PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Escuela de Posgrado

Enhanced Predictive Modelling of Toxicities in Gynecologic Cancer Patients Treated with High Dose Rate Brachytherapy Using Syed-Neblett or Fletcher-Suit-Delclos Tandem and Ovoid Applicators: A Machine Learning Perspective

Tesis para obtener el grado académico de Doctor en Física que presenta:

Andres Portocarrero Bonifaz

Asesor:

Dr. Daniel Francisco Palacios Fernández

Lima, 2024

Informe de Similitud

Yo, Daniel Francisco Palacios Fernández, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la Tesis de Doctorado en Física titulada "*Enhanced Predictive Modelling of Toxicities in Gynecologic Cancer Patients Treated with High Dose Rate Brachytherapy Using Syed-Neblett or Fletcher-Suit-Delclos Tandem and Ovoid Applicators: A Machine Learning Perspective*", del autor Andrés Portocarrero Bonifaz, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 19%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 18/12/2024 11:43a.m. (UTC-0500).
- He revisado con detalle dicho reporte y la Tesis o Trabajo de Suficiencia Profesional, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:
Lima, 18/12/2024

Apellidos y nombres del asesor: *Palacios Fernández, Daniel Francisco*	
CE: *001490304*	Firma
ORCID: *0000-0001-8248-347X*	

Dedicatoria

A mis amores y motivos de vida, Laura y Matteo

A mis padres que me han enseñado y dado todo, Rosa y Carlos

A mi querida hermana Bianca, mi inspiración

A toda mi cariñosa familia

Acknowledgements

To my wife Laura, who supports my research and encourages me in all my professional and personal endeavors, I am deeply grateful. Even while pregnant with our son Matteo, she listened to my ideas and discussed science with me in my area of expertise. Her listening and helpful perspectives have been invaluable, whether I faced technical challenges or emotional hurdles.

To my family, especially my parents Rosa and Carlos; and my sister Bianca, from whom I have learned so much and who have always motivated me to give my best throughout my life, I am deeply grateful. They have played an important role in shaping my vocation for service and my interest in research and teaching. They have helped me become a very empathetic person, fostering in me a genuine interest in others and a strong desire to help.

To my research advisors Scott R. Silva and Keith T. Sowards, who gently shared their knowledge with me, supervised this work and helped me find my niche in the clinic: brachytherapy. Their teachings are invaluable, and they have become excellent friends who always motivate me. They serve as outstanding role models, both personally and clinically.

To all my friends from University of Louisville, who helped me through difficult times and whom I trust unconditionally: Maxwell Kassel, Salman Syed, Reema Ishteiwy, Marjan Shojaei, and all the staff, residents, and faculty.

Finally, I am grateful to Dr. Daniel Palacios, who has been supporting and mentoring me since I was an undergraduate student; and very kindly accepted to be my advisor for this doctoral thesis.

Abstract

This doctoral thesis presents an advanced predictive modeling approach for assessing toxicities in gynecologic cancer patients treated with high-dose-rate (HDR) brachytherapy. Using machine learning algorithms such as Support Vector Machines, Random Forest, and Neural Networks, the study aims to enhance the accuracy of toxicity predictions, thereby allowing the clinician to optimize treatment plans and improving patient outcomes. This research focuses on patients treated with Syed-Neblett or Fletcher-Suit-Delclos tandem and ovoid applicators, commonly used in brachytherapy procedures. Objectives include comparing dosimetric profiles and associated toxicities between the two applicator types, investigating the predictive value of non-dosimetric factors, evaluating the performance of various machine learning models against traditional statistical methods, and identifying the most effective predictive model through rigorous cross-validation and feature selection techniques. A comprehensive dataset, one of the most sizeable in this topic, serves as the basis for training and testing the models. By integrating demographic, treatment, and tumor-related data, the study aims to develop ML models that offer a superior performance compared to existing methods. The findings highlight the potential of machine learning to revolutionize brachytherapy planning by providing physicians with precise, patient-specific risk assessments, ultimately enhancing the quality of care for gynecologic cancer patients. This research not only advances the field of radiation oncology but also contributes valuable insights into the integration of machine learning in clinical practice, paving the way for more effective and personalized cancer treatments.

Resumen

Esta tesis doctoral propone un método innovador para predecir toxicidades en pacientes con cáncer ginecológico tratados con braquiterapia de alta tasa de dosis (HDR) y radioterapia externa (EBRT). Combinando conocimientos de Física de Radiaciones Ionizantes, Oncología y Ciencia de Datos, se emplean algoritmos de Machine Learning, como Support Vector Machines, Random Forest y Redes Neuronales, para entrenar y desarrollar modelos multivariables que integran variables de dosis de radiación, datos demográficos, factores clínicos y características del tratamiento, entre otros. En primer lugar, el estudio, basado en una de las bases de datos más grandes utilizadas en este campo, con más de 12 años de recolección de datos, compara los aplicadores Syed-Neblett y Fletcher-Suit-Delclos, destacando la importancia de crear modelos multivariables en lugar de depender únicamente de la práctica histórica de utilizar tolerancias de dosis derivadas de estudios poblacionales. Posteriormente, se explora el uso del Machine Learning como herramienta predictiva en pacientes con cáncer ginecológico tratados con HDR y EBRT, realizando un análisis exhaustivo sobre cómo entrenar estos modelos de manera óptima para apoyar tratamientos de radiación más personalizados y efectivos. Los hallazgos subrayan el potencial del Machine Learning para revolucionar la planificación de la braquiterapia al proporcionar a los médicos evaluaciones de riesgo precisas y adaptadas a cada paciente, mejorando así la calidad de la atención en pacientes con cáncer ginecológico.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Brachytherapy

Brachytherapy is a form of radiation therapy in which a radiation source is placed near a tumor or adjacent to tissue at risk of harboring microscopic disease. It is a highly conformal treatment that unlike external radiation beam treatments (EBRT), treats the cancer from the inside, minimizing the dose to the organs adjacent to the treatment area (1–3).

1.1.1 Physical Magnitudes and Definitions

Activity (A): It is the number of disintegrations per second for a specific radioactive source. It is commonly measured in Curies (Ci), while in the International System of Units (SI), it is expressed in Becquerels (Bq); 1 Bq = 1 disintegration per second and 1 Ci = 3.7×10^{10} Bq.

Absolute Dose (D): It is the energy deposited per unit mass. It is commonly represented in Grays (Gy) which is equivalent to Joules per Kilogram (J/kg) in the International System of Units (SI).

Air Kerma Strength (U): It is the recommended quantity for specifying a brachytherapy source strength by the American Association of Physicists in Medicine (AAPM). It is defined as the air-kerma rate, in vacuo, and due to photons greater than a threshold energy, at a distance d, times the distance squared (μGy m^2 / h).

Treatment Planning System (TPS): It is a software that allows the user to do multiple operations that are relevant for generating a radiotherapy treatment plan. It incorporates tools for contouring, for inserting models of brachytherapy applicators, and to set active dwell positions to create a desired dose distribution, among others.

High Dose Rate Brachytherapy: When the dose rate of the source used for brachytherapy treatments exceeds 12 Gy per hour it is defined as a High Dose Rate (HDR) source. Common examples are Iridium 192 and Cobalt 60.

Intracavitary: An applicator that is inserted inside a body cavity.

Interstitial: An applicator, such as needles, which are directly implanted into the tumor, usually through tissue.

Catheters: Thin, flexible tubes that are specifically designed to be put within a patient's body. In brachytherapy, the needles that are inserted for interstitial brachytherapy are considered catheters.

Transfer-Guide-Tube (TGT): A long hollow tube that connects the afterloader with the applicator that is inserted inside the patient´s body.

Afterloader: A machine located in the brachytherapy treatment rooms that houses the radiation source providing the necessary shielding when not in used. It controls the position and the amount of time that the source will remain in one specific position. The afterloader has two wires, one dummy wire to do quality assurance tests, and one with the radiation source.

Brachytherapy Applicator: In the context of HDR brachytherapy, an applicator is a medical device that is inserted inside a cavity, implanted directly into tissue, or in contact with a patient's body. Examples of intracavitary applicators are vaginal cylinders or Fletcher-Suit-Delclos Tandem and Ovoids (T&O), in which a wire bearing a radioactive source can be inserted with the help of Transfer Guide Tubes that connect the afterloader to the applicator (Figure 1.1). Interstitial applicators would be the ones in which there are additional elements such as needles, which are implanted in tissue to obtain the desired dose distribution, such as the Syed-Neblett (Figure 1.2).

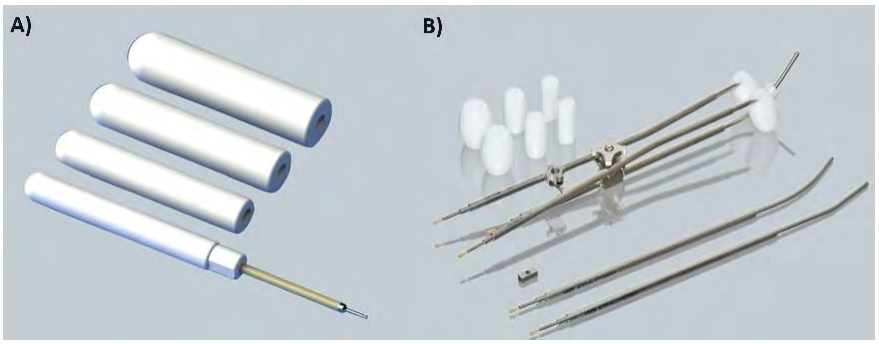


Figure 1.1. A) Vaginal Cylinder B) Fletcher-Suit-Delclos Tandem and Ovoids. Images referenced from the Varian Medical Systems Brachytherapy catalog (4)

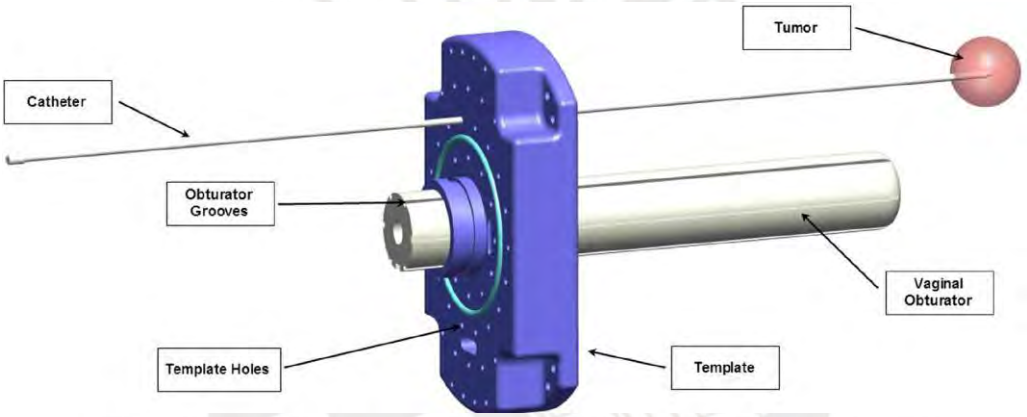


Figure 1.2. Syed Neblett Interstitial Applicator as shown in Damato et al. (5)

1.1.2 TG – 43 Formalism

The updated protocol of the Task Group report No. 43 (TG-43U1) of the American Association of Physicist in Medicine (AAPM) is the most used physics formalism in Treatment Planning Systems (6), and the one employed by Varian´s BrachyVision™ (Varian Medical Systems, Inc., Palo Alto, CA).

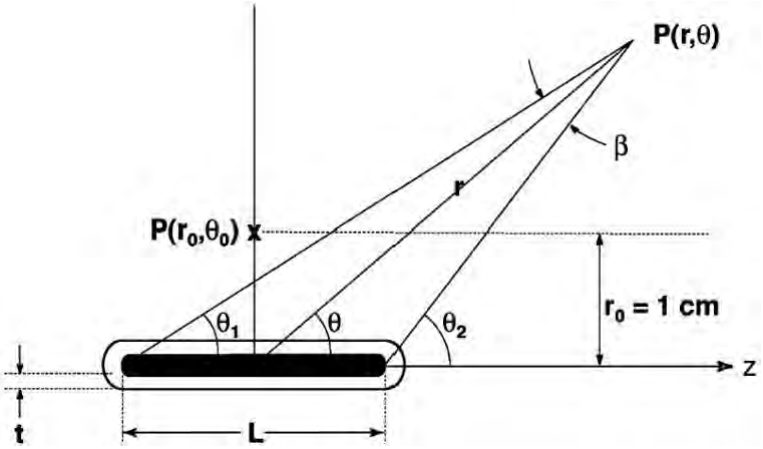


Figure 1.3. Brachytherapy reference coordinate system as specified by TG-43 (7)

This protocol is defined as follows:

$$D(r,\theta) = S_K \cdot \Lambda \cdot \frac{G_L(r,\theta)}{G_L(r_0,\theta_0)} \cdot g_L(r) \cdot F(r,\theta) \quad (1.1)$$

where S_K is the Air-Kerma strength, Λ is the Dose-rate constant in water, $G_L(r, \theta)$ is the Geometry function, $g_L(r)$ is the Radial dose function, and $F(r, \theta)$ is the 2D Anisotropy function.

The Air-Kerma strength is defined by the Air-Kerma rate, in vacuo, and due to photons of an energy greater than a defined threshold intended to exclude contaminant photons arising from characteristic x-rays generated in the outer radioactive source cover. It is calculated at a specified distance from the source center to the point where the Air-Kerma rate is measured, multiplied by the square of the distance, as shown in equation 1.2.

$$S_K = K_\delta(d) \cdot d^2 \quad (1.2)$$

The dose rate constant in water is defined as the ratio between the dose rate at a reference point $r_0 = 1\ cm\ and\ \theta_0 = \pi/2$ and the Air-Kerma strength. This parameter is specific for each source-model and isotope. It is defined as follows:

$$\Lambda = \frac{D(r_0,\theta_0)}{S_K} \quad (1.3)$$

For the geometry function, which is an effective inverse square-law for a hypothetical line source, the TPS uses the formula as defined by equation 1.4.

$$G_L(r,\theta) = \begin{cases} \frac{\beta}{L \cdot r \cdot \sin\theta}, & \theta \neq 0° \\ (r^2 - L^2/4)^{-1}, & \theta = 0° \end{cases} \quad (1.4)$$

where β is the angle subtended between the tips of the line source model and the designated calculation point, denoted as $P(r,\theta)$.

The radial dose function factors in the decrease in the transverse plane caused by attenuation and photon scattering. The influence of the geometry function is completely removed from this parameter. Its formula is shown in equation 1.5.

$$g_L(r) = \frac{D(r,\theta_0)}{D(r_0,\theta_0)} \cdot \frac{G_L(r_0,\theta_0)}{G_L(r,\theta_0)} \quad (1.5)$$

Finally, the 2D anisotropy function describes the variation in dose with respect to the polar angle relative to the transverse plane and is defined by equation 1.6.

$$F(r,\theta) = \frac{D(r,\theta)}{D(r,\theta_0)} \cdot \frac{G_L(r,\theta_0)}{G_L(r,\theta)} \quad (1.6)$$

All the consensus parameters for Λ, $g_L(r)$ and $F(r,\theta)$ are presented in Appendix A. for the GammaMed HDR ^{192}Ir Plus source model and in Appendix B. for the New Varisource HDR VS2000 manufactured by Varian Medical Systems Inc. which were used all throughout this research.

1.1.3 Uncertainty Analysis

An in – depth analysis of the uncertainty analysis can be found in the AAPM TG-43U1 and the Joint Report of AAPM´s TG-138 and GEC-ESTRO (6,8).

For the consensus dataset (a set of critically evaluated dosimetric parameters, endorsed by experts for specific brachytherapy sources) intended for clinical use, Λ

was calculated by taking the average of all the experimentally found values and Monte Carlo simulated value separately, and then by obtaining the mean of those two values as described by equation 1.7. If inconsistencies were present on the reported data, they were excluded from the final Λ determination.

$$\Lambda_{CON} = \frac{\Lambda_{EXP} \pm \Lambda_{MC}}{2} \tag{1.7}$$

For the consensus dataset, on which $F(r, \theta)$ and $g_L(r)$ to use, most of the studied source models were defined by the Monte Carlo data. This consensus dataset was chosen with 3 selection criteria: a) Largest distance range b) highest resolution and c) highest degree of smoothness. The dataset also had to be within an acceptable uncertainty range when checked by experimental methods.

To calculate the uncertainty for each component used to determine these consensus datasets, the Law of Propagation of Uncertainty was used as suggested by the NIST Report 1297 (9).

$$\sigma_y^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_{x_i}^2 + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j}\sigma_{x_i,x_j} \tag{1.8}$$

The second term of the equation was assumed to be zero, as there was no covariance of variables. The total uncertainty is made up of a) Uncertainty of the cross-sections, b) Uncertainty of the source model geometry, and c) Statistical uncertainty due to Monte Carlo.

From equation 1.8, the total percent uncertainty is equal to:

$$\%\sigma_{D(r)} = \sqrt{\%\sigma_{S_K}^2 + \%\sigma_{\Lambda_{CON}}^2 + \%\sigma_{g(r)_{CON}}^2} \tag{1.9}$$

According to AAPM TG-43U1 the total dose calculation uncertainties are 6.7%, 5.7%, and 7.3% at 0.1 cm, 1 cm, and 5 cm, respectively.

For a more detailed, in-depth, and updated uncertainty analysis the reader is referred to the Joint Report of AAPM´s TG-138 and GEC-ESTRO. A summary is shown in the picture below:

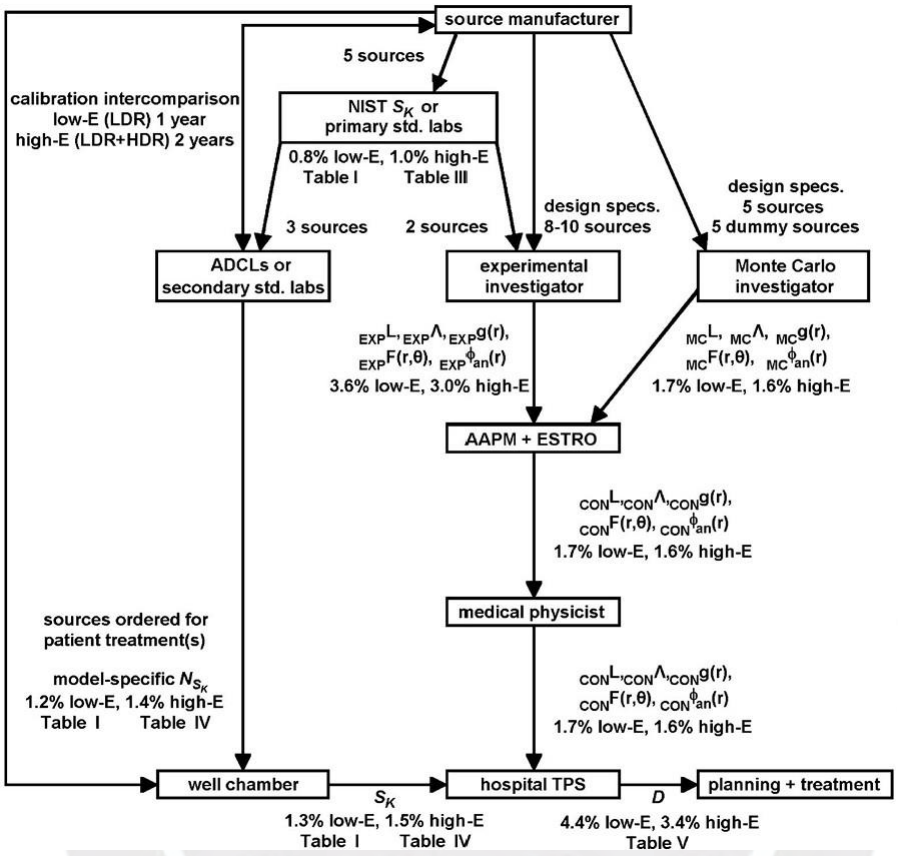


Figure 1.4. Brachytherapy uncertainty budget as shown in TG-138 (8)

Considering the combined uncertainty, TG-138 agrees with TG-43U1 and reports an approximate value of 7% with k = 2 for high energy photon sources such as Ir-192.

1.1.4 Radiation Source

Multiple Iridium 192 radioactive sources housed inside Varian Medical Systems afterloaders were used throughout this research, as they are replaced 3 or 4 times per year for faster patient treatment times. The average activity for a new source is around 10 Ci and 15 Ci for a VariSource iX and Bravos afterloader respectively. Ir-192 has gamma ray spectrum that ranges from 136 keV to 1.06 MeV with a mean energy of 380 keV and a half-life of 73.8 days (10,11). Its decay scheme is shown in Figure 1.5.

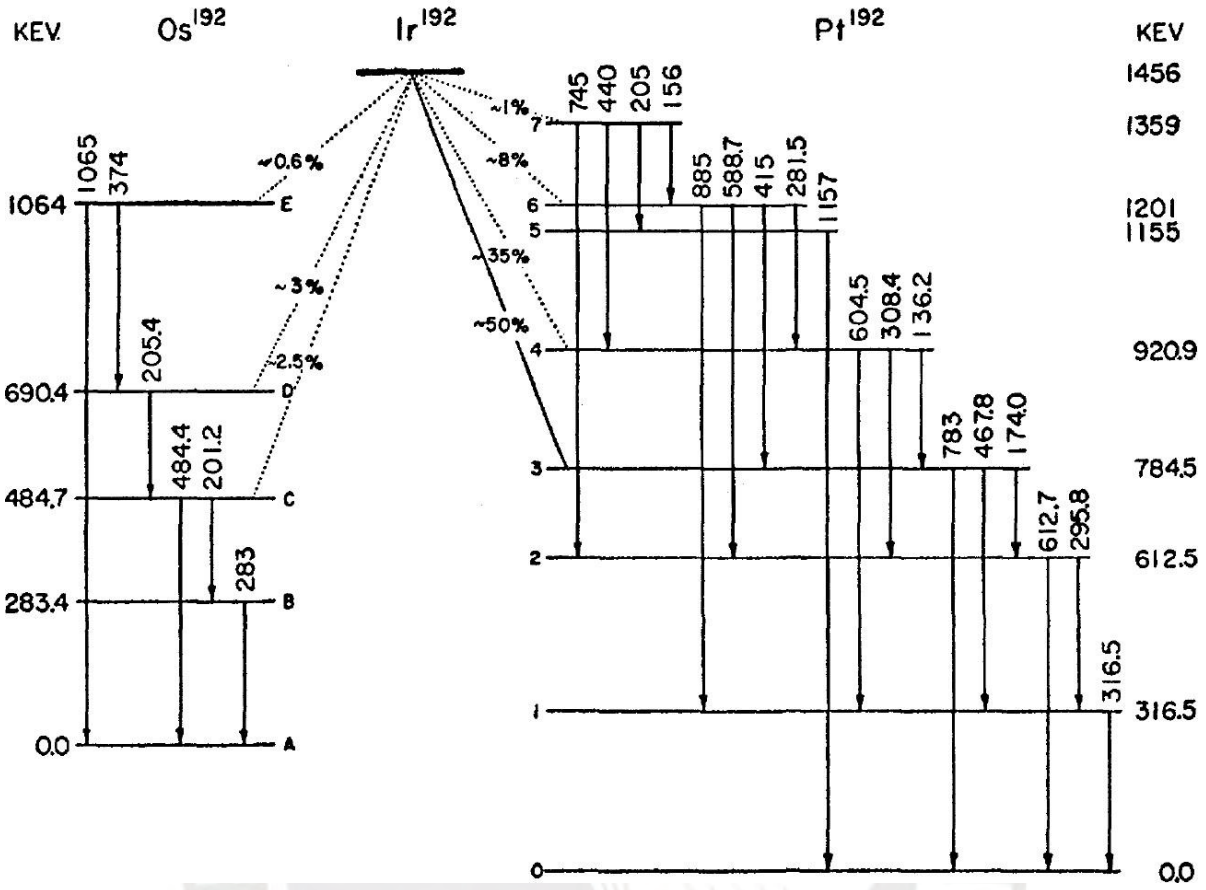


Figure 1.5. Iridium 192 decay scheme as presented by Johns & Nablo (12)

1.1.5 Clinical Procedure

As reported in Portocarrero et al. (13): “At the institution where the research was conducted, brachytherapy is delivered after the completion of primary external beam radiation (EBRT) and concurrent chemotherapy, with the goal of completing EBRT and brachytherapy within 8 weeks. For both Syed and T&O patients, the median primary EBRT dose was 45 Gy. The median HDR doses were 28 Gy and 27.5 Gy for Syed and T&O brachytherapy, respectively. Patients with PET-positive pelvic and/or para-aortic lymph nodes received either a sequential or simultaneous integrated boost to the PET-positive lymph nodes. Prior to 2019, T&O brachytherapy was performed as an outpatient procedure and beginning in 2019 all T&O brachytherapy was performed as two inpatient hospital admissions, with 2–3 treatments per admission, to decrease the total treatment time. All Syed treatments were performed during a 2–3 day inpatient

hospital stay. Patients received either a patient-controlled analgesia (PCA) pump or an MRI-compatible epidural for pain control during their hospital admission. Both T&O and Syed applicators were placed in the operating room under general anesthesia. For Syed applicators, the decision to use a tandem was made by the physician on a case-by-case basis, and no hybrid needles were used with the T&O applicators. Three gold fiducial markers were inserted into the cervix. Transabdominal ultrasound was used to sound the uterus for appropriate tandem placement, and for T&O placement, the ovoid size was chosen based on the width of the vaginal cavity. Standard vaginal packing was used to secure the T&O applicator in place. Fluoroscopic guidance in the operating room was used to confirm appropriate placement and symmetry of the T&O applicator or the Syed interstitial needles in relation to the gold fiducial markers in the cervix. Beginning in 2019, all patients without a contraindication to magnetic resonance imaging (MRI) underwent a pelvic MRI after applicator placement, and the MRI was fused to the CT simulation scan to help with tumor and Organs at Risk (OAR) delineation and brachytherapy planning. Prior to 2022, a Varisource iX remote afterloader (Varian Medical Systems, Palo Alto, CA) was used for brachytherapy delivery, and beginning in 2022, a Bravos remote afterloader (Varian Medical Systems, Palo Alto, CA) was used for brachytherapy treatment."

