Journal of Science and Technique - ISSN 1859-0209, June-2025

MACHINE LEARNING-BASED DOSE RECOMMENDATIONS FOR REMOVAL OF RESIDUAL THYROID TISSUE

Thanh Trung Nguyen[1,*], *Thi Thu Ha Ngo*[1], *Phu Minh Lai*[1]

Abstract

Thyroid cancer, particularly differentiated types such as papillary and follicular carcinomas, presents a significant health challenge globally and in Vietnam, where surgical removal of the thyroid gland (thyroidectomy) is the primary treatment approach. However, residual thyroid tissue often remains post-surgery, necessitating effective ablation to prevent recurrence and complications. Radioactive iodine therapy using I-131 is the standard method for ablating this residual tissue. The accuracy of I-131 dose estimation plays a crucial role in ensuring therapeutic success and patient safety. Traditional methods rely on general guidelines and physician expertise but may lack the precision necessary for individual patients. While several previous studies have attempted to improve dose estimation using machine learning approaches, they have often overlooked the importance of data preprocessing techniques. This lack of attention to data quality has limited the performance of predictive models. To address this problem, our study expands upon prior work by placing a strong emphasis on data preprocessing, aiming to enhance model accuracy and reliability. We implement a multi-stage framework that processes medical records and identifies key features, using seven traditional machine learning models for prediction. Our results demonstrate that the Decision Tree model outperforms other models, achieving the highest True Positive Rate of 0.995, a low False Positive Rate of 0.001, and exceptional performance in Recall, Precision, and F1-score (0.995, 0.996, and 0.995, respectively). Moreover, we developed decision rules generated by the Decision Tree for dose prediction, which use a dataset of clinical information from Vietnamese patients. This model represents a promising tool for improving radiation therapy delivery, ensuring more accurate, data-driven decisions, and ultimately better patient outcomes.

Index terms

Residual thyroid tissue; machine learning; decision rules; dose estimation; Vietnamese patient dataset.

1. Introduction

Thyroid cancer, particularly differentiated types such as papillary and follicular thyroid carcinomas, ranks among the most prevalent cancers globally, and poses a

---

[1] Medical Equipment Department, 108 Military Central Hospital

[*] Corresponding author. email: trung.ntc10@benhvien108.vn

DOI: 10.56651/lqdtu.jst.v14.n01.1046.ict

significant public health concern in Vietnam [1], [2]. Surgical removal of the thyroid gland (thyroidectomy) is the cornerstone of treatment, but thyroidal residual tissue often remains post-surgery [3]. This residual tissue, if not adequately addressed, can increase and lead to disease recurrence or severe complications. Effective ablation of residual thyroid tissue is [4], therefore, a critical step in comprehensive thyroid cancer management, as it enhances therapeutic outcomes, improves survival rates, and supports long-term recovery [5]. Radioactive iodine therapy using I-131 is the standard method for ablating residual thyroid tissue. I-131 selectively accumulates in thyroid cells [6], allowing targeted destruction of cancerous and remnant thyroid tissues without the need for additional surgical interventions. This approach minimizes patient discomfort, reduces recovery time, and lowers healthcare costs by avoiding prolonged hospital stays. Despite its widespread adoption, the success of I-131 therapy largely depends on the accurate estimation of the appropriate dose for each patient. The process of determining the optimal I-131 dose is challenging due to the variability in patient-specific factors, including tumor size, health status, and clinical characteristics [7].

Current methods for dose estimation rely on general guidelines and physician expertise, which, while effective in many cases, may lack precision for individual patients. Inadequate dosing can leave residual cancerous cells untreated, increasing the risk of recurrence, while excessive dosing may result in adverse effects such as damage to healthy tissues and long-term complications. Therefore, improving the precision of I-131 dose estimation is crucial for achieving better therapeutic outcomes and ensuring patient safety. As artificial intelligence (AI) advances [8]–[10], AI-based approaches offer promising solutions to these challenges. Machine learning models are capable of analyzing large-scale clinical datasets to identify patterns and predict individual responses to specific doses of I-131 [11]–[14].

A previous study [15] within the same research project used this dataset to apply machine learning techniques for predicting I-131 ablation doses. It mainly applied basic preprocessing techniques such as Min-Max normalization for quantitative variables and one-hot encoding for categorical features; the results achieved were relatively modest. Our study significantly expands this by integrating advanced data preprocessing, including imputation for missing values, feature selection, and clustering-based outlier removal, leading to higher model accuracy and more robust dose predictions. The study address overlooked challenges such as data heterogeneity and label inconsistency, and demonstrate the impact of preprocessing on model performance. The tool is built upon a multistage framework designed to process patients medical records through sequential steps. In summary, our contributions are as follows:

- A self-collected dataset from the hospital was utilized, containing detailed clinical information of Vietnamese patients, including age, sex, cancer stage, blood test results, iodine uptake capacity, tumor size, and other critical indicators.
- A key set of predictive features was identified to enhance model accuracy and

support data-driven decision-making, thereby contributing to the precise and effective delivery of radiation therapy.
- Data preprocessing techniques, including imputation, feature selection, and clustering, were applied to improve data quality prior to model development.
- Decision rules for dose prediction were derived using a Decision Tree algorithm, providing an interpretable and clinically relevant framework for therapy planning.

2. Background and related work

2.1. Machine learning for tabular data classification

Classification with tabular data involves assigning data samples organized in rows and columns to predefined categories, such as diagnosing patients as diabetic or non diabetic based on features like age, blood pressure, and glucose levels. Traditional machine learning algorithms such as k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Decision Tree (DT), and Logistic Regression (LR) are well-suited for this task due to their efficiency, ease of use, and strong performance on structured data. While deep learning excels with unstructured data, its advantages are less clear for tabular datasets and often come with higher computational costs. As a result, traditional methods are typically preferred for medical and other structured tabular data applications.

