International Journal of Artificial Intelligence in Medical Issues

Volume 3 Issue 1
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v3i1.250

Research Article

Predicting Thyroid Cancer Recurrence After Radioactive Iodine Therapy Using Random Forest and Neural Network Models

Rosmasari [1,*]
[1] Universitas Mulawarman, Samarinda, Kalimantan Timur 75119, Indonesia, rosmasari@unmul.ac.id
Correspondence should be addressed to Rosmasari; rosmasari@unmul.ac.id

Received 15 January 2025; Revised 08 February 2025; Accepted 25 April 2025; Published 30 May 2025

Abstract:
Thyroid cancer recurrence following Radioactive Iodine (RAI) therapy remains a clinical concern, necessitating accurate and timely risk prediction to guide post-treatment management. This study aims to evaluate the effectiveness of machine learning models—Random Forest and Neural Networks—in predicting recurrence using a structured clinical dataset consisting of 383 patient records and 13 diagnostic and pathological attributes. All categorical features were encoded ordinally, and the dataset was partitioned into training and testing sets with appropriate normalization for neural network processing. Both models were evaluated using standard metrics including accuracy, precision, recall, and F1-score. The Random Forest model achieved an accuracy of 97.39%, outperforming the Neural Network which recorded 93.04%. Moreover, Random Forest showed better recall in detecting recurrence cases, making it a more suitable model for clinical application. These results demonstrate that machine learning, particularly ensemble-based methods, can offer a practical and interpretable solution for recurrence prediction, supporting data-driven decision-making in thyroid cancer follow-up care.

Keywords: Machine Learning, Neural Network, Radioactive Iodine, Random Forest, Thyroid Cancer.
Dataset link: https://www.kaggle.com/datasets/shahriar26s/benign-prostate-hyperplasiabph-detection

1. Introduction

Thyroid cancer is one of the most prevalent endocrine malignancies globally, with papillary thyroid carcinoma (PTC) being the most common subtype. Despite generally favorable prognoses and the effectiveness of initial treatments such as thyroidectomy followed by radioactive iodine (RAI) therapy, a significant proportion of patients experience cancer recurrence. The potential for recurrence necessitates long-term monitoring and effective predictive strategies to inform clinical decision-making. Traditionally, risk stratification has relied on the American Thyroid Association (ATA) guidelines, which classify patients into low, intermediate, or high-risk categories based on clinicopathologic factors. However, these stratification methods often lack individual specificity and can be subjective in real-world applications.

In recent years, machine learning (ML) has emerged as a promising approach to enhance recurrence prediction by leveraging complex, multidimensional data. Unlike traditional statistical models, ML algorithms can learn from patterns within large datasets and incorporate numerous features to yield more accurate and personalized predictions [1], [2]. Nonetheless, the application of ML in thyroid cancer recurrence prediction is still developing, with various studies employing different methodologies, datasets, and evaluation criteria. This variance highlights the need for a standardized, reproducible framework that demonstrates the practical utility of ML in clinical oncology [3], [4].

Numerous studies have explored the application of ML in this domain, with notable success. [5] conducted a comprehensive study involving 2,244 PTC patients and demonstrated that Random Forest (RF) and Neural Network (NN) models outperformed conventional ATA models, achieving AUC scores between 0.738 and 0.767, with RF offering superior calibration and interpretability. Similarly, [6] applied several ML classifiers and found RF achieving 98.18% accuracy, underscoring the importance of feature correlation analysis in improving model performance. Further, [7] emphasized the potential of early post-RAI ML-based assessments to predict treatment responses effectively. Meanwhile, [8] explored the integration of 15 years of clinicopathologic data and validated the usability of decision trees, RF, and logistic regression in real-time clinical settings. These advancements collectively affirm the strong potential of RF and NN models as practical tools for post-treatment recurrence monitoring.

Despite these encouraging developments, many prior studies utilize large and often institution-specific datasets, which are not always publicly accessible or reproducible. Furthermore, the interpretability of more complex models, such as deep learning frameworks, can pose challenges in clinical adoption. Thus, there remains a critical gap in the literature regarding the development of interpretable, efficient, and scalable ML models using smaller, structured clinical datasets—especially those that can be openly shared and implemented in varied clinical environments [9], [10].

