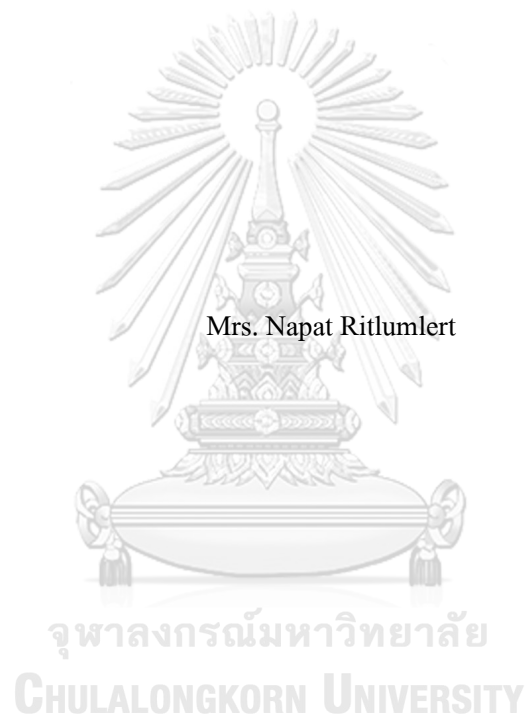


Radiomics-based Prediction of Radiation-induced Hypothyroidism in Nasopharyngeal Cancer
Patients



A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Biomedical Engineering
Faculty Of Engineering
Chulalongkorn University
Academic Year 2023

การใช้เรดิโอมิกส์ทำนายการเกิดภาวะต่อมไทรอยด์ทำงานต่ำหลังจากการฉายรังสีในผู้ป่วยมะเร็ง
โพรงจมูก



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมชีวเวช
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2566

ฉกัทร ฤทธิ์ลีลลล : การใช้เรดิโอมิกส์ทำนายการเกิดภาวะต่อมไทรอยด์ทำงานต่ำหลังจากการฉายรังสีในผู้ป่วยมะเร็งโพรงจมูก. (Radiomics-based Prediction of Radiation-induced Hypothyroidism in Nasopharyngeal Cancer Patients) อ.ที่ปรึกษาหลัก : ศศ. ดร. โยชิน รัทวงษ์ไทย, อ.ที่ปรึกษาร่วม : อ. ดร.เอกพล ช่วงสุนิษ

ภาวะแทรกซ้อนในระยะยาวจากการใช้รังสีเป็นเรื่องที่สำคัญที่ควรคำนึงถึงในการกำหนดแผนการรักษาโรคต่อมไทรอยด์ทำงานต่ำเป็นหนึ่งในภาวะแทรกซ้อนที่พบได้บ่อยจากการใช้รังสีรักษาบริเวณศีรษะและลำคอ ปัจจุบันมีการใช้ข้อมูลทางคลินิกและข้อมูลปริมาณรังสีจากแผนการรักษาช่วยประเมินโอกาสเกิดภาวะต่อมไทรอยด์ทำงานต่ำหลังการฉายรังสีในคนไข้มะเร็งโพรงจมูกแต่ผลการประเมินยังไม่ดีนัก การนำข้อมูลภาพทางการแพทย์เข้ามาช่วยน่าจะเพิ่มข้อมูลที่สำคัญในการทำนายได้ถูกต้องมากขึ้น จุดมุ่งหมายของงานวิจัยนี้เพื่อทำนายภาวะต่อมไทรอยด์ทำงานต่ำในคนไข้มะเร็งโพรงจมูกโดยการใช้เรดิโอมิกส์ของภาพเอกซเรย์คอมพิวเตอร์ร่วมกับข้อมูลทางคลินิกและข้อมูลปริมาณรังสี

การศึกษานี้ทำการรวบรวมข้อมูลจากคนไข้มะเร็งโพรงจมูกจำนวน 220 คน โดยที่ได้รับการวินิจฉัยว่ามีภาวะต่อมไทรอยด์ทำงานต่ำหลังการฉายรังสีในช่วงเวลา 2 ปี เริ่มจากกำหนดขอบเขตของภาพต่อมไทรอยด์และดึงข้อมูลภาพเปลี่ยนเป็นข้อมูลเชิงตัวเลขจากภาพเอกซเรย์คอมพิวเตอร์ก่อนการฉายรังสี พีเจอร์เรดิโอมิกส์ทั้งหมดจะถูกนำไปวิเคราะห์ร่วมกับข้อมูลทางคลินิกและข้อมูลปริมาณรังสีจากแผนการรักษาของคนไข้ โดยทำการสร้างโมเดลในการทำนายโรคด้วยวิธี logistic regression random forest และ gradient boosting เพื่อสร้างโมเดลเรดิโอมิกส์ โมเดลคลินิกและปริมาณรังสี และ โมเดลข้อมูลรวม

ผลการศึกษา พบว่าโมเดลข้อมูลรวมจากการใช้ logistic regression และ random forest มีประสิทธิภาพในการทำนายสูงสุด โดยมีค่า AUC อยู่ที่ 0.80 ± 0.06 และ 0.81 ± 0.06 ตามลำดับ และพบว่าสามารถทำนายได้ดีกว่าการใช้โมเดลคลินิกและปริมาณรังสีอย่างมีนัยสำคัญทางสถิติ ($p\text{-value} < 0.05$) โดยมีค่า AUC อยู่ที่ 0.68 ± 0.07 และ 0.71 ± 0.06 โดยที่เจอร์เรดิโอมิกส์ที่ใช้ในโมเดลข้อมูลรวมนี้ส่วนใหญ่มาจาก พีเจอร์กลุ่ม texture-based และกลุ่ม filtered-based ร่วมกับตัวแปรที่สำคัญทางคลินิกและปริมาณรังสี คือ การแพร่กระจายของต่อมน้ำเหลืองที่คอ, ค่า TSH ก่อนการรักษา, อายุ, ร้อยละของต่อมไทรอยด์ที่ได้รับปริมาณรังสีอย่างน้อย 40 Gy และปริมาณรังสีเฉลี่ยที่ต่อมไทรอยด์ได้รับ

โดยสรุป การใช้ข้อมูลเรดิโอมิกส์จากภาพเอกซเรย์คอมพิวเตอร์ร่วมกับข้อมูลทางคลินิกและปริมาณรังสีสามารถใช้ทำนายโอกาสการเกิดภาวะต่อมไทรอยด์ทำงานต่ำได้และมีประสิทธิภาพดีกว่าการใช้วิธีดั้งเดิมอย่างมีนัยสำคัญ แสดงว่าภาพไทรอยด์นั้นมีข้อมูลที่สำคัญที่ช่วยให้สามารถทำนายภาวะต่อมไทรอยด์ทำงานต่ำจากการใช้รักษาในผู้ป่วยมะเร็งโพรงจมูก

สาขาวิชา วิศวกรรมชีวเวช
ปีการศึกษา 2566

ลายมือชื่อนิติศ
ลายมือชื่อ อ.ที่ปรึกษาหลัก
ลายมือชื่อ อ.ที่ปรึกษาร่วม

6371009921 : MAJOR BIOMEDICAL ENGINEERING

KEYWORD: Hypothyroidism, Nasopharyngeal cancer, Radiomics, CT, Medical images

Napat Ritlumert : Radiomics-based Prediction of Radiation-induced Hypothyroidism in Nasopharyngeal Cancer Patients. Advisor: Asst. Prof. YOTHIN RAKVONGTHAI, Ph.D. Co-advisor: EKAPOL CHUANGSUWANICH, Ph.D.

When planning radiation therapy, late-effect complications due to radiotherapy should be considered. One of the most common complications of head and neck radiotherapy is hypothyroidism. Although clinical and dosimetry data are usually used to assess the risk of hypothyroidism after radiation for nasopharyngeal cancer, the outcome is still unsatisfactory. Medical imaging can provide additional information and increase prediction accuracy. The aim of this study was to predict hypothyroidism in patients with nasopharyngeal cancer using CT radiomics combined with clinical and dosimetric data.

The study included 220 participants who were diagnosed with hypothyroidism within 2 years after radiotherapy. Manual segmentation covered the thyroid gland, and feature extractions were performed from pretreatment CT images. All radiomics features were analyzed with clinical and dosimetry information, and the model was constructed using logistic regression, random forest, and gradient boosting. In addition to the radiomics model, conventional, and combined models were built based on the tree-based predictive algorithms.

The findings of the study demonstrated that the combined model had the highest validation performance, as indicated by AUCs of 0.80 ± 0.06 and 0.81 ± 0.06 in logistic regression and random forest, respectively, which were greater than the conventional mode with the AUCs of 0.68 ± 0.07 and 0.71 ± 0.06 (p -value < 0.05). The combined model used in this study used radiomics features, with the majority of these features coming from texture-based classes and filtered-based classes, while the important clinical and dose factors were bilateral neck metastasis, pretreatment TSH level, age, TR V40, and TR mean.

In conclusion, the combination of CT radiomics with clinical and dose information can predict the RIH in nasopharyngeal cancers and significantly improve the performance of prediction models compared to the conventional method. We contend that pretreatment thyroid images contain valuable information that can be used to predict the risk of hypothyroidism after nasopharyngeal radiotherapy.

Field of Study: Biomedical Engineering

Student's Signature

Academic Year: 2023

Advisor's Signature

Co-advisor's Signature

ACKNOWLEDGEMENTS

Firstly, I'm extremely grateful to HRH Princess Chulabhorn College of Medical Science, Chulabhorn Royal Academy for their tangible support of the PhD Scholarship.

This endeavor would not have been possible without my advisor, Associate Professor Dr. Yothin Rakvongthai, who provided knowledge, suggestions, inspiration, and moral support for my Ph.D. study. I have learnt and developed a variety of talents under his guidance that will be useful in the future. Thanks should also go to my co-advisor, Dr. Ekapol Chuangsuwanich and Dr. Sira Sriswasdi, for his comment in my thesis work, especially in data sciences. I would like to extend my sincere thanks to Assistant Dr. Pakpum Somboon, Associate Professor Dr. Juthamas Ratanavaraporn, Assistant Dr. Peerapat Thongnuek and all other professors, lecturers, or teachers at Chulalongkorn University, who gave advies and suggestions to improve our knowledge based on my research in my Ph.D. work.

I would like to acknowledge Professor Dr. Chawalit Lertbutsanukul, Dr. Anussara Prayongrat, and Dr. Sornjarod Oonsiri for their advice and suggestions in the clinical part of my thesis and their helpfulness in data collection.

Lastly, I would be remiss in not mentioning my family, especially my parents, spouse, and children. Their belief in me gave me a high spirit and motivation during my study. Furthermore, I would like to thank all my labmates from Chulalongkorn University Biomedical Imaging Group (CUBIG) and my classmates from the Radiological Technology Program at Chulabhorn Royal Academy, who have always helped and inspired me both physically and mentally.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Napat Ritlumert

TABLE OF CONTENTS

	Page
.....	iii
ABSTRACT (THAI).....	iii
.....	iv
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
CHAPTER I: INTRODUCTION.....	9
1. Background and rationale.....	9
2. Research questions.....	10
3. Research objectives.....	10
4. Scope.....	11
5. Expected benefits.....	11
CHAPTER II: THEORY.....	12
1. Radiation-induced hypothyroidism.....	12
1.1 Head and neck cancers.....	12
1.2 Treatment of head and neck cancers.....	12
1.3 Radiation treatment planning.....	13
1.4 Toxicity.....	14
1.5 Thyroid gland.....	14
1.6 Synthesis and release of thyroid hormones.....	15
1.7 Hypothyroidism.....	17

2. Radiomics in oncology.....	18
2.1 Applications in oncology	18
2.2 Planning a radiomics study	19
2.3 Radiomics workflow	19
1. Image acquisition.....	19
2. Data curation and image preprocessing.....	19
3. Image segmentation	20
4. Feature extraction	21
5. Model construction	24
3. Machine learning.....	24
CHAPTER III: RELATED LITERATURE REVIEWS.....	28
CHAPTER IV: MATERIALS AND METHODS	35
1. Research design.....	35
2. Research design model.....	35
3. Research conceptual framework	36
4. Materials and methods	37
Population and sample	37
Image acquisition	37
Thyroid segmentation	37
Features extraction	39
Right and left lobe thyroid glands in radiomics feature values.....	40
Clinical information and dosimetry information	40
Model construction	41
Testing the model robustness for variation in segmentation.....	42

Statistical analysis	42
5. Ethical consideration	42
CHAPTER V: RESULTS	43
Model performance	54
Radiomics feature values in right and left-lobe thyroid glands	61
Testing the model robustness for variation in segmentation.....	61
CHAPTER VI: DISCUSSION	67
Limitations	70
Future directions.....	70
CHAPTER VII: CONCLUSION.....	72
Appendix.....	73
REFERENCES	83
VITA	88

CHAPTER I: INTRODUCTION

1. Background and rationale

Nasopharyngeal carcinoma (NPC) is one of the most common head and neck cancers (HNC) with specific etiological, clinical, epidemiological, and genetic characteristics. More than 70% of NPC cases occurred in Southeast Asia and South China [1]. GLOBOCAN 2020 [2] reported that during 2020 Thailand found 2,316 new cases of NPC, which were 1.2% of all cancers and found up to 1482 cancer deaths, 1.2% of which were from NPC. Radiotherapy, chemotherapy, and surgery are the main treatments to control and manage HNC. The major treatment is radiotherapy by using various techniques to provide a high dose at a target volume while maintaining an acceptable low dose to critical organs to avoid complications. Since the thyroid gland's location is anterior to the neck regions, it is exposed to high radiation doses. Moreover, the thyroid gland is a highly radiosensitive organ, which results in a higher risk of complications. According to Kazemi et al. [3], it was reported that hypothyroidism, which occurred in 15%-48% of patients after radiation treatment, is the most common radiation side effect of HNC treatment. Therefore, prevention strategies should be considered because of its effect on the quality of life after the treatment course.

At present, hypothyroidism is diagnosed by laboratories and clinical symptoms only. Several studies [4, 5] investigated predictors for radiation-induced hypothyroidism based on pre-treatment clinical and dose parameters. Lertbusayanukul et al. [4] validated a prior report of dose factors in hypothyroidism after intensity-modulated radiation treatment (IMRT) in patients with nasopharyngeal carcinoma (NPC). It was found that TSH greater than 1.55 $\mu\text{U/ml}$ and VS_{60} less than 10 cm^3 were important predictors. Another study by Peng et al. [5] suggested that using the pre-treatment volume ($< 20 \text{ cm}^3$) and $\text{V}_{30,60}$ ($> 80\%$) of thyroid glands can be predictors with moderate prediction results (AUC = 0.64). Therefore, an effective predictive model is needed to improve the treatment planning and reduce the occurrence of radiation induced hypothyroidism.

Medical imaging plays a key role in cancer diagnosis and treatment since it is needed for treatment planning, treatment monitoring, and evaluation of treatment response. The treatment planning system must use imaging guidance before radiation treatment. CT with other modalities (e.g., MRI, PET) will be examined first to define the dose to the tumor target volume and organ at

risk volume. Rapidly increasing computer technology and imaging methods have enabled the storage and use of medical image data in various techniques and clinical applications. More than the applications of traditional images are used for diagnostic and treatment monitoring by human observation, hidden information in images may be useful for improving treatments. Radiomics refers to the automated extraction of quantitative features from medical images for the development of diagnostic and prognostic biomarkers. Using a statistical analysis approach with radiomics features and clinical data could develop a novel predictive tool in clinical application. There have research studies on predicting locoregional recurrence, treatment response, and survival as well as complications in HNC patients after treatment using radiomics [6]. Sheikh et al. [7] used radiomics features from salivary glands to study post-radiation xerostomia. They suggested that baseline CT and MR imaging features may reflect baseline salivary gland function and potential risk for radiation injury.

The objective of this study was to use radiomics from pretreatment contrast-enhanced CT images, dosimetry parameters, and clinical data to predict hypothyroidism within 2 years. The hypothesis was that radiomics with clinical and dosimetry data can improve the prediction of radiation-induced hypothyroidism and has potential to be a new tool in pre-treatment planning to optimize dose constraints on the thyroid gland to reduce thyroid complication effects.

2. Research questions

- 1) What is the best model to predict radiation-induced hypothyroidism after radiation treatment in nasopharyngeal cancer patients?
- 2) What is the additional gain from using radiomics as compared to using only clinical and dosimetry data?
- 3) What are the radiomics signatures of radiation-induced hypothyroidism after radiation therapy in nasopharyngeal cancer patients?

3. Research objectives

- 1) To develop an effective model for predict radiation-induced hypothyroidism after radiation treatment in nasopharyngeal cancer patients.

2) To evaluate whether the superior performance of the model that uses radiomics is better than clinical and dosimetry data alone.

3) To identify the radiomics signatures of radiation-induced hypothyroidism after radiation therapy in nasopharyngeal cancer patients.

4. Scope

This study focused on using radiomics from pre-treatment contrast enhanced CT to develop a model for predicting radiation-induced hypothyroidism after radiation therapy in nasopharyngeal cancer patients.

5. Expected benefits

1. Radiation-induced hypothyroidism predictive models using CT radiomics combined with clinical and dosimetry data.

2. Propose guidance for the important clinical, dose, and image predictors relevant to RIH that help to adjust the treatment plan and reduce the occurrence of thyroid complications after radiation treatment in nasopharyngeal cancers.

CHAPTER II: THEORY

1. Radiation-induced hypothyroidism

1.1 Head and neck cancers

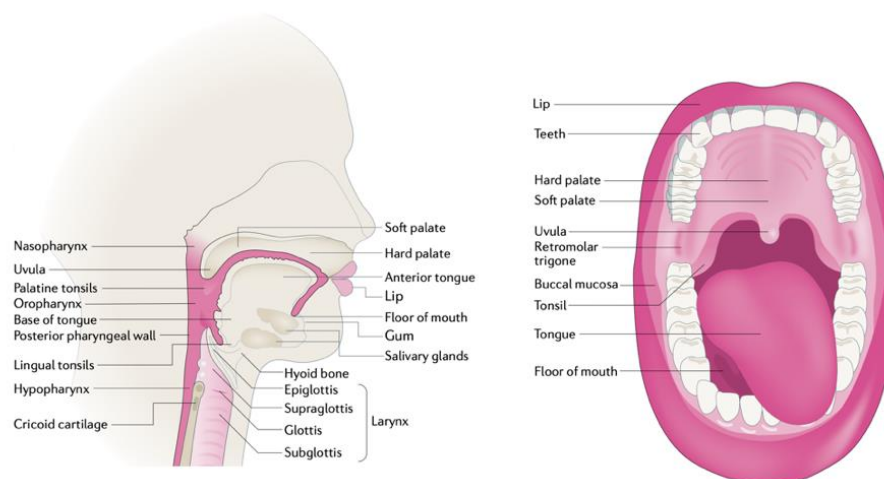


Figure 1 Anatomical site of HNSCC development. HNSCC arises from the mucosal epithelium of oral cavity, nasopharynx, oropharynx, hypopharynx and larynx. [8]

Head and neck squamous cell carcinomas (HNSCC) are the most common malignancies that arise in the head and neck. It is developed from the mucosal epithelium in the oral cavity, pharynx, and larynx as show in Figure 1. The prevalence of HNSCC varies by region and has been linked to tobacco-derived carcinogen exposure, excessive alcohol intake, or both. Moreover, the incidence of HNSCC caused by Human Papillomavirus (HPV) has raised up over the last 30 years, especially for oropharyngeal cancers. This has led to a new subgroup of patients with HNSCC who have different clinical characteristics and molecular biology and more radiosensitivity [9, 10].

1.2 Treatment of head and neck cancers

HNSCC is predominantly a loco-regional disease with only about 5% of patients having distant metastases at the time of diagnosis. Over 60% of patients have locally advanced disease (stage III and VI) at diagnosis. The main treatment consists of surgery, radiotherapy, and recent years with additional of chemotherapy. For early-stage patients, one treatment modality is

sufficient; however, for advanced-stage patients, all treatments must be combined for tumor control, and fractionation can be used to improve the outcome. In the last three decades, technological development in radiotherapy has been improved from 2D-RT based on X-ray images to 3D-CRT (3-dimensional conformal treatment). Three or more treatment fields are used in 3D-CRT as shown in Figure 2 to deliver a homogeneous dosage to the target area. Furthermore, the deployment of multi-leaf collimators and more complex computer algorithms have enabled intensity modulated RT (IMRT) and volumetric modulated arc therapy (VMAT). With these treatment techniques, there is the possibility of conserving normal tissues while increasing radiation dose to the tumor and other target areas [11].