2.2. Data imputation

Missing data refers to situations where one or more values in a dataset are not recorded or unavailable. This can result from data collection errors, technical limitations, or unavailability of information at the time of recording. It's a common issue in large, complex datasets, especially in healthcare, where data comes from various sources and may be incomplete or inconsistent. In healthcare, missing data may occur due to incomplete patient disclosures, skipped tests (because of cost or time), or privacy and data-sharing restrictions between institutions. It can significantly impact machine learning models, leading to biased or inaccurate predictions. Models trained on incomplete data may fail to generalize well to new cases, especially in critical areas like healthcare where high accuracy is vital. To handle missing data, a simple approach is to check the percentage of missing values. If a feature has more than 50% missing, it is often removed. Otherwise, imputation methods like mean imputation, mode imputation.

Mean imputation is a method of handling missing data by replacing missing values with the mean of that variable across other observations. For example, if the attribute represents a patient's blood glucose level and the value is missing, it can be replaced with the average blood glucose level of other patients. This method is simple but may reduce the natural variability of the data and introduce bias into the affected features.

Mode imputation replaces missing values with the most frequent value of that variable. This approach is commonly used for categorical variables. For instance, if information about the type of medication a patient is using is missing, it can be filled in with the most commonly used medication among other patients. Mode imputation tends to preserve variability better than mean imputation, but can still reduce attribute diversity.

2.3. Feature ranking

Feature ranking is a critical process in machine learning that evaluates and prioritizes the importance of features in a dataset. Its primary goal is to identify the features that have the greatest impact on the target variable, improving the overall performance of the model. By conducting feature ranking, users can reduce complexity, avoid overfitting, and enhance model efficiency by selecting only the most relevant features.

Various methods are available for feature ranking, including statistical techniques, machine learning algorithms, wrapper methods, and embedded methods. Statistical methods, such as correlation measures and t-tests, are simple but may not capture complex relationships. Machine learning algorithms like DT and Random Forest (RF) can provide more accurate assessments of feature importance, while wrapper methods test different feature combinations to find the best model. Embedded methods combine feature selection with model training for more efficient optimization.

By using feature ranking, models become more efficient, interpretable, and easier to understand. This is particularly valuable in fields like healthcare, where decisions have significant consequences, and in finance, where model transparency is critical. Additionally, feature ranking uncovers hidden relationships within the data, opening opportunities for further research and improving predictive performance.

2.4. Interpretable machine learning

The interpretability of machine learning models refers to the ability to understand how a model makes decisions and the reasons behind its outputs. This aspect is becoming increasingly important, especially in critical fields such as healthcare, finance, and law, where decisions can significantly impact human lives. In healthcare, interpretability plays a crucial role as the accuracy and reliability of decisions directly affect patient outcomes. It enables healthcare professionals to understand why a model suggests a specific diagnosis or treatment plan, allowing them to combine this information with clinical experience for final decision-making. Moreover, interpretability helps detect and address biases that may arise from non-representative data, ensuring that models perform fairly and effectively across diverse populations.

Decision-making is represented as a branching tree, where each branch reflects a decision based on a specific attribute, making it easy to trace how the model reaches its conclusions. DT are constructed through a process that splits data based on their attributes to optimize prediction accuracy. Two important factors in this process are entropy and information gain. Entropy measures the level of uncertainty or disorder in a dataset, while information gain indicates the reduction in entropy when a particular attribute is used for splitting. The algorithm selects the attribute with the highest information gain to perform the split, and this process is repeated until a complete DT is formed, capable of making precise predictions based on the analyzed attributes.

2.5. Related work

Numerous studies are currently exploring the use of AI to optimize radioactive iodine therapy doses. To ascertain the maximum allowable activity for I-131 therapy, researchers in [12] examined data from 83 adult patients who received I-131 dosimetry between 1996 and 2023. Eight input variables were used in this study: age, patient weight, corresponding blood sample counts at 4 hours, 24 hours, and 48 hours, and normalized anterior/posterior counts, and then using a deep learning neural network to forecast the maximum allowable activity. The paired t-tests were used in a confidence interval 95% to compare the predicted and actual values based on medical internal radiation dose for the test cases, showing positive results with no significant differences. Ri Sa *et al.* [13] also create machine learning models to forecast how patients with differentiated thyroid cancer who do not have structural disease would react to treatment with I-131 therapy and thyrotropin suppression therapy (TSH). Using six algorithms—patients were used to train the models for I-131 and TSH treatments, respectively. It uses Shapley Additive exPlanations to assess how significant elements contributed to a machine learning algorithm's anticipated value. When it came to forecasting successful reactions to both treatments, the RF model performed the best, exhibiting excellent accuracy and Area Under the Curve values.

Additionally, using external radiation dosimetry data at different distances of 1 m, 0.3 m, and 0.1 m, the authors in [14] look for the best model to forecast the real radioactivity of high-dose I-131. To train, they used the hold-out technique to split the data in a 7:3 ratio (609 for the training set and 259 for the test set). They use a variety of machine learning algorithms and processing techniques. To evaluate the models, they calculated R2 (0.99), mean square error (79.10), mean absolute error (6.65), and root mean square error (8.89). The RF model demonstrated the strongest prediction performance. Furthermore, to create an internal dosimetry prediction toolkit for pediatric nuclear medicine, authors [11] presents an approach that uses AI algorithms and Monte Carlo simulations to estimate doses based on individual anatomy accurately. It entailed building a dosimetry database for five radiopharmaceuticals and 28 pediatric models. This data was used to train and test machine learning regression models, which were then optimized for maximum prediction power. With high accuracy (> 90%) and quick findings (< 2 seconds), the toolkit estimates absorbed dose rates in 30 organs. A notable prior study closely related to our work is that of Lai *et al.* [15], which explored the use of machine learning techniques for recommending I-131 ablation doses in patients with differentiated thyroid cancer after thyroidectomy. Using the same clinical dataset, the authors applied a set of traditional preprocessing steps, including Min-Max normalization for continuous variables and one-hot encoding for categorical features. The study developed and evaluated several classic machine learning models, such as DT, k-NN, and Naive Bayes (NB) to predict the recommended dose category based on patient features. While this study confirmed the potential of AI in supporting dose decisions and yielded promising initial results, its preprocessing pipeline was relatively limited, and challenges such as missing data, outlier handling, and label inconsistencies were not fully addressed. Some studies highlight significant improvements in treatment outcomes across different patient populations.

