Hybrid AdaBoost-PSO Model for Thyroid Disease Diagnosis

Ammar A. Kazm[1], Zamen Abood Ramadhan[1], Jaafar Sadiq Alrubaye[1], Iman Kadhim Ajlan[1], Jamal Kh-Madhloom[1], Yasir Hussein Shakir[2], Eshaq Aziz Awadh AL Mandhari[3], Fahad Taha Al-Dhief[4], Nurhizam Safie[4] and Ali Hashim Abbas[5]

[1]Computer Department, College of Education for Pure Sciences, Wasit University, 52001 Alkut, Iraq
[2]College of Graduate Studies, University Tenaga Nasional, 43000 Kajang, Malaysia
[3]Graduate School of Technology, Asia Pacific University of Technology and Innovation, Technology Park Malaysia, Bukit Jalil, 57000 Kuala Lumpur, Malaysia
[4]Center for Software Technology and Management, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[5]Department of Computer Technical Engineering, College of Information Technology, Imam Ja'afar Al-Sadiq University, 66002 Al-Muthanna, Iraq
aawaad@uowasit.edu.iq, Z.ramadaan@uowasit.edu.iq, Jsadiq@uowasit.edu.iq, eajlan@uowasit.edu.iq, jamalkh@uowasit.edu.iq, yasserhessein19855@gmail.com, eshaq.almandhari@utas.edu.om, fahad@ukm.edu.my, nurhizam@ukm.edu.my, alsalamy1987@gmail.com

Keywords: AdaBoost-PSO, Random Under Sampling, SMOTE-SVM, K-Means-SMOTE, Machine Learning, Thyroid Diseases.

Abstract: The endocrine thyroid gland is an essential organ for regulating metabolism, production of primarily triiodothyronine (T3) and thyroxine (T4) which are necessary for digestion, heart rate and their imbalance could result in metabolism disorder. Diagnosis of thyroid diseases could be frustrated, If the lab techniques are used. This study was designed to use classical machine learning (ML) to improve the accuracy and precision of thyroid symptoms diagnosis using UCI thyroid dataset. The applied stage includes handling missing values using the K-Nearest Neighbors (KNN) imputation method. At the second stage to address class imbalance, we applied three data resampling method: Random Under Sampling, SMOTE-SVM, and K-Means-SMOTE, and many of the classical machine learning algorithms were utilized including Logistic Regression, AdaBoost, SGD, SVM, and KNN. Additionally, a hybrid AdaBoost-PSO (Particle Swarm Optimization) model was merged to enhance classification performance in terms of accuracy, precision, recall, F1-score, mcc, roc confusion matrix, true positive, false positive, true negative, and false negative. The results showed that K-Means-SMOTE + AdaBoost + PSO pipeline model achieves accuracy of 99.73%, precision of 0.991, recall of 98.40%, F1-score of 98.80%, MCC of 97.60%, and an AUC score of 98.40%. The corresponding confusion matrix indicated excellent classification capability with 64 true positives, 1 false positive, 2 false negatives, and 1065 true negatives as compared with other models. As well as the heat map analysis showed that thyroid-stimulating hormone (TSH) has a high level and both free thyroxine index (FTI) and T4 have a low level among the numbers of patients.

1 INTRODUCTION

Hypothyroidism is a globally prevalent condition characterized by inadequate thyroid hormone production. The definition of hypothyroidism is mostly biochemical because of the significant variability in clinical symptoms. Ninety-nine percent of primary hypothyroidism cases are associated with a shortage of thyroxine (T4) and triiodothyronine (T3) hormones [1], [2]. A deficiency in T4 and T3 hormones, generated by the thyroid gland, results in elevated production of thyroid-stimulating hormone (TSH) via a negative feedback loop [3]. Hypothyroidism presents with nonspecific symptoms including weight gain, weariness, impaired focus, depression, menstrual irregularities, and constipation, which vary according to age, sex, and other variables. Autoimmune thyroiditis, often known as Hashimoto's disease, is the most prevalent manifestation of this condition.

Recent studies have shown a significant 2% increase in the prevalence of hypothyroidism worldwide, despite the presence of adequate iodine in daily food [4], [5]. Various factors, such as gender, age, family history, and environmental conditions, influence this increase. Several factors have contributed to this increase, including geographical areas, ethnicity, aging, and the amount of iodine intake [3], [6], [7]. One study [8] conducted in Iraq in 2023 indicates that thyroid disorders are a significant public health concern in Iraq, with a notable prevalence of hypothyroidism compared to hyperthyroidism. The results specifically revealed that 68.75% of the patients received a diagnosis of hypothyroidism, while only 31.25% received a diagnosis of hypothyroidism. The most common complications of hypothyroidism are increasing serum cholesterol levels, the risk of coronary artery disease, and cardiovascular mortality [9]. The economic burden of hypothyroidism is fairly high, especially for patients with other underlying diseases such as diabetes and hemodialysis [10]. The common clinical method for diagnosing equally primary hypothyroidism is to check the serum concentration of TSH.

Hypothyroidism can present via several symptoms, including fatigue, decreased cold tolerance, constipation, depression, bradycardia, and weight gain. Physicians modify the medication dosage based on the patient's condition. Hyperthyroidism is characterized by the overproduction of thyroid hormones. Symptoms of hyperthyroidism include sleep disturbances, irritability, agitation, fine, brittle hair, heightened sweating, hand tremors, tachycardia, anxiety, skin atrophy, and muscle weakness [11].

