VIRTUAL PATIENT-SPECIFIC VMAT QUALITY ASSURANCE USING MACHINE LEARNING IN RADIOTHERAPY
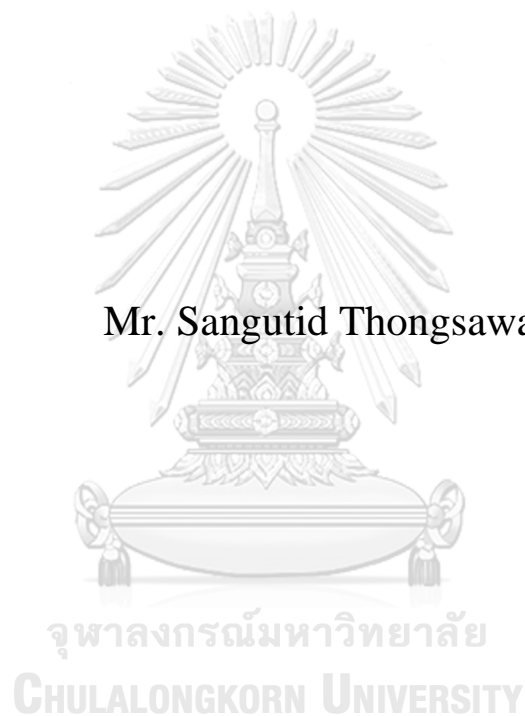
Mr. Sangutid Thongsawad

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Engineering in Nuclear Engineering
Department of Nuclear Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2020
Copyright of Chulalongkorn University

การประกันคุณภาพแผนการรักษาเสมือนของเทคนิคการฉายรังสีปรับความเข้มเชิงปริมาตรเฉพาะ
ผู้ป่วยมะเร็งรายบุคคลด้วยเทคนิคการเรียนรู้ของเครื่องในงานรังสีรักษา

นายแสงอุทิศ ทองสวัสดิ์

Thesis Title	VIRTUAL PATIENT-SPECIFIC VMAT QUALITY ASSURANCE USING MACHINE LEARNING IN RADIOTHERAPY
By	Mr. Sangutid Thongsawad
Field of Study	Nuclear Engineering
Thesis Advisor	Associate Professor Somyot Srisatit
Thesis Co Advisor	Dr. Todsaporn Fuangrod, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Engineering

.... Dean of the FACULTY OF ENGINEERING

(Professor SUPOT TEACHAVORASINSKUN, Ph.D.)

DISSERTATION COMMITTEE

.... Chairman

(Assistant Professor PHONGPHAETH PENGVANICH, Ph.D.)

.... Thesis Advisor

(Associate Professor Somyot Srisatit)

.... Thesis Co-Advisor

(Dr. Todsaporn Fuangrod, Ph.D.)

.... Examiner

(Associate Professor SUNCHAI NILSUWANKOSIT, Ph.D.)

.... Examiner

(Assistant Professor TAWEAP SANGHANGTHUM, Ph.D.)

.... External Examiner

(Assistant Professor Dr. Yudthaphon Vichianin, Ph.D.)

แสงอุทิศ ทองสวัสดิ์ : การประกันคุณภาพแผนการรักษาเสมือนของเทคนิคการฉายรังสีปรับความ
เข้มเชิงปริมาตรเฉพาะผู้ป่วยมะเร็งรายบุคคลด้วยเทคนิคการเรียนรู้ของเครื่องในงานรังสีรักษา. (
VIRTUAL PATIENT-SPECIFIC VMAT QUALITY
ASSURANCE USING MACHINE LEARNING IN
RADIOTHERAPY) อ.ที่ปรึกษาหลัก : สมยศ ศรีสถิตย์, อ.ที่ปรึกษาร่วม : ดร.ทศพร
เฟื่องรอด

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างการประกันคุณภาพแผนการรักษาแบบเสมือนด้วยเทคนิคการ
เรียนรู้ของเครื่องโดยการทำนายผลการประกันคุณภาพแผนการรักษาจากการเคลื่อนที่มัลติลีฟคอลลิเมเตอร์
และแผนผังความเข้มรังสี การเรียนรู้ของเครื่องถูกนำมาใช้ในการสร้างโมเดลสำหรับทำนายแบบคลาสสิฟ
เคชั่นและแบบรีเกรสชั่น โดยข้อมูลอินพุตต์ของโมเดลประกอบด้วยการเคลื่อนที่มัลติลีฟคอลลิเมเตอร์และ
แผนผังความเข้มรังสีซึ่งเป็นข้อมูลสำคัญ ส่วนข้อมูลเอาพุตต์สำหรับโมเดลคือค่าแกมม่าพาสซึ่งเรทของผลการ
ประกันคุณภาพแผนการรักษาการฉายรังสีแบบปรับความเข้มเชิงปริมาตรด้วยอุปกรณ์รับภาพอิเล็กทรอนิคส์
ค่าความไวและความจำเพาะถูกคำนวณเพื่อใช้ในการหาความถูกต้องของโมเดล ผลการวิจัยพบว่าค่าความไว
สูงสุดถูกพบที่โมเดลแบบรีเกรสชั่นโดยมีความไวร้อยละ 81.25 และค่าความจำเพาะสูงสุดถูกพบที่โมเดล
แบบคลาสสิฟเคชั่นโดยมีความจำเพาะร้อยละ 66.67 ความถูกต้องของโมเดลสามารถเพิ่มได้โดยการเพิ่ม
จำนวนประชากรในการเทรนนิ่งให้มากขึ้น จากผลการวิจัยนี้แสดงให้เห็นถึงความเป็นไปได้ในการนำโมเดล
ทำนายการประกันคุณภาพแผนการรักษาแบบเสมือนมาช่วยในการตัดสินใจกรณีแผนการรักษาที่มีโอกาสไม่
ผ่านเกณฑ์ โดยไม่ต้องทำการประกันคุณภาพจริง มากกว่านั้นการประกันคุณภาพแผนการรักษาแบบเสมือน
แสดงให้เห็นว่ามีศักยภาพที่จะสามารถนำมาใช้ร่วมกับกระบวนการวางแผนการรักษาในการควบคุมค่าความ
ซับซ้อนของแผนการรักษาได้

สาขาวิชา	วิศวกรรมนิวเคลียร์	ลายมือชื่อนิสิต
		....
ปีการศึกษา	2563	ลายมือชื่อ อ.ที่ปรึกษาหลัก
		....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม
		....

5771477421 : MAJOR NUCLEAR ENGINEERING
KEYWORD: Machine Learning, patient-specific VMAT QA, gamma prediction

Sangutid Thongsawad : VIRTUAL PATIENT-SPECIFIC VMAT QUALITY ASSURANCE USING MACHINE LEARNING IN RADIOTHERAPY . Advisor: Assoc. Prof. Somyot Srisatit Co-advisor: Dr. Todsaporn Fuangrod, Ph.D.

The purpose of this study was to develop the virtual patient-specific VMAT QA based on the extracted features from the multileaf collimator (MLC) patterns and fluence map. The machine learning (ML) approach was used to develop the prediction (or regression model) and classification models. The input of models applied Multileaf collimator (MLC) patterns and fluence map as the key features, whereas gamma passing rates (GPR) results from patient-specific VMAT QA of electronic portal imaging devices (EPID) as the label or the response class for testing these models. Sensitivity and specificity scores were calculated for models' accuracy. The highest sensitivity score was observed at the regression model, with 81.25% sensitivity, and the highest specificity score was also observed at the classification model with 66.67% specificity. The models' accuracy can be improved by increasing the population in the training data set. This study indicated the virtual patient-specific VMAT QA (model-based prediction) using ML approach has feasibility for determining the VMAT plan risk that could fail the tolerance without actual QA measurement. Moreover, the virtual patient-specific VMAT QA shows the potential to corporate within the VMAT treatment planning process.

Field of Study:	Nuclear Engineering	Student's Signature ....
Academic Year:	2020	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

A_{l-h}	The average proportion of leaf accelerations from a given range
AAA	Anisotropic analytical algorithm
AAV	Aperture area variability
ANN	Artificial neural network
aSi	Amorphous silicon
AUC	Area under curve
AAPM	the American association of physicists in medicine
ASCII	American standard code for information interchange
CC	Collapsed-cone convolution
CP	Control point
CT	Computed thermography
CAM	Converted aperture metric
CAS	Cross-axis score
CLS	Close leaf score
CNN	Conventional neural network
CTV	Clinical target volume
DL	Deep learning
DR	Variations of the nominal dose rate
DT	Decision trees
DTA	Distance-to-agreement analysis
DVH	Dose volume histogram

EM	Edge metric
EAM	Edge are metric
EBRT	External beam radiation therapy
EPID	Electronic portal imaging device
ESTRO	the European Society for Therapeutic Radiology and Oncology
FN	False negative
FP	False positive
GS	Variation of gantry speed
GPR	Gamma passing rate
GTV	Gross tumor volume
HVL	Half value layer
IMRT	Intensity modulated radiation therapy
IEC	the International Electrotechnical Commission
ISO	International organization for standardization
IAEA	International atomic energy agency
ICRP	International commission on radiological protection
keV	Kilo electron volts
k-NN	k-nearest Neighbors
LR	Logistic regression
LSV	Leaf sequence variability
LINAC	Linear accelerator
LOIC	Leaf offset impact on calculation
LTMCS	Combination of leaf transmission and modulation complexity score

MI	Modulation index
ML	Machine learning
MU	Monitors unit
MRI	Magnetic resonance imaging
MAD	Mean asymmetric distance
MCS	Modulation complexity scores
MeV	Mega electron volts
MFA	Mean field area
MIR	Maximum intensity ratio
MLC	Multileaf collimator
MISPORT	Modulation index for station parameter optimized radiation therapy
NB	Naïve bayes
OBI	On board imager
PA	Plan averaged beam area
PI	Plan averaged beam irregularity
PL	Poisson lasso
PBC	Pencil beam convolution
PET	Positron emission thermography
PTV	Planning target volume
PDIP	Portal dose images prediction
PDPC	Preconfigured portal dosimetry package
PIMV	Plan intensity map variation
PMMA	Polymethylmethacrylate
QA	Quality assurance

RF	Random forest
RL	Reinforcement learning
RT	Radiation therapy
ROC	Receiver operating characteristics
S_{l-h}	The average proportion of leaf speeds from a given range
SAS	Small aperture score
SPC	Statistically process control
SRS	Stereotactic radiosurgery
SRT	Stereotactic radiation therapy
SVM	Support vector machine
SBRT	Stereotactic body radiation therapy
SA	Segment area
SIB	Simultaneous intergraded boost
TN	True negative
TP	True positive
TPS	Treatment planning system
VMAT	Volumetric modulated arc therapy

CHAPTER 1

INTRODUCTION

Modern radiation therapy is developed and clinically used to achieve the radiation therapy goal, such as Volumetric Modulated Arc Therapy (VMAT). VMAT is a beam delivery technique performed by varying dose rate, gantry speed, and Multi-Leaf Collimator (MLC) position during gantry rotation. Nowadays, VMAT has become a general technique to treat cancer patients with benefits. However, the comprehensive Quality Assurance (QA) process, such as a patient-specific QA, needs to be performed before treating the patient. Patient-specific QA is a method to verify the agreement between beam delivery and dose calculation in Treatment Planning Systems (TPS). There are many ways to perform patient-specific QA, such as measurement-based with phantom, Electronic Portal Imaging Device (EPID)-based dosimetry, and log-file analysis. EPID is generally used to verify patient position. However, EPID also can be used for measurement a radiation dose with the appropriate calibration method. This research emphasized to assess the potential of EPID-based dosimetry in term of error detection and model-based prediction for VMAT patient-specific QA.

1.1 Motivation

The motivation for this study can be classified into two parts as follows.

1) Patient-specific Volumetric Modulated Arc Therapy (VMAT) Quality Assurance (QA) has become a critical process to detect the errors before beam delivery to patients. Patient-specific VMAT QA was expected

to detect the systematic error such as a missed calibration of MLC positions, missed calibration of gantry angle, and missed calibration of machine output, and to detect the random error such as the gravitational sagging of MLC during gantry rotation. According to VMAT QA tool characteristics, such as detector resolution, phantom shape, the VMAT QA tool's sensitivity to detect errors can differ. Verifying the patient-specific VMAT QA performance, the sensitivity to detect errors need to be investigated.

2) In VMAT, the beams can be modulated by moving MLCs. The beam's high complexity is related to the high fluctuation of beam intensity used to improve a tumor conform and reduce an organ dose. However, high beam complexity can cause MLC position error, which leads to a decrease in the pass of QA results in patient-specific VMAT QA. Moreover, the practical drawbacks of patient-specific QA have been reported, such as time-consuming measurement, resource-intensive, patient scheduling impact, repeated measurement in case of fail QA. Recently in medical physics, several studies have implemented machine learning to predict and classify patient-specific QA errors. Machine learning demonstrated the feasibility method to use for model-based prediction. The prediction model has shown an advantage in determining the risk of a plan that could fail the tolerance level before they re-measure the QA.

1.2 Objectives

1) To determine the error detection sensitivity of various patient-specific VMAT QA systems; including EPID dosimetry, Delta4 and Log file based on the complex error scenarios; uncertainty, mis-calibration, and worse case scenarios.

2) To develop virtual patient-specific VMAT QA using the machine learning algorithm to predict and classify the predicted outcomes without the actual QA measurements.

1.3 Scope

1.3.1 General scope

TrueBeam and Trilogy linacs (Varian Medical System, Palo Alto, CA, USA) were used to deliver radiation for patient-specific QA. EPID images were acquired to perform patient-specific VMAT QA with integrated mode.

1.3.2 Determination of error detection sensitivity of patient-specific VMAT QA systems

Five head and neck and five prostate VMAT plans were used to perform patient specific VMAT QA with EPIDs, Delta4 (ScandiDos AB, Uppsala, Sweden), log file, and then intentional errors were introduced in each plan to investigate the capability of error detection of different QA tools.

1.3.3 Predicting patient-specific quality assurance results VMAT plan using machine learning approach

The previous 53 of head and neck patient-specific VMAT QA with EPID-based dosimetry were collected to study, recorded as Gamma Passing Rates (GPR). The MLC speed and acceleration and texture analysis were extracted from the plan, which was used as the features. The model dataset was composited into two parts; prediction features (MLC speed and acceleration and texture analysis), response classes (GPR). The

model was trained by using machine learning with classification and regression models.

1.4 Expected outcome

There are two expected outcomes according to the objectives. First, this study aims to establish the QA method performance between Delta4, EPID, and log file approaches to detect complex and realistic errors. For the second expected outcome, the virtual patient-specific VMAT QA tool with the prediction model can help the medical physicists determine the VMAT plan's risk that could fail the tolerance level without QA measurements.
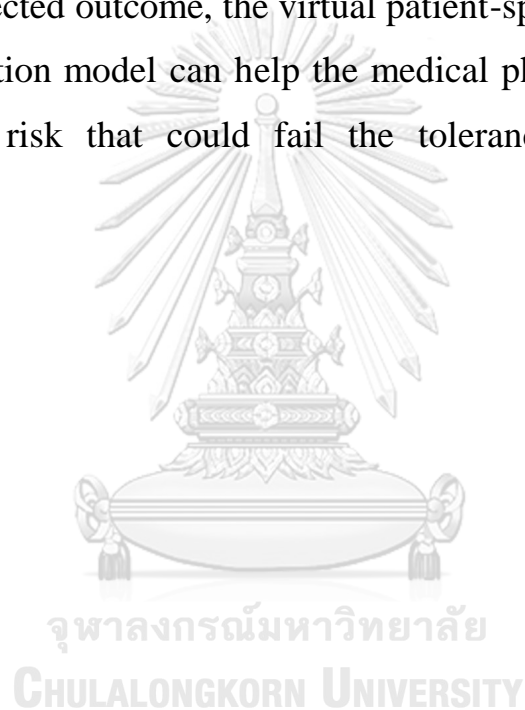
CHAPTER 2

THEORETICAL BACKGROUND

2.1 Radiation Therapy

The cancers can be treated with many methods such as surgery, chemotherapy, hormonal therapy, targeted therapy, and radiation therapy (RT). Cancer patients can be treated with a single method or combined methods, depending on the cancer type and tumor stage. Radiation therapy is one of the commonly used to treat cancer. Although RT destroyed a tumor by breaking the DNA, the normal organ also receives a small dose compared to a tumor. Thereby, the primary goal of RT is to maximize a tumor dose and minimize organ dose. RT can be categorized into two types: External Beam Radiation therapy (EBRT) and Brachytherapy.

2.2 Radiotherapy Process

When cancer patients come to the radiotherapy department, they are determined to treat the oncologist with appropriate RT. The first process is a simulation. In the simulation process, the patient is set up in a position and fixed with an immobilization tool such as a thermoplastic mask, vacuum lock that depends on the cancer site. The immobilization is used to reduce the patient movement during beam delivery and improve patient setup reproducibility between fraction and fraction. After setup and immobilization, the image is acquired for dose calculation using conventional x-ray simulation or computed tomography (CT) simulation that depends on the treatment planning technique used for the patient. The second process is treatment planning, consisting of subprocess as follows; contouring, plan optimization (for Intensity Modulated Radiation Therapy:

IMRT) and dose calculation, and plan evaluation. The treatment planning is performed in Treatment Planning Systems (TPS). In contouring, the oncologist determines the tumor target and normal organs' location by drawing in CT images. In some cases, the other image modality, such as Magnetic Resonance Imaging (MRI) and Positron emission tomography (PET), can help determine tumor staging related to a tumor outline. In plan optimization, the doses are optimized between tumor dose prescription and normal organs dose receiving related to RT's goal using beam modulation. Multileaf collimators (MLCs) position is calculated regarding beam fluence from optimization, and then dose distribution is calculated on CT images. There are many methods to calculate dose distribution such as correction-based, model-based (Anisotropic Analytical Algorithm: AAA, Pencil Beam Convolution: PBC, Collapsed-cone convolution: CC), and Monte Carlo-based. In plan evaluation, the oncologist determines the tumor coverage and normal organ dose with an iso-dose line, color wash displayed, and dose-volume histograms (DVHs), and the appropriated plan is selected to RT. The third process is the patient-specific quality assurances (QA) that especially only for IMRT.

Regarding beam complexity in IMRT, patient-specific QA is established to verify the correct data transfer from TPS to the treatment machine and verify its accuracy. The dose comparison of radiation measurement and TPS are performed to determine the agreement. If the agreement is within the acceptance level, the plan will be approved for treatment. If the agreement is not within the acceptance level, the plan needs to be investigated and re-optimized the new plan. The fourth process is radiation beam delivery. In the radiation beam delivery, the patient is set up and fixed with immobilization, as same as the simulation process. For modern

treatment machines, image guidance such as Electronic Imaging Devices (EPIDs) and On-board Imagers (OBI) can verify patients before treatment. After the patient is shifted to the correct position, radiation will be delivered to the patient as a TPS plan. The end-to-end External Beam Radiation Therapy (EBRT) process can be shown in figure 2.1.
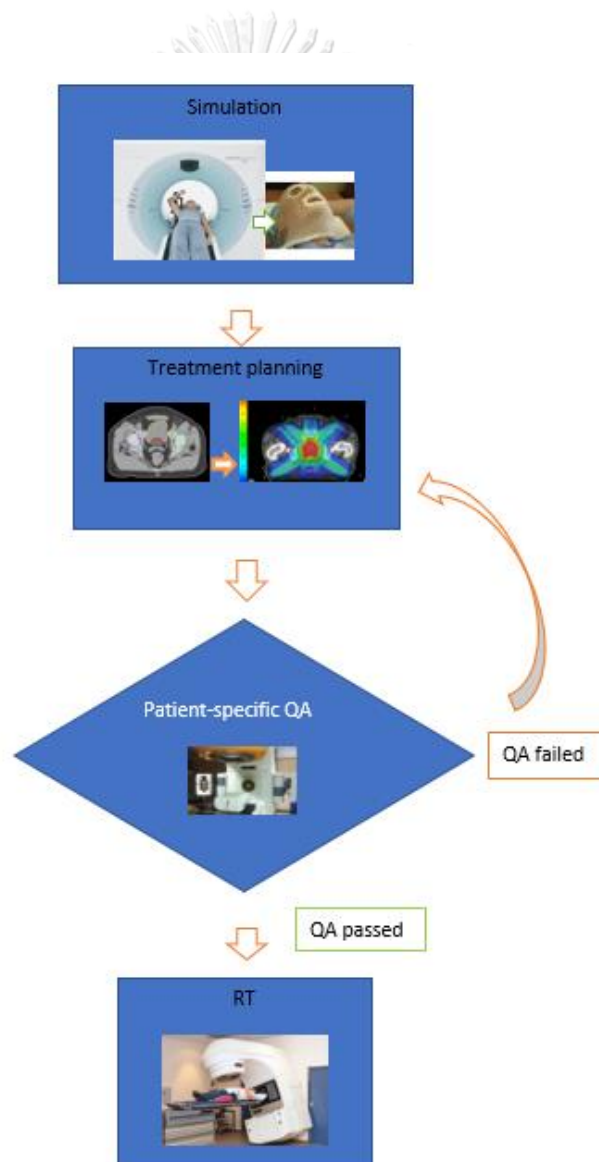
Figure 2.1 Flowchart of External Beam Radiation Therapy (EBRT)

2.3 External Beam Radiation therapy (EBRT) and Brachytherapy

Radiation therapy can be categorized into two types: External Beam Radiation therapy (EBRT) and Brachytherapy.

2.3.1 EBRT

EBRT is one method of RT by delivering radiation outside the patient. Radiation can be used for EBRT, including electromagnetic waves (x-ray, gamma) and charged particles (electron, proton, carbon ion, Helium ion). For RT with an electromagnetic wave, the radiation can be produced using Co-60 radioactive source or Linear accelerator machines (LINACs). Radiation was applied to treat cancer after a few months of x-ray discovered by Wilhelm C. R¨ontgen in 1895. Harold E. John's group conducted the first radiation machine in 1950 with the Co-60 machine. The Co-60 radioactive source produces gamma radiation of 1.25 MeV (1.173MeV, 1.332 MeV). Figure 2.2 shows the decay schematic of Co-60 transformed into Ni-60. The new EBRT passed to LINACs, which were commonly used for cancer treatment. The first LINACs only produced low x-ray energy (kV) and operated with the Van De Graaff generators. The modern LINACs have been developed to generate high x-ray energy (MV).

$^{60}_{27}$Co

5.272 a   0.31 MeV β^- 99.88%

0.12%

1.48 MeV β^- 1.1732 MeV γ

1.3325 MeV γ

$^{60}_{28}$Ni
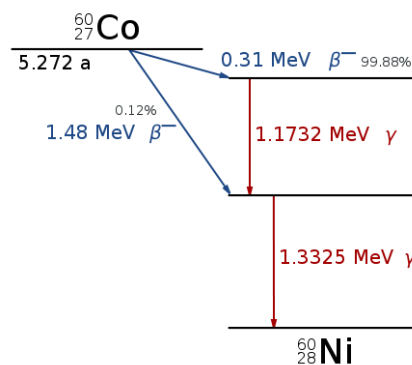
Figure 2.2 Decay schematic of Co-60

The main components of linac can be considered in different parts: the electron gun, accelerating tube, waveguide, microwave source, bending magnet, target, and LINACs head, as shown in Fig.2.3. Firstly, electrons are injected into the accelerating tube by using an electron gun (Fig.2.4). The electrical voltage is applied between the accelerating tube's proximal and distal to accelerate electron (Fig.2.5). The electron speed is magnified using a Radiofrequency (RF) pulse from a microwave source, which is synchronized to an electron bunch. The microwave source can be classified into different types: klystron and magnetron (Fig.2.6). Klystron is one type of microwave source operated with an affectively RF amplifier used in high energy physics and medical linac for high energy (>6 MeV). A magnetron is the other type of microwave source preferred for low energy (4-6 MeV). At the end of the accelerating tube, the electron energy is gained to MeV. For electron treatment, the electron is bent to the isocenter with a bending magnet, and then the scattering foils are used to spread the beam and get a uniform fluence across the treatment field. For photon treatment, the target is hit by an electron to produce the x-ray. The target is generally made of tungsten. There are two types of x-ray generated: characteristic x-ray and bremsstrahlung x-ray. The population of bremsstrahlung x-ray is approximately 80% of all composites. To produce the photon uniform intensity, a flattening filter can be used. The photon field size can be collimated by using the primary and secondary collimator jaws. To improve dose conformity around a tumor, MLCs (Fig.2.7) are also used to shape a tumor, and the other MLCs utilization is to modulate the beam for IMRT.
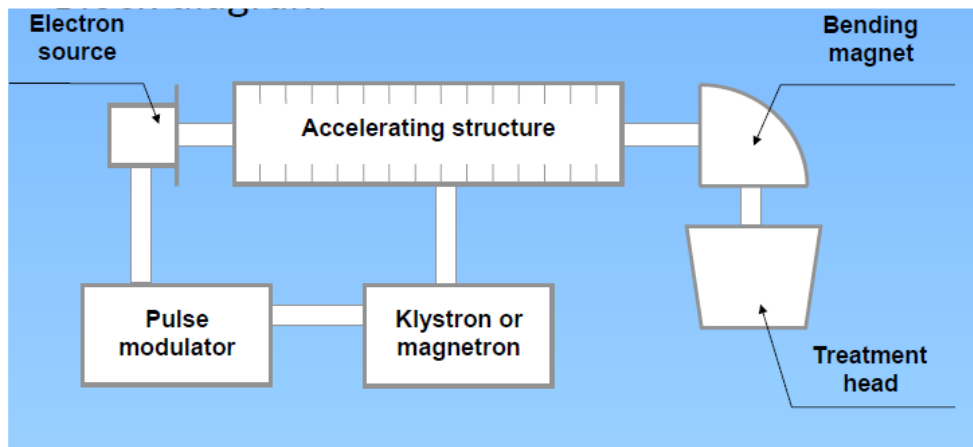
Figure 2.3 The block diagram of medical linacs
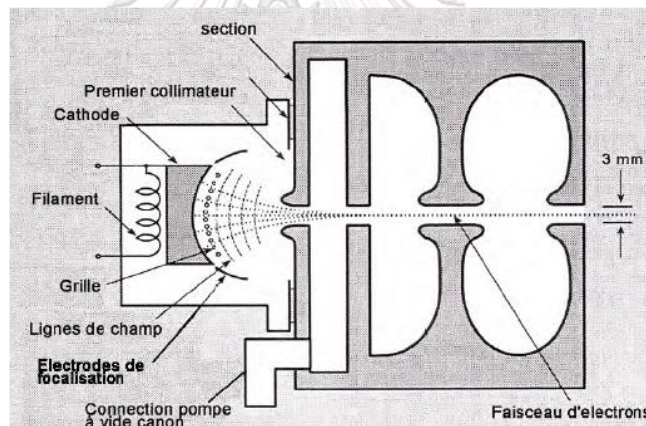
Figure 2.4 The sagittal view of electron gun
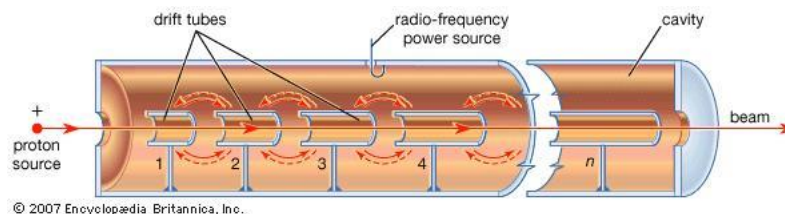
Figure 2.5 The sagittal view of accelerating tube
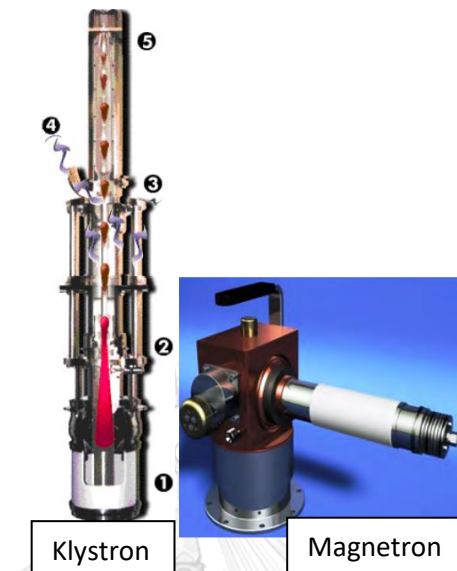
Figure 2.6 Klystron and Magnetron microwave sources
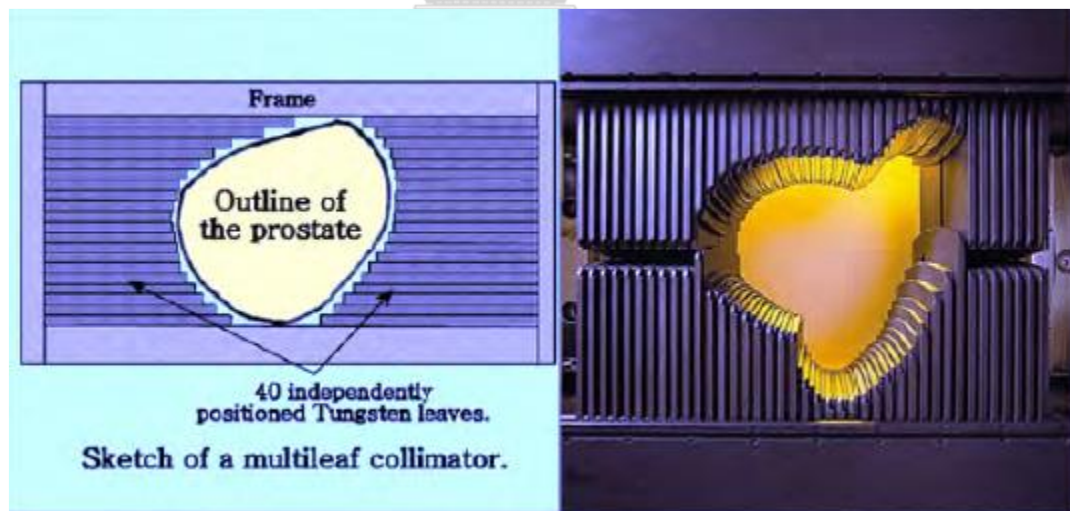
Figure 2.7 The Multi Leaf collimators (MLCs)

2.3.2 Brachytherapy

Brachytherapy is the other RT method by placing a radiation source into tumor volume (intracavitary or interstitial brachytherapy) or tumor volume (surface mold or intraoperative radiotherapy). The type of brachytherapy can be classified in different terms depend on the purpose, such as temporary and permanent (in term of treatment duration), hot loading and afterloading (in term of source loading), low dose rate, medium dose rate, and high dose rate (in term of dose rate). Table 2.1 shows a commonly used isotope for brachytherapy and the characteristics of each isotope.