3. Proposed method

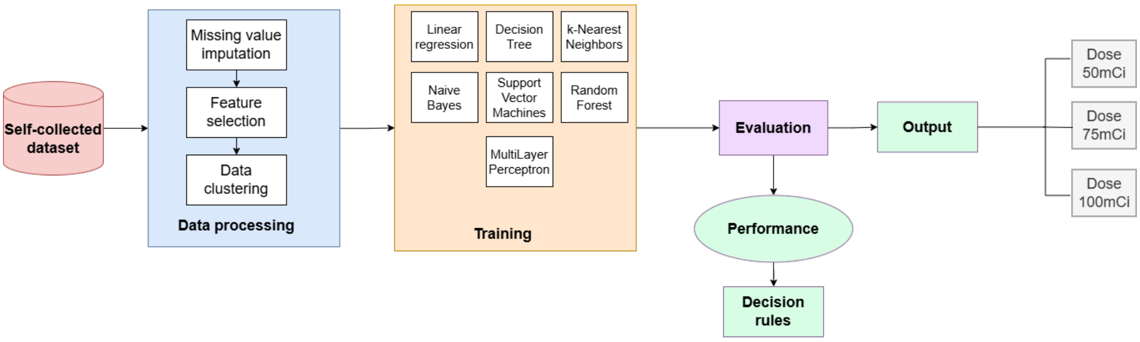

Fig. 1. Flowchart.

We propose a framework with primary stages, as shown in Figure 1. The flowchart illustrates a structured process for predicting optimal doses based on a self-collected dataset. Building upon study [15], this research extends the data preprocessing phase by not only addressing missing values through imputation, but also incorporating advanced techniques in feature selection and clustering to enhance data quality and model performance. Following this, the framework emphasizes feature selection to identify and retain the most relevant variables, which not only reduces the dimensionality and noise within the dataset but also enhances the efficiency and interpretability of the predictive models. Alongside feature selection, data clustering is performed to group similar data points, which helps capture intrinsic data structures and patterns, improving the homogeneity within clusters and allowing models to better learn from more coherent subsets of data. Processed data are then used in the training stage, which employs various machine learning techniques such as LR, DT, k-NN, NB, SVM, RF, and multi-layer perceptron (MLP) to train predictive models. These trained models are evaluated to measure performance and generate decision rules. Finally, in the output stage, the framework delivers specific, actionable dose recommendations derived from the best-performing predictive model, enabling personalized and optimized dosing decisions that can be directly applied in practical, real-world scenarios.

3.1. Self-collected dataset

The data was collected from 1,727 patients in 108 Military Central Hospital. They underwent thyroid cancer surgery, aiming to predict the optimal dose of I-131 radiation therapy required to eliminate any residual thyroid tissue. The dataset includes 40 input features that capture detailed clinical information about the patients, such as age, gender, cancer stage, blood test results, iodine uptake capacity, tumor size, and other critical indicators shown in Table 1. Each feature represents an aspect of the patient's health status, providing a comprehensive view to support machine learning models in identifying key factors influencing the radiation dose decision. The dataset comprises 40 input features that provide detailed clinical information about the patient's health condition, factors related to cancer severity, and various medical test results. These features include age, gender, cancer stage, blood test results, the patient's iodine uptake ability, tumor size, and many other critical indicators.

Each feature reflects a specific aspect of the patient's condition, contributing to a comprehensive overview of their health and supporting the machine learning model in identifying key factors that influence the appropriate radiation dosage. The target features indicate the dose, classified into one of three levels: 50 mCi, 75 mCi, or 100 mCi. These are the most common doses in clinical practice, but selecting between them, especially 50 vs. 75 mCi or 75 vs. 100 mCi, can be challenging due to light differences in patient conditions. To address this issue, a dataset was constructed to develop machine learning models capable of accurately predicting personalized radiation doses based on patients' clinical information. These models support physicians in making data-driven treatment decisions, enhancing the accuracy and consistency of care. By learning from historical cases, the models offer precise recommendations for future patients, contributing to improved treatment outcomes, reduced dependence on clinical intuition, and minimized risks of complications in thyroid cancer therapy.

3.2. Data processing

3.2.1 Missing data imputation

Missing data is a common challenge in machine learning that can significantly affect model performance if not properly addressed [16]. Missing values may arise from various causes, such as incomplete data collection, input errors, or the inability to measure certain desired variables. In this dataset, several attributes have over 50% missing values shown in Table 1, including "Cervical lymph node metastasis size" (66.3%), "Lymph node group" (66.01%), and "Number of metastatic lymph nodes" (66.01%). Retaining such attributes risks introducing bias and insufficient learning. Therefore, attributes with over 50% missing data were excluded, leaving 37 input features to ensure the model focuses on reliable, informative data for improved performance and reduced complexity.

Imputation techniques were applied to handle attributes with minimal missing data, ensuring the dataset's integrity while avoiding the loss of valuable records. For continuous variables, such as "Calcium index" (0.17% missing), "Potassium index" (0.12% missing), and "Tg index" (0.06% missing), mean imputation was employed. This approach replaces missing values with the mean of the available data, preserving the overall distribution and minimizing disruption to the dataset's statistical properties. For categorical variables, including "Surgical approach" (1.74% missing), "Cancer