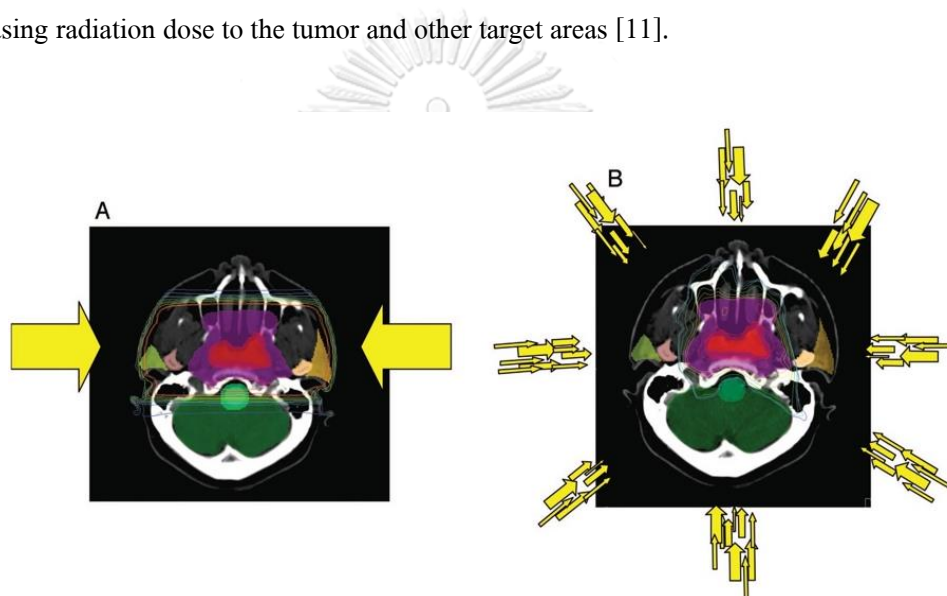


Figure 2 Radiation beam directions. A, Two opposing beams of single intensities, represented by the yellow arrows, create a single-dose distribution through a nasopharynx tumor. B, Intensity modulated radiotherapy allows multiple beams of different intensities [12]

1.3 Radiation treatment planning

Radiation treatment is an external beam radiation using linear accelerators. Intensity-modulated radiation therapy (IMRT) has been successfully implemented because of advances in technology like inverse treatment planning and the multileaf collimator-equipped linear accelerator. These methods allow for the modulation of radiation intensity to give a larger radiation dose to the targets with a sharply conformal target volume coverage while significantly reducing the exposure to the surrounding normal tissues. In daily practice, treatment planning for head and neck cancer patients is initiated with a therapeutic CT scan of the patient to fixate in the

treatment position. In the treatment planning system, the oncologist and the radiologist collaborate to define the tumor volume and clinically relevant target areas. Planning target volumes are then generated, and a dose calculation is performed to ensure that adequate dose distributions are used to cover tumor areas during radiotherapy, and treatment doses of normal tissue are below the tolerance thresholds as indicated by clinical guidelines. At present, treatment planning systems allow a rather precise calculation of the treatment dose for small areas. More information on ionizing radiation tolerance levels in various organs led to RT dose-planning, which restricts a specific dose (dose-constraint) or treated volume to an organ using either a single value, such as the mean dose to the organ, or multiple dose-volume constraints.

1.4 Toxicity

The toxicity of normal tissues is the limiting factor in radiotherapy. Optimization of target dose delivered while minimizing toxicity to normal tissue is an important challenge for radiotherapy. Toxicity has typically been classified into two categories: acute reactions and late reactions. Acute reaction toxicity is defined as effects that occur within 90 days after the start of radiotherapy, whereas late reactions occur more than 90 days (and up to years after RT). However, the relevance of this diagnosis in grading and reporting adverse effects has been challenged internationally. In HNSCC, common acute effects during treatment are pain, mucositis, dysphagia, xerostomia, mucosal edema, and erythema. Common late effects consist of xerostomia, dysphagia, mucosal edema, hypothyroidism, and skin fibrosis. Adverse event reporting in cancer treatment, and specifically radiation oncology have been the Common Terminology Criteria for Adverse Events (CTCAE) and the LENT/SOMA scale [13, 14]. LENT is an acronym for Late Effects Normal Tissues, while SOMA defines toxicity from Subjective, Objective, Management related and Analytic measures (i.e. blood test, CT or the like). Toxicology scores can be converted into binary data, which can be used to analyze radiation dose-response relationships and predict toxicity in patients receiving radiation therapy.

1.5 Thyroid gland

As shown in Figure 3A, the thyroid gland is a butterfly-shaped structure that is located anterior to the trachea and inferior to the larynx, as shown in figure 3A. The gland consists of two lobes, right and left, which are connected through the isthmus. The isthmus, or medial region, is

flanked by wing-shaped left and right lobes. The tissue of the thyroid gland is composed of thyroid follicles. Colloid (figure 3B) is a type of follicle that has a central cavity filled with a sticky fluid. The colloid is surrounded by epithelial follicle cells and is the center of thyroid hormone production. The hormone's production is dependent on the hormone's essential and unique component: iodine [15].

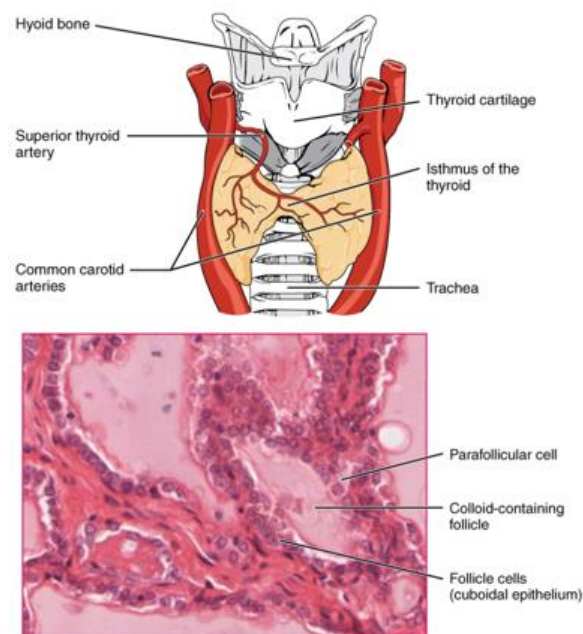


Figure 3 Thyroid anatomy: (a) Thyroid anatomy, (b) Thyroid follicle cells [16]

CHULALONGKORN UNIVERSITY

1.6 Synthesis and release of thyroid hormones

Hormones are created in the colloid when atoms of the mineral iodine bind to a glycoprotein known as thyroglobulin, which is released into the colloid by follicle cells. The hormones are assembled in the following steps:

1. Thyroglobulin synthesis

Thyroglobulin is synthesized and released by the endoplasmic reticulum and Golgi apparatus in follicular cells of the thyroid gland. Thyroglobulin is a large glycoprotein that contains 140 molecules of the amino acid tyrosine and is stored in the follicle.

2. Iodine trapping

Iodide (I^-) from the blood stream enters follicular cells by active transport via an electrochemical gradient called iodide trapping. It passes through follicular cells along with Na^+ by using a sodium-iodide symport pump (iodide pump).

3. Transport of iodine into follicular cavity

After that, iodide enters the follicular cavity by an iodide-chloride pump (pendrin). Iodide was oxidized into iodine (I^0) using thyroid peroxidase as an activator.

4. Iodination of tyrosine

Iodination of tyrosine is iodine combined with tyrosine, which occurs on thyroglobulin. This iodination process was activated by the iodinase enzyme from follicular cells. Then tyrosine was iodized into MIT (mono-iodothyrosine) and DIT (di-iodothyrosine), which are called iodotyrosine residues.

5. Coupling reaction

Iodotyrosine residues were combined into thyroid hormones in three patterns: 1) $DIT + MIT = T3$ (triiodothyrosine), 2) $MIT + DIT =$ reverse T3 (less than 1%), and 3) $DIT + DIT = T4$ (thyroxine). T3, T4 hormones will enter the bloodstream and regulate the whole body.

The production and release of thyroid hormone is controlled by negative feedback, as shown in Figure 4, that involves the hypothalamus, pituitary gland, thyroid gland, and multiple hormones. Start with the hypothalamus releasing thyroid-releasing hormone (TRH), which stimulates the pituitary gland to produce and release thyroid-stimulating hormone (TSH). Then, TSH triggers the thyroid gland to produce T4 and T3. The total amount of hormones that TSH triggers to release is about 80% T4 and 20% T3. When the levels of T3 and T4 increase, they prevent the release of TRH, while when T3 and T4 levels drop, the feedback loop starts again. This system allows the body to maintain a constant level of thyroid hormone balance. An imbalance in the hormones associated with this system can occur if there are problems with the hypothalamus, pituitary gland, or thyroid.

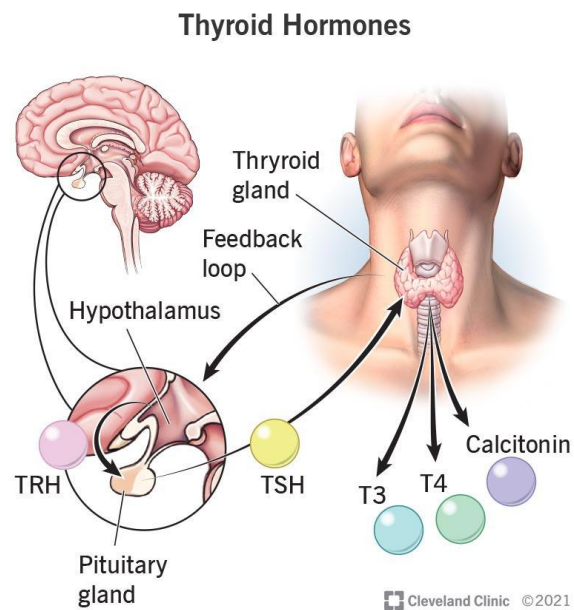


Figure 4 Negative feedback of thyroid hormones. Hypothalamus releases thyrotropin-releasing hormone (TRH), which triggers pituitary gland to release thyroid-stimulating hormone (TSH), which stimulates thyroid to release T3 and T4 [17]

1.7 Hypothyroidism

Several conditions can result from abnormal thyroid hormone levels. Hypothyroidism is one of the most common conditions that is defined as a decreased function of the thyroid gland. It is characterized by TSH levels above the normal range (0.3-4.0 mIU/I). Hypothyroidism can be subclinical or overt. Subclinical hypothyroidism is elevated TSH and normal T4 and T3, while overt hypothyroidism is elevated TSH and T4 and/or T3 below the normal range. Hypothyroidism symptoms can include being tired, gaining weight, and being unable to tolerate cold temperatures. Autoimmune disease, hyperthyroidism therapies, radiation therapy, thyroid surgery, and certain drugs are all possible causes. Hormone replacement therapy is the most common treatment for hypothyroidism. It may not cause noticeable symptoms in the early stages. Untreated hypothyroidism can lead to several health issues over time, including obesity, joint discomfort, infertility, and heart disease.

2. Radiomics in oncology

Radiomics is the extraction of mineable high-dimensional data from radiologic images, and it has been used in oncology to improve diagnosis and prognosis with precision medicine [18]. Since imaging data might have meaningful information about tumor biology, behavior, and pathophysiology. It may reveal information that is not apparent from conventional radiologic images and clinical interpretation. Radiomics quantifies textural information through mathematical extraction of the spatial distribution of signal intensities and pixel relationships, using analysis methods from the field of AI. The workflow of radiomics includes curation of clinical and imaging data, image preprocessing, tumor segmentation, feature extraction, model development, and model validation as show in Figure 5. This research field needs the cooperation of multiple disciplines, including radiologists, image scientists, and data scientists. Radiomics features can be obtained at a single (typically pretreatment) or multiple time points (delta radiomics) and applied to imaging data.

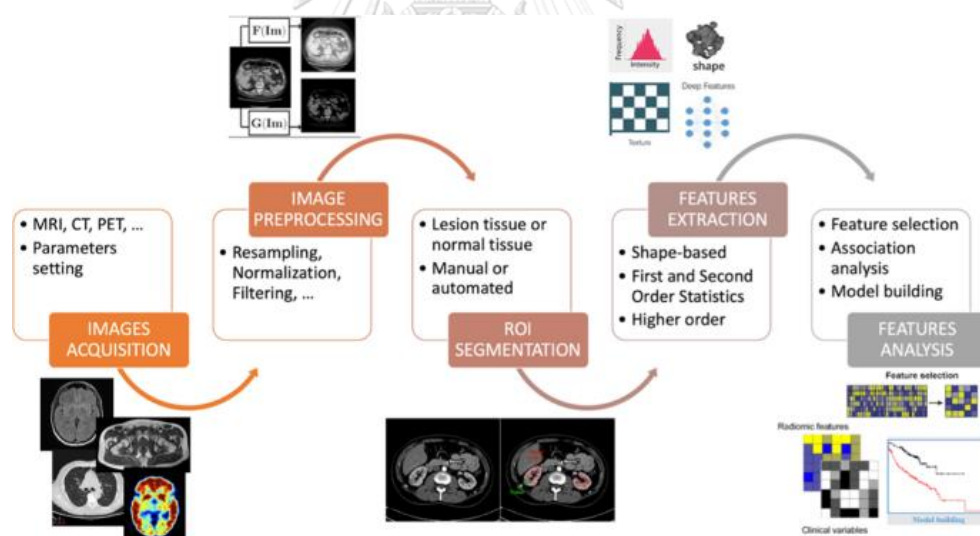


Figure 5 Workflow of radiomics [19]

2.1 Applications in oncology

In oncology, radiomics can be applied to several tasks, for example, classification, treatment response prediction, and side effect prediction. Classification involves dividing subjects into outcome categories such as benign and malignant, normal and disease, tumor stage, presence of metastases. Treatment response prediction models use clinical outcomes to stratify patients into

different risk groups on the basis of clinical endpoint occurrences. For side effect prediction, it involves the treatment toxicity after treatment, such as radiation-induced brain injury and xerostomia.

2.2 Planning a radiomics study

When planning a radiomics study, it needs to propose an interesting clinical task and check how possible it is. The important factor is having adequate data to support the development of a radiomics signature. As a general guideline, in binary classification investigations, 10–15 samples per feature in the final radiomics signature should be obtained. This varies by study, but it is a good rule of thumbs when starting a new project. The rule should be applied to the smaller class if the class sizes are unequal. In fact, when developing a model, it is difficult to control the number of final radiomics signature. It depends on the data and the model algorithm used. After the research question and study population have been selected, collecting pilot data to help detect and minimize potential problems before collecting full data.

2.3 Radiomics workflow

The radiomics workflow is divided into several tasks that are usually performed in this order.

1. Image acquisition

Imaging data acquired from CT, PET, MRI, or US examinations. CT and PET data are signal intensities, which are inherent quantitative data. MRI has good tissue contrast, but several sequence and scanning parameters can affect feature stability. The US is more operator dependent. Therefore, the choice of radiomics image depends on the task and clinical interest.

2. Data curation and image preprocessing

Nonimaging and clinical data are typically collected to be analyzed with radiomics data. Before integrating clinical and radiomics data, curation steps to identify missing or incomplete data can then be taken, also for correction of typographic errors or inconsistencies in data.

Before feature extraction, image data can be enhanced through preprocessing steps. While preprocessing may improve image quality, it must also have an effect on the radiomics signature. In MRI, signal intensity is arbitrary, and hence normalization of signal intensity is recommended. Because of the lack of consensus on this issue, the z-score is a simple way to fix it. The z-score is computed by subtracting the mean signal intensity of the area of

interest (ROI) from the pixel signal intensity and dividing the result by the standard deviation. Bias field correction should also be applied with MRI to correct for the spatial field inhomogeneities encountered. In CT, threshold on voxel Hounsfield units can be applied to image data to exclude voxels that are assumed to contain noninformative tissues, such as low values that may correspond to air within the lung, and high values to bone or calcification. Because some radiomics feature values are influenced by voxel size, all samples should be resampled to the same spatial resolution. The use of linear interpolation is generally advised. Motion correction can be used to correct for misregistration, blurring, or motion artifacts but this additional processing has the potential to impact radiomic information in the images. The use of motion control techniques, such as breath holding, is advised. Before extracting features, image filtration can be used as a preprocessing step to highlight image properties. Wavelet filters, which separate high- and low-spatial-frequency information, and Laplacian of Gaussian (LoG) filters, which emphasize areas of fast change (e.g., edge detection), are two examples.

3. Image segmentation

Tumor delineation is performed by drawing ROIs in tumor areas, tumor subregions, peritumoral zones, or organs of interest according to research hypothesis as show in Figure 6. Radiation therapy tumor volume data used for treatment planning can also be used, but this may differ from ROIs specifically drawn for radiomics analysis. Segmentation can be automatic, manual, or semi-automatic in 2D or 3D. When manual delineation is used, 3D ROIs will capture all the information, but it will be time-consuming to draw. Automatic segmentation is possibly faster and more reproducible, but it may require larger datasets where manual segmentation is not possible. However, this step should be checked by a radiologist for correct radiomics feature calculation. Additionally, in manual segmentation, it may be necessary to perform feature stability by multiple segmentations of the same tumor to evaluate radiomics features independently from the observers.

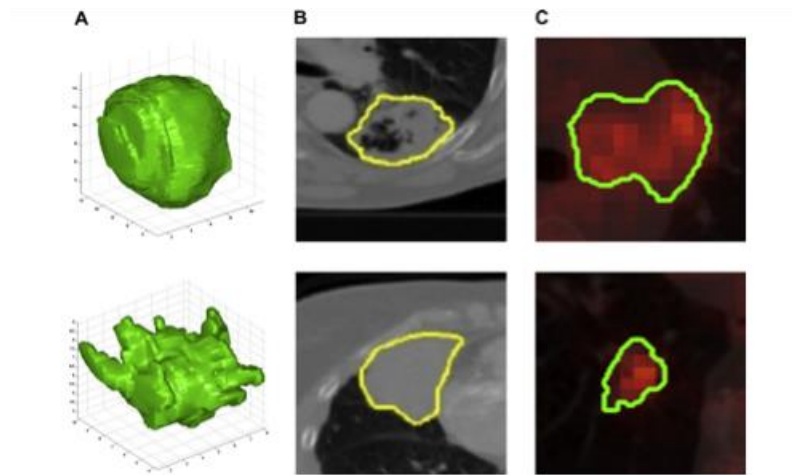


Figure 6 Tumor segmentation. A, two representative 3-D representations of a round tumour (top) and spiky tumour (bottom) measured by computed tomography (CT) imaging. B, Texture differences between non-small cell lung cancer (NSCLC) tumours measured using CT [20]

4. Feature extraction

The final step before model building and validation is feature extraction, which is the calculation of radiomics features from each ROI that will be used in the model. Radiomics features are hand-crafted approaches, whereas deep learning approaches are learned directly from the images. The features were divided into four categories: shape-based, first-order-based, texture-based, and filtered image-based [21].

Shape-based features

This feature type expresses the shape and size of tumors or ROIs. For 3D shape features, these characteristics are only calculated on the non-derived image and mask since they are independent of the gray level intensity distribution in the ROI. The examples of features in this group include mesh volume, voxel volume, surface area to volume ratio, sphericity, etc.

Example: Sphericity is an example of a shape-based feature that is calculated from the equation below.

$$sphericity = \frac{\sqrt[3]{36\pi V^2}}{A},$$

V represents volume and A represents the area of tumors. Sphericity has a value from 0 to 1, so if sphericity equals 1, it means that tumors are circular in shape.

First-order features

First-order features are those that are dependent on the statistical value of voxel intensities in the tumor region, such as mean, standard deviation, skewness, kurtosis, maximum and minimum values, energy, and entropy.

Example: Entropy is defined as a measure of randomness or disorder in the image values.

$$entropy = - \sum_{i=1}^{Ng} p(i) \log_2 (p(i) + \epsilon)$$

Ng is the number of non-zero bins, evenly spaced from 0, with a width specified in the bin width parameter. $P(i)$ be the first order histogram with Ng discrete intensity levels.

