IJMSIT

International Journal of Multidisciplinary Studies and Innovative Technologies

e-ISSN: 2602-4888
dergipark.org.tr/en/pub/ijmsit
Research Article

2024, 8 (2), 59-62
DOI: 10.36287/ijmsit.8.2.2
Received: October 9, 2024; Accepted: November 28, 2024

Thyroid Disease Diagnosis: A Study on the Efficacy of Feature Reduction and Biomarker Selection in Artificial Neural Network Models

Erman ÖZER[1*]

[1] Computer Engineering Department, Recep Tayyip Erdoğan University, Rize, Turkey (erman.ozer@erdogan.edu.tr) (ORCID: 0000-0002-9638-0233)*

***Abstract* –**This study employs ANN to enhance thyroid disease diagnosis while minimizing features and choosing the most biomarkers. The data were analyzed focusing on three key indicators of thyroid function: TSH, TT4, and FTI. All of these biomarkers are vital signs that reflect thyroid activity and are incorporated in ANN models. This is achievable by minimizing the number of features and there by the Billboard ANN models deliver high diagnostic accuracy and high computational effectiveness. Computing with this simplified dataset results in faster computation times while at the same time, maintaining a high degree of diagnostic accuracy. Thus, the profound features of TSH, TT4, and FTI as indices of thyroid disorders, as well as the introduction of these markers into simple diagnostic algorithms, are discussed. The regression values achieved with the complete dataset were 0.66 for the training phase, 0.62 for validation, and 0.61 for testing. Conversely, utilizing the reduced dataset resulted in regression values of 0.67 during training, 0.99 in validation, and 0.80 in testing. Hence this study supports the application of ANN models in medical diagnosis by adding to the existing proof to the strategy. The data suggest that the exclusion of features can enhance the speed and boost the time to obtain a precise result. These improvements could have significant implications for clinical practice, especially in enhancing the management and treatment of thyroid diseases, where precise and prompt diagnosis is essential.

Keywords *–Thyroid; Disease Diagnosis; Artificial Neural Networks; Feature Reduction*

Citation: Özer, E. Thyroid Disease Diagnosis: A Study on the Efficacy of Feature Reduction and Biomarker Selection in Artificial Neural Network Models. International Journal of Multidisciplinary Studies and Innovative Technologies, 8(2): 59-62.

I. INTRODUCTION

Diseases affecting the thyroid, such as hypothyroidism and hyperthyroidism, are influenced by a complex interplay of physiological processes. Accurate diagnosis and effective treatment of these conditions are challenging without robust mathematical models to capture the intricacies of thyroid function. In recent years, Artificial Neural Networks (ANNs) have emerged as a powerful tool to model these complex systems, providing valuable insights into disease diagnosis and management [1-3].

Inspired by the structure and functioning of the brain's network of neurons, ANNs emulate biological learning by creating interconnected networks capable of identifying patterns, behaviors, and pathways in data. These networks are widely applied across various domains, including classification, prediction, and complex system modeling. In the context of thyroid disease, ANNs are particularly useful for predicting outcomes based on patient information, enabling early detection and personalized treatment planning [4,5].

An ANN typically comprises three main components: an input layer, one or more hidden layers, and an output layer. The input layer processes raw data such as hormone levels (e.g., TSH, T4, FT3) and imaging data (e.g., ultrasound images). This information is then transmitted through the hidden layers, where patterns and relationships relevant to thyroid regulation are identified. Finally, the output layer provides predictions or classifications, such as the likelihood of a specific thyroid disorder. Despite their utility, ANNs have limitations, particularly their "black box" nature, which makes it challenging to interpret their decision-making processes. This lack of transparency poses challenges in medical applications, where reproducibility and explainability are critical for clinical acceptance.

To address these challenges, researchers are exploring methods to improve ANN interpretability, such as integrating explainable AI techniques and incorporating domain-specific knowledge into model architectures. By overcoming these barriers, ANNs can become more reliable tools for advancing the diagnosis and treatment of thyroid disorders while maintaining the trust of healthcare professionals.

