SciNexuses

Journal Homepage: sciencesforce.com/scin

SciNexuses Vol. 1 (2024) 139–147

Paper Type: Original Article

Improving Equity in Healthcare: Machine Learning-Based Thyroid Disease Classification

Rizwan Karim [1,*] and Muhammad Imran Asjad [1]

[1] Department of Mathematics, University of Management and Technology, Lahore, Pakistan.
Emails: rizwankarim3@gmail.com; imran.asjad@umt.edu.pk.

Received: 17 Aug 2024 **Revised**: 06 Nov 2024 **Accepted**: 24 Nov 2024 **Published**: 26 Nov 2024

Abstract

The thyroid, an important component of the endocrine system at the tip of the neck, plays an important role in the production of thyroxine, which is essential for overall health Disturbances in thyroid hormone production can lead to insufficient or excess levels. The use of large amounts of data in the healthcare industry has become increasingly complex, necessitating the use of machine learning. Our research focuses on thyroid disease, using a combination of machine learning and unbiased samples. Our main goal is to classify thyroid diseases into hypothyroidism, regular, and hyperthyroidism using machine learning. We used a real-world data set from Kaggle, split into 70% for training and 30% for testing. This division allows one to search for measures of accuracy and unbiasedness, especially with respect to logistic regression values. When reweighting methods were implemented, we saw an increase in accuracy and fairness metrics. Our study demonstrates the effectiveness of machine learning models with unbiased priority, yielding an accuracy of 84.58% with an appropriate value of -0.007. The importance of our work extends to the application of artificial intelligence (AI) in healthcare. By using AI algorithms to identify patterns in data, we demonstrate the potential to enhance medical research and treatment outcomes. Furthermore, our inclusion of justice considerations in model construction highlights the ethical considerations in the use of AI. It emphasizes the importance of fair and transparent decision-making in health care systems. This is consistent with broader AI research objectives, which aim to develop technologies that not only maximize accuracy but also maintain principles of fairness and accountability in their implementation.

Keywords: Fairness; Machine Learning; Thyroid Disease; Classification; Model Evaluation.

1 | Introduction

Thyroid dysfunction (TD) refers to the complicated condition of the thyroid gland, a small, butterfly-like organ lying in the neck, which is responsible for producing important hormones that regulate metabolism, growth, and energy the amount will manifest in various ways when this delicate balance is disrupted [1]. It can, be where the thyroid produces these important hormones in excess or insufficiently. TD encompasses several disorders, each with specific symptoms and symptoms [2]. Hyperthyroidism refers to an excessive amount of thyroid hormone, which causes symptoms such as weight loss, rapid heartbeat, and nervousness [3]. Conversely, hypothyroidism means that thyroid hormones are not being secreted properly, leading to fatigue, weight gain, and feelings of cold. Additionally, thyroiditis can be inflammation of the thyroid gland,

Corresponding Author: rizwankarim3@gmail.com
https://doi.org/10.61356/j.scin.2024.1434

which causes symptoms such as sore throat, fatigue, difficulty swallowing, etc. A specific type of thyroiditis, Hashimoto's thyroiditis, is an autoimmune disorder of the body's immune system the immune system mistakenly attacks the thyroid gland and there is a problem [4]. By understanding the nuances of these conditions, healthcare professionals can diagnose and effectively manage thyroid dysfunction, to restore hormonal balance and improve patients [5].

A large proportion of adult women, estimated at 9 to 15 percent, and fewer men are believed to have thyroid disease. According to one study, thyroid complications affect more than 20 million Americans. Experts estimate about 12% of individuals will develop thyroid dysfunction at some point in their lives [6]. Notably, women are at a higher risk, about five to eight times more likely than men, to develop thyroid disease, and one in eight women are likely to experience thyroid disease in their lifetime Throughout pregnancy, the thyroid gland undergoes significant changes in both size and function [7]. In iodine-rich environments, the vessels dilate by 10%, whereas in iodine-deficient environments, this dilation can reach $20-40\%$ [8]. It should be noted that iodine plays an important role in the metabolism of energy. However, despite its importance, nearly one-third of the world's population lives in iodine-deficient areas [9]. This highlights the prevalence of iodine deficiency and its potential impact on thyroid health worldwide.

