

OPTIMIZING MEDICAL DECISION-MAKING FOR HYPOTHYROIDISM WITH EXPLAINABLE ARTIFICIAL INTELLIGENCE

Caio M. V. Cavalcante

Federal University Rural of the Semi-Arid (UFERSA)
Rio Grande do Norte, Brazil
caio.cavalcante@alunos.ufersa.edu.br

Rosana C. B Rego

Federal University Rural of the Semi-Arid (UFERSA)
Department of Engineering and Technology
Rio Grande do Norte, Brazil
rosana.rego@ufersa.edu.br

ABSTRACT

Precise diagnoses are essential for the effective management of hypothyroidism. However, the subjective nature inherent in the interpretation of diagnostic tests poses a significant challenge. In this paper, we examined and evaluated the utility of the machine learning algorithms SHAPASH and LIME, with explainable AI, in addressing that challenge. The objective was to forecast hypothyroidism, a thyroid gland dysfunction, utilizing attributes derived from blood biomarkers and explain the model's resulting predictions. These attributes include thyroxine, thyroid-stimulating hormone, free thyroxine index, total thyroxine, and triiodothyronine. The results demonstrate the efficacy of employing these algorithms in accurately classifying hypothyroidism and providing diagnostic support, achieving an accuracy rate of 99.16%.

KEYWORDS. Classification. Machine learning. Hypothyroidism.

SA - OR in Health. IA - OR and AI.

1. Introduction

Hypothyroidism occurs when the thyroid gland fails to produce adequate thyroid hormones. It poses significant health issues if not addressed. Untreated cases may reveal conditions, including hypertension, dyslipidemia, infertility, cognitive impairment, as well as neuromuscular dysfunction [Gaitonde et al., 2012]. The diagnosis of hypothyroidism entails a multifaceted approach encompassing clinical assessment, laboratory analyses, and the evaluation of some symptoms and medical history [Matei et al., 2023]. Typically, primary blood tests employed for diagnosis comprise measurements of TSH, Free T4, T3, and Thyroid Peroxidase Antibody (TPOAb) levels [Gaitonde et al., 2012].

Integration of machine learning (ML) models holds potential in the detection of hypothyroidism, facilitating rapid analysis of medical data to identify patterns and critical details [Naeem et al., 2023; Cavalcante and Rego, 2024]. For conditions like hypothyroidism, doctors must understand why an AI model reaches a particular diagnosis. Explainable AI (XAI) techniques offer interpretability by revealing the features and patterns that contribute to a model's decision [Mishra, 2023; Sannala et al., 2023]. Several ML approaches in the scientific literature aim to detect thyroid disorders early [Shahid et al., 2019; Guleria et al., 2022]. However, many previous studies suffer from a lack of models that humans can interpret, as they do not provide feature importance scores, textual or visual explanations, or transparent internal workings for review. In this work, we implemented some ML algorithms to predict hypothyroidism, offering transparency in the decision-making process of the models with explainable AI algorithms, such as SHAPASH [MAIF, 2024] and LIME (Local Interpretable Model-agnostic Explanations) [Ribeiro et al., 2016].

2. Methodology

The Hypothyroid subset of the Thyroid Disease dataset from the UCI machine learning repository [Quinlan, 1987] was used with 3017 instances. It contains 29 patient attributes, including age, sex, medication history, pregnancy, surgeries, and thyroid function test results. Grid search with 5-fold cross-validation was applied to reduce overfitting. Features were selected based on frequency across Recursive Feature Elimination (RFE), Pearson Correlation, and K-means clustering. The selected features were TT4, TT4m, T4Um, T3m, FTI, T3, TSH, T4U, Pregnant, and I131. TT4 indicates total T4, T4U represents free thyroxine levels, and I131 refers to Iodo-131 usage. The dataset was split into 80% training and 20% testing.

