André Filipe dos Santos Oliveira

Predicting Surgery Radiation Exposure using Machine Learning Methods

Biomedical Engineering Master's Dissertation Report

ORIENTADOR
Professor Doutor Miguel López
Professora Doutora Catarina Santos

Dezembro, 2023

André Filipe dos Santos Oliveira

Predicting Surgery Radiation Exposure using Machine Learning Methods

JÚRI

Presidente: Professor Doutor Célio Pina, Instituto Politécnico de Setúbal (IPS/ESTS)

Vogal: Professor Doutor Luís Magalhães, Universidade do Minho, Escola de Engenharia

Vogal: Doutor Nuno Cristino, Hospital da Luz, Setúbal

Vogal: Professor Doutor André Castro, Instituto Politécnico de Setúbal (IPS/ESTS)

Orientador: Professor Doutor Miguel López, Instituto Politécnico de Setúbal (IPS/ESTS)

Dezembro, 2023

Dedico esta dissertação aos meus pais, amigos
e namorada, por todo o apoio nesta jornada.

Acknowledgments

Desde mais quero agradecer primeiramente aos meus orientadores, Professor Miguel López pela paciência em cada chamada e pela dedicação em ensinar-me programação e à Professora Catarina Santos por todos os "puxões de orelhas" e incentivo em adiantar trabalho. Ao Dr. Nuno Cristino por ter confiado em mim para a realização deste estudo, por procurar sempre aprender mais sobre cada novo tema e pela motivação para continuar a tentar novas abordagens, pelo fornecimento dos dados médicos, pela oportunidade de assistir a uma cirurgia no Hospital da Luz de Setúbal e pelos conhecimentos transmitidos. Um agradecimento especial aos Professores Ricardo Baptista e Célio Pina, pela oportunidade de frequentar o Mestrado em Engenharia Biomédica na Escola Superior de Tecnologia de Setúbal, e aos restantes professores, pelos conhecimentos transmitidos e experiências adquiridas.

Agradeço aos meus pais pelo incentivo e apoio que me deram ao longo de todo o meu percurso académico e especialmente nestes últimos meses motivando-me sempre a continuar os meus estudos e a enfrentar os desafios que me fossem postos. Assim como me darem na cabeça e dizerem-me para esta noite não me deitar tão tarde.

Obrigado a todos os meus amigos por me apoiarem, ajudarem e desencaminharem sempre que eu precisei, em especial ao Pedro Pereira, ao Rúben Ferreira e ao meu Mestre Tiago Ferreira por todas as horas em chamada a ouvir-me reclamar e a mostrar o quão piores eles estavam. Não podendo esquecer o Ricardo Evangelista, o meu parceiro de mestrado, de percorrer hospitais, de futuro artigo científico e o maior lagarto que já conheci.

E por fim, mas não menos importante quero agradecer-te Ana Rita Costa, por ao fim de tantos anos decidires entrar nesta aventura na altura certa. Por me acelerares e ajudares nos momentos mais críticos, por zelares pela estética deste documento assim como mil outras coisas. Obrigado por todas as manhãs acordares com uma lista do que eu tinha feito nessa noite e me apoiares em como o trabalho estava a ficar bom. Sem ti não era a mesma coisa, definitivamente.

Resumo

A exposição à radiação é uma preocupação significativa tanto para pacientes quanto para profissionais de saúde durante procedimentos cirúrgicos. O princípio de otimização "*As Low As Reasonably Achievable*" (ALARA) recomenda minimizar a dosagem de radiação ao nível prático mais baixo, mantendo a qualidade da imagem. O Conselho Nacional de Proteção e Medidas de Radiação (NCRP) dos EUA define um Nível de Dose de Radiação Substancial (SRDL) com níveis de referência de *Air Kerma* (AK) = 5 Gy, *Kerma Area Product* (KAP) = 500 Gy.cm2 e Tempo de Exposição = 60 minutos.

Este estudo visa desenvolver métodos de *Machine Learning* (ML) para estimar a exposição à radiação em pacientes submetidos a procedimentos cirúrgicos. Isso envolve a identificação de fatores-chave que influenciam o uso de radiação, o desenvolvimento de modelos de ML para prever a exposição à radiação, a seleção de métricas para medir a precisão do modelo e o fornecimento de feedback aos cirurgiões. O objetivo é contribuir para o avanço da predição da exposição à radiação e aumentar a segurança do paciente, otimizando os resultados cirúrgicos.

O estudo visa investigar a viabilidade de prever a exposição à radiação recebida pelos pacientes durante a imagem intraoperatória usando modelos de regressão. Os parâmetros Tempo de Exposição e Número de Disparos foram identificados como fatores-chave que influenciam a exposição à radiação.

Neste estudo foram explorados cinco modelos de regressão (*Random Forest*, *Support Vector Regression*, *Artificial Neural Network*, *Decision Tree* e *LASSO*), utilizando um conjunto de dados que contêm dados de radiação de casos de pacientes. Primeiramente, foram realizadas análises estatísticas para obter um entendimento mais profundo do conjunto de dados, com foco na identificação dos fatores que influenciam significativamente a exposição à radiação e das características que apresentam a mais forte correlação com as variáveis de exposição à radiação. Em seguida, várias técnicas de ML, incluindo *feature selection* e *hyperparameter tuning*, foram aplicadas para aprimorar o poder preditivo dos modelos. O desempenho dos modelos foi avaliado usando métricas de regressão como *Mean Square Error* (MSE) e Mean Absolute Error (MAE). Além disso, o *Cross Validation* (CV) foi utilizado para garantir a capacidade de generalização dos modelos e minimizar o *overfitting*.