1.1.6 Workflow

Treatments for gynecologic cancers include surgery, chemotherapy, and/or radiotherapy (14). As part of radiotherapy treatment for gynecologic cancers, brachytherapy is often used as a boost to EBRT, and the addition of brachytherapy has demonstrated improved local control and survival (15,16).

A typical radiation oncology workflow consists of two parts, one for A) EBRT, and another for B) Brachytherapy.

A typical workflow for EBRT, consisting of the decision of the clinical plan, CT imaging, planning, and treatment delivery has been described in Figure 1.6:

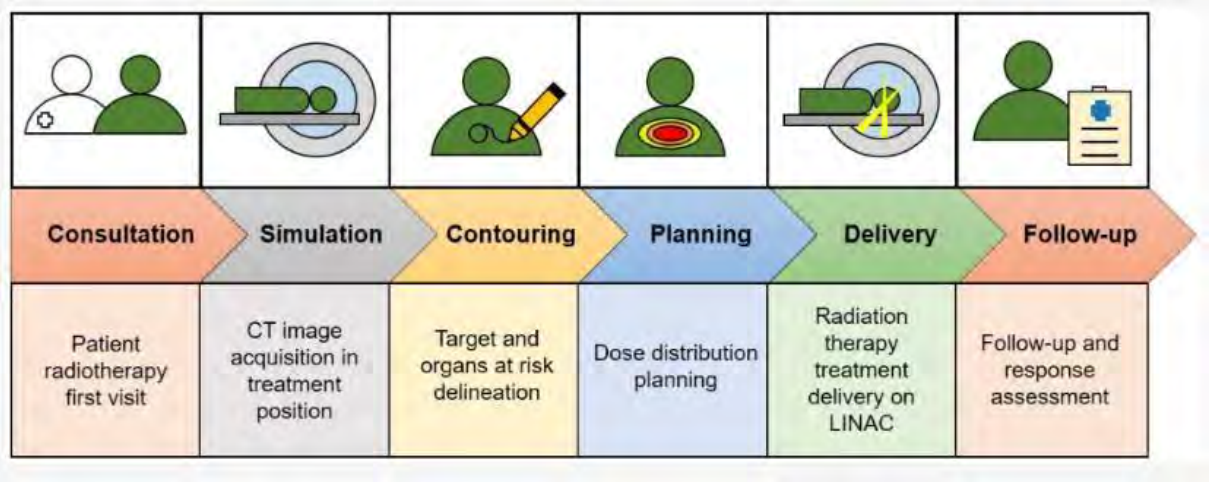


Figure 1.6. Typical EBRT workflow as described by Marvaso et al (17).

In the institution where this research was carried out, brachytherapy takes place after the completion of EBRT as defined in the Clinical Methods section. Common workflows for brachytherapy are described below:

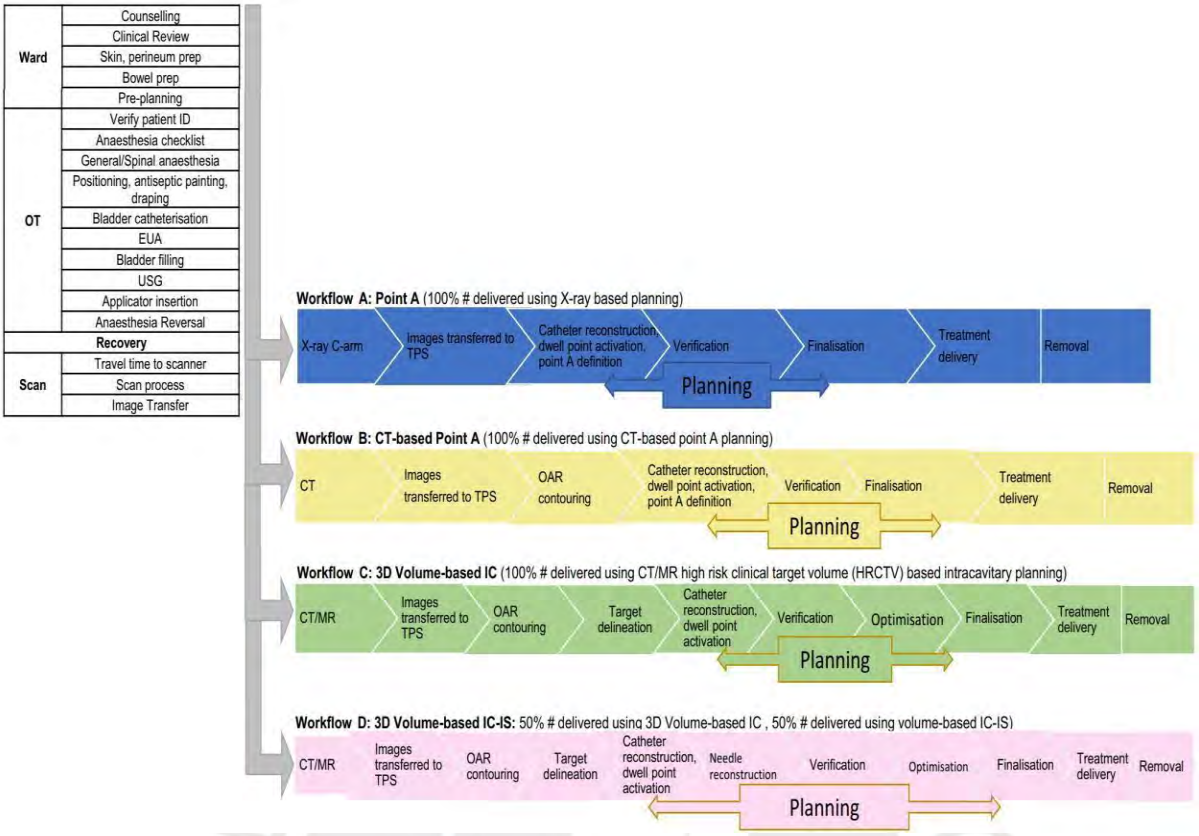


Figure 1.7. Multiple Workflow options for gynecological HDR treatments by Hande et al. (18).

1.2 Radiation therapy Associated Toxicities

1.2.1 Biological Effective Dose and Equivalent Dose in 2 Gy Fractions

The Linear-Quadratic model is the most widely accepted for cell death due to radiation. It consists of two components: A) The linear $S = e^{-\alpha D}$ which occurs due to directly lethal damage, and B) The quadratic $S = e^{-\beta D^2}$ which has the contribution of sublethal damage, which is understood as radiation-induced cellular damage that can be repaired, allowing cell survival (19).

As defined by Barendsen (20), the response due to multiple series of treatment is given by:

$$F_r = n \cdot F(d) \tag{1.10}$$

where n is the number of fractions, $F(d)$ is the frequency of effective lesions per cell after a single dose d, and F_r is the response level, at which when the average value of effective lesions of the irradiated cells reaches this value, they generate a response in a tissue.

But following the linear quadratic model, the frequency of an effect as a function of dose is also given by:

$$F(d) = \alpha d + \beta d^2 \tag{1.11}$$

Combining formulas 1.10 and 1.11 the following expression is found:

$$F_r = n \cdot (\alpha d + \beta d^2) \tag{1.12}$$

But with the relationship from equation 1.12

$$D = nd \tag{1.13}$$

It is found that:

$$F_r = D(\alpha + \beta d) \tag{1.14}$$

where D is the total dose required to attain the effect F_r, and $(\alpha + \beta d)$ equals the “effectiveness per unit dose”.

As shown by Fowler (21) from formula 1.14, the factor F_r/α can be considered a Biological Effective Dose (Gy), which is defined by:

$$BED = D\,(1 + \frac{d}{\alpha/\beta}) \tag{1.15}$$

In addition, the concept of EQD2, which is the equivalent dose as given in 2 Gy fractions (Gy), is obtained by solving:

$$D\,(1 + \frac{d}{\alpha/\beta}) = D_2\,(1 + \frac{2}{\alpha/\beta})$$

In which we obtain:

$$D_2 \equiv EQD2 = D\,(\frac{d+\alpha/\beta}{2+\alpha/\beta}) \tag{1.16}$$

EQD2 is widely used clinically for understanding the total dose that would be required to get the same effect if it was given using a fractionation scheme of 2 Gy per Fraction. The biggest challenge of using BED alone is that these numerical values are not familiar to many medical professionals; so EQD2 is a type of normalization that is easier to understand. These concepts are related as seen in equation 1.17 (22,23).

$$EQD2 = \frac{BED}{1+\frac{2}{\alpha/\beta}} \tag{1.17}$$

This equation is valid when total recovery of normal tissue between fractions is assumed (For modalities such as EBRT and HDR, where the time for half of the DNA

to be repaired is $T_{1/2} = 1.5\ hr$ and the time between fractions is much greater than that). On the contrary, if cellular sub-lethal repair which starts after 15-30 minutes is taken into consideration for cases such as Low-Dose-Rate (LDR) or Pulsed Dose Rate (PDR) brachytherapy, an additional factor g, known as the Lea-Catcheside factor, needs to be included for BED and EQD2, as displayed in equation 1.18 (24).

$$F_r = D(\alpha + g\beta d) \tag{1.18}$$

1.2.2 Therapeutic Ratio

As the dose given to a tumor is escalated, the probability of control increases; but so does the normal tissue damage (radiation side effects incidence and severity) as shown in Figure 1.8.

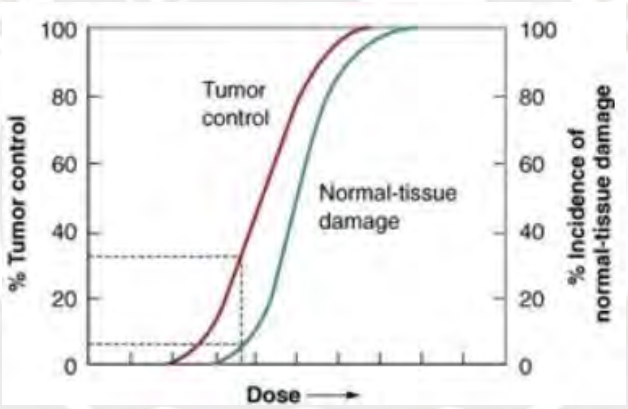


Figure 1.8. Dose response curve for tumor control and normal tissue damage as illustrated in Hall (25).

The therapeutic ratio or index relates the connection between the likelihood of tumor control and the probability of damage to healthy tissue and is denoted by equation 1.19 (26).

$$TR = \frac{Damage\ to\ tumor\ cells}{Damage\ to\ normal\ cells}\Big|_{Equal\ Dose} \tag{1.19}$$

1.2.3 Dose Constraints

Dose-Volume constraints are values routinely used in clinical practice to estimate the risk of toxicities. While multiple dose-volume histograms (DVHs) may meet constraint thresholds, each potentially associated with different risk probabilities, it is only through a multi-point and visual dose analysis that acceptable isodose distributions are achieved and the risk-benefit balance assessed during treatment planning. These constraints do not account for clinically relevant prognostic factors or organ-specific spatial information. Historically, these constraints have been associated with magnitudes such as TD5/5 and TD50/5 indicating the dose levels leading to a 5% risk of severe complications over 5 years and a 50% probability of such complications within the same timeframe post-irradiation. Other publications such as the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) report constraints for the "generally clinically acceptable" rates of injury, specifying the incidence rate, and the endpoint (27–30).

Reports made by the American Society for Radiation Oncology, American Brachytherapy Society, American Society for Clinical Oncology, Society for Gynecologic Oncology and others provide guidance with respect to the ideal and maximum dose constraints to the OARs adjacent to gynecological cancer treatments, for a combination between EBRT and Brachytherapy. Some of the most recent consensus values for these OARs are shown in Figure 1.9.

Organ at risk	Ideal dose Constraint (cGy) ($EQD2_3$)	Maximum* dose constraint (cGy) ($EQD2_3$)	ICRU point (cGy) ($EQD2_3$)	References
Rectum	<6500 D2cc	<7500 D2cc	<7500 point dose	95,102,115,117,118
Bladder	<8000 D2cc	<9000 D2cc	<9000 point dose	115,117-120
Vagina (recto-vaginal point)†	<6500 point dose	<7500 point dose	---	96,116
Sigmoid‡	<7000 D2cc	<7500 D2cc	---	120
Bowel‡	<7000 D2cc	<7500 D2cc†	---	120,121

Figure 1.9. Dose constraints for cervical cancer patients as reported in the ASTRO Guideline on Radiation Therapy for Cervical Cancer (31)

1.3 Machine Learning

As seen in Portocarrero et al., machine learning is fundamentally an optimization problem focused on discovering the most appropriate predictive model for a particular dataset. The recent surge in machine learning's popularity can be attributed to several key developments: the emergence of novel algorithms, advancements in theoretical understanding, expanded data availability, and significant improvements in low cost computational power. In numerous critical medical applications, such as predicting heart failure readmission and mortality, forecasting hepatocellular carcinoma development, and diagnosing glaucoma, machine learning has consistently proven to outperform conventional statistical models in terms of overall predictive metrics. (33–35).

1.3.1 Scoring Metrics

There exists a diverse array of performance metrics for evaluating a model's predictive capability when confronted with previously unseen data. In this particular research, the investigators employed a comprehensive set of evaluation metrics, which included Accuracy, Precision, Recall, F1 score, Matthews Correlation Coefficient (MCC), the area under the receiver operating characteristic curve (AUC ROC), and the area under the precision recall curve (AUC PR). The first four metrics—Accuracy, Precision, Recall, and F1 score—are defined based on the counts of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), as detailed below:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1.20) \qquad Precision = \frac{TP}{TP+FP} \quad (1.21)$$

$$Recall = \frac{TP}{TP+FN} \quad (1.22) \qquad F1\,Score = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \quad (1.23)$$

$$MCC = \frac{(TP \cdot TN)-(FP \cdot FN)}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}} \quad (1.24)$$

In this context, "Positive" or "Negative" describes whether the machine learning model predicts a toxicity event, while "True" or "False" indicates if the model's prediction aligns with or disagrees from the actual patient outcome. Accuracy, as described in formula 1.20, is the proportion of correctly classified instances relative to the total number of patients. Precision, given by formula 1.21, measures the ratio of correctly

identified positive cases to all cases predicted as positive. Recall, or Sensitivity, shown in formula 1.22, calculates the proportion of correctly predicted toxicities among patients who actually experienced them. The F1 score, defined in formula 1.23, provides a combined measure of Precision and Recall, representing their harmonic mean (36). The Matthews Correlation Coefficient (MCC) as shown in formula 1.24, or its normalized form (normMCC), offers a balanced evaluation by incorporating all four key components: TP, FP, TN, and FN (37). Complementing these metrics, AUC-ROC and AUC-PR serve as comprehensive performance indicators that assess the model's predictive capabilities across all potential decision thresholds, offering a holistic view of its classification effectiveness. (38).

1.3.2 Models

Logistic Regression (LR): This model can determine the influence of one or multiple factors, based on the logit odds (represented by the relationship between the probability of the event happening and the probability of non-occurrence, expressed as a ratio) of a 2-level outcome problem, as modeled in equation 1.25 (39,40).

$$Logit(Y) = natural\log(odds) = \ln\left(\frac{Y}{1-Y}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_i x_i \qquad (1.25)$$

where

$$Probability\ of\ outcome(Y_i) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_i x_i}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_i x_i}}$$

and Y_i represents the probability of being in one of two possible outcomes (i) and $e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots\beta_i x_i}$ represents a linear fit equation in the logit scale, to guarantee that the probability range will be between 0 and 1 (41).

Random Forest (RF): It is an ensemble model in which multiple decision trees are generated and then vote for the most popular class. The sci-kit learn library has modified this algorithm to average their probabilistic prediction instead of letting the classifiers vote. The following mathematical definitions for RF and Decision Tree

classifiers are defined in Breiman and the Scikit Learn documentation (42–44). RF classifiers are “the collection of tree-based classifiers” and defined by:

$$\{h(\boldsymbol{x}, \Theta_k),\ \mathrm{k}\ =\ 1, \ldots)\} \tag{1.26}$$

where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input **x**.

These tree-based models are based on training vectors $x_i \in R^n$, $i = 1, \ldots, l$ and label vectors $y \in R^l$ where a decision tree divides the feature space repeatedly, organizing samples with identical or closely related target values into groups, by using the following formalism:

At each node “m”, the data Q_m with n_m samples is partitioned into two groups $Q_m^{left}(\theta)\ and\ Q_m^{right}(\theta)$, where $\theta = (feature, threshold\ for\ partition) = (j, t_m)$, and the group elements are defined by equation 1.27 and 1.28.

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\} \tag{1.27}$$

$$Q_m^{right}(\theta) = Q_m \backslash Q_m^{left}(\theta) \tag{1.28}$$

The candidate split will then be assessed by equation 1.29 by using an impurity function $H()$:

$$G(Q_m, \theta) = \frac{n_m^{left}}{n_m} H(Q_m^{left}(\theta)) + \frac{n_m^{right}}{n_m} H(Q_m^{right}(\theta)) \tag{1.29}$$

Which selects parameters to minimize impurity by using equation 1.30.

$$\theta^* = argmin_\theta G(Q_m, \theta) \tag{1.30}$$

This algorithm is going to continue splitting nodes until $n_m < min_{samples}$ or $n_m = 1$

K-Nearest Neighbors (KNN): It is an algorithm where unassigned data is categorized by determining the class most heavily represented among its k_n neighbors (45).

The mathematical formalism of this algorithm is summarized by Li et al (46). In a classification problem let $T = \{x_m \in R^n\}_{m=1}^{n}$ be a training set with M samples in a feature space with n-dimensions, with l class labels, with each sample x_n having a class label c_n associated to it.

In this case, the kNN algorithm will first compute the distance between points by using equation 1.31 and finding the set of k nearest neighbors $Neighbor\{x_1, x_2, \dots, w_k\}$ for an unclassified point y.

$$d(y, x_i) = \sqrt{(y - x_i)^T (y - x_i)} \tag{1.31}$$

To assign a label based on this set of neighbors, two common strategies are shown in equations 1.32 and 1.33.

$$c = arg \max_i \sum_{x_j \in Neighbor} sign(x_j, c_i) \tag{1.32}$$

or

$$c = arg \max_i \sum_{x_j \in Neighbor} Sim(y, x_j) sign(x_j, c_i) \tag{1.33}$$

where $Sim(y, x_j)$ is a function to measure how similar y and x_j are, and $sign(x_j, c_i)$ indicates to which class x_j belongs and can take the value of 0 or 1.

Support Vector Machines (SVM): As explained by Awad & Khanna and Noble (47,48), this algorithm is composed by four key elements which include the following:

1) Separating hyperplane: A straight line in a high-dimensional space that separates classes.

2) Maximum-margin hyperplane: A separating hyperplane defined by maximizing the margin between the classes.
3) Soft margin: Allows the algorithm to include certain data points to reside within the margin of the separating hyperplane. This allows outliers to fall in the “incorrect” part of the separation boundary.
4) Kernel function: It is a mathematical operation that adds an additional dimension to the data which could result in it becoming separable by a hyperplane in the newly generated space.

In a training set $\{x_i, y_i\}$ where $i = 1,2, \dots, l$ and each sample has d features such that $x_i \in R^d$, and 2 classes $y_i \in \{-1,1\}$. Then all hyperplanes in R^d are parametrized by a vector w orthogonal to the hyperplane, and a constant b as shown in equation 1.34.

$$\boldsymbol{w^T} \cdot \boldsymbol{x} + b = 0 \tag{1.34}$$

Given the defined hyperplane from equation 1.34, the function for classification can be defined as follows in equation 1.35.

$$g(x) = \boldsymbol{w^T} \cdot \boldsymbol{x} + b \tag{1.35}$$

In this case, $\boldsymbol{w^T} \cdot \boldsymbol{x} + b = 1$ for $x \in y_1$ and $\boldsymbol{w^T} \cdot \boldsymbol{x} + b = -1$ for $x \in y_2$ which when we include a soft margin leads to the optimization problem that requires solving as shown in equation 1.36:

$$J(w, b, \zeta) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{N} \zeta_i \tag{1.36}$$

Subject to the following constraints:

$$y_i[\boldsymbol{w_i^T} \cdot \boldsymbol{x_i} + b] \geq 1 - \zeta_i, \qquad i = 1,2, \dots, N$$
$$\zeta_i \geq 0, \qquad i = 1,2, \dots, N$$

Gaussian Naïve Bayes (GNB):

As stated in Zhang et al. and Perez et al. (49,50), Bayes' theorem states that the relationship between a class y and feature vectors x_1 to x_n is given by:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1,\ldots,x_n|y)}{P(x_1,\ldots,x_n)} \tag{1.37}$$

By assuming naïve conditional independence, equation 1.37 is simplified to:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1,\ldots,x_n)} \tag{1.38}$$

As the term in the denominator is constant, then the classification rule is defined as:

$$P(y \mid x_1, \ldots, x_n) \propto P(y)\prod_{i=1}^{n} P(x_i \mid y) \tag{1.39}$$

The class that maximizes the joint probability of observing the features and the class label is found by using equation 1.40.

$$\hat{y} = \arg\max_{y} P(y)\prod_{i=1}^{n} P(x_i \mid y) \tag{1.40}$$

For the case of Gaussian Naïve Bayes, the likelihood of the features is defined by equation 1.41:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \tag{1.41}$$

where σ_y is the variance and μ_y is the mean of the class y.

Multi-Layer Perceptron Neural Network (MLP):

As shown in Hossam et al. and Xingyang et al. (51,52), in a training set $\{x_i, y_i\}$ where $i = 1,2,\ldots,n$ and each sample has m features such that $x_i \in R^m$, and 2 classes $y_i \in \{0,1\}$ the architecture for a MLP model consists of input layers in which there are no computations and simply pass the input feature to the hidden layers; hidden layers

in which elements called neurons apply an activation function to a weighted sum of the inputs and an output layer which usually consist of one element that represents the probability of belonging to a class.

The summation function of the neuron is defined by:

$$S_j = \sum_{i=1}^{m} w_{ij} I_i + \beta_j \tag{1.42}$$

where I_i is the ith-input, w_{ij} is the weight connecting I_i to the neuron, β_j is a bias weight, and m is the number of neuron inputs.

As mentioned above, an active function like the logistic function will be applied to this result to obtain outputs between zero and one.

$$f(x) = \frac{1}{(1+e^{-x})} \tag{1.43}$$

The output of neuron j will be defined by

$$y_j = f_j(\sum_{i=1}^{m} w_{ij} I_i + \beta_j) \tag{1.44}$$

The MLP is optimized for classifications problems by using different metrics such as the Average Cross-Entropy.

$$Loss(\hat{y}, y, W) = -\frac{1}{n}\sum_{i=1}^{n} (y_i ln\hat{y}_i + (1 - y_i)\ln(1 - \hat{y}_i)) + \frac{\alpha}{2n}\|W\|_2^2 \tag{1.45}$$

where $\alpha\|W\|_2^2$ is a L2-regularization term, α is a positive hyperparameter that controls the magnitude of the penalty, y is the actual value, $\hat{y}$ is the estimated value, and n is the number of samples of the training dataset.