A stacking model was implemented using Gradient Boosting (GB), Extra Trees (ET), Decision Tree (DT), and Random Forest (RF). GB, ET, and DT served as base classifiers, while RF performed the final classification. Explanations were provided using SHAPASH and LIME. Model performance was assessed using accuracy, recall, precision, and F1-score. The stacking model achieved 0.9916 accuracy and 0.9973 AUC, with a recall of 0.9863, precision of 0.9969, and F1-score of 0.9915, indicating excellent predictive performance and a strong balance between sensitivity and precision.

3. Discussions and Results

3.1. SHAPASH

SHAPASH offers an interface for generating and visualizing explanations, making to understand the behavior of the model prediction. Fig. 1 (a) depicts the importance of features on SHAPASH method results; for instance, TSH values are highlighted as having a stronger positive impact on the model's predictions. Additionally, features such as FTI, TT4, T3, T3 measure, and T4U emerge as crucial for predicting hypothyroidism. The visualization of points presents an understanding of how each selected feature influences the model's predictions across the entire dataset. Furthermore, SHAPASH provides local interpretability, which is the understanding of the decision-making process for a single sample, enabling the display of contributions from each most important

feature for this sample. Fig. 1 (b), (c), (d) and (e) present the contribution of TSH, FTI, T3, and total T4 to the prediction of hypothyroidism. A zero or negative contribution of TSH indicates a higher likelihood of hypothyroidism (predicted probability 0). A TSH contribution greater than zero suggests a lower likelihood of hypothyroidism (predicted probability 1).

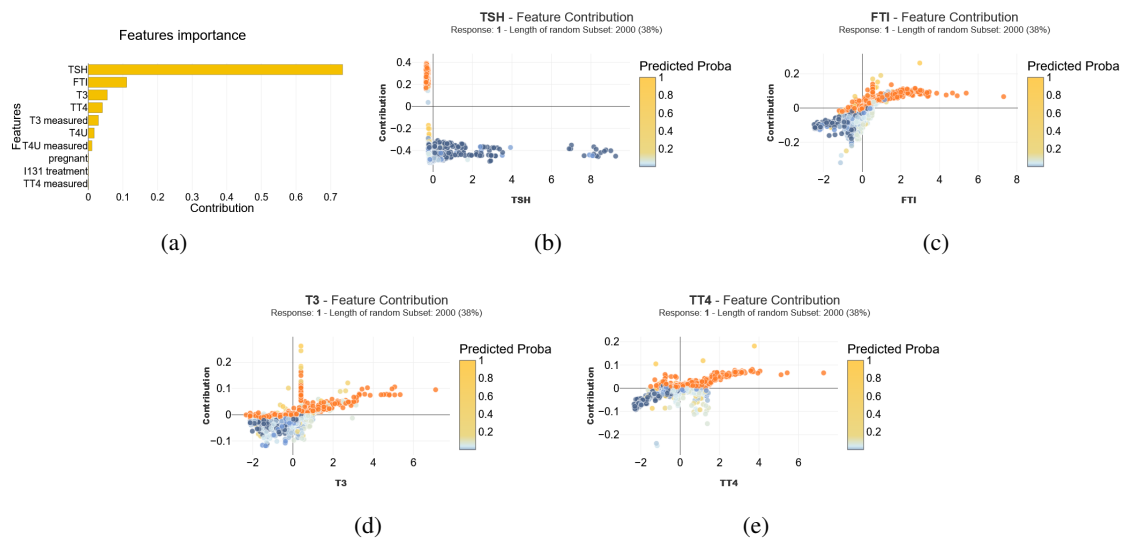


Figure 1: Feature importance and individual feature contributions for hypothyroidism prediction. (a) Overall feature importance ranking, highlighting TSH as the most influential predictor. (b)-(e) SHAP-based contribution plots for TSH, FTI, T3, and TT4, showing their impact on the predicted probability, where 1 corresponds to a normal patient and 0 to hypothyroidism.