Table 2.1 Characteristics of each isotope used for brachytherapy

Isotope	Average photon energy (MeV)	Half-life	HVL in lead (mm)	Γ_{AKR} $(\frac{\mu Gy \cdot m^2}{GBq \cdot h})$	Λ $\left(\frac{cGy \cdot h^{-1}}{cGy \cdot cm^2 \cdot h^{-1}}\right)$
Co-60	1.25	5.26 y	11	309	1.11
Cs-137	0.66	30 y	6.5	77.3	1.11
Au-198	0.41	2.7 d	2.5	56.2	1.13
Ir-193	0.38	73.8 d	3	108	1.12
I-125	0.028	60 d	0.02	-	-
Pd-103	0.021	17 d	0.01	-	-

2.4 Modern RT techniques

The collimators are developed to shield organs at the first radiation therapy technique. However, collimators cannot complete shielding regarding the collimator shape limit in rectangular (Fig.2.8). Then,

cerrobend block was used to reduce organ dose. MLCs were developed for this purpose also. Modern radiation therapy techniques have been developed and clinically used to achieve the radiation therapy goal, such as Intensity Modulated Radiation Therapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT). VMAT is a beam delivery technique performed by varying dose rate, gantry speed, and Multi-leaf collimators (MLCs) position during gantry rotation, while statics IMRT is a beam modulated at a fixed gantry angle. The difference between statics IMRT and VMAT are shown in Fig.2.9.

Additionally, VMAT was developed to reduce treatment time by modulating a beam during gantry rotation. Because of the complexity of beam delivery in advanced therapy techniques, beam delivery accuracy needs to be verified before treatment. This process was called "patient-specific VMAT QA."

Figure 2.8 Collimator jaw shielding normal organs
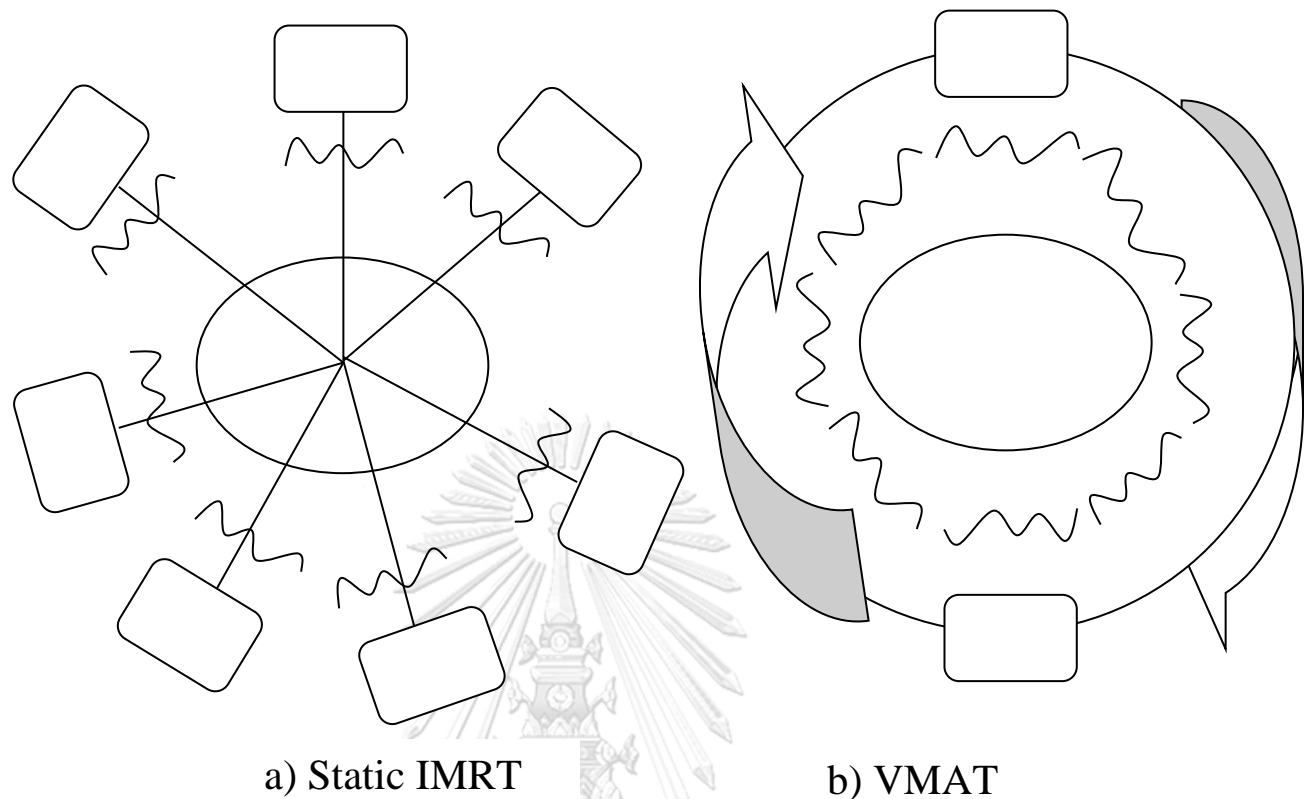
a) Static IMRT   b) VMAT

Figure 2.9 Difference of beam modulation between statics IMRT and VMAT, a) Static IMRT beam modulated at fixed gantry, b) VMAT beam modulated during gantry rotation

2.5 Errors in RT

The Basic Safety Standards [1] defines the radiological accident as a severe unintended event including an operating error, equipment failure, or other mishaps, the consequences of which cannot be ignored from the protection or safety point of view, and which usually leads to potential overexposure or abnormal exposure conditions for the treated patient, staff or the general public. The error in RT can be classified into four primary sources [2]:

1. Human mistakes caused by inattention, misunderstanding, or misjudgment

2. Instrumental mistakes caused by mechanical or electrical failure

3. Random errors due to unknown and/or uncontrolled experimental conditions in the process involved in the planning and delivery of radiation

4. Systematic errors, i.e., biases, in the same set of processes

Coeytaux et al. [3] have reported the radiation overexposure accidents during 1980-2013. They found that the accidence has decreased from 1980-2013, and most of the accidence can be found in radiation therapy and fluoroscopy field with 64%. Figure 2.10 shows a summary of accidence in a different field by Coeytaux and groups.

a) Over exposures

b) Over exposures people
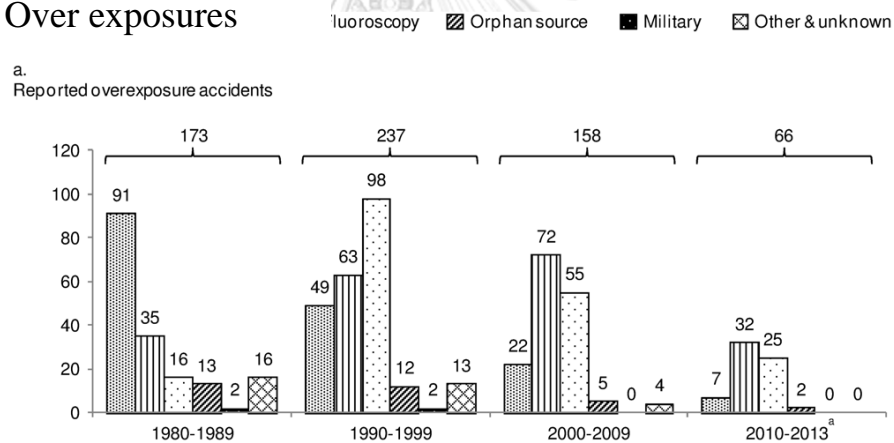
b.
Reported overexposed people
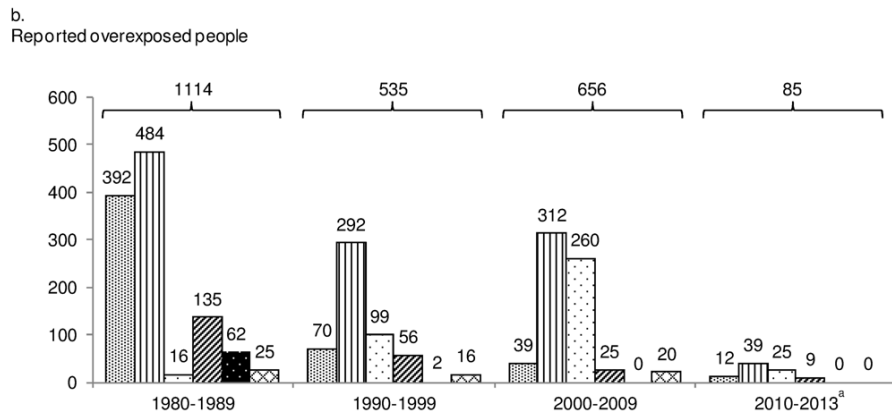
Figure 2.10 The report of radiation accidence during 1980-2013 in term of a) over exposures accidence and b) over exposure people

Several accidences in RT have been reported with different causes. One of RT's severe accidences has been reported at Panama [4] with the wrong dose calculation of Co-60 treatment. The government performed the error investigation, and they found 28 of 56 patients had died in 2005 with 10-110 % overdose. The underexposure of RT for LINACs treatment has been reported in the UK [5]. The cause of the error was an incomplete understanding and testing of a treatment planning system. This accidence can be the cause of tumor recurrence in 492 patients. IAEA and ICRP [5, 6] have reported the accidence in RT around the world, as shown in Table 2.2. They have suggested that the accidence can be prevented by high-quality measurements of the treatment beams and careful calculation of doses to target volumes, supported by an excellent preventive maintenance program for the equipment, i.e., a well-implemented quality assurance program.

Table 2.2 The RT accidence in three-year decades around the world

Type of accident	Country	Year (s)	Patient overdosed or underdosed
Miscalibration of cobalt-60 units	United States	1974-1976	426 overdosed
	Germany	1986-1987	86 overdosed
	United	1988	207 overdosed
	Kingdom	1996	114 overdosed
	Costa Rica		
Hardware/software problem with linear accelerators	Canada and United States	1985-1987	3 overdosed
	Spain	1990	27 overdosed
	Poland	2001	5 overdosed
Low dose rate brachytherapy problems	United Kingdom	1988-1989	14 underdosed
			12 overdosed
		1992	1 overdosed
	Unites States		
Treatment planning errors	United Kingdom	1982-1990	1,045 underdosed
		1987-1988	
	Unites States	2000-2001	33 overdosed
	Panama		28 overdosed

2.6 Quality Assurance of RT

RT's process is sophisticated and involved with different medical physics principles, radiobiology, radiation safety, dosimetry, RT planning, simulation, and interaction of RT with other treatment modalities. Figure

2.11 shows the vital process of QA and Quality Control (QC) that can prevent the error for each step and give high confidence that patients will receive the prescribed treatment correctly. The current paradigm of quality management (QM) in RT emphasizes measuring the functional performance of RT equipment to contain the accuracy within the acceptance level. There are several guidelines provided for QA in RT such as the American Association of Physicists in Medicine (AAPM) in Task Group (TG) 40, 43, 53, 56, 59, 60 and 64[7-13], the American College of Radiology and the American College of Medical Physics in reports on Radiation Therapy Quality Assurance (RTQA) the European Society for Therapeutic Radiology and Oncology (ESTRO) in a report on RTQA[14], the International Electrotechnical Commission (IEC) publications on functional performance of RT equipment; and the International Organization for Standardization (ISO). Since the advanced RT (IMRT, VMAT, Stereotactic Body Radiation Therapy: SBRT, Stereotactic Radiosurgery: SRS, Stereotactic Radiation Therapy: SRT) are introduced to the clinic, the beam complexity also increases. Hence, planning accuracy needs to be investigated. Ishikura [15] has reported the incidents without any known adverse events to patients. He has found 7% of incidents related to the planning stage, 39% related to information transfer, and 19% related to the treatment delivery stage.
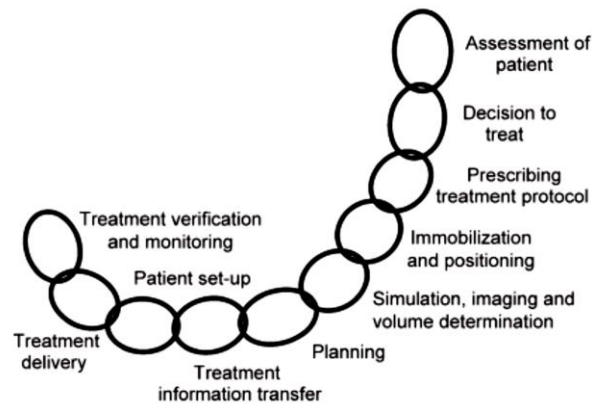
Figure 2.11 Sequential process of planning and delivering radiotherapy to patients

A patient-specific QA process is performed to detect discrepancies between TPS and beam delivery. The primary purpose of patient-specific QA is to verify the correct data transfers. The discrepancy between treatment delivery and planned dose should be within the specified tolerance and the accuracy of dose calculation in TPS. Various methods were used to perform pre-treatment verification, such as measurement-based with phantom (Delta4, ArcCHECK, OCTAVIOUS 4D, MatriXX), Electronics Portal Imaging Devices (EPIDs), and log file method. Figure 2.12 shows a diagram of the patient-specific VMAT QA process.
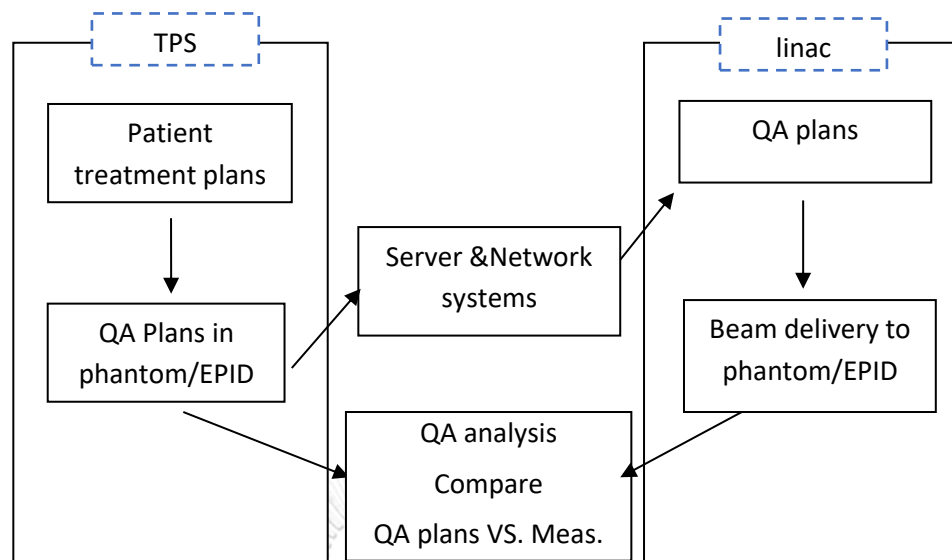
Figure 2.12 Diagram of patient-specific QA process

2.7 QA tool for patient-specific QA

The patient-specific QA tool can be classified into three main types; 1) measurement-based with phantom, 2) EPIDs, 3) Log files.

2.7.1 Measurement-based with phantom

Many commercial products provide a measurement-based with phantom methods for patient specific QA, as shown in Fig.2.13. Each device was developed with a different phantom shape. MapCHECK, OCTAVIOUS 4D, and Mtrixx were designed as cubic shapes by placing the detector in the same plane. Delta4 have placed a detector in two planes by crossing together with X-shape as shown in Fig. 2.14 a), while ArcCHECK was placed a detector around the circle phantom as shown in Fig. 2.14 b)

Figure 2.13 Commercial products of measurement-based with phantom devices for patient-specific QA
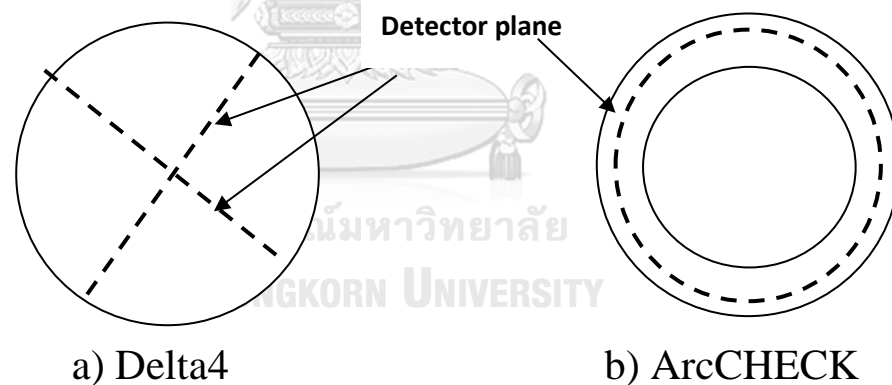
a) Delta4   b) ArcCHECK

Figure 2.14 a) Delta4 Phantom with detector placed at X-cross shape and b) ArcCHECK phantom with detector placed around the circle

2.7.2 Electronics Portal Imaging Devices (EPIDs)

EPID images are generally used to verify patient position [16]. According to the high density of the EPID detector, EPID can be used for dosimetry with an appropriate calibration [17-20]. EPID images for

patient-specific QA can perform with integrated and cine mode. Some TPS vendor provides the dose calculation in EPID such as Eclipse (Eclipse, Varian Medical System, Palo Alto, CA, USA) TPS using an algorithm called Portal Dose Images Prediction (PDIP), or the commercial products provide the dose calculation in water such as EPIDose (Sun Nuclear Corp., Melbourne, FL, USA), Epiqa (EPIdoss.r.o., Bratislava, Slovakia). Many publications [17-24] also developed dose calculation in EPID and water with integrated and cine mode.

2.7.3 Log file analysis

Log file commonly used in engineering filed is the tool to record the machine information during beam delivery, which is used to evaluate the machine performance. Log file information for radiation treatment includes gantry angle, collimator angle, MLC position, MU, collimator jaws position. Many publications [25-32] and commercial software developed patient-specific QA using log files such as Mobius3D software (Mobius Medical Systems, Houston, TX, USA), PerFRACTION (Sun Nuclear Corp., Melbourne, FL, USA). The commercial software available to generate the fluence map from log file data and then calculated dose distribution in CT images. For free software, Dynalog QA [27]and Pylinac [33]can generate the fluence map from log file data.

2.8 Characteristics of patient-specific QA methods

The characteristics of patient-specific QA methods can be summarized in Table 3. The semiconductor QA devices can perform a higher resolution than used ionization detectors, but the semiconductor has accumulated dose-effect to reduce sensitivity [34]. Several vendors of QA

devices also provide software to reconstruct 2D planar dose to 3D dose distribution.

2.8.1 ArcCHECK detector-arrays

The ArcCHECK (Sun Nuclear, Melbourne, FL, USA) is a diode array detector. There are 1,386 diodes placed in cylindrical phantom (Fig. 2.15) at a depth of 2.9 cm. The spacing between diodes is 1 cm, and diodes were placed as a spiral down with 21 cm array length to increase the spatial sampling rate.

Figure 2.15 ArcCHECK commercial QA tool

2.8.2 OCTAVIOUS 4D detector-arrays

The Octavious 4D (PTW, Freiburg, Germany) is an ionization chamber array placed in a rectangular shape (30 cm × 42 cm), which was inserted into a cylindrical phantom (Fig. 2.16). There are 729 ionization chambers with 1 cm spacing between the ionization chamber. To rotate array detectors perpendicular to the incidence beam, a measurement was controlled using a motor-related inclinometer reading.

Figure 2.16 Octavious 4D commercial QA tool

2.8.3 MatriXX detector-arrays

The MatriXX (IBA Dosimetry, Bartlett, TN, USA) is a diode array placed in a square shape (24.4 cm ×24.4 cm), which was inserted into a cubic phantom (Fig.2.17). There are 1,020diodes with 0.72 cm spacing between diodes.

Figure 2.17 MatriXX commercial QA tool

2.8.4 Delta4 detector-arrays

The Delta4 PT (ScandiDos AB, Uppsala, Sweden) is 1,069 p-type diodes, which were placed as x-cross-plane inserted inside cylindrical polymethylmethacrylate (PMMA) phantom (Fig.2.18). The diodes were placed with the spacing of 0.5 cm at the center and 1cm at the peripheral. The detection area of Delta4 is 20 cm×20 cm. To reduce the dose rate effect, the measurement was synchronized to the beam pulse of LINAC.
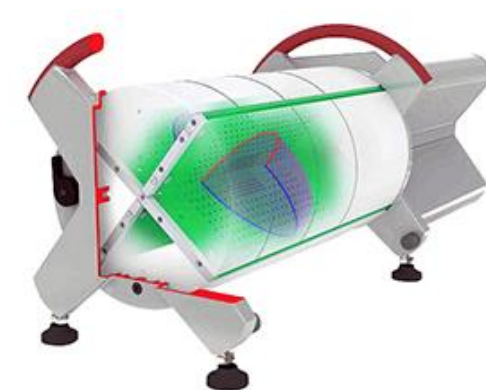
Figure 2.18 Delta4 PT commercial QA tool

2.8.5 EPID

EPID (Portal Vision, Varian Medical System, Palo Alto, CA, USA) is a photodiode material placed array with a resolution of 1,024×768 pixels. The irradiation area at SSD of 100 cm is 40×30 cm. There are three-layer materials for EPID, as shown in Fig. 2.19 a). Figure 2.19 b) explains how EPID can convert photon to signal. The first layer is a copper plate that was used to convert photon to an electron. The second layer is a phosphor screen that was used to convert electrons to light. The third layer is photodiodes (Amorphous silicon: aSi) used to convert light to charge.

a) Composite of EPID for each layer

b) the process to convert radiation to signals of EPID
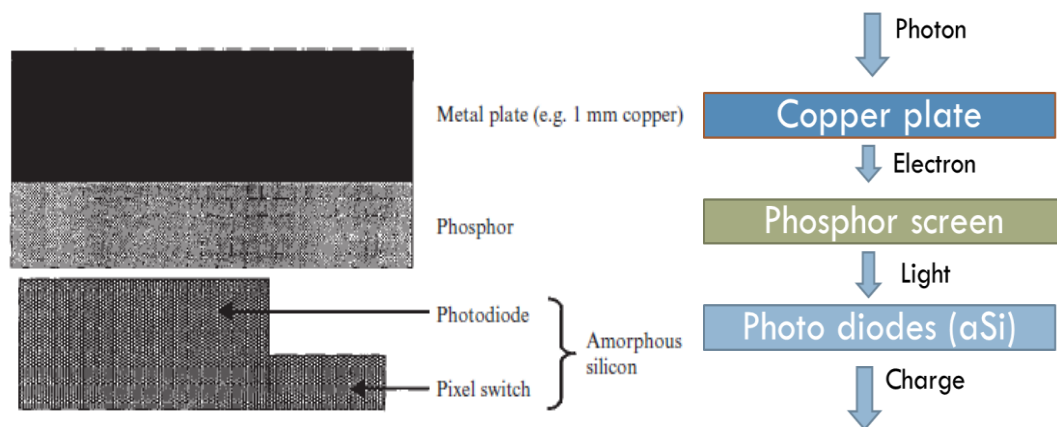
Figure 2.19 a) Composite of EPID for each layer and b) the process to convert radiation to signals of EPID

2.8.6 Log file

The log file is data recorded during beam delivery. For prior modern Varian LINAC (C-series model), data is recorded in ASCII text file as MLC log files (Dynalogs). Dynalogs are created in two separate files, one for MLC bank A and the other one for MLC bank B. The file structures for each bank include header and contents (shown in Table 2.3 a) and b)). Figure 2.20 shows the header and contents parameters, which were recorded in each column for Dynalogs.

Table 2.3 Structure of a) header and b) content for Dynalog file

a) header of Dynalog file

Line	Format	Description
1	B	Letter Indicating Version
2	<Lastname>,<Firstname>,<ID>	Patient information, up to 25 characters
3	<Field Serial Number>	if VARViS 6.2
	or	or
	<PlanUID>,<BeamNumber>	Treat 6.5 or greater
	or	or
	<Plan Filename>	Standalone
4	<long>	Tolerance
5	<long>	Number of Leaves in MLC
6	<long>	Clinac Scale
		0 = Varian Scale
		1 = IEC Scale

b) contents of Dynalog file

Column	Description
1	Current Dose Fraction or Gantry Angle
2	Previous segment number (starting with zero)
3	Beam hold-off state
	2 = LFIMRT carriage group transitions
	1 = MLC beam hold-off signal asserted
	0 = MLC beam hold-off signal not asserted
4	Beam on state
	1 = Clinac beam is on
	2 = beam is off
5	Previous segment dose index or previous segment gantry angle
6	Next segment dose index or next segment gantry angle
7	Gantry rotation in 10th of a degree
8	Collimator rotation in 10th of a degree
9	Upper Y1 jaw position in mm in the isoplane
10	Upper Y2 jaw position in mm in the isoplane
11	Lower X1 jaw position in mm in the isoplane
12	Lower X2 jaw position in mm in the isoplane
13	Carriage expected position in 100th of a mm
14	Carriage actual position in 100th of a mm
	The remaining columns contain the following values
	for each leaf in the carriage. (nLeaf = 0,1,2 and so on)
4*nLeaf + 15	Expected position in 100th of a mm
4*nLeaf + 16	Actual position in 100th of a mm
4*nLeaf + 17	Previous field position
4*nLeaf + 18	Next field position

Figure 2.20 Dose dynamic treatment dynalog file contents

For modern Varian LINAC (TrueBeam model), data is recorded in the binary file as Truebeam trajectory log files. Truebeam trajectory log files are created only one file. The file structures include header, sub-beams, axis data, and CRC. Table 2.4 a) and b) show header and sub-beams data for Truebeam trajectory log files, respectively.

Table 2.4 Structure of a) header and b) sub beams for TrueBeam trajectory log file

a) header of Truebeam trajectory log file

Line	Data Description
1	Signature 'VOSTL'
2	Version '2.1'
3	Header Size (fixed for now at 1024)
4	Sampling Interval in milliseconds
5	Number of axes sampled. (Indicates length of next entry)
6	Axis enumeration Coll Rtn - 0 Gantry Rtn - 1 Y1 - 2 Y2 - 3 X1 - 4 X2 - 5 Couch Vrt - 6 Couch Lng - 7 Couch Rtn - 9 MU - 40 Beam Hold - 41 Control Point - 42 MLC - 50
7	Samples per axis. For MLC, it is the number of leaves and carriages
8	Axis Scale 1 - Machine Scale 2 - Modified IEC 61217
9	Number of subbeams
10	Is Truncated? (Did the plan exceed 20 minutes?)
11	Number of snapshots
12	Reserved

b) sub beams of Truebeam trajectory log file

Line	Data Description
1	Control Point. Internally-defined marker that defines where the plan is currently executing.
2	Dose delivered in units of MU
3	radTime. In units of seconds it is the expected irradiation time of the subbeam. When actual irradiation time exceeds the expected irradiation time the plan is terminated.
4	Sequence number of the subbeam
5	Name of the subbeam
6	Reserved

The specifications for difference patient-specific QA tools have been summarized as shown in Table 2.5.

Table 2.5 Summarization of specification for difference patient-specific QA tools

Charect-eristics	MatriXX	OCTAVIOUS 4D	ArcCHECK	Delta4	EPID (Varian)
Phantom shape	2D array	Cylinder	Cylinder	Cylinder	2D array
Detector	Ion Chamber	Ion Chamber	Diode	Diode	Photo diode
Spacing (mm)	7.62	7.1	10	5/10	0.39
Number	1,020	1,405	1,386	1,069	1024×768
Pattern	Plane	Plane	Cylinder	Ortho. plane	Plane
Max. field size (cm)	24.4×24.4	27×27	27	20	40×30

2.9 The pros and cons for various type of patient-specific QA tools

The pros and cons of the different methods have been discussed, and we can summarize, as shown in Table 2.6. The pros of measurement-based with phantom are the absorbed dose calibration in water equivalent, traditional measurement, but the phantom set-up is required before beam delivery. The pros of EPIDs are no require time for set-up (only imager mode-up), high resolution to detect a small error in MLCs movements, available in modern Linear Accelerator machines (LINACs). However, the method for cine mode requires extensive data. Unlike EPIDs with cine mode, EPIDs with integrated mode requires less data than cine mode. They were comparing with integrated mode, cine mode available to explore for

each control point related to the gantry angle. The pros of log file analysis are no required time for set-up, but the method is not independent analysis.