Texture features

This group explained the texture of voxel intensities; that is, the spatial locations of the signal intensities of two or more pixels are used when computing the features. For example, gray-level co-occurrence matrix (GLCM) features consider the signal intensities of pairs of pixels separated by a given distance and direction, while gray-level size-zone matrix (GLSZM) features consider the sizes of contiguous regions that share the same signal intensity after discretization. After that, it can calculate features such as energy, local homogeneity, or entropy.

Example: The GLCM (Gray Level Co-occurrence Matrix) functions characterize the texture of an image by calculating how often pairs of pixels with specific values and in a specified spatial relationship occur in an image.

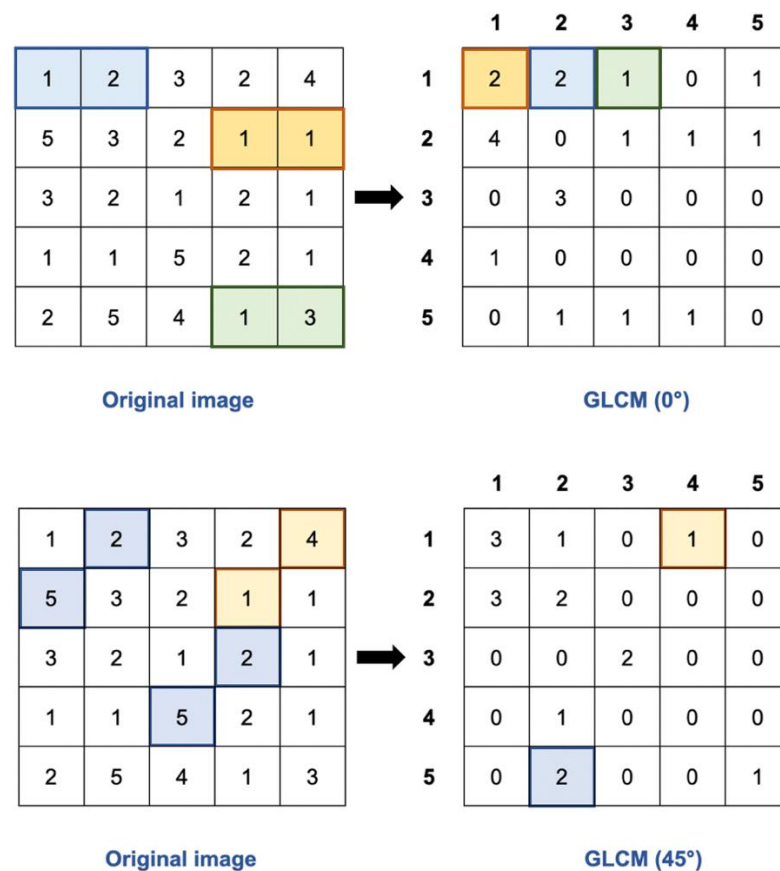


Figure 7 Gray Level Co-occurrence Matrix, which is calculated from the original image in 0° and 45° directions.

Wavelet features

Wavelet is one of the image processing techniques which is used to divide information present on an image (signals) into two discrete components. A signal is passed through two filters, high pass and low pass filters. The image is then decomposed into high frequency (details) and low frequency components. At every level, we get 4 sub-signals as show in Figure 8. The approximation shows an overall trend of pixel values and the details as the horizontal, vertical and diagonal components. After that, it can calculate features using equation of first-order features and texture features.

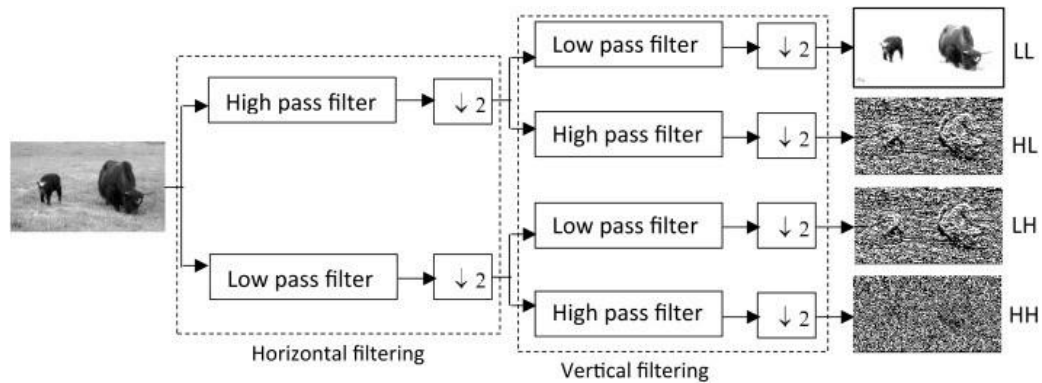


Figure 8 Decomposition of an image 2-D discrete wavelet transform (2-D DWT). [22]

5. Model construction

When clinical and radiomics data are collected, statistical models are built to predict study endpoints, such as tumor type or survival time. A typical model uses input features, including radiomics features and clinical features, to predict outcomes such as benign versus malignant or risk of recurrence. The final models discovered from a radiomics analysis are determined by validating the model on new test data. To avoid model bias, validation data should be kept separate from the model training phase, and final validation should only be performed once. Many models have tuning and optimizing parameters, which is an important step for good model performance. Poorly tuned parameters can lead to overfitting or underfitting of the model.

3. Machine learning

Machine learning is a subfield of artificial intelligence (AI) and computer science that relies on using data and algorithms to simulate human learning processes and progressively increase accuracy. Machine learning is an important component of the development of data science. Using statistical methods, algorithms are trained to make classifications or predictions about the outcome. There are many to ML training methods to choose from including: supervised learning, unsupervised learning, and semi-supervised learning.

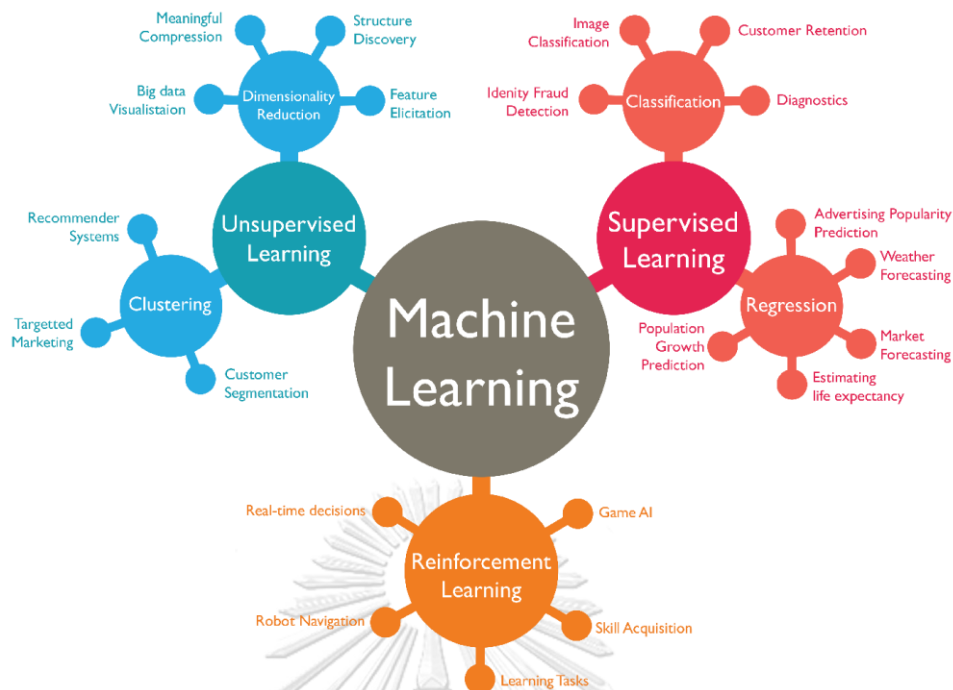


Figure 9 Machine learning training methods [23]

Supervised machine learning algorithms use labeled examples to apply what they have learned in the past to predict future events. The learning method creates an inferred function to predict output values by investigating a known training dataset. After sufficient training, the system can provide targets for any new input. To identify mistakes and make the model more accurate, it can also compare its output with the desired, correct output.

Unsupervised machine learning algorithms are used when the data used to train is neither classified nor labeled. Unsupervised learning investigates how systems might extrapolate a function from unlabeled data to describe a hidden structure. The system cannot be ever certain that the output is correct. Instead, it infers from datasets what the result is supposed to be.

Reinforcement Machine learning algorithms are a type of learning that interactions with its surroundings by taking actions and identifying mistakes or rewards. Trial-and-error learning and delayed rewards are two of reinforcement learning's most important features. With this technique, software agents and machines may automatically decide the best way to proceed in a specific situation in order to enhance performance.

Machine learning algorithms are commonly used, such as linear regression, logistic regression, neural networks, decision trees, random forests, and clustering [24].

Logistic Regression

Logistic Regression is used when the outcome variables are categorical, such as disease or normal, pass or fail, malignant or not. These methods estimate the probability of events by using the logistic function. Logistic regression is developed from linear regression, but linear regression is unbounded and not suitable for classification problems. The outcome value of logistic regression strictly ranges from 0 to 1 [25].

Decision trees

This method is used for both regression and classification problems. As they visually flow like trees, it was given that name. In the classification scenario, they begin at the tree's root and go through binary splits depending on possible outcomes until they reach a leaf node, where the final binary result is given.

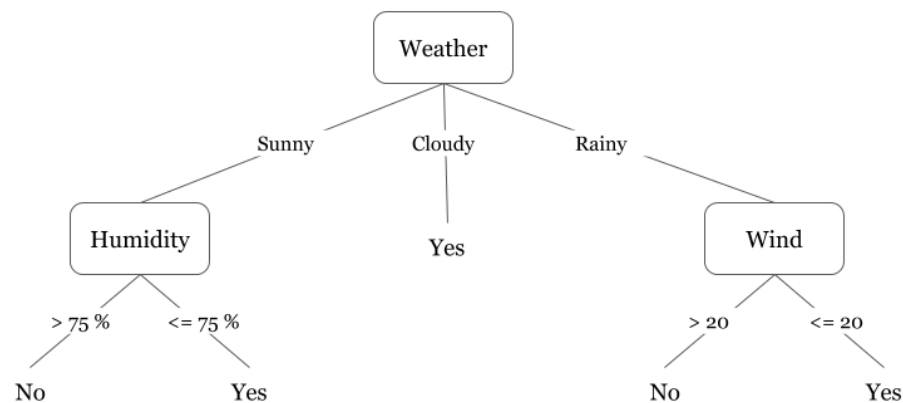


Figure 10 Example of a decision tree [23]

Random Forest

The Random Forest approach uses ensemble learning techniques with the decision tree framework to generate many randomly selected decision trees from the input. The results are

averaged to produce an output that frequently produces accurate predictions and classifications.

Gradient boosting classifier

The primary concept underlying this algorithm is to build models in sequence while attempting to minimize the errors of the prior model. To reduce the errors, building a new model based on the errors or residuals of the old model does this. Gradient Boosting Regressor is used when the target column is continuous; Gradient Boosting Classifier is used when the problem is binary classification [26].



CHAPTER III: RELATED LITERATURE REVIEWS

Late effect complications due to radiotherapy should be considered when planning treatment since they affect quality of life after treatment. Radiation-induced hypothyroidism is one of the most common complications of head and neck cancers. To evaluate the risk of radiation-induced hypothyroidism in nasopharyngeal cancer, we usually use clinical and dosimetric data, but the result was still unsatisfied. For more information, medical images may help improve predictive performance. For the radiotherapy workflow, it is necessary to use imaging to guide treatment planning for tumor delineation, normal structure segmentation, and dose calculation. These imaging data make it possible to enhance the prediction of radiation-induced hypothyroidism. Recently, radiomics has become an interesting research field in many oncologic tasks, including predicting complications due to radiation treatment. For instance, the use of MRI and CT radiomics to predict radiation-induced xerostomia from head and neck cancer treatment and MRI radiomics to early detect radiation-induced brain injury in nasopharyngeal cancer.

In HNC publications, *Haider SP et al* [6]. provided an overview of recent radiomics studies. There has been increasing interest in the application of radiomics for prediction of molecular biomarkers, prognostication, and treatment response in HNC. Classification and regression models are mostly applied to prediction of molecular markers, specification of genomic signatures, diagnostic differentiation of suspected tissue, survival prognostication, and prediction of treatment response. The big data and open-source machine learning algorithms led to the development of new multivariate diagnostic and prognostic biomarkers integrated with radiomics features and clinical information for risk stratification, outcome prediction, and precision treatment planning in HNC. One of the radiomics applications was the prediction of post chemoradiotherapy complications. Side effects that can occur from the treatment include xerostomia, trismus, hearing loss, mucositis, and dermatitis. Identifying patients who are at risk of developing specific side effects may assist oncologists in planning personalized treatment strategies. Many studies have used radiomics biomarkers to predict the occurrence or severity of treatment-related toxicities based on bioimaging features of at-risk organs.

According to *Zhang YM et al* [27], they collected the studies of radiomics for diagnosis and radiotherapy of NPC. The resolution of soft tissue on MRI is better than on CT and PET-CT and clearly shows the range of parapharyngeal space, skull base, and intracranial tumors. Because NPC used MRI as the gold standard for evaluation, most radiomics studies on NPC have focused on MRI images, but CT, PET-CT, and PET-MRI images can also be used. In addition, most studies were retrospective, which provided valuable clinical guidance for diagnosis, differential diagnosis, treatment, recurrence, and prognosis of disease. For NPC diagnosis, the radiomics model that combined clinical data with features extracted from MRI was used to analyze the survival subgroups of early NPC. Moreover, using PET-MRI radiomics revealed the subtle changes in local lesions. Radiomics was helpful not only for the diagnosis of NPC but also for differential diagnoses and treatment response prediction.

For side effect prediction, *Sheikh K et al.* [7] proposed a study for predicting acute radiation-induced xerostomia in HNC using MR and CT radiomics of the parotid and submandibular gland. The patient datasets included 266 HNC patients who were treated with IMRT from 2009–2018. The patients were diagnosed with moderate to severe xerostomia after a 3-month treatment course by physicians using the NCI-CTCAE v4 criteria. The pre-treatment CT and T1W-MR images used for contralateral parotid and submandibular glands delineations by radiation oncologists also by inhouse automate segment software for dose volume histograms (DVH) features. For radiomics extractions, radiomics features were extracted from CT and MR images of bilateral parotid and submandibular glands by pyradiomics software into 5 categories: shape-based, first-order-based, second-order-based (GLCM, GLRLM, and GLSZM), and wavelet filtration based. To develop a prediction model, a generalized linear model with ten-fold cross validation was used for radiation-induced xerostomia. The results showed that the highest model performance was the combined clinical + DVH + CT + MR model with an AUC of 0.79 ± 0.01 in the validation set and 0.68 in the test set, which was no different from combined models without clinical information. The CT + MRI radiomics model was significantly different than the CT or MRI ROC only, and it was approximately equal to the performance of the DVH model. In conclusion, they suggested that the pre-treatment CT and MRI image features combined with the DVH features may reflect baseline salivary gland function and potential risk for radiation injury.

Another radiomics study on side effect prediction was reported by *Zhang B et al* [28]. The study aimed to develop radiomics models for early detection of radiation-induced brain injury in NPC. It included 242 NPC patients who underwent radiotherapy and regular follow-up MRI examinations between January 2006 and August 2016. Radiation-induced temporal lobe injury (RTLTI) was diagnosed based on MRI with a follow-up time of less than 112 months by two independent radiologists. The middle and lower portions of the medial temporal lobe were segmented by open-source software in contrast-enhanced T1-weighted and T2-weighted images. Radiomics features were extracted from the medial temporal lobe, gray matter, and white matter after that feature selection was performed using the relief algorithm. For model construction, they developed three radiomics models, 1, 2, and 3, to predict RTLTI at the last 1, 2, and 3 MRI scans (N-1, N-2, and N-3) before MRI confirmation using random forest and evaluated by AUC values. Predictive performance was compared based on different combinations of segmented tissue and the number of top-ranked features. The results found that the AUC of radiomics models with longitudinal MRI was 0.872, 0.836, and 0.780 for RTLTI in advance. From this result, the radiomics approach allowed us to identify imaging phenotypes and to detect pathophysiological changes. The different MRI measurements contained additional information for which the combination of these measurements may improve the predictive performance of RTLTI.

For radiation-induced hypothyroidism, *Lertbutsayanukul C et al.* [4] validated previously reported dosimetric parameters and clinical factors affecting hypothyroidism after radiotherapy treatment in NPC patients. Participants included 178 NPC patients from October 2010 to September 2015 who were diagnosed with hypothyroidism after radiotherapy treatment. Radiation-induced hypothyroidism was defined as a TSH value outside of the upper limit reference range (reference range 0.3 - 4.2 μ U/ml), with or without reduced FT4 (reference range 0.8 - 1.8 ng/dl), regardless of symptoms. Using cox proportional hazard models with univariate and multivariate analysis to identify the predictors of radiation-induced hypothyroidism. The median latency period of hypothyroidism was 21 months, and the median mean dose of thyroid gland that was received of 53.5 Gy. Female, smaller thyroid volumes, higher pre-treatment TSH more than 1.55 μ U/ml and VS60 <10 cm³ were significantly correlated with radiation-induced hypothyroidism in univariate analysis. In multivariate analysis, only pre-treatment TSH and VS60 were significant predictors. The limitation was mentioned as the follow-

up time that was not long enough to assess late hypothyroidism. Finally, they suggested that a pre-treatment TSH more than $1.55 \mu\text{U/ml}$ should be considered because of the risk of hypothyroidism, and $\text{VS}_{60} < 10 \text{ cm}^3$ is recommended for treatment planning.

Peng L et al. [5] proposed a new model for predicting hypothyroidism after intensity-modulated radiotherapy (IMRT) for NPC. A total of 545 NPC patients treated between 2011 and 2015 were included if they were identified as euthyroid before radiotherapy and were assessed for thyroid function regularly after the treatment course. Dose volume histograms were retrieved from treatment planning, including thyroid volume, dosimetry parameter: V_x (percentage of thyroid volume receiving more than x Gy of radiation) and dosimetry parameter: $V_{a,b}$ (percentage of thyroid volume receiving > 1 Gy, while $< b$ Gy radiation). The outcome was the development of hypothyroidism within two years after IMRT. To identify predictors, they used least absolute shrinkage and selection operator and multivariate logistic regression. The results indicated that the combination of thyroid volume and $V_{30,60}$ could be useful as predictors of radiation-induced hypothyroidism after IMRT in NPC patients by an AUC of 0.643 (0.590-0.695).

Zhai R et al. [29] revealed the importance of clinical and dosimetric factors to predict the risk of hypothyroidism after treatment with IMRT in NPC patients. All 404 non-met NPCs were included in the study. All patients had thyroid function tests before and after radiotherapy. Univariate and multivariate Cox regression analyses were used to identify the main factors for hypothyroidism prediction. The results found that the mean dose of thyroid, V_{30-50} (percentage of thyroid volume receiving dose to thyroid), and VS_{45-60} (the volumes of thyroid spared from various dose levels) were statistically significant in multivariate analyses.

According to **Smczynska U et al.** [30], they developed a radiomics NTCP model for radiation-induced hypothyroidism. Prior to receiving IMRT for OPC, the thyroid function of all 98 patients was normal, and RIH was observed in the patient after treatment for 2 years. CT images and clinical data were used to develop the NTCP model. The results found that radiomics-based models did not outperform state-of-the-art NTCP models ($p > 0.05$).

Based on the aforementioned studies, the use of radiomics with complication effect prediction is possible. Predictive models can be developed using machine learning and statistical approaches. **Sheikh K et al.** and **Zhang B et al.** used a machine learning algorithm which may be

suitable for radiomics features. Since extracted features are numerous when compared to the patient cases, feature selection is important to exclude redundant or uncorrelated features and reduce the probability of overfitting. According to *Peng L et al.* they used the statistical model to predict the probability of hypothyroidism, and the performance was moderate, while *Smyczynska U et al.* showed that CT images did not exceed the NTCP model in oropharyngeal cancers. To improve the performance, it might add radiomics information to the predictive model. For effective personal treatment and side effect reduction, the use of clinical and dosimetric information combined with radiomics data should be a new tool to develop an effective predictive model for radiation-induced hypothyroidism.