II. LITERATURE REVIEW

Research on diagnosing thyroid disorders using advanced computational techniques has shown significant progress. Irina and Liviu proposed a data mining technique to classify thyroid disorders, specifically hyperthyroidism and hypothyroidism[6]. Their research aims to assess the effectiveness of data mining in classifying thyroid diseases.

Similarly, Geetha and Baboo developed a Hybrid Differential Evolution Kernel-Based Naive Bayes algorithm, which reduced the dimensionality of thyroid data, achieving an impressive classification accuracy of 97.97% [7]. These findings underscore the role of sophisticated algorithms in enhancing diagnostic outcomes.

Further studies have validated the effectiveness of machine learning in thyroid disease diagnostics. Doğantekin et al. introduced the ADSTG (Automatic Diagnosis System based on the Thyroid Gland), which employed Principal Component Analysis (PCA) for dimensionality reduction and Least Squares Support Vector Machines (LS-SVM) for binary classification. This hybrid approach achieved a notable accuracy of 97.67%, illustrating the benefits of combining dimensionality reduction techniques with advanced classifiers [8].

Recent advances in deep learning have further pushed the boundaries of diagnostic performance. Yadav et al. reported 99.95% accuracy in detecting thyroid diseases, demonstrating the success of hybrid methods when deep-learning with intricate medical datasets[9]. Usman et al. noted the potential of artificial intelligence algorithms and multiple linear regression for predicting thyroid hormone balance, indicating a greater implication of AI in clinical diagnostics with this regard[10].

Despite these advancements, challenges such as the interpretability of complex models remain. This study addresses these gaps by employing ANNs in a novel configuration. While ANNs excel in identifying intricate patterns, their "black box" nature often limits their interpretability. To overcome this, the current research emphasizes improving both diagnostic accuracy and computational efficiency through advanced techniques.

Building on prior studies, this work utilizes ANN technology to classify thyroid disorders with enhanced accuracy. By leveraging the latest datasets and incorporating a larger sample size, the proposed model achieves significant improvements in computational efficiency. Although the diagnostic accuracy remains comparable to existing methods, the reduction in computational complexity marks a substantial step forward. This innovation contributes to early detection and better management of thyroid diseases, highlighting the transformative potential of ANNs in medical diagnostics.

Our study stands out from previous research by using the latest data and including a larger number of patients. Although the accuracy rate using similar feature reduction techniques stayed the same, our method significantly improved computational efficiency.

III. DATASET AND METHODOLOGY

The thyroid disease dataset, sourced from the UCI Machine Learning Repository, is a valuable resource for machine learning tasks related to thyroid health. The databases from the Garvan Institute include medical data from patients with thyroid disease, each containing around 2,800 training instances and 972 test instances[11]. In total, there are approximately 29 features in each database, classified as Boolean or continuous variables.

Quinlan offers databases that may have missing data. The Aeberhard database has 215 instances, 5 features, and 3 classes of medical data. The Turney database has 3,772 training instances, 3,428 test instances, 5 features, and 3 classes of both medical and cost data. This versatile dataset supports tasks such as diagnosing thyroid disease, assessing treatment efficacy, and cost-effectiveness analysis, making it valuable for clinical practice and research.

Table 1. Thyroid Disease Dataset

Feature	Type	Description
age	Continuous	Patient's age
sex	Binary	Patient's gender
on thyroxine	Binary	Whether the patient is taking thyroid hormone thyroxine
query on thyroxine	Binary	Whether the patient has inquired about thyroxine
on antithyroid medication	Binary	Whether the patient is taking medication to treat hyperthyroidism
sick	Binary	Whether the patient is currently sick
pregnant	Binary	Whether the patient is pregnant
thyroid surgery	Binary	Whether the patient has undergone thyroid surgery
I131 treatment	Binary	Whether the patient has received radioactive iodine treatment
query hypothyroid	Binary	Whether the patient has inquired about hypothyroidism
query hyperthyroid	Binary	Whether the patient has inquired about hyperthyroidism
lithium	Binary	Whether the patient is taking lithium medication
goiter	Binary	Whether the patient has a goiter
tumor	Binary	Whether the patient has a thyroid tumor
hypopituitary	Binary	Whether the patient has hypopituitarism
psych	Binary	Whether the patient has a psychiatric condition
referral source	Categorical	Source of referral for the patient
binary class	Binary	Whether the patient has thyroid disease

Thyroid disease is evaluated using key biomarkers such as Thyroid Stimulating Hormone (TSH), thyroxine (T4), and free triiodothyronine (FT3). Elevated TSH points to hypothyroidism, low TSH indicates hyperthyroidism, and high T4 and FT3 levels are associated with increased thyroid function. Refer to Table 2 in this study for more information on these specific biomarkers.