Table 1. Characteristic of feature name

No.	Feature name	Field ID	Value Domain	Missing value (%)
1	Gender	g_t	Discrete	0
2	Symptoms	t_c		0
3	Type of Surgery	ht_pt		1.74
4	Post-Surgery Condition	s_pt		0
5	Metastasis	dc_hc		6.37
6	Additional Surgery	pt_bs		0
7	Family Medical Cancer	ts_k		0
8	Family History of Goiter	ts_bc		0
9	Diagnosis or Treatment	cd_dt		0
10	Clinical Neck Lymph Nodes	ls_hc		0
11	Disease status	b_l		0
12	Pregnancy Test	test_t		0
13	Other Blood Biochemistry	sh_m		0
14	Complete Blood Count	ct_m		0
15	Thyroid Tissue	m_g		0
16	Abdominal Ultrasound Results	sa_ob		4
17	Diagnostic Scan Dose (mCi)	l_cd		0
18	Pulmonary Ventilation	tkp		0
19	Cancer Type	t_ut		0
20	Cancer Stage	gd_ut		12.74
21	Recurrence Risk	nc_tp		7.7
22	T Index	t		0
23	N Index	n		0
24	M Index	m		0
25	Comorbidities	b_kt		0
26	Tumor Properties	tc_u		35.32
27	Lymph Node Groups	n_h		66.01
28	Number of Metastatic Nodes	sl_hdc		66.01
29	Radiation Dose (mCi)	l_dk		0
30	Heart Rate (bpm)	mach	82.16 ± 13.04	0.35
31	Height (cm)	c_c	155.80 ± 18.74	0.35
32	Weight (kg)	c_n	56.78 ± 16.28	0.41
33	TSH Level (mIU/ml)	tsh	68.66 ± 33.71	0
34	Calcium Level (mmol/l)	canxi	2.27 ± 0.22	0.17
35	Potassium Level (mmol/l)	kali	3.80 ± 1.45	0.12
36	Tg Level	tg	3.67 ± 15.22	0.06
37	Anti-Tg Level	atg	33.53 ± 87.87	0.12
38	Tumor Size	kt_u	9.62±7.45	34.11
39	Cervical Lymph Node Metastasis Size	kt_hdc	0.55 ± 2.46	66.3
40	Systolic Blood Pressure (mmHg)	ha_tt	124.01 ± 17.26	1.56
41	Age	tuoi	46.66 ± 12.29	0.93

stage" (12.74% missing), and "Recurrence risk" (7.7% missing), mode imputation was used to substitute missing values with the most frequently occurring category. This method maintains the categorical nature of these attributes while effectively addressing missing entries.

By employing these imputation strategies, the dataset retains critical information and avoids unnecessary reduction in size, ensuring that the data remains robust for training machine learning models. These techniques enhance the model's ability to accurately predict the optimal I-131 radiation dose, improving both performance and reliability.

3.2.2 Feature selection

The purpose of analyzing the correlation between features and the target variable is to eliminate features with low correlation [17]. Retaining highly correlated features improves the accuracy and efficiency of machine learning models for dose prediction. For instance, attributes like "Cervical lymph node metastasis" (correlation = 0.39605) and "N index" (correlation = 0.43318) are essential factors that significantly influence the treatment decision. Conversely, features such as "Pulse (ck/p)" and "Weight," with zero correlation, were removed as they provide no predictive value and could introduce noise into the model. This process reduces complexity while enhancing predictive performance.

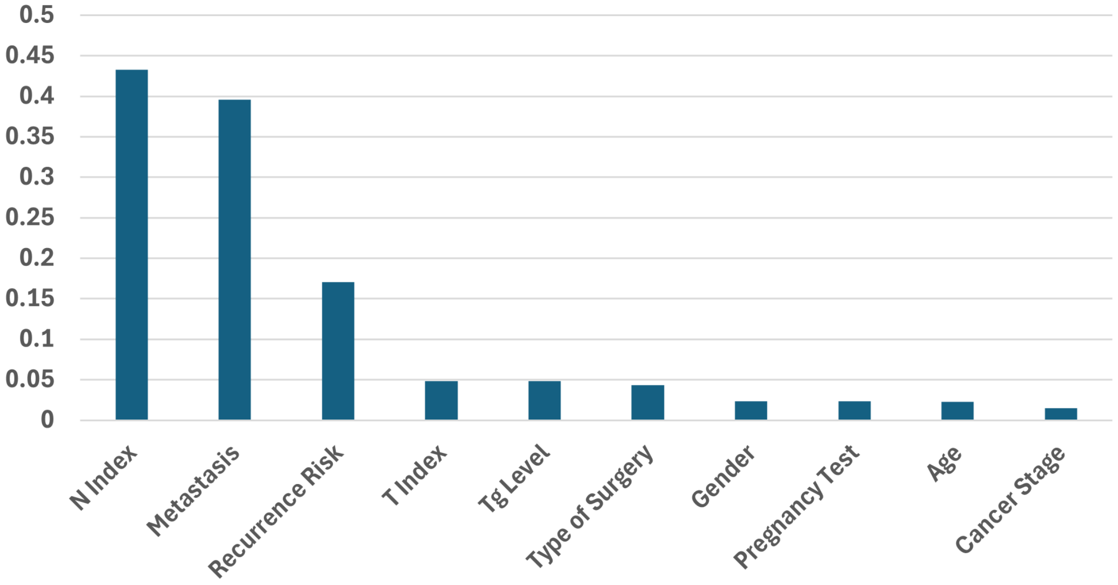

Fig. 2. The correlation of selected features.

In our experiment, Figure 2 shows the correlation of selected features. This approach ensures the model focuses on critical factors, such as "Cervical lymph node metastasis" and "Recurrence risk," which directly impact treatment outcomes. Additional features like "Age," "Gender," and "Surgical approach" were also preserved for their role in providing a holistic understanding of patient health. By refining the dataset, we improve model accuracy and support data-driven decisions in delivering precise and effective radiation therapy.

3.2.3 Data clustering

After filtering out irrelevant features, we applied the K-Means clustering algorithm 1 [18] to group the data into k clusters by iteratively assigning data points to the nearest centroid and updating the centroids. Initially, k centroids are chosen randomly, and the process continues until the centroids stabilize or the maximum number of iterations is reached. This iterative approach helps optimize the partitioning of the data, reducing its complexity and identifying potential mislabeled records for removal.

Algorithm 1: K-Means Clustering

```
Input: Dataset $X = \{x_1, x_2, \ldots, x_n\}$, Number of clusters $k$
Output: Cluster centroids $\{c_1, c_2, \ldots, c_k\}$ and their assignments
Randomly initialize centroids $\{c_1, c_2, \ldots, c_k\}$;
while centroids have significantly changed or maximum number of iterations not reached do
    for each data point $x_i$ do
        Assign $x_i$ to the nearest centroid;
    end
    for each cluster $C_j$ do
        Update centroid;
    end
end
return Final centroids $\{c_1, c_2, \ldots, c_k\}$ and their assignments;
```

To standardize the dataset and minimize confusion, we grouped the records into three clusters: Cluster 1 represents the 50 dose, Cluster 2 represents the 75 dose, and Cluster 3 represents the 100 dose. Each cluster reflects a specific radiation dose level. After performing the clustering, we reviewed the records within each cluster. If any records within the same cluster had different labels, we removed them. This process resulted in a refined dataset with 1,232 records. While this process reduced the dataset size, it significantly improved label consistency and data quality, which are crucial for building reliable machine learning models. This refinement enhances the model's ability to generalize and make accurate predictions, especially important in the context of clinical decision-making for thyroid cancer treatment.