Table 1 Related literature reviews

Authors		Applications	Results
Sheikh K et al. [7]	66	Predicting acute radiation-induced xerostomia in HNC using MR and CT radiomics	The pre-treatment CT and MRI image features combined with the DVH features may reflect baseline salivary gland function and potential risk for radiation injury
Zhang B et al. [28]	42	Develop radiomics models for early detection of radiation-induced brain injury in NPC	The different MRI measurements contained additional information for which the combination of these measurements may improve the predictive performance of RTLI
Lertbutsayanukul C et al. [4]	78	Validated previously reported dosimetric parameters and clinical factors affecting hypothyroidism after radiotherapy treatment in NPC patients	Pre-treatment TSH more than $1.55 \mu\text{U/ml}$ should be considered because of the risk of hypothyroidism, and $\text{VS}_{60} < 10 \text{ cm}^3$ is recommended for treatment planning
Peng L et al. [5]	45	Proposed a new model for predicting hypothyroidism after intensity-modulated radiotherapy (IMRT) for NPC	The combination of thyroid volume and $V_{30,60}$ could be useful as predictors of radiation-induced hypothyroidism after IMRT in NPC patients
Zhai R et al. [29]	04	To find clinical and dosimetric indicators that can	The likelihood of hypothyroidism in NPC

Authors		Applications	Results
		indicate a patient's likelihood of developing hypothyroidism after receiving radiotherapy.	patients following IMRT was strongly predicted by thyroid Vt40.
Smyczynska U et al. [30]	8	Using imaging biomarkers, create a radiomics NTCP model for radiation-induced hypothyroidism.	Compared to the NTCP models currently in use, radiomics models based on CT scans demonstrated comparable ability to predict RIHT in OPC patients.
Liu Y et al. [31]	9	Investigating radiation-induced changes of computed tomography (CT) radiomics in parotid glands (PGs) and saliva amount (SA) can predict acute xerostomia during the RT for nasopharyngeal cancer (NPC)	Radiation-induced acute xerostomia level could be early predicted based on the SA and radiomics changes of the PGs during IMRT delivery.

CHAPTER IV: MATERIALS AND METHODS

1. Research design

This study was a retrospective cohort study.

2. Research design model

This thesis was designed to develop a predictive model for RIH in nasopharyngeal carcinoma patients. The radiomics features of the thyroid gland from CT contrast enhanced before treatment, combined with clinical and dosimetric information, were used in model construction. ROC analyses were evaluated the model performance. The scope of the study was defined by the following steps, as depicted in Figure 11.

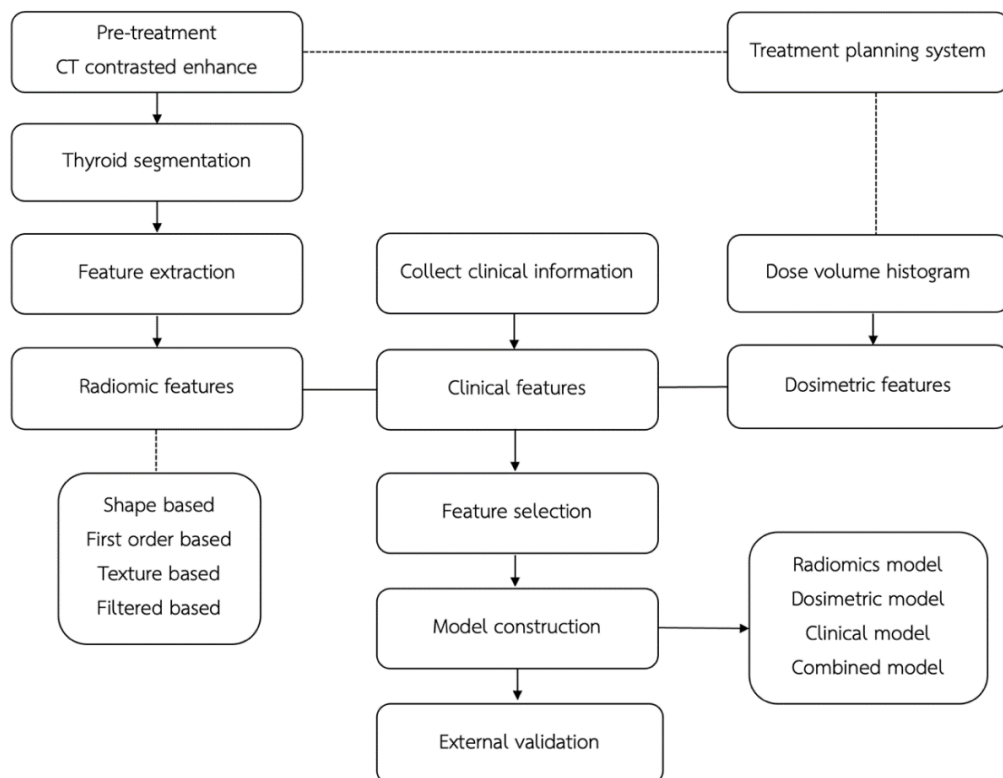


Figure 11 Research design model

Radiomics features

- Collected pretreatment CT contrast-enhanced images from the treatment planning system.

- Manual thyroid segmentation.

- Radiomics features extraction.

Clinical and dosimetry information

- Collected patient information.

- Calculated dosimetry parameters from the treatment planning system.

Model construction and validation

- Feature selected based on the recursive feature elimination technique and fisher information.

- Model construction using radiomics features, clinical features, dosimetry features, and combined features.

- Selected final model based on validation performance.

- Validated in the testing set.

3. Research conceptual framework

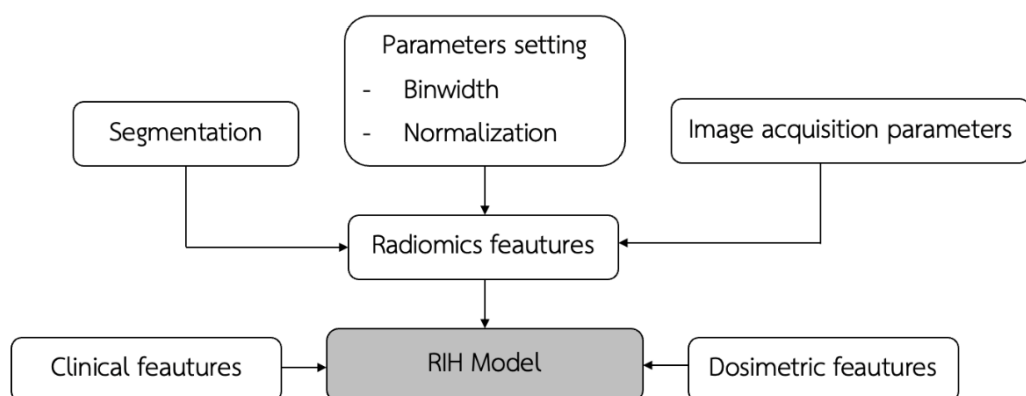


Figure 12 Factors affecting the RIH model

4. Materials and methods

This section discusses the methodology for developing a predictive model. We started with the patient's collection, radiomic feature extraction and the model construction.

Population and sample

A total of 220 patients with NPC who were diagnosed with hypothyroidism within 2 years after radiation therapy treatment from 2010-2020 were included in this study. All participants were chosen from patients with NPC in the Division of Radiation Oncology, Department of Radiology, King Chulalongkorn Memorial Hospital.

The inclusion criteria are as follows:

- 1) NPC patients in King Chulalongkorn Memorial Hospital treated with definite RT (IMRT/VMAT) with or without chemotherapy 70 Gy in 33-35 fractions.
- 2) Age more than 18 years.
- 3) Normal baseline thyroid function test (TFT).
- 4) At least 2-years TFT follow up or less than 2-years TFT follow up but develop RHT

The exclusion criteria are as follows:

- 1) History of pre-existing thyroid disease.
- 2) Abnormal or no baseline TFT.
- 3) History of thyroid surgery.
- 4) History of radiotherapy at neck.

Image acquisition

All NPC patients underwent a CT simulation before radiotherapy treatment. A 64 detector-row CT simulator was used to acquire CT images (Revolution CT; GE Healthcare, Chicago, IL, USA). Acquisition protocols included a noncontrast phase and a contrast-enhanced phase in helical mode at 120 kV with smart mA by 2.5 mm slice thickness.

Thyroid segmentation

Manual 3D segmentation of the thyroid gland was performed by radiation oncologists. The region of interest (ROI) covering the thyroid gland in contrast-enhanced CT images were drawn using the Eclipse Contouring software (Varian Medical System, Inc: version 15.5). To filter out radiomics features that depended on the observer, a multiple delineation test was performed by having three radiologists segmented the thyroid ROI from the same set of thirty

randomly selected patients. The radiomics features were extracted into three sets from the same images to measure the reliability of the feature's value among multiple delineations in thyroid segmentation. The intraclass coefficient (ICC) was used to evaluate the reliability of feature values by comparing the variability of different values for the same subject to the total variation across all ratings and all subjects. Two-way random effects, absolute agreement, and multiple raters or measurements were applied from the open-source Pingouin package [32]. ICC cutoffs of 0.5, 0.75 and 0.9 were then applied to select radiomics features before the model development step as displayed in Figures 13-14.



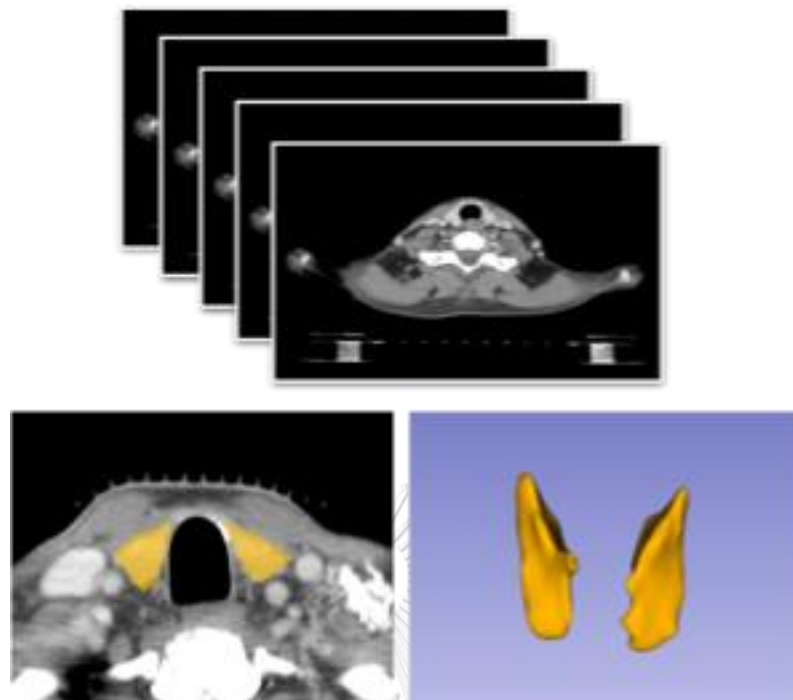


Figure 13 3D thyroid segmentation using 3D slicer software.

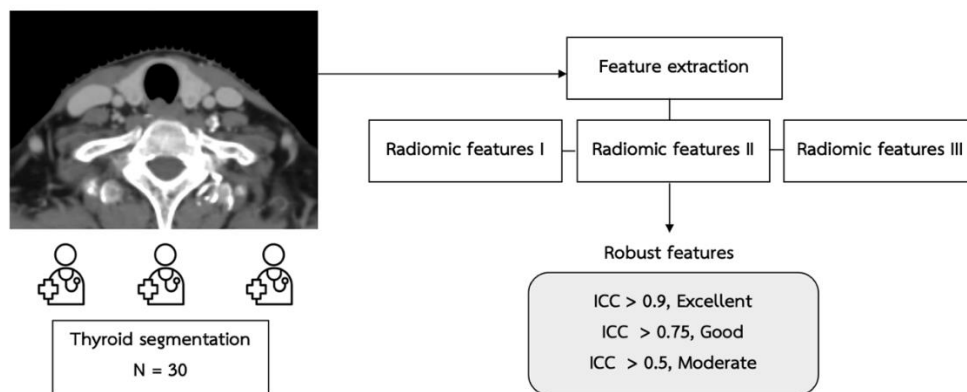


Figure 14 Workflow for ICC testing.

Features extraction

Feature extractions were performed from CT contrast enhanced images using Pyradiomics software package (version 3.0) package accessed using 3D slicer (version 4.11.2) [21, 33]. All radiomics features consisted of 14 shape-based features, 18 first-order statistics features, 73 texture-based features, and 1,183 filtered-based features. The bin width parameter was varied from 0.05, 0.1, 0.15, and 0.2 to achieve the best performance of the radiomics model.

ID	v1	v2	v9	v14	v15	v16	v17	v18	v19	v21	v22	v23	v24	v27	v28	v30	v31
1	0.7945119	0.3482001	30796.151	20786.25	2.8978015	2.610394	1.6716e+11	2.8775746	0.133742	3.1480814	0.0841795	2.5155063	2.5471292	0.0576476	1001.5155	2.089E+10	0.1708948
2	0.1603346	0.156497	94757.667	34797.5	2.1805943	2.5832663	1.797E+11	9.4956814	0.2716756	5.3125587	0.1480167	2.3513991	2.3029072	0.1087592	1002.3512	3.496E+10	0.1094443
3	0.5310416	0.3413896	11234.51	11253.875	2.8022041	3.0659089	0.057E+10	2.8787062	0.1290062	3.3462307	0.0885254	2.9615652	3.0018492	0.0562954	1002.9616	1.132E+10	0.1822967
4	0.7169858	0.351464	15067.417	15088.125	2.5219132	2.7280279	1.213E+11	2.6116478	0.0988217	2.8579994	0.0667632	2.6433179	2.6672845	0.0427484	1002.6433	1.517E+10	0.2080848
5	0.7172086	0.252244	6728.8594	6744.875	2.3613871	1.6003228	5.423E+10	2.9566996	0.1379016	2.9513147	0.0882537	2.4859204	2.5148721	0.0571216	1002.4859	7.79E+09	0.1607386
6	0.7302545	0.3504784	26646.12	28076.875	1.3795486	1.548386	2.14E+11	2.363991	0.0944444	1.9124956	0.0577839	4.4785101	1.5040021	0.0397866	1001.4785	2.67E+10	0.2539853
7	0.6504855	0.2820549	55339.729	15170	2.25236	2.516827	1.219E+11	2.8795634	0.1368155	2.6529238	0.0932378	2.4128622	2.4604762	0.0593603	1002.4129	1.524E+10	0.1851937
8	0.8188396	0.3742291	13157.281	13175	2.6809206	2.5228585	1.06E+11	2.7203904	0.1079435	3.0979915	0.0803383	2.8302931	2.8688643	0.0488959	1002.8303	1.322E+10	0.2086676
9	0.5375503	0.3035125	12970.938	12998.75	1.4603803	1.6690704	1.043E+11	2.737916	0.0896363	2.4025146	0.0690738	1.595247	1.6116522	0.0406828	1001.5953	1.304E+10	0.2240524
10	0.7948307	0.4133155	64249.656	64277.5	1.0440293	1.1606765	5.154E+11	1.9963282	0.0580592	1.5918281	0.0390015	1.1223414	1.1265064	0.0250102	1001.1123	6.442E+09	0.3139071
11	0.5042474	0.4525921	10440.734	10480	2.0004054	2.8963489	8.431E+10	3.2257389	0.1631292	4.5094448	0.1220803	2.7746664	2.802722	0.0699395	1002.7747	1.054E+10	0.1598307
12	0.7198375	0.398893	18312.073	18384.75	1.501746	1.6083979	1.471E+11	1.9164464	0.0555247	1.9265299	0.0385283	1.5648103	1.5795597	0.0235111	1001.5648	1.838E+10	0.3389181
13	0.6967825	0.3669713	34616.552	34645.875	1.2833994	1.4304627	2.779E+11	2.2648975	0.080071	2.3354193	0.0514707	1.3684506	1.3848859	0.0336172	1001.3685	3.474E+10	0.2650421
14	0.8389574	0.3610372	16966.203	16987.5	2.4847475	2.7082659	1.366E+11	2.5451987	0.1046441	3.0223174	0.0719256	2.628383	2.6652013	0.0463551	1002.6284	1.08E+10	0.2374202
15	0.6729358	0.2426435	15027.474	15059.625	1.271346	1.4268457	1.208E+11	2.2490783	0.0775717	1.7542486	0.0527439	1.366595	1.3905025	0.0337884	1001.3666	1.551E+10	0.2716556
16	0.72058	0.322306	18351.432	18382.5	1.4350618	1.5836257	1.473E+11	2.2476146	0.0847035	1.8210833	0.0503095	1.5314515	1.501438	0.0352966	1001.5164	1.844E+10	0.2538056
17	0.6520316	0.2558815	14187.323	14210	1.9841279	2.2546242	1.142E+11	3.0601743	0.144278	2.7522178	0.0935772	1.1434722	2.1811994	0.0613862	1002.1455	1.427E+10	0.1621265
18	0.7249918	0.3850284	9898.8698	9924	2.6235102	2.945238	7.984E+10	3.2001635	0.1973537	3.0442577	0.1118575	2.8092788	2.8498852	0.0800851	1002.8093	9.98E+09	0.1416371
19	0.6869122	0.3275167	12308.651	12334	2.3038559	2.606895	9.916E+10	3.0529194	0.173126	2.9355727	0.1020975	2.4892159	2.5325834	0.0729029	1002.4892	1.24E+10	0.1550866
20	0.5995141	0.2873665	22058.839	22088.75	1.3181275	1.4712478	1.772E+11	2.1526513	0.0620036	1.8400496	0.0479212	1.4151364	1.4365235	0.0291648	1001.4151	2.21E+10	0.2873533
21	0.7795405	0.3561137	11104.245	11135	2.2194412	2.4894956	8.95E+10	2.9600326	0.1510216	2.5495066	0.0884747	2.3677133	2.3962346	0.0631138	1002.3677	1.119E+10	0.1576085
22	0.6768875	0.3283174	7637.0833	7647.5	2.0924638	2.3797232	6.146E+10	3.0604816	0.1678012	2.4711414	0.092065	2.2469094	2.2681312	0.0682858	1002.2469	1.682E+09	0.1378483
23	0.7617016	0.2740392	10677.052	10703.5	2.7103221	2.8795165	8.611E+10	2.4431608	0.0807213	3.2400955	0.0559712	1.804826	2.8223955	0.0349992	1002.8048	1.07E+10	0.2338086
24	0.6816427	0.292341	11699.099	11722.375	1.2990601	1.4762574	9.404E+10	2.2998507	0.0940822	1.7294936	0.0584813	1.4116722	1.4406683	0.041011	1001.4117	1.57E+10	0.2609236
25	0.6470018	0.2418959	15325.875	15362.5	1.0729191	1.293471	1.232E+11	2.6059857	0.1299712	1.4561295	0.0752389	1.20512	1.2359985	0.0539136	1001.2052	1.54E+10	0.2221367
26	0.875086	0.2548271	12012.698	12037.5	2.5490161	2.7817591	9.682E+10	2.6901499	0.1131737	3.4003146	0.0772727	2.6937689	2.7299571	0.0499939	1002.6938	1.21E+10	0.2090152

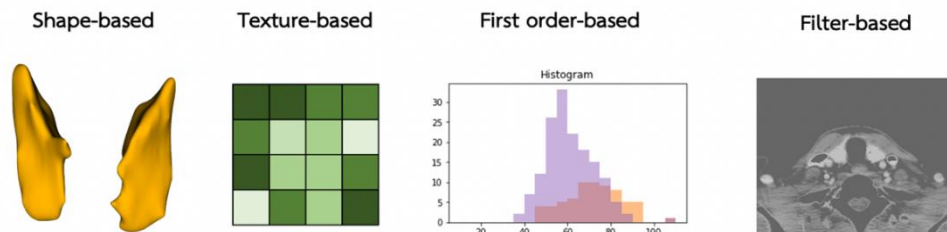


Figure 15 Example of radiomics features from CT images, which were divided into 4 classes.

Right and left lobe thyroid glands in radiomics feature values

To investigate the difference of left and right in the thyroid gland in radiomics feature value, we evaluated the difference in radiomics features between two lobes by randomly selecting 30 cases and calculating the radiomics features in each lobe to compare the average feature value.