Table 2.6 Pros and cons of difference QA methods

QA methods	Pros	Cons
Measurement-based with phantom	-Directly absorbed dose in water -Immediately readout	-Require phantom set up -Less resolution
EPIDs Cine mode Integrated mode	-No require setup -High resolution -Immediately readout	-Require more images (Cine mode) -Not suitable for VMAT QA (Integrated mode) -Available some machine
Log file	-No require setup	-No independent QA

2.10 Patient-specific QA analysis

The comparison of measured and plans dose are performed to analyze the QA results with a different method such as Distance-to-Agreement Analysis (DTA) [35, 36] and gamma index analysis [37]. For patient-specific QA analysis, the result is shown as passing rates compared between measurement and plans dose. DTA analysis was compared by finding the agreement with assigned distance criteria. Gamma index (Y)

analysis is compared using distance and dose agreement together, as explained in Fig. 2.21 and equation 2.1.

$$\gamma = \sqrt{\frac{(distance)^2}{(\Delta d)^2} + \frac{(dose\ difference)^2}{(\Delta D)^2}}$$ Equation 2.1

Where Δd is a distance difference criteria,

And ΔD is dose difference criteria

The passed point can be determined as $\gamma \leq 1$, while fails point can be determined as $\gamma > 1$.
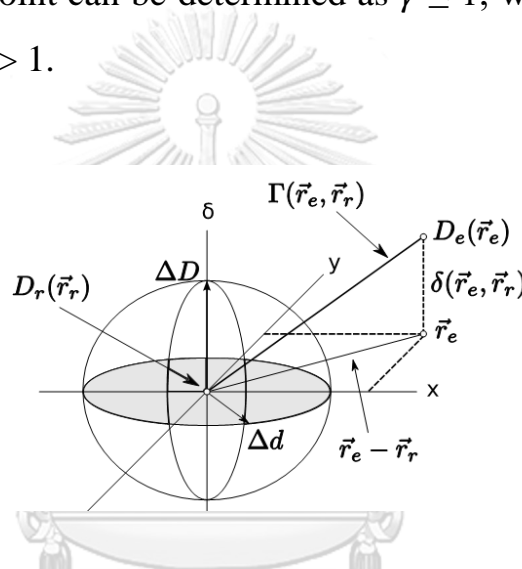
Figure 2.21 Geometric in 2D dose distribution for gamma index analysis method

2.11 Sensitivity and specificity

Generally, sensitivity and specificity are used in a medical diagnostic to measure the accuracy of the test. Sensitivity is the capability of a test to correctly identify the results with a disease, which is called true positive rate (TPR), whereas specificity is the test's capability to correctly identify the results without the disease, which is called true negative rate (TNR). Figure 2.22 shows the confusion matrix of sensitivity and specificity, which are explained the definition of True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN). Sensitivity and

specificity can be calculated, as shown in equation 2.2 and 2.3, respectively.

$$Sensitivity\ (\%) = \frac{TP}{(TP+FN)} \times 100\% \qquad \text{Equation 2.2}$$

$$Specitivity\ (\%) = \frac{TN}{(TN+FP)} \times 100\% \qquad \text{Equation 2.3}$$

The truth

Test result	Disease	No disease
Positive test	True Positives (TP)	False Positives (FP)
Negative test	False Negatives (FN)	True Negatives (TN)

Figure 2.22 Confusion matrix to explain definition of True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN)

The other parameter used to measure the capability of the test and tool are Receiver Operating Characteristics (ROC) analysis, the area under curve (AUC), and overlapping histogram, which are determined using the statistical method [38]. ROC is usually used to determine diagnostic decision-marking, which can be plotted between true positive rate (TPR) and false positive rate (FPR) with various thresholds (Fig. 2.23). The high

sensitivity to detect error can be shown as the curve is close to ideal classification (yellow line in Fig. 2.23). AUC can be determined by calculating the area under ROC curve used to evaluate the sensitivity to detect an error in terms of quantitative. The histogram's overlapping histogram is plotted for true positive and false positive events, as shown in Fig. 2.24. The less overlap can be determined as high sensitivity to detect an error.
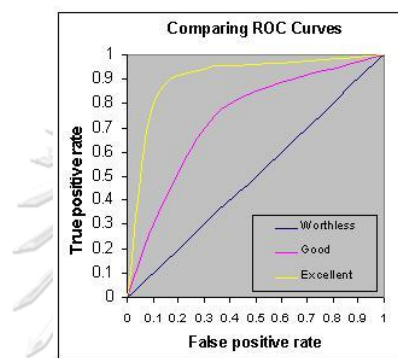
Figure 2.23 ROC curve plotted between true positive rate and false positive rate
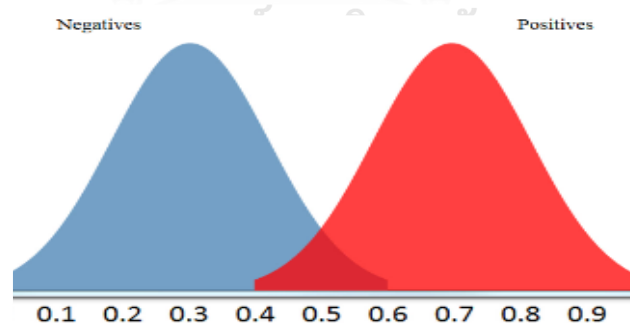
Figure 2.24 Overlapping histogram

2.12 Statistically process control (SPC) analysis

Statistically process control (SPC) analysis is commonly used in industrial quality assurance to improve products' quality [39]. Pawlicki et

al. [40] have applied SPC for radiotherapy quality assurance to set action thresholds separated between random and systematic error sources. Eun Rah et al. [41] have evaluated the SPC analysis to improve the reliability of proton quality assurance (QA) processes and provide an optimal customized tolerance level. Sanghangthum et al.[42] have applied SPC for patient-specific IMRT and VMAT QA to establish the % gamma pass's control limits. The behavior chart is a common method for SPC analysis consisting of four steps: 1) obtain data from a QA process, 2) compute summary statistics of the data, 3) determine process parameters by obtaining limits, 4) use process parameters to characterize process outcomes. For the behavior chart, the average chart can be calculated as three parameters: threshold (Au), centerline (Ac), and lower threshold (Al). The parameter is calculated as equation 2.4, 2.5, and 2.6 for Au, Ac, and Al, respectively.

$$A_u = \bar{\bar{x}} + 3\frac{\bar{R}}{d_2\sqrt{n}} \qquad \text{Equation 2.4}$$

$$A_c = \bar{\bar{x}} \qquad \text{Equation 2.5}$$

$$A_l = \bar{\bar{x}} - 3\frac{\bar{R}}{d_2\sqrt{n}} \qquad \text{Equation 2.6}$$

The factor 3 is added as an economic margin for the limits, is the average of all subgroups, and n is subgroup size.

The range chart can be calculated as three parameters: an upper threshold (Ru), centerline (Rc), and a lower threshold (Rl). The parameter is calculated as equation 2.7, 2.8, and 2.9 for Ru, Rc, and Rl, respectively.

$$R_u = \left(1 + 3\frac{d_3}{d_2}\right)\bar{R} \qquad \text{Equation 2.7}$$

$$R_c = \bar{R} \qquad \text{Equation 2.8}$$

$$R_l = (1 - 3\frac{d_3}{d_2})\bar{R}$$
<div align="right">Equation 2.9</div>

Where d_2 and d_3 are the factors for control limit depend on subgroup size (n) [43] as shown in Table 2.7.

Table 2.7 The d_2 and d_3 related to subgroup sizes (n)

n	d_2	d_3
2	1.128	0.8525
3	1.693	0.8884
4	2.059	0.8798
5	2.326	0.8641
6	2.534	0.8480
7	2.704	0.8332
8	2.847	0.8198
9	2.970	0.8078
10	3.078	0.7971
11	3.173	0.7873
12	3.258	0.7785
13	3.336	0.7704
14	3.407	0.7630
15	3.472	0.7562
20	3.735	0.7287
25	3.931	0.7084
30	4.086	0.6927
40	4.322	0.6692
50	4.498	0.6521
100	5.015	0.6052

The different error sources need to be determined to separate the error type. For example, the data contains only a random error, or the data contains both errors (random and systematic errors).

2.13 Predictive model and Machine learning

Predictive modeling consists of two steps as follows; qualification, validation. Kang et al. [44] have proposed the principles of modeling in radiation oncology: Consider both dosimetric and non-dosimetric predictors, Manually curate predictors before automated analysis, Select a method for automatic predictor selection, Consider how predictor multicollinearity is affecting the model, Correctly use cross-validation to

improve prediction performance and generalization to external data to provide model generalizability with external data sets when possible, Assess multiple models and compare results with established models.

Two main methods can be used as follows; traditional statistical methods and machine learning methods. For traditional statistical methods, Logistic regression (LR) map is commonly used to predict a model in the radiation oncology field that appropriates a few parameter predictions. Figure 2.25 shows the example LR model for one dimension (total dose) and two dimensions (total dose, GTV sizes) where blue dots are local control failure and orange are local control success.
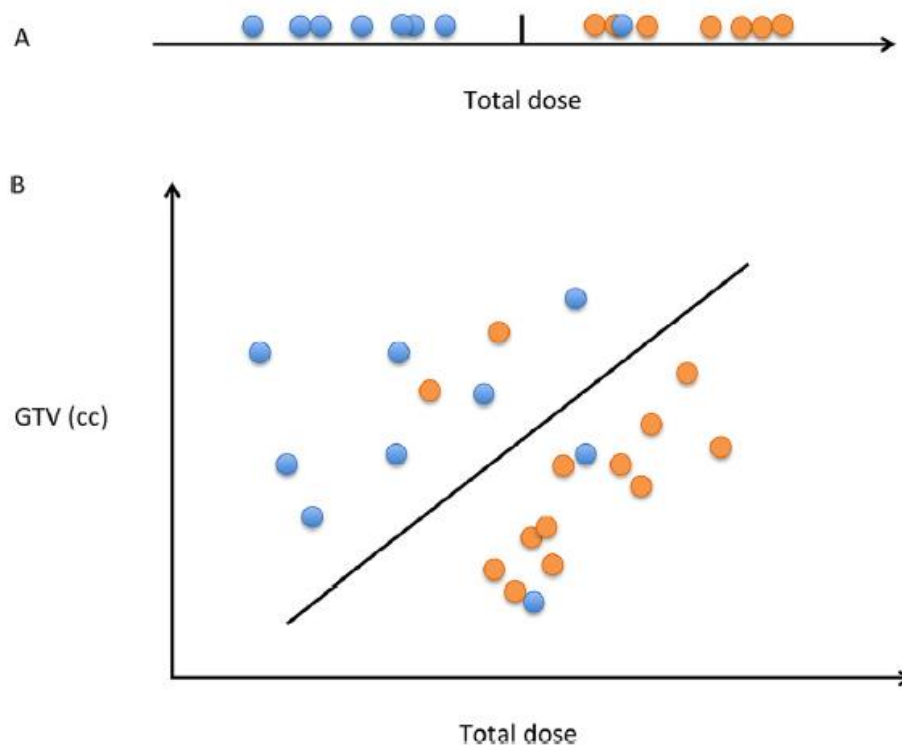
Figure 2.25 LR model in term of a) one dimension and b) two dimensions where blue dots are local control failure and orange are local control success

Machine Learning (ML) is firstly applied in computer science with a different purpose. Samuel [45] developed the ability of a computer to learn without being explicitly programmed. Mitchell [46] developed the computer to learn from experience. Alpaydin [47] programmed computers to optimize a performance criterion using example data or experience. ML have been applied in the related field such as artificial intelligence, probability and statistics, computer science, information theory, neuroscience, psychology, control theory, and philosophy. ML can be categorized into three types according to the underlying task; supervised learning, unsupervised learning, and reinforcement learning (RL). ML has been studied in many medical fields, including computer-aided detection and diagnosis, image segmentation, and outcomes modeling. Several algorithms have been used for ML as follows; Decision Trees (DT), Naïve Bayes (NB) classifiers, k-nearest Neighbors (*k*-NN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Deep Learning (DL). Decision Trees (DT) is a simple algorithm by answering questions in a predefined order. Naïve Bayes (NB) classifiers are the model algorithm to predict output probabilistic related to among variables. K-nearest Neighbors (*k*-NN) is a method to classify closest neighbors in the dataset used for classification and regression. Support Vector Machine (SVM) is a method to predict with multi parameters and non-linear correlation. Artificial Neural Network (ANN) is a method to predict correlation with weight factors in different parameter connects to each layer (parameter). Figure 2.26 shows a diagram of the Artificial Neural Network (ANN) process to predict local control and overall survival. Unlike ANN, Deep Learning (DL) is used to solve the problem with a higher number of hidden layers and perform supervised or unsupervised learning. Bibault et al. [48]

have summarized the benefits and limitations of different machine learning
algorithms, as shown in Table 2.8.
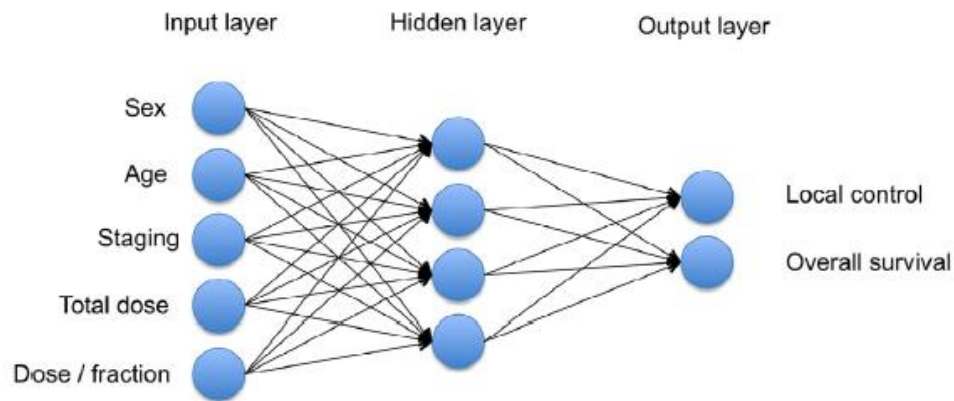
Figure 2.26 Diagram of the Artificial Neural Network (ANN) process to
predict local control and overall survival

Table 2.8 The benefits and limitations of different machine learning
algorithms

Algorithm	Advantages	Limitations
Decision Tree	Easy to understand Fast	Classes must be mutually exclusive Results depend on the order attribute selection Risk of overly complex decision trees
Naïve Baysian	Easy to understand Fast No effect of order on training	Variables must statically independent Numeric attributes must follow a normal distribution Classes must be mutually exclusive

Naïve Baysian		Less accurate
k-nearest Neighbors	Fast and simple Tolerant of noise and missing value in data Can be used for non-linear classification Can be used for both regression and classification	Variable with similar attributes will be sorted in the same class All attributes are equally relevant Requires considerable computer power as number of variable increases
Support Vector Machine	Robust model Limit the risk of error Can be used to model non-linear relations	Slow training Risk of overfitting Output model is difficult to understand
Artificial neural network and deep learning	Tolerant of noise and missing value in data Can be used for both regression and classification Can be easily updated with new data	Output model is difficult to understand (black-box) Risk of overfitting Requires a lot of computer power Require experimentation to find the optional network structure

CHAPTER 3

LITURATURE REVIEWS

3.1 Beam complexity

The beam complexity metric is the score to tell that the beam has a variety of parameters such as MU, leaf sequence, aperture sequence, intensity, etc. related to [49, 50] beam delivery error. Beam complexity can be categorized in various types corresponding to different approaches such as fluence metrics, deliverability metrics, accuracy metrics. Chiavassa et al. [51] have reviewed the beam complexity metrics for IMRT and VMAT and summarized them in table 3.1. Table 3.1 summarizes the different types of plan complexity according to different approaches. The first classical beam complexity was Modulation Index (MI) and Modulation Complexity Scores (MCS). MI introduced by Webb et al. [52] is used to determine the beam's complexity based on the mean and standard deviation of the beam, which can be calculated as equation 3.1.

$$MI\ (F) = \int_0^F z(f)df \qquad \text{Equation} \qquad 3.1$$

Where Z is a histogram which shows the number of adjacent intensity changes that exceeds the fraction.

The number of passing intensity (N) for each fraction can be determined as follows:

$$N_x: \Delta x = \left|I_{i,j} - I_{i+1,j}\right| > f\sigma$$

$$N_y: \Delta y = \left|I_{i,j} - I_{i,j+1}\right| > f\sigma$$

$$N_{xy}: \Delta xy = \left|I_{i,j} - I_{i+1,j+1}\right| > f\sigma$$

$$N_{yx}: \Delta yx = \left|I_{i,j} - I_{i-1,j+1}\right| > f\sigma$$

Where i is beam intensity for each direction (x,y),

f is the fraction of standard deviation,

σ is the standard deviation of beam intensity for all beam.

MCS introduced by Nivenetet et al. [53] is used to determine the beam's complexity based on the variability of leaf positions, aperture area between segments, and segment weight, which can be calculated as equation 3.2.

$$MSC_{beam} = \sum_{i=1}^{I-1} \left\{ \left(\frac{LSV_{CP_i} + LSV_{CP_{i+1}}}{2} \right) \times \left(\frac{AAV_{CP_i} + AAV_{CP_{i+1}}}{2} \right) \times \left(\frac{MU_{i.i+1}}{MU_{beam}} \right) \right\}$$

Equation 3.2

Where i is the number of control point,

LSV is leaf sequence variability,

 AAV is aperture area variability,

MU is the monitors unit for the beam.

Nauta et al. [54] developed a beam complexity score using fractal dimension analysis with the variation, power spectrum, and variogram to quantify the modulation degree in intensity-modulated radiation therapy (IMRT) treatment fields. Masi et al. [55] have modified MSC for VMAT plan by determining gantry rotation as a control point to evaluate the impact of plan parameters on the dosimetric accuracy of VMAT. A Simple method has been introduced to determine the accuracy of beam delivery, such as the ratio between MU and control point (CP) number (MU/CP) [56], MU per Gy [57], and plan normalized monitor unit [58]. Du et al. [59] used the irregular beam shape to determine beam delivery accuracy, called plan averaged beam irregularity (PI). Crowe et al. [60] introduced the new complexity score using the characteristic of aperture and position of leaf such as small aperture score (SAS), close leaf score (CLS), cross-axis

score, and mean asymmetric distance (MAD). Park et al. [61] determined a new metric of MLC speeds and accelerations to predict plan delivery accuracy of VMAT using a fraction of leaf speed and leaf acceleration in different levels. Park et al. [62] introduced the new fluence method to determine beam complexity using texture analysis with five parameters; contrast, correlation, energy, entropy, and homogeneity.

Table 9 Categorization of beam complexity scores according to different approaches [1]

Category	Complexity matrices	Reference
Fluence complexity	FMC - Fluence map complexity	Llacer et al [2]
	MI - modulation index	Web et al. [3]
	2D MI - 2D modulation index	Nicolini et al. [4]
	MIR - Maximum intensity ratio	Coselmon et al. [5]
	PIMV - Plan intensity map variation	
	Fractal dimension analysis methods: the variation, power spectrum and variogram methods	Nauta et al. [6]
	Textural features: ASM, IDM, contrast,	Park et al. [7]

	variance, correlation and entropy	
Deliverability	MU, MU/Gy or PMU - Monitor Unit, monitor unit per Gy or plan normalized monitor unit	Du et al [8], Mohan et al. [9], Masi et al. [10]
	AAV - Aperture area variability	McNiven et al. [11]
	LSV - Leaf sequence variability	
	MCS - Modulation complexity score (combination of LSV and AAV)	
	DR - Variations of the nominal DR	Nicolini et al. [4]
	GS - Variation of gantry speed	
	Degree/MU - The gantry angle per MU	Miura et al. [10]
	mm/MU - Leaf travel per MU	
	MU/CP - Number of Monitor unit per Control Point and	Shen et al. [12]

	proportion of CP with MU <3 (%MU/CP <3)	
	S_{l-h} - The average proportion of leaf speeds from a given range	Park et al. [13]
	A_{l-h} - The average proportion of leaf accelerations from a given range	
	MI_s - Modulation index for speed of MLC	Park et al. [14]
	MI_a - Modulation index for speed and acceleration of MLC	
	MI_t - Modulation index for speed and acceleration of MLC, gantry acceleration and dose rate variation	

	MCS$_v$ or MCS$_{arc}$ - Modulation complexity score for VMAT plans	Masi et al. [10]
	LT - Leaf travel	
	LTMCS - Combination of LT and MCS$_v$	
	MISPORT – Modulation index for station parameter optimized radiation therapy	Li and Xing [15]
Accuracy	Average leaf gap	Nauta et al. [6]
	MFA - Mean field area	Kairn et al. [16],
	SAS - Small aperture score and SAS(x)	Crowe et al. [17]
	CLS - Closed leaf score	
	CAS - Cross-axis score	
	MAD - Mean asymmetry distance	
	SA/CP - Segment area per CP	Shen et al. [12]

	Modulation degree	Heijmen et al. [18]
	PA - Plan averaged beam area	Du et al. [8]
	Segment area/Perimeter or Circumference/area	Carlsson et al. [19], Gotstedt et al. [20]
	EM - Edge metric	Younge et al. [21]
	EAM - Edge area metric	Gotstedt et al. [20]
	CAM - Converted aperture metric	
	LOIC - Leaf offset impact on calculation	Mathot et al. [22], Dechambre et al. [1]
Accuracy/Deliverability	MI_c - Comprehensive modulation index	Park et al. [23]

3.2 Evaluation of Delta4 characteristics

Several authors have evaluated the characteristics of Delta4. Nilsson [76] have introduced and characterized a novel diode-based detector (Delta4). Sadagopan et al. [77] have characterized and evaluated the Delta4 QA tool in terms of day-to-day reproducibility, dose rate independence, and linearity. Feygelmanet al.[78]have evaluated the characteristic of biplanar diode array dosimeter (Delta4) for patient-specific IMRT QA in terms of rotational dependence, the accuracy of point dose, and IMRT measurement accuracy 3D dose distribution. The results

were found that the rotational dependence was ± 3% in ± 1° angular intervals, the point dose difference between measured and calculate dose was -0.2±9%, and IMRT measurement with gamma criteria of 3%, 3mm was more than 95% passed in ten plans and 92% in one plan. They concluded that the Delta4 device has a good characteristic for patient specific IMRT QA. They also suggested that the capabilities and limitations of Delta4 must be evaluated concerning the specific clinical situation.

3.3 The accuracy of MLC

Many publications have investigated the accuracy of MLC position. Losasso et al.[24]have investigated the performance of MLC in IMRT delivery. They found that the accuracy of MLC maybe affected by many factors such as mechanical limitations, motor fatigue, failure of supporting system, specific gantry or collimator angle due to gravity effects. Xia et al. [25]have investigated the accuracy of MLC position, and they found that the error may be affected by calibration errors, differences between light and radiation field, or even communication delay between the treatment console and the MLC controller. Cadman et al. [26]found dosimetric difference between TPS and measurement could be a wrong setting of dosimetric leaf gap in TPS. Oliver et al.[26]categorized MLC error in term of random (Type 1), systematic leaf gap shift (Type2) and systematic open/close (Type 3). Rangel et al.[27]have evaluated the tolerance of MLC accuracy in IMRT delivery. They have suggested to simulate the maximum error of ±2 mm corresponds to the maximum deviation allowed by the MLC controller in treatment delivery system. Budgell et al. [28]reported that variability in leaf position is very small (±1 mm of standard deviation).

3.4 EPID-based dosimetry for patient-specific QA

Many publications have investigated the potential of cine mode EPID for machine QA and pre-treatment VMAT QA. Adamson and Wu [84] have developed cine-EPID images to verify the gantry angle for pre-treatment VMAT QA. With cine-EPID images, the gantry angle could be measured to within $0.0 \pm 0.3°$ for the static gantry and $0.2 \pm 0.2°$ for arc acquisitions. McCurdy and Geer [85] have investigated the dosimetric properties of an amorphous-silicon EPID used in continuous acquisition mode for application to dynamic and arc IMRT. The properties in terms of dose linearity, reproducibility of response, and image stability were evaluated. They found that the EPID with continuous acquisition mode suited to time-resolved dosimetry applications such as arc-IMRT and dynamic IMRT. Bakhtiari et al. [21] have developed a cine-EPID based for patient-specific VMAT QA. Each cine images related to each control point were converted to the edge of MLCs by using a 38% iso-intensity line and then compared to the MLCs for each control point from TPS. With cine-EPID based method, it possible to use for verifying MLCs position in patient-specific VMAT QA. Fuang rod et al. [86] have developed a cine-EPID based method to verify the jaw position in IMRT and VMAT with jaw tracking. EPID images with cine mode were removed scattered radiation for the first step. Then histogram clustering was applied to determine the optimized threshold level used to separate the image region within the jaw and without the jaw. The algorithm's accuracy was evaluated by determining the root mean square (RMS) difference between planed and detected jaw positions in nine IMRT treatment plans (three head and neck, three prostate, and three brain plans and a prostate VMAT plan. The accuracy of jaw detection with cine-EPID based method was ±1 mm RMS

error for statics-jaw IMRT plans, and 1.5 mm RMS error for dynamics-jaw VMAT plan. Liu et al. [24] have developed a cine-EPID based method for pre-treatment VMAT QA. The pixel value of images was converted to images with three steps: pixel value scaling, profile correction, and pixel size scaling. The portal dose image prediction (PDIP) from TPS in VMAT plans was split into multiple sub-arcs to compare with the imaging dose calibrated using this method. In this study, the gantry rotation recorded in the image header was additionally used to determine gantry error in 3D gamma analysis. The accuracy of a cine-EPID based method was analyzed using 2D gamma analysis and 3D gamma analysis (images+ gantry angle axis). The gamma criteria of 3%, 3mm were used for 2D gamma analysis, and the gamma criteria of 3%, 3mm, and 2o angle were used for 3D gamma analysis. Delta4 device was used as a traditional QA tool to compare with cine-EPID based methods. 3D gamma passing rates in 21 plans were 92.8±1.4% for cine-EPID based method and 99.9±0.1% for Delta4 QA. The results indicated that this technique is efficient and accurate for pre-treatment VMAT QA comparable to the other QA tool.

3.5 Utilities of log file QA

Several authors have investigated utilities of the log file for machine performance evaluation and patient-specific QA. Mcgarry et al. [30] have investigated the accuracy of MLCs in different Varian LINACs by using the log file. They found that TrueBeam LINACS had better MLCs position accuracy (<1.0 mm) than 2300IX (2.5 mm). Losasso [87] has studies the Varian MLCs performance, and they found that the old Varian LINACs (pre-TrueBeam) had communication delayed of MLCs around 50 ms. Alonso et al.[88] have investigated the accuracy of MLCs in pre-

TrueBeam and TrueBeam. The delay effect of MLCs communication was determined in this study. They found that the pre-TrueBeam had RMS errors of 0.306 mm (with delay effect) and 0.03 mm (without delay effect), and the accuracy of TrueBeam MLCs are statically better than pre-TrueBeam MLCs. Sun et al. [89] have evaluated the effectiveness and efficiency of independent dose calculation followed by machine log file analysis compared to conventional measurement-based IMRT QA. They found that the independent dose calculation followed by machine log file analysis can detect dose calculation error from TPS and verify data transfers. Agnew et al. [26] have investigated the correlation of phantom-based and log file patient-specific QA. Fluence delivery maps were created from log file data and then compared to planed fluence maps. The gamma passing rates were used to determine the correlation between the phantom-based and log file. They found a strong correlation between the phantom-based and log file in gamma passing rates.

3.6 Error detection of QA tool

Several authors have investigated the error detection capability of different QA method for patient-specific QA. Hauri et al. [90] have investigated the sensitivity of Delta4 to detect gantry error in VMAT plans. In this study, the gantry error was simulated as a sinusoidal form related to the gravitational effect. Heilemann et al. [91] have investigated the sensitivity of Delta4 to detect MLC error. The MLCs error was simulated included open, close, and shift MLCs bank. Liang et al.[92] have investigated the sensitivity of three VMAT QA systems (ArcCHECK, Delta4, in-house developed EPID technique) to machine errors. In this study, 3D gamma analysis (images+ gantry angle axis) was used for the

EPID technique. The machine errors were simulated in the sinusoidal function of gantry angle, MLCs position, and MU. They found that ArcCHECK is more sensitive to detect gantry error and Delta4 is more sensitive to detect MLCs error. EPID technique has a same sensitivity to detect both MLC and gantry errors by adjusting the extra angle-to-agreement parameters. Defoor et al. [28] have investigated the capability of three QA methods (Delta4, MU-EPID, Dynalog QA) to detect gantry angle, MLCs position, and MU errors. They found that the errors were detected at a rate of 60, 27, and 47% for Delta4, MU-EPID, and Dynalog QA, respectively. Maraghechi et al. [93] have investigated the capability of ArcCHECK and TrueBeam EPID to detect MLCs positioning errors. The MLCs errors were simulated with a mean value of 0.25, 0.5, 1, and 2 mm and a standard deviation of 0.1 mm on 25, 50, 75, and 100% of the control points.

3.7 Correlation of plan complexity and QA results

Several authors have evaluated the correlation between QA results and plan complexity with different QA tools and plan-complexity parameters. Chiavassa et al. [51] have reviewed and summarized the study of the correlation between beam complexity and patient-specific QA results, as shown in Table 3.2. Rajasekaran et al. [50] have evaluated the correlation between MSC and 3D GPR of Octavius 4D QA results. Only gamma criteria of 3%, 3mm was used for QA analysis. They found that out of 90 correlation analyses between the MCS and GPR, only 3 had the r value>0.5. Wang et al. [94] have investigated the impact of beam complexities on planar quality assurance. The IMRT QA was performed using portal dosimetry, including 20 dynamics IMRT plans. The

complexity scores include fractal dimension, monitor unit, modulation index, fluence map complexity, the weighted average of the field area, weighted average of field perimeter, and small aperture ratio. They found that the most relation was fractal dimension for DMLC. Masi et al. [58] have evaluated the effect of plan parameters on VMAT dosimetric accuracy to the possibility of scoring plan complexity. The plan complexities were considered the following parameters: average leaf travel (LT), MCS, MU, and a multiplicative combination of LT and MCS (LTMCS). The correlation analysis was performed using Pearson's method. They found that LT, MCS, and LTMCS have significantly correlated with VMAT dosimetric accuracy. Mcgarry et al. [95] have demonstrated the benefit of complexity metrics such as MCS and MU in multi-institutional audits of VMAT delivery. VMAT QA was performed using the Octavius phantom and seven 29-2D arrays. They found that MU and MCS have a role in assessing plan complexity in audits along with plan quality metrics. This study suggests that plan complexity metrics indicate plan deliverability but should be analyzed with plan quality.