3.3. Building machine learning models

Classifying tabular data is a key task in machine learning, where the goal is to categorize data into predefined labels or groups. Machine learning methods for classification are typically divided into traditional approaches and deep learning. Traditional methods perform well with structured tabular data, requiring fewer computational resources and offering faster results, making them ideal for healthcare classification tasks. While deep learning excels in unstructured data areas like image or natural language processing, it is often less efficient and unnecessary for tabular data. Therefore, we implemented seven fundamental machine learning methods to predict the I-131 radiation dose for thyroid cancer patients after surgery: DT [19], kNN [20], NB [21], SVM [22], LR [23], RF [24], and MLP [25].

We applied fine-tuning techniques to optimize the performance of various classification models for I-131 dose prediction in thyroid cancer treatment. Beginning with the DT model, we set a maximum depth of five, a minimum of four samples per

Table 2. Summary of optimal hyperparameters of studied ML models

Model	Hyperparameters	Values
DT	max_depth	5
	min_samples_leaf	4
	min_samples_split	8
KNN	n_neighbors	3
	weights	distance
NB	var_smoothing	0.001
SVM	C	10
	kernel	rbf
LR	C	10
RF	Bootstrap	FALSE
	Max depth	5
	Min samples leaf	4
	Min samples split	8
MLP	learning_rate	1
	hidden_layer_sizes	8 4
	max_iter	300

leaf, and required eight samples to split a node to control overfitting. The kNN algorithm was configured with three neighbors and distance-based weighting, while the NB classifier used a smoothing factor of 0.001 to handle rare values. The SVM employed a kernel with a regularization parameter C = 10, and LR also used C = 10 to ensure generalizability. We tested a MLP with two hidden layers (8 and 4 neurons), a learning rate of 1.0, and trained it over 300 epochs. Finally, the RF ensemble method was evaluated for its ability to improve prediction accuracy. All models were implemented using Python 3.11.9 and scikit-learn 1.2.2, with their optimal hyperparameters summarized in Table 2.

4. Results and discussion

4.1. Evaluation metrics

Table 3. Confusion matrix

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

In our classification task to predict 50, 75, and 100 doses for thyroid cancer patients, we using several evaluation metrics based on the confusion matrix in Table 3, such as True Positive Rate (TPR), Recall, Precision, False Positive Rate (FPR), and F1 score, to assess the model's performance. TPR and Recall focus on the model's ability to correctly identify each dose category (50, 75, or 100). Precision evaluates the accuracy of these predictions, while FPR measures the proportion of incorrect classifications. Finally, the F1 score harmonizes Precision and Recall, offering a comprehensive view of model effectiveness. These metrics are essential in ensuring accurate dose predictions and minimizing classification errors.

4.2. Results

The performance metrics presented in the Table 4 provide a comprehensive evaluation of several machine learning models on a classification task. The results indicate that the DT model demonstrates the highest performance with the highest TPR of 0.995, and a notably low FPR of 0.001. Its Recall, Precision, and F1-score (0.995, 0.996, and 0.995, respectively) also perform excellently, suggesting that DT efficiently identifies positives while maintaining precision. The MLP model follows closely behind, showing strong scores of 0.986 in TPR, 0.007 in FPR, and 0.986 in F1-score, indicating a robust ability to classify correctly with minor errors.

- **Decision Tree**
  The DT model achieved the highest performance. For the 50 mCi class, it recorded a TPR of 0.998 and an F1-score of 0.999, indicating near-perfect classification. Although precision slightly dropped to 0.94 for the dose 75 mCi class, the model still maintained strong performance across all classes, with an average F1-score of 0.995. Its high interpretability makes it especially valuable in clinical settings.
- **k-Nearest Neighbors**
  kNN performed well for the 50 and 100 mCi, but less well with the 75 mCi class, achieving a lower F1-score of 0.862. The overall average F1-score was 0.967. This model is sensitive to class overlap and data distribution, affecting its reliability in imbalanced datasets.
- **Naive Bayes**
  NB showed moderate performance, especially struggling with the 75 mCi class (F1-score of 0.792). It performed better for the dose 100 mCi class with high precision (0.99). The average F1-score was 0.95, reflecting its simplicity and speed but limited accuracy in complex classification problems.
- **Support Vector Machine**
  SVM demonstrated high accuracy for the 50 and 100 mCi classes, with an average F1-score of 0.982. However, its performance dropped for the 75 mCi class (F1-score of 0.871), likely due to class overlap. Overall, SVM is effective but less interpretable than DT.
- **Logistic Regression**
  Logistic Regression provided stable results with an average F1-score of 0.97. It showed limitations in the 75 mCi class (F1-score of 0.85) but performed comparably to SVM and DT in the other classes. Its simplicity makes it suitable for binary classification, though less effective for complex patterns.
- **Multi-Layer Perceptron**
  MLP yielded strong performance across all classes, especially the 75 mCi class (F1-score of 0.905). With an average F1-score of 0.986, it demonstrated robust handling of non-linear relationships, though its “black-box” nature limits interpretability compared to DT.
- **Random Forest**
  Random Forest achieved high accuracy similar to DT, with an average F1-score

of 0.978. It was slightly less effective in the 75 mCi class (F1-score of 0.88) but excelled in classifying 100 mCi cases. RF offers a good balance of performance and robustness, though it is less interpretable than a single DT.

Overall, while DT consistently outperforms the other models in both accuracy and balance, the MLP and SVM provide competitive alternatives, each showing excellent performance across all metrics. The NB model, although effective, tends to underperform relative to others in precision, indicating room for improvement. The results suggest that models like DT, MLP, and SVM are well-suited for high-performance classification tasks, with DT emerging as the most effective in this specific dataset.