Clinical information and dosimetry information

Clinical variables were collected from the hospital information system. Dosimetric variables were calculated from the dose volume histogram via the treatment planning system, which are V_{40} , V_{50} , V_{60} , Pit_{50} , Pit_{55} (V_x : Percentage of thyroid volume that has received at least x Gy radiation, Pit_x : Percentage of pituitary volume that has received at least x Gy radiation), VS_{40} , VS_{50} , VS_{60} (VS_x : Percentage of thyroid volume preserved from x Gy of radiation), the mean dose of thyroid and pituitary gland, the maximum dose of thyroid and pituitary gland, and the minimum dose of thyroid and pituitary gland dose.

Model construction

The radiomics features were screened using an intraclass correlation greater than 0.5, 0.75 and 0.9 to ensure that only robust features were used, given the manual segmentation. For model development, the data was divided into two groups: training and testing, which were split 80:20 %. Three predictive algorithms, logistic regression with regularization, random forest, and gradient boosting classifier were performed by the scikit-learn library [34]. The recursive elimination method and fisher information were utilized in the feature selection process. The models were internally validated using 5-fold cross-validation repeated 20 times. Their performance was evaluated using ROC analysis in terms of area under the curve (AUC). In addition to the radiomics model, clinical, dosimetric and combined models were built based on the two predictive algorithms shown in Figure 16.

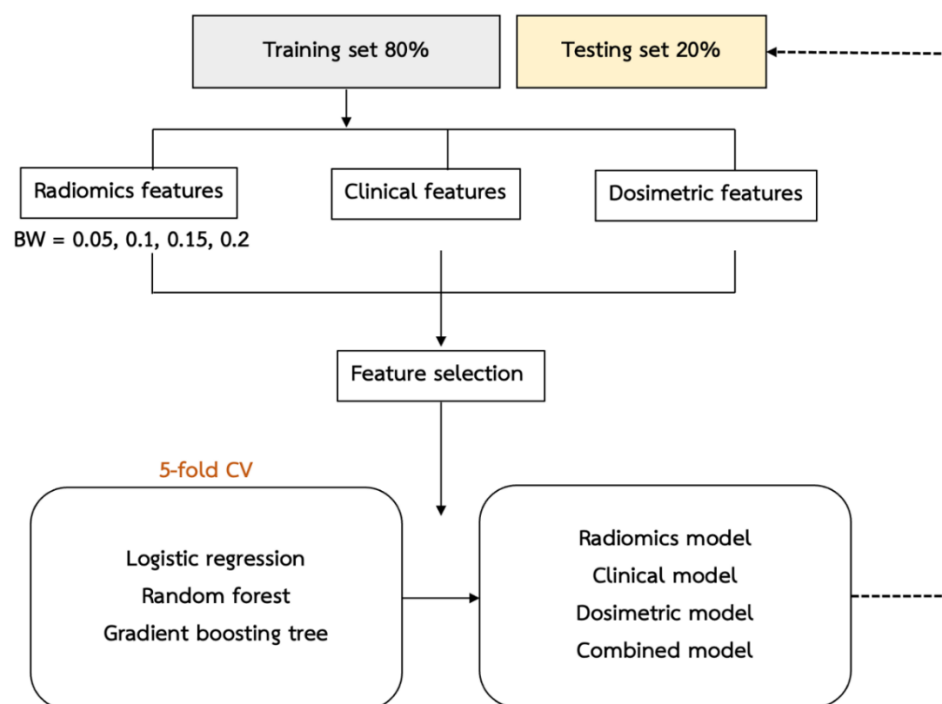


Figure 16 Model construction workflow

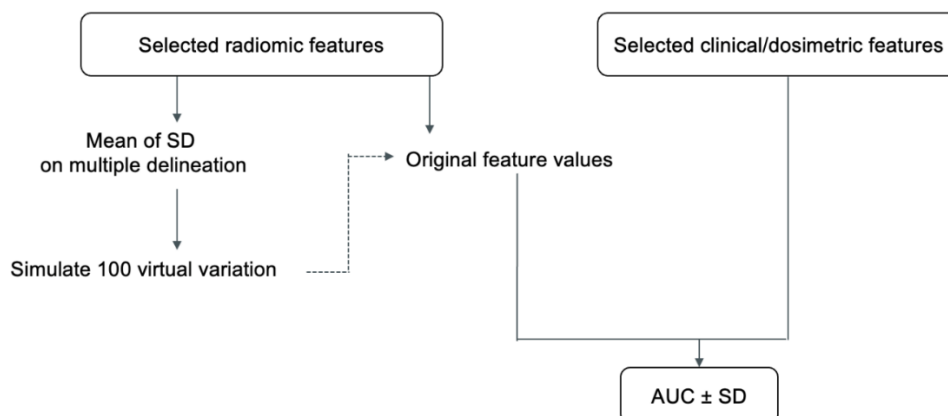


Figure 17 Testing model robustness to variation in segmentation workflow

Testing the model robustness for variation in segmentation

To test the model robustness for variation among observers in segmentation, radiomics features in the final model were perturbed by adding a zero-mean gaussian noise with standard deviation estimated from multiple delineation testing which were randomly selected from 30 patients and 3 observers to segment. The final model was tested with perturbed radiomics features to evaluate the mean and standard deviation of the AUC value as shown in Figure 17.

Statistical analysis

The mean and standard deviation (SD) values were calculated for continuous variables, while the counts and percentages were used to summarize categorical features. The differences in feature values between patients with and without RIH were evaluated using the Mann-Whitney U tests and Chi-square tests. The difference in predictive performance of the models were evaluated using signed rank tests. The Benjamini-Hochberg procedure was performed to control for multiple testing. A p -value cutoff of 0.05 was set to define statistical significance.

5. Ethical consideration

The current investigation was conducted using a retrospective cohort design and was approved by the Institutional Review Board of the Faculty of Medicine of Chulalongkorn University (IRB Number: 745/61). All methods were performed in accordance with relevant guidelines and regulations.

CHAPTER V: RESULTS

Table 2 Patient Characteristics

	mean±SD	RIH (N=106)	No RIH (N=114)	P-value
Age (years)	48.28±11.71 (18-83)	46.53 ± 11.22 (21-70)	49.90 ± 11.96 (18-83)	0.05
Sex (%)				
Male/Female	72.27/27.73	66.98/33.02	77.19/22.81	0.10
T stage (%)				
T1	31.36	33.01	29.82	0.36
T2	33.18	34.91	31.58	
T3	7.73	16.04	19.30	
T4	17.73	16.04	19.30	
N stage (%)				0.06
N0	8.18	4.72	11.40	
N1	28.64	23.58	33.34	
N2	51.82	57.55	46.49	
N3	11.36	14.15	8.77	
Clinical stage (%)				0.73
I	1.82	1.89	1.75	
II	17.27	16.04	18.42	
III	51.82	52.83	50.88	

	mean±SD	RIH (N=106)	No RIH (N=114)	P-value
IVA	19.55	17.92	21.05	
IVB	9.54	11.32	7.90	
FT4 before treatment (ng/dl)	1.25 ± 0.22 (0.42-1.80)	1.24 ± 0.23 (0.42-1.68)	1.25 ± 0.22 (0.68-1.78)	0.45
TSH before treatment (μU/ml)	2.24 ± 5.19 (0.31-3.55)	2.68 ± 7.08 (0.46-3.55)	1.83 ± 2.30 (0.31-2.23)	< 0.05
Thyroid volume before treatment (cm ³)	14.78 ± 6.82 (5.00-61.80)	13.23 ± 6.43 (5.50-46.70)	15.06 ± 7.07 (5.00-61.80)	< 0.05
Pituitary volume before treatment (cm ³)	0.25 ± 0.17 (0.00-1.00)	0.26 ± 0.18 (0.00-1.00)	0.24 ± 0.16 (0.00-0.80)	0.61

Table 3 Dosimetry parameters

Dosimetry parameters	mean±SD	RIH (N=92)	No RIH (N=103)	p-value
Minimum dose to thyroid gland (Gy)	34.41 ± 7.13 (2.64-7.13)	34.82 ± 6.68 (2.64-53.10)	34.02 ± 7.55 (10.94-51.80)	0.41
Maximum dose to the thyroid gland (Gy)	66.34 ± 7.76 (46.69-101.90)	66.93 ± 7.55 (53.60-91.90)	65.79 ± 7.94 (46.69-101.90)	0.19
Mean dose to thyroid gland (Gy)	52.36 ± 6.68 (29.80-74.20)	53.05 ± 6.11 (29.80-71.60)	51.72 ± 7.14 (32.80-74.20)	0.10
TR V ₄₀ (%)	88.93 ± 16.31 (8.90-100.00)	91.16 ± 13.38 (41.20-100.00)	86.86 ± 18.45 (8.90-100.00)	0.19
TR V ₅₀ (%)	64.93 ± 24.45 (0.00-100.00)	67.08 ± 23.10 (10.00-100.00)	62.94 ± 25.57 (0.00-99.70)	0.24
TR V ₆₀ (%)	14.89 ± 19.70 (0.00-83.40)	17.34 ± 20.97 (0.00-83.40)	12.62 ± 18.23 (0.00-78.60)	0.08
TR VS ₄₀ (%)	16.31 ± 21.49 (0.00-71.96)	14.07 ± 20.30 (0.00-56.63)	18.39 ± 22.42 (0.00-71.96)	0.12
TR VS ₅₀ (%)	44.59 ± 19.44 (0.00-74.20)	46.07 ± 18.64 (0.00-71.60)	43.21 ± 20.13 (0.08-74.20)	0.17
TR VS ₆₀ (%)	45.83 ± 17.19 (5.49-74.20)	47.42 ± 16.24 (5.49-71.60)	44.35 ± 17.98 (6.73-74.20)	0.12
Minimum dose to the pituitary gland (Gy)	45.44 ± 20.17 (4.45-96.70)	44.11 ± 20.51 (4.45-75.00)	46.67 ± 19.86 (9.40-96.70)	0.42

Dosimetry parameters	mean±SD	RIH (N=92)	No RIH (N=103)	p-value
Maximum dose to the pituitary gland (Gy)	60.80 ± 15.45 (7.00-102.00)	58.97 ± 16.68 (7.00-84.50)	62.49 ± 14.08 (16.40-102.00)	0.16
Mean dose to the pituitary gland (Gy)	52.91 ± 17.68 (0.50-100.60)	51.34 ± 18.14 (6.10-82.50)	54.36 ± 17.19 (0.50-100.60)	0.28
Pit V ₅₀ (%)	64.98 ± 42.78 (0.00-100.00)	64.33 ± 43.03 (0.00-100.00)	65.58 ± 42.72 (0.00-100.00)	0.85
Pit V ₅₅ (%)	56.12 ± 44.93 (0.00-100.00)	56.14 ± 45.03 (0.00-100.00)	56.11 ± 45.03 (0.00-100.00)	0.86

TR V₄₀, V₄₅, V₅₀, V₆₀ = Percentage of thyroid volume that has received at least 40, 45, 50, 60 Gy

TR VS₄₀, VS₅₀, VS₆₀ = Percentage of thyroid volume preserved from 40, 50, 60 Gy

Pit V₅₀, V₅₅ = Percentage of pituitary volume that has received at least 50, 55 Gy

Table 2 provides patient information for the 220 participants, which consisted of NPC patients who had a RIH of 106 (48.18%) and non-RIH of 114 (51.82%). The average age was 48.28 ± 11.71 years, and men outnumbered women by 159 (72.27%) to 61 (27.73%). For TNM staging, most patients were clinically staged 3, with a mean of 51.82%, a T2 stage of 33.18%, and a N2 stage of 51.82%, respectively. Before treatment, the mean FT4 level was 1.25 ± 0.22 ng/dl, and the mean TSH level was 2.24 ± 5.19 μ U/ml. The mean volumes of the pituitary and thyroid glands before RT were 0.25 ± 0.17 cm³ and 14.78 ± 6.82 cm³, respectively. Figures 18-19 show two clinical variables. TSH level before treatment and thyroid volume before treatment, which showed a statistical difference between patients with RIH and non-RIH patients (p-value < 0.05). As indicated in Table 3, dosimetry data from the dose volume histogram were recorded from treatment planning and did not reveal any differences between the RIH and non-RIH groups.

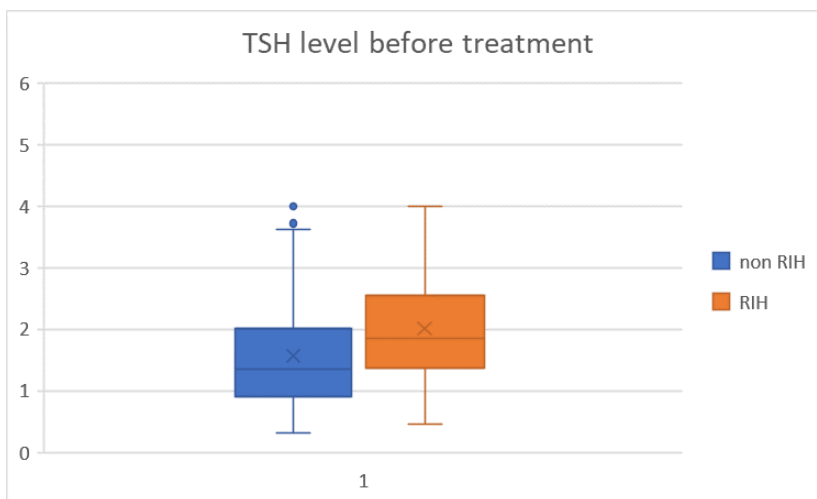


Figure 18 TSH level before treatment

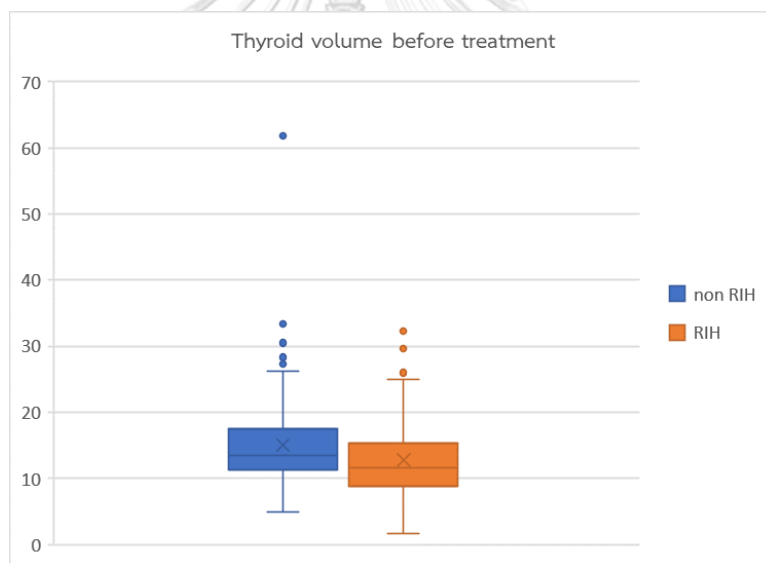


Figure 19 Thyroid volume before treatment

Table 4 Number of radiomics features after filtration from the intraclass correlation test

Feature class	No. of features	ICC > 0.90	ICC > 0.75	ICC > 0.50
Shape based	14	4	4	5
First order statistics	18	14	14	17
Texture based	73	27	44	63
Filter-based	1183	573	776	941
Total	1288	617	838	1026

- Less than 0.50: Poor reliability

- Between 0.5 and 0.75: Moderate reliability

- Between 0.75 and 0.9: Good reliability

- Greater than 0.9: Excellent reliability

Pretreatment CT contrast-enhanced images were extracted into four classes of 1,288 radiomics features. To use features that did not depend on the observer, we kept features based on the ICC value. ICC values greater than 0.5, 0.75, and 0.9 were used to reduce the number of radiomics features to 617, 838 and 1026 radiomics features, as demonstrated in Table 4.

Table 5 Number of radiomics features that differ in RIH and non-RIH groups in each bin width.

Feature class	No. of features that that differ in RIH and non-RIH groups (p -value < 0.05)			
	BW 0.05	BW 0.1	BW 0.15	BW 0.2
Shape based	1	1	1	1
First order based	1	3	1	1
Texture based	3	12	2	2
Filtered base	475	461	489	484
Total	480	477	493	488

Table 6 Radiomics feature univariate analysis

Radiomics features	mean±SD	RIH	non-RIH	AUC
BW 0.05: wavelet- HLL_glcm_MaximumProbability	42.23 x 10 ⁻² ± 0.03	43.23 x 10 ⁻² ± 0.04	41.31 x 10 ⁻² ± 0.03	0.64
BW 0.1: log-sigma-1-0-mm- 3D_ngtdm_Coarseness	13.99 x 10 ⁻⁵ ± 0.00	15.39 x 10 ⁻⁵ ± 0.00	12.68 x 10 ⁻⁵ ± 0.00	0.64
BW 0.15: wavelet-LLH_ngtdm_Strength	9.22 x 10 ⁻⁵ ± 0.00	10.30 x 10 ⁻⁵ ± 0.00	8.18 x 10 ⁻⁵ ± 0.00	0.65
BW 0.2: wavelet-LLH_ngtdm_Strength	8.91 x 10 ⁻⁵ ± 0.00	9.88 x 10 ⁻⁵ ± 0.00	8.02 x 10 ⁻⁵ ± 0.00	0.65

The number of radiomics features with a statistical difference (p -value less than 0.05) in each bin width between RIH and non-RIH groups is shown in Table 5. Only one feature was different in the shaped-based class, 1-3 features in the first-order class, 2–12 features in the texture-based class, and 461-489 features in the filtered-based class. Univariate analysis of radiomics features indicated that while the values of several features significantly differ between patients with and without RIH, they are only moderately predictive of RIH (Table 6, AUC = 0.64-0.65). Highly predictive radiomics features include the wavelet-HLL_glcm_MaximumProbability, log-sigma-1-0-mm-3D_ngtdm_Coarseness, wavelet-LLH_ngtdm_Strength, and wavelet-LLH_ngtdm_Strength.

Table 7 Model performance in training and validation sets based on recursive elimination technique.

Data Type			Train AUC	Validation AUC	p-value
Dose	Clinical	Radiomics			
Logistic Regression					
+	-	-	0.66 ± 0.02	0.63 ± 0.06	< 0.05
-	+	-	0.67 ± 0.01	0.65 ± 0.07	< 0.05
+	+	-	0.74 ± 0.01	0.68 ± 0.07	-
-	-	+	0.82 ± 0.01	0.71 ± 0.07	< 0.05
+	-	+	0.79 ± 0.02	0.71 ± 0.07	< 0.05
-	+	+	0.87 ± 0.01	0.78 ± 0.07	< 0.05
+	+	+	0.88 ± 0.01	0.80 ± 0.06	< 0.05
Random Forest					
+	-	-	0.84 ± 0.01	0.51 ± 0.06	< 0.05
-	+	-	0.83 ± 0.01	0.69 ± 0.07	< 0.05
+	+	-	0.83 ± 0.02	0.71 ± 0.06	-
-	-	+	1.00 ± 0.00	0.78 ± 0.06	< 0.05
+	-	+	1.00 ± 0.00	0.78 ± 0.06	< 0.05
-	+	+	1.00 ± 0.00	0.80 ± 0.06	< 0.05
+	+	+	1.00 ± 0.00	0.81 ± 0.06	< 0.05
Gradient boosting classification					
+	-	-	0.97 ± 0.03	0.49 ± 0.11	< 0.05
-	+	-	0.99 ± 0.01	0.63 ± 0.13	< 0.05

Data Type			Train AUC	Validation AUC	<i>p</i> -value
Dose	Clinical	Radiomics			
+	+	-	0.95 ± 0.04	0.66 ± 0.02	-
-	-	+	0.95 ± 0.04	0.78 ± 0.09	< 0.05
+	-	+	0.95 ± 0.02	0.73 ± 0.05	< 0.05
-	+	+	0.97 ± 0.04	0.71 ± 0.06	< 0.05
+	+	+	0.93 ± 0.04	0.73 ± 0.05	< 0.05



Table 8 Model performance in training and validation sets based on fisher information.

