Thyroid Syndrome Detection using Machine Learning Algorithms: A Comparative Analysis

Caio M. V. Cavalcante[†‖], Vinicius A. Almeida[†¶], Marcos Barros[†], Nathalee Lima[* §], and Rosana C. B. Rego[*‡]
[*]Departament of Engineering and Technology
Federal University Rural of the Semi-Arid, BR-226, Pau dos Ferros RN, Brazil
Email: [‡] rosana.rego@ufersa.edu.br; [§]nathalee.almeida@ufersa.edu.br
[†]Information Technology
Computational Intelligence Laboratory
Federal University Rural of the Semi-Arid, BR-226, Pau dos Ferros RN, Brazil
Email:[¶]vinicius.almeida37366@alunos.ufersa.edu.br; [‖]caio.cavalcante@alunos.ufersa.edu.br

***Abstract*—A thyroid syndrome necessitates early and proper diagnosis to facilitate adequate treatment. However, subjectivity in analyzing test results poses a challenge. In this work, we explored and analyzed the potential of machine learning algorithms. These algorithms include decision trees, random forest, logistic regression, naive Bayes, XGBoost, LightGBM, and a stacking ensemble model. The goal was to classify the euthyroid syndrome, which is a medical condition impacting the thyroid gland, by utilizing attributes obtained from blood tests. These attributes encompass thyroxine, thyroid stimulating hormone, free thyroxine index, total thyroxine, and triiodothyronine. The findings indicate the efficacy of employing these algorithms in accurately classifying the syndrome and providing diagnostic support.**

***Index Terms*—Thyroid syndrome, Machine learning, Classification, Healthcare.**

I. INTRODUCTION

The thyroid gland is located in the neck, and plays a crucial role in regulating various bodily functions. It is connected to the brain through a complex feedback loop involving the pituitary gland [1], as shown in Figure 1. The thyroid gland produces hormones, primarily thyroxine (T4) and triiodothyronine (T3), which are essential for maintaining metabolism, growth, and development [2]. These hormones are synthesized and released in response to signals from the brain [1].

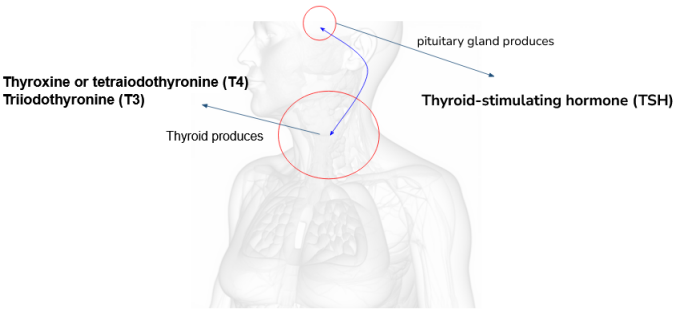


Fig. 1. Illustration of the location of the thyroid gland within the human body and its connection with the pituitary gland situated in the brain.

The connection between the thyroid and the brain begins with the hypothalamus, a region in the brain that releases thyrotropin-releasing hormone (TRH). TRH signals the pituitary gland, a pea-sized gland located at the base of the brain, to release thyroid-stimulating hormone (TSH). As the levels of T4 and T3 increase, they provide negative feedback to the hypothalamus and pituitary gland, reducing the production and release of TRH and TSH. This feedback mechanism ensures that thyroid hormone levels remain within a healthy range [1], [3].

A know thyroid syndrome is the Euthyroid Sick Syndrome (ESS), which is characterized by disrupted hormonal regulation of the thyroid glands in individuals who are concurrently affected by another illness or infection. This condition poses challenges in maintaining normal thyroid function and requires attention due to its impact on overall health. Understanding the mechanisms underlying ESS can provide insights into its diagnosis, management, and potential treatment strategies [4]–[6]. ESS is considered to be provoked by increasing circulating levels of cytokines and other inflammation mediators [7], [8]. These agents have the ability to impede the thyroid axis at various stages, encompassing the pituitary gland, resulting in reduced secretion of TSH, diminished levels of T4 and T3, decreased binding of thyroid hormones, and decreased conversion of T4 to T3 [7], [9].

Despite the thyroid operating in its usual manner, the synthesis of its hormones is influenced by inflammation, infection, or an alternative medical condition [8], [10]. This can result in abnormal levels of thyroid hormones in the blood and symptoms similar to those of thyroiditis or hypothyroidism [11]. Patients who are euthyroid but affected by a nonthyroidal systemic illness exhibit decreased serum levels of thyroid stimulating hormone [7]. In such cases, the management of this condition typically involves addressing the underlying disease, and evaluating thyroid function may be essential for monitoring the progression of the ailment [8], [12].

