Thyroid Disease Multi-class Classification based on Optimized Gradient Boosting Model

Mona Alnaggar[a], Mohamed Handosa[b], T. Medhat[c] and M. Z. Rashad[d]

[a] Department of Robotics and Intelligent Machines, Faculty of Artificial Intelligence, Kafrelsheikh University, Egypt.
[b] Department of Computer science, Faculty of Computers and Information, Mansoura University, Egypt.
[c] Department of Electrical Engineering, Faculty of Engineering, Kafrelsheikh University, Egypt.
[d] Department of Computer science, Faculty of Computers and Information, Mansoura University, Egypt.
[*]Corresponding Author: Mona Alnaggar, E-mail: mona.alnaggar@ai.kfs.edu.eg

Abstract: Human healthcare is one of the most important issues in society to ensure that patients receive the care they require as quickly as possible. One of the disorders that affects the global population and is becoming more prevalent is thyroid disease. Medical information systems are crucial in their capacity to diagnose thyroid disease. Artificial intelligence has recently offered fresh approaches to the existing clinical treatment issues and has demonstrated promising results for individualized diagnosis and therapy planning. Hence, this paper proposes an optimized multi-class classification model, which depends on XGBoost to classify patients with different types of thyroid disease. The main contribution is to (i) propose a Multiclass-Classification for the purpose of diagnosing three different thyroid diseases, (ii) raise the row dataset's feature selection accuracy for classification. (iii) utilize the highly selective XGBoost algorithm for the chosen characteristics, (iv) show that The XGBoost has the best performance and recall, making it the top choice for data analysis in terms of classifying thyroid disease, and (v) improve upon findings from earlier studies by doing the proposed study. XGBoost is trained and tested using UCI machine learning repository dataset of thyroid disease. In addition to build the model with the optimized hyperparameters to achieve and compare the gained results aiming to get the best score of accuracy. From the results, it is shown that the optimized XGBoost achieved 99% accuracy as a win over performing compared with the state of arts models.

Keywords: Human healthcare; Thyroid disease; Artificial intelligence; multi-class classification; XGBoost algorithm; hyperparameters optimization.

1. Introduction

1.1 Overview

Everyone, including males, women, children, adolescents, and the elderly, can get thyroid illness. Blood tests are used to diagnose thyroid disease, but these tests are notoriously challenging to interpret since so much information is needed to predict the findings. The thyroid gland produces a crucial hormone that helps the body function more healthily. Thyroid disease [1] is the term for a situation when excessive or insufficient thyroid hormone production occurs below the threshold. For the thyroid to operate normally, thyroid hormone levels must be normal. Menarche, pubertal growth and development, menstrual physiology, fertility and foetal development, postpartum period, reproductive physiology, and postmenopausal status are all impacted by a woman's thyroid function. All medical professionals have seen irregular menstrual cycles in women with hypothyroidism and hyperthyroidism. The most frequent menstrual complaint was menorrhagia [2]. Thyroid hormones affect body homeostasis [3] and metabolism [4] in a variety of ways, and their concentration may raise the risk of various disorders, mostly cardiovascular in nature [5]. So, the key need in the medical field is a reliable method of diagnosing thyroid illness. Different thyroid conditions can cause weight loss, weight gain, anxiety, irritability, eyesight issues, intolerance to cold or hot temperatures, and other symptoms that have a significant influence on the

human body [6]. Since the 1990s, there have been an increasing number of thyroid illness cases, and in recent years, thyroid cancer has become the malignancy with the fastest rate of increase [7].

While thyroid dysfunction can have a negative impact on bone structures, thyroid hormones are necessary for proper skeletal growth [8] and bone metabolism in adulthood. Congenital hypothyroidism (CH) is brought on by an anomaly in the thyroid gland's development. Hypothyroidism is the most prevalent thyroid condition. Hypothyroidism [9] is a disorder in which the thyroid gland is underactive or produces too little thyroid hormone. Hypo- denotes deficient or under(active).

1.2 Problem statement

Machine learning (ML) algorithms are being used more often in the medical sector. As the name suggests, Software can quickly learn on data using ML to provide effective representations in a semi-automatic manner. Many different formats of data have been applied to ML techniques for the goal of thyroid disease identification and classification. The number of articles published in recent years discussing the use of ML to diagnose thyroid disease has increased for enhancing the medical decision support systems. The ML usage in diagnosing and evaluating Thyroid disease was the subject of earlier investigations. ML algorithms, computer-based statistical techniques, may be trained to search through massive amounts of data for recurrent patterns. Doctors can identify patients based on many characteristics simultaneously by using ML algorithms.

1.3 Objectives

Artificial intelligence approaches [10-14] have created previously unimaginable possibilities and changing innovations in several sectors. Deep learning-based computer assisted diagnostic (CAD) systems in clinical medicine [15] have been found to produce diagnostic accuracy and efficiency comparable to, and occasionally even better to, that of seasoned doctors. so modern societies rely heavily on data and applying Artificial intelligence techniques and frameworks [16-18] which is anticipated to raise the standard, accessibility, affordability, safety, and equity of healthcare. Systems for thyroid computer-aided diagnosis (CAD) [19] have been created to help radiologists do diagnoses more effectively and reliably. These CAD systems' performance is frequently assessed using various datasets, rendering it incomparable. In recent years, medical information systems like the

Internet of Medical Things (IoMT) [20] have attracted a lot of interest. When looking for a certain kind of anomaly, X-ray and MRI scans are crucial repository of information. Machine learning methods[21] might also be used to mine reports based on the photos and lab test results. Medical information systems are crucial in their capacity to diagnose thyroid disorders. To save time and increase the efficacy of treatment, classification is used to identify thyroid disease. Finding the classification strategy that is most useful for detecting thyroid disease is a problem, but research into the pertinent literature shows that a variety of different classification strategies have been used to get greater results. Choosing the correct classification technique might be challenging because it needs to be used with a local dataset.

1.4 Paper contribution

Artificial intelligence models can help healthcare providers to diagnose and predict many different diseases and classify them according to their types and levels. Hence, the main contribution in this study is to:

- Propose a Multiclass-Classification model for the purpose of diagnosing three different classes of thyroid disease.
- Raise the row dataset's feature selection accuracy for classification.
- Utilize the highly selective XGBoost algorithm for the chosen characteristics.
- Show that The XGBoost has the best performance and recall, making it the top choice for data analysis in terms of classifying thyroid disease.
- Building the XGBoost with the hyperparameters optimization to get more accurate results.
- Improve upon findings from earlier studies by doing the suggested study.