Data Type			Train AUC	Validation AUC	p-value
Dose	Clinical	Radiomics			
Logistic Regression					
+	-	-	0.65 ± 0.01	0.63 ± 0.09	< 0.05
-	+	-	0.66 ± 0.01	0.65 ± 0.06	< 0.05
+	+	-	0.68 ± 0.02	0.68 ± 0.10	-
-	-	+	0.63 ± 0.01	0.62 ± 0.07	< 0.05
+	-	+	0.66 ± 0.02	0.64 ± 0.07	< 0.05
-	+	+	0.68 ± 0.02	0.65 ± 0.06	< 0.05
+	+	+	0.71 ± 0.02	0.70 ± 0.07	< 0.05
Random Forest					
+	-	-	0.82 ± 0.00	0.61 ± 0.08	< 0.05
-	+	-	0.83 ± 0.01	0.70 ± 0.07	< 0.05
+	+	-	1.00 ± 0.00	0.48 ± 0.06	-
-	-	+	1.00 ± 0.00	0.72 ± 0.06	< 0.05
+	-	+	0.86 ± 0.01	0.74 ± 0.07	< 0.05
-	+	+	0.97 ± 0.01	0.70 ± 0.07	< 0.05
+	+	+	0.95 ± 0.00	0.76 ± 0.07	< 0.05
Gradient boosting classification					
+	-	-	0.57 ± 0.00	0.48 ± 0.10	< 0.05
-	+	-	0.85 ± 0.00	0.63 ± 0.09	< 0.05
+	+	-	0.92 ± 0.04	0.58 ± 0.11	-

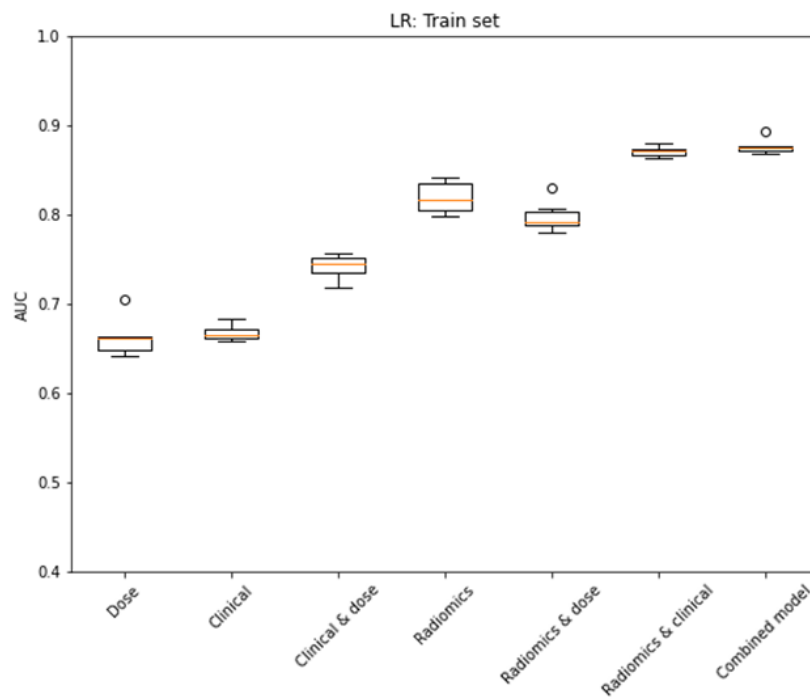
Data Type			Train AUC	Validation AUC	<i>p</i> -value
Dose	Clinical	Radiomics			
-	-	+	0.89 ± 0.06	0.67 ± 0.07	< 0.05
+	-	+	1.00 ± 0.00	0.71 ± 0.07	< 0.05
-	+	+	0.95 ± 0.01	0.64 ± 0.09	> 0.05
+	+	+	1.00 ± 0.00	0.70 ± 0.06	> 0.05



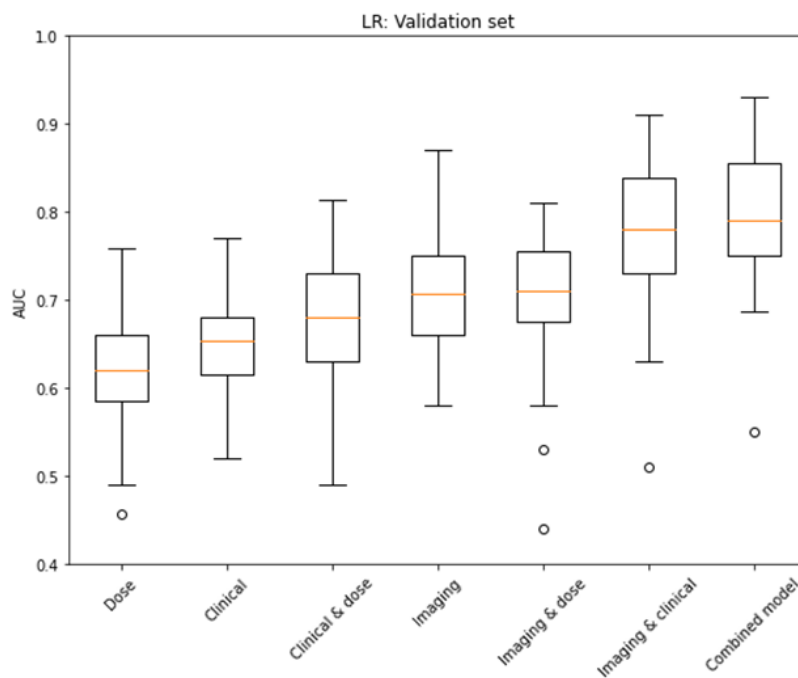
Model performance

Model performance in training and validation sets is displayed in Table 7-8. For recursive elimination technique in feature selection step, the combined model had the best validation performance in logistic regression, as indicated by an AUC of 0.80 ± 0.06 , which was higher than the dose, clinical and clinical & dose, which had AUCs of 0.63 ± 0.06 , 0.65 ± 0.07 , and 0.68 ± 0.07 , respectively. The radiomics had an AUC of 0.71 ± 0.07 , which is comparable to the radiomics & dose, while the AUC for the radiomics & clinical is 0.78 ± 0.07 . The results of the random forest demonstrate the same pattern as the logistic regression. The combined model had the highest AUC of 0.81 ± 0.06 , which was higher than the dose, clinical, and clinical & dose, which had AUCs of 0.51 ± 0.06 , 0.69 ± 0.07 , 0.71 ± 0.06 , respectively. The radiomics in the random forest algorithm received an AUC of 0.78 ± 0.06 , which is equal to the radiomics & dose, while the AUC for the radiomics & clinical was 0.80 ± 0.06 . In gradient-boosting classification, the radiomics model had the highest AUC of 0.78 ± 0.09 , which is greater than the dose, clinical, and clinical & dose models in AUCs of 0.49 ± 0.11 , 0.63 ± 0.13 , and 0.66 ± 0.02 , respectively. The performance of radiomics & dose, radiomics & clinical, and the combined model were better than the conventional model but not greater than the radiomics model.

Based on fisher information in the feature selection step, as shown in Table 8, the combined model had the greatest validation performance with an AUC of 0.70 ± 0.07 in logistic regression, which was slightly more than the dose, clinical, and clinical & dose models by AUCs of 0.63 ± 0.09 , 0.65 ± 0.06 , and 0.68 ± 0.10 , respectively. Radiomics model, radiomics & dose, and radiomics & clinical had AUCs in validation performance of 0.62 ± 0.07 , 0.64 ± 0.07 , and 0.65 ± 0.06 . In random forest, the combined data type also showed the best validation performance with an AUC of 0.76 ± 0.07 , which was greater than the dose, clinical, and clinical & dose models with AUCs of 0.61 ± 0.08 , 0.70 ± 0.07 , and 0.48 ± 0.06 . Radiomics model, radiomics & dose, and radiomics & clinical had AUCs of 0.72 ± 0.06 , 0.74 ± 0.07 , and 0.70 ± 0.07 , respectively. In the gradient boosting classifier, the radiomics & dose model had validation performance comparable to the combined model with an AUC of 0.71 ± 0.07 and 0.70 ± 0.06 which was higher than the dose, clinical, clinical & dose, and radiomics & clinic models with AUCs of 0.48 ± 0.10 , 0.63 ± 0.09 , 0.58 ± 0.11 , and 0.64 ± 0.09 , respectively.

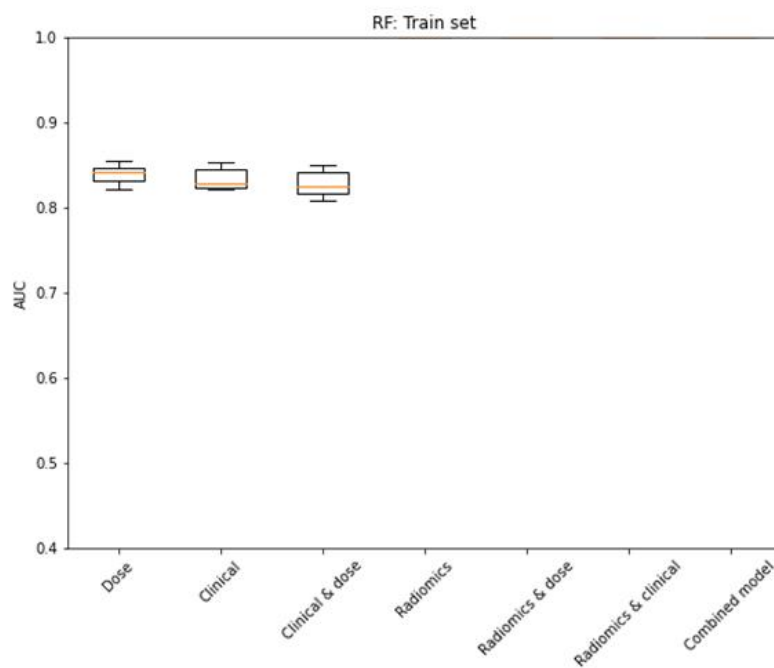


(a)

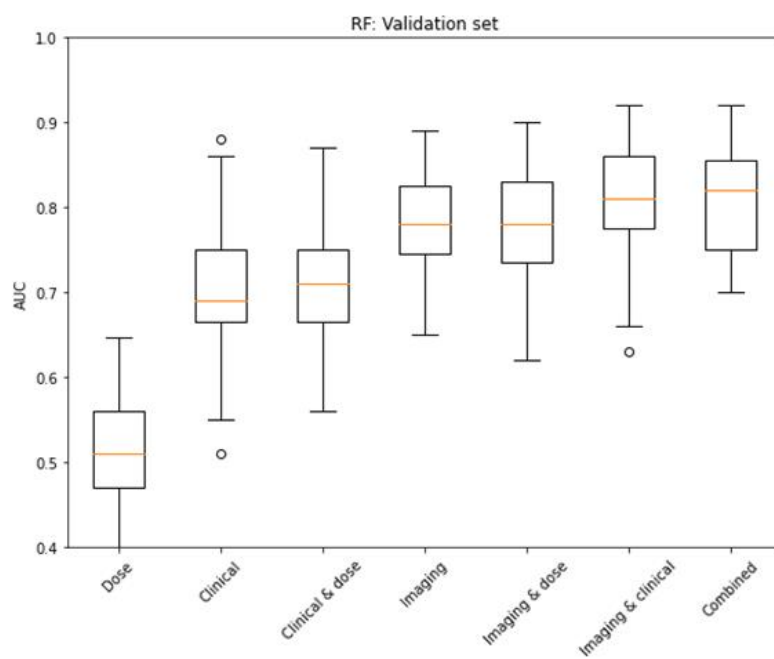


(b)

Figure 20 Performance of the combined model compared to other models in the train set (a) and the validation set (b) in the logistic regression

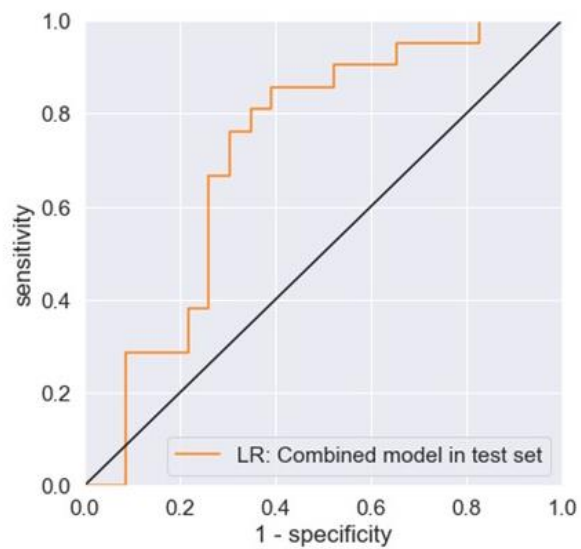


(a)

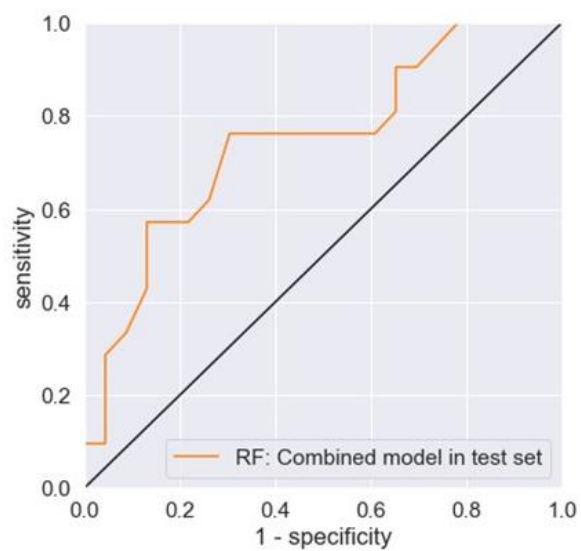


(b)

Figure 21 Performance of the combined model compared to other models in the train set (a) and the validation set (b) in the random forest



(a)



(b)

Figure 22 Performance of the combined model in the test set: (a) logistic regression, (b) random forest

Table 9 Model performance in the test dataset in logistic regression

Model	AUC	accuracy	precision	F1-score	recall	specificity
Dose	0.49	0.50	0.50	0.50	0.52	0.50
Clinical	0.62	0.64	0.68	0.62	0.86	0.44
Clinical & dose	0.63	0.57	0.57	0.57	0.57	0.57
Radiomics	0.70	0.66	0.70	0.65	0.86	0.48
Radiomics & dose	0.71	0.64	0.64	0.64	0.67	0.61
Radiomics & clinical	0.71	0.70	0.72	0.70	0.81	0.61
Combined	0.72	0.72	0.74	0.73	0.81	0.68

Table 10 Model performance in the test dataset in random forest

Model	AUC	accuracy	precision	F1-score	recall	specificity
Dose	0.45	0.45	0.45	0.45	0.43	0.50
Clinical	0.62	0.61	0.62	0.61	0.67	0.57
Clinical & dose	0.67	0.68	0.69	0.68	0.76	0.61
Radiomics	0.70	0.61	0.61	0.61	0.57	0.61
Radiomics & dose	0.70	0.64	0.64	0.64	0.67	0.61
Radiomics & clinical	0.73	0.70	0.71	0.70	0.71	0.70
Combined	0.74	0.73	0.73	0.73	0.76	0.70

		Actual class					
		RIH	non-RIH			RIH	non-RIH
Predicted class	RIH	17	8	Predicted class	RIH	16	5
	Non-RIH	4	15		Non-RIH	7	16

Figure 23 Confusion matrices for the best combined models on the held-out test dataset. (a) The combined logistic regression model. (b) The combined random forest model.

The top performing models, namely the combined logistic regression model and the combined random forest model, were then evaluated on a held-out test dataset (Table 9-10). The combined models again outperformed the models variants that utilized only clinical and dosimetric data in almost all metrics. The only exception is sensitivity where the combined models did not achieve the best performances. It should be noted that although the combined logistic regression model and the combined random forest model achieved similar AUCs (0.72 and 0.74, respectively), they have different tradeoffs. While the combined random forest model achieved higher sensitivity in the high specificity range (>0.70), the opposite is true in the intermediate specificity range. The confusion matrices (Figure 23) also suggested that the combined logistic regression model tends to produce slightly more false positives than false negatives while the combined random forest model behaves in the opposite manner. Hence, multiple metrics should be considered when selecting the best model and cutoff value.

Table 11 Optimal threshold points of combined model

Model type	Threshold	TPR	FPR	1-FPR
LR: Combined model	0.61	0.67	0.30	0.70
RF: Combined model	0.52	0.76	0.30	0.70

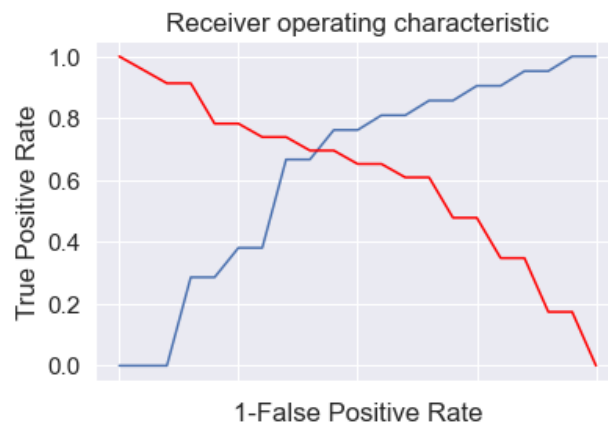


Figure 24 The optimal threshold point for the combined model of logistic regression

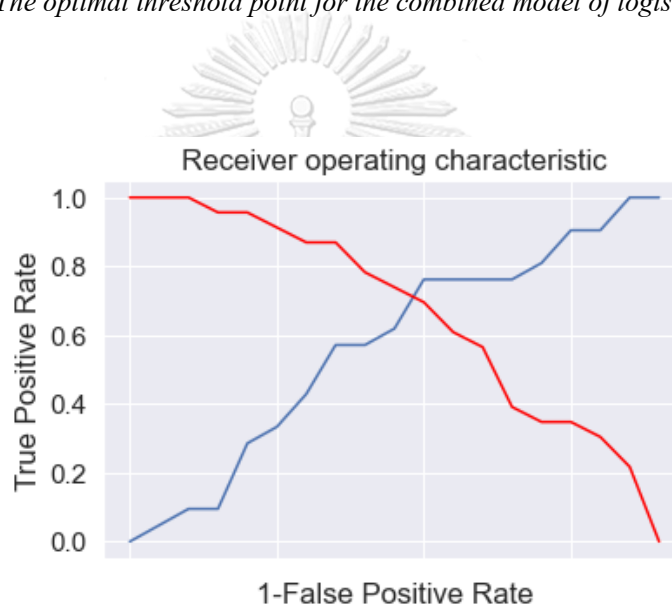


Figure 25 The optimal threshold point for the combined model of random forest

For the combined model in logistic regression, the optimal cutoff point was 0.61 (Table 11), so anything above this can be labeled as 1 else 0. Figure 24 displayed the output that where TPR is crossing 1-FPR, TPR was 67% and FPR was 30%. In the combined model in random forest, the optimal cutoff point was 0.52, where the TPR crossed the 1-FPR and the TPR was 76% and the FPR was 30%, as shown in Figure 25.

The features selected in the combined logistic regression model consisted of three clinical features: bilateral neck metastasis, pretreatment TSH level, and age; one dosimetry feature: TR V_{40} ; and 26 radiomics signatures (Table 12). The radiomics feature's top three high positive coefficients were log-sigma-1-0-mm-3D_ngtdm_contrast, log-sigma-1-0 mm-3D_glszm_LowGrayLevelZoneEmphasis, and log-sigma-2-0-mm-3D_ngtdm_coarseness, while

the top three high negative coefficients were wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis, wavelet-LLH_glrIm_ShortRun LowGrayLevelEmphasis, and wavelet-LLL_firstorder_Minimum. For random forest, the feature selected comes from two clinical and dose variables: pre-treatment TSH level and mean dose to thyroid (TR mean), respectively. Thirty-four radiomics signatures were selected, and the radiomics features of high importance were wavelet-HHL_glszm_SmallAreaEmphasis, log-sigma-2-0-mm-3D_firstorder_TotalEnergy, and log-sigma-3-0-mm-3D_glcM_Idn as shown in Figures 27 and Table 13.

Radiomics feature values in right and left-lobe thyroid glands

The results showed that more than 90% of radiomics features in the right and left lobes had no difference in the mean value of radiomics features (p -value > 0.05). For those features which were selected in the final combined models, more than 95% of the radiomics features had no difference in terms of average feature value in each lobe of the thyroid gland, as displayed in the Appendix Table 1 and Appendix Figure 1. This implied that there were no heterogeneities of radiation-induced thyroid effects on the different sides because the thyroid is a parallel organ that contains a number of thyroid follicles in both lobes that are independent functional subunits. All subunits must be disabled to cause organ failure.