Thyroid hormone plays an important role in the synthesis and release of triiodothyronine (T3) and thyroxine (T4), the only iodine-containing hormones found in vertebrates. These hormones are essential to ensure growth efficiency, cell differentiation, and metabolism [10]. The anterior pituitary gland assumes responsibility for the synthesis of serum thyrotrophin (TSH), which in turn regulates the secretion of this important hormone [11]. Approximately 95% of the thyroid hormone in the blood is composed of T4, which regulates metabolism, mood, and primary body temperature Traditionally, the remaining 5% of the thyroid hormone pool is composed of T3 [12]. Disruption of the delicate balance of thyroid hormone levels can precipitate a variety of side effects. Hyperthyroidism characterized by an overactive thyroid can manifest itself in many complications including vision problems, heart irregularities, osteoporosis, and skin disease. In contrast, hypothyroidism marked by an under-active thyroid can cause a wide range of symptoms, such as inflammation of the thyroid gland known as goiter, which can impair breathing and swallowing Other causes include high cholesterol, extremely fragile heart disease, painful arthritis, birth issues, congenital infant defects and anxiety [13]. Early detection and intervention hold promise for changing the course of thyroid dysfunction, alleviating symptoms, and preventing irreversibility Using machine learning algorithms to accurately predict thyroid dysfunction in the newborn can reduce the severity of complications and enhance patient safety.

Machine learning (ML) algorithms employ a diverse array of statistical, probabilistic, and optimization techniques to distill insights from historical data and discern meaningful patterns within intricate and disordered datasets [14-17]. These algorithms serve as powerful tools for automatically categorizing text, detecting network intrusions, analyzing consumer purchasing behavior, forecasting the onset of diseases, and facilitating decision-making processes, among numerous other applications. By leveraging pre-programmed algorithms, ML systems glean knowledge from input data, continuously refining their understanding and optimizing their performance through iterative evaluation. This iterative process enables ML algorithms to make predictions within an acceptable margin of error, effectively harnessing the vast potential of data to drive informed decision-making and enhance outcomes across a broad spectrum of domains [18].

Many difficult problems can be solved using machine learning methods [19]. Classification is a data extraction technique (machine learning) used to predict and identify many diseases, including thyroid disease, which we studied and classified here because machine learning algorithms play an important role in thyroid disease classification and are high-performing and efficient [20]. While computer learning and artificial intelligence have been used in medicine since their inception [21], there is now a growing emphasis on machine learning-driven healthcare solutions. Analysts believe that machine learning will become widely used in healthcare shortly [22].

According to experts, identifying, diagnosing, and treating illnesses early on is critical to lowering disease progression and death. Early diagnosis and treatment for a variety of illnesses boost the likelihood of effective therapy. Despite several clinical studies, detection is widely regarded as a difficult task [23]. The butterfly-shaped thyroid gland is located near the bottom of the neck. It consists of levothyroxine (T4), two functional thyroid hormones, and triiodothyroxine (T3), which controls blood pressure, heart rate, and body temperature. Thyroid illness is a common condition in the nation, often caused by a lack of iodine. However, other factors might also contribute. The thyroid gland seems to be a hormone-producing endocrine organ that transfers hormones throughout the body. It is located at the front core of the human body. Thyroid hormones are primarily important for metabolism and maintaining the body's moisture balance. There are two forms of thyroid problems: hyperthyroidism and hypothyroidism. ML methods are one of the most effective approaches to a variety of complex issues [24]. Classification appears to be a method of data gathering (machine learning) that is used for detecting and recognizing various diseases, which include thyroid cancer, something that we explored and classified here simply as ml algorithms play a very critical role in characterizing thyroid disease because they are effective and successful and aid in classifying [25].