1.5 Paper organization

The remainder of this paper is structured as follows: Thyroid disease classification literature review is illustrated in Section 2 which includes a presentation of various relevant publications. Section 3 presents the materials and methods with the steps of the proposed model in detail. The experiments and trials are shown in section 4. The achieved results are shown in section 5. Finally, the conclusion and the future work will be discussed in section 6.

2. Related work

This section gives a brief overview of prior studies on thyroid disorders and classification methods in this section. This study used semantic reports and examination data to implement an artificial neural network (ANN) in IoMT systems with the primary objective of improving thyroid problem diagnosis accuracy[20]. A collection of multiple multilayer perceptron (MMLP) neural networks with the capability to back-propagate mistakes is described to improve generalization and prevent over-fitting of ANN during training. The findings considerably improved the categorization of thyroid disease. Using data from the UCI repository, multiclass support vector machine classifiers are trained to develop a model and categorize the thyroid into four different conditions: hypothyroidism, hyperthyroidism, Euthyroid sickness, and Euthyroid. The suggested model's findings show that SVM is a well-known method for solving binary classification issues in the machine learning field. SVM [6] provides excellent accuracy in a high dimensional environment.

The numerical data were preprocessed into nominal data using SPSS Statistics. Age and sex are the two main determinants, according to an analysis of the Top-7 association rules produced by the two algorithms. The findings also indicated that the largest prevalence of thyroid disease recurrence occurs between the ages of 50 and 60. Men are more likely than women to be thyroid disease-free when the gender factor is considered. The experimental findings demonstrated that the AR-ANN performs better, demonstrating the viability and usefulness of the association rule mining method and the BP neural network in the detection of thyroid disease [22]. In this study [23], algorithms such as KNN, Mean/Median, and PCA were tested to pre-process thyroid datasets by removing duplicate and unnecessary variables. The hybrid impute algorithm appeared to be better than the three imputation methods. To create an accurate prediction model for the Thyroid dataset following pre-processing, effective feature selection methods employed, along with a variety of classification algorithms. In this work [24], KNN was utilized to predict thyroid illness using a variety of distance functions. The chi-square approach and L1-based featured selection methods were used to choose the best features. Euclidean and cosine distance functions were applied by the KNN. According to the findings, KNN showed promise in thyroid detection.

The Improved Ensemble Classification Method of Thyroid Disease Based on Random Forest is the authors' suggested piece of work. In this study, a novel random forest-based classification approach for thyroid illness was developed. They employed an ensemble classifier technique based on random forests, which has a 96.16% accuracy rate [25]. Using feature selection methods like Univariate Selection (UFS), Recursive Feature Elimination (RFE), and Tree-Based Feature Selection, the authors' work proposed a model for diagnosing thyroid disease. The classification algorithms used were Naive Bayes, Support Vector Machine (SVM), and Random Forest. They have the maximum accuracy of 92.92% when SVM and RFE are combined. The UCI Machine Learning Repository provided the data set that was used for this study [26].

This study [27] proposed a hybrid optimization algorithm-based feature selection design for thyroid disease classifier with rough type-2 fuzzy support vector machine to improve classification accuracy. To choose the top-n features, a hybrid optimization method is implemented, which combines the firefly algorithm (FA) and butterfly optimization algorithm (BOA). It is assessed using several important metrics, including specificity, accuracy, and sensitivity. This paper [28] includes an experimental investigation for several machine learning techniques, including statistical classifiers to predict thyroid risk, tree-based classifiers, and neural networking classifiers. These algorithms are used to forecast the thyroid risk on a specific dataset, Sick-euthyroid. Precision, recall, F1 score, and accuracy are calculated to evaluate the algorithms used. The ANN classifier performs best among all methods, with an accuracy of 0.9587. With accuracy scores of 0.9538 and 0.9533, the CatBoost and XGBoost classifiers came in second and third, respectively, in this proposed study.

Table 1: Summarized of some previous studies for thyroid diseases classification.

Reference	Year	Dataset	Methodology	accuracy
[6]	2020	thyroid disease dataset from UCI	Multiclass SVM	83%
[29]	2018	KEEL repository thyroid disease dataset	KNN (Euclidean)	96.6%
[30]	2014	thyroid disease dataset from UCI	KNN (Euclidean)	96%
[31]	2018	UCI repository thyroid disease dataset	KNN (Minkowski)	98%
[32]	2019	Bangalore hospital, India	KNN (Euclidean)	97%
[26]	2021	UCI Machine Learning Repository	SVM and RFE are combined	92.92%
[28]	2022	Sick-euthyroid dataset 3,162 records collected is UCI machine learning repository	ANN, CatBoost, XGBoost	95.87%, 95.38 %, 95.33%

3. Materials and Methods

One of the disorders that affects the global population and is becoming more prevalent is thyroid disease. Thyroid dysfunction can have a negative impact on bone structures, thyroid hormones are necessary for proper skeletal growth [8] and bone metabolism in adulthood. The creation of ML models that can help

with diseases prediction can be crucial for making early predictions. In this work, we identify thyroid diseases patients using a shared dataset and XGBoost algorithm. Hyper-parameter optimization permits fine-tuning prior to evaluating a model's performance. To find the ideal values, samples of hyper-parameter values are generated using the XGBoost optimization method. Each hyper-parameter configuration's test set is used to evaluate classification algorithms, and an optimization training set is used to impart knowledge. The configuration of parameter that offers the maximum all-around accuracy is the optimum one. The proposed model is then trained using the optimums from the first training set, and its accuracy is evaluated by the classifier on the test set. Figure 1 will preview the proposed model steps for thyroid disease multi-class classification with the hyperparameters optimization.

The proposed model depends on using XGBoost algorithm which is one of the most potent ML algorithms [33]. It is known for automatically tweaking hundreds of learnable parameters to identify patterns and regularities in the data in addition to its ability for providing a large range of hyperparameters. As a result, it will have huger hyperparameters and additional design considerations. These are fixed parameters that were manually supplied to the algorithm during training. XGBoost is employed with hyperparameter optimization to categorize patients with various thyroid-related disease based on their age, sex, and medical data, including test results for blood levels of thyroid hormone [34]. In figure 2 the proposed model architecture will be displayed.

A collection of variables called hyperparameters is utilized in training and testing to aid in learning. Examples of hyperparameters include learning rate, iterations per batch, batch size, hidden layers, regularization, and activation functions. These parameters might be a continuous variable, a categorical variable, or an integer with values spanning from the lower to upper bounds. As hyperparameters remain constant throughout training, the model's accuracy is increased while training time and memory requirements are concurrently decreased. Different models employ a variety of hyperparameters depending on the issue description. There are no hyperparameters that are optimal for all models.