In the study conducted by McDermott [7], ESS was described as an adaptive response aimed at reducing tissue metabolism and conserving energy during systemic illnesses. The author noted that treatment with thyroid hormone is typically not recommended, except for patients with chronic heart failure who might benefit from it. Nevertheless, the management of ESS remains a topic of debate. To differentiate ESS from genuine hypothyroidism, it is often possible to assess the levels of serum T4, T3, and TSH, as well as resin

T3, commonly referred to as the free thyroxine index [13].

Timely and accurate diagnosis plays a vital role in effectively treating ESS. However, the interpretation of results for serum T4, T3, and TSH levels can be challenging and subjective, potentially leading to misdiagnosis [14]. In this regard, the application of machine learning (ML) algorithms has the potential to assist healthcare professionals in diagnosing patients' conditions [15]–[17]. ML models can learn and adapt to data patterns, enhancing the accuracy of prognosis, as demonstrated by [16]. By utilizing ML techniques for classifying Euthyroid sick syndrome, it becomes feasible to identify individuals who may be in the early stages of developing the condition, leading to more effective treatment [18]. Additionally, employing algorithms for ESS classification can save time and resources compared to traditional methods of data analysis.

The utilization of machine learning algorithms shows promise in detecting ESS at an early stage. This approach enables a rapid and precise examination of data, along with the identification of patterns and trends in clinical information. These insights can then be utilized to enhance the diagnosis and treatment of ESS [17], [19]. Numerous machine learning techniques have been proposed in the literature to facilitate the early detection of thyroid conditions [16]–[19]. Previous studies have not adequately verified whether the models were affected by overfitting. Although a model may achieve high accuracy, it can still suffer from overfitting. Analyzing the error curve helps us determine if the model is experiencing overfitting.

Inspired by the aforementioned discussion, this study undertook a comparative analysis of the random forest (RF), decision tree (DT), logistic regression (RL), XGBoost, light gradient-boosting (LightGBM), Naive Bayes, and a stacking ensemble model consisting of RF and XGBoost algorithms. The goal is to classify ESS using attributes derived from blood tests, including T4, TSH, free T4, and T3.

The manuscript is organized as follows: In Section II, we explained the dataset used, including its sources, characteristics, and any preprocessing steps undertaken to ensure data quality and consistency, from initial feature selection and engineering to the specific algorithms chosen for analysis. In Section III, we discussed about machine learning models, providing a comprehensive overview of the methodologies employed. Following this, Section IV presents the core of the paper, combining discussions and results. Finally, Section V the conclusions is presented.

II. Dataset

We utilized the Euthyroid Sick Syndrome dataset, which was obtained from the publicly available machine learning repository of the University of California Irvine (UCI). This dataset encompasses comprehensive information for each patient, including 25 attributes such as age, sex, medication history, pregnancy status, surgical history, and results from thyroid function tests.

Among the numerous available data points, we focused on extracting specific parameters such as age, sex, sickness status, TSH, T3, T4, Total T4, and free thyroxine index (FTI). The selection of these parameters was based on their correlation coefficients (Figure 2) and Recursive feature elimination (RFE) method (Figure 3), as they hold significant relevance in evaluating a patient's thyroid function and overall health condition. For instance, age and gender have been observed to impact the susceptibility to thyroid disorders. Meanwhile, TSH, T3, T4, total T4, and FTI serve as crucial indicators of various aspects of thyroid function, aiding in the diagnosis of thyroid-related ailments. Furthermore, these parameters can be effectively assessed through a standard blood test.

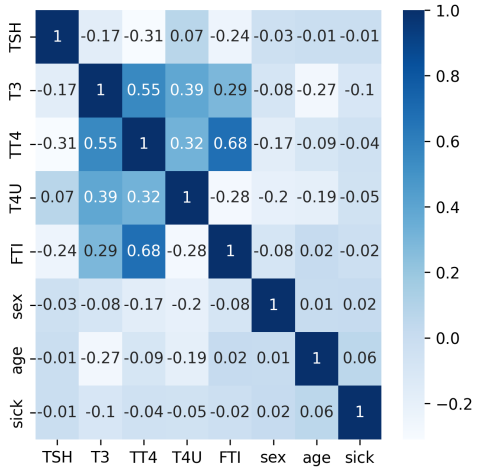
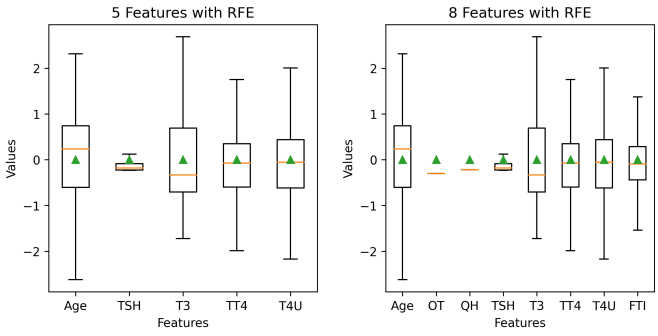