In recent years, artificial intelligence and machine learning methods have garnered heightened interest from biomedical researchers. Machine learning algorithms have proven indispensable in addressing intricate and nonlinear issues when creating predictive models [12] - [14]. Predictive modeling faces challenges due to imbalanced predictions, as most machine learning methods for classification assume an equal number of samples for each class. As a result, models with low predictive performance emerge, especially for the minority group [15], [16]. This is problematic because, in most cases, the minority class is more significant than the majority class, making the problem more susceptible to minority class classification errors than to majority class classification errors [17], [18].

Moreover, existing studies are typically limited in scope, addressing only a few types of thyroid diseases and primarily focusing on binary classification, which limits their applicability to real-world disease detection scenarios [19] - [21]. This study aims to address these limitations by using the oversampling technique combined with a hybrid proposed model.

In other words, this study aims to overcome the issue of class imbalance by augmenting the minority class with synthetic data, while simultaneously ensuring robust performance in hypothyroidism prediction. The contributions of this study are listed as follows:

- Resolved missing data challenges by implementing K-Nearest Neighbors (KNN) imputation and ensuring robust dataset completeness for analysis.
- Evaluated three resampling techniques (RUS, SMOTE-SVM, and K-Means-SMOTE) to mitigate class imbalance and enhance model performance.
- To successfully reduce false negative predictions in healthy individuals by enhancing the generalization capability of the proposed hybrid AdaBoost-PSO models when combined with the K-Means-SMOTE as an oversampling technique.
- The optimal hybrid model, the K-Means-SMOTE-AdaBoost-PSO model, was compared with five other machine learning algorithms as a comprehensive evaluation strategy. The results demonstrated their efficacy in accurately predicting thyroid diseases, thus confirming their reliability and precision.

2 LITERATURE REVIEW

Early prediction of thyroid disease and its classification as hyperthyroidism or hypothyroidism is critical for optimizing patient treatment and recovery outcomes. Recent research has extensively employed deep learning and machine learning techniques to improve the prediction and diagnosis of thyroid diseases. To identify the new various techniques used in the detection of thyroid diseases. This paper provides a detailed overview of various methods used in the prediction and detection of thyroid disorders. The review also highlights whether or not these studies used sampling strategies to ensure more accurate and reliable models in current research investigations.

In [22], authors used several machine learning methods to diagnose thyroid disorders using a dataset obtained from a reputable laboratory in Kashmir. The authors used four classification models: KNN, SVM, DT, and NB. Out of all the models, the DT model exhibited exceptional performance, attaining the highest accuracy rate of 98.89%. This study emphasizes the efficacy of decision trees in diagnosing thyroid illnesses, highlighting their potential to enhance clinical decision-making. However, the absence of specific information on the size of the dataset might potentially impact on the applicability of the findings. The study in [11] introduced a comparative study on thyroid disease detection using KNN and NB classification techniques. The authors utilized a dataset with numerous samples, demonstrating that the NB classifier outperformed KNN in accuracy. The NB model achieved an accuracy of 87.2%, compared to 83.5% for KNN. However, both models may encounter difficulties when dealing with imbalanced datasets, leading to a potential bias in performance towards the majority class. Machine learning models, including a multi-layer perceptron (MLP), SVM, RF, DT, NB, LR, and KNN, are used to predict thyroid disease in [23]. A sample dataset of 1250 is collected from hospitals and labs in Iraq. The RF model obtained the best accuracy of 98.93% among the evaluated models. However, authors have focused on the application of various machine learning algorithms but do not address strategies to handle imbalanced datasets.

The authors in [24] conducted a study on the diagnosis of hypothyroidism using machine learning models. They analyzed clinical data from 1,296 patients who were experiencing hypothyroid symptoms for the first time. The research used many models, including Random Forest, Decision Tree, and Logistic Regression (LR), to forecast hypothyroidism by analyzing clinical signs such as fatigue, abnormal chilly sensations, jaundice, and weight gain. Random Forest had the highest performance among the investigated models, with an accuracy of 83%, a kappa value of 0.46, a sensitivity of 88%, and a specificity of 80%.

In [25], researchers compared the effectiveness of the XGBoost algorithm with the DT, LR, and KNN models. The results indicated that the XGBoost method enhances precision by 0.02% compared to the KNN algorithm. Another study [26] used the XGBoost algorithm to specifically predict the recurrence of thyroid cancer. The dataset consisted of clinical records of individuals with thyroid cancer, including a range of clinicopathological markers that are significant for predicting recurrence. Their research revealed that the XGBoost model surpassed conventional models with a remarkable accuracy rate of 97.74%. Furthermore, they reported the model's F1 score of 95.94%. In [27], the authors used an enhanced model of the XGBoost algorithm to predict thyroid disease. The authors used clinical data from fine-needle aspiration biopsies and gene expression analysis. The research conducted a comparison of several machine learning models, such as C5.0, CART, CHAID, Quest, LSVM, and random tree models. Out of all the models, the improved XGBoost model demonstrated exceptional performance, with an accuracy of 94.6%. Although the authors in [24] - [27] provide a comprehensive methodology, they do not address whether they utilized balanced data or took any steps to address data imbalance, resulting in performance accuracy below 97.8%. Besides using various machine learning models, some researchers specifically focus on employing deep learning algorithms with machine learning models for thyroid disease prediction.