The four kinds of XGBoost hyperparameters are: (i)general parameters, (ii)booster parameters, (iii)learning task parameters, and (iv)command line

parameters. The XGBoost model's general operation is guided by these parameters. It can decide which attributes are most crucial for training and utilizes them to create its tree. The proposed hyperparameter tuning steps of the used algorithm are shown in algorithm 1 and figure 2.
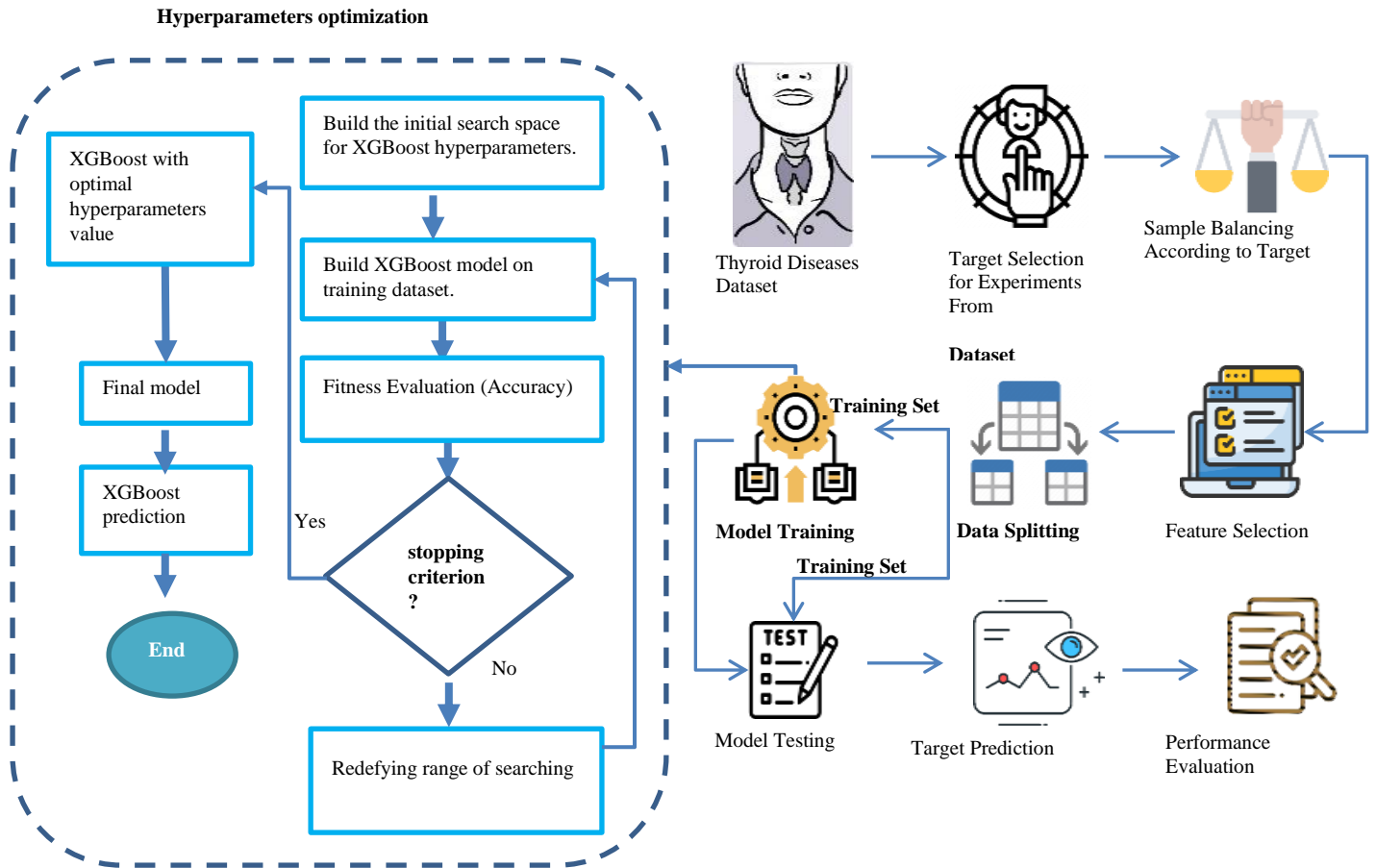

Figure 1: The proposed model for Thyroid disease classification

Algorithm 1: proposed algorithm for hyperparameter tuning

Input: *initial values for XGBoost Algorithm (learning_rate, reg_lambda, early stopping, clsample_bytree) and training dataset*

Output: *The highest accuracy and optimal values achieved by the optimized hyperparameter of XGBoost Algorithm*

Steps:

1- *Train all dataset with initial values for (learning_rate, reg_lambda, early stopping, clsample_bytree)*
2- *Test the algorithm.*
3- *Calculate accuracy (ACC)*

$$Acc = \frac{correctly\ classified}{all\ classified}$$

   optAcc = Acc, where optAcc is the optimal accuracy
4- *For each subset of datasets:*
   Create new values for (learning_rate, reg_lambda, early stopping, clsample_bytree)
   Train the algorithm with new values for (learning_rate, reg_lambda, early stopping, clsample_bytree)
   Calculate the new accuracy (acc_new)
   If acc_new > ACC
      Then optAcc = acc_new

   End if

   Save the new values for (learning_rate, reg_lambda, early stopping, clsample_bytree)

5. *End for*

The model starts with building and running the base model with default hyperparameters. Then it was run with the optimized hyperparameters to achieve and compare the gained results to finally get the best score of accuracy.