Fig. 2. Correlation matrix.

Figure 3 (a) displays the features that were picked using the recursive feature elimination method. These chosen features consist of age, TSH, T3, TT4, and T4U, making up a set of five selected features. In Figure 3 (b), a total of eight features were opted for: Age, On thyroxine (OT), Query hypothyroid (QH), TSH, T3, TT4, T4U, and FTI.



(a) (b)

Fig. 3. Recursive Feature Elimination: (a) RFE with 5 features selected (b) RFE with 8 features selected.

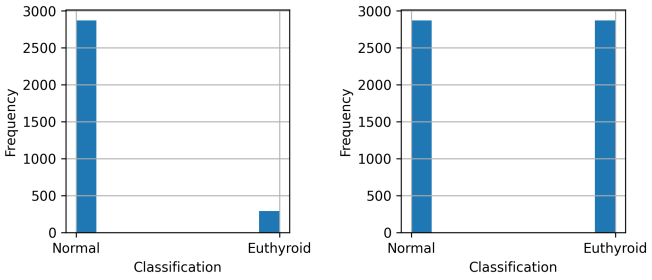
Table I shows the first five rows of the utilized database. The classification parameter value of 1 indicates that the patient

has euthyroid syndrome, while a value of 0 signifies that the patient does not have it.

TABLE I
FIRST 6 ROWS OF THE DATABASE.

Class	Age	Sex	Sick	TSH	T3	Total T4	T4	FTI
1	45	0	0	1.90	1.0	82	0.73	112
1	64	0	1	0.09	1.0	101	0.82	123
1	56	1	0	0.00	0.8	76	0.77	99
0	57	1	0	1	1.6	73	0.97	75
0	69	0	0	21	0.8	53	0.86	61
0	68	1	0	0.00	2.8	86	1.05	82

In order to prepare the data for utilization in the machine learning algorithm, it is necessary to address the issue of data imbalance, as depicted in Figures 4 (a) and (b).



(a) (b)

Fig. 4. Database: (a) unbalanced data and (b) balanced data.

The balancing procedure was carried out using the Synthetic Minority Over-sampling Technique (SMOTE) [20]. This technique tackles the data imbalance by generating additional data points for the minority class. The generation of these new instances is accomplished through the implementation of the K Nearest Neighbors (KNN) algorithm. Following the formatting, cleaning, and balancing steps, the data is now prepared for integration into the machine learning model. Figure 4 (b) shows the database after the application of SMOTE method.

III. MACHINE LEARNING MODELS

This section introduces various machine learning models employed for the classification task, including decision tree, random forest, logistic regression, XGBoost, LightGBM, Naive Bayes, and a stacking ensemble comprising RF and XGBoost.

A. Decision Trees

The decision tree model is a machine learning algorithm that uses the structure of trees to make predictions. The model structure is composed of nodes, sub-trees, and leaves, with the root level symbolizing the entire dataset. At each branch, a decision is made, and the number of questions asked is known as the depth of the tree. Upon reaching the sheet, we have the final classification [21], [22].

For the implementation of the decision trees model, we used the entropy function to evaluate the data divisions in each tree. In addition, the maximum depth of a tree was selected as 6. Moreover, weights were inversely proportional to the frequencies of the classes in the input data.

B. Random Forest

The random forest model is a supervised machine learning algorithm employed for classification or regression tasks [23]. This algorithm combines the simplicity of decision trees with an element of randomness. By constructing multiple decision trees, referred to as a forest, and aggregating their outputs, the model enhances accuracy. However, it is crucial to control the growth of the forest to prevent potential issues of overfitting or underfitting [24].