Algorithms like gradient descent compute the gradient $\nabla Loss_W$ with respect to the weight and subtract it from W until the algorithm reaches a set value of iterations, or

when the improvement in loss reaches a certain threshold. This is repeated in i iteration steps as shown by:

$$W^{i+1} = W^{i} - \varepsilon \nabla Loss^{i}{}_{W} \tag{1.46}$$

where ε (> 0) is the learning rate.

1.3.3 Balancing Techniques

In scenarios where the training dataset does not have equally distributed classes (50% Negative, 50% Positive outcome), an oversampling balancing technique can be used to create synthetic samples for the minority class; with the goal of obtaining a better model performance. Some of these algorithms include the Synthetic Minority Over-sampling Technique (SMOTE), Support Vector Machine - Synthetic Minority Over-sampling Technique (SVM-SMOTE) and ADASYN (Adaptive Synthetic Sampling) (53).

Readers seeking a thorough understanding of these algorithms can access the original publications by Chawla et al. (54), Nguyen et al. (55), and He et al. (56); a summary of the algorithm descriptions as found in these references are shown in Appendix C.

1.3.4 Feature Selection/Extraction

Feature selection and extraction are commonly used in situations where many variables are available, but not all are relevant for a specific situation. Feature selection methods choose a subset of features such that only the most relevant ones are used for prediction, whereas feature extraction involves transforming the original data using mathematical techniques to reduce the dimensionality of the problem and generate new, more significant features. The goal of these two methods is to simplify models to make them more interpretable, reduce training time, decrease the probability of overfitting and to remove redundant data (57).

Common feature selection methods include A) Filters, which select variables based on their statistical significance with the target variable, B) Wrapper methods, that find the best sub-set of variables based on the training performance (Such as Sequential Feature Selection) and C) Embedded methods, which incorporate the feature selection process as part of the training of the model (58,59)

On the other hand, feature extraction methods include algorithms such as A) Principal Component Analysis, which is a non-parametric method that applies a linear transformation to the data to minimize redundancy and B) Linear Discriminant Analysis, which finds a transformation to reduce the dimensionality of the feature vectors that maximizes class separability, etc. (60)

1.4 Cancer Epidemiology

1.4.1 Gynecological Cancer

Gynecological cancers represent a significant global health challenge, being among the most prevalent malignancies impacting women worldwide (61). According to current projections for the United States, the year 2024 is expected to witness approximately 116,930 new gynecological cancer diagnoses, accompanied by an estimated 36,250 fatalities resulting from these malignant conditions. (62).

1.4.2 Cervical Cancer

In 2024, an estimated 13,820 women in the United States will be diagnosed with cervical cancer, leading to around 4,360 deaths (62). Since the early 1970s, the incidence of cervical cancer in the U.S. has been decreasing (63) due to increased Papanicolaou (Pap) screening and human papillomavirus (HPV) vaccination (64,65). However, the mortality from cervix cancer has remained largely unchanged over the past 50 years (66). On a global scale, cervical cancer ranks fourth in frequency among women, with about 604,127 new cases and 341,831 deaths reported in 2020 (67).

1.5 Research Aim and Objectives

Accurately forecasting toxicities in gynecologic cancer patients receiving external beam radiation therapy (EBRT) alongside high-dose-rate (HDR) brachytherapy is key to refining treatment strategies and minimizing side effects. Routine clinical practice uses specific dosimetric variables as predictors. While practical, this approach is insufficient for enabling patient-specific radiation therapy treatments and is limited in its predictive power to dose metrics values. Supervised machine learning offers a promising solution, utilizing various algorithms to analyze and predict outcomes based on intricate patterns within the data.

This work aims to first gather data to support the hypothesis that not only dosimetric values are useful for predicting toxicities. A conventional statistical comparison was conducted for cervical cancer patients who received 3D-based brachytherapy, with a volumetric prescription to the HR-CTV, using two commonly used applicators. Doing this analysis, only for cervical patients, will be useful for assessing any dosimetric and toxicity differences between groups. Due to the patients having the same type of cancer, this study will explore if toxicities might be related not only to dose but also to other variables such as tumor volume and stage. Moreover, by not including patients that had a point-based prescription, which was commonly used in the past, the assessment of modern-day brachytherapy treatments and their outcomes can be done.

This study will leverage a dataset of 233 patients to develop and compare various machine learning models aimed at predicting grade 3 or higher toxicities in gynecologic cancer patients undergoing EBRT and HDR brachytherapy. The entire dataset, one of the largest to date which spans over 12 years, will be used to train machine learning models to investigate their utility in predicting toxicities. This approach involves incorporating demographic, treatment, and tumor-related variables; employing data-balancing techniques, feature selection such as sequential feature selection and rigorous cross-validation to enhance model generalizability and performance. In contrast to previous studies, multiple machine learning models will be trained with one of the largest datasets for this field of study. Additionally, an in-depth analysis of the workflow and scoring metrics will be provided.

The goal is to identify the most effective predictive model to be able to provide personalized and precise radiation therapy plans, ultimately improving patient outcomes. Moreover, these models will also present more information to physicians to enable them to better communicate risks effectively to patients.

1.5.1 Main Objectives

Estimate the occurrence of different grade toxicities in patients that undergo high dose rate brachytherapy treatments for gynecological cancers through multi-variable machine learning-based predictive models.

1.5.2 Specific Objectives

1. To analyze and compare the dosimetric profiles and toxicity outcomes of cervix cancer patients that have undergone HDR brachytherapy with Syed-Neblett and Fletcher-Suit-Delclos Tandem & Ovoid (T&O) applicators, with a focus on the impact of non-dosimetric factors in predicting toxicity.
2. Analyze non-dosimetric features such as age at completion, BMI, Charlson Comorbidity Index, Karnofsky Performance Status (KPS), length of treatment, type of applicator, concurrent chemotherapy, an additional boost of EBRT, tumor size, stage, histology, MRI assisted planning and volume of the HR-CTV as being potential indicators for different grade toxicities using ML models.
3. Compare how well machine learning algorithms predict secondary toxicities associated with brachytherapy in contrast to traditional statistical models.
4. Do an in-depth analysis of which scoring metrics (accuracy, precision, recall, F1 scores and specificity) are appropriate when comparing the predictive power of machine learning model for toxicity prediction.

Chapter 2

Dosimetric and Toxicity Comparison Between HDR Applicators

2.1. General Remarks

This chapter is based on Portocarrero et al (13). The standard approach for treating locally advanced cervical cancer involves EBRT with concurrent cisplatin-based chemotherapy followed by brachytherapy (68,69). Cause-specific survival and overall survival are significantly improved in FIGO stage IB2-IVA cervix cancer patients treated with both EBRT and brachytherapy compared to treatment with EBRT alone (70). Brachytherapy can be administered using Syed-Neblett and tandem & ovoid applicators (71,72). However, there may be dosimetric differences between these applicators due to variations in the spatial arrangement of the dwell positions with each design or because of differences in the surgical implantation procedure between intracavitary and interstitial brachytherapy (73).

While intracavitary and interstitial brachytherapy are beneficial for target coverage and sparing normal tissue (74,75), they can cause secondary toxicities such as fistulas, vaginal stenosis, bleeding, local fibrosis, ulceration, necrosis, and incontinence (76). These side effects can significantly impact patients´ quality of life (QoL) (77–80). The GEC ESTRO suggests reporting dose-volume parameters of the organs at risk (OARs) for predicting morbidities (81).

This study aims to describe and compare the dosimetric profiles and toxicities in cervical cancer patients treated with Syed-Neblett versus Fletcher-Suit-Delclos Tandem & Ovoid (T&O) applicators.

2.2. Patient Characteristics and Study Design

A retrospective chart review was performed for patients who underwent volume based (3D) high dose rate (HDR) brachytherapy for cervical cancer from 2011 - 2023 at a single institution. Collected data included demographics, tumor and treatment variables, dosimetric data, and toxicities developed during and after treatment.

Patients were excluded if they had prior brachytherapy, were treated with more than one brachytherapy applicator, had missing dosimetric information, had a dose prescription to point A or H, were undergoing EBRT for a different nearby disease site or had a combination of LDR and HDR treatments. For the assessment of toxicities, an exhaustive review of documentation was conducted, encompassing records from various departments within the institution. This included comprehensive examination of documentation from radiation oncology, medical oncology, gynecologic oncology, the emergency department, nursing notes, as well as from other relevant departments within the hospital; external data from other physicians was also considered. To maintain data integrity, the database was reviewed three times by both a physician and a medical physicist. The study received approval from the institutional review board (IRB 22.0117).

Both Syed-Neblett and Fletcher-Suit-Delclos Tandem & Ovoid implants are widely used in HDR brachytherapy, but they show differences in dosimetric profiles and clinical outcomes. This study examines these differences and their effects on treatment planning and side effect management. To the author´s knowledge, this is the first known comparison of dosimetry, toxicities and local control between T&O and Syed-Neblett applicators in the management of cervical cancer.

Modern volumetric based brachytherapy calculates a 3D dose distribution and optimizes dwell positions based on acceptable volumetric target coverage and OAR constraints, as those suggested in the ASTRO Guideline on Radiation Therapy for Cervical Cancer (82,83). A study by Paul et al., involving 48 cervical cancer patients initially planned to point A and re-normalized to the HR-CTV, demonstrated an 8 - 37% reduction in the D2cc metrics for the OARs per brachytherapy fraction with an acceptable dose coverage by today’s standards (84). Due to higher doses to OARs with point A based planning and because this method is no longer utilized in our clinical practice, this analysis only evaluated patients who underwent volumetric (3D) based brachytherapy to appropriately compare Syed and T&O brachytherapy.

2.3. EQD2 Calculation

To calculate the total radiation dose received, combining EBRT and brachytherapy, the following process was followed:

First, the biologically effective dose (BED) was calculated using the linear - quadratic formula (1):

$$BED = nd\ (1\ +\ \frac{d}{\alpha/\beta}) \qquad (1)$$

After obtaining the BED, it was converted into the equivalent dose in 2 Gy fractions by using formula (2):

$$EQD2 = \frac{BED}{(1+\frac{2}{\alpha/\beta})} \qquad (2)$$

This calculation was performed using an Eclipse Scripting Application Programming Interface (ESAPI) program developed at this institution (Figure 2.1), with the following assumptions:

- The alpha/beta ratio was 10 for the HR-CTV.
- The alpha/beta ratio was 3 for the OARs.
- The full prescription of the EBRT was delivered to the OARs and the HR-CTV (for both primary and boost treatments).
- An EBRT boost was only considered for the calculation if it was delivered to the cervix, parametrium or uterus.
- An EBRT boost to the lymph nodes was not included in the EQD2 calculation.
- Image registration was not performed for brachytherapy plans to correct for inter-fraction movement of OARs and HR-CTV in plans with multiple CT scans.

The ESAPI program calculates the EQD2 dose after inserting the EBRT prescription dose and number of fractions, selecting the brachytherapy plans and inputting their respective number of fractions and matching the structure names in the program display to the plan structures. As recommended by ICRU 89 and the GEC-ESTRO report, for the HR-CTV the $V_{100\%}$, $D_{50\%}$, $D_{90\%}$ and $D_{98\%}$ values were compiled, whereas for the OARs the $D_{0.1cc}$, D_{1cc} and D_{2cc} values were collected (76,81).

Figure 2.1. Display of the ESAPI script used to calculate the EQD2 doses for the target volume and organs at risk.

2.4. Statistical Analysis

Patient characteristics were analyzed by comparing demographics and treatment variables between the Syed and T&O applicators. Categorical variables were presented as frequency counts and percentages, with statistical comparisons conducted using the Fisher exact test. Continuous variables that did not follow a normal distribution were expressed as medians with accompanying interquartile ranges (IQR), and statistical differences were evaluated using the non-parametric Mann-Whitney test. Toxicities affecting the gastrointestinal (GI), genitourinary (GU), and vaginal systems were systematically graded according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0, with a comprehensive five-tier severity scale ranging from G1 (Mild) to G5 (Death). (85). Acute toxicity grade was determined by the most severe toxicity within 6 months of treatment, while late toxicity consisted of toxicities observed after 6 months. The presence of toxicity (grade one or higher vs. no toxicity) by applicator was compared using the Fisher exact test;

and the distribution of toxicities (among those with grade one or higher toxicity) was compared between groups using the Mann-Whitney test. Local control was assessed using standard Kaplan-Meier methods and compared between applicators using the log-rank test. Throughout, a significance level of alpha = 0.05 was used to define statistical significance, with hypothesis testing conducted using p-values < alpha as the criterion to establish statistical significance. Missing data was handled by performing an available case analysis that included all patients with data for the given analysis. Statistical analysis was performed using R statistical software (86), version 4.2.2.

2.5. Demographic, treatment, and tumor comparison

The demographic, treatment, and tumor characteristics detailed in Table 2.1 revealed several significant differences between patients receiving Syed and T&O brachytherapy. Patients treated with T&O brachytherapy exhibited notably smaller tumors, with a mean tumor diameter of 4.6 cm compared to 6.0 cm in the Syed group ($p < 0.001$). Similarly, the median high-risk clinical target volume (HR-CTV) was substantially lower in the T&O group at 28 cc versus 54 cc in the Syed group ($p < 0.001$). Notably, patients undergoing Syed brachytherapy demonstrated a significantly higher proportion of advanced disease, with 76.5% presenting at FIGO stages 3 or 4, in contrast to 34.4% in the T&O brachytherapy group ($p < 0.001$). Furthermore, the median follow-up period differed markedly, with Syed brachytherapy patients having a shorter follow-up of 4.3 months compared to 8.8 months for T&O brachytherapy patients ($p = 0.013$). Moreover, 46% of Syed brachytherapy patients had follow-up durations exceeding 6 months, compared to 64% of those treated with T&O brachytherapy.

	Syed		T&O		
	n=115	66%	n=58	34%	p-value
Follow-up (months)	4.3	IQR: [1.2 – 18.2]	8.8	IQR: [3.2 – 20.9]	**0.013**
Age at Completion	52.3	STD: 13.7	49.7	STD: 14	0.256
Non-Caucasian	13	11%	6	10%	1.000
BMI	26.8	STD: 8.6	28.5	STD: 7.5	0.181
Charlson Comorbidity Index	3.5	STD: 1.7	3.3	STD: 1.7	0.763
KPS	87	STD: 11	88	STD: 10	0.604
Concurrent Chemo	106	92%	57	98%	0.168
Tumor Size (cm)	6.0	STD: 2.2	4.6	STD: 1.7	***< 0.001***
HR-CTV Volume (cc)	54	IQR: [32 - 100]	28	IQR: [20 - 43]	***< 0.001***
Cancer Stage					***< 0.001***
Stage 0	1	1%	0	0%	
Stage 1	8	7%	19	33%	
Stage 2	18	16%	19	33%	
Stage 3	72	63%	18	31%	
Stage 4	16	14%	2	3%	
MRI Fused	66	57%	30	52%	0.519
Number of Needles	13.7	STD: 3.5			

Table 2.1. Demographic, treatment and tumor variables of patients treated with Syed-Neblett and Tandem and Ovoid applicators.

2.6. Dosimetric Analysis

As shown in Table 2.2, the dosimetric parameters V100%, D50%, D90%, and D98% of the HR-CTVs showed no statistically significant differences in target coverage

between Syed and T&O brachytherapy. However, there were significant differences in the dosimetric values for the OARs:

- For the bladder, Syed brachytherapy resulted in significantly lower median D0.1cc, D1cc, and D2cc values (95.9 Gy EQD2 vs. 107.6 Gy EQD2, $p = 0.003$; 83.6 Gy EQD2 vs. 90.4 Gy EQD2, $p < 0.001$; and 78.6 Gy EQD2 vs. 84.3 Gy EQD2, $p < 0.001$, respectively).

- For the small bowel, Syed brachytherapy also resulted in significantly reduced median D0.1cc, D1cc and D2cc values (66.0 Gy EQD2 vs. 76.7 Gy EQD2, $p = 0.03$; 60.6 Gy EQD2 vs. 67.0 Gy EQD2, $p = 0.027$; and 58.2 Gy EQD2 vs. 63.8 Gy EQD2, $p = 0.025$, respectively).

- For the sigmoid colon, the median D2cc was significantly lower with Syed applicators (64.2 Gy EQD2 vs. 66.7 Gy EQD2, $p = 0.046$). Other dosimetric values to the sigmoid colon were lower with Syed brachytherapy but not statistically significant.

- For the rectum, the median D0.1cc, D1cc, and D2cc doses were higher with Syed applicators; with only the D0.1cc showing statistical significance (83.3 Gy EQD2 vs. 79.7 Gy EQD2, $p = 0.013$).

	Syed N=115 Median	IQR	T&O N=58 Median	IQR	p-value
HR-CTV V100 (cc)	70.0	[51 - 119.2]	88.3	[65.3 - 131.3]	0.066
HR-CTV D50 (Gy)	113.2	[104.1 - 119.9]	112.5	[106.5 - 118.4]	0.982
HR-CTV D90 (Gy)	84.3	[80.5 - 87.7]	82.3	[80.3 - 85.5]	0.154
HR-CTV D98 (Gy)	75.2	[70.6 - 78.9]	74.3	[72.8 - 77.4]	0.999
D0.1cc - Bladder (Gy)	95.9	[83.5 - 110.1]	107.6	[92.9 - 122.7]	***0.003***
D1cc - Bladder (Gy)	83.6	[76.2 - 92]	90.4	[80.6 - 102.7]	***< 0.001***
D2cc - Bladder (Gy)	78.6	[72.3 - 85]	84.3	[75.5 - 94.7]	***< 0.001***
D0.1cc - Small Bowel (Gy)	66.0	[53.8 - 81.9]	76.7	[58 - 88.1]	***0.030***
D1cc - Small Bowel (Gy)	60.6	[51.8 - 70.9]	67.0	[55.4 - 77.1]	***0.027***
D2cc - Small Bowel (Gy)	58.2	[50.8 - 68]	63.8	[54.3 - 72.8]	***0.025***
D0.1cc - Sigmoid Colon (Gy)	74.5	[62.2 - 87.7]	79.0	[71.9 - 84.8]	0.170
D1cc - Sigmoid Colon (Gy)	67.5	[57.5 - 75.5]	70.0	[64.5 - 76]	0.053
D2cc - Sigmoid Colon (Gy)	64.2	[56 - 71.7]	66.7	[61.9 - 71.6]	***0.046***
D0.1cc - Rectum (Gy)	83.3	[76.1 - 89.2]	79.7	[74.6 - 83.8]	***0.013***
D1cc - Rectum (Gy)	73.2	[65.8 - 79.3]	70.6	[67.3 - 73.7]	0.099
D2cc - Rectum (Gy)	69.8	[61.8 - 75]	66.8	[63.7 - 70.4]	0.258

Table 2.2. Tumoral target coverage and dosimetric values to the OARs for Syed and T&O planning.

Syed brachytherapy demonstrated equal or improved dosimetry to the surrounding OARs compared to T&O brachytherapy although patients treated with a Syed applicator tended to have larger tumors and more advanced FIGO stage 3 or 4 disease. The median doses to the rectum were lower with T&O applicators, but only the D0.1cc value was statistically significant. The lower median dose to the rectum with T&O brachytherapy may be attributed by this institution standard practice of packing the vagina with gauze to maintain the T&O applicators in place and to create a separation between the T&O applicator and the rectum (87) as seen in Figure 2.2.

When compared with other studies, the local tumor control, dose metrics and toxicities for both applicators were within the reported ranges. The local tumor control rate of 84.4% at 3 years post-treatment in this study closely aligns with the findings of the EMBRACE-I trial (88). For Syed applicators, the bladder dose was reported to be the highest by Hanania et al. (89) and Kim et al. (90); which contrasts with Thibault et al. (91) who found higher rectum EQD2 doses. For T&O brachytherapy, the D2cc for the bladder was also the highest among the OARs, consistent with the research by Levin et al. (92), Rangarajan (93), Atasever Akkas and Bedri Altundag (94). The toxicities observed in this work were consistent with reports by Koh et al. (95) and Romano et al. (96); they were on the lower end of the range reported by the American Brachytherapy Society (97) for acute toxicities and on the higher end of the range for late toxicities. In accordance with the GEC-ESTRO findings from the EMBRACE trials, MRI-based brachytherapy was associated with a decreased risk of severe toxicities (98).

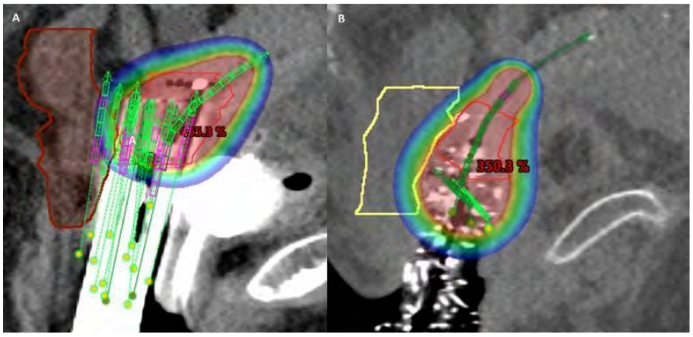


Figure 2.2. A) A Syed applicator with no separation between the applicator and rectum and B) a Tandem and Ovoid applicator with vaginal packing using radio-opaque gauze separating the applicator from the rectum.

2.7. Local Control and Toxicity

There were statistically significant differences in acute GI toxicities, as shown in Table 2.3. Syed patients were significantly more likely to have no GI toxicities within 6 months of treatment compared to T&O patients (44% vs 21%, p = 0.003). Among those who did develop acute toxicities, most were grade 1 or 2 and there was no significant difference in toxicity grade between the applicators (p = 0.457). For toxicities occurring 6 months or more post-treatment, a larger percentage of T&O patients were free of toxicities (51% vs 32%, p = 0.082). Among those who experienced late GI toxicities, there was a trend toward a significant difference with Syed patients having more grade 3 or higher GI toxicities (33% vs 17%, p = 0.058).

Similar trends were seen for GU toxicities. Within 6 months, Syed patients were significantly more likely to be free of any genitourinary toxicity (58% vs 36%, p = 0.010). Acute toxicities were generally grade 1 or 2, but Syed patients had a higher rate of grade 2 or higher GU toxicities compared to T&O patients (p = 0.009). After 6 months, both applicators demonstrated statistically similar rates of GU toxicities of any grade (Syed 51% vs T&O 35%, p = 0.196). Among the patients who developed late GU toxicities, a significantly higher percentage of Syed patients experienced grade 3 or higher toxicities compared to T&O patients (52% vs 15%, p = 0.014). The risk of developing any GU toxicity at any time was significantly lower for patients who underwent MRI-based brachytherapy compared to those receiving CT-based brachytherapy with either Syed (p = 0.024) or T&O applicators (p = 0.004), as shown in Appendix D.

Regarding vaginal toxicities, Syed patients were more likely to be free of any acute vaginal toxicity (60% vs 33%, p = 0.001); but both applicators had a similar distribution of acute toxicity grades. There were no statistically significant differences in the rate of late vaginal toxicities between applicators; though Syed patients had a higher rate of any late toxicity (70% vs 57%, p = 0.264). Among brachytherapy patients that experienced late vaginal toxicity of any grade, Syed patients were more likely to have grade 3 or higher vaginal toxicity (43% vs 24%, p = 0.342).

		Combined		Syed		T&O		p-value for difference
		n	%	n	%	n	%	
Acute GI	No Toxicity	63	36	51	44	12	21	***0.003***
	Grade 1 or higher	110	64	64	56	46	79	
	1	52	47	32	50	20	43	
	2	48	44	27	42	21	46	0.457
	3	10	9	5	8	5	11	
Late GI	No Toxicity	36	40	17	32	19	51	0.082
	Grade 1 or higher	54	60	36	68	18	49	
	1	23	43	12	33	11	61	
	2	16	30	12	33	4	22	0.058
	3	14	26	11	31	3	17	
	4	1	2	1	3	0	0	
Acute GU	No Toxicity	88	51	67	58	21	36	***0.010***
	Grade 1 or higher	85	49	48	42	37	64	
	1	51	60	23	48	28	76	
	2	33	39	24	50	9	24	***0.009***
	3	1	1	1	2	0	0	
Late GU	No Toxicity	50	56	26	49	24	65	0.196
	Grade 1 or higher	40	44	27	51	13	35	
	1	17	43	8	30	9	69	
	2	7	18	5	19	2	15	***0.014***
	3	15	38	13	48	2	15	
	4	1	3	1	4	0	0	
Acute Vaginal	No Toxicity	88	51	69	60	19	33	***0.001***
	Grade 1 or higher	85	49	46	40	39	67	
	1	54	64	26	57	28	72	
	2	24	28	17	37	7	18	0.247
	3	6	7	3	7	3	8	
	4	1	1	0	0	1	3	
Late Vaginal	No Toxicity	32	36	16	30	16	43	0.264
	Grade 1 or higher	58	64	37	70	21	57	
	1	26	45	16	43	10	48	
	2	11	19	5	14	6	29	0.342
	3	20	34	15	41	5	24	
	4	1	2	1	3	0	0	

Table 2.3. Acute and late vaginal, GI and GU toxicities distributed by grade for Syed-Neblett and Tandem and Ovoid applicators.