Unbalanced datasets need to be handled using an appropriate model, therefore XGBoost would be a suitable choice. XGBoost is an optimizing gradient tree boosting that builds decision trees in a sequential fashion.
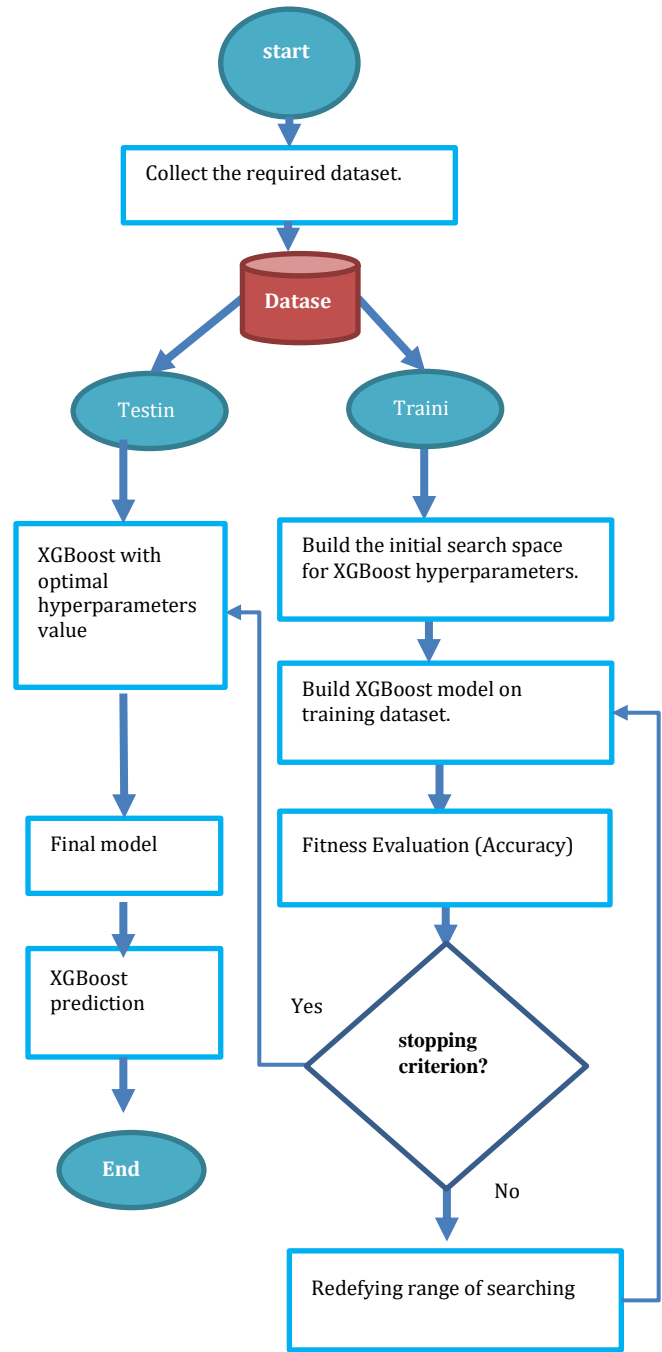

Figure 2 The proposed steps to build the XGBoost classification model.

4. Experimental results

In this section, the dataset description, and the specification of the used device.

4.1 The used dataset

The suggested model relies on the use of a thyroid disease dataset [35] that the Garvan institute

acquired from the UCI machine learning repository. Most earlier studies on thyroid disease were reported on the UCI machine learning repository as their shared dataset. It contains data of 7200 patients with 21 input variables. It includes three categories for classifying a patient's thyroid gland function, placing them in either the normal functioning, hyperthyroidism, or hypothyroidism categories. 166 of the 7200 individuals have normal thyroid function, 368 have hyperthyroidism, and 6666, as mentioned before, have hypothyroidism. Table 2 shows the original thyroid disease dataset input and output properties. Table 3 shows the description of dataset attributes after the preparation stage. Table 4 will also show the distributions of numeric variables.

Table 2: The Thyroid Dataset

| No. of Instances	7200
No. of Input Attributes	21
No. of Output Classes	3
Class Label	1-Normal 2-Hyperthyroidism 3-Hypothyroidism

Table 3: The description of dataset attributes

Attribute	Domain
Age	[0.01, 0.97]
Sex	[0,1]
On thyroxine	[0,1]
Query On thyroxine	[0,1]
On antithyroid medication	[0,1]
sick	[0,1]
pregnant	[0,1]
Thyroid surgery	[0,1]
I131_treatment	[0,1]
Query hypothyroid	[0,1]
Query hyperthyroid	[0,1]
lithium	[0,1]
Goitre	[0,1]
Tumor	[0,1]
Hypopituitary	[0,1]
Psych	[0,1]
Thyroid stimulating hormone (TSH)	[0.0, 0.53]
Triiodothyronine (T3)	[0.0005, 0.18]
Thyroid transcription factor (TT4)	[0.0020, 0.6]
Thyroxine utilization (T4U)	[0.017, 0.233]
Free thyroxine index (FTI)	[0.0020, 0.642]
Class normal	[0,1]
Class hyperthyroid	[0,1]
Class hypothyroid	[0,1]

Table 4: The statistical analysis of the numerical variables

	age	TSH	T3	TT4	T4U	FTI	TBG
count	7546.00	6824.00	5337.00	7192.00	6870.00	6877.00	259.00
mean	78.013	5.42	2.02	105.20	0.97	110.57	22.95
std	1305.26	26.08	0.81	32.60	0.16	36.60	6.09
min	1.00	0.005	0.05	2.00	0.19	1.40	0.10
25%	37.00	0.57	1.60	87.00	0.87	93.00	20.00
50%	55.00	1.40	2.00	103.00	0.96	108.00	23.00
75%	67.00	2.70	2.30	121.00	1.06	125.00	27.00
max	65526.00	530.00	18.00	430.00	2.12	839.00	45.00

ML methods have been utilized in this work to predict thyroid disease. To begin with, the experiment looks at the distribution of hormone levels in blood for each of the target classes using Exploratory Data Analysis (EDA). EDA gives us a better insight of the prospective predictive value of each of the dataset features. As shown in figure 3 There are some features which are considered more effective than others.
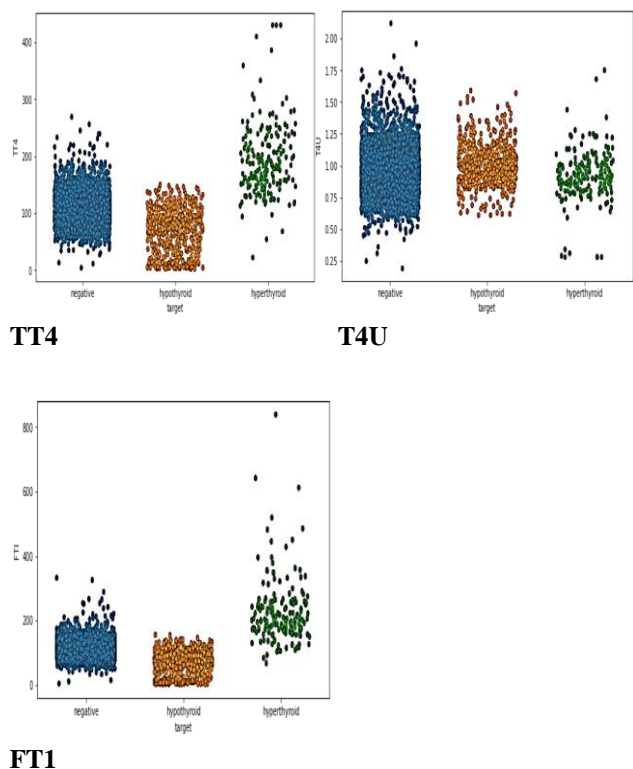