For the implementation of the RF model, we incorporated a forest consisting of 10 trees, each with a maximum depth of 10. To assess the data partitions and evaluate the model, we employed the logarithmic function. Furthermore, we set the minimum requirement of 5 samples for a leaf node and a minimum of 2 samples for dividing an internal node. Similar to the decision tree model, we assigned weights inversely proportional to the class frequencies in the input data.

C. Extreme gradient boosting

Extreme gradient boosting or XGBoost, introduced by [25], is a gradient boosting algorithm. This method leverages gradients to train the decision trees within the ensemble, incorporating the gradient of the loss function to update tree parameters. As a result, the algorithm exhibits improved speed and scalability compared to traditional approaches.

XGBoost has the capability to handle imbalanced data, large datasets, and showcases remarkable computational efficiency [25], [26]. We utilized the official XGBoost library to implement this algorithm. Initially, we employed the grid search method to fine-tune the model parameters. However, we encountered a challenge where the grid search returned parameters that excessively matched the training data, leading to overfitting. Therefore, we took a cautious approach to manually adjust the parameters, ensuring that overfitting was avoided.

D. Light gradient boosting

Similar to XGBoost, LightGBM is a gradient boosting technique that utilizes decision trees. This method is applicable for both classification and regression tasks. LightGBM constructs decision trees that grow leaf-wise, meaning that, based on the gain, only one leaf is split at a time given a condition. Although leaf-based trees can be prone to overfitting, particularly with smaller datasets, the issue can be mitigated by imposing constraints on the tree depth. By limiting the depth of the tree, the risk of overfitting can be alleviated [26].

For the implementation of the LightGBM model, we utilized a boosting learning rate of 0.3. Additionally, we specified a maximum tree depth of 15 and incorporated 5 boosted trees. Furthermore, we set the maximum number of tree leaves for base learners to be 15.

E. Logistic Regression

Logistic regression is a classification algorithm that utilizes multivariate analysis to assess the likelihood of a particular event occurring based on the identification of specific characteristics within each category resulting from the division of the defined area [27]. This technique enables the inclusion of non-linear effects and involves direct statistical tests [28].

For algorithm implementation, we employed the scikit-learn library and adopted the one-vs-rest (OvR) training scheme [29]. To handle class imbalances, we assigned weights inversely proportional to the frequencies of each class in the input data. Furthermore, we utilized the random state mode to control the random number generator for reproducibility purposes.

F. Stacking ensemble

Stacking is an ensemble technique in machine learning that involves training multiple models to generate predictions on the same dataset. These individual models' outputs are then used to train a meta-model that produces final predictions. The key concept behind stacking is to leverage the strengths of different models while mitigating their weaknesses [30], [31].

We employed a voting-based stacking approach, where the predictions from each base model were considered as votes, and a meta-classifier made the final classification decision based on these votes. In our stacking implementation, we combined the random forest model with XGBoost classifier. To implement this approach, we utilized the Stacking Classifier model available in the scikit-learn library.

G. Naive Bayes

Naive Bayes is a classification algorithm in machine learning that utilizes Bayes' theorem to make predictions about the category of a new entry [32]. It assumes that the features are independent of each other and that each feature follows a known probability distribution. The algorithm is rooted in Bayes' theorem, originally formulated by the English mathematician Thomas Bayes [32], [33].

The algorithm begins by calculating the probability of each category based on the features of the new input, and subsequently selects the category with the highest probability [33]. This is achieved by utilizing the Bayes equation to compute the posterior probability of each category, given the features of the new input. To implement the classifier, the scikit-learn library was employed. Additionally, the grid search method was utilized to fine-tune the algorithm's parameters.

IV. DISCUSSIONS AND RESULTS

A. Models evaluation

In order to assess the count of patients accurately categorized using decision tree, random forest, logistic regression, XGBoost, LightGBM, Naive Bayes, and a stacking ensemble consisting of RF and XGBoost model, we employed various measurements including confusion matrix, classification error curve, accuracy, recall, precision, and F1-score.

We executed pipeline with various machine learning algorithm, such as stacking ensemble called Random-XGBoost, XGBoost, Random Forest (RF), LightGBM, Decision Tree (DT), Extra Trees (ET), Gradient Boosting (GBC), Ada Boost (ADA), K Neighbors (KNN), SVM with a linear kernel, Logistic Regression (LR), Ridge Classifier, Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Quadratic Discriminant Analysis (QDA). Table II depict the models compiled and its respective metrics. In order to fine-tune the models, we employed a grid search approach coupled with cross-validation to ensure that overfitting is minimized.