Table 2. Key Biomarkers and Their Clinical Significance İn Thyroid Disease Diagnosis

Feature	Type	Description
TSH measured	Measured TSH value	Input
T3 measured	Measured T3 value	Input
TT4 measured	Measured TT4 value	Input
T4U measured	Measured T4U value	Input
FTI measured	Measured FTI value	Input
binaryClass	Thyroid disease status: 0 = normal, 1 = disease present	Output

The formula for regression is provided as follows:

$$\beta_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

- X_i represents the individual values of the predictor variable X,
- $\bar{X}$ denotes the mean of X
- Y_i represents the individual values of the target variable Y
- $\bar{Y}$ denotes the mean of Y

IV. Results

TSH, total TT4, and FTI are important biomarkers for diagnosing thyroid disease because they accurately reflect thyroid gland activity. A model using these values demonstrated high accuracy in diagnosing thyroid disease, confirming the effectiveness of these biomarkers for clinical use. These results highlight the critical role of TSH, TT4, and FTI in evaluating thyroid function, offering valuable insights to enhance diagnostic accuracy.

The pituitary gland releases TSH to signal the thyroid to release T4 and T3. T4 is converted to the active form, FT3. The FTI measurement reflects levels of TT4 and T3. The best validation performance of the entire dataset (Fig. 1) and the ANN results (Fig. 2) were both obtained using the complete dataset.