TSH   **T3**

TT4   **T4U**

FT1
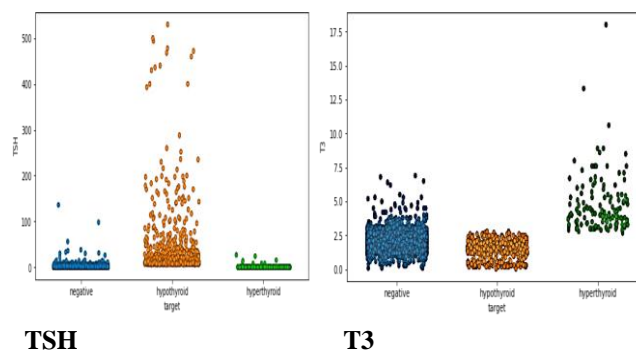
Figure 3: Numerical attributes vs target

It can be hypothesized that FTI, T3, and TT4 will be useful feature additives to diagnose thyroid disease. TSH appears to have potential as well, but before taking any action, we must deal with the outliers for the "target" hypo and conduct further research on the attribute distribution. This information is consistent with what we learned about hormone level testing from our preliminary research. it was obvious how unbalanced the dataset is, with so many negative "targets" compared to hypothyroid or hyperthyroid. The distribution analysis of the characteristics is shown in figure 4. It is a graphical representation of a dataset's distribution that displays the frequency of data points within various intervals. For displaying the distribution of numerical data, we proceed by making a pairplot tool of the numerical variables and looking for any clusters that may be emerging between the variables. We looked at the histogram, a common graphing technique used to combine discrete and continuous data collected on an interval scale, for the features shown in this figure. It is widely employed to include both continuous and discrete data that was captured on an interval scale. The distributions of each numerical variable in relation to one another may be seen in the pairplot's diagonals. With so many negative "targets" compared to

hypothyroid or hyperthyroid, the dataset's imbalance is clear. It was shown that some hormone tests generate beautiful clusters when compared to others. This is promising since it shows how well they distinguish between each of our target classes.

There is severe target class imbalance. It is normal for this type of data. Hence, XGBoost will be a good option for building the proposed model. To investigate the feature correlation between all our attributes between one another. An open-source package dython is used as a multi-variate correlation exploration tool. This is not necessarily the most thorough way to explore relationships between variables of different types within the dataset, but it is straight-forward and easy to use. It will help to get an idea of variable correlations. The correlation heatmap analysis for the features of the used dataset will be shown in figure 5.
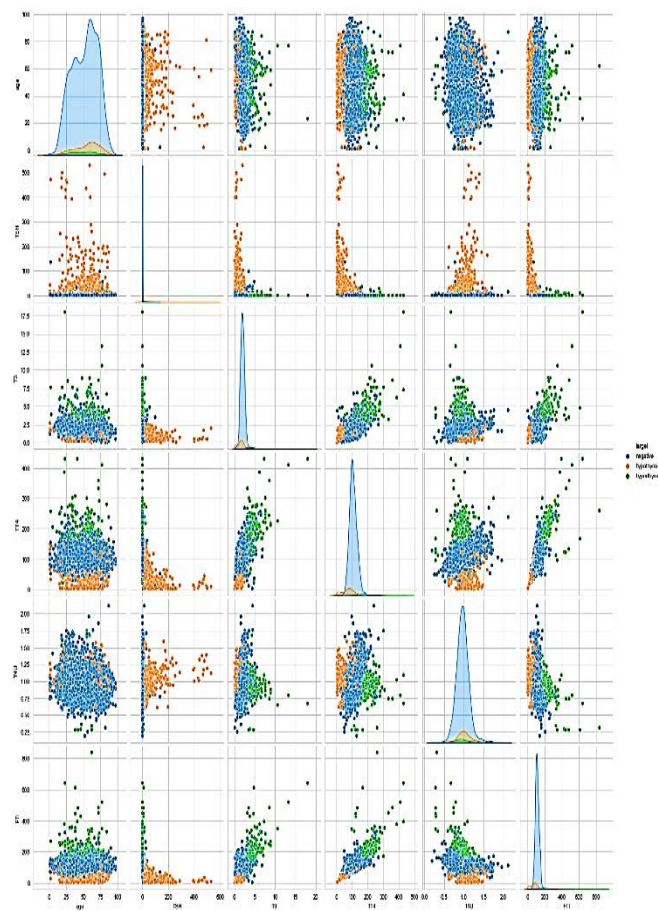

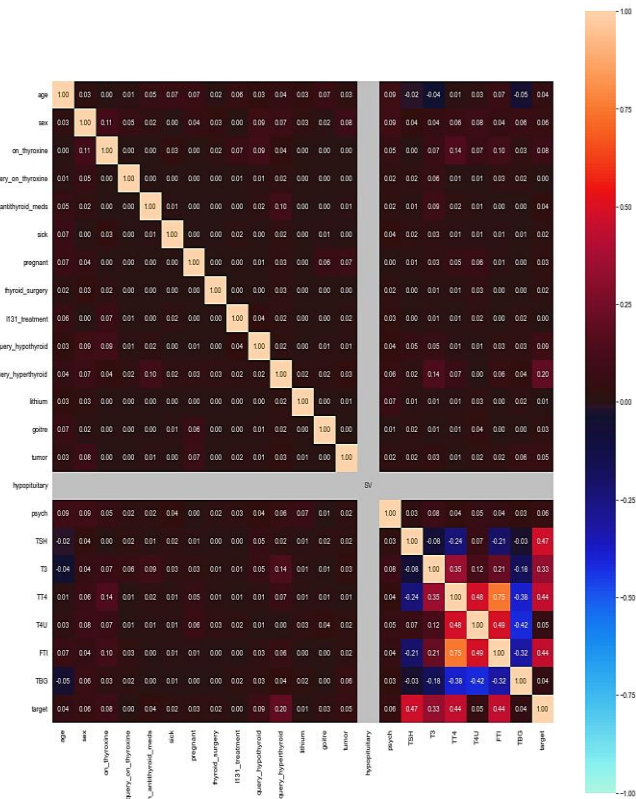**Figure 4:** Distribution analysis of the feature