Table 10 Main correlations studies between complexity metrics and patient-specific QA results [1]

Reference	Technique (localization)	QA system	Analysis `	Evaluated metrics	Correlation coefficients
Crowe et al. [29]	52 ss-IMRT plans (multisite)	EPID with Epiqa system (EPIdos)	γ 3%/3 mm and 2%/2 mm	MU, AAV, CAS, MAD, MCS, MI, SAS1, SAS5 and SAS10	Significant (F)

				MFA, CLS and FMC	Not significant (F)
Crowe et al. [29]	70 VMAT plans (multisite)	EPID with Epiqa system (EPIdos)	γ 3%/3 mm and 2%/2 mm	MAD, MCS, MI and SAS1	Significant (F)
				MU, MFA, AAV, CLS, CAS, FMC, SAS5 and SAS10	Not significant (F)
Du et al. [8]	65 ss-IMRT plans (prostate, head and neck, and spine) and 26 VMAT plans (prostate)	Ion chamber and radiographic films	Dose difference and γ 5%/3 mm	PA, PI, PM and PMU	None (s)
Park et al. [7]	40 VMAT plans (20 prostate, 20 h&N)	MapCHECK (Sun Nuclear)	γ local and global 2%/2 mm	$Contrast_{1,5,10}$, ASM_1, $IDM_{1,5}$, $Variance_{1,5}$, $correlation_{5,10}$	Moderate (s)
				ADM_5, IDM_{10}, $Variance_{10}$, $correlation_1$, $Entropy_{1,5,10}$	Weak (s)
				ASM_{10}	None (s)

McNiven et al. [11]	243 IMRT plans (multisite)	MapCheck (Sun Nuclear)	γ local 3%/3 mm and 2%/2 mm	MCS and MU	None (NA)
Masi et al. [10]	142 VMAT plans (multisite)	Delta4 phantom (Scandidos)	γ local 3%/3 mm and 2%/2 mm	LT, MCSv and LTMCS PMU	Moderate (p) Weak (p)
Shen et al. [12]	71 VMAT (nasopharyngeal cancer)	ArcCHECK (Sun Nuclear)	γ Individual volume-based 3D	MU/CP, SA/CP	Significant (c)
				%MU/CP < 3, %SA < 5.5, MCSv/arc and LT	Not significant (c)
Park et al. [13]	40 VMAT plans (20 prostate, 20 h&n)	MapCHECK (Sun Nuclear)	γ global and local 1%/2 mm	Mean MLC speeds and mean MLC accelerations	Weak and moderate (s)
Park et al. [14]	40 VMAT plans (20 prostate, 20 h&n)	MapCHECK (Sun Nuclear)	γ local 2%/2 mm	MIs, Mia and MIt	Moderate (s)
				MIsport	Weak (s)
				LTMCS	None (s)
Crown et al. [17]	122 ss-IMRT beams (prostate)	MapCHECK (Sun Nuclear)	γ 3%/3 mm and 2%/2 mm	MI, MFA, SAS5 and SAS10	Significant (F)

				MCS, LSV, AAV, FMC, CLS, CAS, MAD, SAS2 and SAS20	Not significant
Godtstedt et al. [20]	30 Artificial IMRT / VMAT beams	EPID	Dose difference pass rate (3 and 5%)	CAM, EAM, EM, Circumference/area and MU/Gy	Strong (p)
				Aperture area	Moderate (p)
				MCS and Aperture Irregularity	Weak (p)
Godtstedt et al. [20]	30 Artificial IMRT / VMAT beams	Rradiochromic films	Dose difference pass rate (3 and 5%)	CAM, EAM, EM, Circumference/area and MU/Gy	Strong (p)
				MCS and Aperture Irregularity	Moderate (p)
				Aperture area	Weak (p)
Dechambre et al. [1]	93 VMAT plans (multisite)	ArcCHECK (Sun Nuclear)	γ global 3%/3 mm and 2%/2 mm	LOIC, CAS and MCSv SAS and PMU MFA, MAD and CLS	Moderate (p) Weak (p) None (p)
Park et al. [23]	52 VMAT plans (22 prostate, 30 h&n - two institutions)	MapCHECK (Sun Nuclear)	γ local 2%/2 mm	MIt, MIc and MIsport MIc, LTMCS, variance, PI and PMMCSv,	Strong (s) Moderate (s) Weak (s) None (s)

				contrast, contrast$_{edge}$ and PA PMU	
Agnew et al. [30]	30 VMAT (prostate, prostate and pelvic node, H&N)	OCTAVIUS 4D and 729 array (PTW)	γ global 2%/2 mm	MCSv	Moderate (p)
				MU	None (p)
Glenn et al. [31]	343 IMRT/ VMAT plans (H&N - 312 institutions)	TLD and Radiochromic films	Dose difference (5%) and γ 7%, 4 mm	MU, MCS, EM, PI, PM, MIs, MIa, MIt, LT, Mean DR variation, Mean GS var. Mean MLC speed var.	None (s)
Jurado-Bruggeman et al. [32]	36 VMAT plans (2 prostate and 2 h&n done by nine institutions)	ArcCHECK (Sun Nuclear)	γ global 3%/3 mm and 2%/2 mm	MU	Strong (p)
				PI, MCSv, MIt	None (p)
McGarry et al. [33]	39 VMAT plans (virtual volumes – 34 institutions)	OCTAVIUS II and 729 array (PTW)	γ global and local 1%/2 mm	MCSv and MU	Weak (p) but Moderate (p) for Varian linacs
Park et al. [34]	202 IMRT plans (multisite)	ArcCHECK and MapCHECK (Sunnuclear)	γ global 2%/2 mm	PI and MCS	Moderate (s)

				MIs, PA and PM	None (s)

Wang et al. [35]	20 IMRT plans (10 nasopharyngeal cancer and 10 prostate) with intentional MLC leaf errors	ArcCHECK (Sun Nuclear)	γ 2%/2 mm	MCS	Strong (s)
Park et al. [36]	240 VMAT plans (multisite)	ArcCHECK (Sun Nuclear)	γ local 2%/2 mm	MIt, MIc, LTMCS, MIsport, PI and PM MCSv and PMU PA	Moderate (s) Weak (s) None (s)

Statistical analysis methodologies used are Spearman (s), Pearson (p), Canonical (c), or F-test (F). Correlation for Spearman and Pearson coefficients were considered as strong $r \geq 0.7$, moderate for $0.5 \leq r < 0.7$, weak for $0.4 \leq r < 0.5$ and none for $r < 0.4$. All presented data are associated with a p-value ≤ 0.05.

3.8 Machine learning-based prediction

Machine learning (ML) have been introduced to the medical physics field to solve the different class of problems such as the automation of the process and predictive model. Many publications have been utilized the benefit of ML to develop model-based prediction in patient-specific QA process. Table 3.3 summarizes ML application for model-based prediction

of patient-specific QA with different QA tools and different criteria. Tomori et al. [103] developed a model-based model to predict patient-specific QA results using DL with the CNN algorithm. Results indicated that DL could provide a useful method for model-based prediction. Nyflot et al. [104] demonstrated the feasibility of using a conventional neural network (CNN) to classify errors in the patient-specific QA process. Interian et al. [105] compared the performance of CNN and expert-designed features. They found CNN with transfer learning can predict IMRT QA passing rates by automatically designing features from the fluence maps without human expert supervision. Vlades et al. [106] validated the ML approach to virtual IMRT QA, and they demonstrated that virtual IMRT QA could predict passing rates using different measurement techniques and across multiple institutions. Li et al. [107] assessed the accuracy of ML to predict and classify QA results for VMAT plans, and they found Poisson lasso (PL) regression model could accurately predict QA results, while the Random forest (RF) model was preferred for QA results classification. Vlades et al. [108] developed an algorithm for predicting IMRT QA results called virtual IMRT QA. They found that the algorithm can detect the fail QA results due to setup errors and mall differences between matched linacs.

Table 11 Research of ML for patient-specific QA model-based prediction

Reference	Study size	QA systems	Analysis	Algorithm
Tomori et al. [37]	60 plans of IMRT prostates	GAFCHROMIC EBT-3 film	2%(global)/2mm, 3%(global)/2mm, 2%(global)/3mm, and 3%(global)/3 mm.	DL with CNN
Nyflot et al. [38]	23 IMRT plans including head and neck cancer, lung cancer, rectal cancer, sarcoma, and glioblastoma	EPID	3%(global)/3 mm.	DL with CNN
Interian et al. [39]	498 IMRT plans including breast, central nervous system, gastrointestinal, genitourinary, gynecologic, head and neck, lung, and pediatric	2D diode-arrays (MapCHECK2)	3% (local) /3 mm	DL with CNN
Vlades et al. [40]	489 IMRT plans (Multiple institutes)	EPID, 2D diode-arrays (MapCHECK2)	3%(global)/3 mm	ML with PL regression
Li et al. [41]	303 VMAT including GYN and H&N	Ion chamber array (MatriXX)	2%(global)/2mm, 3%(global)/2mm,	ML with PL regression and

			and 3%(global)/3 mm.	RF classification
Vlades et al. [42]	498 IMRT plans including breast, central nervous system, gastrointestinal, genitourinary, gynecologic, head and neck, lung, and pediatric	2D diode-arrays (MapCHECK2)	3% (local) /3 mm	ML with PL regression

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY
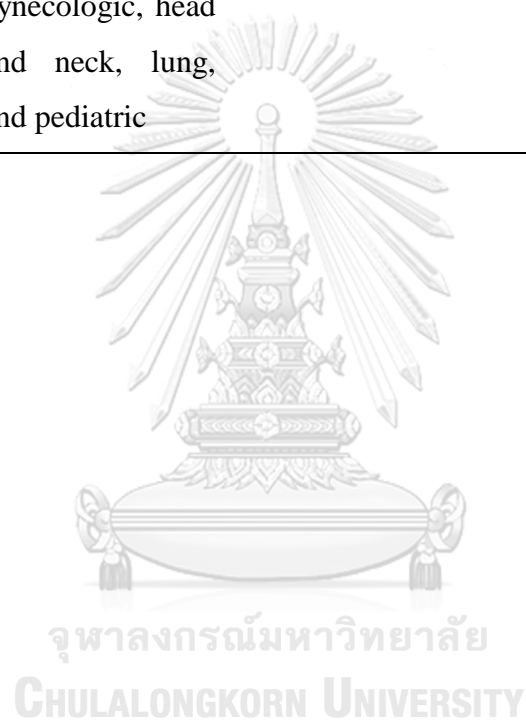
CHAPTER 4

MATERIAL AND METHODS

4.1 Overview and research workflow

Figure 4.1 shows the workflow of this research. The first step of the study was assessment the error detection in three QA tools (EPID-based dosimetry, Delta4, Log file-based) to find the highest QA tool that can detect the various type of errors. Complex artificial errors were introduced to the original plans to determine the error detection sensitivity of QA tools. These intended errors were simulated into three possible scenarios: uncertainty, miss-calibration, and worst-case scenario. For the next step of the study, the highest QA tool to detect error was selected from the first step to develop the predicting gamma evaluation model for patient-specific VMAT QA. ML approach with classification and regression model was used for model-based prediction. The accuracy of the model was validated by using sensitivity and specificity scores.
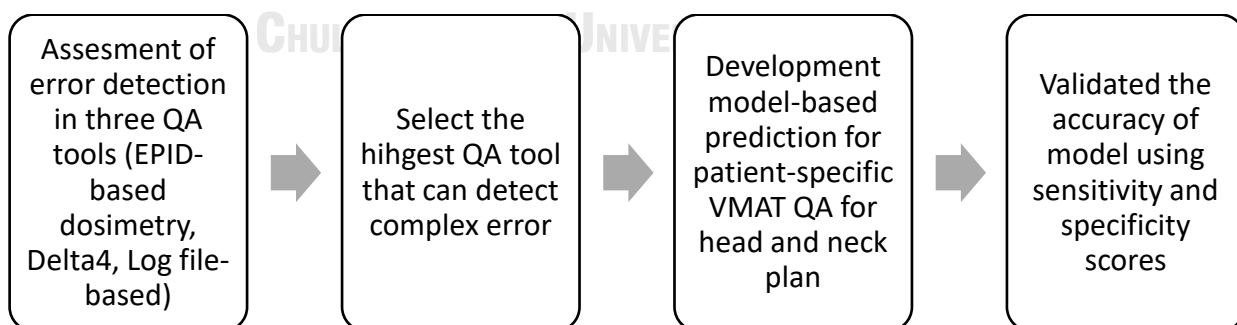
Figure 27 Workflow of this research

4.2 Material and Instruments

4.2.1 Linac and TPS

In this study, Eclipse TPS (Eclipse V.10.1, Varian Medical System, Palo Alto, CA, USA) was used for dose calculation in VMAT plans. Figure 4.2 shows the Eclipse TPS computer used for this study. Trilogy and TruBeam linacs (Varian Medical System, Palo Alto, CA, USA) were used to deliver radiation for patient-specific VMAT QA as shown in Fig. 4.3 a), and EPID (Portal Vision, Varian Medical System, Palo Alto, CA, USA) images with pixel sizes of 1024×768 aSi-1000 was used for image acquisition as shown in Fig. 4.3 b). EPID images were acquired with integrated mode using Image Acquisition System -1200 (IAS-1200). Linacs created the trajectory log file during beam delivery, and information is captured every 20 ms. Delta4 PT (ScandiDos AB, Uppsala, Sweden) with 1,069 p-type diodes were used for this study.

Figure 28 Eclipse TPS (Eclipse V.10.1, Varian Medical System, Palo Alto, CA, USA)
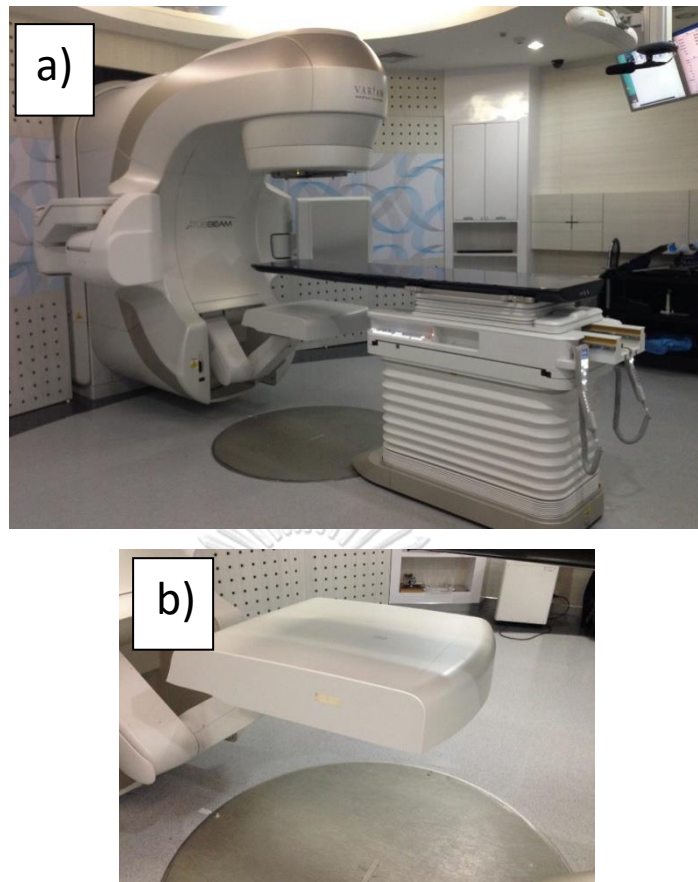
Figure 29 a) TrueBeam linacs (Varian Medical System, Palo Alto, CA, USA), and b) EPID (Portal Vision, Varian Medical System, Palo Alto, CA, USA)

Figure 30 Delta4 PT (ScandiDos AB, Uppsala, Sweden)

4.2.2 MATLAB software

The MATLAB software program (The Math Works Inc., Natick, MA, USA) was used for EPID images model, log file analysis, generating errors in VMAT plans, extracting beam complexity matrices, and model-based prediction.

4.2.3 QA methods

Three QA methods were used to perform patient-specific QA in this study; phantom-based measurement with Delta4, EPID-based dosimetry, Log file analysis.

4.2.3.1 Delta4

Our study performed measurement-based with the phantom method using Delta4 (Fig. 4.4). Delta4 was calibrated follow a manual guide [109] before measurement of patient-specific VMAT QA. The calibration was included as relative, absolute, and directional calibration.

4.2.3.2 Integrated mode for EPID-based dosimetry

In this study, in-house software for EPID-based dosimetry was used to reconstruct the flat phantom's dose. EPID images were acquired with integrated mode. For integrated mode, EPID captures a single image consisting of the average number of frames acquired during radiation delivery. Then the delivery fluence map was compared with the planned fluence map. Dose agreement was analyzed with the gamma passing rates for each arc. For 2D dose distribution, EPID images were converted to dose distribution in flat water. Figure 4.5 shows the dose model calculation in this study. The measured EPID image was converted to a fluence map in

the air by deconvolution EPID image with air kernel in the first step. In the second step, the air fluence map was converted to dose distribution in the water at depth 10 cm by convolution in air fluence map with dose-to-water-kernel.
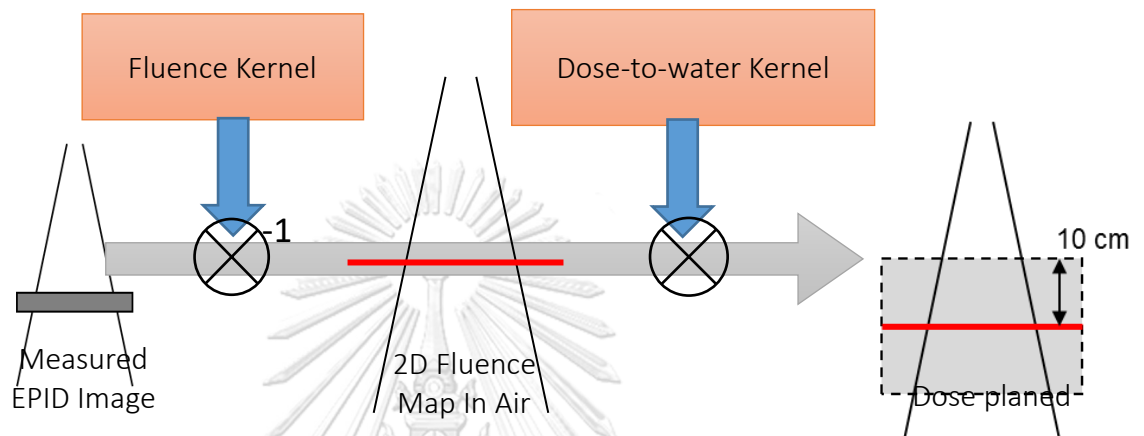
Figure 31 Flowchart of model calculation for EPID-based dosimetry

4.2.3.3 Log file

Log file of patient-specific VMAT QA from linacs was exported to MATLAB programs. Log file data included MU, MLC position, dose rate, and gantry angle, which were related to each control point, were used to calculate the delivery fluence map. Gamma passing rates (GPR) were used to analyze patient-specific VMAT QA by comparing the delivery fluence map and the planned fluence map.

4.3 Research methods

This research was divided into two steps according to the objectives. For the first step, the sensitivity of error detection was evaluated to determine error detection baselines in three QA tools (EPID-based dosimetry, Delta4, and Log file). The highest error detection QA tool was

selected for model-based prediction in patient-specific QA results. For the second step, model-based prediction for VMAT patient-specific QA was developed, and then model accuracy was assessed using sensitivity and specificity scores.

4.3.1 Method of patient-specific VMAT QA sensitivity and accuracy

4.3.1.1 Plan population

Figure 4.6 shows the flowchart of this study. Five head and neck VMAT plans were randomly selected for this study. Patient-specific VMAT QA was performed using three QA tools (Delta4, EPID-based dosimetry, and log file) with five original plans. To create an error plan, the intention errors were introduced to the original plan. Table 4.1 shows the error simulation for this study. These errors were simulated into the three possible scenarios: uncertainty, miss-calibration, and worst case. MLC's random errors at $\pm$ 2.0 mm and gantry angle at $\pm$ 2.0 degrees were introduced to the original plan for the uncertainty scenario. For the miss-calibration scenario, the systematic error of +2 MU and random MLC errors at $\pm$ 2.0 mm and gantry angle at $\pm$ 2.0 degrees were introduced to the original plan. For worst-case scenarios, errors were integrated between systematic and random variation of MLC and gantry angle at 2$\pm$0.5 mm and 2$\pm$0.5 degrees, respectively.
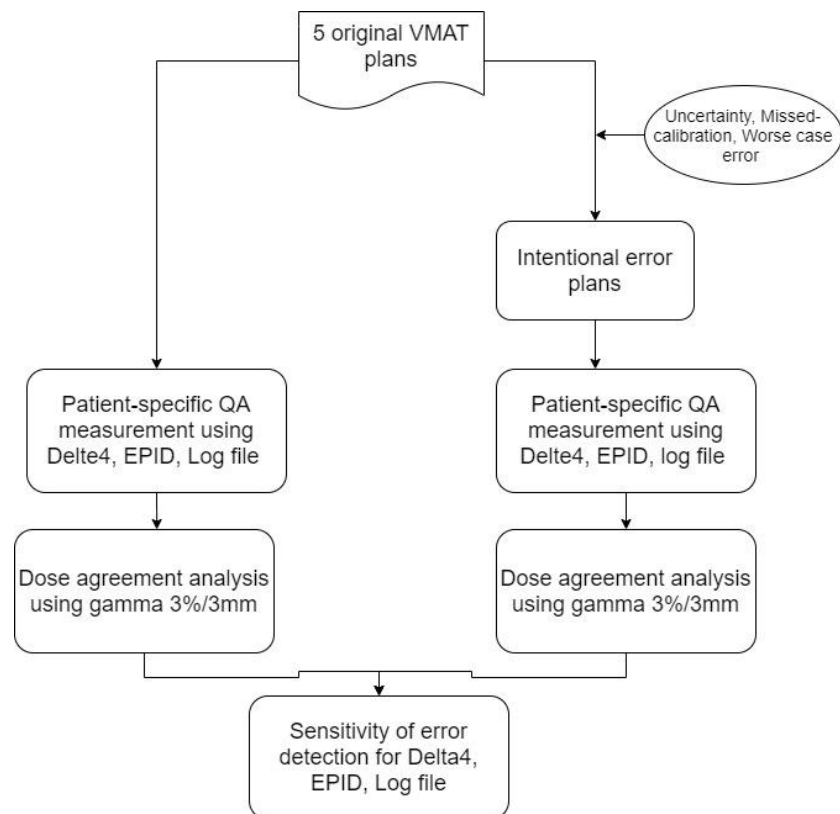
Figure 32 Flowchart of error detection sensitivity

Table 12 Error simulation in a different type of scenario

Scenario types	Error parameters
	Uncertainty
	MLC (random) + Gantry angle (random)
	Miss-calibration
Combination error (realistic error scenarios)	MU (systematic) + MLC (random) + Gantry angle (random)
	Worst case (max.)
	MU (systematic + random) + MLC (systematic + random) + Gantry angle (systematic + random)

4.3.1.2 Patient-specific QA measurement

Delta4 PT

Delta4 PT (ScandiDos AB, Uppsala, Sweden) was calibrated before following a manual guide before the VMAT patient-specific QA performance. Delta4 PT contains 1,069 p-type diodes placed at cross-plane inserted inside a cylindrical phantom made from polymethylmethacrylate (PMMA). To compare dose distribution between measurement and TPS calculation, plan for patient-specific QA were transferred to Delta4 phantom (Cylinder phantom), and then calculate dose distribution in the phantom. For the measurement process, a phantom position error was corrected by Delta4 PT software to reduce the error from the phantom setup. Figure 4.7 shows the phantom setup for patient-specific QA measurement. Three-dimension dose distributions were calculated to other planes with the interpolation method.

Figure 33 Delta4 phantom setup for measurement

EPID-based dosimetry

Plan for patient-specific QA measurement was transferred to flat phantom (30×30×40 cm: width×height×length), and then the dose distribution was calculated by using AAA algorithm form Eclipse TPS. Amorphous silicon (a-Si) 1000 EPID was used to acquire an image for VMAT patient-specific QA with a spatial resolution of 1024×768 pixels. EPID images were converted to dose at ten cm-depth in water using our model, as explained in session 4.2.1.2.

Log file

Dynalog [110] script code was used to extract log file to the fluence map. Log file data contains MU information, MLC position, dose rate, and gantry angle, which are related to each control point. When patient-specific QA measurement was finished, the treatment log file was recorded in the treatment console workstation. The log file was exported to MATLAB software for extracting data, and then the data was reconstructed to fluence

map using Dynalog script code. The dose agreement was determined by comparing between delivery fluence map and the planned fluence map.

4.3.1.3 Data analysis

Dose agreement analysis

After patient-specific VMAT QA was performed, the dose agreement between TPS calculation and measurement was analyzed using gamma analysis with gamma criteria of 3%, 3 mm, and a cut-off threshold at 10%.

Error detection assessment

Results of patient-specific QA for original and introduced error plans were evaluated as GPR scores. The sensitivity of error detection was determined using the ratio between GPR of original and GPR of introduced error plan, as shown in equation 4.1. The higher value in equation 4.1 was related to a higher sensitivity to detect the error.

$$\text{Error detection sensitivity} = \frac{\text{GPR}_{\text{original}}}{\text{GPR}_{\text{error}}} \qquad \text{Equation 4.1}$$

Where GPR $_{\text{original}}$ was GPR from original plan, and GPR $_{\text{error}}$ was GPR form introduced error plan.

4.3.2 Method of model-based patient-specific VMAT QA

In this study, EPID-based dosimetry was emphasized for patient-specific VMAT QA to benefit from easy to use, high spatial resolution, eliminating potential setup errors. Hence, the patient-specific VMAT QA

results of EPID-based dosimetry were used for the model-based patient-specific VMAT QA.

4.3.2.1 Clinical data collection

Fifty-three VMAT plans (a total of 168 arcs) of the head and neck were randomly collected over a time period of 2018 – 2019 from a single center. All plans were treated with a 6 MV photon. Table 4.2 shows the plan information. The simultaneous intergraded boost (SIB) technique was used with different dose prescriptions according to the staging and type of tumor. Dose prescription can be categorized in six prescription protocols. The first protocol was prescribed with 70 Gy (2.12 Gy × 33 fractions) for a gross tumor, 59.4 Gy (1.8 Gy × 33 fractions) for high-risk nodes, and 54 Gy (1.64 Gy × 33 fractions) for low-risk nodes. The second protocol was prescribed with 70 Gy (2.12 Gy × 33 fractions) for a gross tumor and 59.4 Gy (1.8 Gy × 33 fractions) for high-risk nodes. The third protocol was prescribed with 70 Gy (2.12 Gy × 33 fractions) for a gross tumor, 66 Gy (2 Gy × 33 fractions) for high-risk nodes, 59.4 Gy (1.8 Gy × 33 fractions) for intermediate-risk nodes, and 54 Gy (1.64 Gy × 33 fractions) for low-risk nodes. The fourth protocol was prescribed with 70 Gy (2 Gy × 35 fractions) for a gross tumor and 59.5 Gy (1.7 Gy × 35 fractions) to high-risk nodes. The fifth protocol was prescribed with 70 Gy (2 Gy × 35 fractions) for a gross tumor and 63 Gy (1.8 Gy × 35 fractions) for high-risk nodes. The sixth protocol was prescribed with 56 Gy (2 Gy × 28 fractions) for a gross tumor. The range of arc numbers for the VMAT plan was 2 to 4 arcs per plan. The dose constraints used to evaluate the plan are shown in Table 4.3.

Table 13 Summarization of the randomly selected VMAT head and neck plans used

Tumor region (n = plan)
Nasopharynx (n = 42)
Supraglottic (n = 3)
Floor of mouth (n = 1)
Tongue (n = 1)
Base of tongue (n = 1)
Neck nodes (n = 1)
Tonsil (n = 1)
Glottis and thyroid (n = 1)
Thyroid (n = 1)
Buccal (n = 1)

Dose prescription (n = plan)
First protocol (n = 37)
Second protocol (n = 9)
Third protocol (n = 4)
Fourth protocol (n = 1)
Fifth protocol (n = 1)
Sixth protocol (n = 1)
Total (n = 53)

Table 14 Dose constraints used for the plan evaluation

Target/Organs	Dose constraint
Target	
CTV	$D_{100} > 100\%$ prescribed dose
PTV	$D_{95} > 100\%$ prescribed dose
	Max. $<110\%$ prescribed dose
Organs	
Temporal lobe	Max. < 60 Gy, $V_{65} < 1\%$
Brain stem	Max. < 54 Gy, $V_{60} < 1\%$
Eye	Mean < 35 Gy
Lens	Max. < 10 Gy
Optic nerves	Max. < 54 Gy, $V_{60} < 1\%$
Optic chiasm	Max. < 54 Gy, $V_{60} < 1\%$
Parotid grands	Mean < 26 Gy (at least 1 grand), $D_{50\%} < 30$ Gy
Spinal cord	Max. < 45 Gy, $V_{50} < 1$ cc
Mandible and TM joint	Max. < 70 Gy, $V_{75} < 1$ cc
Oral cavity	Mean < 40 Gy (excluding PTV)
Cochlear and IAC	Mean < 45 Gy
Glottic larynx	Mean < 45 Gy
Mucosa	Max. < 30 Gy
Constrictor muscle	Max. < 50 Gy
Brain	Max. < 60 Gy
Pharynx	Mean < 50 Gy

4.3.2.2 Patient-specific VMAT QA using EPID-base dosimetry

Patient-specific VMAT QA was performed for all plans prior to treatment using an EPID dosimetry technique. All selected plans were delivered with a TrueBeam linear accelerator (Varian Medical Systems, Palo Alto, CA). The planned 2D dose distribution at the EPID level (Source–imager distance = 100 cm) was generated using Portal Dose Image Prediction (PDIP) from the Eclipse treatment planning system (TPS) (Version 13.6, Varian Medical Systems, Palo Alto, CA). The beam delivery data as an integrated EPID image per arc was collected from aSi-1000 EPID, which had a spatial resolution of 1024×768 pixels with a pixel size spacing of 0.39 mm. These measured EPID images were then used to compare the planned EPID images generated from PDIP in TPS using the 2D gamma evaluation method. In this study, gamma evaluation was determined using two sets of criteria based on recommendations of AAPM Task Group No. 218 [43]. The first criteria was 2%/2 mm with a 10% global aperture threshold for detecting subtle regional errors, and the second criteria was 3%/2 mm with a 10% global aperture threshold for general purpose. The outcome of this comparison was GPR.