In the cohort, 16 local recurrences were observed: 11 (10%) with Syed applicators and 5 (9%) with T&O applicators. As illustrated In Figure 2.3, the three-year local control rate for the entire cohort was 84.4% (95% CI: 77.0% to 92.5%). Comparing the applicators, T&Os showed slightly better local control, but the difference was not statistically significant (p = 0.329). Local control at 3 years post treatment was 86.1% (95% CI: 75.2% to 98.6%) for patients treated with T&O brachytherapy and 84.0% (95% CI: 75% to 94.2%) for patients who underwent Syed brachytherapy.

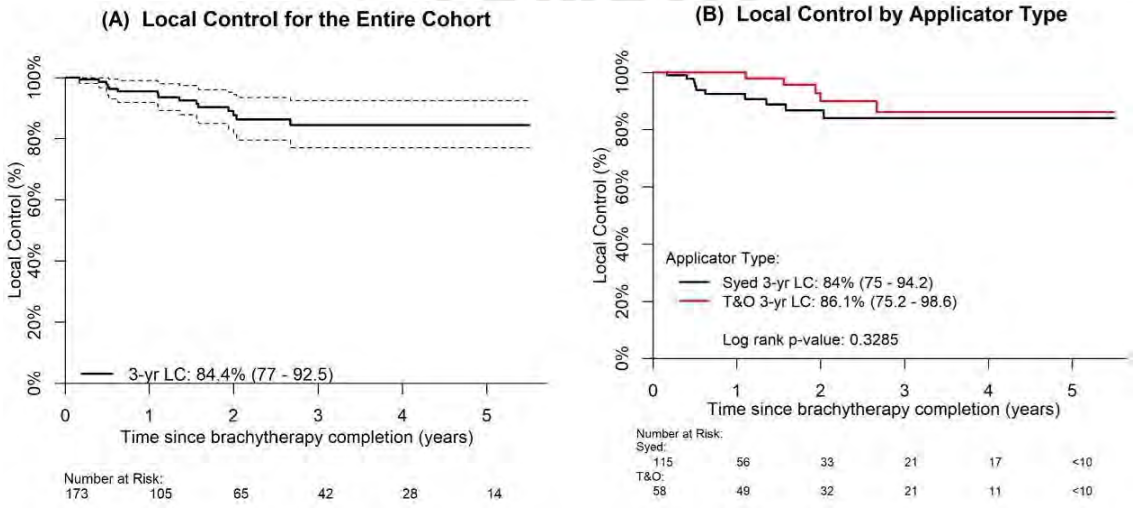


Figure 2.3. A) A Kaplan Meier plot for local control of the entire patient cohort and B) Kaplan Meier plots of local control by applicator type.

The toxicity distribution shows that although Syed-Neblett applicators deliver less radiation to most of the OARs apart from the rectum, only the acute toxicities are significantly lower. After 6 months of follow-up, the rate of late toxicities is lower for T&O applicators and Syed brachytherapy patients are more likely to experience higher grade toxicities. Our findings indicate that while Syed brachytherapy patients were significantly more likely to be free of any acute toxicities, those who developed either acute or late toxicities experienced equivalent or more severe toxicities compared to T&O patients. A possible explanation for this observation is that because Syed patients tend to have larger and higher stage tumors, they are more predisposed to severe toxicities.

Chapter 3

Machine Learning for Gynecological HDR Toxicities Prediction

3.1. Supervised Machine Learning

Supervised machine learning has been widely applied in healthcare for classification tasks, including predicting COVID-19 patient diagnosis and prognosis, forecasting hospitalizations related to heart disease, and estimating outcomes for infectious diseases (99–101). In this study, ML models are trained and assessed to predict grade 3 or higher toxicities in patients with gynecologic malignancies treated with external beam radiation therapy (EBRT) and high-dose-rate (HDR) brachytherapy. Various algorithms and preprocessing methods were utilized to determine the most accurate model for toxicity prediction, and the results of their performance comparisons are detailed below.

3.2. Study Design

A single-institution retrospective review evaluated 233 patients who received high-dose-rate (HDR) brachytherapy for gynecological malignancies (cervix, endometrium, vagina, or vulva) between 2009 and 2023. Treatments employed Syed-Neblett or Fletcher-Suit-Delclos Tandem and Ovoid (T&O) applicators. Researchers gathered demographic information, tumor characteristics, treatment specifics, dosimetric details (including a external beam radiotherapy boost), and gastrointestinal (GI), genitourinary (GU), and vaginal (VAG) toxicities observed during and after treatment. Exclusion criteria encompassed a history of brachytherapy, use of more than one applicator type, inconsistent dosimetric records, concurrent EBRT for a separate adjacent disease site, or combination low-dose-rate (LDR) and HDR therapy. Toxicities were graded using the Common Terminology Criteria for Adverse Events (CTCAE) v5.0 (85). A physician and a medical physicist independently verified the database three times to ensure accuracy and reliability. For preliminary analysis, patients with grade 2 or lower toxicity were compared to those with grade 3 or higher toxicity at any time post-EBRT initiation,

applying a significance threshold of $p \leq 0.05$. Kaplan-Meier curves for disease-free survival and local control were generated to characterize the patient cohort.

3.3. Data Preprocessing

The machine learning workflow was implemented in Python 3 using a Jupyter Notebook (running the IPython kernel) and leveraged the Scikit-learn (v1.3.2) library (44). The Charlson Comorbidity Index was segmented into approximate quartiles labeled as “Low” [0–2], “Medium” [3], “High” [4–5], and “Very High” [>5]. The Karnofsky Performance Status (KPS) classification, guided by clinical judgment, was “Bad” [50–70], “Normal” [80], or “Good” [90–100]. Data pre-processing was carried out in four phases: (A) Encoding, (B) Missing-Value Imputation, (C) Class Balancing, and (D) Normalization.

Data encoding converted both categorical and ordinal variables into numeric representations. Subsequently, the dataset was divided using a stratified approach grounded in the target variable, maintaining comparable toxicity event proportions across subsets. This procedure produced a 75% training set (n = 174) and a 25% testing set (n = 59).

Missing feature values were imputed based on the variable type, addressing missingness primarily caused by unobtainable patient records or unrecorded measurements, which suggests the data is Missing at Random (MAR). A K-nearest neighbors (KNN) imputer configured with a single nearest neighbor was utilized for categorical and ordinal variables, ensuring that missing values mapped to only one class. In contrast, numerical continuous features were imputed via a KNN imputer using five neighbors, where the mean of those neighbors replaced the missing data. These parameters were chosen after extensive experimentation. KNN imputation is particularly suitable for MAR when missingness is moderate (15%-30%), as it matches individuals with missing data to similar patients based on observed data, although its advantages may be less pronounced with lower levels of missingness. The imputers identify their nearest neighbors by calculating the Euclidean distance between data points (excluding missing data) and were fitted using only the training data before being applied to both the training and testing datasets (102).

Grade 3 or higher toxicity, defined as the positive class, was observed infrequently (24%, 56 out of 233 patients), leading to an imbalanced dataset. To mitigate this issue, the SVM-SMOTE balancing technique (55) was employed exclusively during model training. Preliminary evaluations demonstrated that SVM-SMOTE performed better than SMOTE (54) and ADASYN (56). Consequently, from the original 174 training samples, 90 synthetic positive samples were generated, yielding a total of 264 training examples (132 positive, 132 negative).

The final stage of preprocessing addressed normalization and standardization. Based on experimental comparisons, the Standard Scaler was adopted for continuous numeric features. For categorical and ordinal variables, both Target Encoding and One-Hot Encoding were initially explored. Target Encoding was ultimately selected to avoid inflating the number of features, an important consideration given the limited sample size and the associated risk of overfitting. Other approaches, including MinMax Scaler and Robust Scaler, were also tested but not reported here due to poorer results. All fitted Scalers and Target Encoding objects were then saved as a Joblib file and applied to the test set.

Potential collinearity among input features was examined using Pearson's correlation coefficient. Any pair of features with a correlation exceeding 0.85 was regarded as collinear, and one was removed accordingly. Testing other thresholds (0.7 and 0.95) yielded worse outcomes. In the event of correlated dose metrics, D2cc and D90 were retained in the final model because of their frequent clinical usage (31).

3.4. Baseline for Comparison

In this research, the F1, recall, and precision metrics are reported specifically for the positive class (patients experiencing toxicity). The AUC-ROC curve uses a random classifier as a baseline, depicted by a diagonal line with an AUC-ROC value of 0.5. For the precision-recall (PR) curve, the baseline assumes every instance is classified as positive, represented by a horizontal line whose vertical position varies based on the dataset's characteristics (103,104). All these metrics are calculated and presented for both the training set (excluding the synthetic cases generated by SVM-SMOTE) and the test set (after imputing missing values in both).

3.5. Dimensionality Reduction

High-dimensional datasets, common in healthcare and other domains, can challenge both model training and interpretability for algorithms like Logistic Regression, Support Vector Machines, K Nearest Neighbors, and Gaussian Naive Bayes (105,106). Feature extraction and feature selection (107) are common strategies to address these problems. In this work, we initially evaluated several sequential feature selection techniques: Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Forward Floating Selection (SFFS), and Sequential Backward Floating Selection (SBFS). All of these methods used the same algorithm planned for final model training as their estimator (108–110). After experimentation, Sequential Forward Selection was applied to LR, SVM, KNN, and GNB models due to its faster computation time and superior performance. This approach adds one predictor at a time, choosing the feature that maximizes the target metric, which in our case was the F1 score. A 10-fold Stratified Shuffle Split cross-validation was employed on the balanced training set to help prevent overfitting and to reliably measure performance (111,112). Finally, for the top three models, a SHAP analysis or a variable importance plot (for RF) was performed to study the relevance of the features selected.

3.6. Binary classification for Grade 3 or higher HDR toxicities

Six machine learning models were trained and examined to explore how well they could predict grade 3 or worse toxicities in gynecologic cancer patients who received EBRT plus either interstitial or T&O brachytherapy. These were: Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Gaussian Naïve Bayes (GNB), and a Multi-Layer Perceptron (MLP). These six models were selected because they are commonly applied in this field. They represent standard solutions for classification tasks and help identify patients at higher risk. The baseline for the precision–recall curve was 0.237, using a simple method that labeled every sample in the hold-out test set as positive. After choosing the most important features via the Sequential Feature Selection procedure for each suitable model, the hyperparameters of all six methods were optimized through Grid Search with a 10-fold Stratified Shuffle Split cross-validator, aiming to improve predictive ability. The Python scripts and the complete study data have been shared for reference. All study data

can be accessed at https://github.com/AndresPB95/ML-Model-Gynecological-HDR-G3Plus-Toxicities.

The dataset consisted of demographic and clinical details for 233 individuals, of whom 56 (24%) experienced grade 3 or higher toxicity. Table 3.1 provides a summary of demographic, treatment, and tumor-related information. Those who developed grade 3 or higher toxicity had a longer median follow-up (12.4 months vs. 3.8 months), were more inclined to have either low or very high comorbidity scores, and showed notably larger HR-CTV volumes (median 50 cc vs. 39 cc, p = 0.041).

	Full Cohort		No Grade 3+ Toxicity		Grade 3+ Toxicity		
	n=233	100%	n=177	76%	n=56	24%	p-value
Length of Follow-Up (mo)	6.1	IQR: [1.4 - 18.2]	3.8	IQR: [1.2 - 16.4]	12.4	IQR: [7.1 - 22.1]	***< 0.001***
Age at Completion	53.6	STD: 14.8	54.4	STD: 14.7	50.8	STD: 14.8	0.107
Non-Caucasian	25	11%	17	10%	8	14%	0.328
BMI	28.0	STD: 8.3	28.0	STD: 8.6	27.9	STD: 7.6	0.969
Charlson Comorbidity Index							***0.014***
Low [0-2]	87	37%	60	34%	27	48%	
Medium [3]	43	18%	33	19%	10	18%	
High [4-5]	60	26%	54	31%	6	11%	
Very High [>5]	43	18%	30	17%	13	23%	
KPS							0.369
Good [90-100]	147	63%	115	65%	32	57%	
Normal [80]	62	27%	43	24%	19	34%	
Bad [50-70]	23	10%	18	10%	5	9%	
Treatment Days	60	IQR: [52 - 71]	60	IQR: [52 - 69]	61	IQR: [52 - 74]	0.504
Applicator: T&O	80	34%	63	36%	17	30%	0.521
Concurrent Chemo	201	86%	153	86%	48	86%	1.000
Type of Boost							0.681
None	139	60%	108	61%	31	55%	
Sequential	54	23%	39	22%	15	27%	
SIB	40	17%	30	17%	10	18%	
Tumor Size (cm)	5.4	STD: 2.1	5.4	STD: 2.0	5.6	STD: 2.5	0.622
HR-CTV (cc)	43	IQR: [27 - 74]	39	IQR: [25 - 71]	50	IQR: [34 - 77]	***0.041***
Tumor Site							0.864

	Cervix	194	83%	147	83%	47	84%	
	Endometrium	16	7%	13	7%	3	5%	
	Other	23	10%	17	10%	6	11%	
Cancer Stage								0.163
	Stage 1	45	19%	37	21%	8	14%	
	Stage 2	57	25%	41	23%	16	29%	
	Stage 3	107	46%	84	48%	23	41%	
	Stage 4	22	10%	13	7%	9	16%	
Histology: SCC		180	77%	136	77%	44	79%	0.857
MRI Fused		105	45%	84	47%	21	38%	0.219

Table 3.1. Comparison of patient, treatment, and tumor characteristics between groups with and without grade 3 or higher toxicities.

Table 3.2 compares median dose coverage to the tumor (V100%, D50%, D90%, and D98%) and the dose to the organs at risk by toxicity status. Patients who developed grade 3 or higher toxicities had significantly higher rectal D2cc doses ($p = 0.043$), while all other dose comparisons were not statistically significant. The HR-CTV V100, D1cc - Rectum, and doses to the sigmoid colon were slightly higher for the group with grade 3 or higher toxicities , but these differences did not reach significance.

	Full Cohort		No Grade 3+ Toxicity		Grade 3+ Toxicity		
	N=233	(100%)	N=177	(76%)	N=56	(23%)	
	Median	IQR	Median	IQR	Median	IQR	p-value
HR-CTV V100 (cc)	80.1	[54.3 - 129.3]	74.5	[50.8 - 123.3]	88.3	[60.3 - 132.7]	0.104
HR-CTV D50 (Gy)	110.0	[101.2 - 119.1]	110.8	[101.4 - 119.1]	109.8	[100.3 - 118.9]	0.628
HR-CTV D90 (Gy)	83.1	[79.9 - 87.7]	83.1	[80.0 - 87.7]	83.1	[79.2 - 86.7]	0.967
HR-CTV D98 (Gy)	75.3	[70.9 - 79.9]	75.1	[70.9 - 79.8]	75.7	[69.7 – 80.0]	0.837
D0.1cc - Bladder (Gy)	97.5	[83.4 - 114.5]	97.0	[83.3 - 112.9]	98.6	[85.2 - 120.8]	0.455
D1cc - Bladder (Gy)	84.9	[74.9 - 95.4]	84.6	[75.2 - 94.7]	85.5	[74.6 - 98.1]	0.570
D2cc - Bladder (Gy)	79.7	[71.3 - 89.2]	79.2	[71.8 - 88.5]	81.9	[70.5 - 91.1]	0.569
D0.1cc - Small Bowel (Gy)	67.2	[52.8 - 83.8]	68.0	[52.7 - 84.5]	65.2	[53.9 - 78.2]	0.838
D1cc - Small Bowel (Gy)	60.9	[51.5 - 73.5]	61.1	[51.2 - 73.7]	59.6	[52.4 - 70.1]	0.852
D2cc - Small Bowel (Gy)	59.3	[50.7 - 69.6]	59.4	[50.6 – 70.0]	57.8	[51.0 - 66.4]	0.898
D0.1cc - Sigmoid Colon (Gy)	74.3	[62.0 – 86.0]	73.9	[59.9 - 86.3]	76.6	[65.2 - 85.4]	0.277
D1cc - Sigmoid Colon (Gy)	67.0	[57.2 - 75.4]	66.5	[56.1 – 75.0]	69.1	[60.3 - 76.3]	0.182
D2cc - Sigmoid Colon (Gy)	64.2	[55.3 - 71.6]	63.8	[54.3 - 71.2]	65.6	[58.0 - 72.2]	0.172
D0.1cc - Rectum (Gy)	80.8	[74.3 - 87.6]	80.7	[74.0 - 87.6]	82.3	[76.8 - 86.9]	0.361
D1cc - Rectum (Gy)	71.4	[66.2 - 78.4]	71.1	[65.7 - 76.6]	73.7	[68.5 - 79.9]	0.066
D2cc - Rectum (Gy)	67.6	[62.5 - 74.4]	67.3	[62.2 - 72.6]	71.1	[65.3 - 75.9]	***0.043***

Table 3.2. HR-CTV and OAR dosimetric values between groups with and without grade 3 or higher toxicities.

3.7. Model Performance

The six machine learning models were fitted using all variables listed in Table 3.1 and 3.2 as described in sections 3.3 – 3.6. A visual representation of the workflow is shown in Figure 3.1, and a summary of the inputs/outputs of the model as well as the variable types and number of missing data is shown in Table 3.3. The results are shown in Figures 3.2 and 3.3, which depict the performance of these models on the withheld test data; and in Figures 3.4 through 3.6, which analyze the most relevant features for the top 3 models. Numeric comparisons as well as the 95% confidence intervals for both the class-imbalanced training data and withheld test data are shown in Table 3.4. Confidence intervals of 95% were calculated using bootstrap resampling with replacement over 1000 iterations. In each iteration, the model was retrained using balanced data, and metrics were recalculated on the unbalanced training and testing datasets. The standard deviation of the metric values across iterations was used to estimate the confidence interval under the assumption of a normal distribution. The top three models for predicting grade 3 or higher toxicities were Support Vector Machines (SVM), Random Forests (RF), and Logistic Regression (LR) with F1 testing scores of 0.63, 0.57 and 0.52, normMCC testing scores of 0.75, 0.77 and 0.71, and Accuracy testing scores of 0.80, 0.85 and 0.81, respectively. All values shown in Table 3.4 assume a classification threshold value of 0.5 for toxicity prediction. For some models, such as MLP and KNN, there is a significant discrepancy between training and test data performance measures, indicating severe overfitting in the training data. Table 3.5 and 3.6 list the hyperparameter search space, and the most relevant features and the hyperparameters selected by the GridSearchCV optimization algorithm for the training data. The top features consistently selected among these three models are Chemotherapy, Charlson Comorbidity Index, KPS, D2cc - Small Bowel, Stage, Histology, and Follow-Up Time.

N	Input				Output
	Feature Name	Feature Type	Missing Data Points	Number of Categories[a]	
1	Had Chemotherapy	Categorical	0	2	Develops toxicity grade 3 or higher
2	Body Mass Index (BMI)	Numerical	0	-	
3	Charlson Comorbidity Index (CCI)	Ordinal	0	4	

4	Karnofsky Performance Status (KPS)	Ordinal	1	3	
5	Ethnicity	Categorical	0	2	
6	Age at Completion [years]	Numerical	0	-	
7	EBRT Type of Boost	Categorical	0	3	
8	Treatment Days	Numerical	26	-	
9	Type of Applicator	Categorical	0	2	
10	MRI	Categorical	0	2	
11	V100 [cc]	Numerical	5	-	
12	D50 [Gy]	Numerical	6	-	
13	D90 [Gy]	Numerical	6	-	
14	D98 [Gy]	Numerical	6	-	
15	D0.1Bladder [Gy]	Numerical	8	-	
16	D1Bladder [Gy]	Numerical	8	-	
17	D2Bladder [Gy]	Numerical	8	-	Does not develop toxicity grade 3 or higher
18	D0.1SmallBowel [Gy]	Numerical	10	-	
19	D1SmallBowel [Gy]	Numerical	10	-	
20	D2SmallBowel [Gy]	Numerical	10	-	
21	D0.1Sigmoid [Gy]	Numerical	7	-	
22	D1Sigmoid [Gy]	Numerical	7	-	
23	D2Sigmoid [Gy]	Numerical	7	-	
24	D0.1Rectum [Gy]	Numerical	6	-	
25	D1Rectum [Gy]	Numerical	6	-	
26	D2Rectum [Gy]	Numerical	6	-	

27	Tumor Site	Categorical	0	3	
28	Cancer Stage	Ordinal	2	4	
29	Histology	Categorical	0	2	
30	Maximum Tumor Length [cm]	Numerical	39	-	
31	HRCTV Volume [cc]	Numerical	0	-	
32	Follow Up [months]	Numerical	0	-	

a The number of categories is also shown for categorical and ordinal variable types.

Table 3.3. Summary of input features and output from models. The variable type and number of missing data points for each input is shown.