Figure 5: The heatmap correlation between features of the dataset

A box plot for the feature distribution analysis is shown in Figure 4. For displaying the distribution of numerical data, it is a helpful tool. It may be used to show the distribution of each feature when examining the distribution of features in a dataset. The most important aspects of the applicable thyroid dataset are shown in this figure by the enrolled features. Box plots divide the data into sections, each of which contains around 25% of the total data. Because they provide a visual summary of the data, box plots are useful because they make it simple to identify mean values, dataset dispersion, and skewness. Figure 6 will preview the distributed analysis of some features using the box plot.
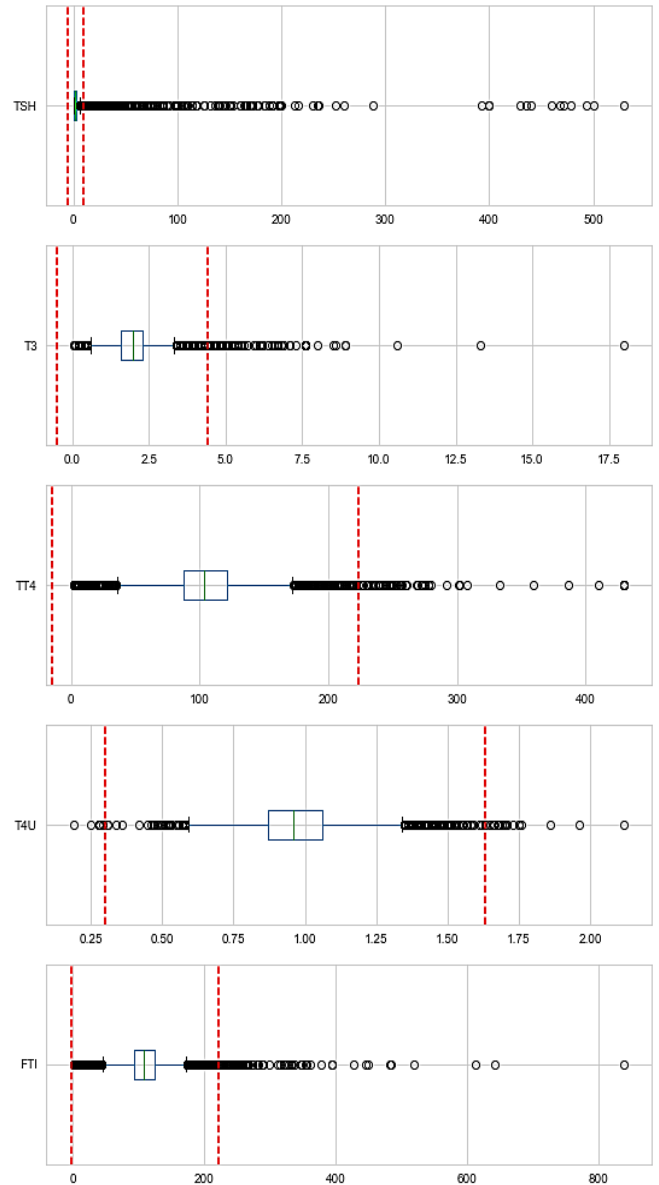

Figure 6: The distribution analysis of the feature using the box plot.

4.2 The used device specifications

To support the implementation of our suggested modal, our experiments were implemented in Python notebook version (6.4.8) and several open-source libraries. It was carried out using laptop with a 2.30GHz, 11th-generation Intel Core (TM) i7-11800H processor, and 16.0 GB (15.8 GB usable).

4.3 Evaluation matrix

This section presents the experiments of the proposed model and evaluates its performance using

AUC. By epochs it plots AUC, loss, and confusion matrix of the four classes. Because of the imbalance of these four classes in the training phase, the AUC has been chosen instead of accuracy according to the providing equations [36]. The suggested model's assessment process was specified in equations (1) through (5). Together with accuracy, additional model performance metrics such as precision, recall, and sensitivity were also evaluated. The true positive, true negative, false positive, and false negative values must first be established to assess the model's performance. True positive and negative numbers are what the model really predicted to be positive and negative, respectively. On the other side, false positives and negatives happen when the model mistakenly classifies positive and negative predictions as positive and negative, respectively.

i. **Accuracy:**

$$Accuracy = \frac{True\ Negative + True\ Positive}{Total\ Data} \quad (1)$$

ii. **Precision:**

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (2)$$

iii. **Recall:**

$$Recall\ (Sensitivity) = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

iv. **Specificity:**

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (4)$$

v. **F1 score**

$$F1 = \frac{2x\ Precision\ x\ recall}{Precision + recall} \quad (5)$$

4.4 The feature importance

The model starts with building and running the base model with default hyperparameters. Then the model was run with the optimized hyperparameters to achieve and compare the gained results to finally get the best score of accuracy. After execution with the optimized hyperparameters, the features importance of best XGBoost model can be shown in figure 7.
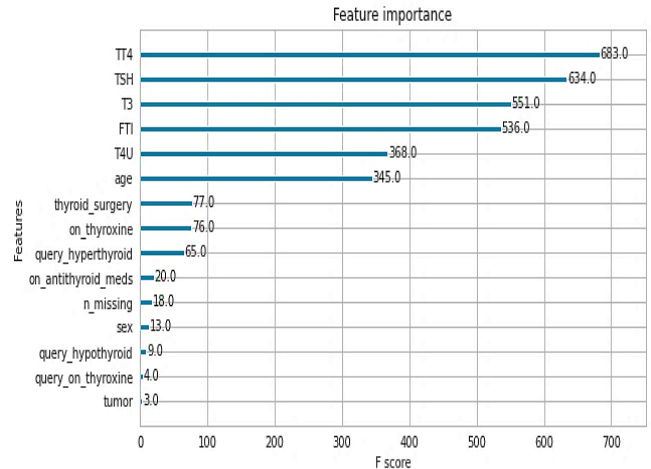

Figure 7: The features importance of XGBoost

5. Results and discussions

To get the model that fits the data, the dataset is trained using a variety of parameters. It is obvious that to get the best results from a ML or DL model, it is preferred to tune the hyperparameters for the model. In our proposed model, the trials demonstrated that XGBoost with the hyperparameter optimization is more accurate compared with the same algorithms without using the hyperparameter optimization among the used classifiers. The performance of the models was evaluated using the dataset both before and after the hyperparameter optimization technique. Figure 8 shows the difference in the performance of XGBoost using various hyperparameter values.
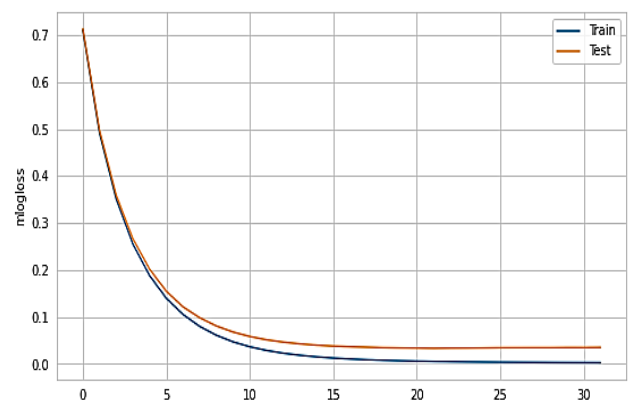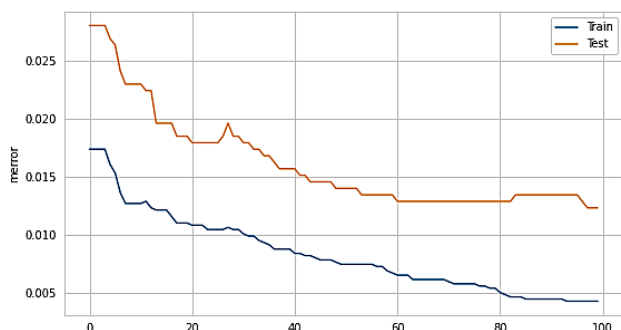