According to Table II, XGBoost attained an accuracy of 0.9843 and an AUC of 0.9966. It exhibited high recall, precision, and F1-score values, indicating its effectiveness in classification. The RF model showcased an accuracy of 0.9837 and an AUC of 0.9973. Similar to XGBoost, RF demonstrated good recall, precision, and F1-score values. The algorithm LightGBM achieved an accuracy of 0.9837 and an AUC of 0.9971. It had a high recall score, although its precision and F1-score were slightly lower compared to RF and XGBoost. Decision Tree model demonstrated an accuracy of 0.9817 and an AUC of 0.9884. It exhibited high recall, precision, and F1-score values, making it a reasonable choice for classification. Overall, while the Random-XGBoost stacking ensemble was the best-performing model, other models such as XGBoost, Random Forest, and Decision Tree also demonstrated strong classification performance.

The confusion matrices for each model were depicted in Figures 5 (a), (b), (c), (d), and (e). In Figure 5 (a), it was observed that the random forest correctly labeled 579 patients as being in good health and 550 patients as being ill, while incorrectly classifying 9 healthy patients as ill and 10 ill patients as healthy. In Figure 5 (b), the logistic regression categorized 534 patients as healthy and 522 patients as ill, with an erroneous classification of 54 healthy patients as ill and 38 ill patients as healthy.

XGBoost accurately categorized 580 patients as healthy and 552 patients as ill, but it made 8 incorrect predictions by classifying healthy patients as ill and 8 ill patients as healthy. In contrast, LightGBM classified 576 patients as healthy and 545 patients as ill, with a misclassification of 12 healthy patients as ill and 15 ill patients as healthy. Hence, each model exhibited distinct outcomes and performance. Logistic regression classified 534 patients as normal and 522 patients as sick, with misclassifications of 54 normal patients as sick and 38 sick patients as normal. On the other hand, the Bayes model classified 455 patients as normal and 520 patients as sick, with misclassification of 133 normal patients as sick and 40 sick patients as normal. It is possible to notice that each model had different results.

As illustrated in Figure 5 (e), the random-XGBoost model accurately categorized 581 patients as healthy and 553 patients as ill, with only 7 misclassifications of healthy patients as ill and 7 ill patients as healthy. Consequently, the random-XGBoost model showcased the highest performance in terms of classification accuracy.