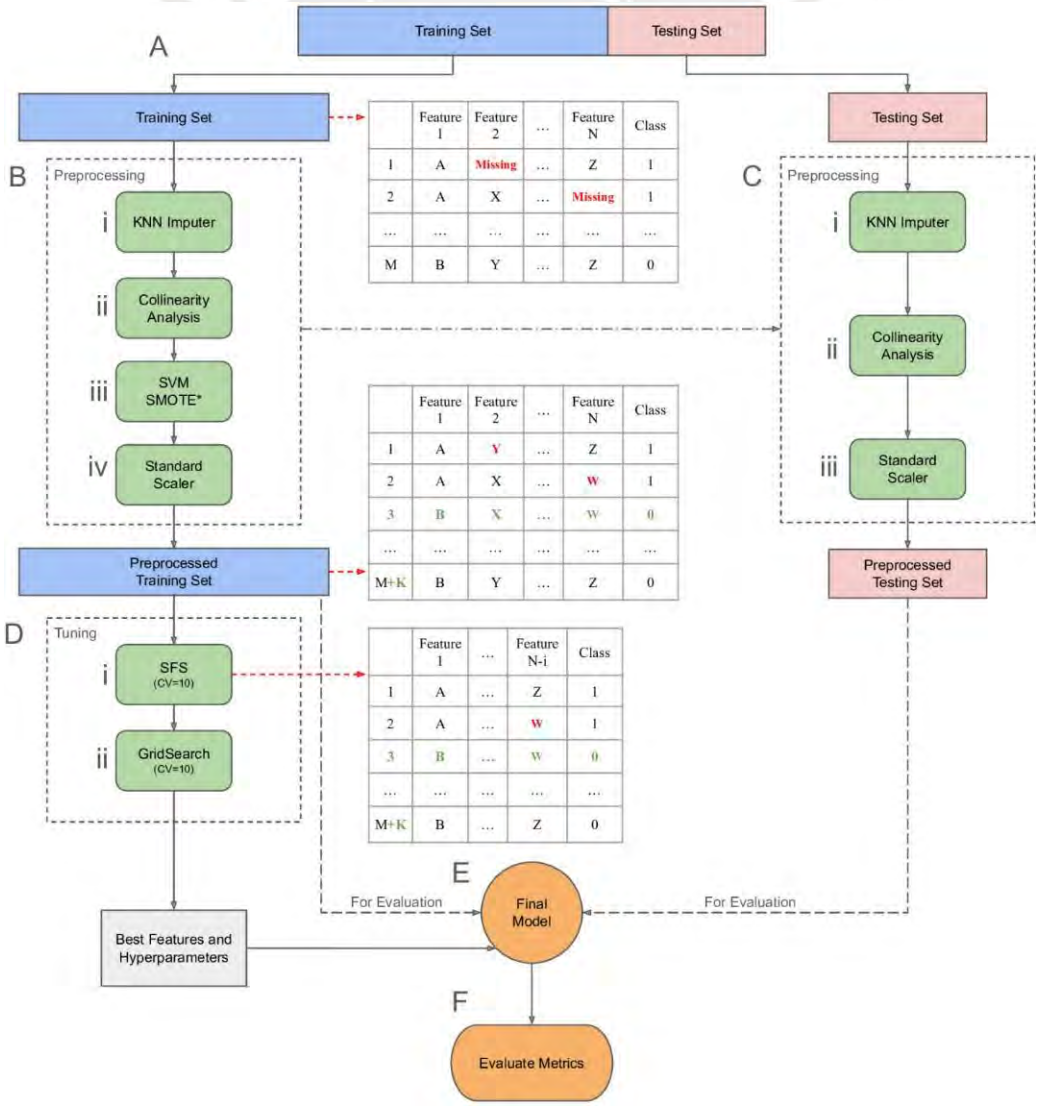


Figure 3.1. Machine Learning flowchart outlining the steps used when training and evaluating the different models. Full description available in Portocarrero et al. (33)

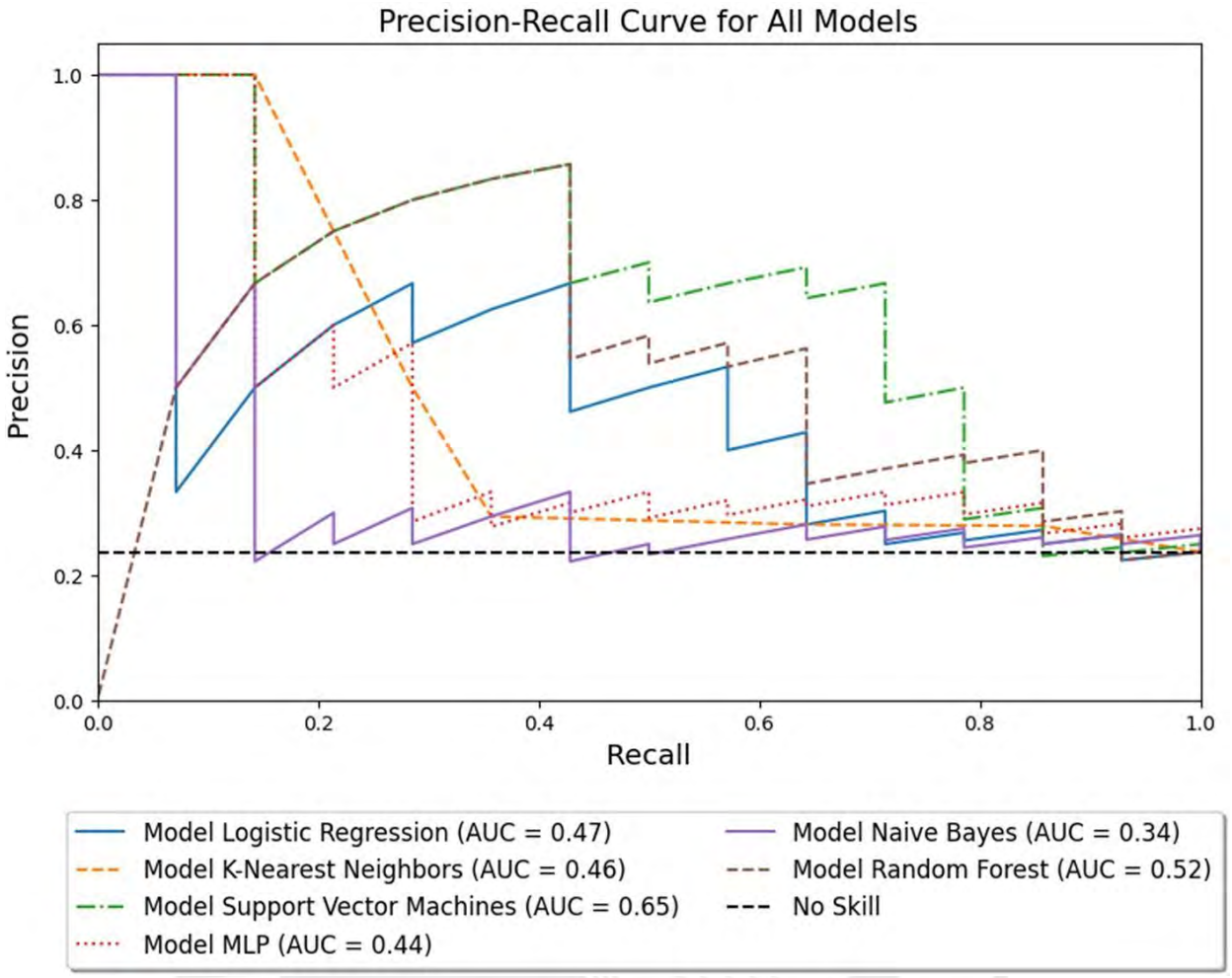


Figure 3.2. Precision-Recall curves comparing six machine learning models and a baseline value. The PR curves are computed using the withheld test data. The Support Vector Machine model has the highest area under the curve.

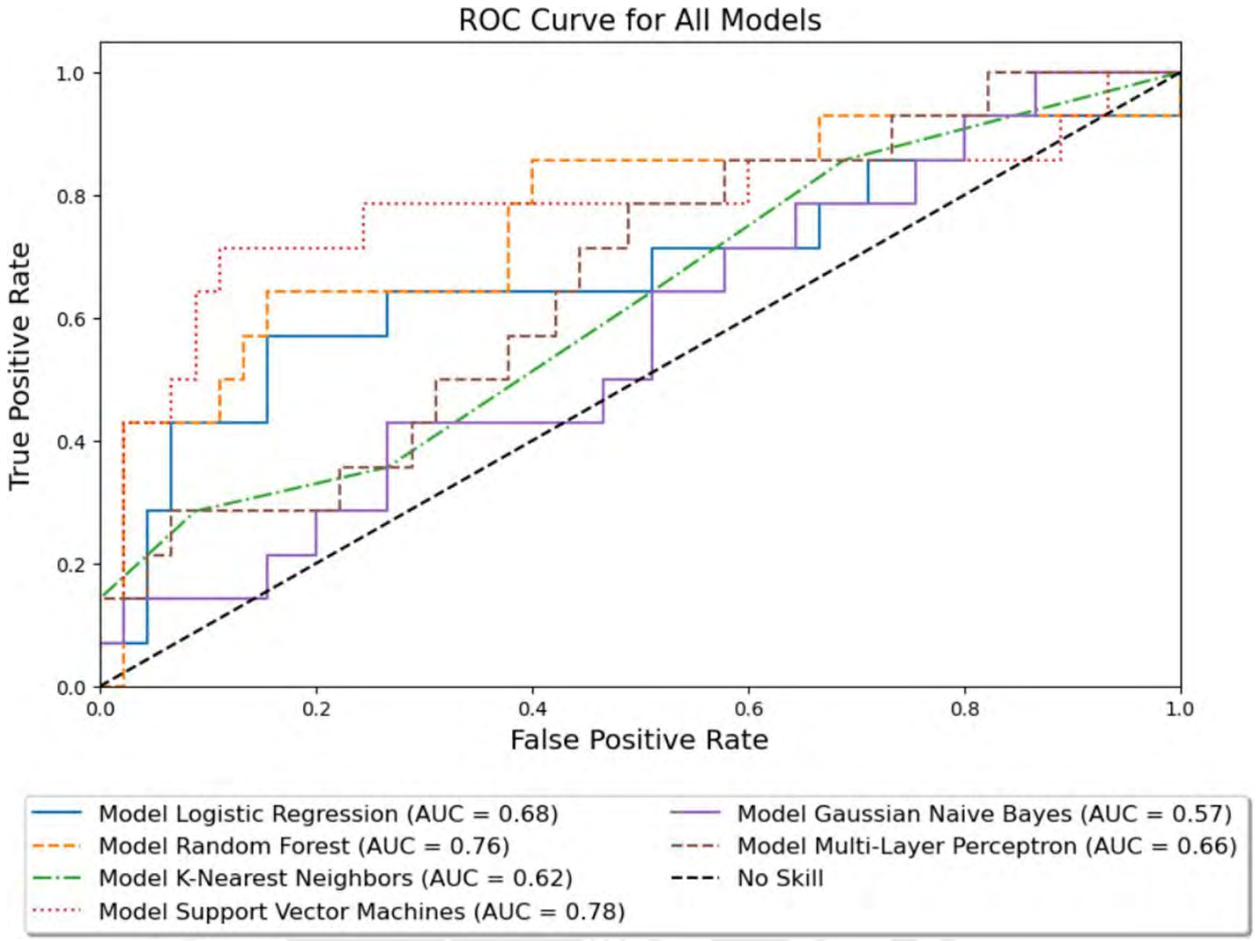


Figure 3.3. Receiver Operating Characteristics curves for six machine learning models and a baseline value. The ROC curves are computed using the withheld test data. The Support Vector Machine model has the highest area under the curve.

Model	Metric	Training Dataset		Testing Dataset	
		Mean	95% CI	Mean	95% CI
SVM	F1	0.61	[0.55 - 0.67]	**0.63**	[0.51 - 0.74]
	Accuracy	0.82	[0.74 - 0.89]	0.80	[0.69 - 0.91]
	normMCC	0.75	[0.70 - 0.79]	0.75	[0.65 - 0.84]
	Precision	0.63	[0.55 - 0.70]	0.56	[0.46 - 0.66]
	Recall	0.60	[0.46 - 0.73]	**0.71**	[0.54 - 0.89]
	AUC-ROC	0.87	[0.83 - 0.91]	**0.78**	[0.68 - 0.87]
	AUC-PR	0.60	[0.51 - 0.68]	**0.65**	[0.50 - 0.80]
RF	F1	0.82	[0.75 - 0.89]	0.57	[0.39 - 0.75]
	Accuracy	0.92	[0.86 - 0.97]	**0.85**	[0.74 - 0.95]
	normMCC	0.89	[0.84 - 0.94]	**0.77**	[0.65 - 0.89]
	Precision	0.89	[0.79 - 0.99]	**0.86**	[0.70 - 1.00]
	Recall	0.76	[0.67 - 0.86]	0.43	[0.18 - 0.68]
	AUC-ROC	0.97	[0.94 - 1.00]	0.76	[0.65 - 0.87]
	AUC-PR	0.93	[0.86 - 1.00]	0.52	[0.35 - 0.69]
LR	F1	0.45	[0.40 - 0.50]	0.52	[0.43 - 0.61]
	Accuracy	0.75	[0.68 - 0.81]	0.81	[0.73 - 0.89]
	normMCC	0.64	[0.60 - 0.69]	0.71	[0.65 - 0.78]
	Precision	0.47	[0.42 - 0.53]	0.67	[0.56 - 0.77]
	Recall	0.43	[0.33 - 0.53]	0.43	[0.31 - 0.55]
	AUC-ROC	0.72	[0.68 - 0.76]	0.68	[0.63 - 0.72]
	AUC-PR	0.43	[0.38 - 0.48]	0.47	[0.38 - 0.56]
MLP	F1	1.00	[0.93 - 1.00]	0.39	[0.26 - 0.52]
	Accuracy	1.00	[0.95 - 1.00]	0.63	[0.53 - 0.73]
	normMCC	1.00	[0.96 - 1.00]	0.57	[0.48 - 0.67]
	Precision	1.00	[0.91 - 1.00]	0.32	[0.21 - 0.43]
	Recall	1.00	[0.93 - 1.00]	0.50	[0.31 - 0.69]
	AUC-ROC	1.00	[0.95 - 1.00]	0.66	[0.54 - 0.77]
	AUC-PR	1.00	[0.85 - 1.00]	0.44	[0.31 - 0.57]
KNN	F1	0.72	[0.66 - 0.78]	0.32	[0.19 - 0.46]
	Accuracy	0.86	[0.81 - 0.91]	0.64	[0.53 - 0.76]
	normMCC	0.81	[0.77 - 0.86]	0.54	[0.44 - 0.65]
	Precision	0.68	[0.62 - 0.74]	0.29	[0.19 - 0.40]
	Recall	0.76	[0.66 - 0.87]	0.36	[0.15 - 0.57]
	AUC-ROC	0.91	[0.87 - 0.95]	0.62	[0.51 - 0.73]
	AUC-PR	0.78	[0.71 - 0.85]	0.46	[0.34 - 0.57]
GNB	F1	0.33	[0.26 - 0.41]	0.24	[0.08 - 0.39]
	Accuracy	0.79	[0.73 - 0.86]	0.78	[0.67 - 0.89]
	normMCC	0.66	[0.61 - 0.71]	0.62	[0.51 - 0.72]
	Precision	0.75	[0.69 - 0.81]	0.67	[0.55 - 0.79]

	Recall	0.21	[0.04 - 0.39]	0.14	[0.00 - 0.41]
	AUC-ROC	0.67	[0.62 - 0.72]	0.57	[0.47 - 0.67]
	AUC-PR	0.40	[0.33 - 0.47]	0.34	[0.23 - 0.45]

Table 3.4. Training and testing performance metrics for the considered machine learning models.

Model	Hyperparameter Search Space
SVM	**kernel**: linear,rbf,poly,sigmoid **Gamma**: Scale, Auto **C**: [1 - 10]
RF	**min_samples_leaf**: [5 - 10] **n_estimators**: [10, 15, 20, 25, 30, 35, 40, 45, 50,100, 200] **min_samples_split**: [5 - 10] **max_features**: sqrt, log2
LR	**penalty**: l1, l2 **solver**:liblinear, lbfgs **C**: [0.1, 0.3, 0.5, 0.7, 1]
MLP **2 Hidden Layers: (16,8)**	**learning_rate**: constant, invscaling, adaptive **activation**: logistic, tanh, relu **solver**: lbfgs, sgd **learning_rate_init**: [0.001, 0.1, 0.15, 0.2, 0.3, 0.5, 1] **alpha**: [0.0001, 0.0005, 0.001] **batch_size**: [20, 32, 50] **tol**: [0.00001, 0.0001, 0.0005]
KNN	**n_neighbors**: [1 - 10]
GNB	**var_smoothing**: [5e-08, 1e-09, 5e-09]

Table 3.5. Hyperparameter Search Space and MLP architecture.

SVM		RF		LR	
Features	HyperParameters	Features	HyperParameters	Features	HyperParameters
Chemotherapy	C: 1	All	n_estimators: 15	Chemotherapy	C: 1
Charlson	kernel: rbf		max_features: log2	Charlson	penalty: l2
KPS	gamma: scale		min_samples_leaf: 5	KPS	solver: lbfgs
MRI			min_samples_split: 5	Ethnicity	
D2cc Small Bowel				Type of Boost	
D2cc Sigmoid				Applicator	
Stage				D2cc Small Bowel	
Histology				D2cc Rectum	
HR-CTV				Tumor Site	
Follow-Up Time				Stage	
				Histology	
				Follow-Up Time	

Table 3.6. Most important features as selected by the Sequential Feature Selection algorithm (where appropriate) and found optimal hyperparameters for the top 3 scoring models.

3.8. Conventional Statistics Models vs Machine Learning

Historically, constraints in radiation therapy have been established using traditional statistical analyses such as log-rank tests on actuarial estimates, univariate logistic regression analysis, and others. These methods have provided a foundational understanding of the relationship between radiation dose and adverse effects. However, they often fall short in capturing the complexity and variability of individual patient responses, and even in the literature used to set these constraints, the development of multi-variable models is suggested (113). Machine learning (ML) techniques offer a more nuanced approach, leveraging data to identify intricate

patterns and interactions that traditional methods might miss. This can lead to more accurate and personalized dose constraints. An analysis of the references shown in the ASTRO report (31) that were used to suggest dose constraints is shown in Table 3.7:

Guideline Reference Number	Author	OAR	Method
95	Mazeron et al. (114)	Rectum	Incidences, Prevalence, Mean doses, logistic regression analysis (Probit), Log - Rank tests on actuarial estimates.
96	Kirchheiner et al. (115)	Vagina	Incidences, Prevalence, actuarial estimates.
102	Jensen et al. (116)	Rectum	Incidences, Prevalence, actuarial estimates.
115	Georg et al. (117)	Bladder and Rectum	Incidences, Mean doses, multi-variate analysis with general linear models, actuarial estimates
116	Kirchheiner et al. (118)	Vagina	Mean and median doses, Incidences, univariate and multivariate analysis with cox proportional Hazard model, Log - Rank tests on actuarial estimates.
117	Viswanathan et al. (119)	Rectum and Bladder	ABS recommendations. Literature review.
118	ICRU 89 (76)	Rectum and Bladder	ICRU 89. Literature Review.
119	Manea et al. (120)	Bladder	Incidences, logistic regression analysis (Probit), Log - Rank tests on actuarial estimates, univariate and multivariate analysis with cox proportional Hazard model.
120	Potter et al. (121)	Bladder, Sigmoid, Bowel	Embrace II. Literature Review.
121	Petit et al. (122)	Bowel	Mean doses, logistic regression analysis (Probit), and Log - Rank tests on actuarial estimates

Table 3.7. Analysis of references used to set dose constraints for cervical cancer brachytherapy treatments by ASTRO.

A side-by-side comparison of patients with and without grade 3 toxicities, conducted using marginal statistical methods, revealed only a few key significant distinctions, such as HR-CTV, Charlson Comorbidity Index, length of follow-up, and rectal D2cc. Some of these parameters, notably HR-CTV and rectal D2cc, have previously been identified as predictors of severe (grade 3 or higher) toxicity in HDR brachytherapy. For example, Lee et al. (75) reported that individuals experiencing grade 3–4 toxicities had a median HR-CTV of 111 cc, compared to just 43 cc in those with grade 0–2 toxicities. Likewise, Mesko et al. (123) documented a statistically significant difference between patients with and without grade 3 toxicities, citing median volumes of 93.8 cc and 51 cc, respectively. Additionally, Mazeron et al. (114) determined that rectal D2cc values of at least 75 Gy EQD2 were linked to both higher severity and greater incidence of toxicity in MRI-guided adaptive brachytherapy for locally advanced cervical cancer. Unlike traditional statistics, machine learning models consider nonlinear interactions between variables (124). The top-performing model selected 10 features, with Length of Follow-Up, D2cc to the Small Bowel, Charlson Comorbidity Index, HR-CTV volume, and D2cc to the sigmoid being the most relevant (Figure 3.4), agreeing with the discussed literature and the statistical analysis. Notably, Length of Follow-Up and Charlson Comorbidity Index were also significant predictors in our second and third best-performing models (Figures 3.5 and 3.6). Higher D2cc doses to the small bowel and sigmoid indicate greater radiation exposure, increasing the risk of severe toxicity, while a higher Charlson Comorbidity Index reflects poorer overall health and greater susceptibility to adverse effects. Larger HR-CTVs may necessitate more aggressive treatment, heightening toxicity risk, and Length of Follow-Up is crucial for capturing late-onset toxicities. It should be acknowledged that Sequential Forward Selection (SFS) can overlook variables, especially dosimetric ones, that are readily adjustable during treatment planning. Two main factors may lead to this outcome: (1) certain hyperparameter settings may constrain SFS's ability to recognize important feature interactions, or (2) some feature combinations may yield stronger predictions than incorporating actual dosimetric data. Although a model without dosimetric variables remains useful for forecasting toxicity risk, it does not give

clinicians the flexibility to adjust the treatment plan and potentially reduce that risk.

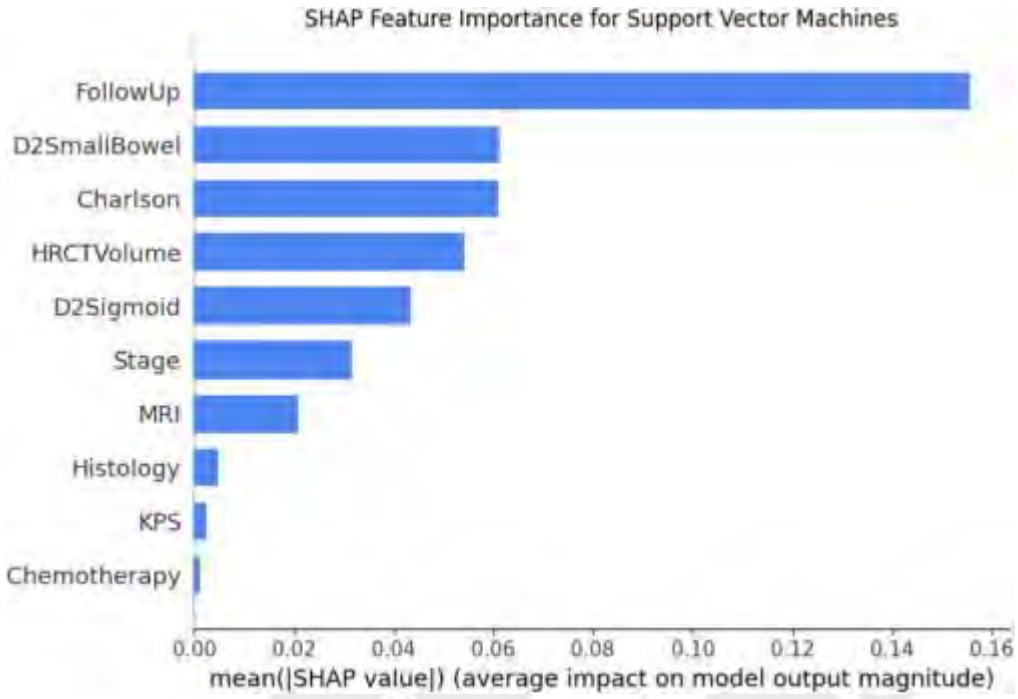


Figure 3.4. SHAP analysis for Support Vector Machine.

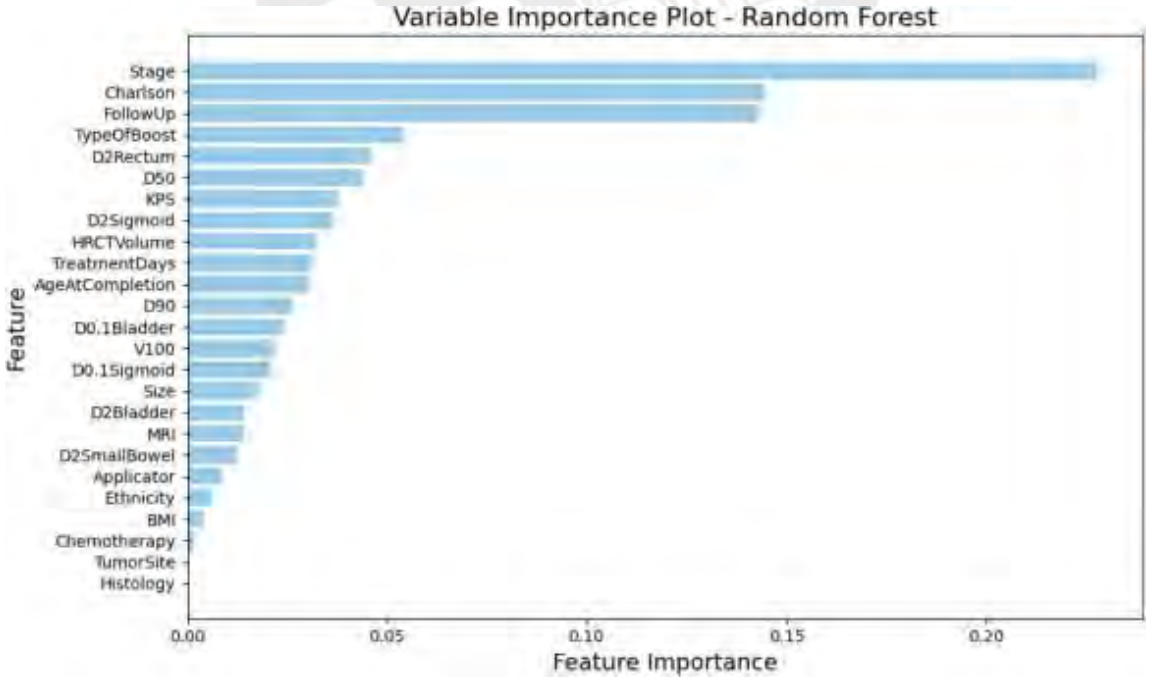


Figure 3.5. Variable Importance plot for Random Forest.