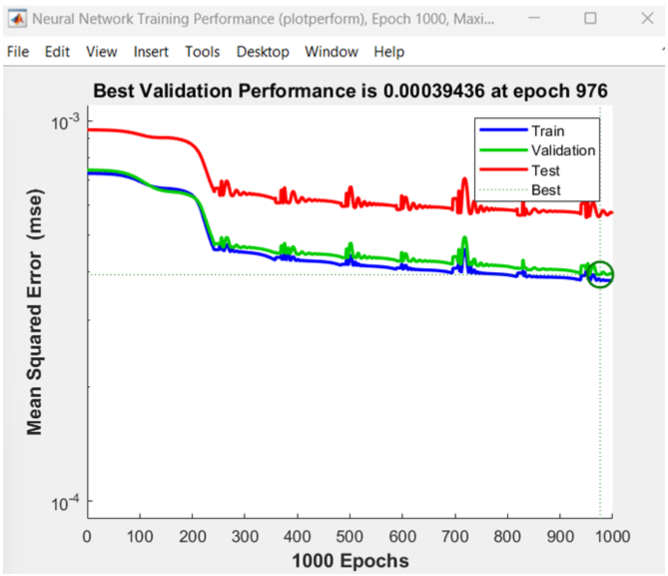

Fig. 1. Best Validation Performance of the Whole Dataset

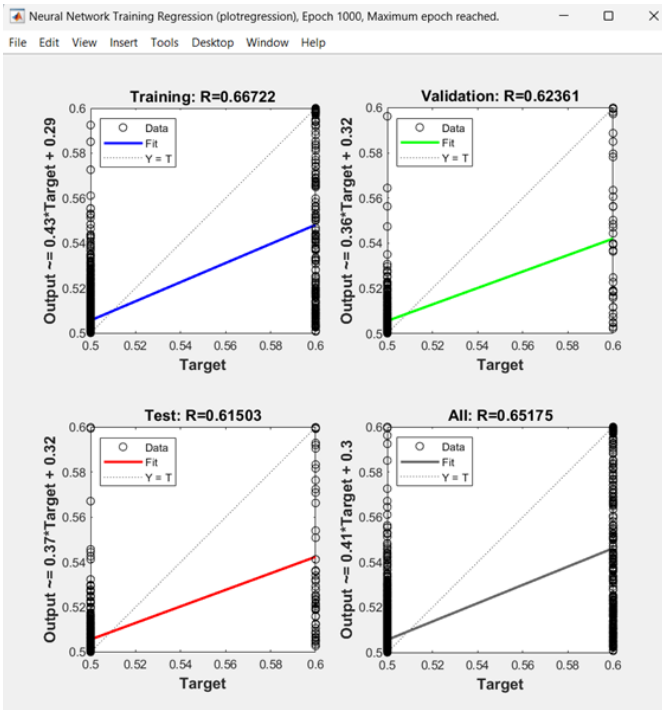

Fig. 2. ANN results of the whole dataset

Data selection was based on the correlation coefficient, which measures the relationship between variables. Strong correlations were found among TSH, TT4, and FTI values, suggesting a close connection among these biomarkers. This indicates that these variables collectively offer a thorough evaluation of thyroid function.

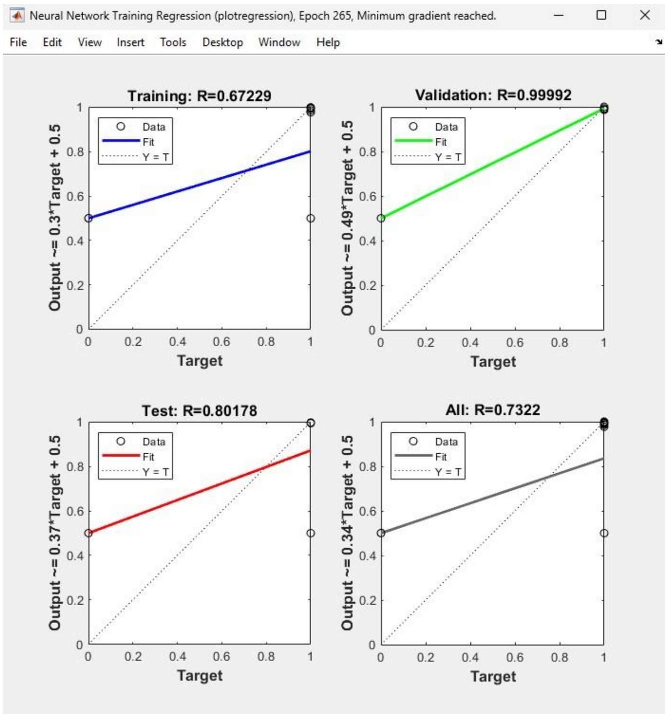

Fig. 3. ANN results of the reduced dataset

Reducing the dataset features led to improved results and saved significant time. The study achieved approximately 7% faster results.

V. CONCLUSION

This study highlights the effectiveness of ANNs in diagnosing thyroid diseases with improved computational efficiency and diagnostic accuracy. By leveraging key biomarkers such as TSH, TT4, and FTI, the proposed model demonstrates a significant enhancement in both performance and speed. Utilizing a reduced dataset further improved validation results, achieving a regression value of 0.99, and reduced computational time by approximately 7%, underscoring the practical benefits of feature reduction.

ANNs, inspired by the structure and functionality of the human brain, offer robust capabilities for modeling complex medical systems. They are invaluable tools in medical diagnostics, particularly in tasks such as classification and prediction, enabling early detection and effective management of conditions like hyperthyroidism and hypothyroidism. The strengths of ANNs include their ability to handle incomplete data, tolerate faults, and process information in parallel. However, their reliance on specialized hardware, unpredictable behavior, and the challenge of optimizing network architecture present areas for further research.

This study contributes to the growing body of evidence supporting the use of ANN models in clinical applications, with a focus on achieving a balance between computational efficiency and diagnostic accuracy. While ANN models exhibit limitations in interpretability, integrating explainable AI techniques in future work could address these challenges, further enhancing their clinical relevance. Overall, the findings of this study pave the way for more efficient and precise diagnostic tools, offering significant implications for improving the management and treatment of thyroid diseases.