3.1.1. LIME

LIME is used for its ability to provide intuitive and easily understandable explanations for individual predictions. Fig. 2 demonstrates the application of the LIME method in making predictions.

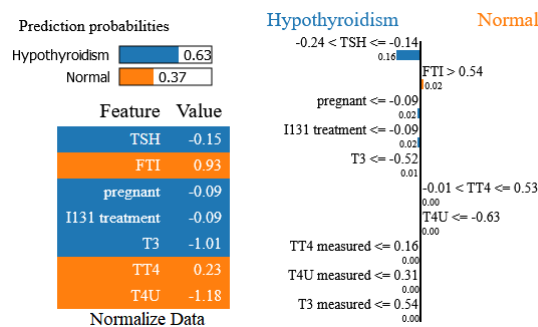


Figure 2: Features that impacted in the predicted value and their respective meanings.

In Fig. 2, the model predicted a 63% probability of hypothyroidism. Key features influencing this prediction include TSH, T3, and pregnancy status. These methods permit the explainability of the ML model. The XAI plays a crucial role in ensuring transparency and trustworthiness in AI systems.

4. Conclusions

ML algorithms offer significant promise in identifying hypothyroidism by uncovering patterns and trends within clinical data. These insights play a key role in advancing the diagnosis and treatment of hypothyroidism, eventually enhancing patient care and outcomes. This research highlights the potential of combining medical data with ML algorithms to refine diagnostic processes. Additionally, it emphasizes the importance of employing transparent, interpretable, and understandable models for doctors, facilitated by explainable AI methods like SHAPASH and LIME.

References

- Cavalcante, C. M. and Rego, R. C. (2024). Early prediction of hypothyroidism based on feature selection and explainable artificial intelligence. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, p. 49–60. SBC.
- Gaitonde, D. Y., Rowley, K. D., and Sweeney, L. B. (2012). Hypothyroidism: an update. *South African Family Practice*, 54(5):384–390.
- Guleria, K., Sharma, S., Kumar, S., and Tiwari, S. (2022). Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning. *Measurement: Sensors*, 24:100482.
- MAIF (2024). Shapash: User-friendly explainability and interpretability to develop reliable and transparent machine learning models. <https://github.com/MAIF/shapash>.
- Matei, M., Vlad, M. M., Golu, I., Dumitru, C. È., De Scisciolo, G., and Matei, S.-C. (2023). Can routine laboratory tests be suggestive in determining suspicions of malignancy in the case of thyroid nodules? *Medicina*, 59(8):1488.
- Mishra, P. (2023). Introducing explainability and setting up your development environment. In *Explainable AI Recipes: Implement Solutions to Model Explainability and Interpretability with Python*, p. 1–15. Springer.
- Naeem, A. B., Senapati, B., Chauhan, A. S., Makhija, M., Singh, A., Gupta, M., Tiwari, P. K., and Abdel-Rehim, W. M. (2023). Hypothyroidism disease diagnosis by using machine learning algorithms. *International Journal of Intelligent Systems and Applications in Engineering*, 11(3): 368–373.
- Quinlan, R. (1987). Thyroid Disease. UCI Machine Learning Repository. DOI: <https://doi.org/10.24432/C5D010>.
- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Why should i trust you? explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, p. 1135–1144.
- Sannala, G. S., Rohith, K., Vyas, A. G., and Kavitha, C. (2023). Explainable artificial intelligence-based disease prediction with symptoms using machine learning models. In *International Conference on IoT Based Control Networks and Intelligent Systems*, p. 523–538. Springer.
- Shahid, A. H., Singh, M. P., Raj, R. K., Suman, R., Jawaid, D., and Alam, M. (2019). A study on label tsh, t3, t4u, tt4, fti in hyperthyroidism and hypothyroidism using machine learning techniques. In *2019 International Conference on Communication and Electronics Systems (ICCES)*, p. 930–933. IEEE.