Another study uses a deep neural network to predict and classify thyroid disease in [28]. To make thyroid predictions more accurate, the authors looked at several machine learning techniques, such as Artificial Neural Networks (ANN), Extra Trees, CatBoost, LightGBM, KNN, SVM, RF, DT, XGBoost, and Gaussian Naive Bayes (GaussianNB). The experimental results of the study indicated a 96% accuracy rate with an ANN classifier. However, the authors don't highlight the challenges posed by unbalanced datasets, where the majority class significantly outnumbers the minority class. This imbalance can lead to models that perform poorly, particularly for the minority class, which is often more critical in medical predictions. Another study [29] performed research focused on identifying hyperthyroidism and hypothyroidism, the two predominant thyroid illnesses. The categorization was performed using two methodologies: multinomial LR models and ANN. The study included 310 patients, and the models used demographic and hormonal information as input. The findings indicated the superior performance of the neural network model, achieving an average accuracy of 96.3%, compared to multinomial LR, which attained a mean accuracy of 91.4%, across all scenarios.

On the other hand, a small number of studies have applied sampling techniques combined with machine learning algorithms to the problem of thyroid prediction and classification because of their ability to enhance performance by balancing data. For example. The study [30] focuses on predictive therapy for thyroid disease using various machine learning techniques combined with the SMOTE technique to enhance performance prediction. The study employs a method to determine whether to escalate, diminish, or maintain the patient's current treatment level based on thyroid hormone metrics and other clinical data. The findings of the extra tree classifier (ETC) indicated the best-obtained results at an F1-Score of 84%, precision of 84%, recall of 84%, and accuracy of 84%.

In [31], the authors modified the dataset to ensure accurate predictions for thyroid disease. The authors conducted the categorization on balanced datasets after applying the down sampling technique, which provided more accurate performance. However, combining the down sampled technique with the RF model yielded an accuracy of only 94.8% in the experiment results, marking the highest accuracy achievable with this approach. Moreover, the authors don't mention details about the type of down sampled technique used in their work.

In [32], the authors applied data mining techniques to improve the diagnosis of hypothyroidism using a dataset from the UCI machine learning repository, consisting of 3163 samples, 151 of which were hypothyroid cases. The authors developed three classification models: LR, KNN, and SVM. Among these, the LR model trained on the balanced dataset performed the best, achieving an accuracy of 97.8%, an F-score of 82.26%, an AUC of 93.2%, and a Matthews correlation coefficient of 81.8%. However, the authors did not mention the specifics of the various sampling methods used in balance methods combined with machine learning models.

3 MATERIAL AND METHODS

The methodology outlined in this work consists of stages, as shown in Figure 1. In the first stage, we process the data to determine whether a missing value exists. In the second stage, the characteristics are detected, and any irrelevant properties are removed. The next stage involves suitably splitting the dataset. To address the imbalance in the training dataset, the proposed oversampling approach is implemented. The proposed hybrid prediction model, as well as other machine learning methods for hypothyroidism prediction, is implemented in the fifth step. Finally, the outcomes were compared and analyzed in terms of performance measures. As a result, we anticipate that our findings will be more consistent and dependable.

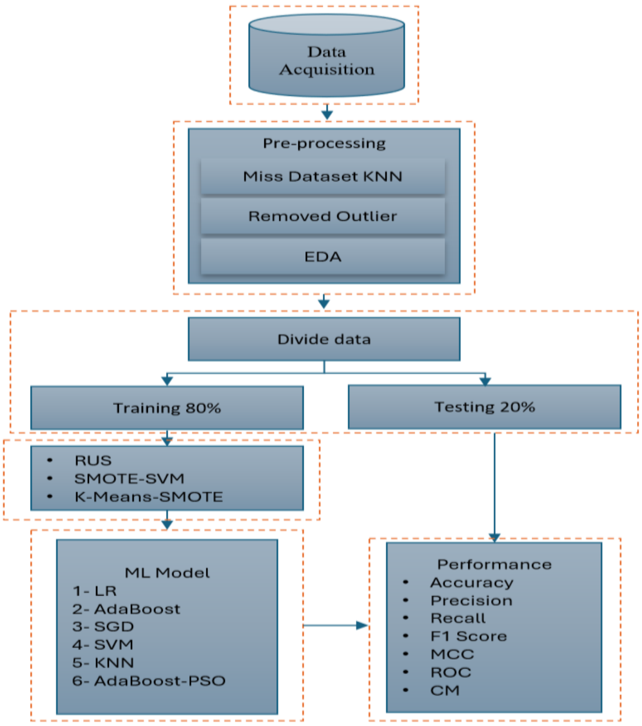

Figure 1: The flowchart of the proposed study.

3.1 Data Acquisition

The dataset used in this study has been collected and obtained from the Kaggle site [33]. This dataset includes 3772 patient samples, and 30 features related to patients' medical information for laboratory thyroid analysis. True (T) and false (F) are unambiguously used to represent the twenty-one categorical qualities out of the 30. Other characteristics are also real values. The individuals are split into two categories: normal status, represented by thyroid-negative (N), and hypothyroid condition, characterized by thyroid-positive (P).