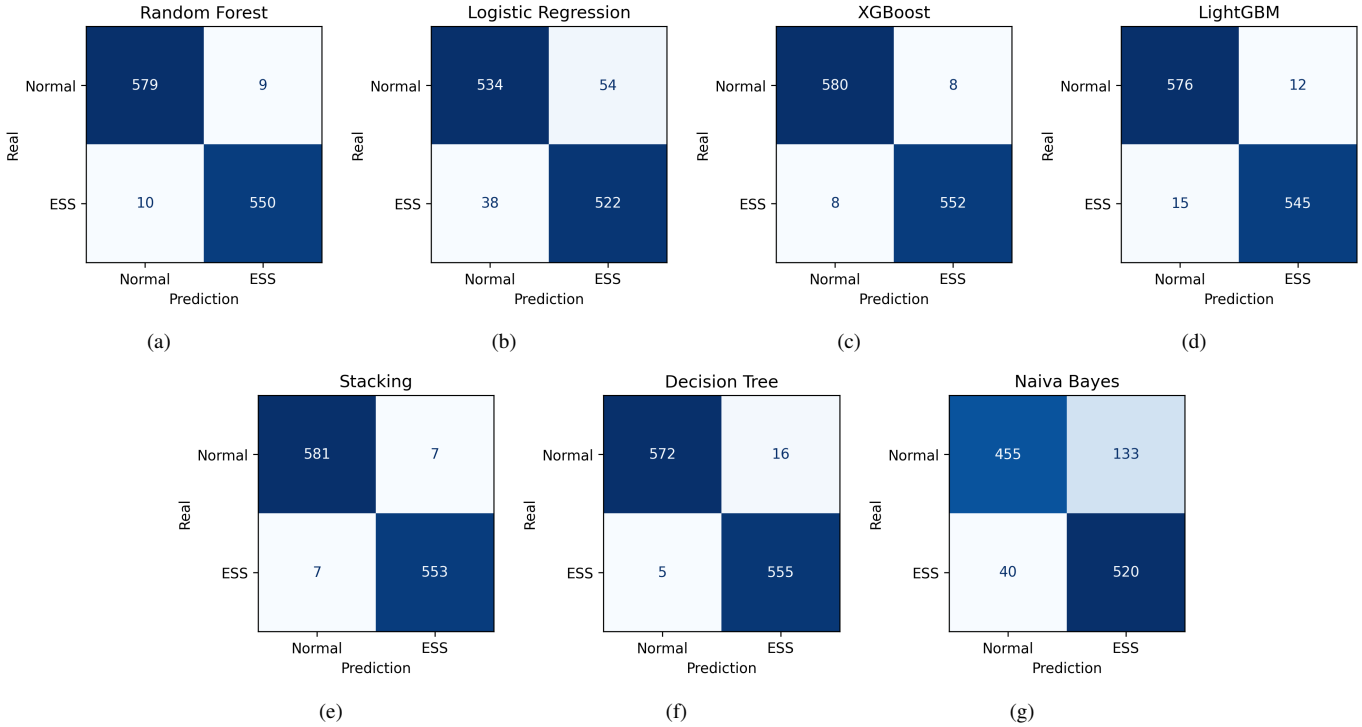


Fig. 5. Confusion matrix: (a) RF model, (b) LR model, (c) XGBoost model, (d) LightGBM, and (e) Random-XGBoost.

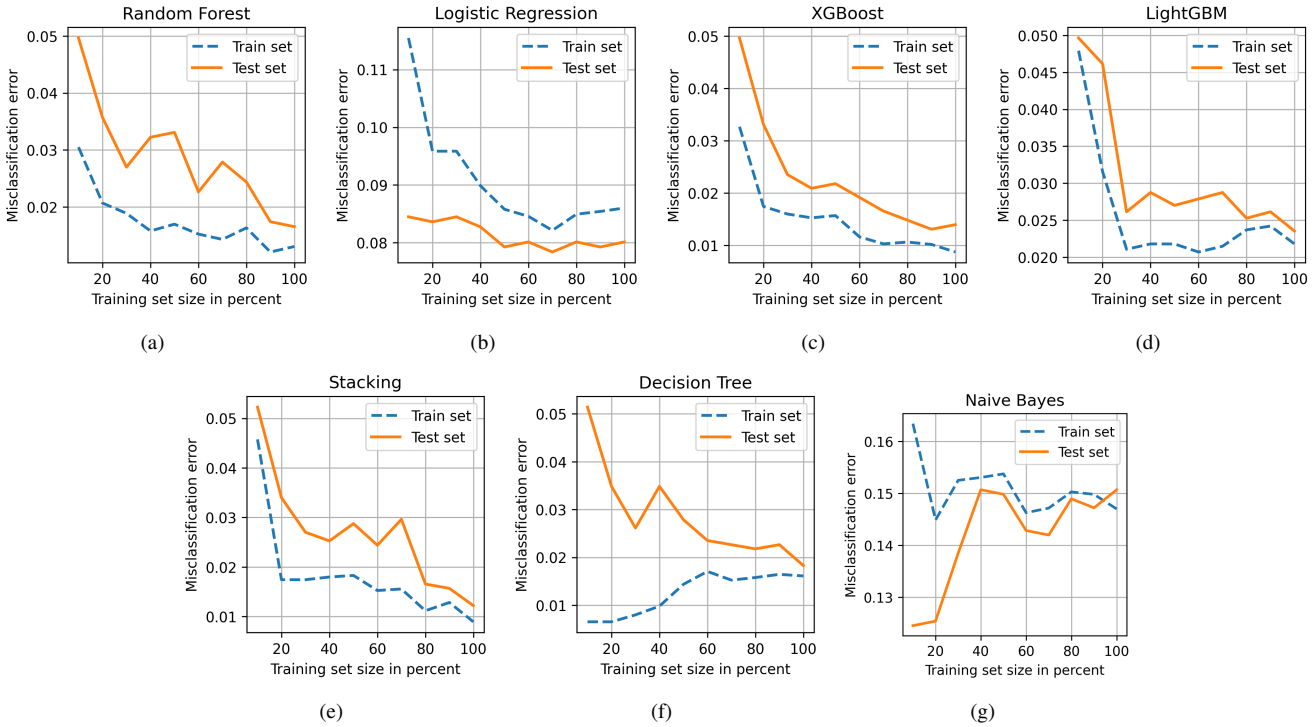


Fig. 6. Misclassification error: (a) RF model, (b) LR model, (c) XG Boost model, (d) LightGBM, and (e) Random-XGBoost.