4.3.2.3 Overall process and feature extraction

Figure 4.8 demonstrates the process used in this study. There were 53 VMAT plans of head and neck randomly retrieved. Eighty percent of the total plans (134 arcs) were used as a training data set and the remaining 20% of the total plans (34 arcs) were used as the testing data set. The features were extracted from MLC patterns in the VMAT plan including leaf speed, leaf acceleration, and texture analysis. These features were set as the predictor, and the GPR related to the plan was set as the response.

The input from extracted features and corresponding GPR were used for generating the predictive model. The evaluation method was performed to determine the accuracy of the prediction model using a testing data set. All processes were implemented in MATLAB software version 2019b (The Mathworks, Inc, Natick, MA) with machine learning Toolbox 11.6 (classification learner and regression learner application).
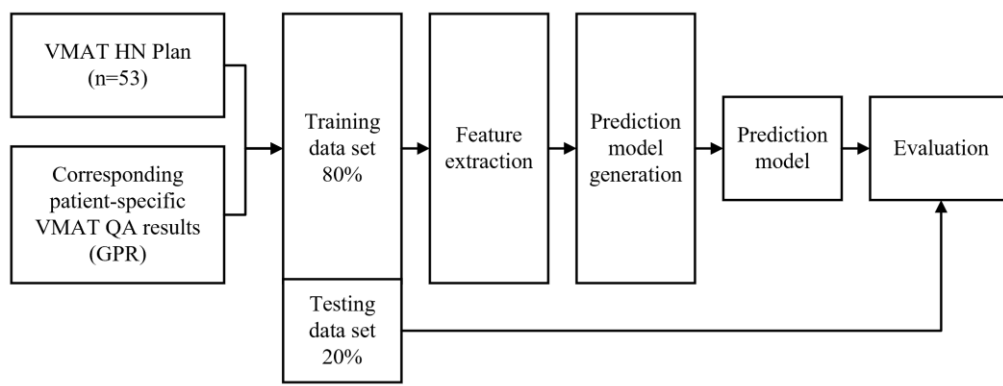
Figure 34 Flowchart diagram for model-based prediction

Table 4.4 shows the list of features used for training the model that can be classified into two main groups: 1) features of leaf speeds and accelerations and 2) texture analysis from the fluence map. Leaf speed and acceleration were calculated according to the study of Park et al. [13]. Leaf speed and acceleration can be calculated using equation 4.2 and 4.3, respectively.

$$Leaf\ speed_i = \frac{Leaf_i + Leaf_{i+1}}{Time_i} \qquad\qquad \text{Equation 4.2}$$

$Leaf_i$ is the position of the leaf at the ith CP, and $Time_i$ is the time between the ith CP and $(i+1)^{th}$ CP.

$$Leaf\ accelleration_i = \frac{Leaf\ speed_i + Leaf speed_{i+1}}{Time_i}$$ Equation 4.3

For texture analysis, Park's method [62] was implemented to generate a gray-level co-occurrence matrix (GLCM). The MLC data and MU for each control point were reconstructed using the integrated intensity fluence map by accumulating all control point fluence maps, then GLCM was generated for each arc. In this study, the particular displacement distances (d) of 1 pixel and four angles (θ) of 0°, 45°, 90°, and 135° were used to calculate the GLCM. The extracted features of the texture analysis were contrast, correlation, energy, entropy, and homogeneity.

Table 15 Summarization of the features used for model-based prediction and classification

Leaf speed and acceleration (Bank A and B)
1) Maximum (max.) of leaf speed
2) Mean of leaf speed
3) Standard deviation (sd.) of leaf speed
4) Fraction of leaf speed for 0 to 4 mm/s
5) Fraction of leaf speed for 4 to 8 mm/s
6) Fraction of leaf speed for 8 to 12 mm/s
7) Fraction of leaf speed for 12 to 16 mm/s
8) Fraction of leaf speed for 16 to 20 mm/s
9) Maximum (max.) of leaf acceleration
10) Mean of leaf acceleration
11) Standard deviation (sd.) of leaf acceleration
12) Fraction of leaf acceleration for 0 to 4 mm/s^2
13) Fraction of leaf acceleration for 4 to 8 mm/s^2
14) Fraction of leaf acceleration for 8 to 12 mm/s^2

15) Fraction of leaf acceleration for 12 to 16 mm/s^2

16) Fraction of leaf acceleration for 16 to 20 mm/s^2

Texture analysis

17) Contrast

18) Correlation

19) Energy

20) Entropy

21) Homogeneity

4.3.2.4 Predictive model generation Classification and regression

The training model dataset included 21 features (16 features from leaf speed and acceleration parameters and 5 features from texture analysis parameters) that were used as the predictors. The class of GPRs was used as the responses. The model was separated into two different gamma criteria; 2%/2 mm with a 10% threshold and 3%/2 mm with a 10% threshold. Cross-validation with five folds was used to protect the overfitting effect.

Per the AAPM Task Group 218 recommendation, the action levels of the patient-specific QA results were set to 90% GPR for the 2%/2 mm criteria and 95% GPR for the 3%/2 mm criteria [111]. However, in our clinical experience, most of the patient-specific VMAT QA in the head and neck seldom exceeded the recommended action level. Therefore, we set the institutional tolerance level for patient-specific VMAT QA in the head and neck to 93.7% and 97.0% for the 2%/2 mm and 3%/2 mm criteria, respectively. The GPRs in the training data set were grouped into two

classes, pass, and warning stages, based on the tolerance level. To classify GPR into the pass and warning stages, the pass stages were labeled as 1, and the warning stages were labeled as 0. For the regression model, the initial result was a GPR value. However, the GPR value was converted into an institutional tolerance level to classify the QA results as pass or warning stages.

The population of the pass and warning stages in the training data set can be summarized as follows: for the gamma criteria of 2%/2 mm with a 10% threshold, the pass, and warning stages were 63.43% (85 arcs) and 36.57% (49 arcs), respectively; for the gamma criteria of 3%/2 mm with a 10% threshold, the pass, and warning stages were 68.66% (92 arcs) and 31.34% (42 arcs), respectively.

4.3.2.5 Model selection

In this study, the classification and regression methods were used to generate the model. Different models were included to train the data, and the highest accuracy was selected to verify accuracy in the testing dataset. For the classification method, the different models were decision trees, support vector machine (SVM), discriminant analysis, logistic regression, Naive Bayes, nearest neighbor, and ensemble classification. For the regression method, the different models were linear regression, regression trees, SVM, Gaussian process, and ensemble of trees.

4.3.2.6 Model accuracy

In this study, model accuracy was investigated using the testing data set. The population of the pass and warning stage in the testing data set can summarize as follows: for the gamma criteria of 2%/2 mm with a 10%

threshold, the pass and warning stages were 52.94% (18 arcs) and 47.06% (16 arcs), respectively; for the gamma criteria of 3%/2 mm with a 10% threshold, the pass and fail tolerance level were 50% (17 arcs) and 50% (17 arcs), respectively.

The model accuracy was investigated in terms of sensitivity and specificity. The sensitivity and specificity scores were calculated as shown in equation 4.4 and 4.5, respectively. Agreement of the fail tolerance level between the prediction and QA results were identified as a true positive (TP), while the disagreement was a false positive (FP). Agreement of the pass tolerance level between the prediction and QA results were identified as true negative (TN), while disagreement was a false negative (FN).

$$Sensitivity\ (\%) = \frac{TP}{(TP+FN)} \times 100\% \qquad \text{Equation 4.4}$$

$$Specitivity\ (\%) = \frac{TN}{(TN+FP)} \times 100\% \qquad \text{Equation 4.5}$$

The sensitivity represents the probability of the model to detect a failure tolerance level. Alternatively, the specificity represents the probability of the model to detect a pass tolerance level. The sensitivity and specificity were also determined for the regression model; however, the model results were predicted as the GPR value. Hence, GPR was classified as a pass or warning stage before calculation sensitivity and specificity using the institutional tolerance level, as explained in the predictive model generation session. In addition, the paired t-test was used to determine the accuracy of the GPR prediction in the regression model by comparing the

prediction and known QA results. In this study, $p > 0.05$ was considered statistically significant.

CHAPTER 5

RESULTS

5.1 Results of patient-specific VMAT QA sensitivity and accuracy

Table 5.1 shows the average GPR (gamma criteria of 3%, 3mm) ± 1SD for all scenarios. EPID has the lowest %GPR with 84.94±2.21, 82.05±1.53, 77.51±1.26 %GPR for uncertainties, miss-calibration, worst-case scenario, respectively. Log file has the intermediate %GPR with 92.42±0.56, 89.59±0.67, 86.87±1.21 %GPR for uncertainties, miss-calibration, worst-case scenario, respectively. Delta4 has the highest %GPR with 97.53±1.11, 94.12±0.78, 90.71±0.93 %GPR for uncertainties, miss-calibration, worst-case scenario, respectively. Figure 35 shows error detection sensitivity in terms of the GPR ratio between the original and introduced error plan for all scenarios. The results in figure 5.1 indicated that EPID has the highest sensitivity to detect the error with 9.2%, 12.09%, 16.63% GPR drop in uncertainty, miss-calibration, and worst-case scenarios, respectively. The log file has intermediate sensitivity to detect error with 4.34%, 7.17%, 9.89% GPR drop in the same order. Finally, Delta4 has the lowest sensitivity on error detection with 1.8%, 5.21%, 8.62% GPR drop.

Table 16 Percentage of average gamma pass-rate ±1SD (3%, 3mm) for all scenarios

Scenario	Delta4	EPID	Log file
Original plans	99.33±0.58	94.14±1.82	96.76±0.87
Uncertainties	97.53±1.11	84.94±2.21	92.42±0.56
Miss-calibration	94.12±0.78	82.05±1.53	89.59±0.67
Worst case	90.71±0.93	77.51±1.26	86.87±1.21
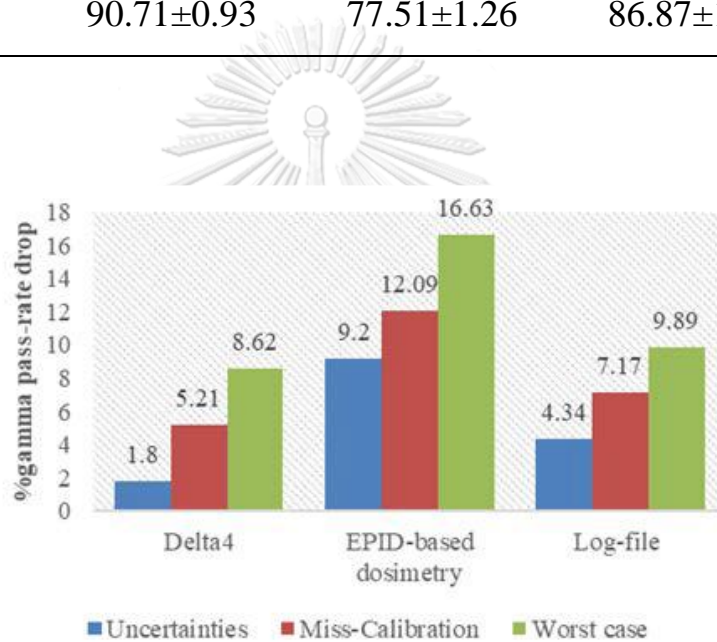
Figure 35 GPR drop from original plan for all scenarios

Figure 5.2 shows the example of patient-specific VMAT QA results in three QA systems (log file, EPID-based dosimetry, and Delta4) for miss-calibration scenario.

Log file map   Gamma   Planned fluence

(b) EPID-based dosimetry

(c) Delta4

Meas. VS. TPS   Gamma evaluation
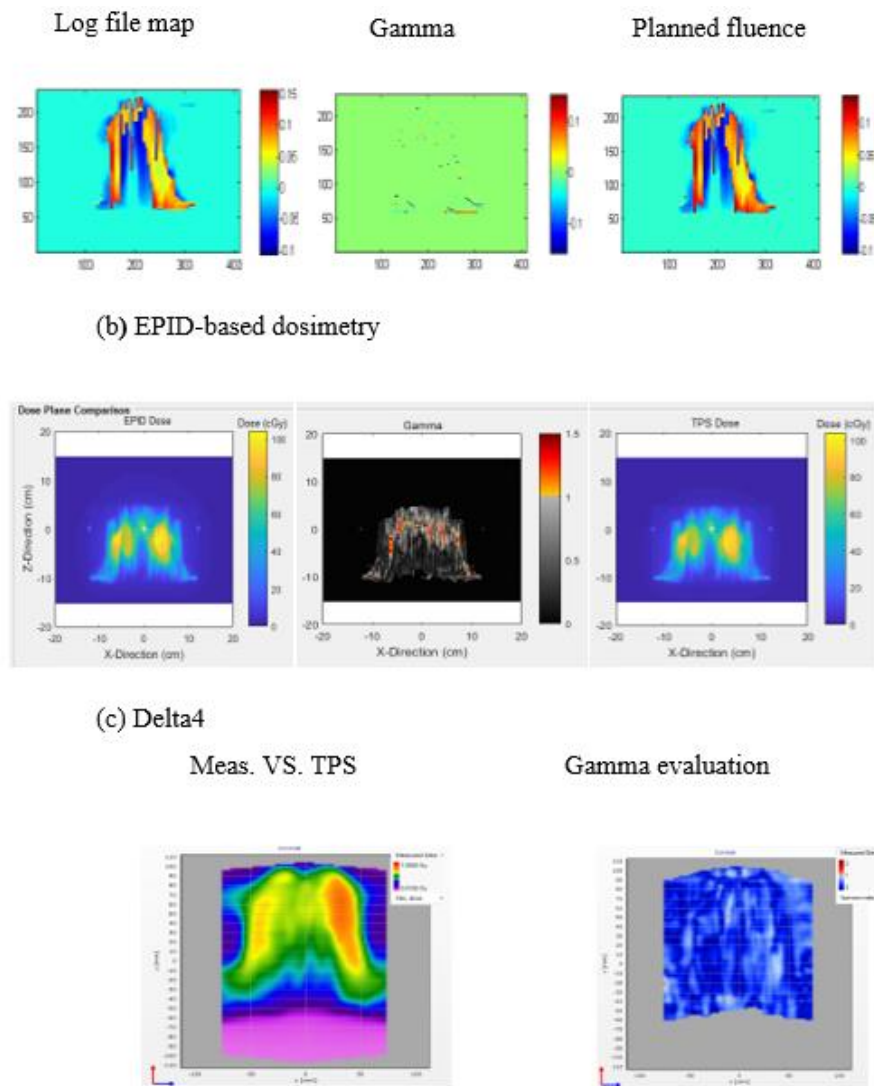
Figure 36 Example of patient-specific VMAT QA results in three QA systems a) Log file, b) EPID-based dosimetry, and c) Delta4 for miss-calibration scenario

5.2 Results of model-based prediction for VMAT patient-specific QA

5.2.1 Model selection

Table 5.2 shows the accuracy of various model algorithms. For gamma criteria 3%/2mm of the classification model, the highest accuracy was found at nearest neighbor with70.9% accuracy, and the lowest accuracy was found at discriminant analysis with 65.7% accuracy. For

gamma criteria 3%/2mm of the regression model, the highest accuracy was found at ensemble of tree with 2.1867 RMS, and the lowest accuracy was found at linear regression with 2.5164 RMS. For gamma criteria 2%/2mm of the classification model, the highest accuracy was found at nearest neighbor with 69.4% accuracy, and the lowest accuracy was found at discriminant analysis and Naive Bayes with 59.7% accuracy. For gamma criteria 2%/2mm of the regression model, the highest accuracy was found at Gaussian process with 2.8276 RMS, and the lowest accuracy was found at linear regression with 3.337 RMS.

Table 17 Model accuracy result of various model algorithms

Gamma criteria	Classification model (% accuracy)	Regression model (Root Mean Square error: RMS)
3%/2mm	Decision trees = 66.4% SVM = 69.4% Discriminant analysis = 65.7% Logistic regression = 67.2% Naive Bayes = 66.4% Nearest neighbor = 70.9 % Ensemble classification = 68.7%	Linear regression = 2.5164 RMS Regression trees = 2.2672 RMS SVM = 2.2434 RMS Gaussian process = 2.2911 RMS Ensemble of trees = 2.1867 RMS
2%/2mm	Decision trees = 63.4 % SVM = 67.2% Discriminant analysis = 59.7 % Logistic regression	Linear regression = 3.337 RMS Regression trees = 2.8688 RMS SVM = 2.9141 RMS

	= 61.9 % Naive Bayes = 59.7 %	Gaussian process = <u>2.8276</u> RMS
	Nearest neighbor = <u>69.4 %</u>	Ensemble of trees = 2.8743
	Ensemble classification = 64.9 %	RMS

Figure 5.3 shows the confusion matrix of the neighbor classifier model in the gamma criteria.

a) Gamma criteria of 2%/2 mm
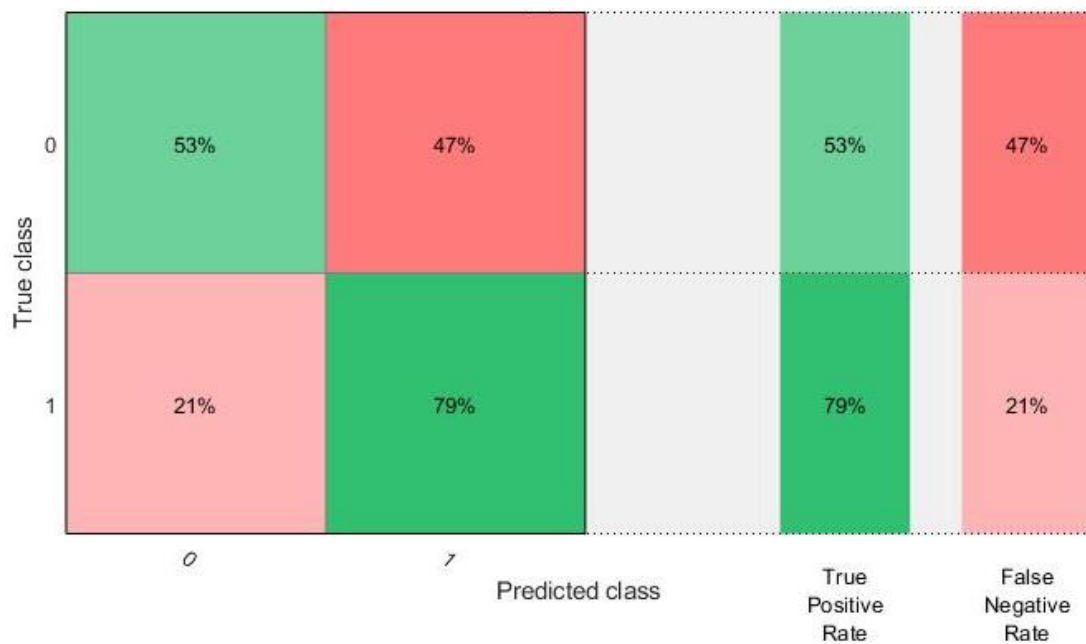
b) Gamma criteria of 3%/2 mm
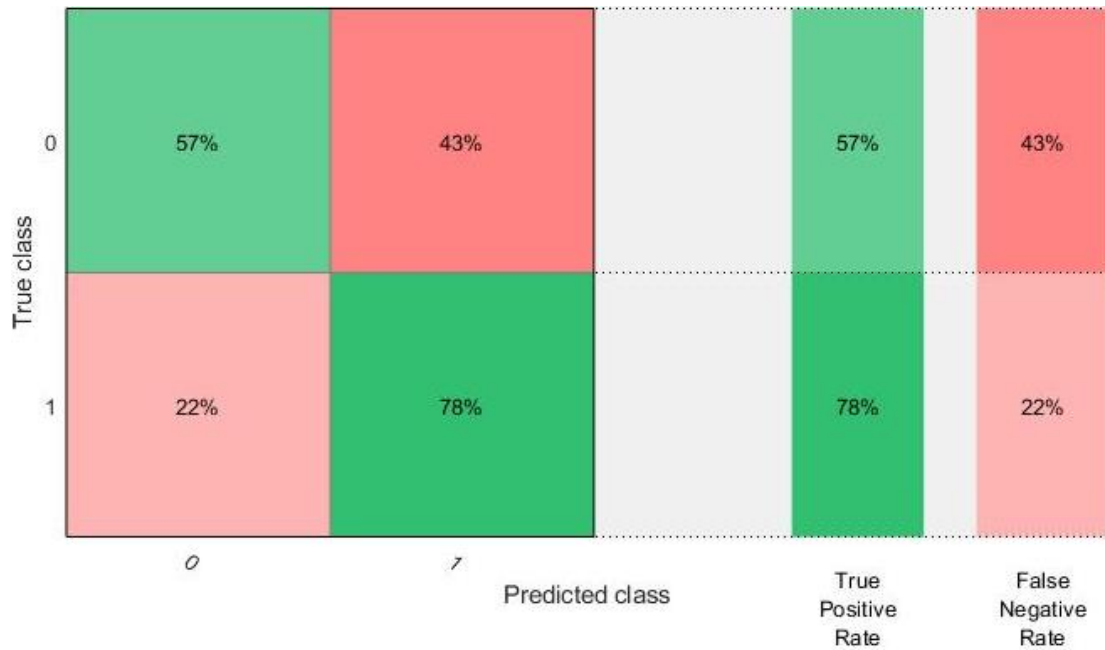
Figure 37 Confusion matrix of neighbor classifier model for gamma criteria of a) 2%/2 mm and b) 3%/2 mm

Unlike the classification model, the accuracy of the regression model shows a relation between the prediction and true response, and between the residual error and true response. Figure 5.4 shows the relationship between the prediction and true response, and a relation between the residual error and true response of a) the Gaussian process model for the gamma criteria of 2%/2 mm and b) ensemble of trees model for the gamma criteria of 3%/3 mm.

a) Gamma criteria of 2%/2 mm

b) Gamma criteria of 3%/2 mm
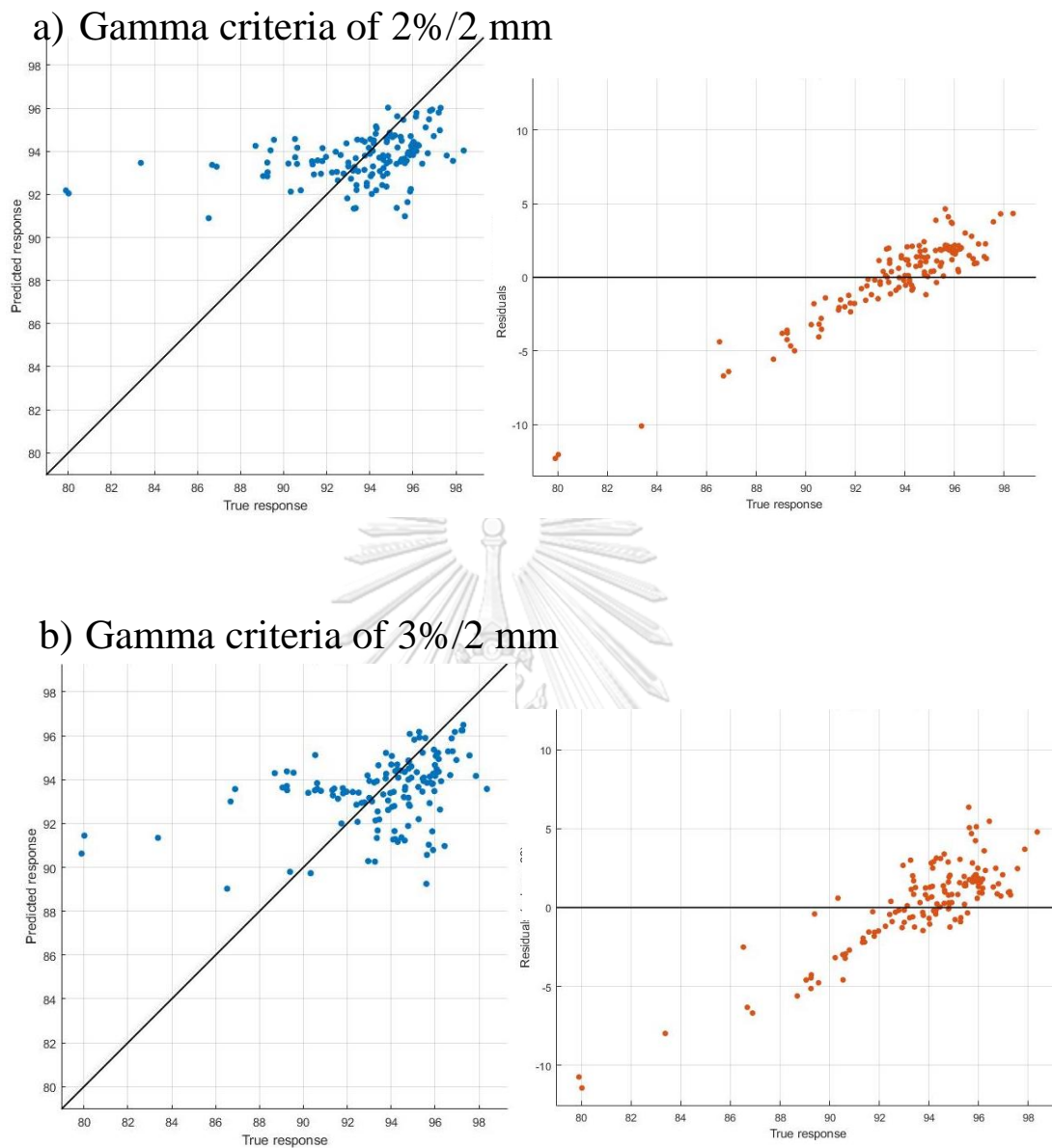
Figure 38 Relation between the prediction and true response, and between the residual error and true response of a) the Gaussian process model for the gamma criteria of 2%/2 mm and b) ensemble of trees model for the gamma criteria of 3%/3 mm

5.2.2 Model accuracy

Table 5.2 summarizes the results of the sensitivity and specificity of the testing data set for the classification and regression models with different gamma criteria (2%/2 mm with a 10% threshold and 3%/2 mm

with a 10% threshold). For gamma criteria of 2%/2 mm with a 10% threshold, the highest sensitivity of 81.25% was observed with a regression model, while the highest specificity of 66.67% was observed with a classification model. For gamma criteria of 3%/2 mm with a 10% threshold, the highest sensitivity of 76.47% was observed with the regression model, while the highest specificity of 64.71% was observed with the classification model. A paired t-test was used to determine the statistical difference of the GPR prediction with the QA results. There was no significant difference for the gamma criteria of 2%/2 mm with a 10% threshold (P-value = 0.1784) or the gamma criteria of 3%/2 mm with a 10% threshold (P-value = 0.761). Figure 5.5 shows the GPR prediction error results for the gamma criteria sets of 2%/2 mm and 3%/2 mm.

Table 18 Summarization of the sensitivity and specificity in the testing data set for classification and regression models with two gamma criteria (2%/2 mm with a 10% threshold, and 3%/2 mm with a 10% threshold)

Model	Classification		Regression	
	Sensitivity	Specificity	Sensitivity	Specificity
	(TP/Fail	(TN/Pass	(TP/Fail	(TN/Pass
Gamma	number)	number)	number)	number)
criteria				
2%/2 mm with a 10% threshold	68.75% (11/16)	66.67% (12/18)	81.25% (13/16)	44.44% (8/18)
3%/2 mm with a 10% threshold	64.71% (11/17)	64.71% (11/17)	76.47% (13/17)	58.82% (10/17)
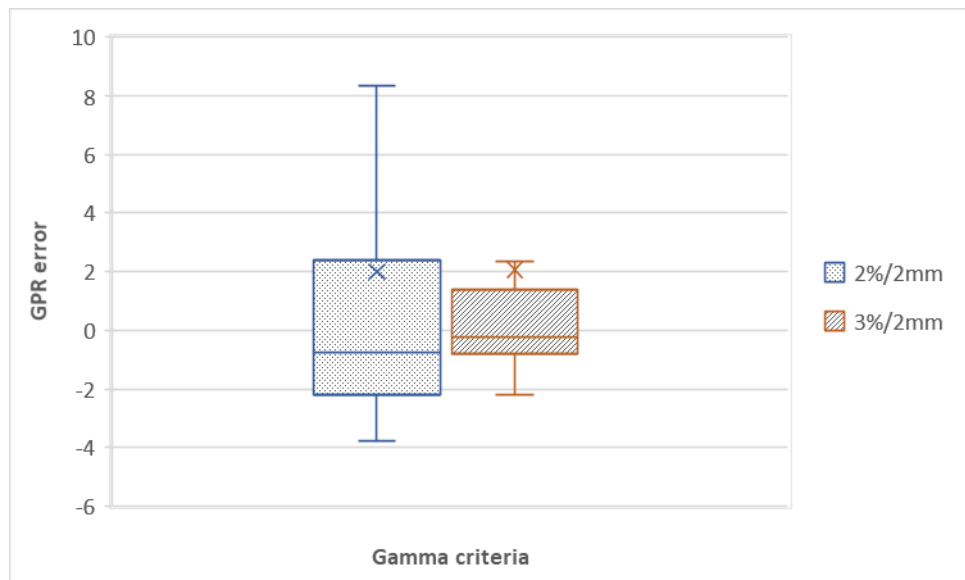
Figure 39 Gamma prediction error of the testing data set using the regression model

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

CHAPTER 6

DICUSSION AND CONCLUSIONS

6.1 Discussion of patient-specific VMAT QA sensitivity and accuracy

Uncertainty scenario has a small effect on error detection when compared to the miss-calibration scenario. This may imply that random errors have less effect than systematic errors. The worst-case scenario showed the most considerable effect on error detection because random and systematic errors were integrated. Woon et al. [112] reported that the sensitivity to detect errors depends on detectors' spatial resolution. According to the high spatial resolution of EPID (1024×768 pixels), EPID-based dosimetry showed higher sensitivity to detect error than Delta4 (1,069 diodes). Zhang et al. [113] reported that 3D analyses are sensitive to detect errors less than the 2D plane. In this study, EPID-based dosimetry and log file were analyzed with a 2D plane, whereas Delta4 was analyzed with 3D interpolation. In low spatial resolution detectors, AAPM task group no.218 [111] recommended that error detection sensitivity be increased by reducing gamma criteria.