Figure 8. a) The base model using XGBoost mlgloss
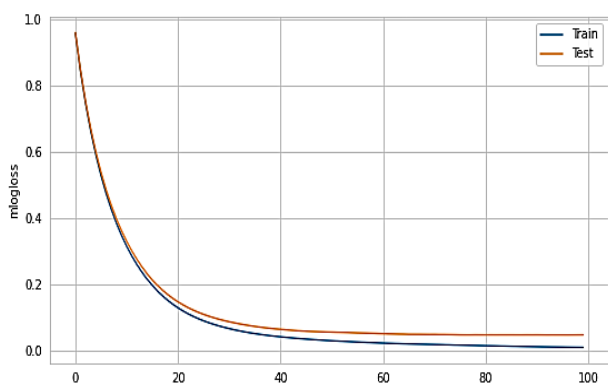
Figure 8. b) The base model using XGBoost merror

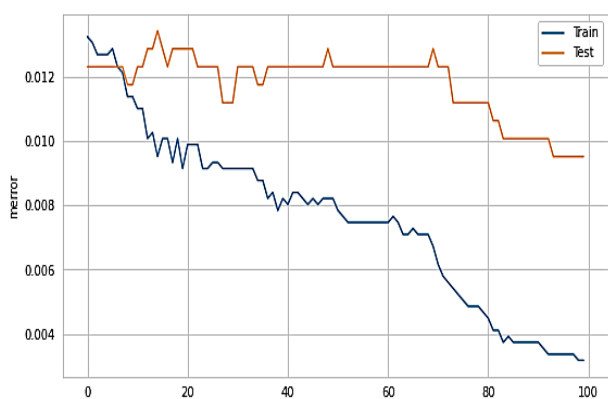
Figure 8. c) Optimized XGBoost mlgloss

Figure 8. d) Optimized XGBoost merror

Figure 8: The results using the XGBoost depending on optimized hyperparameters

The finding results of the proposed model can be compared with previous studies using the same dataset in table 5.

Table 5 The comparison between the proposed models and state of the art algorithms

Reference	Year	Dataset	Methodology	accuracy
[30]	2014	thyroid disease dataset from UCI	KNN (Euclidean)	96%
[29]	2018	KEEL repository thyroid disease dataset	KNN (Euclidean)	96.6%
[31]	2018	UCI repository thyroid disease dataset	KNN (Minkowski)	98%
[32]	2019	Bangalore hospital, India	KNN (Euclidean)	97%
[6]	2020	thyroid disease dataset from UCI	Multiclass SVM	83%
[26]	2021	UCI Machine Learning Repository	SVM and RFE are combined	92.92%
[28]	2022	Sick-euthyroid dataset 3,162 records collected is UCI machine learning repository	ANN, CatBoost, XGBoost	95.87%, 95.38 %, 95.33%
Proposed	2023	UCI Machine Learning Repository	XGBoost with optimized hyperparameter	99%

From table 5, it is shown that the optimized XGBoost achieved a satisfied performance compared with the state-of-the-art models.

6. Conclusion

One of the disorders that affects the global population and is becoming more prevalent is thyroid disease. Thyroid dysfunction can have a negative impact on bone structures. Thyroid hormones are necessary for proper skeletal growth and bone metabolism in adulthood. This paper presented an optimized multiclass classification algorithm to classify the Thyroid disease. The main contribution is to (i) propose a Multiclass-Classification for the purpose of diagnosing three different thyroid diseases, (ii) raise the row dataset's feature selection accuracy for classification. (iii) utilize the highly selective XGBoost algorithm for the chosen characteristics, (iv) show that The XGBoost has the best performance and recall, making it the top choice for data analysis in terms of classifying thyroid disease, and (v) improve upon findings from earlier studies by doing the suggested study. XGBoost is trained and tested using UCI. The trials demonstrated that XGBoost with the hyperparameter optimization is more accurate compared with the same algorithms without using the hyperparameter optimization among the used classifiers. From the results, it is shown that the optimized XGBoost achieved a satisfied performance compared with the state of the arts models. It achieved 99% accuracy. The dataset used to categorize Thyroid's disease may be expanded in future research to include more varied and representative sample populations. Moreover, using more sophisticated machine learning methods, like deep learning, might produce even more accurate and effective outcomes. Using various feature selection techniques to determine the most crucial features for the classification assignment might be another area for development. The study's potential future directions include generalization to a broader population and eventual integration into a bigger healthcare system, employing fog computing and the Internet of Medical Things.

Data availability: This dataset is taken from publicly available database [35].

Conflicts of Interest: The authors declare no conflict of interest.

References:

[1]   G. Tamagno, G. Federspil, and G. Murialdo, "Clinical and diagnostic aspects of encephalopathy associated with autoimmune thyroid disease (or Hashimoto's encephalopathy)," *Internal and Emergency Medicine,* vol. 1, pp. 15-23, 2006.

[2]   S. Y. Lee and E. N. Pearce, "Assessment and treatment of thyroid disorders in pregnancy and the postpartum period," *Nature Reviews Endocrinology,* vol. 18, pp. 158-171, 2022.

[3]   A. P. Delitala, M. Manzocco, F. G. Sinibaldi, and G. Fanciulli, "Thyroid function in elderly people: The role of subclinical thyroid disorders in cognitive function and mood alterations," *International Journal of Clinical Practice,* vol. 72, p. e13254, 2018.

[4]   A. P. Delitala, G. Delitala, P. Sioni, and G. Fanciulli, "Thyroid hormone analogs for the treatment of dyslipidemia: past, present, and future," *Current medical research and opinion,* vol. 33, pp. 1985-1993, 2017.

[5]   A. P. Delitala, "Subclinical hyperthyroidism and the cardiovascular disease," *Hormone and Metabolic Research,* vol. 49, pp. 723-731, 2017.

[6]   H. H. 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)*, 2020, pp. 836-841.