3.2 Data Preprocessing

Preparation of the data is an essential step in data mining because it has a beneficial effect on the improved accuracy of the prediction of types of disease. Pre-processing includes the following steps: In the first step, data preparation and data cleaning are achieved, etc. In this process, to address missing data, we employed the K-Nearest Neighbor (KNN) imputation technique for the variables age, sex, thyroid-stimulating hormone (TSH), thyroxine (TT4), (T4U), and free thyroxine index (FTI). The KNN imputation estimates a missing value $\hat{x}_i$ by averaging the values of its K nearest neighbors in the feature space following equation:

$$\hat{x}_i = \frac{1}{K} \sum_{j \in N_k(i)} x_j. \quad (1)$$

Where $N_k(i)$ denotes the set of K nearest neighbors of instance i based on a chosen distance metric. Additionally, outlier values for age were removed by excluding records where age exceeded 100 years, reducing the maximum age in the dataset from 455 to 94. Visualizing correlation patterns through heatmap analysis is crucial for exploring datasets and understanding variable interactions. This method provides insights beyond formal modeling, helping to assess the suitability of statistical techniques for data analysis, particularly in the context of thyroid function measurements.

3.3 Dividing and Balancing Dataset

The dataset is divided into two subsets: training and test sets, with data employed to learn the model, measure the model's accuracy using unseen testing data, and lastly assess the model utilizing unseen testing data. We divided the dataset (3772 instances) into two subsets: 70% for training and 30% that were used to test and assess the models' overall performance. The binary prediction algorithms experience bias-to-majority and ignore the minority class when handling an imbalanced dataset without appropriate optimization phases. This will lead to the model's inability to distinguish between the two classes, achieving high accuracy but yielding low AUC (area under the curve) results. In this study, our main objective is to achieve class balance in the training data before feeding it into the classification model and evaluating prediction performance with the best sampling technique used. The main objective of class balancing is to either augment the occurrence of the minority class or diminish the occurrence of the dominant class. This ensures that the number of samples in each class is approximately equal.

To address the class imbalance problem, we applied three resampling strategies that were employed: Random Under-Sampling (RUS), Support Vector Machine Synthetic Minority Oversampling Technique (SMOTE-SVM), and K-Means SMOTE. These techniques were applied to either oversample the minority class or under sample the majority class, thus generating a balanced training set. By increasing the representation of the minority class, these methods help prevent paradoxical or misleading outcomes that can arise from imbalanced datasets and ensure more consistent and reliable classification results.

3.4 Machine Learning Models

3.4.1 Logistic Regression (LR)

Logistic Regression (LR) is the most utilized statistical learning algorithm in binary classification. LR estimates the probability that the input under consideration belongs to the target class through the transformation of the linear combination of the attributes into the [0,1] interval via the logistic (or sigmoid) function. LR assumes that there is a linear relationship connecting independent variables with the log-odds of the response variable. LR is computationally efficient and interpretable and can be applied in medical applications such as the diagnosis of thyroid disorders. LR, however, is inefficient with nonlinear or unbalanced data and therefore requires the application of preprocessing or resampling techniques in order to boost the classification efficiency.

3.4.2 Stochastic Gradient Descent (SGD)

Stochastic Gradient Descent (SGD) is the optimization algorithm mainly used to train linear classifiers like logistic regression and the linear support vector machines. Unlike the batch gradient descent, which utilizes the full dataset in order to update the model's parameters, the SGD updates the parameters using the training sample incrementally, so the algorithm is computationally efficient and is scalable applicable with large datasets. The algorithm's efficiency is highly dependent upon the hyperparameter selection like the learning rate, the penalty terms, and the loss function. Though the algorithm is highly dependent on tuning, the algorithm is suited well with sparse, high-dimensional datasets and is efficient when adequately regularized.

3.4.3 Support Vector Machines (SVM)

Support Vector Machines (SVM) are powerful supervised learning algorithms that attempt to create the optimum hyperplane that separates multiple classes with the greatest margin. Through the use of the kernel functions such as the linear, polynomial, or radial basis function (RBF), SVM effectively handles both the linear and nonlinear classification issues. For medical applications, SVM is particularly helpful during the separation of normal from disease conditions despite the limited and highly imbalanced data. Despite the high accuracy of SVM, with large data, this is computationally intensive and requires judicious parameter tweaking in order to work at its optimal best.

3.4.4 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a non-parametric instance-based learner that gives the sample the most frequent classification of the (k) nearest neighbors in the space of features. The Manhattan or Euclidean distance is often applied in order to discover neighbor similarity. KNN is interpretable, easy to code, and efficient in classification in the medical field. KNN is, nevertheless, computationally expensive during the prediction phase and is noisy, irrelevant-feature, and (k)-sensitive. Its efficiency is considerably boosted when used in combination with the use of scaling of features as well as with data balancing schemes.

3.4.5 Adaptive Boosting (AdaBoost)

Several fundamental learning algorithms are typically used in ensemble learning and are naturally produced from training data. AdaBoost is a robust ensemble learning algorithm that builds a strong classifier by iteratively adjusting the weights of weak classifiers to focus on challenging instances in the training data [34], [35]. Its adaptability and ability to create a robust ensemble make it a popular choice in machine-learning applications [36], [37]. With its strong theoretical foundation and extreme simplicity, its variations have been successfully used in some fields. The AdaBoost algorithm was [38].

3.5 PSO Algorithm

The PSO algorithm is an advanced technique that has been improved recently. The PSO method is predicated on particles continuously accelerating toward the optimal solution [39]. The swarm intelligence probability random optimization technique is a typical PSO proposed in 1998 and has garnered the attention of various academics due to its simplicity and ease of implementation [40], [41]. Furthermore, it is utilized to solve large and complex optimization problems that do not require a solid mathematical foundation or prior knowledge, such as machine learning, signal processing, and vehicle scheduling. Each particle in PSO has the basic properties of position and velocity and presents the best solution optimum. During each iteration, the particle is modified based on its inertial weight (w), best position (pbest), and current global best position (gbest).