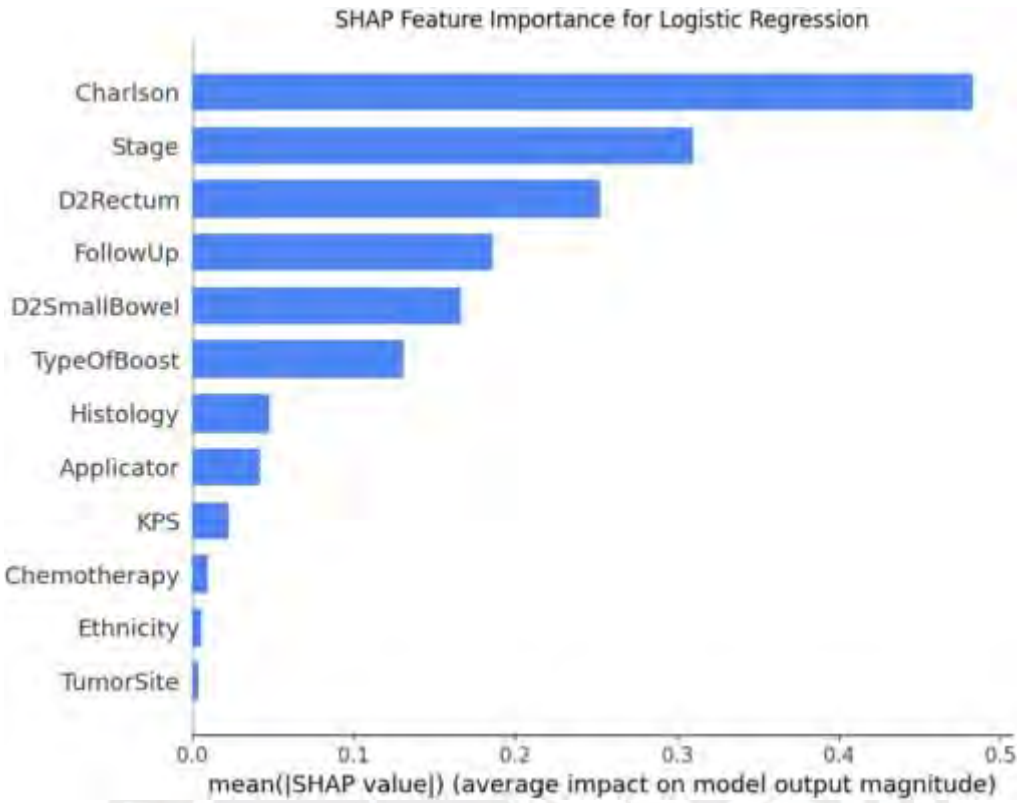


Figure 3.6. SHAP analysis for Logistic Regression.

3.9. Comparison against other ML models

Several studies have employed different machine learning strategies for predicting brachytherapy-related toxicities, showing diverse approaches and results. Tian et al. created a model to predict fistula formation using the SMOTE algorithm combined with an SVM model employing a radial basis kernel. Their results showed a recall of 97.1% and an AUC of 0.904, but the study was limited by a small dataset of 35 patients (7 positive cases), the absence of a separate test set, overfitting risks, and reliance on a single model (125). For rectal toxicity prediction, Chen et al. applied SVM, while Zhen et al. used convolutional neural networks, achieving cross-validated recall and AUC scores of 0.85 and 0.91 for Chen et al., and 0.75 and 0.89 for Zhen et al. Both studies included dose map features and used a dataset of 42 patients, with 12 experiencing grade 2 or higher rectal toxicity (126,127). Lucia et al. developed Normal Tissue Complication Probability (NTCP) models for predicting acute and late gastrointestinal,

genitourinary, and vaginal toxicities using radiomic features from 102 patients, but their work was limited to logistic regression. Their models achieved balanced accuracy scores between 63.99 and 78.41 (128). Cheon et al. applied deep learning methods to predict late bladder toxicities, outperforming traditional multivariable logistic regression. (129). Their models, trained on 281 patients, achieved an F1 score of 0.76. Unlike prior studies, our work leverages the largest dataset to predict grade 3 or higher toxicities. Similar to earlier approaches, we applied data-balancing techniques to maintain model stability during training. To improve model efficiency, we implemented feature selection using the Sequential Feature Selection Algorithm for all models except MLP and RF. This strategy aligns with established methodologies reported in previous research.

Overfitting occurs when a model becomes overly specialized to the training data, capturing random noise instead of meaningful patterns, which reduces its ability to perform well on unseen data (130). To mitigate this problem, a separate test set is used to evaluate how well the model generalizes and to detect overfitting (131). It is essential to report both training and test scores to provide a complete evaluation of the model's performance. Table 3.4 highlights overfitting in the MLP and KNN models, where their training F1 scores reached 1.00 and 0.72, respectively, but dropped significantly to 0.39 and 0.32 on the test data. These results reveal that the models fail to generalize effectively to new data. Future improvements will focus on expanding the range of hyperparameters explored and enhancing data preprocessing methods to reduce overfitting. These refinements will be addressed in subsequent studies.

Based on the scoring metrics in this study, Support Vector Machines (SVM) proved to be the most effective model for predicting grade 3 toxicities. On the training dataset the SVM model achieved a F1 score of 0.61, an accuracy of 0.82, a precision of 0.63, a recall of 0.6, an AUC-ROC of 0.87, and an AUC-PR of 0.6. For the same model, the test data results included an F1 score of 0.63, an accuracy of 0.80, a precision of 0.56, a recall of 0.71, an AUC-ROC of 0.78, and an AUC-PR of 0.65. Among the 14 patients in the test set who experienced toxicities, the model correctly identified 71% (10 true positives), while 56% of the predicted cases were actual toxicity events, avoiding false positives (8 false positives). Considering the inherent challenges in accurately predicting patient toxicities, this performance is deemed acceptable. In clinical

practice, both precision and recall play equally critical roles. It is not enough to identify true positives; minimizing false positives is equally important to ensure reliable decision-making. A model that recommends reducing doses to specific organs at risk (OARs) to prevent severe radiation-induced toxicities must carefully balance these two metrics. While high recall ensures that most toxicities are detected, an excess of false positives could lead to unnecessary dose reductions in patients who are not at risk. Such reductions may compromise tumor coverage, potentially lowering tumor control rates and jeopardizing treatment effectiveness. Therefore, the F1 score, which combines precision and recall into a single, balanced metric, is particularly suited for evaluating the performance of models in this context. Moving forward, future research could explore situations where either precision or recall takes precedence, as their importance may vary depending on clinical priorities. Specificity, on the other hand, is often less informative in imbalanced datasets, where the majority of outcomes are true negatives, as it can result in overly optimistic evaluations of model performance (132). Once the most effective model has been identified, further validation through multi-institutional clinical trials will be essential to confirm its robustness and ensure its practical utility in routine clinical settings.

Chapter 4

Conclusions and Future Outlooks

4.1 Conclusions

1. Syed-Neblett and tandem & ovoid (T&O) applicators are effective for treating cervical cancer but are better suited for different patient groups. Syed-Neblett applicators, often used in interstitial brachytherapy, are advantageous for treating larger tumors as they reduce bladder and sigmoid doses while maintaining adequate tumor coverage. On the other hand, T&O applicators are associated with higher rates of acute gastrointestinal (GI), genitourinary (GU), and vaginal toxicities, whereas late toxicities are more frequently observed with Syed interstitial brachytherapy. These findings highlight that dose metrics alone cannot fully predict clinical outcomes. As demonstrated in existing research, incorporating non-dosimetric features into predictive models improves the accuracy of toxicity predictions (125,133–135).

2. Various machine learning models were trained and tested to forecast grade 3 or higher toxicities in gynecologic cancer patients treated with EBRT combined with either interstitial or T&O brachytherapy, yielding encouraging results for the top-performing model. This innovative approach to forecasting toxicity may redefine clinical standards and support more personalized radiation therapy. Efforts to enhance model training methods and address common machine learning challenges, such as limited sample sizes, will require collaboration across the field. In the future, prospective validation of these models in different clinical environments will be crucial.

3. A comprehensive evaluation of performance metrics, including accuracy, precision, recall, F1 scores, and specificity, was conducted to assess the model's performance. Among these, the F1 score and precision were given particular emphasis, as they provide a practical measure of the model’s ability to predict toxicities. This focus is especially critical in clinical applications, where

accurately identifying toxicities while minimizing false positives is essential for optimizing patient care and treatment outcomes.

4. Machine learning models can capture nonlinear interactions between variables demonstrating their potential over conventional statistical methods. The practical importance of the selected features can vary, and their inclusion in the model does not automatically reflect their overall relevance to the outcome. This emphasizes the need for careful interpretation of feature significance in machine learning. Using conventional statistical models, HR-CTV volume, Charlson Comorbidity Index, Length of follow up, and the minimum dose to the hottest D2cc of the rectum were shown to be predictors for Grade 3 toxicities.

5. While this study provides valuable insights, there are areas that can be further improved. First, temporal bias was observed, as some earlier patient data could not be retrieved. This challenge was further amplified by the fact that many brachytherapy patients were referred from outside facilities, making it difficult to obtain complete records for EBRT treatments and follow-up appointments. Second, the study was limited to two applicator types, and hybrid intracavitary-interstitial devices were not included, which could be addressed in future research. Third, as with many machine learning studies, increasing the dataset size would likely enhance predictive accuracy, improve model generalizability, and reduce overfitting. Additionally, expanding the hyperparameter search space, exploring ensemble models such as XGBoost, and developing multi-feature combined models would further strengthen the results. Furthermore, while this study considered dosimetric, treatment, and tumor-related variables, incorporating spatial information such as dose maps could provide a more comprehensive analysis. Finally, other libraries such Optuna could be explored to better assess the importance of individual hyperparameters, adding another layer of refinement to the study.

4.2 Future Outlooks

1. Develop Machine Learning models for predicting Grade 2 or higher toxicities and plot how the probability of developing a toxicity varies with dose.
2. Optimize the decision threshold level for machine learning models and assess the variation in performance.

3. Explore other scoring metrics that could be used during the training for ML models such as the f-beta, log-loss, etc.
4. Incorporate additional features and techniques that may be of relevance for the predictive model such as dose maps, and nested cross-validations.
5. Design a multi-center brachytherapy database for the creation of more generalizable Machine Learning models.
6. Study the variation in risk as a function of dose, employing these multi-variable models.

References

1. Tanderup K, Ménard C, Polgar C, Lindegaard JC, Kirisits C, Pötter R. Advancements in brachytherapy. Adv Drug Deliv Rev. 2017 Jan 15;109:15–25.

2. Han K, Viswanathan AN. Brachytherapy in Gynecologic Cancers: Why Is It Underused? Curr Oncol Rep. 2016 Apr 1;18(4).

3. Petereit DG, Frank SJ, Viswanathan AN, Erickson B, Eifel P, Nguyen PL, et al. Brachytherapy: Where has it gone? Journal of Clinical Oncology. 2015 Apr 1;33(9):980–2.

4. Varian Medical Systems. Applicators & Accessories [Internet]. 2023 [cited 2024 Apr 17]. Available from: https://varian.widen.net/s/wwyx9dru1f/brachy_catalog_rad10536a_september2019

5. Damato AL, Cormack RA, Viswanathan AN. Characterization of implant displacement and deformation in gynecologic interstitial brachytherapy. Brachytherapy. 2014;13(1):100–9.

6. Rivard MJ, Coursey BM, DeWerd LA, Hanson WF, Huq MS, Ibbott GS, et al. Update of AAPM Task Group No. 43 Report: A revised AAPM protocol for brachytherapy dose calculations. Med Phys. 2004;31(3):633–74.

7. Nath R, Anderson LL, Luxton G, Weaver KA, Williamson JF, Meigooni AS. Dosimetry of interstitial brachytherapy sources: Recommendations of the AAPM Radiation Therapy Committee Task Group No. 43. Med Phys. 1995;22(2):209–34.

8. DeWerd LA, Ibbott GS, Meigooni AS, Mitch MG, Rivard MJ, Stump KE, et al. A dosimetric uncertainty analysis for photon-emitting brachytherapy sources: Report of AAPM Task Group No. 138 and GEC-ESTRO. Med Phys. 2011;38(2):782–801.

9. Taylor BN, Kuyatt CE. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results [Internet]. Gaithersburg; 1994 [cited 2024 Apr 22]. Available from: http://physics.nist.gov/TN1297

10. Huang L, Gaballa H, Chang J. Evaluating dosimetric accuracy of the 6 MV calibration on EBT3 film in the use of Ir-192 high dose rate brachytherapy. J Appl Clin Med Phys. 2022 May 1;23(5).

11. Rijnders A. Photon sources for brachytherapy. In: Lemoigne Y, Caner A, editors. Radiotherapy and Brachytherapy [Internet]. Brussels: Springer; 2009. p. 185–93. Available from: http://www.nato.int/science

12. Johns MW, Nablo S V. Disintegration of Ir 192 and Ir 194. Physical Review. 1954;96(6):1599–607.

13. Portocarrero-Bonifaz A, Syed S, Kassel M, et al. Dosimetric and toxicity comparison between Syed-Neblett and Fletcher-Suit-Delclos Tandem and Ovoid applicators in high dose rate cervix cancer brachytherapy. *Brachytherapy*. 2024;23(4):397-406.

14. Kehoe S. Treatments for gynaecological cancers. Best Pract Res Clin Obstet Gynaecol. 2006 Dec;20(6):985–1000.

15. Tanderup K, Eifel PJ, Yashar CM, Pötter R, Grigsby PW. Curative radiation therapy for locally advanced cervical cancer: Brachytherapy is NOT optional. Int J Radiat Oncol Biol Phys. 2014 Mar 1;88(3):537–9.

16. Chargari C, Deutsch E, Blanchard P, Gouy S, Martelli H, Guérin F, et al. Brachytherapy: An overview for clinicians. CA Cancer J Clin. 2019 Sep;69(5):386–401.

17. Marvaso G, Pepa M, Volpe S, et al. Virtual and augmented reality as a novel opportunity to unleash the power of radiotherapy in the digital era: a scoping review. Appl Sci. 2022 Nov;12(22):11308.

18. Hande V, Chopra S, Polo A, Mittal P, Kohle S, Ghadi Y, et al. Transitioning India to advanced image based adaptive brachytherapy: a national impact analysis of upgrading National Cancer Grid cervix cancer guidelines. The Lancet Regional Health - Southeast Asia. 2023 Sep 1;16.

19. Chargari C, Van Limbergen E, Mahantshetty U, Deutsch, Haie-Méder C. Radiobiology of brachytherapy: The historical view based on linear quadratic model and perspectives for optimization. Cancer/Radiotherapie. 2018 Jun 1;22(4):312–8.

20. Barendsen GW. Dose Fractionation, Dose Rate And Iso-Effect Relationships For Normal Tissue Responses. nternational Journal of Radiation Oncology* Biology* Physics. 1982;8(11):1981–97.

21. Fowler JF. The British Journal of Radiology The linear-quadratic formula and progress in fractionated radiotherapy. Br J Radiol. 1989;62(740):679–94.

22. Fowler JF. Sensitivity Analysis of Parameters in Linear-Quadratic Radiobiologic Modeling. Int J Radiat Oncol Biol Phys. 2009 Apr 1;73(5):1532–7.

23. Dale R, Plataniotis G, Jones B. A generalised method for calculating repopulation-corrected tumour EQD2 values in a wide range of clinical situations, including interrupted treatments. Physica Medica. 2024 Feb 1;118.

24. Tornero-López AM, Guirado D. Radiobiological considerations in combining doses from external beam radiotherapy and brachytherapy for cervical cancer. Reports of Practical Oncology and Radiotherapy. 2018 Nov 1;23(6):562–73.

25. Hall EJ, Giaccia AJ. Radiobiology for the radiologist. 8th Edition. Philadelphia: Wolters Kluwer Health; 327–327 p.

26. Zindler JD, Thomas CR, Hahn SM, Hoffmann AL, Troost EGC, Lambin P. Increasing the Therapeutic Ratio of Stereotactic Ablative Radiotherapy by Individualized Isotoxic Dose Prescription. J Natl Cancer Inst. 2016 Feb 1;108(2).

27. Prabhakar R, Rath G. A simple plan evaluation index based on the dose to critical structures in radiotherapy. J Med Phys. 2011 Oct;36(4):192–7.

28. Emami B, Lyman J, Brown ~ A, Coia L, Goitein M, Munzenrider JE, et al. Tolerance of normal tissue to therapeutic irradiation. Int J Radiat Oncol Biol Phys. 1991;21(1):109–22.

29. Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of Normal Tissue Complication Probability Models in the Clinic. Int J Radiat Oncol Biol Phys. 2010 Mar 1;76(3 SUPPL.).

30. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. Int J Radiat Oncol Biol Phys. 2010 Mar 1;76(3 SUPPL.).

31. Chino J, Annunziata CM, Beriwal S, Bradfield L, Erickson BA, Fields EC, et al. Radiation Therapy for Cervical Cancer: Executive Summary of an ASTRO Clinical Practice Guideline. Pract Radiat Oncol. 2020 Jul 1;10(4):220–34.

32. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: Diagnosis, drug development, and treatment. Medicina (Lithuania). 2020 Sep 1;56(9):1–10.

33. Portocarrero Bonifaz A, Syed S, Kassel M, McKenzie G, Shah V, Forry B, Gaskins J, Sowards K, Babitha Avula T, Masters A, Schneider J. Advancing patient care: Machine learning models for predicting grade 3+ toxicities in gynecologic cancer patients treated with HDR brachytherapy. medRxiv. 2024:2024-10.

34. Saravanan R, Sujatha P. A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification. In: International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE; 2018. p. 945–9.

35. Kotsiantis S, Zaharakis I, Pintelas P. Machine learning: A review of classification and combining techniques. Artif Intell Rev. 2006 Nov;26(3):159–90.

36. Vujović Ž. Classification Model Evaluation Metrics. International Journal of Advanced Computer Science and Applications. 2021;12(6):599–606.

37. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Min. 2023 Dec 1;16(1).

38. Bradley AE. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30(7):1145–59.

39. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. Journal of Educational Research. 2002;96(1):3–14.

40. LaValley MP. Logistic regression. Circulation. 2008 May;117(18):2395–9.

41. Stoltzfus JC. Logistic regression: A brief primer. Academic Emergency Medicine. 2011 Oct;18(10):1099–104.

42. Breiman L. Random Forests. Mach Learn. 2001;45:5–32.

43. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Belmont: Wadsworth; 1984.

44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Grisel O, Blondel M, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research [Internet]. 2011 [cited 2024 May 1];12:2825–30. Available from: https://scikit-learn.org/stable/modules/tree.html#mathematical-formulation

45. Cover TM, Hart PE. Nearest Neighbor Pattern Classification. IEEE Trans Inf Theory. 1952;24(1):335–42.

46. Li Y, Yang Y, Che J, Zhang L. Predicting the Number of Nearest Neighbor for kNN Classifier. IAENG Int J Comput Sci. 2019;46(4):662–9.

47. Noble WS. What is a support vector machine? Nat Biotechnol. 2006;24(12):1565–7. Available from: http://www.nature.com/naturebiotechnology

48. Awad M, Khanna R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature; 2015. 39–49 p.

49. Pérez A, Larrañaga P, Inza I. Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. International Journal of Approximate Reasoning. 2006 Sep;43(1):1–25.

50. Zhang H, Su J. Naive Bayes for optimal ranking. Journal of Experimental and Theoretical Artificial Intelligence. 2008 Jun;20(2):79–93.

51. Faris H, Aljarah I, Mirjalili S. Training feedforward neural networks using multi-verse optimizer for binary classification problems. Applied Intelligence. 2016 Sep 1;45(2):322–32.

52. Ni X, Fang L, Huttunen H. Adaptive L2 regularization in person Re-identification. In: Proceedings - International Conference on Pattern Recognition. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 9601–7.

53. Gosain A, Sardana S. Handling class imbalance problem using oversampling techniques: A review. In: international conference on advances in computing, communications and informatics (ICACCI). IEEE; 2017. p. 79–85.

54. Chawla N V, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. Vol. 16, Journal of Artificial Intelligence Research. 2002.

55. Nguyen HM, Cooper EW, Kamei K. Borderline over-sampling for imbalanced data classification. Vol. 3, Int. J. Knowledge Engineering and Soft Data Paradigms. 2011.

56. He H, Bai Y, Garcia E, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks . IEEE; 2008. p. 1322–8.

57. Khalid S. A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning. In: Science and Information Conference. IEEE; 2014. p. 372–8.

58. Uncu Ö, Türkşen IB. A novel feature selection approach: Combining feature wrappers and filters. Inf Sci (N Y). 2007 Jan 15;177(2):449–66.

59. Aparna U, Paul S. Feature selection and extraction in data mining. In: Online international conference on green engineering and technologies (IC-GET). IEEE; 2016. p. 1–3.

60. Sharma A, Paliwal KK. Linear discriminant analysis for the small sample size problem: an overview. International Journal of Machine Learning and Cybernetics. 2015 Jun 1;6(3):443–54.

61. Costa M, Lai C. Coordinated efforts to harmonise gynaecological cancer care. Lancet Oncol. 2022;23(8):971–2.

62. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024 Jan;74(1):12–49.

63. Ward KK, Shah NR, Saenz CC, McHale MT, Alvarez EA, Plaxe SC. Changing demographics of cervical cancer in the United States (1973-2008). Gynecol Oncol. 2012 Sep;126(3):330–3.

64. Walker TY, Elam-Evans LD, Yankey D, Markowitz LE, Williams CL, Fredua B, et al. Morbidity and Mortality Weekly Report National, Regional, State, and Selected Local Area Vaccination

Coverage Among Adolescents Aged 13-17 Years-United States, 2018. Morbidity and Mortality Weekly Report. 2019;68(33):718–23.

65. Bhatla N, Aoki D, Sharma DN, Sankaranarayanan R. Cancer of the cervix uteri. International Journal of Gynecology and Obstetrics. 2018 Oct 1;143:22–36.

66. Jemal A, Ward EM, Johnson CJ, Cronin KA, Ma J, Ryerson AB, et al. Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. J Natl Cancer Inst. 2017 Sep 1;109(9).

67. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209–49.

68. Small W, Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, et al. Cervical cancer: A global health crisis. Cancer. 2017 Jul 1;123(13):2404–12.

69. Pötter R, Tanderup K, Schmid MP, Jürgenliemk-Schulz I, Haie-Meder C, Fokdal LU. MRI-guided adaptive brachytherapy in locally advanced cervical cancer (EMBRACE-I): a multicentre prospective cohort study. Articles Lancet Oncol. 2021;22:538–85.

70. Han K, Milosevic M, Fyles A, Pintilie M, Viswanathan AN. Trends in the utilization of brachytherapy in cervical cancer in the United States. Int J Radiat Oncol Biol Phys. 2013 Sep 1;87(1):111–9.

71. Otter S, Coates A, Franklin A, Cunningham M, Stewart A. Improving dose delivery by adding interstitial catheters to fixed geometry applicators in high-dose-rate brachytherapy for cervical cancer. Brachytherapy. 2018 May 1;17(3):580–6.

72. Harkenrider MM, Alite F, Silva SR, Small W. Image-based brachytherapy for the treatment of cervical cancer. Int J Radiat Oncol Biol Phys. 2015 Jul 15;92(4):921–34.

73. Serban M, Kirisits C, de Leeuw A, Pötter R, Jürgenliemk-Schulz I, Nesvacil N, et al. Ring Versus Ovoids and Intracavitary Versus Intracavitary-Interstitial Applicators in Cervical Cancer Brachytherapy: Results From the EMBRACE I Study. Int J Radiat Oncol Biol Phys. 2020 Apr 1;106(5):1052–62.

74. Ma JK, Mourad WF, Allbright R, Packianathan S, Harrell LM, Chinchar E, et al. Short-term clinical outcome and dosimetric comparison of tandem and ring versus tandem and ovoids intracavitary applicators. J Contemp Brachytherapy. 2015;7(3):218–23.

75. Lee LJ, Damato AL, Viswanathan AN. Clinical outcomes of high-dose-rate interstitial gynecologic brachytherapy using real-time CT guidance. Brachytherapy. 2013 Jul;12(4):303–10.

76. Bentzen SM, Boone JM, Brandan ME, Burns DT, Dawson P, Fantuzzi E, et al. ICRU Report No. 89. J ICRU. 2016;13(1–2). Available from: http://jicru.oxfordjournals.org/

77. Spampinato S, Fokdal LU, Pötter R, Haie-Meder C, Lindegaard JC, Schmid MP, et al. Risk factors and dose-effects for bladder fistula, bleeding and cystitis after radiotherapy with imaged-guided adaptive brachytherapy for cervical cancer: An EMBRACE analysis. Radiotherapy and Oncology. 2021 May 1;158:312–20.

78. Spampinato S, Tanderup K, Lindegaard JC, Schmid MP, Sturdza A, Segedin B, et al. Association of persistent morbidity after radiotherapy with quality of life in locally advanced cervical cancer survivors. Radiotherapy and Oncology. 2023 Apr 1;181.

79. Nout RA, Putter H, Jürgenliemk-Schulz IM, Jobsen JJ, Lutgens LCHW, Van Der Steen-Banasik EM, et al. Quality of life after pelvic radiotherapy or vaginal brachytherapy for endometrial cancer: First results of the randomized PORTEC-2 trial. Journal of Clinical Oncology. 2009 Jul 20;27(21):3547–56.

80. Kirchheiner K, Pötter R, Tanderup K, Lindegaard JC, Haie-Meder C, Petrič P, et al. Health-Related Quality of Life in Locally Advanced Cervical Cancer Patients after Definitive Chemoradiation Therapy Including Image Guided Adaptive Brachytherapy: An Analysis from the EMBRACE Study. Int J Radiat Oncol Biol Phys. 2016 Apr 1;94(5):1088–98.