Clearly, a comprehensive study explicitly dedicated to checking the assumption of unbiasedness in the machine learning models currently used to predict thyroid dysfunction. To fill that gap large this address, the present study Is a purposeful retrospective cohort study. The assumption underlying this analysis is that even in models that exhibit exceptional skill in classification accuracy, heterogeneities found in specific sub-populations may persist subpopulations may be defined by specific traits or traits of this, such as patients' T3 test levels. The main objective of this study is to shed light on the possibility of bias or imbalance in model forecasts that may have hitherto gone unnoticed due to the focus on general measures of forecast accuracy because of the background. By exploring these gaps further, we aim to contribute to the growing body of knowledge surrounding the relevant ethical implications of machine learning for thyroid disease prediction. Through careful analysis and understanding of these differences, we seek to enhance our understanding of the complex interplay between machine learning processes and issues of fairness and to inform actions that are just and ethical in health prediction modeling.

2 |Data Set

The process of creating the training and test datasets for our thyroid disease prediction model with fairness included the use of a TD-specific dataset. This TDs dataset is an open-source resource gathered from Kaggle, with a total of 3772 observations, each representing an individual observation. Most of the 3481 samples included in this data set fall into the "negative" category, indicating the absence of thyroid dysfunction. 194 Classified as "compensated hypothyroid", which means that such a condition is brief but the throat-royd hormone levels are all adequately depleted in addition to the other 95 events: primary hypothyroid is given, and only 2 examples: Classified as "secondary hypothyroid", referring to thyroid dysfunction caused by issues outside the gland itself. In total, the data set contains 30 items, each representing a specific attribute or feature associated with the observation. For our analysis, however, we focus our analysis on a subset of the 12 attributes we consider most relevant to the research at hand. This data set uses the binary values ' T ' and ' F ' to represent Boolean properties, where ' T ' represents true and ' F ' represents false. These properties can support the predictive modeling process, playing an important role in informing the classification of models in their categories. Through careful consideration of these attributes, we aim to gain deeper insights into the underlying patterns and relationships within the data, ultimately contributing to a more comprehensive understanding of thyroid dysfunction and its predictive indicators.

3 |Methodology

The main objective of this project is to develop an advanced system that can accurately predict the probability of developing thyroid disease, using state-of-the-art computer prediction algorithms developed specifically for thyroid disease, with a special focus on ensuring accuracy throughout the prognostic process holds significant promise, aiming to transform the early detection and management of thyroid conditions. Our strategy for achieving this ambitious goal is to use a variety of machine learning classification models to critically analyze thyroid disease data, prioritizing fairness and equity in our approach. This study represents the culmination of an extensive research effort aimed at refining and refining our prediction model. An important aspect of our approach is to prepare a well-structured data set to increase the robustness of our analysis methods. This preparation phase includes a comprehensive data cleaning process, aimed at identifying and correcting any inconsistencies or errors in the data structure in addition to systematically eliminating unnecessary or irrelevant information methods to facilitate the search process and improve the overall quality of the data. In addition, some elements of the data set undergo standardization processes to improve the uniformity and accuracy of the data, thereby increasing the effectiveness of our predictive model This careful focus on data quality ensures that our model is accurate and reliable. After the preparation phase, our model is trained on the processed data, using the collective insights gleaned from our rigorous analytical approach. We expect that this new approach will not only provide more accurate forecasts but also significantly improve the performance of the overall model. Objective evidence supporting the efficiency and reliability of our method is presented in Figure 1, highlighting the obvious utility of our method in improving the prognostic field of thyroid disease. Through this combined effort, we seek to set a new standard for predictive modeling in healthcare, one that prioritizes fairness, equity, and patient-centered care. After the preparation phase, our model is trained on the processed data, using the collective insights gleaned from our rigorous analytical approach. We expect that this new approach will not only provide more accurate forecasts but also significantly improve the performance of the overall model.