VI. LIMITATIONS AND FUTURE WORK

This study is limited by the size and diversity of the dataset used, as it may not fully represent the broad range of thyroid disease cases encountered in clinical practice. Additionally, the demographic distribution of the dataset could introduce biases, potentially affecting the generalizability of the results. Another limitation is the "black box" nature of the ANN model, which restricts its interpretability and could hinder its acceptance in critical clinical settings where decision transparency is vital.

Future work could address these limitations by incorporating larger, more diverse datasets to improve model robustness and applicability. Exploring advanced techniques, such as explainable AI (XAI), could enhance the interpretability of the ANN model, making it more suitable for clinical decision-making. Real-time data integration, such as wearable device outputs or continuous monitoring systems, could further refine diagnostic capabilities and expand the model's utility. Lastly, extending the methodology to include multi-disease diagnostic capabilities would make the model more versatile and impactful for broader healthcare applications.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Authors' Contributions

The authors' contributions to the paper are equal.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The authors declare that this study complies with Research and Publication Ethics

REFERENCES

[1] M. A. -A. -R. Asif et al., "Computer Aided Diagnosis of Thyroid Disease Using Machine Learning Algorithms," 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2020, pp. 222-225, doi: 10.1109/ICECE51571.2020.9393054.

[2] A. R. Rao and B. S. Renuka, "A Machine Learning Approach to Predict Thyroid Disease at Early Stages of Diagnosis," 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangluru, India, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298252.

[3] M. Riajuliislam, K. Z. Rahim and A. Mahmud, "Prediction of Thyroid Disease(Hypothyroid) in Early Stage Using Feature Selection and Classification Techniques," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), Dhaka, Bangladesh, 2021, pp. 60-64, doi: 10.1109/ICICT4SD50815.2021.9397052.

[4] A. Begum and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 342-345, doi: 10.1109/ICACCS.2019.8728320.

[5] E. Özer, N. Sevinçkan and E. Demiroğlu, "Comparative Analysis of Computational Intelligence Techniques in Financial Forecasting: A Case Study on ANN and ANFIS Models," 2024 32nd Signal Processing and Communications Applications Conference (SIU), Mersin, Turkiye, 2024, pp. 1-4, doi: 10.1109/SIU61531.2024.10600769.

[6] A. Begum and A. Parkavi, "Prediction of thyroid Disease Using Data Mining Techniques," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 342-345, doi: 10.1109/ICACCS.2019.8728320.

[7] K. Geetha and C. S. S. Baboo, "An Empirical Model for Thyroid Disease Classification using Evolutionary Multivariate Bayesian Prediction Method", Glob. J. Comput. Sci. Technol. E Network, Web Secur., 16:1, 242-250.

[8] Esin Dogantekin, Akif Dogantekin, Derya Avci, An automatic diagnosis system based on thyroid gland: ADSTG, Expert Systems with Applications, Volume 37, Issue 9, 2010, Pages 6368-6372, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2010.02.083.

[9] Yadav, D.C., Pal, S. Prediction of thyroid disease using decision tree ensemble method. Hum.-Intell. Syst. Integr. 2, 89–95 (2020). https://doi.org/10.1007/s42454-020-00006-y

[10] Usman, Abdullahi & Alhosen, Mohamed & Degm, Ali & Alsharksi, Ahmed & Muhammed Naibi, Aishat & Abba, Sani & Muhammad, Umar Ghali. (2020). Applications of Artificial Intelligence-Based Models and Multi- Linear Regression for the Prediction of Thyroid Stimulating Hormone Level in the Human Body.

[11] Quinlan,Ross. Thyroid Disease. UCI Machine Learning Repository. https://doi.org/10.24432/C5D010.

[12] Sheehan MT. Biochemical Testing of the Thyroid: TSH is the Best and, Oftentimes, - A Review for Primary Care. Clin Med Res. 2016 Jun;14(2):83-92. doi: 10.3121/cmr.2016.1309. Epub 2016 May 26. PMID: 27231117; PMCID: PMC5321289.

[13] Feldt-Rasmussen U, Klose M. Clinical Strategies in the Testing of Thyroid Function. [Updated 2020 Nov 20]. In: Feingold KR, Anawalt B, Blackman MR, et al., editors. South Dartmouth (MA): ncbi.nlm.nih.gov/books/NBK285558/