81. Pötter R, Haie-Meder C, Van Limbergen E, Barillot I, De Brabandere M, Dimopoulos J, et al. Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy - 3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology. Radiotherapy and Oncology. 2006 Jan;78(1):67–77.

82. Prisciandaro J, Zoberi J, Cohen G, Kim Y, Johnson P, Paulson E, et al. AAPM task group report 303 endorsed by the ABS: MRI implementation in HDR brachytherapy—Considerations from simulation to treatment. Med Phys. 2022 Aug 1;49(8):e983–1023.

83. Chino J, Annunziata CM, Beriwal S, Bradfield L, Erickson BA, Fields EC, et al. Radiation Therapy for Cervical Cancer: Executive Summary of an ASTRO Clinical Practice Guideline. Pract Radiat Oncol. 2020 Jul 1;10(4):220–34.

84. Paul AG, Nalichowski A, Abrams J, Paximadis P, Zhuang L, Miller S. Dosimetric evaluation of point A and volume-based high-dose-rate plans: a single institution study on adaptive brachytherapy planning for cervical cancer. J Contemp Brachytherapy. 2018;10(3):202–10.

85. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) [Internet]. 2018 [cited 2024 May 7]. Available from: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm

86. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna; 2022 [cited 2024 May 7]. Available from: https://www.R-project.org/

87. Dhanapalan R, Pandjatcharam J, Saravanan K, Patil NH. Comparison of vaginal gauze packing technique with or without balloon in high-dose-rate brachytherapy of uterine cervical cancer: A crossover randomized controlled trial. J Contemp Brachytherapy. 2022;14(6):551–9.

88. Schmid MP, Lindegaard JC, Mahantshetty ; Umesh, Tanderup K, Ina ;, Urgenliemk-Schulz J¨, et al. Risk Factors for Local Failure Following Chemoradiation and Magnetic Resonance Image-Guided Brachytherapy in Locally Advanced Cervical Cancer: Results From the EMBRACE-I Study. J Clin Oncol. 2023;41:1933–42.

89. Hanania AN, Myers P, Yoder AK, Bulut A, Henry Yu Z, Eraj S, et al. Inversely and adaptively planned interstitial brachytherapy: A single implant approach. Gynecol Oncol. 2019 Feb 1;152(2):353–60.

90. Kim Y, Kim YJ, Kim JY, Lim YK, Jeong C, Jeong J, et al. Toxicities and dose–volume histogram parameters of MRI-based brachytherapy for cervical cancer. Brachytherapy. 2017 Jan 1;16(1):116–25.

91. Thibault I, Lavallée MC, Aubin S, Laflamme N, Vigneault éric. Inverse-planned gynecologic high-dose-rate interstitial brachytherapy: Clinical outcomes and dose-volume histogram analysis. Brachytherapy. 2012 May;11(3):181–91.

92. Levin D, Menhel J, Rabin T, Pfeffer MR, Symon Z. Dosimetric Comparison of Tandem and Ovoids vs. Tandem and Ring for Intracavitary Gynecologic Applications. Medical Dosimetry. 2008 Dec;33(4):315–20.

93. Rangarajan R. Dosimetric evaluation of image based brachytherapy using tandem ovoid and tandem ring applicators. Reports of Practical Oncology and Radiotherapy. 2018 Jan 1;23(1):57–60.

94. Akkas EA, Altundag B. Long-term clinical outcome and dosimetric comparison of tandem and ring versus tandem and ovoids intracavitary application in cervical cancer. JBUON. 2021;26(3):698–706.

95. Koh V, Choo BA, Lee KM, Tan TH, Low JHJ, Ng SYJ, et al. Feasibility study of toxicity outcomes using GEC-ESTRO contouring guidelines on CT based instead of MRI-based planning in locally advanced cervical cancer patients. Brachytherapy. 2017 Jan 1;16(1):126–32.

96. Romano KD, Hill C, Trifiletti DM, Peach MS, Horton BJ, Shah N, et al. High dose-rate tandem and ovoid brachytherapy in cervical cancer: Dosimetric predictors of adverse events. Radiation Oncology. 2018 Jul 16;13(1).

97. Mayadev J, Viswanathan A, Liu Y, Li CS, Albuquerque K, Damato AL, et al. American Brachytherapy Task Group Report: A pooled analysis of clinical outcomes for high-dose-rate brachytherapy for cervical cancer. Vol. 16, Brachytherapy. Elsevier Inc.; 2017. p. 22–43.

98. Pötter R, Tanderup K, Kirisits C, de Leeuw A, Kirchheiner K, Nout R, et al. The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies. Vol. 9, Clinical and Translational Radiation Oncology. Elsevier Ireland Ltd; 2018. p. 48–60.

99. Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA. Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset. SN Comput Sci. 2021 Feb 1;2(1).

100. Dai W, Brisimi TS, Adams WG, Mela T, Saligrama V, Paschalidis IC. Prediction of hospitalization due to heart diseases by supervised learning methods. Int J Med Inform. 2015 Mar 1;84(3):189–97.

101. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. American Journal of Medicine. 2019 Jul 1;132(7):795–801.

102. Micci-Barreca D. A Preprocessing Scheme for High-Cardinality Categorical Attributes in Classification and Prediction Problems. SIGKDD Explor. 2001;3(1):27–32.

103. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. 2013;4(2):627–35.

104. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015 Mar 4;10(3).

105. Guyon I, De AM. An Introduction to Variable and Feature Selection André Elisseeff. Journal of Machine Learning Research. 2003;3:1157–82.

106. Zekić-Sušac M, Pfeifer S, Šarlija N. A Comparison of Machine Learning Methods in a High-Dimensional Classification Problem. Business Systems Research Journal. 2014 Sep 1;5(3):82–96.

107. Pham HT, Awange J, Kuhn M. Evaluation of Three Feature Dimension Reduction Techniques for Machine Learning-Based Crop Yield Prediction Models. Sensors. 2022 Sep 1;22(17).

108. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. ACM Comput Surv. 2017 Dec 1;50(6):1–45.

109. Pudil P, Novovi J. Floating search methods in feature selection. Pattern Recognition Letters . 1994;15(11):1119–25.

110. Molina LC, Belanche L, Nebot À. Feature Selection Algorithms: A Survey and Experimental Evaluation. In: IEEE International Conference on Data Mining. IEEE; 2002. p. 306–13.

111. Prusty S, Patnaik S, Dash SK. SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. Frontiers in Nanotechnology. 2022 Aug 19;4.

112. Xu Y, Goodacre R. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. J Anal Test. 2018 Jul 1;2(3):249–62.

113. Mazeron R, Fokdal LU, Kirchheiner K, Georg P, Jastaniyah N, Šegedin B, et al. Dose–volume effect relationships for late rectal morbidity in patients treated with chemoradiation and MRI-guided adaptive brachytherapy for locally advanced cervical cancer: Results from the prospective multicenter EMBRACE study. Radiotherapy and Oncology. 2016 Sep 1;120(3):412–9.

114. Mazeron R, Fokdal LU, Kirchheiner K, Georg P, Jastaniyah N, Šegedin B, et al. Dose–volume effect relationships for late rectal morbidity in patients treated with chemoradiation and MRI-guided adaptive brachytherapy for locally advanced cervical cancer: Results from the prospective multicenter EMBRACE study. Radiotherapy and Oncology. 2016 Sep 1;120(3):412–9.

115. Kirchheiner K, Nout RA, Tanderup K, Lindegaard JC, Westerveld H, Haie-Meder C, et al. Manifestation pattern of early-late vaginal morbidity after definitive radiation (Chemo)therapy and image-guided adaptive brachytherapy for locally advanced cervical cancer: An analysis from the embrace study. Int J Radiat Oncol Biol Phys. 2014 May 1;89(1):88–95.

116. Jensen NBK, Pötter R, Kirchheiner K, Fokdal L, Lindegaard JC, Kirisits C, et al. Bowel morbidity following radiochemotherapy and image-guided adaptive brachytherapy for cervical cancer: Physician- and patient reported outcome from the EMBRACE study. Radiotherapy and Oncology. 2018 Jun 1;127(3):431–9.

117. Georg P, Lang S, Dimopoulos JCA, Dörr W, Sturdza AE, Berger D, et al. Dose-volume histogram parameters and late side effects in magnetic resonance image-guided adaptive cervical cancer brachytherapy. Int J Radiat Oncol Biol Phys. 2011 Feb 1;79(2):356–62.

118. Kirchheiner K, Nout RA, Lindegaard JC, Haie-Meder C, Mahantshetty U, Segedin B, et al. Dose-effect relationship and risk factors for vaginal stenosis after definitive radio(chemo)therapy with image-guided brachytherapy for locally advanced cervical cancer in the EMBRACE study. Radiotherapy and Oncology. 2016 Jan 1;118(1):160–6.

119. Viswanathan AN, Thomadsen B. American Brachytherapy Society consensus guidelines for locally advanced carcinoma of the cervix. Part I: General principles. Brachytherapy. 2012 Jan;11(1):33–46.

120. Manea E, Escande A, Bockel S, Khettab M, Dumas I, Lazarescu I, et al. Risk of Late Urinary Complications Following Image Guided Adaptive Brachytherapy for Locally Advanced Cervical Cancer: Refining Bladder Dose-Volume Parameters. Int J Radiat Oncol Biol Phys. 2018 Jun 1;101(2):411–20.

121. Pötter R, Tanderup K, Kirisits C, de Leeuw A, Kirchheiner K, Nout R, et al. The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the EMBRACE studies. Vol. 9, Clinical and Translational Radiation Oncology. Elsevier Ireland Ltd; 2018. p. 48–60.

122. Petit C, Dumas I, Chargari C, Martinetti F, Maroun P, Doyeux K, et al. MRI-guided brachytherapy in locally advanced cervical cancer: Small bowel D0.1cm3 and D2cm3 are not predictive of late morbidity. Brachytherapy. 2016 Jul 1;15(4):463–70.

123. Mesko S, Swamy U, Park SJ, Borja L, Wang J, Demanes DJ, et al. Early clinical outcomes of ultrasound-guided CT-planned high-dose-rate interstitial brachytherapy for primary locally advanced cervical cancer. Brachytherapy. 2015 Sep 1;14(5):626–32.

124. Li R, Shinde A, Liu A, Glaser S, Lyou Y, Yuh ; Bertram, et al. Machine Learning-Based Interpretation and Visualization of Nonlinear Interactions in Prostate Cancer Survival. JCO Clin Cancer Inform. 2020;4:637–46.

125. Tian Z, Yen A, Zhou Z, Shen C, Albuquerque K, Hrycushko B. A machine-learning–based prediction model of fistula formation after interstitial brachytherapy for locally advanced gynecological malignancies. Brachytherapy. 2019 Jul 1;18(4):530–8.

126. Chen J, Chen H, Zhong Z, Wang Z, Hrycushko B, Zhou L, et al. Investigating rectal toxicity associated dosimetric features with deformable accumulated rectal surface dose maps for cervical cancer radiotherapy. Radiation Oncology. 2018 Jul 6;13(1).

127. Zhen X, Chen J, Zhong Z, Hrycushko B, Zhou L, Jiang S, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. Phys Med Biol. 2017 Oct 11;62(21):8246–63.

128. Lucia F, Bourbonne V, Visvikis D, Miranda O, Gujral DM, Gouders D, et al. Radiomics analysis of 3d dose distributions to predict toxicity of radiotherapy for cervical cancer. J Pers Med. 2021 May 1;11(5).

129. Cheon W, Han M, Jeong S, Oh ES, Lee SU, Lee SB, et al. Feature Importance Analysis of a Deep Learning Model for Predicting Late Bladder Toxicity Occurrence in Uterine Cervical Cancer Patients. Cancers (Basel). 2023 Jul 1;15(13).

130. Peng Y, Nagata MH. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. Chaos Solitons Fractals. 2020 Oct 1;139.

131. El Naqa I, Boone JM, Benedict SH, Goodsitt MM, Chan HP, Drukker K, et al. AI in medical physics: guidelines for publication. Med Phys. 2021 Sep 1;48(9):4711–4.

132. Ali MM, Paul BK, Ahmed K, Bui FM, Quinn JMW, Moni MA. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Comput Biol Med. 2021 Sep 1;136.

133. Chen S, Zhou S, Yin FF, Marks LB, Das SK. Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis. Med Phys. 2007;34(10):3808–14.

134. Isaksson LJ, Pepa M, Zaffaroni M, Marvaso G, Alterio D, Volpe S, et al. Machine Learning-Based Models for Prediction of Toxicity Outcomes in Radiotherapy. Vol. 10, Frontiers in Oncology. Frontiers Media S.A.; 2020.

135. Araújo ALD, Moraes MC, Pérez-de-Oliveira ME, Silva VM da, Saldivia-Siracusa C, Pedroso CM, et al. Machine learning for the prediction of toxicities from head and neck cancer treatment: A systematic review with meta-analysis. Vol. 140, Oral Oncology. Elsevier Ltd; 2023.

136. Perez-Calatayud J, Ballester F, Das RK, Dewerd LA, Ibbott GS, Meigooni AS, et al. Dose calculation for photon-emitting brachytherapy sources with average energy higher than 50 keV: Report of the AAPM and ESTRO. Med Phys. 2012;39(5):2904–29.

Appendix A

GammaMed HDR ^{192}Ir Plus Brachytherapy Source Model

All the data shown below, including pictures of the source model specifications, and the available dosimetric data is found in the repository of the BRAchytherapyPHYsics Quality Assurance System (BRAPHYQS) working group of GEC-ESTRO available in: https://www.estro.org/About/ESTRO-Organisation-Structure/Committees/GEC-ESTRO-Committee/GEC-ESTRO-BRAPHYQS/Ir-192-HDR. The original publication and specifications on how these values were obtained was done by Perez-Calatayud et al. (136)

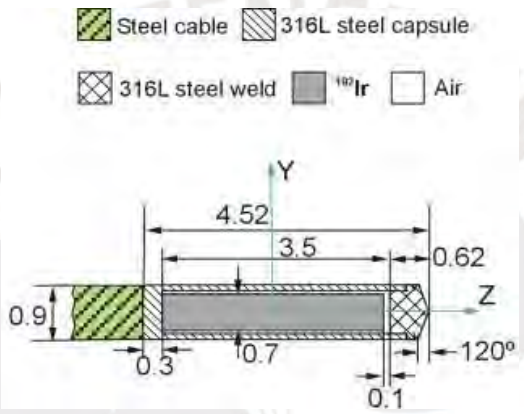


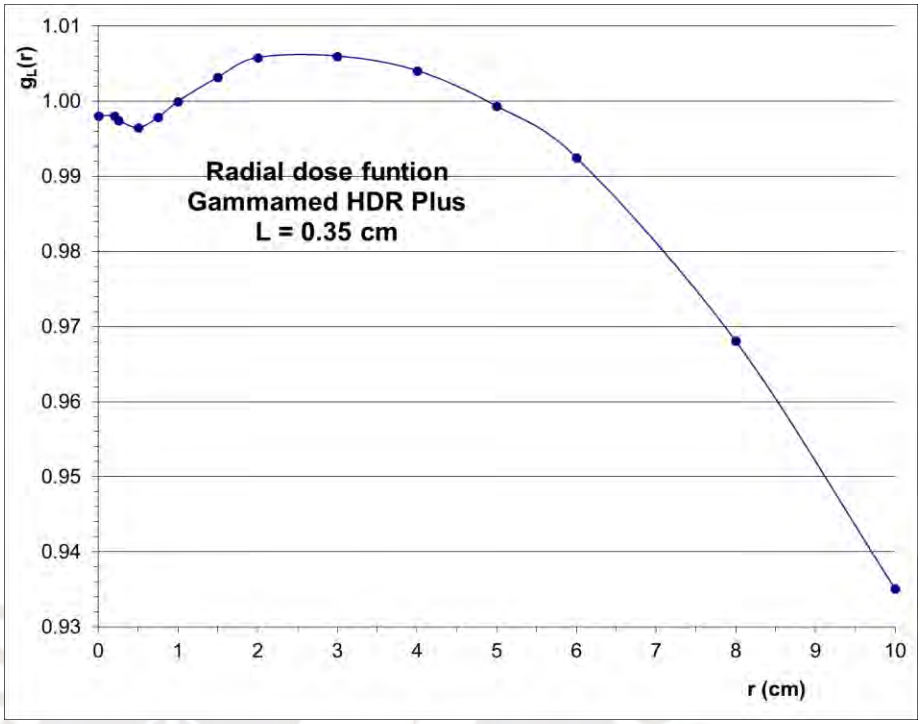
Figure A. GammaMed HDR ^{192}Ir Plus Brachytherapy Source produced by Varian Medical Systems

$$\Lambda = 1.1165 \frac{\text{cGy}}{h \cdot U}$$

Interpolated/extrapolated data are boldface/underlined. Values inside the source are in italics.

The radial dose function $g_L(r)$ for this L = 0.35 cm source model is as follows.

r (cm)	$g_L(r)$
0.00	*0.998*
0.20	0.998
0.25	**0.997**
0.50	**0.996**
0.75	**0.998**
1.00	1.000
1.50	1.003
2.00	1.006
3.00	1.006
4.00	1.004
5.00	0.999
6.00	0.993
8.00	0.968
10.00	0.935

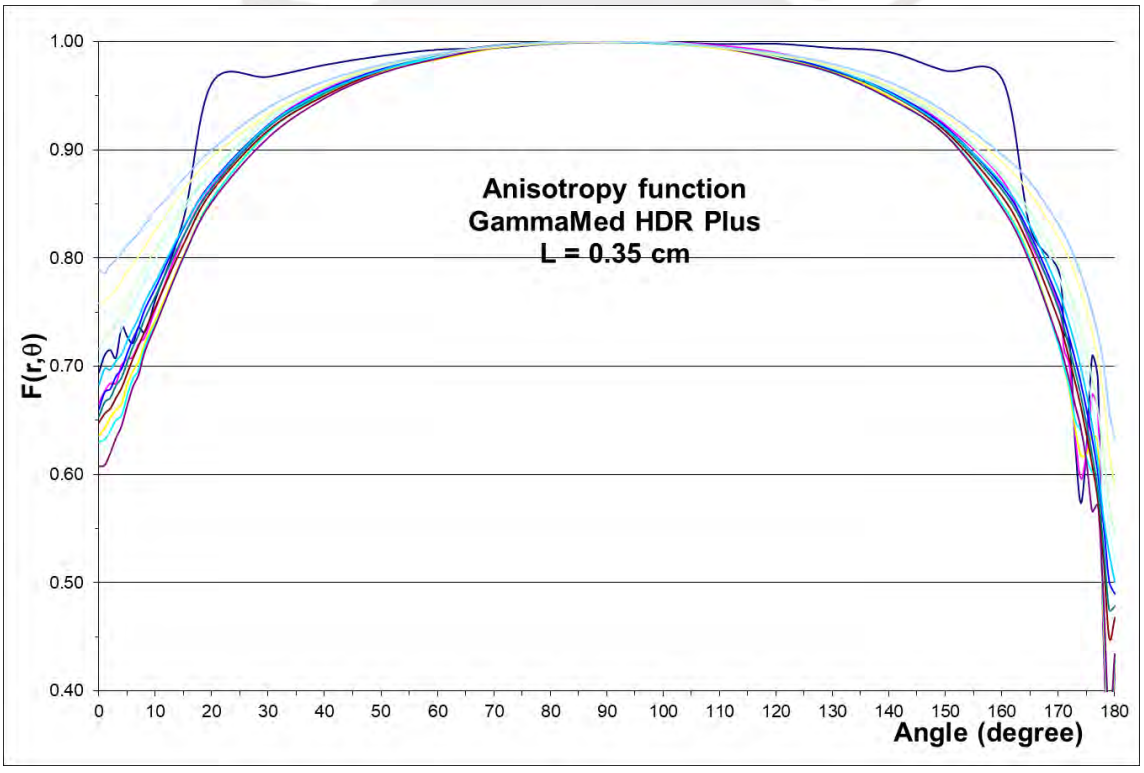


The 2D Anisotropy $F(r,\theta)$ is shown below:

Distance from Active Source Center (cm)

q (°)	0	0.20	0.40	0.60	0.80	1.00	1.25	1.50	1.75	2.00	2.50	3.00	3.50	4.00	5.00	6.00	8.00	10.00
0	*0.695*	*0.695*	0.666	0.636	0.630	0.608	0.615	0.634	0.625	0.629	0.648	0.654	0.660	0.683	0.702	0.716	0.758	0.789
1	*0.711*	*0.711*	0.677	0.643	0.632	0.609	0.614	0.632	0.626	0.633	0.656	0.667	0.676	0.698	0.716	0.727	0.762	0.786
2	*0.715*	*0.715*	0.684	0.653	0.640	0.620	0.624	0.638	0.634	0.641	0.661	0.671	0.679	0.697	0.717	0.730	0.764	0.794
3	*0.708*	*0.708*	0.684	0.660	0.650	0.634	0.637	0.647	0.646	0.653	0.671	0.682	0.691	0.705	0.726	0.739	0.771	0.798
4	*0.736*	*0.736*	0.701	0.666	0.654	0.645	0.652	0.662	0.663	0.668	0.680	0.690	0.697	0.711	0.736	0.750	0.779	0.805
5	*0.728*	*0.728*	0.706	0.684	0.673	0.664	0.668	0.674	0.676	0.682	0.693	0.704	0.712	0.724	0.748	0.762	0.788	0.811
6	*0.722*	*0.722*	0.709	0.696	0.688	0.681	0.683	0.686	0.691	0.697	0.707	0.718	0.725	0.734	0.757	0.770	0.795	0.817
7	*0.736*	*0.736*	0.720	0.705	0.697	0.692	0.697	0.700	0.707	0.712	0.719	0.729	0.736	0.745	0.769	0.780	0.802	0.823
8	*0.732*	*0.732*	0.726	0.720	0.715	0.712	0.713	0.713	0.719	0.724	0.731	0.743	0.751	0.758	0.779	0.789	0.810	0.830
9	*0.744*	*0.744*	0.738	0.733	0.729	0.726	0.728	0.727	0.734	0.738	0.743	0.754	0.762	0.769	0.790	0.799	0.818	0.837
10	*0.762*	*0.762*	0.753	0.743	0.738	0.738	0.741	0.740	0.748	0.752	0.755	0.765	0.772	0.778	0.799	0.808	0.826	0.844
15	*0.837*	*0.837*	0.820	0.803	0.801	0.802	0.804	0.802	0.809	0.811	0.813	0.821	0.828	0.829	0.844	0.848	0.863	0.873
20	*0.962*	0.962	0.866	0.855	0.854	0.852	0.853	0.852	0.858	0.858	0.862	0.865	0.870	0.869	0.878	0.880	0.893	0.900
30	*0.968*	0.968	0.923	0.917	0.916	0.912	0.913	0.912	0.918	0.918	0.918	0.921	0.923	0.923	0.927	0.930	0.933	0.939
40	*0.979*	0.979	0.956	0.951	0.952	0.948	0.949	0.946	0.950	0.951	0.950	0.953	0.955	0.954	0.958	0.958	0.959	0.963
50	*0.987*	0.987	0.977	0.973	0.973	0.971	0.972	0.971	0.972	0.973	0.973	0.974	0.975	0.974	0.977	0.976	0.978	0.978
60	*0.993*	0.993	0.987	0.984	0.985	0.985	0.987	0.985	0.987	0.988	0.989	0.989	0.988	0.988	0.989	0.988	0.988	0.988
70	*0.994*	0.994	0.995	0.994	0.994	0.995	0.996	0.993	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.995	0.995	0.996
80	*0.998*	0.998	0.999	0.999	0.998	0.999	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999	0.999
90	*1.000*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
100	*1.000*	1.000	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.999	0.999	0.999	0.999	1.000	0.999	1.000	0.999

