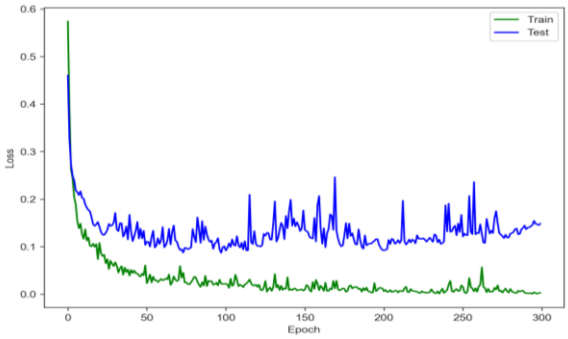


Fig. 17: The training and testing loss for the DNN employing RU sampling with RS.

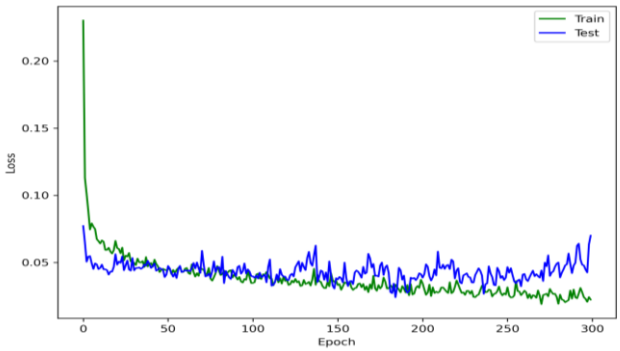


Fig. 18: The training and testing loss for the DNN employing RO sampling with SS.

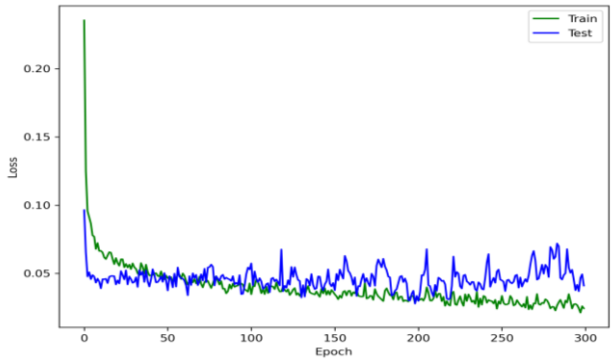


Fig. 19: The training and testing loss for the DNN employing RO sampling with RS.

The figures depicting the accuracy and loss curves for the training and testing sets of the hybrid deep neural network techniques with class balancing methods reveal that random over sampling exhibits higher accuracy and reduced loss compared to random under sampling, thereby delivering enhanced performance in the diagnosis and prediction of thyroid disorders.

To highlight the effectiveness of the proposed hybrid deep learning approach, a comparative analysis was conducted against recent deep learning methods applied to thyroid disease prediction on imbalanced datasets. Table 7 presents a side-by-side comparison of the proposed model with four relevant studies using key performance metrics: accuracy, precision, and recall. Additionally, Fig. 20 visually illustrates these comparisons to provide a clearer understanding of relative performance across models.

4. CONCLUSION

The identification of thyroid disorders has gained significant importance in the medical field, particularly due to the necessity of early diagnosis and appropriate treatment to prevent complications and reduce mortality rates. One of the significant techniques for disease prediction in recent years has been the application of artificial intelligence methods, such as machine learning and deep learning. These methods have attracted considerable attention from researchers due to their ability to enhance predictive accuracy. This study utilized a publicly available thyroid dataset comprising 3772 imbalanced samples from the UCI Machine Learning Repository. A hybrid robust neural network technique was employed to predict thyroid disease, incorporating class-balancing algorithms such as Random Over (RO) sampling and Random Under (RU) sampling. To address the presence of outliers in the thyroid dataset, the Robust Scaling method was applied to enhance the algorithm's performance. Four hybrid deep neural network techniques were evaluated on this dataset. Upon evaluating and comparing the performance of these hybrid techniques using standard classification metrics, it was observed that the deep neural network combined with RO sampling outperformed the model employing RU sampling in terms of classification accuracy, although at the expense of higher computational time. The findings highlight the superiority of the hybrid deep neural network technique by leveraging random over sampling combined with robust scaling, achieving an accuracy of 98.71%, recall of 100%, precision of 97.45%, F1-score of 98.71%, AUC of 99.31%, and a loss function value of 0.07.