Testing the model robustness for variation in segmentation

The results of the prediction of the final model from multiple segmentation by simulated variation in radiomics feature values found that the combined model in logistic regression had an average AUC of 0.81 ± 0.06 and the combined model in random forest had an AUC of 0.82 ± 0.04 .

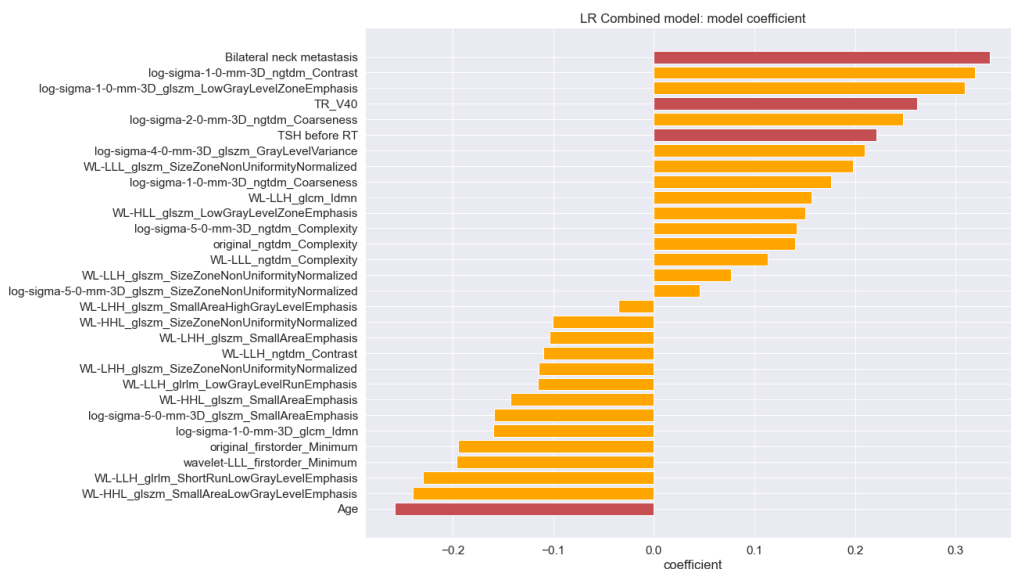


Figure 26 Feature selected in the combined model from logistic regression (The red bars represent the clinical and dose selected features)

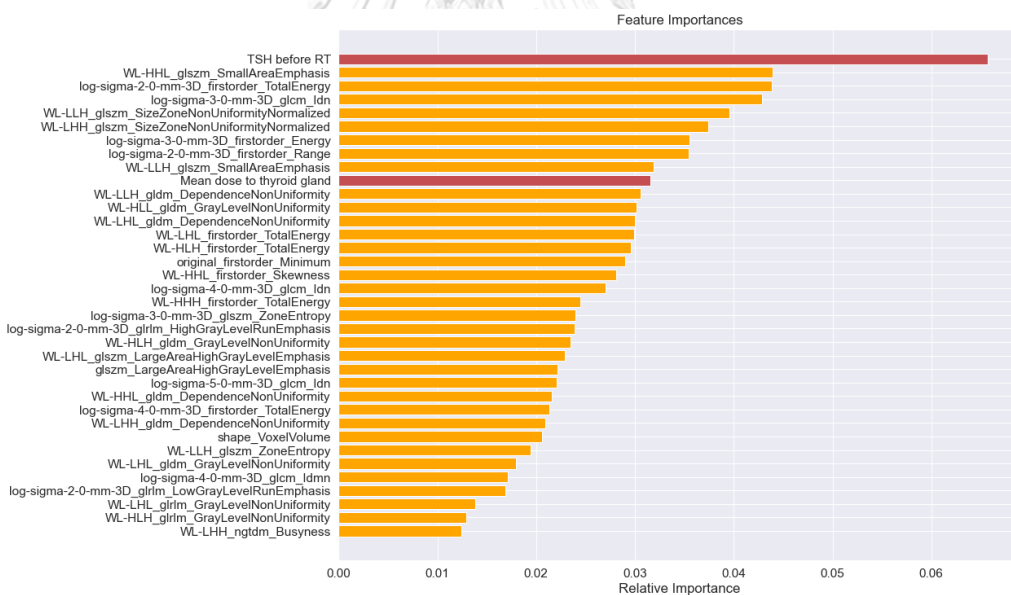


Figure 27 Feature selected in the combined model from random forest (The red bars represent the clinical and dose selected features)

Table 12 Radiomics signatures of the combined model in logistic regression

Feature name	Feature class
original_firstorder_Minimum	First order statistic
original_ngtdm_Complexity	Texture-based
wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis	Filter-based/ Texture-based
wavelet-LLH_glrlm_ShortRunLowGrayLevelEmphasis	Filter-based/ Texture-based
wavelet-LLL_firstorder_Minimum	Filter-based/ Texture-based
wavelet-HHL_glszm_SmallAreaEmphasis	Filter-based/ Texture-based
wavelet-LLH_glrlm_LowGrayLevelRunEmphasis	Filter-based/ Texture-based
wavelet-LHH_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
wavelet-LLH_ngtdm_Contrast	Filter-based/ Texture-based
wavelet-LHH_glszm_SmallAreaEmphasis	Filter-based/ Texture-based
wavelet-HHL_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
wavelet-LHH_glszm_SmallAreaHighGrayLevelEmphasis	Filter-based/ Texture-based
wavelet-LLH_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
wavelet-LLL_ngtdm_Complexity	Filter-based/ Texture-based
wavelet-HLL_glszm_LowGrayLevelZoneEmphasis	Filter-based/ Texture-based
wavelet-LLH_glcm_Idmn	Filter-based/ Texture-based
wavelet-LLL_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
log-sigma-1-0-mm-3D_glcm_Idmn	Filter-based/ Texture-based
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	Filter-based/ Texture-based
log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based

Feature name	Feature class
log-sigma-5-0-mm-3D_ngtdm_Complexity	Filter-based/ Texture-based
log-sigma-1-0-mm-3D_ngtdm_Coarseness	Filter-based/ Texture-based
log-sigma-4-0-mm-3D_glszm_GrayLevelVariance	Filter-based/ Texture-based
log-sigma-2-0-mm-3D_ngtdm_Coarseness	Filter-based/ Texture-based
log-sigma-1-0-mm- 3D_glszm_LowGrayLevelZoneEmphasis	Filter-based/ Texture-based
log-sigma-1-0-mm-3D_ngtdm_Contrast	Filter-based/ Texture-based

* Feature descriptions are shown in the appendix.

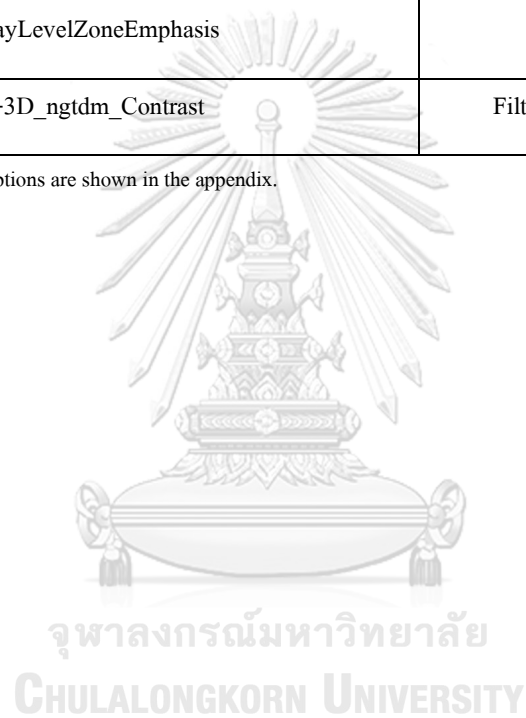


Table 13 Radiomics signatures of the combined model in random forest

Feature name	
original_shape_VoxelVolume	Shape-based
original_firstorder_Minimum	First order statistic
original_glszm_LargeAreaHighGrayLevelEmphasis	Texture-based
wavelet-LLH_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
wavelet-LLH_glszm_SmallAreaEmphasis	Filter-based/ Texture-based
wavelet-LLH_glszm_ZoneEntropy	Filter-based/ Texture-based
wavelet-LLH_gldm_DependenceNonUniformity	Filter-based/ Texture-based
wavelet-LHL_firstorder_TotalEnergy	Filter-based
wavelet-LHL_glrIm_GrayLevelNonUniformity	Filter-based/ Texture-based
wavelet-LHL_glszm_LargeAreaHighGrayLevelEmphasis	Filter-based/ Texture-based
wavelet-LHL_gldm_DependenceNonUniformity	Filter-based/ Texture-based
wavelet-LHL_gldm_GrayLevelNonUniformity	Filter-based/ Texture-based
wavelet-LHH_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based
wavelet-LHH_gldm_DependenceNonUniformity	Filter-based/ Texture-based
wavelet-LHH_ngtdm_Busyness	Filter-based/ Texture-based
wavelet-HLL_gldm_GrayLevelNonUniformity	Filter-based/ Texture-based
wavelet-HLH_firstorder_TotalEnergy	Filter-based

Feature name	
wavelet-HLH_glrIm_GrayLevelNonUniformity	Filter-based/ Texture-based
wavelet-HLH_gldm_GrayLevelNonUniformity	Filter-based/ Texture-based
wavelet-HHL_firstorder_Skewness	Filter-based
wavelet-HHL_glszm_SmallAreaEmphasis	Filter-based/ Texture-based
wavelet-HHL_gldm_DependenceNonUniformity	Filter-based/ Texture-based
wavelet-HHH_firstorder_TotalEnergy	Filter-based
log-sigma-2-0-mm-3D_firstorder_Range	Filter-based
log-sigma-2-0-mm-3D_firstorder_TotalEnergy	Filter-based
log-sigma-2-0-mm-3D_glrIm_HighGrayLevelRunEmphasis	Filter-based/ Texture-based
log-sigma-2-0-mm-3D_glrIm_LowGrayLevelRunEmphasis	Filter-based/ Texture-based
log-sigma-3-0-mm-3D_firstorder_Energy	Filter-based
log-sigma-3-0-mm-3D_glcm_Idn	Filter-based/ Texture-based
log-sigma-3-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	Filter-based/ Texture-based

* Feature descriptions are shown in the appendix.

CHAPTER VI: DISCUSSION

The main findings of the investigation are reviewed in this chapter, including the study limitations and study direction.

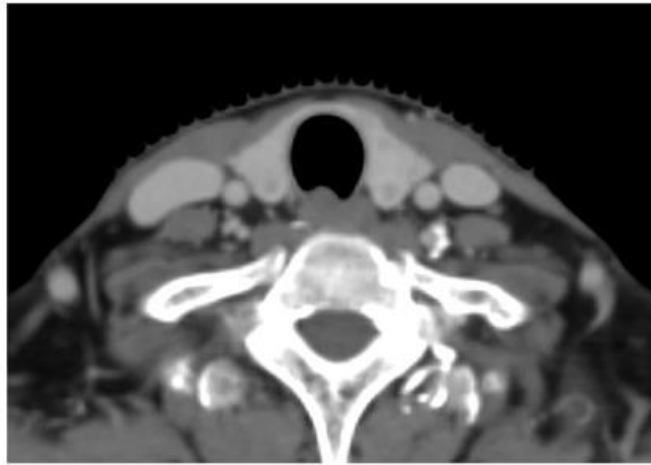
This research aimed to develop and compare three models - a conventional model, a radiomics model, and a combination model - using data from 220 patients with NPC to predict RIH. The results showed that the combination model outperformed the conventional model in logistic regression, random forest, and gradient boosting classifier. The combined model in logistic regression and random forest using the recursive elimination technique showed the best of the model's performance from the combination of data types, while the combined model in the gradient-boosting classifier had lower performance. Hence, the two combined models from logistic regression and random forest were chosen to be further explored.

For the model performance in the test dataset, the combined model had satisfactorily high AUC, accuracy, precision, and F1-scores in both logistic regression and random forest. The sensitivity in logistic regression was high but specificity was quite low. In random forest, the combined model had high sensitivity and specificity. In this clinical application, high sensitivity might be preferred over specificity because the task is to detect the risk of side effects after treatment with a new technique or another modality. In the case of a false-positive prediction, the patient still receives the benefit because they would be treated for the target tumors with a different technique or different modality. For model prediction from variation of delineation, the results showed that the final model did not depend on multiple observers' segmentation, with the coefficient of variation ranging from 8.7% to 7.4% and 7.4% to 4.9% in logistic regression and random forest, respectively.

In the final combined logistic regression model, the important clinical and dose factors were bilateral neck metastasis, pretreatment TSH level, age, and TR V40, while in the random forest, pretreatment TSH level and TR mean were significant features for prediction. These findings were consistent with several other studies [4, 29, 35, 36]. For clinical factors, sex, age, clinical stage, and TSH value were reported as relevant factors in RIH by Zhou L et al. In this study, younger age, positive nodes, and high pretreatment TSH levels were identified as having a higher risk of developing RIH. For dosimetry information, TR V40 and TR mean were significant

features for prediction, which was in agreement with the study by Zhai R et al [29]. They suggested that TR V40 was highly predictive for hypothyroidism occurrence after treatment in patients with NPC. Chow et al [37]. also suggested TR mean as one of the most often suggested dosimetric predictors.

The combined model used in this study incorporated radiomic features, with the majority of these features coming from texture classes and filtered classes, such as log-sigma_ngtmdm_contrast, log-sigma_ngtmdm_coarseness, and wavelet-HHL_glszm_SmallAreaEmphasis. The texture of an image reflects the spatial distribution of intensity levels in a region of interest and can be characterized as fine, coarse, grainy, or smooth, which can help to identify the characteristics of organs or malignancies. Filtered images enhanced the border or information that would not have been visible in a traditional image. Configuration and thyroid texture were interesting points to study in relation to the side effects of radiation injury. Ishibashi N et al [38]. reported that decreased thyroid gland CT density and increased TSH levels before and after radatiotherapy result in hypothyroidism, suggesting that a low thyroid CT intensity before RT might be relevant to a higher risk of RIH. Since it was difficult to distinguish the textures of the thyroid in a typical image by naked eyes, radiomics was used to extract hidden characteristics within the images. Figure 28 shows a thyroid gland CT image in patients with RIH and without RIH after treatment, each with a different radiomics feature value that is not detectable by the human eyes. NGTDM (Neighboring Gray Tone Difference Matrix) complexity is a texture-based class calculated from the variance in gray values between a gray value and its neighbors. The high complexity value means there are several quick fluctuations in the gray level intensity, and the image is not homogeneous. Wavelet-HHL glszm small area emphasis is a filter-based and texture-based class that indicates how small areas are distributed, with a larger number indicating more small areas and smoother textures [21].



Ngtdm Complexity = 41.036
Wavelet-HHL glszm SmallAreaEmphasis = 0.228597

(a)



Ngtdm Complexity = 28.232
Wavelet-HHL glszm SmallAreaEmphasis = 0.298126

(b)

Figure 28 CT image of the thyroid gland in patients who developed RIH (a) and those who did not develop RIH (b) after treatment with different radiomics feature values.

Medical images are one of the most important tools for the success of cancer treatment, as they provide information on anatomy, tumor location, and organs at risk, including baseline organ morphology which are crucial in individualizing patient dose during treatment planning. The application of radiomics for radiation complication prediction has gained attentions recently, as it has been shown to enhance model performance as compared to clinical data [6, 7]. Smyczynska U et al. [30] reported that radiomics analysis did not significantly improve the predictive model for RIH, which contradicts to our findings. One of the possible reasons could be the difference in cancer type. Their predictive model was for oropharyngeal cancer, while we built the model for nasopharyngeal cancer. These two cancer types are well-known for having different etiologies and characteristics, which may have influenced the radiomics and clinical features.

In clinical practice, the probability of radiation-induced hypothyroidism should be considered before planning to reduce the volume of the thyroid gland exposed, notify patients of possible side effects after treatment, or change the treatment modality to proton therapy to reduce the risk of hypothyroidism. Although the mechanism of RIH remains unclear, ionizing radiation can cause damage to thyroid gland, resulting in alterations in the morphology, vessel structure, and immune response. The results of our study demonstrated that the combination of CT imaging with clinical and dose information can significantly improve the performance of prediction models for RIH. We argue that pretreatment thyroid images contain valuable information that can be used to predict the risk of hypothyroidism.

Limitations

This study had certain limitations. First, the retrospective design of the study did not allow for control of CT acquisition protocols, which could have affected radiomics features. Moreover, this study required a long follow-up period, which resulted in a relatively small sample size.

Future directions

Collecting data in multiple centers to increase the sample size and test the model's generalizations in a different population should be studied. Another medical image, such as CT

non-contrast images, might affect radiomics features, which affect prediction performances, and should be explored for better results. In addition, the ensemble model and deep learning network might be study further to increase the prediction performance in this task.



CHAPTER VII: CONCLUSION

In this work, we developed a model to predict hypothyroidism after radiation treatment in nasopharyngeal cancers patients using pre-treatment CT images. For effective prediction, this research used clinical and dosimetry information from treatment planning combined with CT images before treatment.

Firstly, images were acquired from CT simulation before treatment and then manually segmented to cover the thyroid gland area. Second, CT images were extracted into numerical features, which consisted of 1,288 radiomics features divided into 4 classes: shape-based, first-order statistic, texture-based, and filtered-based. For the feature filtration step, the robust features were considered using ICC criteria from variation in the segmentation test. Subsequently, in the model development step, we split 80% of the data for model training and 20% for the test set. Three machine learning techniques were chosen to build the predictive model: logistic regression, random forest, and gradient boosting classifier. The hyperparameters of the model were tuned to get the best performances in the validation set and then performed in the test set. The results were compared between the dose model, clinical model, clinical and dose, radiomics model, radiomics and dose model, radiomics and clinical model, and combined model.

In summary, this study demonstrated that incorporating radiomics with clinical and dosage information had the highest performance and could significantly improve RIH prediction performance in NPC patients when compared to the conventional methods. We identified the significant clinical and dosimetry predictors as TSH level before treatment, age, positive nodes, percentage of thyroid volume that has received at least 40 Gy, and mean dose to thyroid. For radiomics signatures in RIH prediction, there were 26 variables in logistic regression and 30 variables in random forest, which came from first-order statistic classes, texture-based classes, and filtered-base classes. These findings could potentially be used in pre- treatment planning to optimize dose constraints on the thyroid gland and reduce the risk of hypothyroidism.