110	*0.998*	0.998	0.996	0.994	0.995	0.995	0.995	0.993	0.995	0.994	0.995	0.995	0.995	0.995	0.995	0.994	0.995	0.995
120	*0.998*	0.998	0.990	0.985	0.984	0.985	0.986	0.984	0.987	0.987	0.987	0.987	0.988	0.989	0.988	0.987	0.989	0.989
130	*0.994*	0.994	0.976	0.973	0.971	0.972	0.972	0.970	0.972	0.973	0.974	0.974	0.974	0.975	0.976	0.977	0.977	0.979
140	*0.991*	0.991	0.958	0.950	0.947	0.948	0.947	0.947	0.949	0.950	0.952	0.953	0.954	0.953	0.956	0.958	0.959	0.962
150	*0.973*	0.973	0.923	0.914	0.914	0.914	0.914	0.912	0.915	0.916	0.917	0.919	0.921	0.920	0.925	0.927	0.931	0.935
160	*0.966*	0.966	0.873	0.851	0.850	0.847	0.850	0.848	0.853	0.856	0.857	0.863	0.867	0.870	0.876	0.878	0.889	0.895
165	*0.828*	*0.828*	0.814	0.801	0.798	0.796	0.798	0.801	0.802	0.806	0.809	0.818	0.822	0.829	0.838	0.843	0.860	0.870
170	*0.789*	*0.789*	0.756	0.723	0.720	0.725	0.725	0.729	0.730	0.734	0.743	0.754	0.761	0.771	0.784	0.795	0.819	0.831
171	*0.730*	*0.730*	0.715	0.700	0.699	0.706	0.706	0.710	0.712	0.716	0.725	0.735	0.744	0.755	0.768	0.782	0.809	0.822
172	*0.722*	*0.722*	0.699	0.676	0.679	0.686	0.686	0.689	0.692	0.696	0.705	0.716	0.725	0.738	0.752	0.768	0.798	0.811
173	*0.627*	*0.627*	*0.635*	0.642	0.649	0.664	0.663	0.666	0.671	0.675	0.684	0.695	0.705	0.719	0.735	0.752	0.785	0.799
174	*0.574*	*0.574*	*0.596*	0.618	0.640	0.643	0.641	0.643	0.650	0.653	0.662	0.672	0.684	0.698	0.715	0.733	0.769	0.784
175	*0.627*	*0.627*	*0.623*	*0.619*	0.615	0.611	0.615	0.616	0.625	0.627	0.636	0.646	0.659	0.674	0.692	0.713	0.752	0.769
176	*0.710*	*0.710*	*0.674*	*0.638*	0.602	0.566	0.568	0.585	0.597	0.598	0.608	0.618	0.632	0.645	0.666	0.688	0.731	0.750
177	*0.687*	*0.687*	*0.658*	*0.630*	*0.601*	*0.572*	0.536	0.500	0.565	0.564	0.575	0.584	0.600	0.610	0.634	0.659	0.706	0.729
178	*0.450*	*0.450*	*0.459*	*0.467*	*0.476*	*0.485*	*0.496*	*0.507*	0.518	0.529	0.521	0.533	0.553	0.560	0.591	0.620	0.672	0.702
179	*0.329*	*0.329*	*0.340*	*0.350*	*0.361*	*0.371*	*0.384*	*0.397*	*0.410*	*0.423*	*0.449*	*0.475*	0.501	0.527	0.548	0.572	0.621	0.659
180	*0.417*	*0.417*	*0.421*	*0.426*	*0.430*	*0.434*	*0.440*	*0.445*	*0.451*	*0.456*	*0.467*	*0.478*	*0.489*	*0.501*	*0.523*	*0.545*	0.589	0.633



Appendix B

New VariSource HDR VS2000 Brachytherapy Source Model

All the data shown below, including pictures of the source model specifications, and the available dosimetric data is found in the repository of the BRAchytherapyPHYsics Quality Assurance System (BRAPHYQS) working group of GEC-ESTRO available in: https://www.estro.org/About/ESTRO-Organisation-Structure/Committees/GEC-ESTRO-Committee/GEC-ESTRO-BRAPHYQS/Ir-192-HDR. The original publication and specifications on how these values were obtained was done by Perez-Calatayud et al. (136)

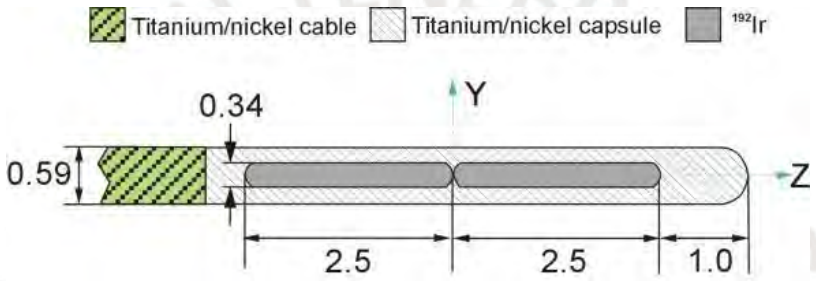


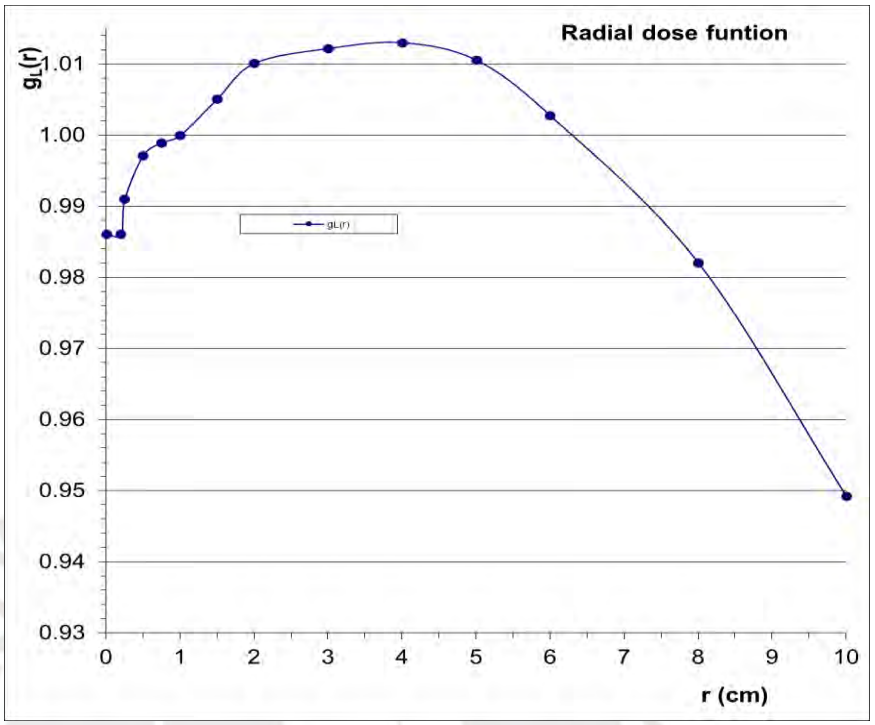
Figure B. New VariSource HDR VS2000 source model produced by Varian Medical Systems

$$\Lambda = 1.1 \frac{\text{cGy}}{h \cdot U}$$

Interpolated/extrapolated data are boldface/underlined. Values inside the source are in italics.

The radial dose function $g_L(r)$ for this L = 0.5 cm source model is as follows.

r (cm)	$g_L(r)$
0	*0.986*
0.20	0.986
0.25	0.991
0.50	0.997
0.75	0.999
1.00	1.000
1.50	1.005
2.00	1.010
3.00	1.012
4.00	1.013
5.00	1.011
6.00	1.003
8.00	0.982
10.00	0.949

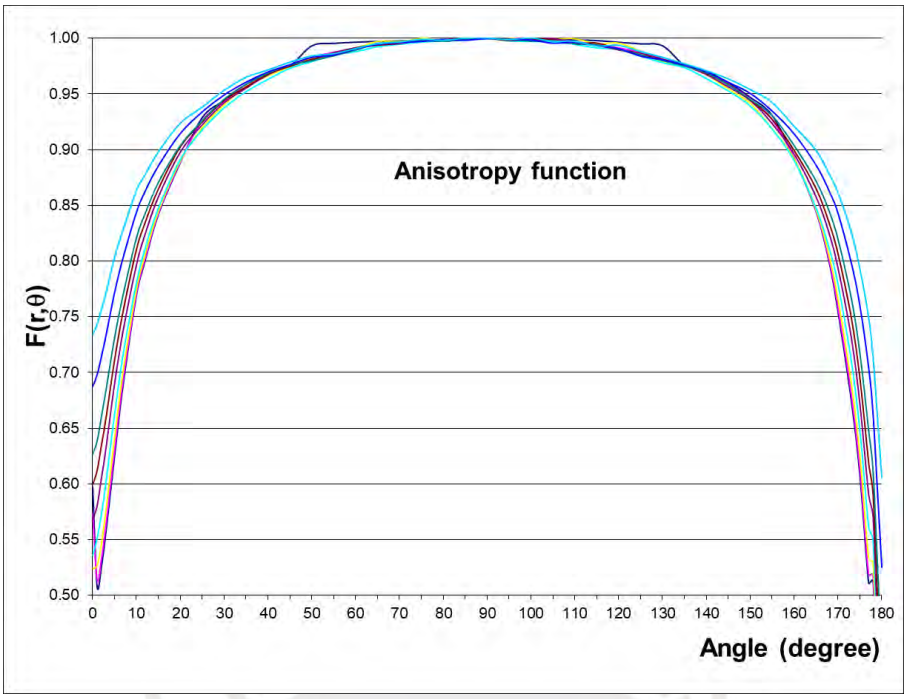


The 2D Anisotropy $F(r,\theta)$ is shown below:

Distance from Active Source Center (cm)

q (°)	0	0.25	0.50	1.00	2.00	3.00	4.00	5.00	7.50	10.00
0	*0.598*	*0.598*	0.573	0.524	0.536	0.568	0.599	0.627	0.688	0.734
1	*0.508*	*0.508*	*0.514*	0.527	0.551	0.580	0.612	0.638	0.698	0.743
2	*0.526*	*0.526*	*0.533*	0.546	0.573	0.606	0.635	0.661	0.715	0.757
3	*0.556*	*0.556*	*0.563*	0.576	0.601	0.631	0.658	0.684	0.733	0.771
5	*0.626*	*0.626*	*0.631*	0.642	0.663	0.690	0.712	0.732	0.772	0.804
7	*0.691*	*0.691*	*0.695*	0.703	0.718	0.739	0.756	0.771	0.804	0.829
10	*0.770*	*0.770*	*0.772*	0.775	0.781	0.797	0.811	0.821	0.844	0.864
12	*0.801*	*0.801*	*0.803*	0.806	0.812	0.825	0.835	0.843	0.863	0.878
15	*0.842*	*0.842*	*0.843*	0.845	0.848	0.858	0.865	0.871	0.886	0.898
20	*0.889*	*0.889*	*0.889*	0.890	0.892	0.898	0.902	0.904	0.915	0.924
25	*0.929*	*0.929*	0.926	0.920	0.918	0.922	0.925	0.926	0.934	0.938
30	*0.945*	*0.945*	0.943	0.940	0.938	0.942	0.944	0.946	0.949	0.953
35	*0.959*	*0.959*	0.958	0.957	0.952	0.955	0.956	0.959	0.961	0.965
40	*0.969*	*0.969*	0.968	0.966	0.963	0.967	0.968	0.969	0.970	0.971
45	*0.976*	*0.976*	0.975	0.974	0.973	0.975	0.975	0.976	0.977	0.978
50	*0.993*	0.993	0.983	0.982	0.979	0.981	0.981	0.980	0.983	0.984
55	*0.996*	0.996	0.987	0.986	0.986	0.987	0.985	0.985	0.985	0.986
60	*0.997*	0.997	0.991	0.991	0.988	0.992	0.991	0.990	0.991	0.991

65	*0.997*	0.997	0.994	0.997	0.993	0.994	0.994	0.996	0.995	0.995
70	*0.998*	0.998	0.997	0.997	0.995	0.995	0.996	0.997	0.996	0.996
75	*0.998*	0.998	0.998	1.000	0.997	0.998	0.998	0.998	0.997	0.999
80	*0.999*	0.999	0.998	1.002	0.998	0.998	0.997	0.997	0.999	0.998
85	*0.999*	0.999	1.000	1.001	0.998	0.999	0.999	1.000	0.999	1.000
90	*1.000*	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
95	*0.999*	0.999	1.001	1.000	0.998	0.999	0.998	1.000	0.999	0.999
100	*0.999*	0.999	0.998	1.005	0.997	0.999	0.999	0.998	0.999	1.000
105	*0.999*	0.999	0.998	1.000	0.997	1.000	0.998	0.998	0.996	0.997
110	*0.999*	0.999	0.997	0.999	0.995	0.995	0.995	0.995	0.997	0.997
115	*0.998*	0.998	0.994	0.995	0.992	0.994	0.995	0.994	0.994	0.993
120	*0.997*	0.997	0.991	0.996	0.990	0.990	0.990	0.990	0.991	0.993
125	*0.995*	0.995	0.987	0.988	0.984	0.987	0.985	0.985	0.984	0.988
130	*0.993*	0.993	0.981	0.979	0.978	0.981	0.981	0.980	0.981	0.983
135	*0.976*	*0.976*	0.976	0.977	0.974	0.975	0.976	0.975	0.976	0.977
140	*0.971*	*0.971*	0.969	0.967	0.964	0.967	0.968	0.969	0.970	0.971
145	*0.960*	*0.960*	0.958	0.955	0.953	0.956	0.958	0.958	0.961	0.963
150	*0.947*	*0.947*	0.945	0.941	0.940	0.944	0.944	0.945	0.951	0.954
155	*0.931*	*0.931*	0.927	0.918	0.919	0.923	0.927	0.931	0.935	0.942
160	*0.892*	*0.892*	*0.892*	0.892	0.891	0.896	0.899	0.904	0.914	0.921
165	*0.845*	*0.845*	*0.845*	0.846	0.848	0.858	0.865	0.872	0.886	0.899
168	*0.799*	*0.799*	*0.801*	0.804	0.812	0.825	0.834	0.845	0.863	0.878
170	*0.755*	*0.755*	*0.758*	*0.765*	0.779	0.794	0.806	0.820	0.843	0.861
173	*0.676*	*0.676*	*0.681*	*0.691*	0.712	0.732	0.750	0.767	0.799	0.826
175	*0.608*	*0.608*	*0.613*	*0.625*	*0.648*	0.671	0.695	0.717	0.760	0.793
177	*0.511*	*0.511*	*0.518*	*0.533*	*0.561*	*0.590*	0.619	0.648	0.704	0.749
178	*0.514*	*0.514*	*0.519*	*0.529*	*0.550*	*0.570*	*0.591*	*0.611*	0.662	0.714
179	*0.402*	*0.402*	*0.409*	*0.422*	*0.449*	*0.475*	*0.501*	*0.528*	0.594	0.660
180	*0.291*	*0.291*	*0.299*	*0.315*	*0.348*	*0.380*	*0.412*	*0.444*	*0.525*	*0.606*


Anisotropy function
F(r,θ)
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
Angle (degree)

Appendix C

Balancing Class Algorithms Pseudocode

SMOTE:

Algorithm *SMOTE*(T, N, k)
Input: Number of minority class samples T; Amount of SMOTE $N\%$; Number of nearest neighbors k
Output: $(N/100)$ * T synthetic minority class samples

1. (∗ *If N is less than 100%, randomize the minority class samples as only a random percent of them will be SMOTEd.* ∗)
2. **if** $N < 100$
3. **then** Randomize the T minority class samples
4. $T = (N/100) * T$
5. $N = 100$
6. **endif**
7. $N = (int)(N/100)$ (∗ *The amount of SMOTE is assumed to be in integral multiples of 100.* ∗)
8. k = Number of nearest neighbors
9. $numattrs$ = Number of attributes
10. $Sample[\,][\,]$: array for original minority class samples
11. $newindex$: keeps a count of number of synthetic samples generated, initialized to 0
12. $Synthetic[\,][\,]$: array for synthetic samples
(∗ *Compute k nearest neighbors for each minority class sample only.* ∗)
13. **for** $i \leftarrow 1$ **to** T
14. Compute k nearest neighbors for i, and save the indices in the $nnarray$
15. Populate(N, i, $nnarray$)
16. **endfor**

$Populate(N,\ i,\ nnarray)$ (∗ *Function to generate the synthetic samples.* ∗)

17. **while** $N \neq 0$
18. Choose a random number between 1 and k, call it nn. This step chooses one of the k nearest neighbors of i.
19. **for** $attr \leftarrow 1$ **to** $numattrs$
20. Compute: $dif = Sample[nnarray[nn]][attr] - Sample[i][attr]$
21. Compute: gap = random number between 0 and 1
22. $Synthetic[newindex][attr] = Sample[i][attr] + gap * dif$
23. **endfor**
24. $newindex{+}{+}$
25. $N = N - 1$
26. **endwhile**
27. **return** (∗ *End of Populate.* ∗)

End of Pseudo-Code.

SVM-SMOTE:

Table 1 Borderline over-sampling algorithm

Input:

- $\boldsymbol{X}$: Training set;
- N: Over-sampling level, such as 100, 200, 300, … (%);
- k: Number of minority nearest neighbours used for over-sampling;
- m: Number of nearest neighbours used to decide how to over-sample (extrapolation or interpolation);

Output:

- $\boldsymbol{X}_{new}$: Over-sampled training set;

Variables:

- $\boldsymbol{X}^+$: Set of minority training instances;
- $\boldsymbol{SV}^+$: Set of minority support vectors;
- T: Total number of the artificial minority instances to be generated;
- ***amount***: Array that contains the number of times to over-sample each minority support vector;
- ***nn***: Array that contains k minority nearest neighbours of each minority support vector;

Begin

1 Extract $\boldsymbol{X}^+$ from $\boldsymbol{X}$;

2 $T \leftarrow (N/100) \times |\boldsymbol{X}^+|$;

3 Compute $\boldsymbol{SV}^+$ by training an SVM on $\boldsymbol{X}$;

4 Compute ***amount*** by evenly distributing T among $\boldsymbol{SV}^+$;

5 Compute ***nn*** from $\boldsymbol{X}^+$;

6 For each $\boldsymbol{sv}_i^+ \in \boldsymbol{SV}^+$, find m nearest neighbours from $\boldsymbol{X}$. If the number of majority nearest neighbours $< m/2$, along the extension lines joining $\boldsymbol{sv}_i^+$ to its k minority nearest neighbours (in order of the first to the k-th nearest neighbour), extrapolate ***amount***$[i]$ artificial minority instances using the following formula:

$$\boldsymbol{x}_{new}^+ = \boldsymbol{sv}_i^+ + \rho(\boldsymbol{sv}_i^+ - \boldsymbol{nn}[i][j])$$

where ***nn***$[i][j]$ is the j-th minority nearest neighbour of $\boldsymbol{sv}_i^+ (1 \le j \le k)$, ρ is a random number in [0, 1]. Otherwise, interpolate using the following formula:

$$\boldsymbol{x}_{new}^+ = \boldsymbol{sv}_i^+ + \rho(\boldsymbol{nn}[i][j] - \boldsymbol{sv}_i^+);$$

7 $\boldsymbol{X}_{new} = \boldsymbol{X} \cup \{\boldsymbol{x}_{new}^+\}$;

End

ADASYN:

[Algorithm - ADASYN]

Input

(1) Training data set D_{tr} with m samples $\{\mathbf{x_i}, y_i\}$, $i = 1, ..., m$, where $\boldsymbol{x_i}$ is an instance in the n dimensional feature space $\boldsymbol{X}$ and $y_i \in Y = \{1, -1\}$ is the class identity label associated with $\boldsymbol{x_i}$. Define m_s and m_l as the number of minority class examples and the number of majority class examples, respectively. Therefore, $m_s \leq m_l$ and $m_s + m_l = m$.

Procedure

(1) Calculate the degree of class imbalance:

$$d = m_s/m_l \tag{1}$$

where $d \in (0,\ 1]$.

(2) If $d < d_{th}$ then (d_{th} is a preset threshold for the maximum tolerated degree of class imbalance ratio):

(a) Calculate the number of synthetic data examples that need to be generated for the minority class:

$$G = (m_l - m_s) \times \beta \tag{2}$$

Where $\beta \in [0,\ 1]$ is a parameter used to specify the desired balance level after generation of the synthetic data. $\beta = 1$ means a fully balanced data set is created after the generalization process.

(b) For each example $\boldsymbol{x_i} \in minorityclass$, find K nearest neighbors based on the Euclidean distance in n dimensional space, and calculate the ratio r_i defined as:

$$r_i = \Delta_i/K, \quad i = 1, ..., m_s \tag{3}$$

where Δ_i is the number of examples in the K nearest neighbors of $\boldsymbol{x_i}$ that belong to the majority class, therefore $r_i \in [0,\ 1]$;

(c) Normalize r_i according to $\hat{r_i} = r_i / \sum_{i=1}^{m_s} r_i$, so that $\hat{r_i}$ is a density distribution ($\sum_i \hat{r_i} = 1$)

(d) Calculate the number of synthetic data examples that need to be generated for each minority example $\boldsymbol{x_i}$:

$$g_i = \hat{r_i} \times G \tag{4}$$

where G is the total number of synthetic data examples that need to be generated for the minority class as defined in Equation (2).

(e) For each minority class data example $\boldsymbol{x_i}$, generate g_i synthetic data examples according to the following steps:

Do the **Loop** from 1 to g_i:

(i) Randomly choose one minority data example, $\boldsymbol{x_{zi}}$, from the K nearest neighbors for data $\boldsymbol{x_i}$.

(ii) Generate the synthetic data example:

$$\boldsymbol{s_i} = \boldsymbol{x_i} + (\boldsymbol{x_{zi}} - \boldsymbol{x_i}) \times \lambda \tag{5}$$

where $(\boldsymbol{x_{zi}} - \boldsymbol{x_i})$ is the difference vector in n dimensional spaces, and λ is a random number: $\lambda \in [0,\ 1]$.

End **Loop**

Appendix D

Comparison of Toxicity Data between Syed-Neblett and T&O Applicators, Segregated by MRI Imaging Status

		Combined		No MRI		MRI		p-value for Difference
		n	%	n	%	n	%	
GI	No Toxicity	41	36	16	33	25	38	0.694
	Grade 1 or higher	74	64	33	67	41	62	
	1	30	26	14	29	16	24	0.977
	2	28	24	11	22	17	26	
	3	15	13	8	16	7	11	
	4	1	1	0	0	1	2	
GU	No Toxicity	57	50	18	37	39	59	***0.024***
	Grade 1 or higher	58	50	31	63	27	41	
	1	21	18	13	27	8	12	0.691
	2	22	19	9	18	13	20	
	3	14	12	9	18	5	8	
	4	1	1	0	0	1	2	
Vag	No Toxicity	54	47	22	45	32	48	0.711
	Grade 1 or higher	61	53	27	55	34	52	
	1	24	21	8	16	16	24	0.223
	2	18	16	9	18	9	14	
	3	18	16	10	20	8	12	
	4	1	1	0	0	1	2	

Table A.1. Comparative distribution of toxicities over time (acute and late) by grade for non-MRI versus MRI with a Syed-Neblett applicator.

		Combined		No MRI		MRI		p-value for Difference
		n	%	n	%	n	%	
GI	No Toxicity	11	19	4	14	7	23	0.508
	Grade 1 or higher	47	81	24	86	23	77	
	1	17	29	10	36	7	23	0.421
	2	23	40	11	39	12	40	
	3	7	12	3	11	4	13	
GU	No Toxicity	17	29	3	11	14	47	***0.004***
	Grade 1 or higher	41	71	25	89	16	53	
	1	29	50	18	64	11	37	0.973
	2	10	17	5	18	5	17	
	3	2	3	2	7	0	0	
Vag	No Toxicity	13	22	4	14	9	30	0.212
	Grade 1 or higher	45	78	24	86	21	70	
	1	27	47	11	39	16	53	***0.023***
	2	10	17	6	21	4	13	
	3	7	12	6	21	1	3	
	4	1	2	1	4	0	0	

Table A.2. Comparative distribution of toxicities over time (acute and late) by grade for non-MRI versus MRI with a tandem and ovoid applicator.

Appendix E

Academic Achievements

Publications

Submitted to Journal "PLOS ONE": Advancing Patient Care: Machine Learning Models for Predicting Grade 3+ Toxicities in Gynecologic Cancer Patients treated with HDR Brachytherapy

Portocarrero-Bonifaz, A., Syed, S., Kassel, M., McKenzie, G.W., Shah, V.M., Forry, B.M., Gaskins, J.T., Sowards, K.T., Avula, T.B., Masters, A. and Silva, S.R. (2024). Dosimetric and toxicity comparison between Syed-Neblett and Fletcher-Suit-Delclos Tandem and Ovoid applicators in high dose rate cervix cancer brachytherapy. Brachytherapy. https://doi.org/10.1016/j.brachy.2024.03.003

Events

ESTRO 2024. May 2024, Glasgow – United Kingdom. "*Dosimetric comparison between different applicators for high dose rate gynecologic brachytherapy*"

ARS 105th Annual Meeting. May 2023, Hawaii - USA. "*Dosimetric Comparison Between Syed-Neblett and Fletcher-Suit-Delclos Applicators in High Dose Rate Gynecologic Brachytherapy*"