As the AAPM task group no.218 report, the method for patient-specific QA measurement can affect GPR, categorized into three methods; true composite, perpendicular field-by-field perpendicular composite. For true composite method, the measurement is performed using the beam direction similar to the actual treatment plan, and the QA device is placed at a fixed position, such as the phantom-based measurement method. The true composite method can provide actual dose simulated closely to the treatment plan, but the method cannot sample every part of the beam. For perpendicular field-by-field measurement, the measurement is performed

using the beam direction similar to the actual treatment plan or setting to gantry 0 degree, and the QA device is rotated perpendicular with beam such as EPID measurement method. In addition, the measurement is performed field-by-field separately. The perpendicular field-by-field measurement can sample every part of the field, but the summation of 3D dose distribution is not available in this method. For perpendicular composite measurement, the measurement is performed using the beam direction similar to the actual treatment plan or setting to gantry 0 degree, and the QA device is rotated perpendicular with the beam. Finally, the measurement was summed together. The perpendicular composite measurement can sample every field and fast acquisition, but the regional error is not available to explore in this method. We note that in this study, the EPID-based dosimetry and the log file was performed with perpendicular field-by-field methods, whereas Delta4 measurement was performed with true composite method.

6.2 Conclusion of patient-specific VMAT QA sensitivity and accuracy

From the investigation of three patient-specific QA systems in this work, we can conclude that EPID-based dosimetry was the most sensitive QA tool to detect the three possible scenarios (uncertainty, miss-calibration, and worst-case scenario) in patient-specific VMAT QA. Log file was the second-best method, whereas Delta4 was the worst method to detect the three possible scenario errors.

6.3 Discussion of model-based prediction for VMAT patient-specific QA

For patient-specific QA measurements, the out-of-control GPR values are removed from the training and testing data set to reduce error in the model using statistical process control [40].

In this study, the feature selection technique was tried before with the filter method of a correlation coefficient, but the results seem similar to the all feature used technique. The feature selection for this study was briefly explained here. The correlation between features and GPR was determined using Spearman's correlation, and the heat map correlation (as shown in Fig. 6.1) was also used to determine which feature correlated to each other. A total of 11 features were selected as the variable that included a leaf acceleration bank A fraction 0-40 mm/s^2, leaf acceleration bank B fraction 0-40 mm/s^2, leaf speed bank B fraction 0-4 mm/s, leaf acceleration bank A fraction 160-200 mm/s^2, leaf acceleration bank B fraction 160-200 mm/s^2, leaf speed bank A fraction 0-4 mm/s, leaf speed bank A fraction 8-12 mm/s, leaf speed bank A fraction 8-12 mm/s, leaf speed bank B fraction 8-12 mm/s, leaf speed bank A fraction 12 -16 mm/s, leaf speed bank A fraction 16-20 mm/s. The model accuracy was found 69.4% accuracy for classification and 2.9 RMS for regression model. Hence, the feature selection was not used for training in this study.
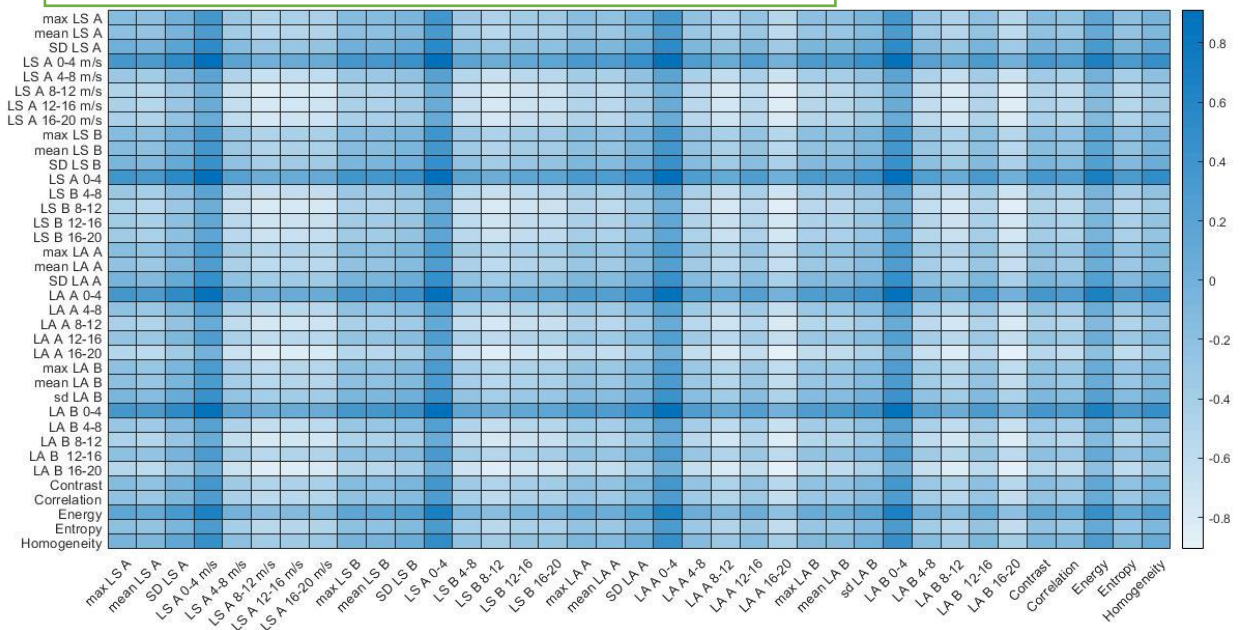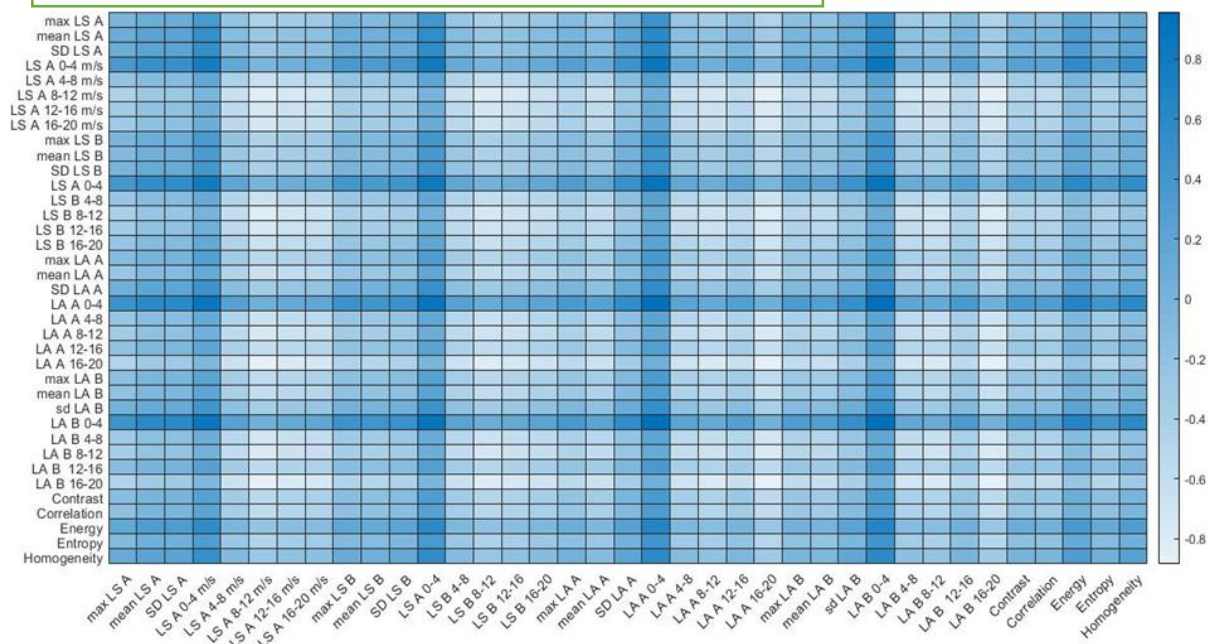
Figure 40 Heat map of Spearman's correlation of feature used for classification and regression model, a) gamma criteria of 2%/2mm with 10% threshold, b) gamma criteria of 3%/2mm with 10% threshold

The overfitting effect typically occurred when small data used to train the model. Cross-validation is an important method to protect against the overfitting effect. Hence, cross-validation was performed in our training model as explained as follows; the fold number was initially from 1 to 10 folds, and the model accuracy of different folds was recorded; the highest model accuracy was selected to train the model. However, 100% accuracy for classification model or 0 RMS for regression model was recognized as the overfitting effect, which was not selected for training in this study. The five-fold cross-validation was found the highest accuracy for our training model.

For the classification model, the fail and pass tolerance levels influence the model accuracy as described by Valdes et al. [108] and Tomori et al. [103]. A large number of fail tolerance levels (TP) can increase the model accuracy to predict the failure tolerance level. Because of the small data size in this study, the universal tolerance level recommended by the AAPM Task Group No. 218 [111] cannot be used to classify the pass and fail tolerance levels. Therefore, the model accuracy was improved using the institution tolerance for classifying pass and warning stages to collect more population in warning stages. Our study was compared with Jiaqi et al. [107] that had a large population. Table 6.1 shows a comparison between our study and Jiaqi's study. They observed the highest sensitivity using the Random Forest (RF) classification model with 66.67% sensitivity (4/6) by setting the action limit to 90% GPR for the gamma criteria of 3%/2 mm, and 100% sensitivity (5/5) by setting the action limit to 80% GPR for the gamma criteria of 2%/2 mm. The highest specificity was observed in the Poisson Lasso (PS) regression model with 100% specificity (42/42) for the gamma criteria of 3%/2 mm, and 43/43

for the gamma criteria of 2%/2 mm. Unlike in our study, the highest sensitivity of 76.47% (13/17) was observed in the regression model by setting warning stages at 97.01% GPR for the gamma criteria of 3%/2 mm, and 81.25% sensitivity (13/16) by setting warning stages at 93.68% GPR for the gamma criteria of 2%/2 mm. The highest specificity of 64.71% (11/17) was observed with the classification model for the gamma criteria of 3%/2 mm, and 66.67% specificity (11/17) for the gamma criteria of 2%/2 mm. This difference could be contributed by various factors such as the size of the pass and warning stage population in the training and testing data sets, the difference in models used for training (we used the nearest neighbor classifier for classification, and Gaussian process and assemble trees model for regression), and the settings of GPR for classification.

Table 19 Comparison between our study and Jiaqi's study

Parameter		our study	Jiaqi's study
Sensitivity accuracy	Gamma criteria 3%/2 mm	found at regression model with 76.47% (13/17) using 97.01% GPR tolerance	found at RF classification model with 66.67% (4/6) using 90.0% GPR tolerance
	Gamma criteria 2%/2 mm	found at regression model 81.25% (13/16) using 93.68% GPR tolerance	found at RF classification model with 100% (5/5) using 80.0% GPR tolerance
Specificity accuracy	Gamma criteria 3%/2 mm	Found at classification model with 64.71% (11/17)	Found at PS regression model with 100%

	using 97.01% GPR tolerance	(42/42) using 90.0% GPR tolerance
Gamma criteria 2%/2 mm	Found at classification model with 66.67% (11/17) using 93.68% GPR tolerance	Found at classification model with 100% (43/43) using 80.0% GPR tolerance

Kruse et al. [114] indicated that the sensitivity to detect error can be further explored using tighter criteria. Similar to this study, the gamma criteria of 2%/2 mm has greater sensitivity than the gamma criteria of 3%/2 mm as demonstrated by the 68.75% and 64.71% sensitivity, respectively, in the classification model and 81.25% and 76.47% in the regression model.

The advantage of the predictive model in patient-specific QA is it aids medical physicists in evaluating the risk of the plan fails QA. If there is a risk, the medical physicist can re-optimize the plan by changing a plan parameter such as the MLC speed and acceleration and intensity fluence variance (contrast, correlation, energy, entropy, and homogeneity in texture analysis). Additionally, the predictive model can reduce the iterative process, i.e., re-measurement if a failed QA result, which can decrease the change in patient scheduling delay as reported by Abolaban et al. [115]. Additionally, the predictive model can be implemented in the online-adaptive radiation therapy workflow because the patient-specific QA process cannot be performed before delivering the beam to the patient due to the time limitations. In online adaptive radiation, only independent

dose calculations can be performed before treatment, and other solutions, such as the transit dose measurement [115-117] and log-file analysis [118], can be used to monitor the dose delivery during treatment. There is no additional process to predict the risk of failed QA results for online adaptive radiation therapy; therefore, the predictive model can solve this problem. Figure 6.2 illustrates how the predictive model of patient-specific QA can be implemented as an online-adaptive radiation therapy workflow. First, the adaptive plan is optimized by controlling the key parameters (feature prediction) to reduce the risk of failed QA results. Second, after optimization, the features are extracted to predict GPR. Third, if the GPR value has a pass tolerance level, the patient will be treated with an adaptive plan: if the GPR value is at the fail tolerance level, the adaptive plan will be investigated to determine the cause of error.

As reported by Vial et al. [119], disagreement between the EPID measurement and PDIP can be caused by various variables such as differences between the profile correction at the EPID calibration and PDIP, and if the change in the beam spectrum from MLC attenuation does not consider the PDIP. Therefore, this error may affect the accuracy of the model. Vlades et al. [106] also validated a machine learning approach for predicting GPR using different QA devices, particularly diode-array detectors, and portal dosimetry, and determined that the accuracy of the prediction model at diode-array detectors was greater than the accuracy at the portal dosimetry (3% compared with 3.5% accuracy) because the portal dosimetry had large disagreements in the low-dose regions. In this study, the EPID measurement was performed using an integrated mode with rotated gantry during measurement to consider the MLC error from the sagging effect, and the IsoCal systems on the Varian TrueBeam were used

to correct imager sag during gantry rotation as previously described by Gao et al. [120].
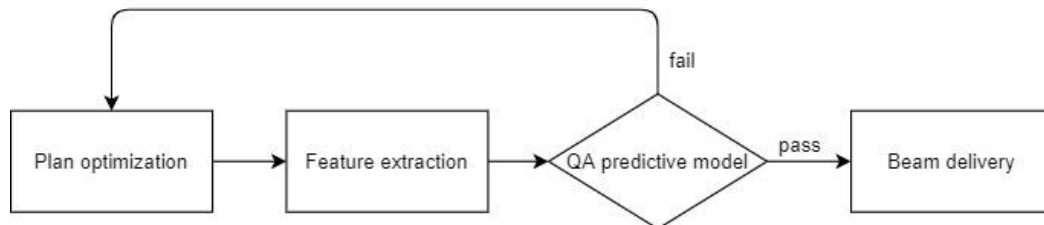
Figure 41Implementation of the patient-specific VMAT QA predictive model for on-line adaptive therapy

Leaf speed and acceleration were collected as the predictor for this study because a previous study by Park et al. [66] showed a strong correlation of leaf speed and acceleration to GPR with a range between -0.458 to -0.511 for leaf speed fraction, and -0.225 to 0.477 for leaf acceleration fraction. Additional features collected in this study were texture analysis parameters because another study by Park et al. [121] showed a strong correlation in the range of -0.475 and 0.213. Only displacement distances (d) of 1 pixel were used in this study because a previous study [121] showed the best correlation at 1 pixel. Our study, the correlation between feature and GPR was also determined using Spearman's correlation, and the result related Park's study with correlation range between -0.2257 to 0.4558 for leaf speed fraction, -0.1244 to 0.4785 for leaf acceleration fraction, and -0.0084 to 0.2331 for texture analysis (as shown in Table 6.2).

Table 20 Results of Spearman's correlation in different features

Features	Spearman Correlation	
	Gamma 2%/2mm	Gamma 3%/2mm
max. leaf speed bank A	-0.0720	-0.0230
mean leaf speed bank A	-0.1384	0.1139
SD. leaf speed bank A	0.0869	0.1139
Leaf speed bank A fraction 0-4 mm/s	<u>0.4430</u>	0.3828
Leaf speed bank A fraction 4-8 mm/s	-0.2527	-0.2257
Leaf speed bank A fraction 8-12 mm/s	-0.4149	<u>-0.4205</u>
Leaf speed bank A fraction 12 -16 mm/s	-0.3719	-0.3320
Leaf speed bank A fraction 16-20 mm/s	-0.3629	-0.3031
max. leaf speed bank B	-0.0721	-0.0231
mean leaf speed bank B	-0.0853	-0.0560
SD. leaf speed bank B	0.0121	0.0044
Leaf speed bank B fraction 0-4 mm/s	<u>0.4558</u>	<u>0.4156</u>
Leaf speed bank B fraction 4-8 mm/s	-0.2637	-0.2359
Leaf speed bank B fraction 8-12 mm/s	-0.4048	-0.3804
Leaf speed bank B fraction 12 -16 mm/s	-0.3074	-0.3251
Leaf speed bank B fraction 16-20 mm/s	-0.2870	-0.2490
max. leaf acceleration bank A	-0.1198	-0.1108

mean leaf acceleration bank A	-0.1758	-0.2405
SD. leaf acceleration bank A	0.0223	0.0704
Leaf acceleration bank A fraction 0-40 mm/s^2	0.4326	<u>0.4785</u>
Leaf acceleration bank A fraction 40-80 mm/s^2	-0.1763	-0.2505
Leaf acceleration bank A fraction 80-120 mm/s^2	-0.3643	-0.3297
Leaf acceleration bank A fraction 120-160 mm/s^2	-0.1823	-0.1896
Leaf acceleration bank A fraction 160-200 mm/s^2	<u>-0.4516</u>	<u>-0.4409</u>
max. leaf acceleration bank B	-0.1509	-0.1498
mean leaf acceleration bank B	-0.1458	-0.1920
SD. leaf acceleration bank B	-0.0070	0.0368
Leaf acceleration bank B fraction 0-40 mm/s^2	0.4350	<u>0.4708</u>
Leaf acceleration bank B fraction 40-80 mm/s^2	-0.2237	-0.2947
Leaf acceleration bank B fraction 80-120 mm/s^2	-0.3811	-0.3491
Leaf acceleration bank B fraction 120-160 mm/s^2	-0.1291	-0.1244
Leaf acceleration bank B fraction 160-200 mm/s^2	<u>-0.4530</u>	-0.4423
Contrast	-0.0884	-0.1184
Correlation	-0.1721	-0.1557

Energy	<u>0.2331</u>	<u>0.1846</u>
Entropy	-0.1428	-0.0595
Homogeneity	0.0333	0.1198

6.4 Conclusion of model-based prediction for VMAT patient-specific QA

The developed model's feasibility to predict patient-specific QA of head and neck VMAT plans was demonstrated based on the MLC effect and texture analysis using a machine learning approach. The accuracy of the model was validated using the testing data set, and the highest sensitivity of 81.25% was observed with the regression model at the gamma criteria of 2%/2 mm, and the highest specificity of 66.67% was observed with the classification model at the gamma criteria of 2%/2 mm. The prediction model can help medical physicists determine the risk that a plan could fail the tolerance level before re-measurement of the QA. Future studies will include an implementation method to control beam complexity scores in the optimization and the dose calculation process to reduce the risk of failed QA results.

6.5 Overall discussion

In radiation therapy, patient-specific QA is the important process to verify agreement between dose calculation from TPS and dose delivery from LINAC machine. With drawback of patient-specific QA [106, 107, 115, 122] such as time consuming, labour tasks, re-measurement process, many researchers have developed method to solve this problem. A few years ago, the study of model-based prediction for patient-specific QA was increased, and the research demonstrated the

possibility of model-based can help medical physicist to determine risk of fail QA results [103, 105, 106, 108]. Moreover, the model-based prediction has possibility to implement in online-adaptive radiation therapy workflows because the patient-specific QA process cannot be performed before delivery the beam to patient due to time limitation. In online-adaptive radiation, only independent dose calculation can be performed before treatment. The other solution can monitor dose delivery during treatment, such as transit dose measurement [116, 117] and log-file analysis [118]. There is no additional process to predict the risk of failed QA results for online adaptive radiation therapy; hence, the predictive model has the feasibility to solve this problem.

6.6 Future work

We have planned to implement our model-based prediction to radiation therapy clinical workflow for reducing the re-measurement process in case of fail QA results. In the future study, we will implement the solution to control plan parameter features in TPS to reduce the risk of fail QA results. Figure 6.3 shows the process of future work. Briefly explain here, the prediction features were calculated during plan optimization and minimize beam complexity score by setting as the one constrains parameter used to reduce the risk of fail QA results.
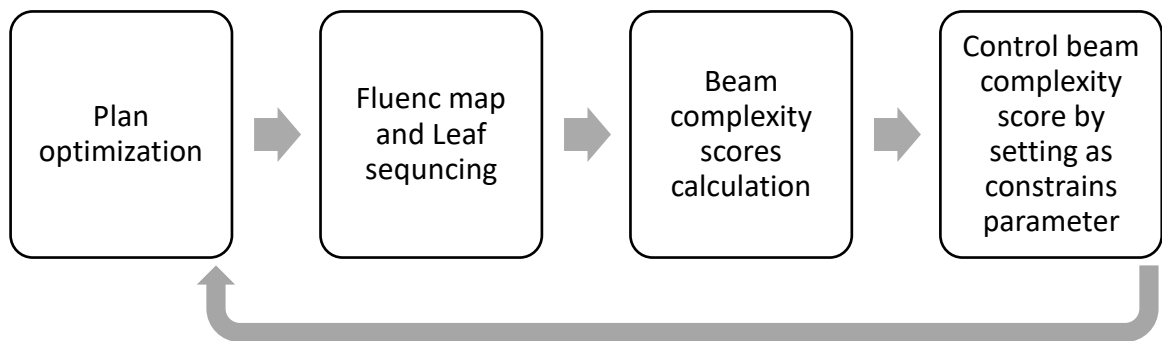
Figure 42 Process of complexity score control during plan optimization

6.7 Overall conclusions

In this study, we have verified the capability to detect errors of three QA tools (EPID-based dosimetry, Delta4, Log file), and the highest error detection QA tool was found at EPID-based dosimetry, which was selected to use for model-based prediction. We also have demonstrated the feasibility of using model-based prediction for patient-specific VMAT QA with ML approach. This study indicated the virtual patient-specific VMAT QA (model-based prediction) using ML approach has a possibility to use for determining the risk that a plan could fail the tolerance level before re-measurement of the QA. Moreover, the virtual patient-specific QA also has a possibility to replace the traditional patient-specific QA by increasing the population in the training model, which was improved the model accuracy. The typical process of predictive model was explained in Figure 6.4. Here is a briefly explained the typical process; the extracted features corresponded to GPR were collected for the training model; for classification model, the tolerance level was labeled first for classifying between pass and fail tolerance; the data was divided into two sets, first set for training (typically 70-80% of total data) and second set or testing data set for verifying the model accuracy (20-30% of total data); the model was

firstly trained with the quick-all algorithm to find the highest accuracy, and fine-tune the model; the highest accuracy model was used to predict in the testing data set; the model accuracy was assessed by comparing between predicted QA results and the truth QA results. The model accuracy for predicting patient-specific QA results depended on two main parameters, the QA process, and the training process. The QA process parameter included QA tool type, the method to perform QA, gamma criteria, the machine output fluctuation, and the other subprocess that involved a patient-specific QA process. For example, the spatial resolution related to the QA tool type could affect the capability to detect the errors. This may influence the model accuracy in terms of the correlation between complexity metrics and QA results. In this case, the model accuracy can improve by decreasing the gamma criteria to explore more errors. EPID-based dosimetry has inherent errors, such as a disagreement between predicted portal dosimetry and EPID measurement, the accumulated dose-effect reducing the signal sensitivity. The disagreement between predicted portal dosimetry and EPID measurement issue can eliminate by taking into account Varian's preconfigured portal dosimetry package (PDPC) which applied 2D profile correction to the predicted portal dosimetry algorithm. However, our model was not imported PDPC into the predicted portal dosimetry algorithm because our TPS version is not supported PDPC. Hence, we suggest that the model accuracy can improve by taking into account PDPC if TPS version is available for PDPC. The accumulated dose effect of EPID can eliminate by calibrating the dose every day before measurement patient-specific QA. The QA results for our model were collected from the previous EPID-based dosimetry QA with monthly dose calibration. To eliminate the accumulated dose-effect and machine output

fluctuation, we suggest performing the daily dose calibration before measurement patient-specific QA which can improve the model accuracy. For the training process, the model accuracy depended on each step, such as the pre-processing step, training step, validation step, and testing step. For the pre-processing step, the data need to be normalized, reduced the dimension before the training step. The training algorithms is a key success for training the model. Many new algorithms were developed to improve model accuracy, such as AdaBoost, XGBoost (eXtreme Gradient Boosting). If the advanced algorithm is available, we suggested implementing that for the training model. Each new algorithm was explained as follows. AdaBoost is the iteration learning of the decision tree, which corrects the weak learner from the previous and improves the accuracy by reducing the large errors. XGBoost is the class of gradient boost regression tree method, which combines the weak decision tree using gradient descent optimization.
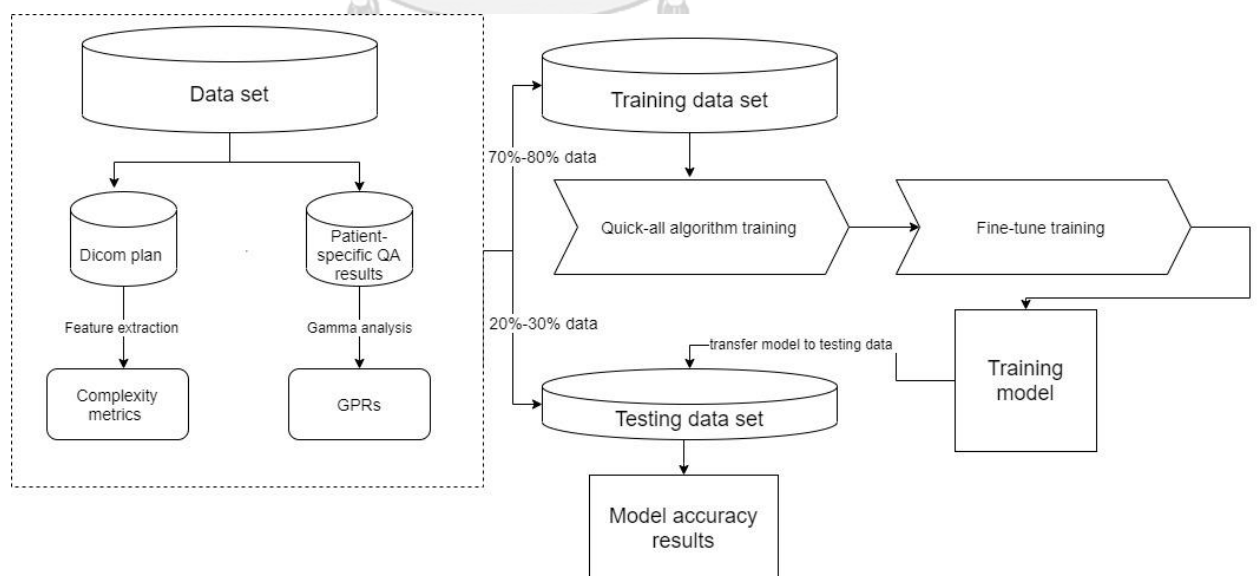
Figure 43 Typical process of patient-specific QA predictive model

REFERENCES

1. International Atomic Energy Agency (IAEA). Basic Safety Standards. IAEA. 1994, Vienna :IAEA.

2. International Commission on Radiological Protection (ICRP) Publication 60. Recommendation of the International Commission on Radiological Protection, in Annals of the ICRP. 1991, Oxford: Pergamon Press.

3. Coeytaux, K., et al., Reported radiation overexposure accidents worldwide, 1980-2013: a systematic review. PLoS One, 2015. **10**(3): p. e0118709.

4. Borras, C., Overexposure of radiation therapy patients in Panama: problem recognition and follow-up measures. Rev Panam Salud Publica, 2006. **20**(2-3): p. 173-87.

5. International Atomic Energy Agency, Vienna (Austria) (2000). Lessons learned from accidental exposures in radiotherapy. International Atomic Energy Agency (IAEA): IAEA. 2000, Vienna: IAEA.

6. Ortiz Lopez, P., et al., ICRP publication 112. A report of preventing accidental exposures from new external beam radiation therapy technologies. Ann ICRP, 2009. **39**(4): p. 1-86.

7. Fraass, B., et al., American Association of Physicists in Medicine Radiation Therapy Committee Task Group 53: quality assurance for clinical radiotherapy treatment planning. Med Phys, 1998. **25**(10): p. 1773-829.

