

Machine Learning Models for Predicting Hypothyroidism: Utilizing Synthetic Data for Improved Accuracy

Vinutha N Sajjan¹, Varsha S², Sheela S³

Undergraduate Students, Department of Computer Science and Technology^{1,2}

Assistant Professor, Department of Computer Science and Technology³

Global Academy of Technology, Bengaluru, India

Abstract: *This study presents a novel approach to early hypothyroidism detection by integrating synthetic data generation with machine learning (ML) techniques. Facing the challenge of limited and imbalanced healthcare datasets, we employ synthetic data to augment training sets, ensuring a richer and more diverse data pool for ML application. Key indicators of early hypothyroidism are distilled through feature selection, optimizing ML model inputs. We test various ML classifiers, including Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM), demonstrating enhanced diagnostic accuracy with our approach. Initial outcomes suggest that combining synthetic data with ML significantly boosts early detection capabilities, offering a promising direction for overcoming traditional data scarcity in medical diagnostics.*

Keywords: Decision tree, Hypothyroidism Prediction, Machine learning, NB, Random Forest, SVM

REFERENCES

- [1] Pan, Qiao, Yuanyuan Zhang, Min Zuo, Lan Xiang, and Dehua 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), pp. 567-571. IEEE, 2016.
- [2] Pavva, K., and B. Srinivasan. "Feature selection algorithms to improve thyroid disease diagnosis." In 2017 International conference on innovations in green energy and healthcare technologies (IGEHT), pp. 1-5. IEEE, 2017.
- [3] Tyagi, Ankita, Ritika Mehra, and Aditya Saxena. "Interactive thyroid disease prediction system using machine learning technique." In 2018 Fifth international conference on parallel, distributed and grid computing (PDGC), pp. 689-693. IEEE, 2018.
- [4] Shahid, Afzal Hussain, Maheshwari Prasad Singh, Rahul Kumar Raj, Rashmi Suman, Drakhshan Jawaid, and Muqtadir Alam. "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), pp. 930-933. IEEE, 2019.
- [5] Rao, Amulya R., and B. S. Renuka. "A machine learning approach to predict thyroid disease at early stages of diagnosis." In 2020 IEEE international conference for innovation in technology (INOCON), pp. 1-4. IEEE, 2020.
- [6] Mir, Yasir Iqbal, and Sonu Mittal. "Thyroid disease prediction using hybrid machine learning techniques: An effective framework." International Journal of Scientific & Technology Research 9, no. 2 (2020): 2868-2874.
- [7] Riajuliislam, Md, Khandakar Zahidur Rahim, and Antara Mahmud. "Prediction of thyroid disease (hypothyroid) in early stage using feature selection and classification techniques." In 2021 International conference on information and communication technology for sustainable development (ICICT4SD), pp. 60-64. IEEE, 2021.
- [8] Chaubey, Gyanendra, Dhananjay Bisen, Siddharth Arjaria, and Vibhash Yadav. "Thyroid disease prediction using machine learning approaches." National Academy Science Letters 44, no. 3 (2021): 233-238.
- [9] Chaganti, Rajasekhar, Furqan Rustam, Isabel De La Torre Díez, Juan Luis Vidal Mazón, Carmen Lili Rodríguez, and Imran Ashraf. "Thyroid disease prediction using selective features and machine learning techniques." Cancers 14, no. 16 (2022): 3914.

- [10] Guleria, Kalpna, Shagun Sharma, Sushil Kumar, and Sunita Tiwari. "Early prediction of hypothyroidism and multiclass classification using predictive machine learning and deep learning." *Measurement: Sensors* 24 (2022): 100482.
- [11] Devi, Munisamy Shyamala, Venkatesan Dhilip Kumar, Adrian Brezulianu, Oana Geman, and Muhammad Arif. "A Novel Blunge Calibration Intelligent Feature Classification Model for the Prediction of Hypothyroid Disease." *Sensors* 23, no. 3 (2023): 1128.
- [12] Hossain, Md Bipul, Anika Shama, Apurba Adhikary, Avi Deb Raha, KM Aslam Uddin, Mohammad Amzad Hossain, Intia Islam, Saydul Akbar Murad, Md Shirajum Munir, and Anupam Kumar Bairagi. "An Explainable Artificial Intelligence Framework for the Predictive Analysis of Hypo and Hyper Thyroidism Using Machine Learning Algorithms." *Human-Centric Intelligent Systems* (2023): 1-21.
- [13] Sa, Ri, Taiyu Yang, Zexu Zhang, and Feng Guan. "Random Forest for Predicting Treatment Response to Radioiodine and Thyrotropin Suppression Therapy in Patients With Differentiated Thyroid Cancer But Without Structural Disease." *The Oncologist* 29, no. 1 (2023): e68-e80.
- [14] Naeem, Awad Bin, Biswaranjan Senapati, Alok Singh Chauhan, Mukta Makhija, Arpita Singh, Meghna Gupta, Pradeep Kumar Tiwari, and Wael MF Abdel-Rehim. "Hypothyroidism disease diagnosis by using machine learning algorithms." *International Journal of Intelligent Systems and Applications in Engineering* 11, no. 3 (2023): 368-373.
- [15] Akash, K. Thirumala, F. Mohammed Usman, T. Nitesh Kumar, Mohammed Riyaz Ahmed, and Raveendra Gudodagi. "Predicting Thyroid Dysfunction Using Machine Learning Techniques." In *2023 12th International Conference on Advanced Computing (ICoAC)*, pp. 1-8. IEEE, 2023.