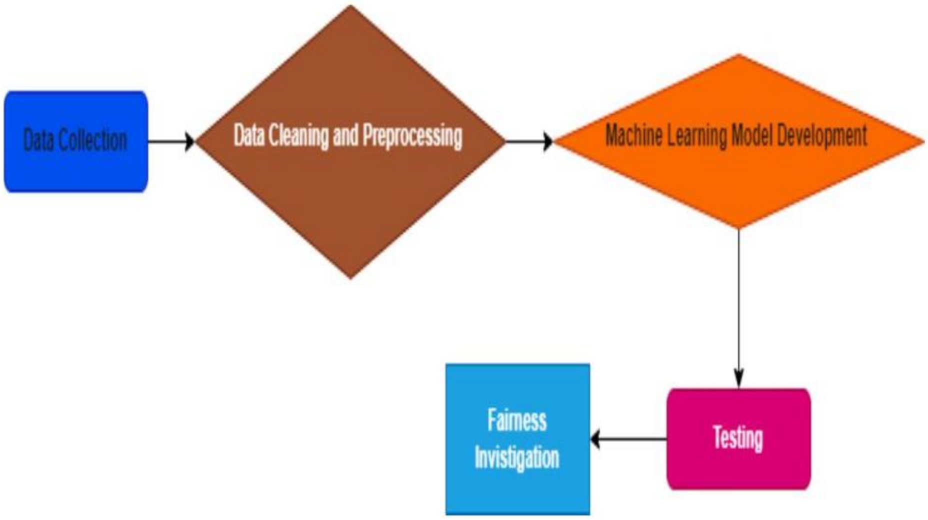


Figure 1. Workflow.

3.1 | Machine Learning Model

In this study, we employed a Logistic regression classifier to predict thyroid outcomes, with a particular focus on considering gender as a significant factor across all patient cases. Our initial assessment revealed an overall accuracy of 0.8458, indicating the model's proficiency in making accurate predictions. However, a slight discrepancy of -0.007 was observed in terms of equal opportunity, suggesting a potential area for improvement in ensuring fairness across gender categories. Figure 2 visually depicts the distribution of accuracies and Table 2 provides a ranking of these accuracies relative to the disparities in equal opportunity observed at different data points. To address this imbalance, we fine-tuned the model by adjusting the hyperparameter value to 0.5, resulting in an enhanced accuracy of 0.8471 and a concurrent improvement in the fairness metric to 0.0003. Interestingly, further adjustments to the hyperparameter, specifically to a value of 0.7, did not yield any significant changes in the model's accuracy.

Employing reweighting techniques in conjunction with logistic regression allowed us to restore the model's accuracy to its original level of 0.8458, while also minimizing the equal opportunity difference to 0.0011.

Figure 3 and Table 4 provide a comprehensive summary of the accuracy rankings and equal opportunity differences after applying the reweighting technique.

Further experimentation involved modifying the hyper-reweighing value to 1, which resulted in a slight decrease in accuracy to 0.8410, albeit with a marginal increase in the fairness metric to 0.008. Conversely, reducing the hyper-reweighing value to 0.00001 led to a notable decrease in both accuracy (0.8074) and fairness (-0.0001), as illustrated in the accompanying table. These findings underscore the sensitivity of the model's performance to adjustments in hyperparameters and emphasize the critical importance of striking a delicate balance between accuracy and fairness in modeling practices. By meticulously fine-tuning these parameters, we can optimize model performance while ensuring equitable outcomes across diverse demographic groups, thereby advancing the ethical integrity and effectiveness of predictive modeling in healthcare.

4 |Fairness

4.1 |Investigating Bias

emphasis on T3 levels. For this design, we classified groups as privileged or unprivileged based on sex and prevalence of thyroid issues. Table 1 provides a detailed breakdown of the privileged and unprivileged classes of protected objects corresponding to each class. We used the Python library "FairLearn" to find and quantify the correctness of our model.

Table 1. Relationship between gender and privileged or unprivileged groups.

Protected Attribute	Privileged Group	Unprivileged Group
Gender	Male	Female

Our method involves estimating critical parameters for each non-opportunistic category and any performance criteria associated with our model as described in Eq. (1). These parameters are a key determinant of fit in model performance. Where a value close to one indicates normal performance [26]. We used a criterion ranging from 0 to 1 to identify any biases in our models across sites [27]. In our analysis, we adopted a critical threshold of 0.7 to identify biases in our samples, which we matched with the widely accepted heuristic known as the " 70% rule" [28]. We applied this threshold we extended the acceptable range for model performance coefficients from 0.7 to 1.15 with $\epsilon = 0.7$ [29]. We were able to establish that, as a result, if the coefficients estimated by Eq. (1) fall within that range within this predefined range (0.7 to 1.15), the model is considered unbiased for that particular measurement and associated components [30].

$$\forall_{i \varepsilon a,b,c,\dots,\dots,z}$$

$$\varepsilon < \frac{metric_i}{metric_{privileged}} < \frac{1}{\varepsilon} \tag{1}$$

Table 2. Accuracy and fairness with different values of test hyper.