Table 4. Performance of machine learning models in predicting I-131 dose

Classifier	**Class**	**TPR**	**FPR**	**Recall**	**Precision**	**F1**
DT	50	0.998	0.000	0.998	1.000	0.999
	75	0.989	0.004	0.989	0.940	0.962
	100	0.993	0.002	0.993	0.998	0.996
	Average	**0.095**	**0.001**	**0.995**	**0.996**	**0.995**
kNN	50	0.994	0.048	0.994	0.938	0.965
	75	0.825	0.005	0.825	0.916	0.862
	100	0.964	0.000	0.964	1.000	0.982
	Average	**0.968**	**0.021**	**0.968**	**0.969**	**0.967**
NB	50	0.970	0.041	0.970	0.947	0.958
	75	0.857	0.025	0.857	0.760	0.792
	100	0.941	0.010	0.941	0.990	0.964
	Average	**0.948**	**0.025**	**0.948**	**0.955**	**0.950**
SVM	50	0.996	0.017	0.996	0.978	0.987
	75	0.845	0.005	0.845	0.912	0.871
	100	0.990	0.005	0.990	0.995	0.992
	Average	**0.983**	**0.010**	**0.983**	**0.983**	**0.982**
LR	50	0.987	0.026	0.987	0.964	0.975
	75	0.825	0.008	0.825	0.897	0.850
	100	0.979	0.015	0.979	0.986	0.982
	Average	**0.970**	**0.019**	**0.970**	**0.971**	**0.970**
MLP	50	0.991	0.009	0.991	0.989	0.990
	75	0.913	0.007	0.913	0.908	0.905
	100	0.993	0.005	0.993	0.995	0.994
	Average	**0.986**	**0.007**	**0.986**	**0.987**	**0.986**
RF	50	0.996	0.027	0.996	0.966	0.981
	75	0.829	0.004	0.829	0.945	0.880
	100	0.985	0.003	0.985	0.997	0.991
	Average	**0.979**	**0.014**	**0.979**	**0.979**	**0.978**

Table 5. Comparison of model performance

Reference	**Best Method**	**Precision**	**Recall**	**F1-Score**
[15]	AdaBoost	0.66	0.88	0.75
Our research	DT	**0.996**	**0.995**	**0.995**

The results demonstrate that our proposed approach yields promising performance, showing strong predictive ability even with relatively simple machine learning models. Compared to previous research [15] using the same dataset, the current study was designed to significantly enhance and extend prediction performance through a more comprehensive framework. As shown in Table 5, the DT method proposed in our study significantly outperforms the AdaBoost approach. Our model achieves a Precision of 0.996, a Recall of 0.995, and an F1-Score of 0.995, while AdaBoost achieves only 0.66, 0.88, and 0.75, respectively. These results demonstrate that our model not only better identifies positive cases but also maintains high precision, thereby improving the overall classification effectiveness.

4.3. Discussion

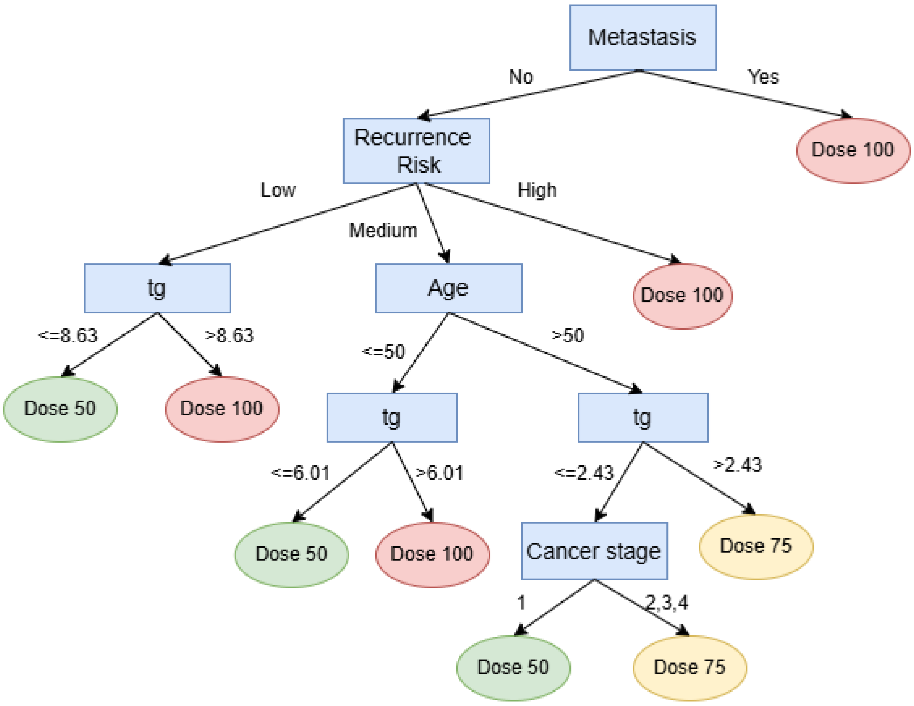

Fig. 3. Decision tree for estimating radiation therapy dose for Vietnamese doctors.

After identifying the DT as the best model, we outline the pathway to estimate the radiation dose based on the tree's classification rules in Figure 3. This method determines the appropriate radiation dose for thyroid cancer patients by considering several clinical factors. The process starts with assessing whether the patient has metastasis. If metastasis is detected, a dose of 100 is recommended. If there is no metastasis, the recurrence risk is evaluated. For low-risk patients, the decision is based on the "Tg" marker, with a threshold of 8.63 used to determine whether the dose should be 50 or 100. For medium-risk patients, age plays a key role in determining treatment. Patients aged 50 or younger are assessed using a "Tg" value with a threshold of 6.01, while those over 50 are further evaluated based on "Tg" (threshold of 2.43) and cancer stage. For patients in stage 1, the recommended dose is 50, while for stages 2, 3, and 4, the dose is increased to 75.

For high-risk patients-whether due to metastasis or high recurrence risk-the DT directly assigns a dose of 100. This structured, rule-based approach ensures that each patient receives an individualized and clinically appropriate radiation dose, tailored to their specific clinical features. The DT offers clear, interpretable guidelines that healthcare professionals can easily follow, minimizing inconsistencies in decision-making and optimizing treatment outcomes for patients.