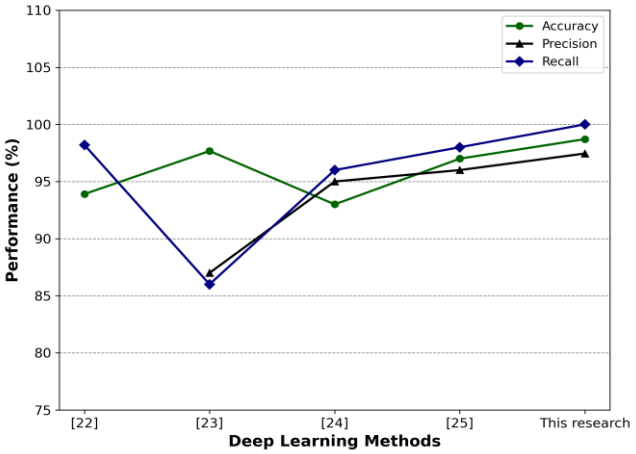


Fig. 20: Performance comparison of the proposed technique with recent deep learning methods on the imbalanced thyroid datasets

Table 7: Comparison of the proposed technique with recent deep learning methods for Thyroid disease prediction on the imbalanced dataset.

Ref.	Deep Learning Method	Balancing Technique	Accuracy (%)	Precision (%)	Recall (%)
[22]	RF, DNN	SMOTE	93.90	-	98.20
[23]	RNN	SMOTE	97.67	87.00	86.00
[24]	LSTM	SMOTE-NC	93.00	95.00	96.00
[25]	SVM, LSTM	SMOTE	97.00	96.00	98.00
This research	DNN	RO Sampling	98.71	97.45	100.00

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Elahe Moradi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

DECLARATION OF COMPETING INTEREST

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The ethical issues; including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy has been completely observed by the authors.

REFERENCES

[1] M. L. Raymer, T. E. Doom, L. A. Kuhn, and W. F. Punch, "Knowledge discovery in medical and biological datasets using a hybrid bayes classifier/evolutionary algorithm," *IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics)*, vol. 33, no. 5, pp. 802–813, Oct. 2003.

[2] A. Mehrno, R. Oktaş, and M. S. Odabas, "Feature selection of thyroid disease using deep learning: A literature survey," *Black Sea Journal of Engineering and Science*, vol. 3, no. 3, pp. 109–114, Jun. 2020.

[3] A. Koliji, S. MihanDoost, N. Ezzati, and E. Mostafapour, "Improving the diagnosis of sudden cardiac death by using non-linear features of the ECG signal and hybrid RBF," *J. Appl. Res. Electr. Eng.*, vol. 2, no. 2, pp. 120-126, 2023.

[4] P. Vafadoost Sabzevar, A. Hajipour, and H. Tavakoli, "Investigating robust tracking of type 1 diabetes control using model-free controllers," *J. Appl. Res. Electr. Eng.*, vol. 3, no. 2, pp. 232-239, 2024.

[5] E. Moradi, "Diagnosis of coronary heart disease via robust artificial neural network classifier by adaptive synthetic sampling approach," *Iranian Journal of Electrical and Electronic Engineering (IJEEE)*, vol. 20, no. 4, pp. 55–67, 2024.

[6] N. M. Xi, L. Wang, and C. Yang, "Improving the diagnosis of thyroid cancer by machine learning and clinical data," *Scientific Reports*, vol. 12, no. 1, Jul. 2022.

[7] G. Mollica et al., "Classification of thyroid diseases using machine learning and bayesian graph algorithms," *IFAC-PapersOnLine*, vol. 55, no. 40, pp. 67–72, Jan. 2022.

[8] M. Parishani and M. Rasti-Barzoki, "CWBCM method to determine the importance of classification performance evaluation criteria in machine learning: Case studies of COVID-19, diabetes, and thyroid disease," *Omega*, vol. 127, p. 103096, Sep. 2024.

[9] X. Ren et al., "Machine learning reveals salivary glycopatterns as potential biomarkers for the diagnosis and prognosis of papillary thyroid cancer," *International Journal of Biological Macromolecules*, vol. 215, pp. 280–289, Aug. 2022.

[10] Z. Chen et al., "Machine learning-aided metallomic profiling in serum and urine of thyroid cancer patients and its environmental implications," *The Science of the Total Environment*, vol. 895, p. 165100, Oct. 2023.

[11] K. Chandel, V. Kunwar, S. Sabitha, T. Choudhury, and S. Mukherjee, "A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques," *CSI Transactions on ICT*, vol. 4, no. 2–4, pp. 313–319, Dec. 2016.

[12] S. S. Islam, Md. S. Haque, M. S. U. Miah, T. B. Sarwar, and R. Nugraha, "Application of machine learning algorithms to predict the thyroid disease risk: an experimental comparative study," *PeerJ Computer Science*, vol. 8, p. e898, Mar. 2022.

[13] V. Manimekalai, S.Gomathi alias Rohini, "Enhanced accuracy in thyroid disease cassification: A comparative analysis of random forest and decision tree methods," *Tuijin Jishu/Journal of Propulsion Technology*, vol. 44, no. 6, 2023.