	Test hyper value with 1	Test hyper value with 0.00001
Accuracy	0.8471731448763251	0.9390459363957597
Fairness	0.00037678087837045293	0.03426573426573429

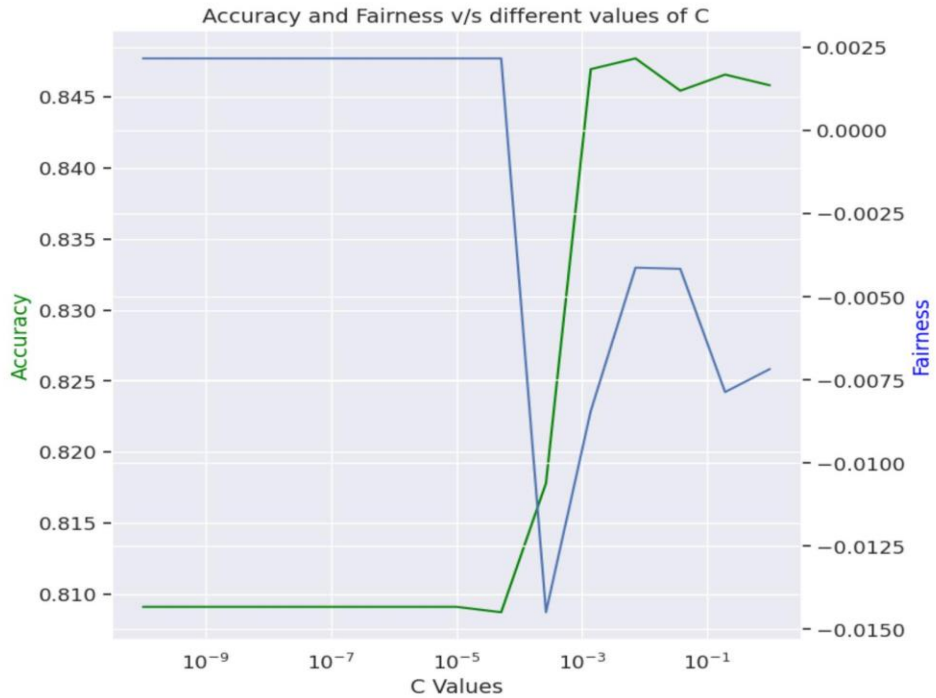
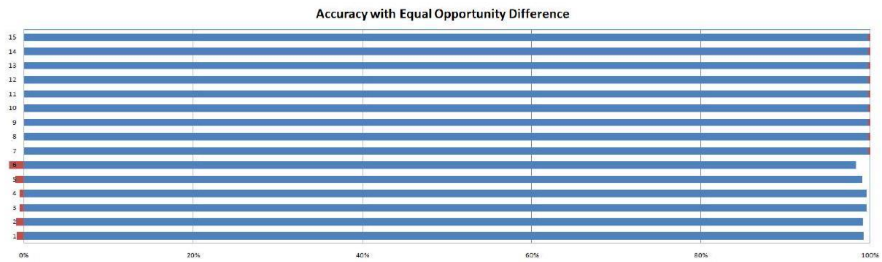


Figure 2. Accuracy with fairness.

4.2 | Mitigating Bias

An additional weighting method was used to reduce the algorithmic bias in our prediction models. This approach involves changing the model weights at each feature-outcome combination during the training phase of the model [31]. The goal was to fine-tune these weights to reduce bias and increase the fairness of the model's predictions [32].