[7]   X. Zhang, V. Lee, J. Rong, J. C. Lee, J. Song, and F. Liu, "Multi-channel deep convolutional neural networks for multi-classifying thyroid disease," *arXiv preprint arXiv:2203.03627,* 2022.

[8]   A. P. Delitala, A. Scuteri, and C. Doria, "Thyroid hormone diseases and osteoporosis," *Journal of clinical medicine,* vol. 9, p. 1034, 2020.

[9]   L. Chiovato, F. Magri, and A. Carlé, "Hypothyroidism in context: where we've been and where we're going," *Advances in therapy,* vol. 36, pp. 47-58, 2019.

[10] S. Kaur, J. Singla, L. Nkenyereye, S. Jha, D. Prashar, G. P. Joshi, *et al.*, "Medical diagnostic systems using artificial intelligence (ai) algorithms: Principles and perspectives," *IEEE Access,* vol. 8, pp. 228049-228069, 2020.

[11] M. Alshamrani, "IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey," *Journal of King Saud University-Computer and Information Sciences,* vol. 34, pp. 4687-4701, 2022.

[12] B. Alhayani, A. S. Kwekha-Rashid, H. B. Mahajan, H. Ilhan, N. Uke, A. Alkhayyat, *et al.*, "5G standards for the Industry 4.0 enabled communication systems using artificial intelligence: Perspective of smart healthcare system," *Applied nanoscience,* pp. 1-11, 2022.

[13] A. Sujith, G. S. Sajja, V. Mahalakshmi, S. Nuhmani, and B. Prasanalakshmi, "Systematic review of smart health monitoring using deep learning and Artificial intelligence," *Neuroscience Informatics,* vol. 2, p. 100028, 2022.

[14] A. I. Siam, M. A. El-Affendi, A. Abou Elazm, G. M. El-Banby, N. A. El-Bahnasawy, F. E. Abd El-Samie, *et al.*, "Portable and Real-Time IoT-Based Healthcare Monitoring System for Daily Medical Applications," *IEEE Transactions on Computational Social Systems,* 2022.

[15] Y. A. Kadhim, M. U. Khan, and A. Mishra, "Deep Learning-Based Computer-Aided Diagnosis (CAD): Applications for Medical Image Datasets," *Sensors,* vol. 22, p. 8999, 2022.

[16] A. I. Siam, A. Abou Elazm, N. A. El-Bahnasawy, G. El Banby, F. E. Abd El-Samie, and F. Abd El-Samie, "Smart health monitoring system based on IoT and cloud computing," in *Proc. 1st International Conference on Electronic Engineering ICEEM2019, Menoufia University, Egypt*, 2019, pp. 37-42.

[17] M. M. Eisa and M. H. Alnaggar, "Hybrid Rough-Genetic Classification Model for IoT Heart Disease Monitoring System," in *Digital Transformation Technology*, ed: Springer, 2022, pp. 437-451.

[18] M. Alnaggar, A. I. Siam, M. Handosa, T. Medhat, and M. Rashad, "Video-based Real-Time Monitoring for Heart Rate and Respiration Rate," *Expert Systems with Applications,* p. 120135, 2023.

[19] E. Y. Jeong, H. L. Kim, E. J. Ha, S. Y. Park, Y. J. Cho, and M. Han, "Computer-aided diagnosis system for thyroid nodules on ultrasonography: diagnostic performance and reproducibility based on the experience level of operators," *European radiology,* vol. 29, pp. 1978-1985, 2019.

[20] M. Hosseinzadeh, O. H. Ahmed, M. Y. Ghafour, F. Safara, S. Ali, B. Vo, *et al.*, "A multiple multilayer perceptron neural network with an adaptive learning algorithm for thyroid disease diagnosis in the internet of medical things," *The Journal of Supercomputing,* vol. 77, pp. 3616-3637, 2021.

[21] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, and J. Qadir, "Explainable, trustworthy, and ethical machine learning for healthcare: A survey," *Computers in Biology and Medicine,* p. 106043, 2022.

[22] D. Li, D. Yang, J. Zhang, and X. Zhang, "Ar-ann: incorporating association rule mining in artificial neural network for thyroid disease knowledge discovery and diagnosis," *IAENG International Journal of Computer Science,* vol. 47, pp. 25-36, 2020.

[23] R. Vanitha and K. Perumal, "Hybrid Model of KNN and PCA to Pre-processor of Thyroid Dataset using Machine Learning," *Journal of Pharmaceutical Negative Results,* pp. 1091-1096, 2022.

[24] H. Abbad Ur Rehman, C.-Y. Lin, and Z. Mushtaq, "Effective K-Nearest Neighbor

Algorithms Performance Analysis of Thyroid Disease," *Journal of the Chinese Institute of Engineers,* vol. 44, pp. 77-87, 2021.

[25] Q. Pan, Y. Zhang, M. Zuo, L. Xiang, and D. Chen, "Improved ensemble classification method of thyroid disease based on random forest," in *2016 8th International Conference on Information Technology in Medicine and Education (ITME)*, 2016, pp. 567-571.

[26] P. Duggal and S. Shukla, "Prediction of thyroid disorders using advanced machine learning techniques," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2020, pp. 670-675.

[27] V. Sureshkumar, S. Balasubramaniam, V. Ravi, and A. Arunachalam, "A hybrid optimization algorithm-based feature selection for thyroid disease classifier with rough type-2 fuzzy support vector machine," *Expert Systems,* vol. 39, p. e12811, 2022.

[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.

[29] R. Pal, T. Anand, and S. K. Dubey, "Evaluation and performance analysis of classification techniques for thyroid detection," *International Journal of Business Information Systems,* vol. 28, pp. 163-177, 2018.

[30] P. Chalekar, S. Shroff, S. Pise, and S. Panicker, "Use of K-nearest neighbor in thyroid disease classification," *International Journal of Current Engineering and Scientific Research,* vol. 1, pp. 2394-2697, 2014.

[31] A. Tyagi, R. Mehra, and A. Saxena, "Interactive thyroid disease prediction system using machine learning technique," in *2018 Fifth international conference on parallel, distributed and grid computing (PDGC)*, 2018, pp. 689-693.

[32] V. S. Vairale and S. Shukla, "Classification of Hypothyroid Disorder using Optimized SVM Method," in *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019, pp. 258-263.

[33] S. Li and X. Zhang, "Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm," *Neural Computing and Applications,* vol. 32, pp. 1971-1979, 2020.

[34] Y. Jiang, G. Tong, H. Yin, and N. Xiong, "A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters," *IEEE Access,* vol. 7, pp. 118310-118321, 2019.

[35] Thyroid disease datasets. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/thyroid+disease

[36] M. J. Zaki, W. Meira Jr, and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*: Cambridge University Press, 2014.