The misclassification error curves for the models were shown in Figures 6 (a), (b), (c), (d), and (e). Analyzing these curves allows for an understanding of whether the models are experiencing overfitting or underfitting. The random forest, XGBoost, random-XGBoost, and LightGBM models exhibited a desirable fit, as indicated by initially high error rates during both training and testing. However, as more data was added, the error gradually decreased and eventually reached a

value close to zero. In contrast, the logistic regression model demonstrated a different pattern. It benefited from the addition of more data during both training and testing, initially reducing the error. However, as the model continued to train, it started to suffer from overfitting, indicated by a rise in the error rate during testing.

V. CONCLUSIONS

The application of ML algorithms is a promising approach for the early detection of ESS, as it allows a quick and accurate analysis of data, as well as the identification of patterns and trends in clinical data that can be used to improve the diagnosis and treatment of ESS. The findings indicated that employing machine learning algorithms for the classification of Euthyroid Sick Syndrome proves to be a valuable asset in facilitating the accurate diagnosis of the condition. Among the models tested, the random-XGBoost model demonstrated the highest classification accuracy performance with 98.78% . This study highlights the promising potential of integrating medical data with machine learning techniques as a means to enhance diagnostic procedures and potentially enhance the quality of life for patients.

The ability to enhance diagnostic procedures through machine learning has the potential to significantly improve patient outcomes. For instance, it could lead to quicker and more accurate diagnoses, enabling timely interventions and treatments. Ultimately, this could contribute to an overall improvement in the quality of life for patients, as they receive more targeted and effective medical care. In essence, this study showcases how the synergy between medical data and machine learning can open up new avenues for advancements in healthcare, showing promise for more accurate diagnostics and better patient well-being.

ACKNOWLEDGMENT

To PICI/UFERSA for financial support in granting a Scientific Initiation scholarship and UFERSA/PROPPG 65/2022 (PAPC) support for research groups.

REFERENCES

[1] J. R. Arthur and G. J. Beckett, "Thyroid function," *British medical bulletin*, vol. 55, no. 3, pp. 658–668, 1999.

[2] R. Pitt-Rivers and J. R. Tata, *The thyroid hormones*. Elsevier, 2013, vol. 1.

[3] V. M. C. d. Costa, D. G. Moreira, and D. Rosenthal, "Thyroid function and aging: gender-related differences," 2001.

[4] D. Sidebotham, *Cardiothoracic critical care*. Elsevier Health Sciences, 2007.

[5] S. Lee and A. P. Farwell, "Euthyroid sick syndrome," *Comprehensive Physiology*, vol. 6, no. 2, pp. 1071–1080, 2011.

[6] R. Negro, R. Attanasio, E. V. Nagy, E. Papini, P. Perros, and L. Hegedüs, "Use of thyroid hormones in hypothyroid and euthyroid patients; the 2019 italian survey," *European Thyroid Journal*, vol. 9, no. 1, pp. 25–31, 2020.

[7] M. T. McDermott, *Endocrine Secrets E-Book*. Elsevier Health Sciences, 2019.

[8] J. J. Zimmerman and B. P. Fuhrman, *Pediatric Critical Care E-Book*. Elsevier Health Sciences, 2011.

[9] V. Prasad, T. S. Rao, and M. S. P. Babu, "Thyroid disease diagnosis via hybrid architecture composing rough data sets theory and machine learning algorithms," *Soft Computing*, vol. 20, pp. 1179–1189, 2016.

[10] M. Zacharin, *Practical pediatric endocrinology in a limited resource setting*. Academic Press, 2013.

[11] F. E. Wondisford and S. Radovick, *Clinical Management of Thyroid Disease E-Book*. Elsevier Health Sciences, 2009.

[12] O. Gumieniak, T. S. Perlstein, P. N. Hopkins, N. J. Brown, L. J. Murphey, X. Jeunemaitre, N. K. Hollenberg, and G. H. Williams, "Thyroid function and blood pressure homeostasis in euthyroid subjects," *The Journal of Clinical Endocrinology & Metabolism*, vol. 89, no. 7, pp. 3455–3461, 2004.

[13] D. B. Allen, S. A. Hagen, and A. L. Carrel, "Disorders of the endocrine system relevant to pediatric critical illness," in *Pediatric Critical Care*. Elsevier, 2006, pp. 1105–1124.

[14] F.-s. Ouyang, B.-l. Guo, L.-z. Ouyang, Z.-w. Liu, S.-j. Lin, W. Meng, X.-y. Huang, H.-x. Chen, H. Qiu-Gen, and S.-m. Yang, "Comparison between linear and nonlinear machine-learning algorithms for the classification of thyroid nodules," *European journal of radiology*, vol. 113, pp. 251–257, 2019.