To assess the effectiveness of this procedure, we compared the degree of bias in the revised models with those found in the original original models [33]. The reweighting process requires that a new model was developed in which the model weights were rescaled based on the distribution of samples in the non-privileged and privileged groups. Therefore, by rebalancing the training data we we aimed to correct any existing biases and to promote fairness and equality in model predictions.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Accuracy	0.845833333	0.846590909	0.845454545	0.847727273	0.846969697	0.81780303	0.808712121	0.809090909	0.809090909	0.809090909	0.809090909	0.809090909	0.809090909	0.809090909	0.809090909
Equal Opportunity Difference	-0.00716836	-0.007861637	-0.004154112	-0.004117542	-0.008429644	-0.014484753	0.002175521	0.002175521	0.002175521	0.002175521	0.002175521	0.002175521	0.002175521	0.002175521	0.002175521

Figure 3. Accuracy with Fairness.

We used the reweighting method as a reliable strategy in our attempts to address biases in prediction models. This method allows us to systematically change the weights of the data, lowering the algorithmic bias in the original models. By carefully adjusting the weights on various combinations of qualities and results throughout training, we hope to produce a more fair representation of the wealthy and marginalized groups. We seek improved samples to uncover any lingering biases and compare them robustly to their original equivalents. The goal of this iterative procedure is to review the model's training profile, encourage unbiasedness, and perfect its predictions. In addition to correcting preexisting biases, our methodical approach demonstrated our dedication to advancing equity and inclusion in predictive modeling procedures in the healthcare industry.

Table 3. Accuracy and fairness with different values of test hyper.

	Test hyper value with 1	Test hyper value with 0.00001
Accuracy	0.8480565371024735	0.8074204946996466
Fairness	0.008995643471093828	-0.00016484163428709397

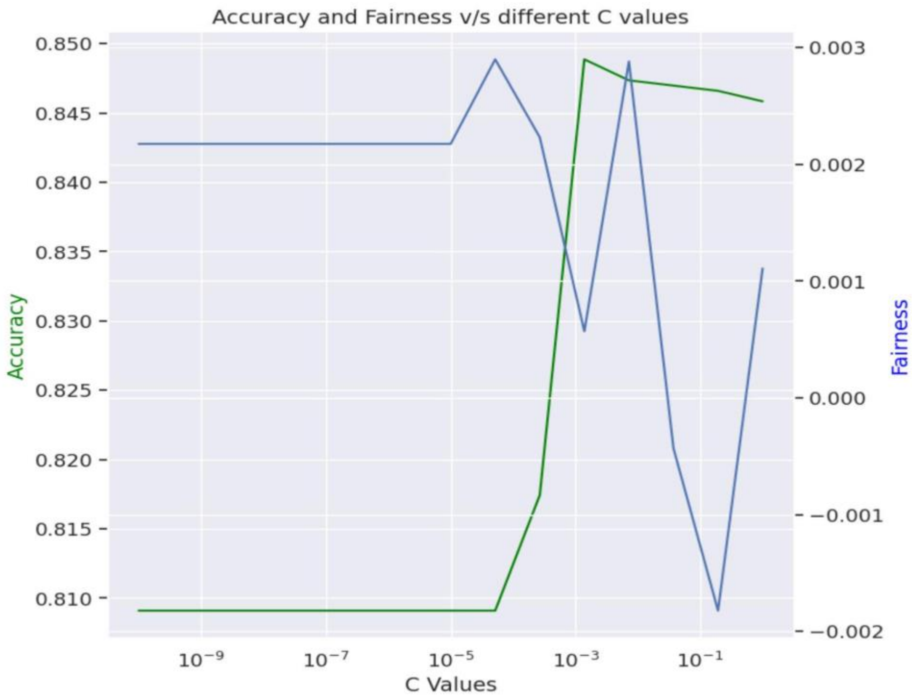


Figure 4. Accuracy with Fairness after reweighting.