This study aims to address that gap by developing and comparing the performance of Random Forest and Neural Network models in predicting thyroid cancer recurrence using a publicly available dataset comprising 383 patients with 13 clinical and pathological features. The main objectives are: (1) to pre-process and encode categorical clinical data effectively for ML input; (2) to train and evaluate RF and NN classifiers; and (3) to identify which model offers superior performance in terms of accuracy and generalization for recurrence prediction. Through this approach, the research seeks to contribute to the growing body of knowledge on ML applications in oncology and provide a practical foundation for future real-time, data-driven decision support tools.

2. Method

Research Design:

This study utilizes a supervised machine learning classification approach to predict the recurrence of thyroid cancer in patients who have undergone Radioactive Iodine (RAI) therapy. The research workflow consists of data pre-processing, feature encoding, data normalization, model training, and performance evaluation as in **Figure 1**. Two machine learning algorithms—Random Forest (RF) and Neural Network (Multi-Layer Perceptron, MLP)—are employed due to their effectiveness in handling structured medical data and nonlinear classification tasks [11]–[13].

.

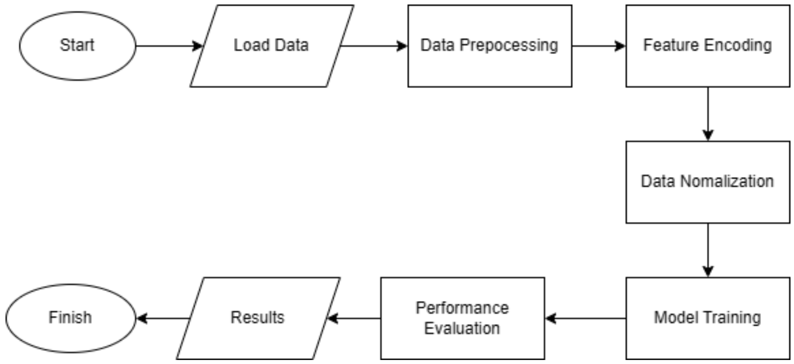

Figure 1: Research Workflow for Thyroid Cancer Recurrence Prediction Using Machine Learning

Dataset and Feature Selection:

The dataset comprises 383 records of thyroid cancer patients and includes 13 attributes encompassing demographic, clinical, pathological, and treatment response features. The target variable is cancer recurrence (`Recurred`), which is binary (Yes or No). The predictor variables include:

a Demographics: Age, Gender
b Clinical history: History of Radiotherapy, Adenopathy, Focality
c Pathological features: Pathology type, Tumor staging (T, N, M), Stage, Risk classification
d Treatment outcome: Response to RAIRandom horizontal flipping

No missing values were found, allowing for direct use in modeling.

Data Pre-processing:

All categorical attributes were converted into numerical values using ordinal encoding, where the order of categories reflects their clinical or biological severity. For instance, staging values such as Stage I, II, III, and IV were encoded in ascending order, as were treatment response categories from "Excellent" to "Biochemical Incomplete". This transformation ensures that the machine learning models can interpret categorical distinctions appropriately.

Data Partitioning and Scaling:

The dataset was partitioned into training (70%) and testing (30%) subsets using a random sampling approach to ensure representativeness. For the neural network model, numerical standardization was applied using Z-score normalization, which transforms features to have zero mean and unit variance. This step enhances convergence and stability of the neural network during training. The Random Forest model, being scale-invariant, was trained on the original unscaled data.

Machine Learning Models

a. Random Forest (RF) is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions [14], [15]. Its strengths include robustness to noise, ability to handle categorical data, and high interpretability through feature importance [16]–[18].

b. Neural Network (MLP Classifier) is a feedforward artificial neural network with one hidden layer. It is capable of modelling complex nonlinear relationships between features and the output class [19]–[21]. The model was trained with a maximum iteration threshold and optimized using a backpropagation algorithm [22], [23].

Performance Evaluation

Model performance was assessed on the test set using several classification metrics [24]–[26]:

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recal

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F1-Score

$$\text{F1-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

TP: True Positive,

TN: True Negative,

FP: False Negative,

FN: False Negative.

Additionally, confusion matrices were generated for each model to visualize prediction errors, and a bar chart comparison was used to illustrate differences in model accuracy. These evaluations provide a comprehensive view of model reliability and generalizability.

3. Result and Discussion

This section presents the comparative performance of the Random Forest (RF) and Neural Network (NN) models in predicting thyroid cancer recurrence after RAI therapy. The models were evaluated on a test set comprising 115 patient records.