[15] M. L. Gild, M. Chan, J. Gajera, B. Lurie, Z. Gandomkar, and R. J. Clifton-Bligh, "Risk stratification of indeterminate thyroid nodules using ultrasound and machine learning algorithms," *Clinical Endocrinology*, vol. 96, no. 4, pp. 646–652, 2022.

[16] Y.-T. Lu, H.-J. Chao, Y.-C. Chiang, and H.-Y. Chen, "Explainable machine learning techniques to predict amiodarone-induced thyroid dysfunction risk: Multicenter, retrospective study with external validation," *Journal of Medical Internet Research*, vol. 25, p. e43734, 2023.

[17] M. Karmeni, E. B. Abdallah, K. Boukadi, and M. Abed, "Towards an accurate stacked ensemble learning model for thyroid earlier detection," in *2022 IEEE/ACS 19th International Conference on Computer Systems and Applications (AICCSA)*. IEEE, 2022, pp. 1–8.

[18] 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.

[19] A. Mashonga, L. KudzaiNyandoro, and K. Zvarevashe, "A comparative analysis of the effectiveness of feature engineering techniques on thyroid disease prediction," in *2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT)*. IEEE, 2022, pp. 1–6.

[20] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[21] Y. Freund and L. Mason, "The alternating decision tree learning algorithm," in *icml*, vol. 99, 1999, pp. 124–133.

[22] Y. Ben-Haim and E. Tom-Tov, "A streaming parallel decision tree algorithm." *Journal of Machine Learning Research*, vol. 11, no. 2, 2010.

[23] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," *Systems Science & Control Engineering: An Open Access Journal*, vol. 2, no. 1, pp. 602–609, 2014.

[24] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," *Ensemble machine learning: Methods and applications*, pp. 157–175, 2012.

[25] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[26] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.

[27] F. Thabtah, N. Abdelhamid, and D. Peebles, "A machine learning autism classification based on logistic regression analysis," *Health information science and systems*, vol. 7, pp. 1–11, 2019.

[28] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[29] Y. Alhessi and R. Wicentowski, "Swatac: A sentiment analyzer using one-vs-rest logistic regression," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 2015, pp. 636–639.

[30] D. H. Wolpert, "Stacked generalization," *Neural networks*, vol. 5, no. 2, pp. 241–259, 1992.

[31] L. Breiman, "Stacked regressions," *Machine learning*, vol. 24, pp. 49–64, 1996.

[32] G. I. Webb, E. Keogh, and R. Miikkulainen, "Naïve bayes." *Encyclopedia of machine learning*, vol. 15, pp. 713–714, 2010.

[33] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 403, p. 412, 2018.

TABLE II
MODEL METRICS.

Acronym	Model	Accuracy	AUC	Recall	Precision	F1	Execution time (second)
Random-XGBoost	Stacking ensemble	0.9878	0.9978	0.9875	0.9875	0.9875	0.455
XGboost	Extreme Gradient Boosting	0.9843	0.9966	0.9849	0.9737	0.9792	0.376
RF	Random Forest Classifier	0.9837	0.9973	0.9808	0.9758	0.9783	0.449
LightGBM	Light Gradient Boosting Machine	0.9837	0.9971	0.9855	0.9715	0.9784	0.227
DT	Decision Tree Classifier	0.9817	0.9884	0.9910	0.9719	0.9814	0.027
ET	Extra Trees Classifier	0.9832	0.9976	0.9849	0.9709	0.9778	0.321
GBC	Gradient Boosting Classifier	0.9780	0.9951	0.9739	0.9678	0.9708	0.660
ADA	Ada Boost Classifier	0.9612	0.9900	0.9385	0.9574	0.9477	0.338
KNN	K Neighbors Classifier	0.9521	0.9845	0.9774	0.9030	0.9387	0.058
SVM	SVM - Linear Kernel	0.9327	0.0000	0.9234	0.9005	0.9116	0.026
LR	Logistic Regression	0.9255	0.9572	0.9135	0.8912	0.9020	0.029
RIDGE	Ridge Classifier	0.8998	0.0000	0.8839	0.8546	0.8688	0.014
LDA	Linear Discriminant Analysis	0.8981	0.9499	0.8845	0.8503	0.8669	0.021
NB	Naive Bayes	0.8450	0.9227	0.9141	0.7370	0.8157	0.015
QDA	Quadratic Discriminant Analysis	0.8140	0.9217	0.9222	0.6889	0.7883	0.024