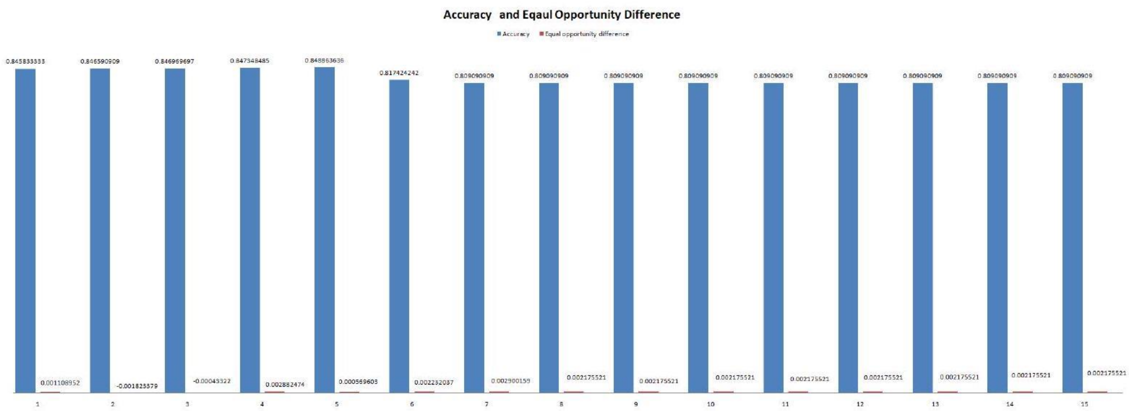


Figure 5. Accuracy with Fairness after reweighting.

5 | Conclusion

In conclusion, our study highlights the important role of machine learning in improving the understanding and management of thyroid diseases in the healthcare sector. Using machine learning along with unbiased models, we were able to classify thyroid disease and improve the accuracy and unbiasedness of measurement. Achieving a precision rate of 84.58% and incorporating fairness considerations highlights the potential of AI to revolutionize medical diagnosis and treatment Not only does our work contribute to improving healthcare practices but it also highlights the ethical importance given that accuracy is also emphasized by AI-driven decision-making. Going forward, it is important to continue to explore machine learning and health communication and to continue to emphasize fairness and accountability. By incorporating these principles into the application of AI, we can strive towards a future where technology not only enhances medical outcomes but also supports the core ethical values of patient care.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Brent, G. A. (2012). Mechanisms of thyroid hormone action. The Journal of clinical investigation, 122(9), 3035-3043.
[2] Liu, W., Lin, C., Chen, D., Niu, L., Zhang, R., & Pi, Z. (2024). Shape-margin knowledge augmented network for thyroid nodule segmentation and diagnosis. Computer Methods and Programs in Biomedicine, 244, 107999.
[3] Devereaux, D., & Tewelde, S. Z. (2014). Hyperthyroidism and thyrotoxicosis. Emergency Medicine Clinics, 32(2), 277-292.
[4] Tamer G, Arik S, Tamer I, Coksert D (2011) Relative vitamin D insufciency in Hashimoto's thyroiditis. Thyroid 21(8):891-896.
[5] Majety, P., & Hennessey, J. V. (2022). Acute and subacute, and Riedel's thyroiditis. Endotext [Internet].
[6] Arshad, S. H., & Babu, K. S. (2008). Asthma. Oxford University Press.
[7] Vanderpump, M. P. (2011). The epidemiology of thyroid disease. British medical bulletin, 99(1).
[8] Risher, J., & Keith, S. (2009). Iodine and inorganic iodides: human health aspects (No. 72). World Health Organization.
[9] Andersson, M., Karumbunathan, V., & Zimmermann, M. B. (2012). Global iodine status in 2011 and trends over the past decade. The Journal of nutrition, 142(4), 744-750.
[10] Obregon, M. J., del Rey, F. E., & de Escobar, G. M. (2005). The effects of iodine deficiency on thyroid hormone deiodination. Thyroid, 15(8), 917-929.
[11] Mariotti, S., & Beck-Peccoz, P. (2021). Physiology of the hypothalamic-pituitary-thyroid axis. Endotext [Internet].
[12] Iacovides, S., Maloney, S. K., Bhana, S., Angamia, Z., & Meiring, R. M. (2022). Could the ketogenic diet induce a shift in thyroid function and support a metabolic advantage in healthy participants? A pilot randomized-controlled-crossover trial. PloS one, 17(6), e0269440.