Random Forest Performance

The Random Forest model achieved an impressive accuracy of 97.39%, with a precision of 0.98 and recall of 0.99 for the non-recurrence class, and a precision of 0.97 and recall of 0.94 for the recurrence class. The overall macro average F1-score was 0.97, indicating high performance across both classes.

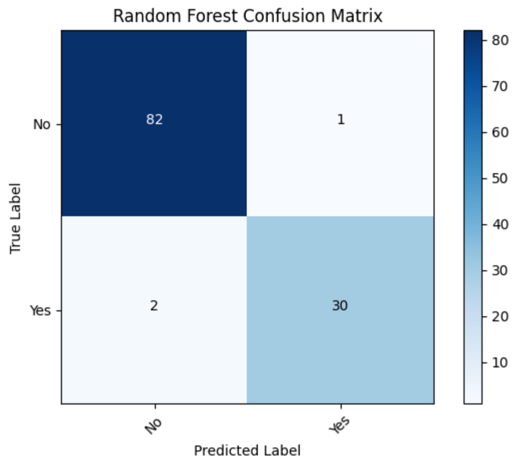

Figure 2: Confusion Matrix for Random Forest

The **Figure 2** illustrates that the model correctly classified 82 non-recurrence cases and 30 recurrence cases, with only 3 misclassifications in total (1 false positive and 2 false negatives). This reflects excellent sensitivity and specificity, which is critical in a medical context where false negatives may delay necessary interventions.

Neural Network Performance

The Neural Network model yielded a slightly lower accuracy of 93.04%. It showed good precision for the non-recurrence class (0.94) and moderate performance for the recurrence class with precision 0.90 and recall 0.84. The macro average F1-score was 0.91, suggesting that while the model performs adequately, its recall for the recurrence class was notably lower than the RF model.

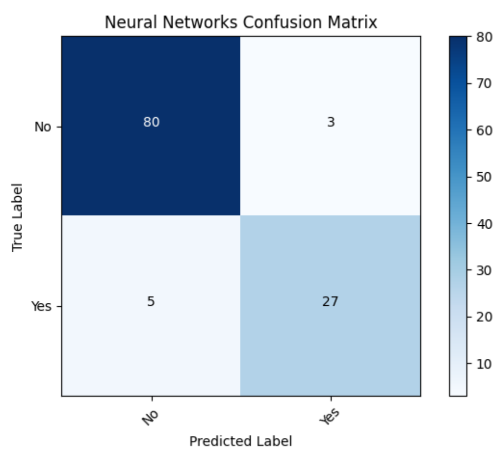

Figure 3: Confusion Matrix for Neural Networks

As depicted in the **Figure 3**, 80 non-recurrence cases and 27 recurrence cases were correctly classified. However, 8 instances were misclassified—3 false positives and 5 false negatives—indicating a greater tendency to under-predict recurrence compared to RF.

Model Comparison

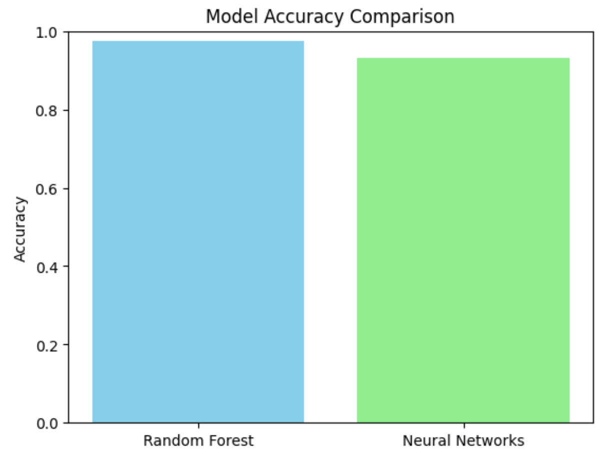

Figure 4: The comparison of model accuracies

Figure 4 highlights the superiority of Random Forest in terms of classification accuracy. Additionally, **Figure 5** demonstrates effective learning and convergence across 300 iterations, although performance plateaued after approximately 200 iterations, suggesting limited further gain from additional training epochs.