5. Conclusion and future work

The results demonstrate that our proposed approach yields promising performance, showing strong predictive ability even with relatively simple machine learning models. Compared to previous research [15] using the same dataset, the current study was designed to significantly enhance and extend prediction performance through a more comprehensive framework. Specifically, we implemented an advanced data processing pipeline that includes intelligent imputation of missing values, guided feature selection, and clustering-based outlier removal. The application of machine learning not only improves accuracy, addresses the limitations of earlier work, and optimizes the treatment process. Studies have shown that machine learning models can personalize the treatment process, thereby improving clinical outcomes for patients with thyroid cancer in Vietnam. A key highlight of the DT model is the DT model introduced as an effective solution to classify data and predict I-131 doses for patients with thyroid cancer. The DT not only provides a visual approach but also offers high interpretability, enabling users to understand the factors influencing treatment decisions. This is particularly important in the medical field, where transparency in decision-making is crucial.

While we present a promising approach to dose estimation using machine learning, there are several avenues for future work. This includes exploring advanced algorithms to improve prediction accuracy for complex data, incorporating more clinical variables and long-term outcomes to enhance model robustness, and extending the approach to real-time clinical decision support systems. Collaboration with hospitals for data collection and validation can further refine the model’s applicability. Furthermore, developing user-friendly applications that facilitate easy deployment and interpretation in clinical settings will be essential to maximizing the impact of this technology.

Acknowledgements

This research is supported by the ministry-level scientific project funded by the Ministry of National Defence under grant number 2025.89.47, titled "Application of artificial intelligence in whole-body scintigraphy image analysis and support for determining I-131 ablative dose in patients with Differentiated thyroid cancer".

References

[1] M. E. Cabanillas, D. G. McFadden, and C. Durante, "Thyroid cancer," *The Lancet*, vol. 388, no. 10061, pp. 2783–2795, 2016. DOI: 10.1016/S0140-6736(16)30172-6

[2] C. M. Kitahara and A. B. Schneider, "Epidemiology of thyroid cancer," *Cancer Epidemiology, Biomarkers and Prevention*, vol. 31, no. 7, pp. 1284–1297, 2022. DOI: 10.1158/1055-9965.EPI-21-1440

[3] R. Prasad, V. Rao, A. Subash, K. S. Majumdar, P. Sinha, K. Kallur, and R. C. Nayar, "Residual thyroid tissue on postoperative diagnostic 131 I radioactive whole-body scan after surgery in differentiated thyroid cancer: A tertiary referral centre experience," *Indian Journal of Surgical Oncology*, vol. 13, no. 1, pp. 109–114, 2022. DOI: 10.1007/s13193-021-01357-x

[4] S. A. Doi, N. J. Woodhouse, L. Thalib, and A. Onitilo, "Ablation of the thyroid remnant and I-131 dose in differentiated thyroid cancer: A meta-analysis revisited," *Clin Med Res.*, vol. 5, no. 2, pp. 87–90, 2007. DOI: 10.3121/cmr.2007.763

[5] A. M. Avram and Y. K. Dewaraja, "Thyroid cancer radiotheragnostics: The case for activity adjusted 131 I therapy," *Clinical and Translational Imaging*, vol. 6, no. 5, pp. 335–346, 2018. DOI: 10.1007/s40336-018-0291-x

[6] Z.-H. Zhao, F.-Q. Li, J.-K. Han, and X.-J. Li, "Effect of 131 I 'clear residual thyroid tissue'after surgery on the function of parathyroid gland in differentiated thyroid cancer," *Experimental and Therapeutic Medicine*, vol. 10, no. 6, pp. 2079–2082, 2015. DOI: 10.3892/etm.2015.2812

[7] C. M. Hong and B.-C. Ahn, "Factors associated with dose determination of radioactive iodine therapy for differentiated thyroid cancer," *Nuclear Medicine and Molecular Imaging*, vol. 52, no. 4, pp. 247–253, 2018. DOI: 10.1007/s13139-018-0522-0

[8] P. Szolovits, R. S. Patil, and W. B. Schwartz, "Artificial intelligence in medical diagnosis," *Annals of Internal Medicine*, vol. 108, no. 1, pp. 80–87, 1988. DOI: 10.7326/0003-4819-108-1-80

[9] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, vol. 69, pp. S36–S40, 2017. DOI: 10.1016/j.metabol.2017.01.011

[10] S. Chakradhar, "Predictable response: finding optimal drugs and doses using artificial intelligence," *Nature Medicine*, vol. 23, no. 11, pp. 1244–1248, 2017. DOI: 10.1038/nm1117-1244

[11] V. Eleftheriadis, G. Savvidis, V. Paneta, K. Chatzipapas, G. C. Kagadis, and P. Papadimitroulas, "A framework for prediction of personalized pediatric nuclear medical dosimetry based on machine learning and monte carlo techniques," *Physics in Medicine & Biology*, vol. 68, no. 8, 2023. DOI: 10.1088/1361-6560/acc4a5

[12] M. F. Georgiou, J. A. Nielsen, R. Chiriboga, and R. A. Kuker, "An artificial intelligence system for optimizing radioactive iodine therapy dosimetry," *Journal of Clinical Medicine*, vol. 13, no. 1, p. 117, 2023. DOI: 10.3390/jcm13010117

[13] R. Sa, T. Yang, Z. Zhang, and F. 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*, vol. 29, no. 1, pp. e68–e80, 2024. DOI: 10.1093/oncolo/oyad252

[14] Y.-W. You, C.-W. Lee, and J.-S. Kim, "A study on predictive modeling of I-131 radioactivity based on machine learning," *Journal of Radiological Science and Technology*, vol. 46, no. 2, pp. 131–139, 2023.

[15] M. Lai *et al.*, "Ablation dosage recommendation for thyroid cancer treatment following thyroidectomy using machine learning," *Journal of Military Science and Technology*, vol. 96, pp. 137–144, 2024. DOI: 10.54939/1859-1043.j.mst.96.2024.137-144

[16] P. D. Allison, "Missing data," in *The SAGE Handbook of Quantitative Methods in Psychology*. Thousand Oaks, CA: SAGE Publications, 2009, vol. 23, pp. 72–89.