[13] Liu, Y., Shan, Z., Endocrine Metabolic Diseases Group of the Chinese Geriatrics Society, Thyroid Group of the Chinese Society of Endocrinology, Chinese Medical Association, Cheng, M., Cao, C. X., ... & Zuo, C. L. (2021). Expert consensus on diagnosis and treatment for elderly with thyroid diseases in China (2021). Aging Medicine, 4(2), 70-92.

[14] Zhong, N. N., Wang, H. Q., Huang, X. Y., Li, Z. Z., Cao, L. M., Huo, F. Y., ... & Bu, L. L. (2023, July). Enhancing head and neck tumor management with artificial intelligence: Integration and perspectives. In Seminars in Cancer Biology. Academic Press.

[15] Islam, M. A., Majumder, M. Z. H., Hussein, M. A., Hossain, K. M., & Miah, M. S. (2024). A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. Heliyon.

[16] Zhong, N. N., Wang, H. Q., Huang, X. Y., Li, Z. Z., Cao, L. M., Huo, F. Y., ... & Bu, L. L. (2023, July). Enhancing head and neck tumor management with artificial intelligence: Integration and perspectives. In Seminars in Cancer Biology. Academic Press.

[17] Kouba, P., Kohout, P., Haddadi, F., Bushuiev, A., Samusevich, R., Sedlar, J., ... & Mazurenko, S. (2023). Machine learning-guided protein engineering. ACS catalysis, 13(21), 13863-13895.

[18] Shankarlal, B., Dhivya, S., Rajesh, K., & Ashok, S. (2024). A hybrid thyroid tumor type classification system using feature fusion, multilayer perceptron and bonobo optimization. Journal of X-Ray Science and Technology, (Preprint), 1-25.

[19] Aswad, S. A., & Sonuç, E. (2020, October). Classification of VPN network traffic flow using time related features on Apache Spark. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 1-8). IEEE.

[20] Akash, K. T., Usman, F. M., Kumar, T. N., Ahmed, M. R., & Gudodagi, R. (2023, August). Predicting Thyroid Dysfunction Using Machine Learning Techniques. In 2023 12th International Conference on Advanced Computing (ICoAC) (pp. 1-8). IEEE.

[21] Singh, T., Sahu, A. K., Dubey, S., Sharma, M. P., Verma, S., & Kumar, C. (2022). Treatment of thyroid disease through machine learning predictive model. International Journal of Health Sciences, 6(S8), 3176-3188.

[22] Vanitha, R., & Perumal, K. (2022). Hybrid Model of KNN and PCA to Pre-processor of Thyroid Dataset using Machine Learning. Journal of Pharmaceutical Negative Results, 1091-1096.

[23] Geerthana, P., Priyadharshini, A., & Anandhi, S. (2020). A SURVEY ON EARLY PREDICTION OF THYROID DISEASE USING DATA MINING TECHNIQUES.

[24] Alsaadawi, M., & Şehirli, E. (2022, October). The Efficiency of Ensemble Techniques in Predicting Thyroid Disorder: A Comparative Study. In 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 834-840). IEEE.

[25] Riajuliislam, M., Rahim, K. Z., & Mahmud, A. (2021, February). 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.

[26] Verma, S., & Rubin, J. (2018, May). Fairness definitions explained. In Proceedings of the international workshop on software fairness (pp. 1-7).

[27] Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015, August). Certifying and removing disparate impact. In proceedings of the 21 th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 259-268).

[28] Lee, N. T., Resnick, P., & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings Institute: Washington, DC, USA, 2.

[29] Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of global health, 9(2).

[30] Blattner, L., & Nelson, S. (2021). How costly is noise? Data and disparities in consumer credit. arXiv preprint arXiv:2105.07554.

[31] Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of global health, 9(2).

[32] Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. The Journal of Machine Learning Research, 11, 2079-2107.

[33] Tantithamthavorn, C., McIntosh, S., Hassan, A. E., & Matsumoto, K. (2018). The impact of automated parameter optimization on defect prediction models. IEEE Transactions on Software Engineering, 45(7), 683-711.