4 EXPERIMENTAL RESULTS

The experimental analysis was designed to evaluate the influence of different resampling strategies on the classification of thyroid disease using multiple machine learning algorithms. The classifiers included Logistic Regression, AdaBoost, Stochastic Gradient Descent, Support Vector Machine, K-Nearest Neighbor, and the hybrid AdaBoost-PSO. These were tested across three balancing techniques, namely Random Under Sampling, SVM-SMOTE, and K-Means-SMOTE. The performance was measured using accuracy, precision, recall, F1-score, Matthews correlation coefficient, and AUC, in addition to confusion matrix components. These measurements have been widely used in the evaluation part. The distribution of the target variable before and after applying different resampling techniques, including Random Under-Sampling (RUS), SVM-SMOTE, and K-Means-SMOTE. The dataset exhibited a substantial class imbalance with 3,480 positive cases (0) and only 291 negative cases (1), and after applied SVM-SMOTE and K-Means-SMOTE the two classes were balanced each including 2,414 instances resulting in a total of 4,828 observations. The RUS method also achieved balance but with a reduced total of 450 instances (225 per class). This balance corrects the previously existing class disparity, enabling the construction of a more reliable and accurate predictive model.

Table 1 shows the experimental results of the proposed model. The outcomes revealed that the hybrid AdaBoost-PSO model consistently outperformed the other algorithms across all resampling strategies. Under K-Means-SMOTE in particular, it reached an accuracy of 99.73%, precision of 0.9914, recall of 0.9844, F1-score of 0.9878, MCC of 0.9757, and an AUC of 0.984. The corresponding confusion matrix showed 64 true positives, one false positive, two false negatives, and 1065 true negatives, highlighting its capacity to achieve near-perfect classification with only minimal misclassifications. Compared with this performance, KNN demonstrated the weakest results, especially under Random Under Sampling, with an accuracy of only 87.95% and MCC of 0.4657. This outcome suggests that distance-based classifiers have difficulty handling the complexity of imbalanced thyroid data, even after resampling.

Furthermore, Table 2 presents the comparison between the proposed model with other methods. When comparing the proposed method with previous studies, the hybrid K-Means-SMOTE and AdaBoost-PSO model achieved superior results. While earlier works employing methods such as Naïve Bayes, Decision Trees, Logistic Regression, Random Forest, or XGBoost reported accuracies in the range of 83% to 98.9%, the current study exceeded these with an accuracy of 99.73%. This establishes the hybrid framework as an advancement over traditional single-model or ensemble-based approaches. The integration of metaheuristic optimization through PSO with the AdaBoost ensemble significantly enhanced the robustness of predictions, and coupling it with K-Means-SMOTE oversampling effectively addressed the class imbalance issue inherent in the dataset.

Taken together, the results suggest that the proposed hybrid pipeline is an efficient and reliable approach for diagnosing hypothyroidism. It not only surpasses existing techniques in predictive accuracy but also demonstrates the potential for real-world deployment in clinical decision support systems. By providing consistent, high-quality predictions based on patient data, this model could offer a less invasive alternative to traditional laboratory methods for thyroid disease detection.

Table 1: Performance comparison of classifiers across resampling techniques.

Resampling	Classifier	Acc	Pre	Rec	F1	MCC	AUC	TP	FP	FN	TN
RUS	LR	97.09	0.8581	0.9576	0.9004	0.8097	0.958	49	19	3	684
	AdaBoost	99.21	0.9546	0.9868	0.9701	0.9409	0.987	51	5	1	698
	SGD	96.69	0.8778	0.8576	0.8674	0.7351	0.858	38	11	14	692
	SVM	96.82	0.8500	0.9473	0.8914	0.7914	0.947	48	20	4	683
	KNN	87.95	0.6607	0.8373	0.7030	0.4657	0.837	41	80	11	623
	AdaBoost-PSO	99.47	0.9715	0.9883	0.9797	0.9596	0.988	51	3	1	700
SVM-SMOTE	LR	97.26	0.8521	0.9357	0.8886	0.7833	0.936	59	24	7	1042
	AdaBoost	99.20	0.9447	0.9887	0.9655	0.9324	0.989	65	8	1	1058
	SGD	96.73	0.8752	0.8050	0.8359	0.6765	0.805	41	12	25	1054
	SVM	97.08	0.8476	0.9206	0.8800	0.7647	0.921	57	24	9	1042
	KNN	95.23	0.7761	0.8681	0.8141	0.6376	0.868	51	39	15	1027
	AdaBoost-PSO	99.65	0.9908	0.9768	0.9837	0.9675	0.977	63	1	3	1065
K-Means-SMOTE	LR	96.38	0.9029	0.7320	0.7915	0.6115	0.732	31	6	35	1060
	AdaBoost	97.97	0.8731	0.9821	0.9194	0.8482	0.982	65	22	1	1044
	SGD	96.82	0.8732	0.8197	0.8441	0.6908	0.820	43	13	23	1053
	SVM	96.29	0.9202	0.7102	0.7760	0.5945	0.710	28	4	38	1062
	KNN	96.11	0.8715	0.7306	0.7823	0.5854	0.731	31	9	35	1057
	AdaBoost-PSO	99.73	0.9914	0.9844	0.9878	0.9757	0.984	64	1	2	1065

Table 2: Comparison of the proposed model with other techniques of previous studies.