8. Kubo, H.D., et al., High dose-rate brachytherapy treatment delivery: report of the AAPM Radiation Therapy Committee Task Group No. 59. Med Phys, 1998. **25**(4): p. 375-403.

9. Kutcher, G.J., et al., Comprehensive QA for radiation oncology: report of AAPM Radiation Therapy Committee Task Group 40. Med Phys, 1994. **21**(4): p. 581-618.

10. Nath, R., et al., Intravascular brachytherapy physics: report of the AAPM Radiation Therapy Committee Task Group no. 60. American Association of Physicists in Medicine. Med Phys, 1999. **26**(2): p. 119-52.

11. Nath, R., et al., Code of practice for brachytherapy physics: report of the AAPM Radiation Therapy Committee Task Group No. 56. American Association of Physicists in Medicine. Med Phys, 1997. **24**(10): p. 1557-98.

12. Rivard, M.J., et al., Update of AAPM Task Group No. 43 Report: A revised AAPM protocol for brachytherapy dose calculations. Med Phys, 2004. **31**(3): p. 633-74.

13. Yu, Y., et al., Permanent prostate seed implant brachytherapy: report of the American Association of Physicists in Medicine Task Group No. 64. Med Phys, 1999. **26**(10): p. 2054-76.

14. Thwaites, D., et al., Quality assurance in radiotherapy. European Society for Therapeutic Radiology and Oncology Advisory Report to the Commission of the European Union for the 'Europe Against Cancer Programme'. Radiother Oncol, 1995. **35**(1): p. 61-73.

15. Ishikura, S., Quality Assurance of Radiotherapy in Cancer Treatment: Toward Improvement of Patient Safety and Quality of Care. Japanese Journal of Clinical Oncology, 2008. **38**(11): p. 723-729.

16. Ramm, U., et al., Feasibility study of patient positioning verification in electron beam radiotherapy with an electronic portal imaging device (EPID). Physica Medica, 2014. **30**(2): p. 215-220.

17. Monville, M.E., Z. Kuncic, and P.B. Greer, Simulation of real-time EPID images during IMRT using Monte-Carlo. Phys Med, 2014. **30**(3): p. 326-30.

18. Richart, J., et al., QA of dynamic MLC based on EPID portal dosimetry. Phys Med, 2012. **28**(3): p. 262-8.

19. Van Elmpt, W., et al., A literature review of electronic portal imaging for radiotherapy dosimetry. Radiother Oncol, 2008. **88**(3): p. 289-309.

20. Van Esch, A., T. Depuydt, and D.P. Huyskens, The use of an aSi-based EPID for routine absolute dosimetric pre-treatment verification of dynamic IMRT fields. Radiother Oncol, 2004. **71**(2): p. 223-34.

21. Bakhtiari, M., et al., Using an EPID for patient-specific VMAT quality assurance. Med Phys, 2011. **38**(3): p. 1366-73.

22. Boutry, C., et al., Technical Note: A simple algorithm to convert EPID gray values into absorbed dose to water without prior knowledge. Med Phys, 2017.

23. Gustafsson, H., et al., Direct dose to water dosimetry for pretreatment IMRT verification using a modified EPID. Med Phys, 2011. **38**(11): p. 6257-64.

24. Liu, B., et al., A novel technique for VMAT QA with EPID in cine mode on a Varian TrueBeam linac. Phys Med Biol, 2013. **58**(19): p. 6683-700.

25. Agnew, A., et al., Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. Physics in Medicine & Biology, 2014. **59**(9): p. N49.

26. Agnew, C.E., D.M. Irvine, and C.K. McGarry, Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT. J Appl Clin Med Phys, 2014. **15**(6): p. 4994.

27. Aydogan, B., Smith, B. and Li, J., DynaLog File Analysis for IMRT Delivery Verification. Med. Phys., 2010. **37**: p. 3228–3229.

28. Defoor, D.L., et al., Investigation of error detection capabilities of phantom, EPID and MLC log file based IMRT QA methods. J Appl Clin Med Phys, 2017. **18**(4): p. 172-179.

29. Dinesh Kumar, M., et al., QA of intensity-modulated beams using dynamic MLC log files. Journal of Medical Physics / Association of Medical Physicists of India, 2006. **31**(1): p. 36-41.

30. McGarry, C.K., et al., The use of log file analysis within VMAT audits. Br J Radiol, 2016. **89**(1062): p. 20150489.

31. Okamoto, H., et al., Log-file analysis of accuracy of beam localization for brain tumor treatment by CyberKnife. Practical Radiation Oncology. **6**(6): p. e361-e367.

32. Rangaraj, D., et al., Catching errors with patient-specific pretreatment machine log file analysis. Practical Radiation Oncology. **3**(2): p. 80-90.

33. Zaila, A., M. Adili, and S. Bamajboor, Pylinac: A toolkit for performing TG-142 QA related tasks on linear accelerator. Physica Medica, 2016. **32**(Supplement 3): p. 292-293.

34. Van Dam, J., G. Leunens, and A. Dutreix, Correlation between temperature and dose rate dependence of semiconductor response; influence of accumulated dose. Radiotherapy and Oncology. **19**(4): p. 345-351.

35. Harms, W.B., Sr., et al., A software tool for the quantitative evaluation of 3D dose calculation algorithms. Med Phys, 1998. **25**(10): p. 1830-6.

36. Van Dyk, J., et al., Commissioning and quality assurance of treatment planning computers. Int J Radiat Oncol Biol Phys, 1993. **26**(2): p. 261-73.

37. Low, D.A., et al., A technique for the quantitative evaluation of dose distributions. Med Phys, 1998. **25**(5): p. 656-61.

38. Carlone, M., et al., ROC analysis in patient specific quality assurance. Med Phys, 2013. **40**(4): p. 042103.

39. Hotelling, H., Journal of the American Statistical Association, 1932. **27**(178): p. 215-217.

40. Pawlicki, T., M. Whitaker, and A.L. Boyer, Statistical process control for radiotherapy quality assurance. Med Phys, 2005. **32**(9): p. 2777-86.

41. Rah, J.-E., et al., Feasibility study of using statistical process control to customized quality assurance in proton therapy. Medical Physics, 2014. **41**(9): p. 092105-n/a.

42. Sanghangthum, T., et al., Statistical process control analysis for patient-specific IMRT and VMAT QA. Journal of Radiation Research, 2013. **54**(3): p. 546-552.

43. Wheeler, D. and D. Chambers, Understanding Statistical Process Control. 1992, Knoxville: SPC Press.

44. Kang, J., et al., Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. Int J Radiat Oncol Biol Phys, 2015. **93**(5): p. 1127-35.

45. Samuel, A.L., Some studies in machine learning using the game of checkers. II—Recent progress. Annual Review in Automatic Programming, 1969. **6**(Part 1): p. 1-36.

46. Hierons, R., Machine learning. Tom M. Mitchell. Published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997. ISBN: 0-07-115467-1, 414 pages. Price: U.K. £22.99, soft cover. Software Testing, Verification and Reliability, 1999. **9**(3): p. 191-193.

47. Alpaydin, E., Introduction to Machine Learning. 2010: The MIT Press. 584.

48. Bibault, J.E., P. Giraud, and A. Burgun, Big Data and machine learning in radiation oncology: State of the art and future prospects. Cancer Lett, 2016. **382**(1): p. 110-117.

49. Park, J.M., et al., The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT. The British Journal of Radiology, 2015. **88**(1049): p. 20140698.

50. Rajasekaran, D., et al., A study on the correlation between plan complexity and gamma index analysis in patient specific quality assurance of volumetric modulated arc therapy. Reports of Practical Oncology and Radiotherapy, 2015. **20**(1): p. 57-65.

51. Chiavassa, S., et al., Complexity metrics for IMRT and VMAT plans: A review of current literature and applications. The British Journal of Radiology, 2019. **92**: p. 20190270.

52. Webb, S., Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal

unmodulated radiotherapy. Phys Med Biol, 2003. **48**(14): p. 2051-62.

53.   McNiven, A.L., M.B. Sharpe, and T.G. Purdie, A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys, 2010. **37**(2): p. 505-15.

54.   Nauta, M., J.E. Villarreal-Barajas, and M. Tambasco, Fractal analysis for assessing the level of modulation of IMRT fields. Medical Physics, 2011. **38**(10): p. 5385-5393.

55.   Mathot, M. and D. Dechambre, 6 VMAT complexity metrics can reduce patient QA workload. Physica Medica, 2018. **56**: p. 3-4.

56.   Shen, L., et al., Multidimensional correlation among plan complexity, quality and deliverability parameters for volumetric-modulated arc therapy using canonical correlation analysis. Journal of Radiation Research, 2018. **59**(2): p. 207-215.

57.   Mohan, R., et al., The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy. Medical Physics, 2000. **27**(6): p. 1226-1237.

58.   Masi, L., et al., Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. Med Phys, 2013. **40**(7): p. 071718.

59.   Du, W., et al., Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. Medical Physics, 2014. **41**(2): p. 021716.

60.   Crowe, S.B., et al., Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results. Australasian Physical & Engineering Sciences in Medicine, 2014. **37**(3): p. 475-482.

61.   Park, J.M., et al., The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT. The British Journal of Radiology, 2015. **88**(1049): p. 20140698.

62.   Park, S.Y., et al., Texture analysis on the fluence map to evaluate the degree of modulation for volumetric modulated arc therapy. Med Phys, 2014. **41**(11): p. 111718.

63.   Llacer, J., T.D. Solberg, and C. Promberger, Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning. Physics in Medicine and Biology, 2001. **46**(10): p. 2637-2663.

64.   Giorgia, N., et al., What is an acceptably smoothed fluence? Dosimetric and delivery considerations for dynamic sliding window IMRT. Radiation Oncology, 2007. **2**(1): p. 42.

65. Coselmon, M.M., et al., Improving IMRT delivery efficiency using intensity limits during inverse planning. Med Phys, 2005. **32**(5): p. 1234-45.

66. Park, J.M., et al., The effect of MLC speed and acceleration on the plan delivery accuracy of VMAT. The British journal of radiology, 2015. **88**(1049): p. 20140698-20140698.

67. Park, J.M., et al., Modulation indices for volumetric modulated arc therapy. Physics in Medicine and Biology, 2014. **59**(23): p. 7315-7340.

68. Li, R. and L. Xing, An adaptive planning strategy for station parameter optimized radiation therapy (SPORT): Segmentally boosted VMAT. Medical Physics, 2013. **40**(5): p. 050701.

69. Kairn, T., et al., Predicting the likelihood of QA failure using treatment plan accuracy metrics. Journal of Physics: Conference Series, 2014. **489**: p. 012051.

70. Heijmen, B., et al., Fully automated, multi-criterial planning for Volumetric Modulated Arc Therapy – An international multi-center validation for prostate cancer. Radiotherapy and Oncology, 2018. **128**(2): p. 343-348.

71. Carlsson, F., Combining segment generation with direct step-and-shoot optimization in intensity-modulated radiation therapy. Med Phys, 2008. **35**(9): p. 3828-38.

72. Götstedt, J., A. Karlsson Hauer, and A. Bäck, Development and evaluation of aperture-based complexity metrics using film and EPID measurements of static MLC openings. Med Phys, 2015. **42**(7): p. 3911-21.

73. Younge, K.C., et al., Penalization of aperture complexity in inversely planned volumetric modulated arc therapy. Med Phys, 2012. **39**(11): p. 7160-70.

74. Mathot, M. and D. Dechambre, 6 VMAT complexity metrics can reduce patient QA workload. Physica Medica: European Journal of Medical Physics, 2018. **56**: p. 3-4.

75. Park, J.M., S.-Y. Park, and H. Kim, Modulation index for VMAT considering both mechanical and dose calculation uncertainties. Physics in Medicine and Biology, 2015. **60**(18): p. 7101-7125.

76. Nilsson, G., SU-FF-T-135: Delta4 — A New IMRT QA Device. Medical Physics, 2007. **34**(6Part9): p. 2432-2432.

77. Sadagopan, R., et al., Characterisation, commissioning and evaluation of DELTA4 IMRT QA system. Vol. 34. 2007.

78. Feygelman, V., et al., Evaluation of a biplanar diode array dosimeter for quality assurance of step-and-shoot IMRT. J Appl Clin Med Phys, 2009. **10**(4): p. 3080.

79. LoSasso, T., C.S. Chui, and C.C. Ling, Comprehensive quality assurance for the delivery of intensity modulated radiotherapy with a multileaf collimator used in the dynamic mode. Med Phys, 2001. **28**(11): p. 2209-19.

80. Xia, P., C.F. Chuang, and L.J. Verhey, Communication and sampling rate limitations in IMRT delivery with a dynamic multileaf collimator system. Med Phys, 2002. **29**(3): p. 412-23.

81. Cadman, P., et al., Dosimetric considerations for validation of a sequential IMRT process with a commercial treatment planning system. Vol. 47. 2002. 3001-10.

82. Rangel, A. and P. Dunscombe, Tolerances on MLC leaf position accuracy for IMRT delivery with a dynamic MLC. Med Phys, 2009. **36**(7): p. 3304-9.

83. Budgell, G.J., et al., Requirements for leaf position accuracy for dynamic multileaf collimation. Phys Med Biol, 2000. **45**(5): p. 1211-27.

84. Justus, A. and W. Qiuwen, Independent verification of gantry angle for pre-treatment VMAT QA using EPID. Physics in Medicine & Biology, 2012. **57**(20): p. 6587.

85. McCurdy, B.M. and P.B. Greer, Dosimetric properties of an amorphous-silicon EPID used in continuous acquisition mode for application to dynamic and arc IMRT. Med Phys, 2009. **36**(7): p. 3028-39.

86. Fuangrod, T., et al., A cine-EPID based method for jaw detection and quality assurance for tracking jaw in IMRT/VMAT treatments. Phys Med, 2015. **31**(1): p. 16-24.

87. Losasso, T., IMRT delivery performance with a varian multileaf collimator. Int J Radiat Oncol Biol Phys, 2008. **71**(1 Suppl): p. S85-8.

88. Olasolo-Alonso, J., et al., Evaluation of MLC performance in VMAT and dynamic IMRT by log file analysis. Vol. 33. 2017.

89. Sun, B., et al., Evaluation of the efficiency and effectiveness of independent dose calculation followed by machine log file analysis against conventional measurement based IMRT QA. J Appl Clin Med Phys, 2012. **13**(5): p. 3837.

90. Hauri, P., et al., Clinical evaluation of an anatomy-based patient specific quality assurance system. Vol. 15. 2014. 4647.

91. Heilemann, G., B. Poppe, and W. Laub, On the sensitivity of common gamma-index evaluation methods to MLC misalignments in Rapidarc quality assurance. Med Phys, 2013. **40**(3): p. 031702.

92. Liang, B., et al., Comparisons of volumetric modulated arc therapy (VMAT) quality assurance (QA) systems: sensitivity analysis to machine errors. Radiation Oncology, 2016. **11**(1): p. 146.

93. Maraghechi, B., et al., The sensitivity of gamma index analysis to detect multileaf collimator (MLC) positioning errors using Varian TrueBeam EPID and ArcCHECK for patient-specific prostate volumetric-modulated arc therapy (VMAT) quality assurance. 2017. 1-12.

94. Wang, J., et al., Are simple IMRT beams more robust against MLC error? Exploring the impact of MLC errors on planar quality assurance and plan quality for different complexity beams. J Appl Clin Med Phys, 2016. **17**(3): p. 147-157.

95. McGarry, C.K., et al., The role of complexity metrics in a multi-institutional dosimetry audit of VMAT. The British Journal of Radiology, 2016. **89**(1057): p. 20150445.

96. Crowe, S.B., et al., Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results. Physics in Medicine and Biology, 2015. **60**(6): p. 2587-2601.

97. Agnew, A., et al., Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. Physics in Medicine and Biology, 2014. **59**(9): p. N49-N63.

98. Glenn, M.C., et al., Treatment plan complexity does not predict IROC Houston anthropomorphic head and neck phantom performance. Physics in Medicine & Biology, 2018. **63**(20): p. 205015.

99. Jurado-Bruggeman, D., et al., Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol, 2017. **124**(2): p. 302-310.

100. Park, S.-Y., et al., Assessment of the modulation degrees of intensity-modulated radiation therapy plans. Radiation Oncology, 2018. **13**(1): p. 244.

101. Wang, Y., et al., Correlation between gamma passing rate and complexity of IMRT plan due to MLC position errors. Physica Medica, 2018. **47**: p. 112-120.

102. Park, J.M., J.I. Kim, and S.Y. Park, Modulation indices and plan delivery accuracy of volumetric modulated arc therapy. J Appl Clin Med Phys, 2019. **20**(6): p. 12-22.

103. Tomori, S., et al., A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance. Med Phys, 2018.

104. Nyflot, M.J., et al., Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks. Med Phys, 2019. **46**(2): p. 456-464.

105. Interian, Y., et al., Deep nets vs expert designed features in medical physics: An IMRT QA case study. Med Phys, 2018. **45**(6): p. 2672-2680.

106. Valdes, G., et al., IMRT QA using machine learning: A multi-institutional validation. J Appl Clin Med Phys, 2017. **18**(5): p. 279-284.

107. Li, J., et al., Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy. Int J Radiat Oncol Biol Phys, 2019. **105**(4): p. 893-902.

108. Valdes, G., et al., A mathematical framework for virtual IMRT QA using machine learning. Med Phys, 2016. **43**(7): p. 4953835.

109. Scandidos, Delta 4 PT Getting Started manual Guide, Uppsala Science Park, Sweden. 2011.

110. Litzenberg, D.W., J.M. Moran, and B.A. Fraass, Verification of dynamic and segmental IMRT delivery by dynamic log file analysis. J Appl Clin Med Phys, 2002. **3**(2): p. 63-72.

111. Miften, M., et al., Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys, 2018. **45**(4): p. e53-e83.

112. Woon, W., et al., A study on the effect of detector resolution on gamma index passing rate for VMAT and IMRT QA. Vol. 19. 2018.

113. Zhang, D., et al., Comparison of 3D and 2D gamma passing rate criteria for detection sensitivity to IMRT delivery errors. Journal of applied clinical medical physics, 2018. **19**(4): p. 230-238.

114. Kruse, J.J., On the insensitivity of single field planar dosimetry to IMRT inaccuracies. Med. Phys., 2010. **37**: p. 2516.

115. Abolaban, F., et al., Changes in Patterns of Intensity-modulated Radiotherapy Verification and Quality Assurance in the UK. Clin Oncol, 2016. **28**(8): p. 13.

116. Fuangrod, T., et al., Investigation of a real-time EPID-based patient dose monitoring safety system using site-specific control limits. Radiat Oncol, 2016. **11**(1): p. 106.

117. Lim, S.B., et al., Investigation of a Novel Decision Support Metric for Head and Neck Adaptive Radiation Therapy Using a Real-Time In Vivo Portal Dosimetry System. Technol Cancer Res Treat, 2019. **18**: p. 1533033819873629.

118. Menten, M.J., et al., Automatic reconstruction of the delivered dose of the day using MR-linac treatment log files and online MR imaging. Radiother Oncol, 2020. **145**: p. 88-94.

119. Vial, P., et al., The impact of MLC transmitted radiation on EPID dosimetry for dynamic MLC beams. Med Phys, 2008. **35**(4): p. 1267-77.

120. Gao, S., et al., Evaluation of IsoCal geometric calibration system for Varian linacs equipped with on-board imager and electronic portal imaging device imaging systems. Journal of applied clinical medical physics, 2014. **15**(3): p. 4688-4688.

121. Park, J.M., J.-i. Kim, and S.-Y. Park, Prediction of VMAT delivery accuracy with textural features calculated from fluence maps. Radiation Oncology, 2019. **14**(1): p. 235.

122. Van Esch, A., et al., Acceptance tests and quality control (QC) procedures for the clinical implementation of intensity modulated radiotherapy (IMRT) using inverse planning and the sliding window technique: experience from five radiotherapy departments. Radiother Oncol, 2002. **65**(1): p. 53-70.

APPENDIX I

DATA AND RESULTS OF MODEL-BASED PREDICTION
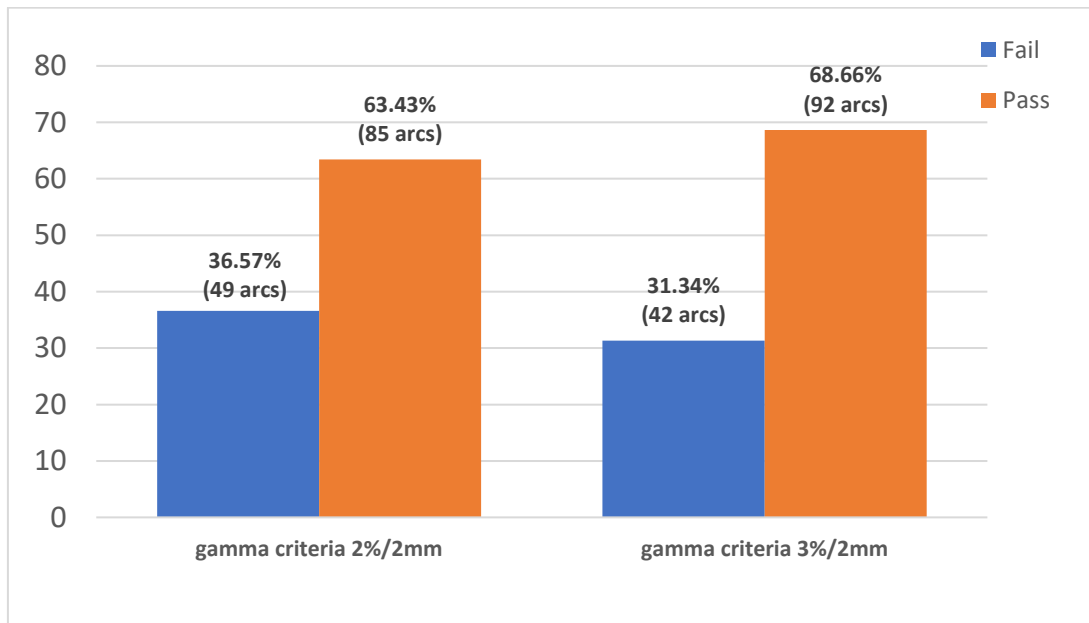
Figure A.1 Population of pass and fail tolerance level for training dataset of 2%/2mm with 10% threshold, and 3%/2mm with 10% threshold
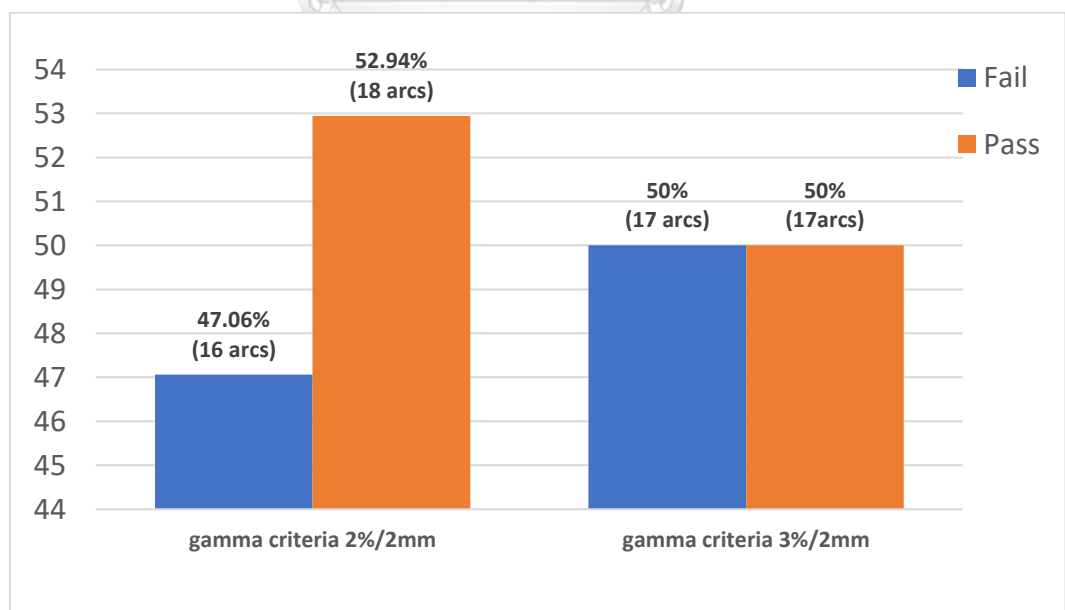
Figure A.2 Population of pass and fail tolerance level for testing dataset of 2%/2mm with 10% threshold, and 3%/2mm with 10% threshold

a) Gamma criteria of 2%/2mm with 10% threshold
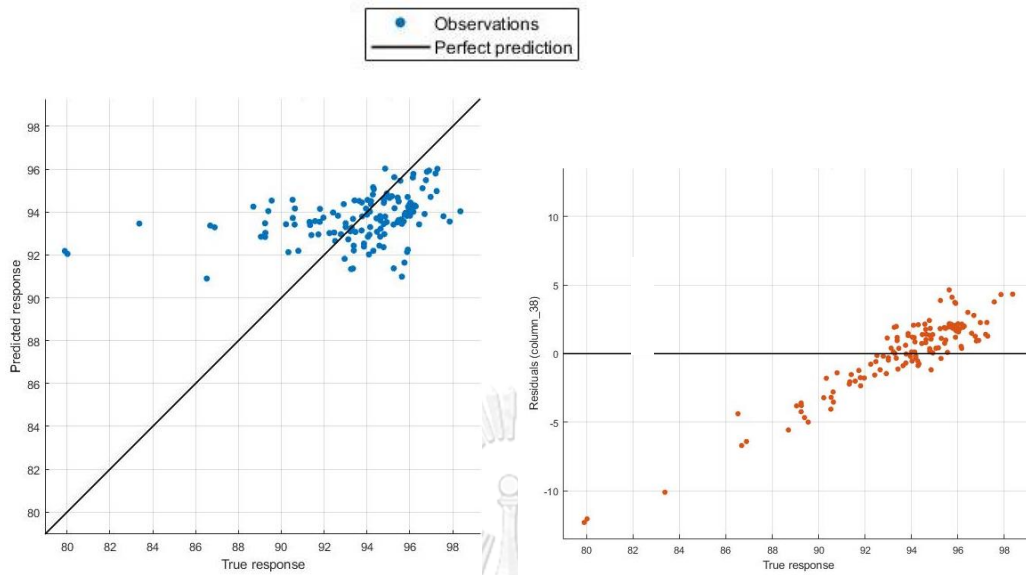
b) Gamma criteria of 3%/2mm with 10% threshold
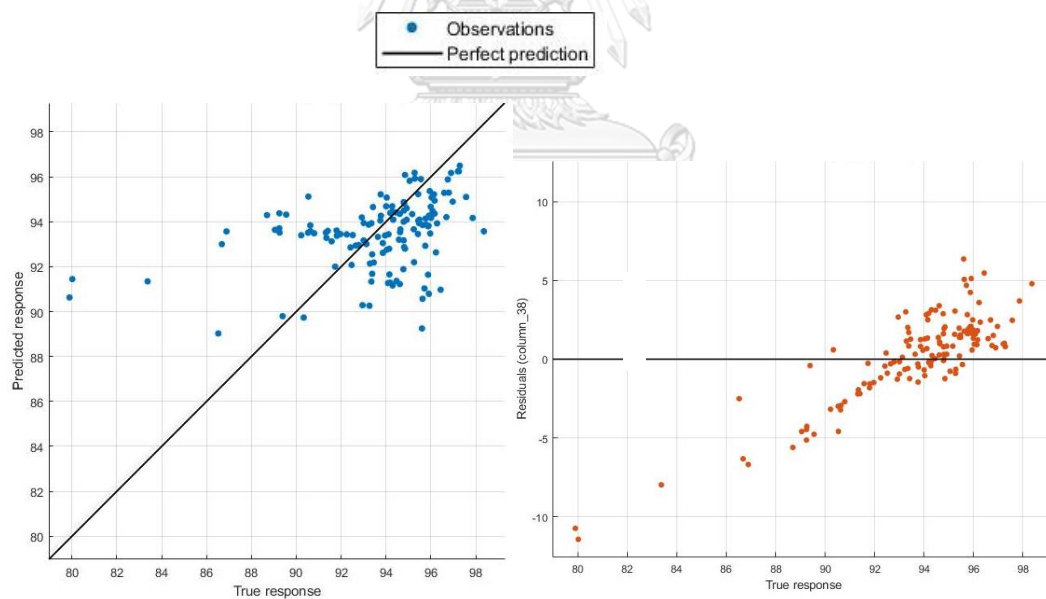
Figure A.3 Relation between prediction and true response, residual error and true response (training dataset) for a) gamma criteria of 2%/2mm with 10% threshold, and b) gamma criteria of 3%/2mm with 10% threshold

a) Gamma criteria of 2%/2mm with 10% threshold
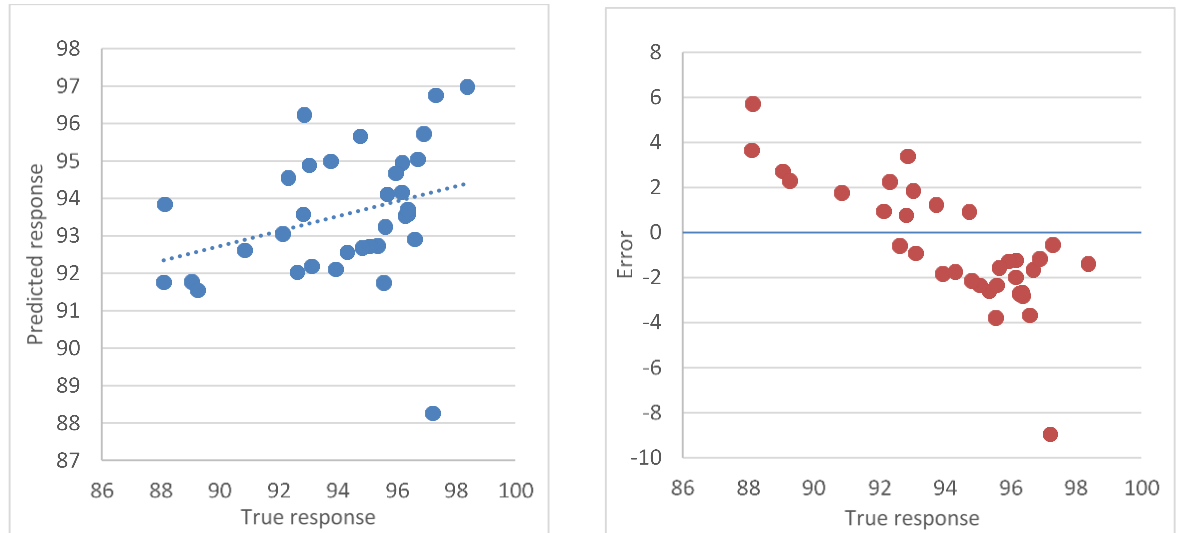
b) Gamma criteria of 3%/2mm with 10% threshold
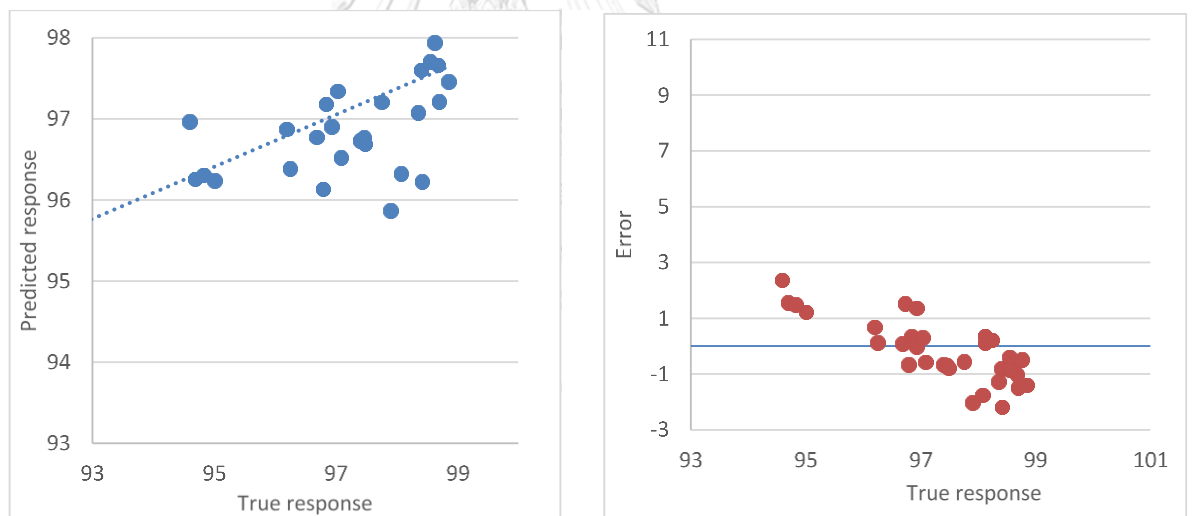
Figure A.4 Relation between prediction and true response, residual error and true response (validation dataset) for a) gamma criteria of 2%/2mm with 10% threshold, add b) gamma criteria of 3%/2mm with 10% threshold