[17] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–45, 2017. DOI: 10.1145/3136625

[18] A. Ahmad and L. Dey, "A k-mean clustering algorithm for mixed numeric and categorical data," *Data & Knowledge Engineering*, vol. 63, no. 2, pp. 503–527, 2007. DOI: 10.1016/j.datak.2007.03.016

[19] B. De Ville, "Decision Trees," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 6, pp. 448–455, 2013. DOI: 10.1002/wics.1278

[20] L. E. Peterson, "K-Nearest Neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009. DOI: 10.4249/scholarpedia.1883

[21] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *Journal of Information Science*, vol. 44, no. 1, pp. 48–59, 2016. DOI: 10.1177/0165551516677946

[22] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of Support Vector Machine learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41–51, 2018. DOI: 10.21873/cgp.20063

[23] E. Y. Boateng, D. A. Abaye *et al.*, "A review of the logistic regression model with emphasis on medical research," *Journal of Data Analysis and Information Processing*, vol. 7, no. 4, pp. 190–207, 2019. DOI: 10.4236/jdaip.2019.74012

[24] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324

[25] H. Taud and J.-F. Mas, "Multilayer perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*. Springer, 2018, pp. 451–455.

Manuscript received 19-03-2025; Accepted 15-06-2025.

■

Thanh Trung Nguyen received Bachelor's degree in Biomedical Engineering from Le Quy Don Technical University, Vietnam, in 2010, Master's degree in Biomedical Engineering in 2015, and the PhD. degree in Electronic Engineering from Hanoi University of Science and Technology in 2020. He is currently the Vice Head of the Department of Equipment at the 108 Military Central Hospital and a lecturer at the Department of Diagnostic Imaging, Institute of Clinical Medical and Pharmaceutical Sciences 108. His research interests include biomedical instrumentation, medical imaging, and electronic systems in healthcare.
Email: trung.ntc10@benhvien108.vn

Thi Thu Ha Ngo obtained her Bachelor's degree in Advanced Biomedical Engineering from Hanoi University of Science and Technology (HUST) in September 2024. She is currently pursuing her Master's degree in the same field at HUST and works at Department of Equipment at the 108 Military Central Hospital, Hanoi, Vietnam. Her main research interests lie in the areas of image processing, video analysis, and signal processing, with a particular focus on applying computer vision and artificial intelligence techniques in the biomedical domain.
E-mail:ngottha110@gmail.com

Phu Minh Lai graduated with a Bachelor's degree in Electronics and Telecommunications from HUST, Vietnam, in 2015. He earned his PhD. in Biomedical Electronic Engineering in 2025 from the School of Electrical and Electronic Engineering, HUST. His doctoral research focused on the integration of biomedical signal processing and artificial intelligence in clinical applications. His current research focuses on artificial intelligence in medical imaging, signal processing, and the translation of AI solutions into clinical practice.
E-mail: phubinh1455@gmail.com

ĐỀ XUẤT KHUYẾN NGHỊ LIỀU XOÁ MÔ GIÁP CÒN SÓT LẠI SAU PHẪU THUẬT DỰA TRÊN MÔ HÌNH HỌC MÁY

Nguyễn Thành Trung, Ngô Thị Thu Hà, Lại Phú Minh

Tóm tắt

Ung thư tuyến giáp, đặc biệt là các loại phân biệt như ung thư biểu mô nang và biểu mô tủy, là một thách thức sức khỏe lớn ở Việt Nam cũng như trên toàn cầu, trong đó phẫu thuật cắt bỏ tuyến giáp (cắt tuyến giáp) là phương pháp điều trị chính. Tuy nhiên, sau phẫu thuật, mô giáp thường vẫn còn sót lại, đòi hỏi phải tiến hành tiêu diệt hiệu quả để ngăn ngừa tái phát và các biến chứng. Điều trị bằng i-ốt phóng xạ I-131 là phương pháp tiêu chuẩn để tiêu diệt mô giáp còn sót lại, tuy nhiên việc ước lượng liều I-131 phù hợp cho từng bệnh nhân vẫn là thách thức lớn. Các phương pháp truyền thống hiện tại chủ yếu dựa vào hướng dẫn chung và kinh nghiệm của bác sĩ, nhưng đôi khi không đảm bảo độ chính xác cần thiết cho từng bệnh nhân cụ thể. Mặc dù một số nghiên cứu trước đã cố gắng cải thiện việc ước lượng liều bằng các phương pháp học máy, nhưng thường bỏ qua tầm quan trọng của các kỹ thuật tiền xử lý dữ liệu. Việc chưa chú trọng đến chất lượng dữ liệu này đã làm hạn chế hiệu quả dự đoán của các mô hình. Để khắc phục vấn đề đó, nghiên cứu của chúng tôi mở rộng các công trình trước đây bằng cách tập trung vào giai đoạn tiền xử lý dữ liệu, với mục tiêu nâng cao độ chính xác và độ tin cậy của mô hình. Chúng tôi áp dụng các phương pháp trích xuất đặc trưng nhằm lựa chọn những đặc trưng nổi bật và có ảnh hưởng lớn trong việc dự đoán liều, sau đó sử dụng bảy mô hình học máy truyền thống để thực hiện dự đoán. Kết quả cho thấy mô hình Cây quyết định *(Decision Tree - DT)* vượt trội so với các mô hình khác, đạt tỷ lệ dương tính thật (TPR) cao nhất là 0,995; tỷ lệ dương tính giả (FPR) thấp là 0,001 và hiệu suất xuất sắc ở các chỉ số Recall, Precision và F1-score (0,995; 0,996 và 0,995 tương ứng). Bên cạnh đó, chúng tôi sử dụng quy tắc quyết định từ mô hình Cây quyết định nhằm hỗ trợ việc dự đoán liều một cách rõ ràng và dễ diễn giải trên bộ dữ liệu thông tin lâm sàng từ bệnh nhân tại Việt Nam. Phương pháp này mang lại hứa hẹn lớn trong việc cải thiện hiệu quả điều trị bức xạ, đảm bảo quyết định chính xác hơn dựa trên dữ liệu, và cuối cùng nâng cao kết quả điều trị cho bệnh nhân.

Từ khóa

Mô giáp còn sót lại; học máy; quy tắc quyết định; ước lượng liều; tập dữ liệu bệnh nhân Việt Nam.