Study	Dataset	Sample Size	Methods	Results
[11]	Not specified	Not specified	KNN, NB	Accuracy: 87.2%, Precision: 86.5%, Recall: 88.0%
[22]	Healthcare lab	Not specified	KNN, SVM, DT, NB	Accuracy: 98.89%
[32]	UCI	3163	LR, KNN, SVM	Accuracy: 97.8%, F-Score: 82.26%, AUC: 93.2 %
[29]	Patients recorded hospital	310	multinomial LR models and ANN	Accuracy: 96.3% AUC: 0.925%
[30]	Combination of hospital records and detailed visited diaries	2784	DT, NB, KNN, RF, EXT, MLP, XGB	Accuracy: 90.7%, Precision: 89.3%, Recall: 91.5%
[23]	Laboratories and hospitals	1250	RF, DT, NB, LR, KNN, MLP, SVM	Accuracy: 98.93%
[25]	UCI	215	XGB,LR, KNN,DT	Accuracy: 98.59 %
[31]	UCI Thyroid Disease Repository	7200	RF, KNN, NB, ANN	Accuracy: 94.8% Specificity: 91%.
[28]	UCI	2,870	DT,ANN,CatBoos, LightGBM,RF, KNN, Extra-Trees	Accuracy: 95.7%
[24]	Clinical and dosimetric data combined with pre-treatment CT	1296 individuals	RF, DT, and LR	Accuracy: 0.83 Sensitivity: 0.88; Specificity: 0.80 Kappa: 0.46
[27]	Clinical data	Not specified	C5.0, CART, CHAID, Quest, LSVM, RT, Modified XGBoost	Accuracy: 94.6%. %
[26]	Patients with hypothyroidism symptoms	500 patients	XGBoost	Accuracy: 0.9774 F1-Score: 0.9594
This work	UCI	3772 patients	LR, AdaBoost, SGD, SVM, KNN, AdaBoost-PSO three balancing methods: RUS, SVM-SMOTE, and KMeans-SMOTE.	K-Means-SMOTE and AdaBoost-PSO: Accuracy 99.73%, Precision 99.44 %, Recall 98.44 %, F score 98.78%, MCC 97.57%, AUC 98.44 %

4 CONCLUSIONS

This study introduced a comprehensive machine learning framework for diagnosing hypothyroidism using the UCI thyroid disease dataset. The framework integrated several crucial stages, beginning with the treatment of missing values through KNN imputation, followed by class balancing using Random Under Sampling, SMOTE-SVM, and K-Means-SMOTE, and the application of established classifiers such as Logistic Regression, SGD, SVM, and KNN. The central contribution was the development of a hybrid AdaBoost-PSO model, which was designed to enhance classification accuracy and robustness. The experimental evaluation demonstrated that although conventional algorithms produced competitive results, the hybrid pipeline combining K-Means-SMOTE with AdaBoost-PSO clearly outperformed all alternatives. The model achieved a testing accuracy of 99.73%, precision of 0.991, recall of 98.40%, F1-score of 98.80%, MCC of 97.60%, and an AUC of 98.40%. These metrics indicate that the hybrid approach not only delivers exceptional predictive accuracy but also minimizes misclassifications, as reflected in the low false positive and false negative rates. The accompanying heatmap analysis further underscored clinical insights, showing elevated TSH levels alongside reduced FTI and T4 values among hypothyroid patients, aligning with known medical patterns.

Overall, the integration of resampling techniques with hybrid optimization provides a powerful and efficient method for thyroid disease diagnosis, offering more accurate, and less costly predictions compared with conventional laboratory testing. The findings confirm the potential of such models to strengthen clinical decision support systems and improve early detection of hypothyroidism. Future research could extend this framework to multiclass thyroid disorder prediction and test its scalability on larger and more diverse datasets to further validate its generalizability.

ACKNOWLEDGMENTS

Faculty Income Generation Grant (FTM1) of Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia.

REFERENCES

[1] S. Shroff, S. Pise, P. Chalekar, and S. S. Panicker, "Thyroid disease diagnosis: A survey," in 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), IEEE, 2015, pp. 1-6, [Online]. Available: https://doi.org/10.1109/ISCO.2015.7282384.

[2] Y. Ma et al., "A study of machine learning models for rapid intraoperative diagnosis of thyroid nodules for clinical practice in China," Cancer Medicine, 2024, [Online]. Available: https://doi.org/10.1002/cam4.6854.

[3] X. Chai, "Diagnosis method of thyroid disease combining knowledge graph and deep learning," IEEE Access, vol. 8, pp. 149787-149795, 2020, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3016676.

[4] S. Talebi et al., "Trace element status and hypothyroidism: A systematic review and meta-analysis," Biological Trace Element Research, vol. 197, pp. 1-14, 2020, [Online]. Available: https://link.springer.com/article/10.1007/s12011-019-01963-5.

[5] Y. Qing et al., "Circulatory trace element variations in Alzheimer's disease: A systematic review and meta-analysis," Environmental Sciences Europe, vol. 36, no. 1, p. 148, 2024, [Online]. Available: http://dx.doi.org/10.1186/s12302-024-00980-z.

[6] U. R. Acharya et al., "Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images," Knowledge-Based Systems, vol. 107, pp. 235-245, 2016, [Online]. Available: https://doi.org/10.1016/j.knosys.2016.06.010.