Appendix

Table 14 The difference of left and right in the thyroid gland in radiomics feature value

Radiomics features	Number of radiomics features: mean of feature value no difference in the right and left sides/total features (p-value > 0.05)	%
bw 0.05	1166/1288	90.53 %
bw 0.1	1174/1288	91.15 %
bw 0.15	1179/1288	91.54 %
bw 0.2	1183/1288	91.84 %
LR: combined model	25/26	96.15 %
RF: combined model	33/34	97.05 %

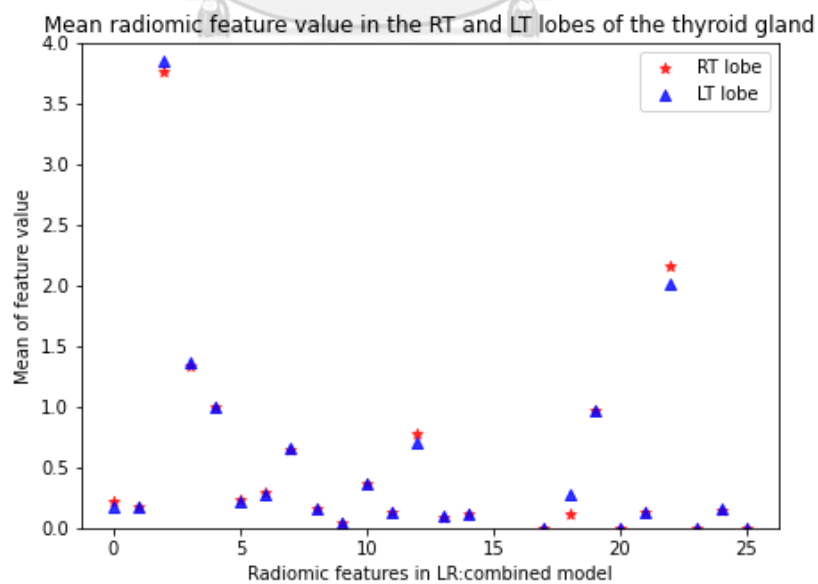


Figure 29 Mean of radiomics features in right and left lobed thyroid in logistic regression combined model

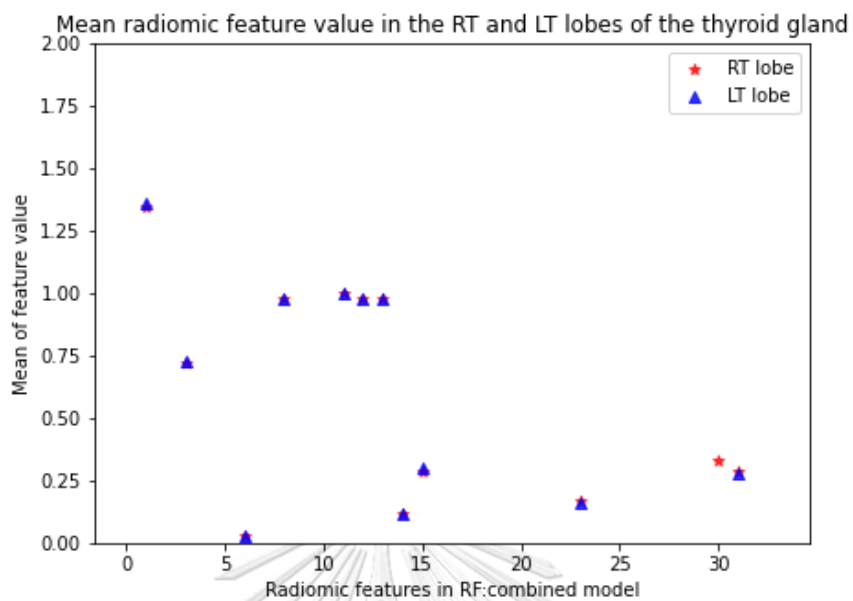


Figure 30 Mean of radiomics features in right and left lobed thyroid in random forest combined model

Table 15 Feature descriptions of radiomics signatures in combined logistic regression model

Feature name	Feature descriptions
original_firstorder_Minimum	Minimum value of voxel intensities in ROI
original_ngtdm_Complexity	Mean difference of voxel intensities among the surrounding area and the center voxel from ngtdm (Neighbouring Gray Tone Difference Matrix)
wavelet-HHL_glszm_SmallAreaLowGrayLevelEmphasis	The percentage of the combined distribution of zones with smaller sizes and lower gray-level values in the image from glszm (Gray Level Size Zone Matrix) after wavelet-HHL transformed
wavelet-LLH_glrml_ShortRunLowGrayLevelEmphasis	The joint distribution of low intensities values and shorter run lengths from glrlm (Gray Level Run Length Matrix) after wavelet-LLH transformed
wavelet-LLL_firstorder_Minimum	Minimum value of voxel intensities in ROI after wavelet-LLL transformed
wavelet-HHL_glszm_SmallAreaEmphasis	How small size zones are distributed from glszm (Gray Level Size Zone Matrix) after wavelet-HHL transformed
wavelet-LLH_glrml_LowGrayLevelRunEmphasis	How low gray-level values are distributed from glrlm (Gray Level Run Length Matrix) after wavelet-LLH transformed
wavelet-LHH_glszm_SizeZoneNonUniformityNorm	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after wavelet-

Feature name	Feature descriptions
alized	LHH transformed
wavelet-LLH_ngtdm_Contrast	The alteration in spatial intensity from ngtdm (Neighbouring Gray Tone Difference Matrix) after wavelet-LLH transformed
wavelet-LHH_glszm_SmallAreaEmphasis	How small size zones are distributed glszm (Gray Level Size Zone Matrix) after from wavelet-LHH transformed
wavelet-HHL_glszm_SizeZoneNonUniformityNormalized	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after wavelet-HHL transformed
wavelet-LHH_glszm_SmallAreaHighGrayLevelEmphasis	The proportion of the joint distribution of smaller size zones with higher gray-level intensities from wavelet-LHH_glszm (Gray Level Size Zone Matrix)
wavelet-LLH_glszm_SizeZoneNonUniformityNormalized	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after wavelet-LLH transformed
wavelet-LLL_ngtdm_Complexity	Mean difference of voxel intensities among the surrounding area and the center voxel from ngtdm (Neighbouring Gray Tone Difference Matrix) after wavelet-LLL transformed
wavelet-HLL_glszm_LowGrayLevelZoneEmphasis	Lower gray-level size zones distributions from glszm (Gray Level Size Zone Matrix) after wavelet-HLL

Feature name	Feature descriptions
	transformed
wavelet-LLH_glcM_Idmn	Calculate the local homogeneity of an image from glcM (Gray Level Co-occurrence Matrix) after wavelet-LLH transformed
wavelet-LLL_glszm_SizeZoneNonUniformityNormalized	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after wavelet-LLL transformed
log-sigma-1-0-mm-3D_glcM_Idmn	Calculated the local homogeneity of an image after log transformed
log-sigma-5-0-mm-3D_glszm_SmallAreaEmphasis	How small size zones are distributed from glszm (Gray Level Size Zone Matrix) after log transformed
log-sigma-5-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after log transformed
log-sigma-5-0-mm-3D_ngtdm_Complexity	Mean difference of voxel intensities among the surrounding area and the center voxel from ngtdm (Neighbouring Gray Tone Difference Matrix) after transformed
log-sigma-1-0-mm-3D_ngtdm_Coarseness	Calculate of mean difference among the center voxel and its surrounding areas from ngtdm (Neighbouring Gray Tone Difference Matrix) after log transformed
log-sigma-4-0-mm-	Measure variance in gray level intensities from glszm

Feature name	Feature descriptions
3D_glszm_GrayLevelVariance	(Gray Level Size Zone Matrix) after log transformed
log-sigma-2-0-mm-3D_ngtdm_Coarseness	Calculate of mean difference among the center voxel and its surrounding areas from ngtdm (Neighbouring Gray Tone Difference Matrix) after log transformed
log-sigma-1-0-mm-3D_glszm_LowGrayLevelZoneEmphasis	How lower gray-level size zone are distributed from glszm (Gray Level Size Zone Matrix) after log transformed
log-sigma-1-0-mm-3D_ngtdm_Contrast	The alteration in spatial intensity from ngtdm (Neighbouring Gray Tone Difference Matrix) after log transformed

Table 16 Feature descriptions of radiomic signatures in combined random forest model

Feature name	Feature descriptions
original_shape_VoxelVolume	Volume of the region of interest
original_firstorder_Minimum	Minimum value of voxel intensities in ROI
original_glszm_LargeAreaHighGrayLevelEmphasis	The percentage of joint distribution between larger size zones and higher gray-level values from glszm (Gray Level Size Zone Matrix)
wavelet-LLH_glszm_SizeZoneNonUniformityNormalized	The variability of size zone volumes in the ROI from glszm (Gray Level Size Zone Matrix) after wavelet-LLH transformed
wavelet-LLH_glszm_SmallAreaEmphasis	How small size zones are distributed from glszm (Gray Level Size Zone Matrix) after wavelet-LLH transformed
wavelet-LLH_glszm_ZoneEntropy	Randomness in the distribution of hues and zone sizes from glszm (Gray Level Size Zone Matrix) after wavelet-LLH transformed
wavelet-LLH_gldm_DependenceNonUniformity	The similarity in the image from gldm (Gray Level Dependence Matrix) after wavelet-LHL transformed
wavelet-LHL_firstorder_TotalEnergy	Size of an image's voxel values that scaled by the volume of the voxel in cubic after wavelet-LHL transformed
wavelet-LHL_glrIm_GrayLevelNonUniformity	Similarity of Gray-level intensity values from

Feature name	Feature descriptions
	glrlm (Gray Level Run Length Matrix) after wavelet-LHL transformed
wavelet-LHL_glszm_LargeAreaHighGrayLevelEmphasis	The proportion of the joint distribution of higher intensity values and larger size zones from glszm (Gray Level Size Zone Matrix) after wavelet-LHL transformed
wavelet-LHL_gldm_DependenceNonUniformity	Similarity of dependence in the image from gldm (Gray Level Dependence Matrix) after wavelet-LHL transformed
wavelet-LHL_gldm_GrayLevelNonUniformity	Similarity of gray-level intensity values in the image from gldm (Gray Level Dependence Matrix) after wavelet-LHL transformed
wavelet-LHH_glszm_SizeZoneNonUniformityNormalized	The alteration of size zone volumes in the image from glszm (Gray Level Size Zone Matrix) after wavelet-LHH transformed
wavelet-LHH_gldm_DependenceNonUniformity	Calculate the similarity of dependence in the image from gldm (Gray Level Dependence Matrix) after wavelet-LHH transformed
wavelet-LHH_ngtdm_Busyness	Calculate the alteration of pixel to its neighbour from ngtdm (Neighbouring Gray Tone Difference Matrix) after wavelet-LHH transformed

Feature name	Feature descriptions
wavelet-HLL_gldm_GrayLevelNonUniformity	Similarity of gray-level intensity values in the image from gldm (Gray Level Dependence Matrix) after wavelet-HLL transformed
wavelet-HLH_firstorder_TotalEnergy	Size of an image's voxel values that scaled by the volume of the voxel in cubic after wavelet-HLH transformed
wavelet-HLH_glrmlm_GrayLevelNonUniformity	Similarity of gray-level intensity values in the image from glrmlm (Gray Level Run Length Matrix) after wavelet-HLH transformed
wavelet-HLH_gldm_GrayLevelNonUniformity	Similarity of gray-level intensity values in the image from gldm (Gray Level Dependence Matrix) after wavelet-HLH transformed
wavelet-HHL_firstorder_Skewness	The asymmetry of the range of values surrounding the mean after wavelet-HHL transformed
wavelet-HHL_glszm_SmallAreaEmphasis	Calculate the distribution of small size zones from glszm (Gray Level Size Zone Matrix) after wavelet-HHL transformed
wavelet-HHL_gldm_DependenceNonUniformity	Calculate the similarity of dependence in the image from gldm (Gray Level Dependence Matrix) after wavelet-HHL transformed
wavelet-HHH_firstorder_TotalEnergy	Size of an image's voxel values that scaled by

Feature name	Feature descriptions
	the volume of the voxel in cubic after wavelet-HHH transformed
log-sigma-2-0-mm-3D_firstorder_Range	Grayscale values in the ROI after log transformed
log-sigma-2-0-mm-3D_firstorder_TotalEnergy	Size of an image's voxel values that scaled by the volume of the voxel in cubic after log transformed
log-sigma-2-0-mm-3D_glrml_HighGrayLevelRunEmphasis	Determine how the higher gray-level values are distributed from glrlm (Gray Level Run Length Matrix) after log transformed
log-sigma-3-0-mm-3D_firstorder_Energy	Calculate the size of an image's voxel values after log transformed
log-sigma-3-0-mm-3D_glcm_Idn	Calculate the local homogeneity of an image after log transformed
log-sigma-3-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	The alteration of size zone volumes in the image from glszm (Gray Level Size Zone Matrix) after log transformed
log-sigma-2-0-mm-3D_glrml_LongRunHighGrayLevelEmphasis	Calculate the joint distribution of long run lengths with higher gray-level values from glrlm (Gray Level Run Length Matrix) after log transformed

REFERENCES

1. Bai R, Sun J, Xu Y, Sun Z, Zhao X. Incidence and mortality trends of nasopharynx cancer from 1990 to 2019 in China: an age-period-cohort analysis. *BMC Public Health*. 2022;22(1):1351.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2021;71(3):209-49.
3. Kazemi E, Zayeri F, Baghestani A, Bakhshandeh M, Hafizi M. Radiation-induced complication after radiotherapy in patients with head-and-neck cancers. *Clinical Cancer Investigation Journal*. 2019;8(6).
4. Lertbutsayanukul C, Kitpanit S, Prayongrat A, Kannarunimit D, Netsawang B, Chakkabat C. Validation of previously reported predictors for radiation-induced hypothyroidism in nasopharyngeal cancer patients treated with intensity-modulated radiation therapy, a post hoc analysis from a Phase III randomized trial. *Journal of Radiation Research*. 2018;59(4):446-55.
5. Peng L, Mao Y-P, Huang C-L, Guo R, Ma J, Wen W-P, Tang L-L. A new model for predicting hypothyroidism after intensity-modulated radiotherapy for nasopharyngeal carcinoma. *Frontiers in Oncology*. 2020;10:551255.
6. Haider SP, Burtneß B, Yarbrough WG, Payabvash S. Applications of radiomics in precision diagnosis, prognostication and treatment planning of head and neck squamous cell carcinomas. *Cancers of the head & neck*. 2020;5:1-19.
7. Sheikh K, Lee SH, Cheng Z, Lakshminarayanan P, Peng L, Han P, et al. Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands. *Radiation Oncology*. 2019;14(1):1-11.
8. Marur S, Forastiere AA, editors. *Head and neck squamous cell carcinoma: update on epidemiology, diagnosis, and treatment*. Mayo Clinic Proceedings; 2016: Elsevier.
9. Bhat GR, Hyole RG, Li J. *Head and neck cancer: Current challenges and future perspectives*. *Advances in cancer research*. 152: Elsevier; 2021. p. 67-102.
10. Johnson D, Burtneß B, Leemans C, Lui V, Bauman J, Grandis J. Head and neck squamous cell carcinoma. *Nature reviews Disease primers*. 2020; 6 (1): 92. Epub 2020/11/28. <https://doi.org/10.1038/s41572-020-0111-1>.

org/10.1038/s41572-020-00224-3 PMID: 33243986.

11. Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, et al. Head and neck cancers, version 2.2020, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*. 2020;18(7):873-98.
12. Bucci MK, Bevan A, Roach III M. Advances in radiation therapy: conventional to 3D, to IMRT, to 4D, and beyond. *CA: a cancer journal for clinicians*. 2005;55(2):117-34.
13. Program NCICTE. Common Terminology Criteria for Adverse Events:(CTCAE): Cancer Therapy Evaluation Program; 2003.
14. Mornex F, Pavy J, Denekamp J, Bolla M. Scoring system of late effects of radiations on normal tissues: the SOMA-LENT scale. *Cancer radiotherapie: journal de la Societe francaise de radiotherapie oncologique*. 1997;1(6):622-68.
15. Yen PM. Physiological and molecular basis of thyroid hormone action. *Physiological reviews*. 2001;81(3):1097-142.
16. OpenStaxCollege. Anatomy & Physiology [Internet]. Rice University2013. [cited 2022].
17. Clinic C. Thyroid Hormone: Cleveland Clinic; 2022 [
18. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JP, et al. Radiomics in oncology: a practical guide. *Radiographics*. 2021;41(6):1717-32.
19. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *La radiologia medica*. 2021;126(10):1296-311.
20. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*. 2012;48(4):441-6.
21. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer research*. 2017;77(21):e104-e7.
22. Parida P, Bhoi N. Wavelet based transition region extraction for image segmentation. *Future Computing and Informatics Journal*. 2017;2(2):65-78.
23. Minaphinant V. 2018. [cited 2021]. Available from: <https://medium.com/about>.
24. IBM. What is machine learning? : IBM; 2020 [updated November 20, 2020. Learn about the history of machine learning along with important definitions, applications, and concerns within

businesses today]. Available from:

<https://www.ibm.com/topics/machine-learning>.

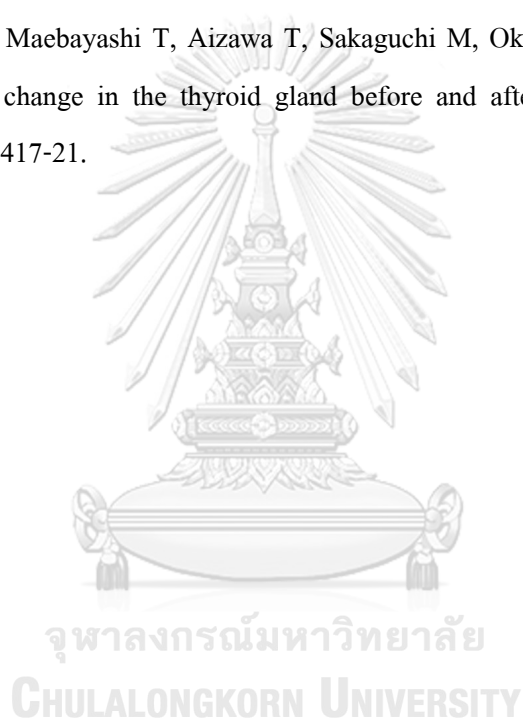
25. Swaminathan S. Logistic Regression: Towards Data Science; 2018 [Logistic regression - detailed overview]. Available from: <https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>
26. Saini A. Gradient Boosting Algorithm: Analytics Vidhya; 2021 [updated August 2, 2023. A Complete Guide for Beginners]. Available from: <https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/>.
27. Zhang Y-m, Gong G-z, Qiu Q-t, Han Y-w, Lu H-m, Yin Y. Radiomics for diagnosis and radiotherapy of nasopharyngeal carcinoma. *Frontiers in Oncology*. 2022;11:767134.
28. Zhang B, Lian Z, Zhong L, Zhang X, Dong Y, Chen Q, et al. Machine-learning based MRI radiomics models for early detection of radiation-induced brain injury in nasopharyngeal carcinoma. *BMC cancer*. 2020;20:1-9.
29. Zhai R, Lyu Y, Ni M, Kong F, Du C, Hu C, Ying H. Predictors of radiation-induced hypothyroidism in nasopharyngeal carcinoma survivors after intensity-modulated radiotherapy. *Radiation Oncology*. 2022;17(1):1-11.
30. Smyczynska U, Grabia S, Nowicka Z, Papis-Ubych A, Bibik R, Latusek T, et al. Prediction of radiation-induced hypothyroidism using radiomic data analysis does not show superiority over standard normal tissue complication models. *Cancers*. 2021;13(21):5584.
31. Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, et al. Early prediction of acute xerostomia during radiation therapy for nasopharyngeal cancer based on delta radiomics from CT images. *Quantitative imaging in medicine and surgery*. 2019;9(7):1288.
32. Vallat. R. Intraclass correlation.
: Sphinx 2022 [
33. Pieper S, Halle M, Kikinis R, editors. 3D Slicer. 2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No 04EX821); 2004: IEEE.
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*. 2011;12:2825-30.
35. Zhou L, Chen J, Tao C-J, Chen M, Yu Z-H, Chen Y-Y. Research progress of radiation-

induced hypothyroidism in head and neck cancer. *Journal of Cancer*. 2021;12(2):451.

36. Murthy V, Narang K, Ghosh-Laskar S, Gupta T, Budrukkar A, Agrawal JP. Hypothyroidism after 3-dimensional conformal radiotherapy and intensity-modulated radiotherapy for head and neck cancers: Prospective data from 2 randomized controlled trials. *Head & neck*. 2014;36(11):1573-80.

37. Chow JC, Cheung K-M, Cheung GT, Tam AH, Lui JC, Lee FK, et al. Dose-volume predictors of post-radiation primary hypothyroidism in head and neck cancer: A systematic review. *Clinical and Translational Radiation Oncology*. 2022;33:83-92.

38. Ishibashi N, Maebayashi T, Aizawa T, Sakaguchi M, Okada M, Matsushita J. Computed tomography density change in the thyroid gland before and after radiation therapy. *Anticancer research*. 2018;38(1):417-21.





จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME Napat Ritlumlert

DATE OF BIRTH 20 February 1984

PLACE OF BIRTH Nakhonratchasima

INSTITUTIONS ATTENDED 1. M.Sc. in Medical Imaging, Faculty of Medicine, Chulalongkorn University, Thailand
2. B.Sc. in Radiological Technology, Faculty of Medical Technology, Mahidol University, Thailand

HOME ADDRESS 23/95 Kanasiri Village, Khukhot, Lumlukka, Pathumthani, Thailand

PUBLICATION Ritlumlert N, Wongwattananard S, Prayongrat A, Oonsiri S, Kitpanit S, Kannarunimit D, Chakkabat C, Lertbutsayanukul C, Sriswasdi S, Rakvongthai Y. Improved prediction of radiation-induced hypothyroidism in nasopharyngeal carcinoma using pre-treatment CT radiomics. *Scientific Reports*. 2023 Oct 14;13(1):17437.

AWARD RECEIVED Ph.D. Scholarship, HRH Princess Chulabhorn College of Medical Science, Chulabhorn Royal Academy