Table A.5 Results of regression model for training dataset at gamma criteria of 2%/2mm with 10% threshold

Regression model (Gaussian process)					
Predicted class (GPR)	Actual class (GPR)	Absolute difference	Predicted class (label)	Actual class (label)	Error
96.76953	97.20459	-0.43506	1	1	
96.9946	97.29004	-0.29544	1	1	
96.88438	96.89941	-0.01504	1	1	
96.69076	96.16699	0.523767	1	1	
94.44497	93.76221	0.682759	1	1	
94.53125	95.94727	-1.41602	1	1	
94.85903	95.05615	-0.19712	1	1	
94.95954	98.36426	-3.40471	1	1	
94.10277	94.92188	-0.8191	1	1	
93.85779	93.77441	0.083376	1	1	
94.27398	97.86377	-3.58979	1	1	
95.14107	95.55664	-0.41557	1	1	
93.97291	94.7876	-0.81469	1	1	
93.58434	93.42041	0.163929	0	0	
95.25227	95.27588	-0.02361	1	1	
93.79117	95.62988	-1.83871	1	1	
92.79284	91.40625	1.386587	0	0	
91.94456	86.68213	5.26243	0	0	

92.62538	92.51709	0.108291	0	0	
93.40975	94.58008	-1.17033	0	1	+
94.94845	96.03271	-1.08427	1	1	
93.85685	95.47119	-1.61434	1	1	
93.49486	93.00537	0.489491	0	0	
94.42094	95.76416	-1.34322	1	1	
92.57829	95.25146	-2.67317	0	1	+
92.81041	94.6167	-1.80629	0	1	+
93.08967	93.28613	-0.19646	0	0	
92.51132	89.24561	3.265712	0	0	
91.78838	89.05029	2.738091	0	0	
94.8325	96.69189	-1.8594	1	1	
93.40707	91.79688	1.610192	0	0	
94.21545	94.18945	0.025993	1	1	
94.40218	95.4834	-1.08122	1	1	
93.44	90.53955	2.900447	0	0	
93.29239	89.55078	3.741609	0	0	
93.77208	94.4458	-0.67372	1	1	
95.72243	96.97266	-1.25022	1	1	
94.75172	94.7998	-0.04808	1	1	
95.0032	94.81201	0.191193	1	1	
94.95905	94.28711	0.671944	1	1	

95.58133	96.14258	-0.56125	1	1	
91.90621	93.34717	-1.44096	0	0	
91.81297	94.0918	-2.27883	0	1	+
93.93701	94.84863	-0.91162	1	1	
92.64115	94.29932	-1.65816	0	1	+
92.90478	92.78564	0.119138	0	0	
91.55269	89.24561	2.307088	0	0	
92.29364	93.38379	-1.09015	0	0	
92.1204	92.95654	-0.83614	0	0	
92.51976	92.46826	0.051497	0	0	
88.26686	79.89502	8.371842	0	0	
91.22838	89.39209	1.836293	0	0	
94.14607	95.28809	-1.14201	1	1	
92.61945	92.24854	0.370914	0	0	
94.92553	95.16602	-0.24048	1	1	
95.28235	95.43457	-0.15222	1	1	
94.67218	94.7876	-0.11542	1	1	
96.2993	96.75293	-0.45363	1	1	
96.40413	96.80176	-0.39763	1	1	
93.86195	95.23926	-1.37731	1	1	
93.93714	94.6167	-0.67956	1	1	
95.26021	94.84863	0.411575	1	1	

93.18448	92.41943	0.765046	0	0	
93.51431	95.91064	-2.39634	0	1	+
94.41551	96.0083	-1.59279	1	1	
92.75322	93.12744	-0.37422	0	0	
92.60284	93.38379	-0.78095	0	0	
93.43416	91.96777	1.466384	0	0	
92.3471	91.73584	0.611261	0	0	
91.68795	90.33203	1.355915	0	0	
92.54978	95.62988	-3.0801	0	1	+
92.2851	93.26172	-0.97662	0	0	
93.43623	90.63721	2.799027	0	0	
94.50074	95.80078	-1.30005	1	1	
95.82294	96.59424	-0.7713	1	1	
93.37085	93.00537	0.365479	0	0	
92.3316	86.88965	5.441947	0	0	
93.39429	94.47021	-1.07592	0	1	+
94.56857	96.16699	-1.59842	1	1	
94.7659	96.27686	-1.51096	1	1	
90.79306	83.37402	7.419038	0	0	
93.22995	91.80908	1.420863	0	0	
94.09344	94.00635	0.08709	1	1	
94.33302	95.71533	-1.38231	1	1	

93.51358	90.52734	2.986236	0	0	
92.48433	91.34521	1.139114	0	0	
93.44358	94.15283	-0.70926	0	1	+
92.70149	93.45703	-0.75554	0	0	
92.81423	90.625	2.189235	0	0	
93.86683	94.64111	-0.77428	1	1	
93.37636	94.7876	-1.41124	0	1	+
94.22122	95.92285	-1.70163	1	1	
93.80536	93.34717	0.458195	1	0	+
94.41558	95.87402	-1.45844	1	1	
94.2246	96.22803	-2.00342	1	1	
94.08216	93.95752	0.124638	1	1	
93.3289	88.69629	4.632607	0	0	
94.88416	94.92188	-0.03771	1	1	
94.99041	95.89844	-0.90803	1	1	
94.17878	94.14063	0.038158	1	1	
89.43038	86.52344	2.906945	0	0	
88.9412	80.01709	8.924111	0	0	
94.21958	92.91992	1.299658	1	0	+
93.99451	95.42236	-1.42785	1	1	
95.19212	95.43457	-0.24245	1	1	
94.0028	94.62891	-0.62611	1	1	

93.94404	93.64014	0.3039	1	0	+
93.38216	92.65137	0.730794	0	0	
92.73955	95.75195	-3.01241	0	1	+
92.81485	93.85986	-1.04502	0	1	+
92.81012	91.57715	1.232972	0	0	
93.14787	95.60547	-2.4576	0	1	+
93.46794	93.2251	0.242847	0	0	
93.36411	94.11621	-0.7521	0	1	+
94.03963	96.43555	-2.39592	1	1	
93.59173	93.75	-0.15827	0	1	+
94.28042	95.97168	-1.69126	1	1	
94.58806	94.03076	0.557296	1	1	
94.89836	94.32373	0.574628	1	1	
93.3156	94.04297	-0.72737	0	1	+
93.61112	94.81201	-1.20089	0	1	+
91.94243	89.25781	2.684622	0	0	
94.67886	97.5708	-2.89194	1	1	
92.6744	94.77539	-2.10099	0	1	+
92.65513	93.85986	-1.20474	0	1	+
91.86424	90.7959	1.068341	0	0	
92.4413	95.87402	-3.43272	0	1	+
94.55157	94.18945	0.362112	1	1	

94.76974	94.2749	0.494834	1	1
96.81177	97.25342	-0.44165	1	1
94.8853	96.08154	-1.19624	1	1
94.71514	95.99609	-1.28096	1	1
92.88085	91.3208	1.560049	0	0
92.88994	90.22217	2.667777	0	0

Table A.6 Results of regression model for training dataset at gamma criteria of 3%/2mm with 10% threshold

Regression model					
(Assembly of trees)					
Predicted class (GPR)	Actual class (GPR)	Absolute difference	Predicted class (label)	Actual class (label)	Error
98.4537	98.9624	-0.5087	1	1	
98.49356	98.82813	-0.33457	1	1	
98.49021	98.49854	-0.00832	1	1	
97.60313	98.24219	-0.63906	1	1	
98.0049	98.12012	-0.11522	1	1	
98.01489	98.12012	-0.10523	1	1	
98.32543	98.40088	-0.07545	1	1	
98.21969	99.18213	-0.96244	1	1	

97.96929	98.7793	-0.81001	1	1	
97.09489	97.65625	-0.56136	1	1	
97.60561	98.88916	-1.28355	1	1	
98.19183	98.07129	0.12054	1	1	
97.17663	97.32666	-0.15003	1	1	
97.3168	96.74072	0.57608	1	0	+
98.25436	98.0957	0.158656	1	1	
97.60024	97.97363	-0.37339	1	1	
97.0119	96.75293	0.258973	1	0	+
95.7564	93.37158	2.384814	0	0	
96.60242	96.45996	0.142463	0	0	
97.08679	97.86377	-0.77698	1	1	
98.30398	98.46191	-0.15793	1	1	
97.5511	98.12012	-0.56902	1	1	
97.3767	97.60742	-0.23072	1	1	
96.92864	97.98584	-1.0572	0	1	+
96.89883	98.03467	-1.13584	0	1	+
96.52382	98.05908	-1.53526	0	1	+
96.26878	97.05811	-0.78933	0	1	+
96.7565	95.58105	1.175444	0	0	
96.23569	95.00732	1.228368	0	0	
97.73205	98.4375	-0.70545	1	1	

97.30031	96.65527	0.645035	1	0	+
96.64882	98.05908	-1.41026	0	1	+
98.06512	98.19336	-0.12824	1	1	
96.58472	94.64111	1.943606	0	0	
96.68913	94.7876	1.901537	0	0	
97.00381	97.41211	-0.4083	0	1	+
98.36211	98.68164	-0.31953	1	1	
97.35591	97.54639	-0.19048	1	1	
97.47806	97.72949	-0.25143	1	1	
97.26491	97.14355	0.121358	1	1	
98.21339	98.36426	-0.15087	1	1	
96.21335	96.43555	-0.2222	0	0	
96.38547	97.11914	-0.73367	0	1	+
96.81435	97.44873	-0.63438	0	1	+
96.62107	97.53418	-0.91311	0	1	+
96.96164	97.00928	-0.04764	0	0	
96.2571	94.68994	1.567155	0	0	
96.65201	97.0459	-0.39389	0	1	+
96.35837	96.82617	-0.46781	0	0	
96.26959	96.75293	-0.48334	0	0	
92.16787	82.00684	10.16104	0	0	
92.29918	94.4458	-2.14662	0	0	

97.99467	97.64404	0.350625	1	1	
96.62242	96.45996	0.162457	0	0	
98.04328	98.15674	-0.11346	1	1	
98.05538	97.87598	0.179408	1	1	
97.8481	97.79053	0.057574	1	1	
98.54845	98.74268	-0.19423	1	1	
98.43331	98.49854	-0.06523	1	1	
97.16826	97.81494	-0.64668	1	1	
97.0516	97.66846	-0.61686	1	1	
98.03304	97.20459	0.828445	1	1	
96.89383	97.30225	-0.40841	0	1	+
96.73959	98.51074	-1.77115	0	1	+
98.18082	98.57178	-0.39095	1	1	
96.83999	97.3877	-0.5477	0	1	+
95.59766	96.91162	-1.31396	0	0	
97.23198	96.38672	0.845262	1	0	+
96.09088	95.39795	0.692931	0	0	
92.93304	95.00732	-2.07429	0	0	
96.57094	97.66846	-1.09752	0	1	+
96.169	96.64307	-0.47407	0	0	
96.76169	95.60547	1.156223	0	0	
97.74384	98.19336	-0.44952	1	1	

98.31587	98.57178	-0.2559	1	1	
96.94646	97.25342	-0.30696	0	1	+
96.32127	93.50586	2.815406	0	0	
96.63489	97.64404	-1.00915	0	1	+
97.60307	98.46191	-0.85884	1	1	
97.98347	98.49854	-0.51507	1	1	
93.27128	86.54785	6.723432	0	0	
96.92413	97.36328	-0.43916	0	1	+
97.55225	97.58301	-0.03075	1	1	
96.63351	98.52295	-1.88944	0	1	+
96.93007	95.28809	1.641987	0	0	
96.14763	96.06934	0.078296	0	0	
96.49791	97.70508	-1.20717	0	1	+
96.06994	98.14453	-2.07459	0	1	+
96.66266	95.10498	1.557675	0	0	
97.08791	97.25342	-0.16551	1	1	
96.91813	97.53418	-0.61605	0	1	+
97.61844	98.12012	-0.50167	1	1	
97.11912	96.99707	0.122047	1	0	+
97.78366	98.13232	-0.34867	1	1	
97.19397	98.73047	-1.5365	1	1	
97.51046	97.31445	0.196003	1	1	

96.36597	93.78662	2.579353	0	0	
97.93241	98.03467	-0.10225	1	1	
97.65644	98.18115	-0.52471	1	1	
96.05999	97.85156	-1.79157	0	1	+
92.99855	91.6748	1.323741	0	0	
92.9846	85.18066	7.803935	0	0	
97.24858	96.7041	0.544479	1	0	+
97.48715	98.14453	-0.65738	1	1	
98.1325	98.21777	-0.08528	1	1	
97.12397	97.66846	-0.54448	1	1	
96.7673	97.18018	-0.41288	0	1	+
96.87359	96.72852	0.145079	0	0	
97.14959	98.57178	-1.42219	1	1	
96.80842	97.18018	-0.37175	0	1	+
96.72954	95.69092	1.038623	0	0	
94.92871	97.98584	-3.05713	0	1	+
97.56129	97.98584	-0.42455	1	1	
96.99633	96.93604	0.060297	0	0	
96.61626	98.47412	-1.85786	0	1	+
97.15174	96.99707	0.154672	1	0	+
97.51717	98.10791	-0.59074	1	1	
98.22847	98.59619	-0.36772	1	1	

98.16614	98.21777	-0.05163	1	1	
95.91071	97.11914	-1.20843	0	1	+
96.01078	97.37549	-1.36471	0	1	+
95.54435	95.3125	0.231849	0	0	
98.12547	98.85254	-0.72707	1	1	
95.71168	97.46094	-1.74925	0	1	+
94.93682	96.94824	-2.01143	0	0	
95.383	95.22705	0.155949	0	0	
95.23459	97.9126	-2.67801	0	1	+
97.38279	97.30225	0.08054	1	1	
96.99559	97.13135	-0.13576	0	1	+
98.34914	98.64502	-0.29588	1	1	
97.66461	98.36426	-0.69965	1	1	
97.62398	98.25439	-0.63042	1	1	
96.69614	95.86182	0.834327	0	0	
96.68446	94.83643	1.848039	0	0	

Table A.7 Results of classification model for testing dataset at gamma criteria of 2%/2mm with 10% threshold

Classification model (Neighbour classifier)		
Predicted class (label)	Actual class (label)	Error
0	0	
0	0	
0	0	
0	0	
0	0	
0	1	+
0	0	
0	0	
1	0	+
1	0	+
1	1	
1	0	+
0	0	
1	0	+
1	1	
0	1	+
1	1	
0	1	+

0	1	+
1	1	
1	0	+
1	1	

Table A.8 Results of classification model for testing dataset at gamma criteria of 3%/2mm with 10% threshold

Classification model		
(Neighbour classifier)		
Predicted class (label)	Actual class (label)	Error
0	0	
0	0	
0	0	
0	0	
0	0	
0	1	+
0	0	
0	0	
1	0	+
1	0	+
1	1	
1	0	+
0	0	

1	0	+
1	1	
0	1	+
0	1	+
1	1	
0	1	+
1	1	
1	0	+
1	1	

Table A.9 Results of regression model for testing dataset at gamma criteria of 2%/2mm with 10% threshold

Regression model (Gaussian process)					
Predicted class (GPR)	Actual class (GPR)	Absolute difference	Predicted class (label)	Actual class (label)	Error
91.78838	89.05029	2.738091	0	0	
91.55269	89.24561	2.307088	0	0	
88.26686	79.89502	8.371842	0	0	
89.32699	75.08545	14.24154	0	0	
90.39376	60.15625	30.23751	0	0	
93.63325	66.65039	26.98286	0	0	
93.11177	92.30957	0.802195	0	0	

93.5909	92.81006	0.780839	0	0	
92.74387	95.33691	-2.59304	0	1	+
91.75989	95.54443	-3.78454	0	1	+
93.5836	96.37451	-2.79091	0	1	+
92.11259	93.9209	-1.80831	0	1	+
92.62598	90.84473	1.781249	0	0	
92.72343	95.05615	-2.33273	0	1	+
93.25185	95.58105	-2.32921	0	1	+
93.85633	88.12256	5.733773	1	0	+
94.99548	93.73779	1.257691	1	1	
94.8953	93.01758	1.877721	1	0	+
96.24728	92.84668	3.400603	1	0	+
94.67492	95.94727	-1.27235	1	1	
93.5431	96.26465	-2.72154	0	1	+
95.05636	96.69189	-1.63553	1	1	
91.76764	88.09814	3.669496	0	0	
92.03789	92.61475	-0.57685	0	0	
92.19198	93.10303	-0.91105	0	0	
95.66327	94.73877	0.924499	1	1	
92.57578	94.29932	-1.72353	0	1	+
93.06134	92.11426	0.947086	0	0	
92.68824	94.81201	-2.12377	0	1	+

95.72625	96.88721	-1.16095	1	1	
94.17758	96.15479	-1.97721	1	1	
94.11538	95.6543	-1.53892	1	1	
93.71996	96.3501	-2.63014	1	1	
92.91171	96.58203	-3.67032	0	1	+

Table A.10 Results of regression model for testing dataset at gamma criteria of 3%/2mm with 10% threshold

Regression model					
(Assembly of trees)					
Predicted class (GPR)	Actual class (GPR)	Absolute difference	Predicted class (label)	Actual class (label)	Error
96.23569	95.00732	1.228368	0	0	
96.2571	94.68994	1.567155	0	0	
92.16787	82.00684	10.16104	0	0	
93.16752	80.24902	12.9185	0	0	
92.67384	65.66162	27.01222	0	0	
97.56729	70.97168	26.59561	1	0	+
96.74028	96.61865	0.121628	0	0	
96.90531	96.92383	-0.01852	0	0	
97.20577	97.75391	-0.54814	1	1	
96.33055	98.07129	-1.74074	0	1	+
96.22408	98.41309	-2.18901	0	1	+

96.52075	97.08252	-0.56177	0	1	+
96.87179	96.19141	0.680381	0	0	
96.73097	97.3999	-0.66894	0	1	+
97.94184	98.62061	-0.67876	1	1	
96.96714	94.59229	2.374856	0	0	
97.18206	96.83838	0.343679	1	0	+
97.34273	97.03369	0.309037	1	1	
98.25734	96.72852	1.528821	1	0	+
97.21217	98.69385	-1.48168	1	1	
97.46019	98.85254	-1.39234	1	1	
98.15091	98.54736	-0.39645	1	1	
96.30944	94.82422	1.485224	0	0	
96.38794	96.25244	0.135498	0	0	
96.13949	96.78955	-0.65006	0	0	
98.29906	96.92383	1.375232	1	0	+
96.76879	97.46094	-0.69215	0	1	+
96.7761	96.67969	0.09641	0	0	
96.69248	97.48535	-0.79287	0	1	+
98.29496	98.76709	-0.47213	1	1	
97.66132	98.68164	-1.02032	1	1	
97.08176	98.35205	-1.27029	1	1	
97.70355	98.54736	-0.84381	1	1	

| 95.87206 | 97.90039 | -2.02833 | 0 | 1 | + |

APPENDIX II

MATLAB CODE

B1. MATLAB code for extracting the feature of leaf speed and acceleration

```
function Calculation_leaf_speed_and_acceleration()
load ('mlc data.mat'); % load mlc data from plan dicom
time_LS = repmat (0.4167, [177 60]);
diff_A = abs(diff (A,1,1));
diff_B = abs(diff (B,1,1));
% calculation for Leaf Speed (LS)in bank A amd B
LS_A = diff_A./time_LS;
LS_B = diff_B./time_LS;
[a b] = size (LS_A);
%let cal.LS bank A
LS_A0_4 =  sum(sum (LS_A>=0 & LS_A <= 4));
LS_A4_8 =  sum(sum (LS_A>=4 & LS_A <= 8));
LS_A8_12 =  sum(sum (LS_A>=8 & LS_A <= 12));
LS_A12_16 =  sum(sum (LS_A>=12 & LS_A <= 16));
LS_A16_20 =  sum(sum (LS_A>=16));
LS_A0_4per = LS_A0_4/(a*b);
LS_A4_8per = LS_A4_8/(a*b);
LS_A8_12per = LS_A8_12/(a*b);
LS_A12_16per = LS_A12_16/(a*b);
LS_A16_20per = LS_A16_20/(a*b);
%let cal.LS bank B
LS_B0_4 =  sum(sum (LS_B>=0 & LS_B <= 4));
LS_B4_8 =  sum(sum (LS_B>=4 & LS_B <= 8));
LS_B8_12 =  sum(sum (LS_B>=8 & LS_B <= 12));
LS_B12_16 =  sum(sum (LS_B>=12 & LS_B <= 16));
LS_B16_20 =  sum(sum (LS_B>=16));
LS_B0_4per = LS_B0_4/(a*b);
LS_B4_8per = LS_B4_8/(a*b);
LS_B8_12per = LS_B8_12/(a*b);
LS_B12_16per = LS_B12_16/(a*b);
LS_B16_20per = LS_B16_20/(a*b);
% calculation for Leaf Acceleration(LA)in bank A amd B
time_LA = repmat (0.4167, [176 60])
```

```
diff_LS_A = abs(diff(LS_A, 1, 1));
LA_A = diff_LS_A./time_LA;
diff_LS_B = abs(diff(LS_B, 1, 1));
LA_B = diff_LS_B./time_LA;
[a b] = size (LA_A);
%let cal.LA bank A
LA_A0_4 =  sum(sum (LA_A>=0 & LA_A <= 4));
LA_A4_8 =  sum(sum (LA_A>=4 & LA_A <= 8));
LA_A8_12 =  sum(sum (LA_A>=8 & LA_A <= 12));
LA_A12_16 =  sum(sum (LA_A>=12 & LA_A <= 16));
LA_A16_20 =  sum(sum (LA_A>=16));
LA_A0_4per = LA_A0_4/(a*b);
LA_A4_8per = LA_A4_8/(a*b);
LA_A8_12per = LA_A8_12/(a*b);
LA_A12_16per = LA_A12_16/(a*b);
LA_A16_20per = LA_A16_20/(a*b);
%let cal.LA bank B
LA_B0_4 =  sum(sum (LA_B>=0 & LA_B <= 4));
LA_B4_8 =  sum(sum (LA_B>=4 & LA_B <= 8));
LA_B8_12 =  sum(sum (LA_B>=8 & LA_B <= 12));
LA_B12_16 =  sum(sum (LA_B>=12 & LA_B <= 16));
LA_B16_20 =  sum(sum (LA_B>=16));
LA_B0_4per = LA_B0_4/(a*b);
LA_B4_8per = LA_B4_8/(a*b);
LA_B8_12per = LA_B8_12/(a*b);
LA_B12_16per = LA_B12_16/(a*b);
LA_B16_20per = LA_B16_20/(a*b);
clear
close
```

B2. MATLAB code for extracting the feature of texture analysis

```
function Texture = Texure_analysis(A,B, MU_Weight)
load('bank_A.mat')          %%import MLC position
load('bank_B.mat')
load('MU_Weight.mat')        %%import MU weight
leafWidth = zeros(1,60);
leaves_diff = round(A-B);
maxLeaves_diff = max(max(leaves_diff));
A_max = round(max(max(abs(A))));
```

```
B_max = round(max(max(abs(B))));
leafWidth(1,1:10) = 10;
leafWidth(1,11:50) = 5;
leafWidth(1,51:60) = 10;
maxsize = max([A_max B_max]);
fieldWidth = (maxsize+10)*2;
mid = fieldWidth/2;
fieldHeight = sum(leafWidth);
fluence = zeros(fieldWidth, fieldHeight);
for i = 1:178
  for j = 1:60
    if leaves_diff(i,j) == 0;
      A(i,j) = 0;
      B(i,j) = 0;
    end
  end
end
for r=1:178                      %% arc
  flu = A(r,:);
  B_flu = A(r,:);                %% A leaf position of arc(r)
  flu_dis = leaves_diff(r,:);       %% leaf distance A to B of arc(r)
  t = MU_Weight(1,r);
    for k = 1:60                  %% check each leaf
      if (k<=10)
        if flu(1,k)~=0            %%
          m = flu(1,k);          %% A position
          n = flu_dis(1,k);      %% distance
          p = round(mid+m);       %% mid + A position
            for j = 0:9
              eachwidth = (10*k)-j;
              for q = 1:(n-1)
                openfield = p-q;
        fluence(openfield,eachwidth)=fluence(openfield,eachwidth) + t;
              end
            end
        end
      end
      if (k>=11) && (k<=50)
        if flu(1,k)~=0
          m = flu(1,k);          %% A position
          n = flu_dis(1,k);       %% distance
```

```
            p = round(mid+m);           %% mid + A position
            for j = 0:4
                eachwidth = 50+((5*k)-j);
                for q = 1:(n-1)
                    openfield = p-q;
    fluence(openfield,eachwidth) = fluence(openfield,eachwidth) + t;

                end
            end
        end
      end
      if (k>=51)
        if flu(1,k)~=0
          m = flu(1,k);              %% A position
          n = flu_dis(1,k);          %% distance
          p = round(mid+m);          %% mid + A position
          k = k-50;
            for j = 0:9
                eachwidth = 300+(10*k)-j;
                for q = 1:(n-1)
                    openfield = p-q;
    fluence(openfield,eachwidth) = fluence(openfield,eachwidth) + t;
                end
            end
          end
        end
      end
    end
offsets = [0 1; -1 1;-1 0;-1 -1]; %convert matrix to grey scale for texture
%analysis
    [glcms,SI] = graycomatrix(fluence,'Offset',offsets);
    SI = uint8(rescale(SI, 0, 255));
    Stats = graycoprops(SI);
    Entropy = entropy (SI);
        Save ('Stats.mat','-struct', 'Stats');
    save ('Entropy.mat', 'Entropy');
    save ('fluence.mat', 'fluence');
imshow(rescale(fluence))
imshow(rescale(SI))
```

<div align="center">

VITA

</div>

NAME	Sangutid Thongsawad
DATE OF BIRTH	10 January 1982
PLACE OF BIRTH	Srisaket province
INSTITUTIONS ATTENDED	HRH Princess Chulabhorn College of Medical Science (PCCMS), Chulabhorn Royal Academy (CRA), Bangkok, Thailand, 10210
HOME ADDRESS	199/186, Casa city (Donmuang-Srisaman) village, Sikan, Donmuang, Bangkok, 0210
PUBLICATION	Thongsawad S, Srisatit S, Fuangrod T. Error detection sensitivity test using complex errors on three patient-specific VMAT QA systems. J Phys Conf Ser. 2019; 1285: 12030.
AWARD RECEIVED	Best Oral Presentation Award of International Nuclear Science and Technology conference 2019 (INST 2019)