[7] J. Y. Noh et al., "An evaluation of the efficacy of machine learning in predicting thyrotoxicosis and hypothyroidism: A comparative assessment of biochemical test parameters used in different health checkups," Internal Medicine, 2024, [Online]. Available: https://doi.org/10.2169/internalmedicine.2825-23.

[8] H. Mousa and A. Zoori, "Prevalence of thyroid disorders in Nasiriya City, Iraq," University of Thi-Qar Journal of Science, vol. 10, no. 1, pp. 122-127, 2023, [Online]. Available: http://dx.doi.org/10.32792/utq/utjsci/v10i1.1040.

[9] K. Sindhya, "Effective prediction of hypothyroid using various data mining techniques," EPRA International Journal of Research and Development, vol. 5, no. 2, 2020, [Online]. Available: https://doi.org/10.36713/epra2016.

[10] S. Gothane, "Data mining classification on hypothyroids detection: Association women outnumber men," International Journal of Recent Technology and Engineering, vol. 8, no. 16, pp. 601-604, 2020, [Online]. Available: http://dx.doi.org/10.35940/ijrte.F7261.038620.

[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, pp. 313-319, 2016, [Online]. Available: https://doi.org/10.1007/s40012-016-0100-5.

[12] F. T. Al-Dhief, N. M. A. Latiff, M. M. Baki, N. N. N. A. Malik, N. Sabri, and M. A. A. Albadr, "Voice pathology detection using support vector machine based on different number of voice signals," in 2021 26th IEEE Asia-Pacific Conference on Communications (APCC), Kuala Lumpur, Malaysia, 2021, pp. 1-6, doi: 10.1109/APCC49754.2021.9609830.

[13] M. A. A. Albadr, F. T. Al-Dhief, Man, L. et al., "Online sequential extreme learning machine approach for breast cancer diagnosis," Neural Computing and Applications, vol. 36, pp. 10413-10429, 2024, [Online]. Available: https://doi.org/10.1007/s00521-024-09617-x.

[14] N. A. N. Za'im, F. T. Al-Dhief, M. Azman, M. R. M. Alsemawi, N. M. A. Abdul Latiff, and M. Mat Baki, "The accuracy of an online sequential extreme learning machine in detecting voice pathology using the Malaysian voice pathology database," Journal of Otolaryngology-Head & Neck Surgery, vol. 52, no. 1, 2023, [Online]. Available: https://doi.org/10.1186/s40463-023-00661-6.

[15] H. Lu and S. Uddin, "Unsupervised machine learning for disease prediction: A comparative performance analysis using multiple datasets," Health and Technology, pp. 1-14, 2023, [Online]. Available: https://doi.org/10.1016/j.rineng.2022.100778.

[16] F. Dehkordi, K. Manochehri, and V. Aghazarian, "Internet of Things (IoT) intrusion detection by machine learning (ML): A review," Asia-Pacific Journal of Information Technology & Multimedia, vol. 12, no. 1, 2023, [Online]. Available: https://doi.org/10.17576/apjitm-2023-1201-02.

[17] N. Ananthi et al., "Detecting six different types of thyroid diseases using deep learning approaches," in 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, 2022, pp. 1-8, [Online]. Available: https://doi.org/10.1109/ACCAI53970.2022.9752581.

[18] P. Gupta et al., "Detecting thyroid disease using optimized machine learning model based on differential evolution," International Journal of Computational Intelligence Systems, vol. 17, no. 1, p. 3, 2024, [Online]. Available: https://doi.org/10.1007/s44196-023-00388-2.

[19] N. M. A. A. Latiff et al., "Voice pathology detection using machine learning algorithms based on different voice databases," Results in Engineering, vol. 25, p. 103937, 2025, [Online]. Available: https://doi.org/10.1016/j.rineng.2025.103937.

[20] M. A. A. Albadr, M. Ayob, S. Tiun, R. Z. Homod, F. T. Al-Dhief, and M. H. Mutar, "Parkinson's disease diagnosis by voice data using particle swarm optimization-extreme learning machine approach," Multimedia Tools and Applications, vol. 84, no. 23, pp. 26843-26876, 2025, [Online]. Available: https://doi.org/10.1007/s11042-024-20108-y.

[21] F. T. Al-Dhief et al., "Voice pathology detection and classification by adopting online sequential extreme learning machine," IEEE Access, vol. 9, pp. 77293-77306, 2021, , doi: 10.1109/ACCESS.2021.

[22] D. Umar Sidiq, S. M. Aaqib, and R. A. Khan, "Diagnosis of various thyroid ailments using data mining classification techniques," International Journal of Scientific Research in Computer Science and Information Technology, vol. 5, pp. 131-136, 2019, [Online]. Available: https://doi.org/10.32628/CSEIT195119.

[23] E. Sonuç, "Thyroid disease classification using machine learning algorithms," in Journal of Physics: Conference Series, vol. 1963, no. 1, IOP Publishing, 2021, p. 012140, [Online]. Available: http://dx.doi.org/10.1088/1742-6596/1963/1/012140.

[24] S. R. Rad, Z. H. Mohammadi, M. J. Zadeh, M. A. Mosleh-Shirazi, and T. Dehesh, "Identification of important symptoms and diagnostic hypothyroidism patients using machine learning algorithms," Annals of Medicine and Surgery, vol. 86, no. 6, pp. 3233-3241, 2024, [Online]. Available: http://dx.doi.org/10.1097/MS9.0000000000002068.