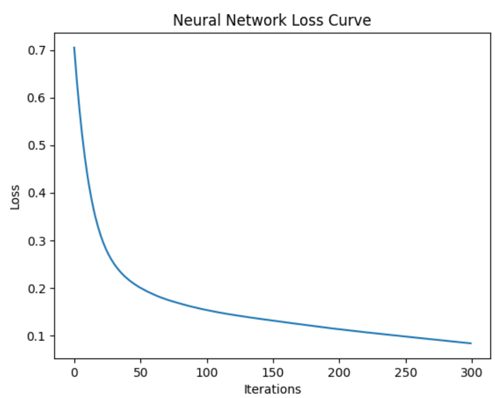

Figure 5: The loss curve of the neural network

Table 1. Comparative evaluation metrics between Random Forest and Neural Networks

Metric	Random Forest	Neural Networks
Accuracy	97.39%	93.04%
Precision (class 1)	0.97	0.9
Recall (class 1)	0.94	0.84
F1-score (class 1)	0.95	0.87

Discussion

The experimental results reveal that the Random Forest model outperforms the Neural Network model across all key evaluation metrics, particularly in sensitivity to recurrence prediction. This is crucial given that accurate early detection of recurrence can significantly improve patient prognosis through timely intervention. The superior performance of Random Forest can be attributed to its ensemble nature, which effectively reduces variance and handles categorical features with greater robustness. Its interpretability also allows for the extraction of feature importance, a valuable asset in clinical settings for explaining model decisions.

On the other hand, although Neural Networks demonstrated satisfactory overall accuracy and loss minimization, their relatively lower recall for the recurrence class suggests that deeper or more complex architectures might be needed, possibly in combination with feature engineering or class imbalance handling techniques. These findings align with recent literature, where Random Forest is frequently reported as a reliable and interpretable classifier in medical data analysis. Moreover, the current results validate that even with a relatively small and structured dataset, classical machine learning algorithms can provide high-performing predictive models.

4. Conclusion

This study investigated the application of machine learning techniques, specifically Random Forest and Neural Networks, in predicting thyroid cancer recurrence following Radioactive Iodine (RAI) therapy. By employing a structured dataset consisting of 383 patient records with 13 relevant clinical and pathological features, the research demonstrated that both models were capable of producing reliable predictive outcomes. Among the two, the Random Forest model achieved the highest performance, recording an accuracy of 97.39%, while the Neural Network model achieved 93.04%. The Random Forest classifier also demonstrated superior precision and recall in identifying both recurrence and non-recurrence cases, making it a more effective model in this specific clinical prediction task.

These results highlight the strength of ensemble learning methods like Random Forest in handling clinical datasets, which often contain mixed data types and potentially complex inter-variable relationships. The model's ability to maintain high performance, even with a relatively small dataset, underscores its practical value in real-world medical settings where data availability may be limited. Furthermore, this study successfully established a complete and reproducible pipeline—from data pre-processing and encoding to model evaluation using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrices.

In conclusion, the study confirms that machine learning models, particularly Random Forest, offer promising tools for enhancing post-therapy monitoring of thyroid cancer patients by accurately predicting the risk of recurrence. This predictive capability can support clinicians in making informed decisions regarding follow-up care and intervention strategies. Future research may consider expanding the dataset, applying explainable AI techniques to improve model transparency, and exploring time-to-event predictive models for estimating recurrence intervals. The insights provided by this study contribute to the growing body of work that seeks to integrate machine learning into evidence-based oncology practices.

References:

[1] K. Bharath, "Predicting Recurrence in Differentiated Thyroid Cancer: A Machine Learning Approach," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024*. 2024, doi: 10.1109/ADICS58448.2024.10533649.

[2] M. A. S. Ahmad, "An Explainable AI Model for Predicting the Recurrence of Differentiated Thyroid Cancer," *2024 2nd Jordanian International Biomedical Engineering Conference, JIBEC 2024*. pp. 84–89, 2024, doi: 10.1109/JIBEC63210.2024.10932125.

[3] M. G. Alfarizi, "Optimized Random Forest Model for Remaining Useful Life Prediction of Experimental Bearings," *IEEE Trans. Ind. Informatics*, vol. 19, no. 6, pp. 7771–7779, 2023, doi: 10.1109/TII.2022.3206339.

[4] N. Aida, "Comparison of Extreme Machine Learning and Hidden Markov Model Algorithm in Predicting The Recurrence Of Differentiated Thyroid Cancer Using SMOTE," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 6, no. 4, pp. 429–444, 2024, doi: 10.35882/jeeemi.v6i4.467.

[5] H. Wang *et al.*, "Development and validation of prediction models for papillary thyroid cancer structural recurrence using machine learning approaches," *BMC Cancer*, vol. 24, no. 1, p. 427, Apr. 2024, doi: 10.1186/s12885-024-12146-4.