[14] H. H. S. Kumar, "A novel approach of SVM based classification on thyroid disease stage detection," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020.

[15] T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad, "Empirical method for thyroid disease classification using a machine learning approach," *BioMed Research International*, vol. 2022, pp. 1–10, Jun. 2022.

[16] P. Poudel, A. Illanes, E. J. G. Ataide, N. Esmaeili, S. Balakrishnan, and M. Friebe, "Thyroid ultrasound texture classification using autoregressive features in conjunction with machine learning approaches," *IEEE Access*, vol. 7, pp. 79354–79365, Jan. 2019.

[17] D. Qiao et al., "Exploring the potential of thyroid hormones to predict clinical improvements in depressive patients: A machine learning analysis of the real-world based study," *Journal of Affective Disorders*, vol. 299, pp. 159–165, Feb. 2022.

[18] L. Aversano et al., "Thyroid disease treatment prediction with machine learning approaches," *Procedia Computer Science*, vol. 192, pp. 1031–1040, Jan. 2021.

[19] A. Sharifi and K. Alizadeh, "Comparison of the particle swarm optimization with the genetic algorithms as a training for multilayer perceptron technique to diagnose

thyroid functional disease," *Shiraz E-Medical Journal*, vol. 22, no. 1, Jun. 2020.

[20] S. Prathibha, D. Dahiya, C. R. R. Robin, C. V. Nishkala, and S. Swedha, "A novel technique for detecting various thyroid diseases using deep learning," *Intelligent Automation & Soft Computing*, vol. 35, no. 1, pp. 199–214, Jan. 2023.

[21] D. K. Gummalla, S. Ganesan, S. Pokhrel, and N. Somasiri, "Enhanced early detection of thyroid abnormalities using a hybrid deep learning model: A sequential CNN and K-means clustering approach," *Journal of Innovative Image Processing*, vol. 6, no. 3, pp. 244–261, Jun. 2024.

[22] Y. Attiga et al., "Applying deep learning to public health: using unbalanced demographic data to predict thyroid disorder," in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, IEEE, 2018, pp. 851–58.

[23] S. Tabassum, S. F. F. Rumky, and M. F. Shahariar, "Thyroid Disease Analysis and Prediction by Using Machine Learning and Deep Learning: A Comparative Approach," B.Sc. thesis, Dept. of Information and Communication Engineering, East West University, Dhaka, Bangladesh, Sep. 2022.

[24] A. Raza et al., "Enhanced interpretable thyroid disease diagnosis by leveraging synthetic oversampling and machine learning models" *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, Nov. 2024, p. 364.

[25] K. Anitha et al., "Exploring the experimental effectiveness of a hybrid deep learning approach in detecting thyroid disease using intelligent predictions," in *2024 International Conference on Computing and Data Science (ICCDS)*, IEEE, 2024, pp. 1–5.

[26] Q. Ross, Thyroid Disease. *UCI Machine Learning Repository*. 1987 [Online]. Available: https://archive.ics.uci.edu/dataset/102/thyroid+disease.

[27] E. Moradi, "A data-driven based robust multilayer perceptron approach for fault diagnosis of power transformers," in *20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Feb. 2024.

[28] E. Moradi et al., "Robust deep neural network-based internet of things for power transformer fault diagnosis under imbalanced data and uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 168, July 2025.

[29] N. X.-Y. Liu, N. J. Wu, and N. Z.-H. Zhou, "Exploratory undersampling for class-imbalance learning," *IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics)*, vol. 39, no. 2, pp. 539–550, Apr. 2009.

[30] T. Zhu, Y. Lin, and Y. Liu, "Synthetic minority oversampling technique for multiclass imbalance problems," *Pattern Recognition*, vol. 72, pp. 327–340, Jul. 2017.

BIOGRAPHY



Elahe Moradi received her Bachelor's degree in Electrical-Electronic Engineering from K. N. Toosi University of Technology, her Master's degree in Electrical Control Engineering from the Islamic Azad University, South Tehran Branch, and her Ph.D. in Electrical Control Engineering from the Islamic Azad University, Science and Research Branch in Tehran. Her research activities and interests lie in robust control, delay systems, optimal control, fault detection and diagnosis, and artificial intelligence, including machine learning and deep learning. She is currently serving as an Assistant Professor in the Department of Electrical and Computer Engineering at the Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch of Islamic Azad University, Tehran, Iran.

Copyrights
© 2025 Licensee Shahid Chamran University of Ahvaz, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution –NonCommercial 4.0 International (CC BY-NC 4.0) License (http://creativecommons.org/licenses/by-nc/4.0/).