[25] S. Sankar, A. Potti, G. N. Chandrika, and S. Ramasubbareddy, "Thyroid disease prediction using XGBoost algorithms," Journal of Mobile Multimedia, vol. 18, no. 3, pp. 1-18, 2022, [Online]. Available: http://dx.doi.org/10.1088/1742-6596/1963/1/012140.

[26] T. A. Alawiyah, T. Wibisono, and Y. S. Mulyani, "The prediction of thyroid cancer recurrence with the XGBoost method: The clinicopathological feature-based approach," Journal of Computer Networks, Architecture and High Performance Computing, vol. 6, no. 3, pp. 1035-1045, 2024, [Online]. Available: https://doi.org/10.47709/cnahpc.v6i3.4101.

[27] S. Dalal et al., "Enhancing thyroid disease prediction with improved XGBoost model and bias management techniques," Multimedia Tools and Applications, 2024, [Online]. Available: https://doi.org/10.1007/s11042-024-19713-8.

[28] S. S. Islam, M. 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, 2022, [Online]. Available: http://dx.doi.org/10.7717/peerj-cs.898.

[29] S. Borzouei, H. Mahjub, N. A. Sajadi, and M. Farhadian, "Diagnosing thyroid disorders: Comparison of logistic regression and neural network models," Journal of Family Medicine and Primary Care, vol. 9, no. 3, pp. 1470-1476, 2020, [Online]. Available: http://dx.doi.org/10.4103/jfmpc.jfmpc_910_19.

[30] L. Aversano et al., "Thyroid disease treatment prediction with machine learning approaches," Procedia Computer Science, vol. 192, pp. 1031-1040, 2021, [Online]. Available: https://doi.org/10.1016/j.procs.2021.08.106.

[31] 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, 2022, [Online]. Available: https://doi.org/10.1155/2022/9809932.

[32] G. Akgül, A. A. Çelik, Z. E. Aydin, and Z. K. Öztürk, "Hipotiroidi hastalığı teşhisinde sınıflandırma algoritmalarının kullanımı," Bilişim Teknolojileri Dergisi, vol. 13, no. 3, pp. 255-268, 2020, [Online]. Available: https://doi.org/10.17671/gazibtd.710728.

[33] Y. H. Shakir, "Thyroid disease dataset," [Online]. Available: https://www.kaggle.com/datasets/yasserhessein/thyroid-disease-data-set, [Accessed: 25 May 2025].

[34] Y. Sun, J. Zhang, and Y. Zhang, "Adaboost algorithm combined multiple random forest models (Adaboost-RF) is employed for fluid prediction using well logging data," Physics of Fluids, vol. 36, no. 1, 2024, [Online]. Available: http://dx.doi.org/10.1063/5.0179422.

[35] J. C. Ho, M. Sotoodeh, W. Zhang, R. L. Simpson, and V. S. Hertzberg, "An AdaBoost-based algorithm to detect hospital-acquired pressure injury in the presence of conflicting annotations," Computers in Biology and Medicine, vol. 168, p. 107754, 2024, [Online]. Available: https://doi.org/10.1016/j.compbiomed.2023.107754.

[36] A. Mahkamov, T. Jumayev, D. Tuhtanazarov, and A. Dadamuxamedov, "Using AdaBoost to improve the performance of simple classifiers," in Artificial Intelligence, Blockchain, Computing and Security, CRC Press, 2024, pp. 755-760.

[37] S. Amaouche, C. Hazman, A. Guezzaz, S. Benkirane, and M. Azrour, "Intrusion detection framework using AdaBoost algorithm and Chi-squared technique," in Blockchain and Machine Learning for IoT Security, Chapman and Hall/CRC, 2024, pp. 92-111.

[38] G. M. Salama, A. Mohamed, and M. K. Abd-Ellah, "COVID-19 classification based on a deep learning and machine learning fusion technique using chest CT images," Neural Computing and Applications, pp. 1-19, 2023, [Online]. Available: https://doi.org/10.1007/s00521-023-09346-7.

[39] S. K. Gharghan, R. Nordin, M. Ismail, and J. Abd Ali, "Accurate wireless sensor localization technique based on hybrid PSO-ANN algorithm for indoor and outdoor track cycling," IEEE Sensors Journal, vol. 16, no. 2, pp. 529-541, 2015, [Online]. Available: https://doi.org/10.1109/JSEN.2015.2483745.

[40] Y. A. Ali, E. M. Awwad, M. Al-Razgan, and A. Maarouf, “Hyperparameter search for machine learning algorithms for optimizing the computational complexity,” Processes, vol. 11, no. 2, p. 349, 2023, [Online]. Available: https://doi.org/10.3390/pr11020349.

[41] H. M. Alaidi, Z. A. Ramadhan, J. S. Alrubaye, H. T. S. Alrikabi, H. A. Mutar, and I. Svyd, “AI-based monkeypox detection model using Raspberry Pi 5 AI Kit,” Sustainable Engineering and Innovation, vol. 7, no. 1, pp. 1-14, 2025.

[42] H. A. Hashim, S. L. Mohammed, and S. K. Gharghan, “Path loss model-based PSO for accurate distance estimation in indoor environments,” Journal of Communications, vol. 13, no. 12, pp. 712-722, 2018, [Online]. Available: https://doi.org/10.1007/s44196-023-00388-2.