[6] D. Iqbal and A. Shahzad, "Prediction of thyroid cancer recurrence with machine learning models," *Pakistan J. Nucl. Med.*, pp. 49–55, 2024, doi: 10.24911/PJNMed.175-1721068107.

[7] O. Bülbül and D. Nak, "Predicting excellent response to radioiodine in differentiated thyroid cancer using machine learning," *Acta Otorhinolaryngol. Ital.*, vol. 44, no. 4, pp. 261–268, Aug. 2024, doi: 10.14639/0392-100X-N3029.

[8] B. K and A. Sai Sabitha, "Predicting Recurrence in Differentiated Thyroid Cancer: A Machine Learning Approach," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, Apr. 2024, pp. 1–6, doi: 10.1109/ADICS58448.2024.10533649.

[9] Y. Camgözlü, "Leaf Image Classification Based on Pre-trained Convolutional Neural Network Models," *Nat. Eng. Sci.*, vol. 8, no. 3, pp. 214–232, 2023, doi: 10.28978/nesciences.1405175.

[10] L. Lei, "An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction," *Energy*, vol. 288, 2024, doi: 10.1016/j.energy.2023.129795.

[11] S. M. Simon, "Interpreting random forest analysis of ecological models to move from prediction to explanation," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-30313-8.

[12] S. Sharma, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," *Procedia Computer Science*, vol. 218. pp. 357–366, 2022, doi: 10.1016/j.procs.2023.01.018.

[13] H. Bhatt, "A Convolutional Neural Network ensemble model for Pneumonia Detection using chest X-ray images," *Healthc. Anal.*, vol. 3, 2023, doi: 10.1016/j.health.2023.100176.

[14] U. K. Lilhore, "Hybrid model for precise hepatitis-C classification using improved random forest and SVM method," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-36605-3.

[15] L. Hua, "An evolutionary deep learning soft sensor model based on random forest feature selection technique for penicillin fermentation process," *ISA Trans.*, vol. 136, pp. 139–151, 2023, doi: 10.1016/j.isatra.2022.10.044.

[16] S. Das, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," *Expert Syst. Appl.*, vol. 213, 2023, doi: 10.1016/j.eswa.2022.118914.

[17] A. Elbeltagi, "Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models," *Environ. Sci. Pollut. Res.*, vol. 30, no. 15, pp. 43183–43202, 2023, doi: 10.1007/s11356-023-25221-3.

[18] F. Huang, "Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models," *Int. J. Coal Sci. Technol.*, vol. 10, no. 1, 2023, doi: 10.1007/s40789-023-00579-4.

[19] G. Ali, "A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification From Fundus Images," *IEEE J. Transl. Eng. Heal. Med.*, vol. 11, pp. 341–350, 2023, doi: 10.1109/JTEHM.2023.3282104.

[20] M. Gupta, "A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network," *Knowledge-Based Syst.*, vol. 259, 2023, doi: 10.1016/j.knosys.2022.110070.

[21] S. Sharma, "A Deep Learning Model for Early Prediction of Pneumonia Using VGG19 and Neural Networks," *Lecture Notes in Networks and Systems*, vol. 588. pp. 597–612, 2023, doi: 10.1007/978-981-19-7982-8_50.

[22] X. Zhao, "Model-Assisted Multi-source Fusion Hypergraph Convolutional Neural Networks for intelligent few-shot fault diagnosis to Electro-Hydrostatic Actuator," *Inf. Fusion*, vol. 104, 2024, doi: 10.1016/j.inffus.2023.102186.

[23] S. Suneel, "Quantum mesh neural network model in precise image diagnosing," *Opt. Quantum Electron.*, vol. 56, no. 4, 2024, doi: 10.1007/s11082-023-06245-y.

[24] A. Zermane, "Predicting fatal fall from heights accidents using random forest classification machine learning model," *Saf. Sci.*, vol. 159, 2023, doi: 10.1016/j.ssci.2022.106023.

[25] D. M. Ge, "Estimation of rapid chloride permeability of SCC using hyperparameters optimized random forest models," *J. Sustain. Cem. Mater.*, vol. 12, no. 5, pp. 542–560, 2023, doi: 10.1080/21650373.2022.2093291.

[26] P. Celard, "A survey on deep learning applied to medical images: from simple artificial neural networks to generative models," *Neural Computing and Applications*, vol. 35, no. 3. pp. 2291–2323, 2023, doi: 10.1007/s00521-022-07953-4.