O modelo *Random Forest* (RF) demonstrou desempenho superior em comparação aos outros quatro modelos, produzindo consistentemente valores mais próximos de 100% de precisão para ambas as métricas de avaliação. Como principal contribuição, os resultados do

estudo demonstram a viabilidade de prever a exposição à radiação usando características obtidas pré cirurgia, com potenciais aplicações em diversas especialidades médicas. No entanto, é necessária uma pesquisa mais aprofundada para validar esses modelos em populações de pacientes maiores e mais diversificadas, incorporando características adicionais relacionadas à experiência do cirurgião e do técnico, e através do desenvolvimento de modelos para outras técnicas de imagiologia.

Palavras-Chave: Machine Learning; Exposição à Radiação; KAP*; AK; Tempo de Exposição

Abstract

Radiation exposure is a significant concern for both patients and healthcare professionals during surgical procedures. The “As Low As Reasonably Achievable” (ALARA) optimization principle recommends minimizing radiation dosage to the lowest practical level while maintaining image quality. The National Council on Radiation Protection and Measurements (NCRP) defines a Substantial Radiation Dose Level (SRDL) with reference levels of Air Kerma (AK)=5 Gy, Kerma Area Product (KAP)=500 Gy.cm2, and Exposure Time=60 minutes. This study aims to develop Machine Learning (ML) methods to estimate radiation exposure for patients undergoing surgical procedures. This will involve identifying key factors influencing radiation usage, developing ML models to predict radiation exposure, selecting metrics to measure model accuracy, and providing feedback to surgeons. The goal is to contribute to the advancement of radiation exposure prediction and enhance patient safety by optimizing surgical outcomes.

The study aims to investigate the feasibility of predicting radiation exposure received by patients during intraoperative imaging using regression models. *Exposure Time* and *Number of Shots* parameters were identified as key factors influencing radiation exposure.

Five regression models were explored in this study (Random Forest, Support Vector Regression, Artificial Neural Network, Decision Tree, and LASSO), employing a dataset comprising patients’ cases radiation data. First, statistical analyses were conducted to gain a deeper understanding of the dataset, with a focus on identifying the factors that significantly influence radiation exposure and the characteristics that exhibit the strongest correlation with radiation exposure variables. To follow, several machine learning techniques, including feature selection and hyperparameter tuning, were applied to enhance the predictive power of the models. The performance of the models was evaluated using regression metrics such as mean squared error (MSE) and mean absolute error (MAE). Additionally, cross-validation (CV) was employed to ensure the generalization ability of the models and minimize overfitting.

The Random Forest (RF) model demonstrated superior performance compared to the other four models, consistently producing values closer to 100% of accuracy for both evaluation metrics. As a main contribution, the study's findings demonstrate the feasibility of predicting radiation exposure using preoperatively obtained features, with potential applications in various medical specialties. However, further research is needed to validate these models in larger and more diverse patient populations, incorporating additional features related to surgeon and technician experience, and developing models for other imaging techniques.

Keywords: Machine Learning; Radiation Exposure; KAP*; AK; Exposure Time

Index

List of Figures

List of Tables

List of Acronyms

AK: Air Kerma

ALARA: As Low As Reasonably Achievable

ANN: Artificial Neural Networks

BMI: Body Mass Index

CT: Computed Tomography

CV: Cross-Validation

DAP: Dose Area Product

DT: Decision Tree

Gy: Gray

IQR: Interquartile Range

KAP: Kerma Area Product

LR: Linear Regression

LASSO: Least Absolute Shrinkage and Selection Operator

MAE: Mean Absolute Error

mGy: Miligray

ML: Machine Learning

MRI: Magnetic Resonance Imaging

MSE: Mean Squared Error

NCRP: National Council on Radiation Protection and Measurements

OR: Operating Room

OTS: Osteosyndissertation

R^2: R-Squared

RBF: Radial Basis Function

ReLU: Rectified Linear Unit

RF: Random Forest

RMSE: Root Mean Squared Error

SD: Standard Deviation

SRDL: Substantial Radiation Dose Level

SV: Sievert

SVR: Support Vector Regression

Chapter 1

Chapter 1 presents an introduction about radiation and its use in intraoperative imaging, explaining the motivation for this study, as well as the identification of the problem. Other topics were addressed, such as this dissertation contributions and its structure.

1. Introduction

Surgical procedures requiring radiation support carry significant risks for both patients and healthcare professionals due to cumulative radiation exposure. These risks can lead to the development of long-term diseases [1]. However, because ionizing radiation is necessary in medical practice, monitoring radiation exposure doses becomes imperative [1, 2]. To minimize risk exposure, the optimization principle known as "As Low As Reasonably Achievable" (ALARA) should be followed whenever possible. ALARA involves reducing the radiation dose to the lowest reasonable extent while maintaining image quality [3]. In accordance with the recommendations of the National Council on Radiation Protection and Measurements (NCRP) of USA, a Substantial Radiation Dose Level (SRDL) should be defined, with reference Air Kerma (AK) set at 5 Gy, Kerma Area Product (KAP) at 500 $Gy.cm^2$, and Exposure Time of 60 minutes, during a single procedure. To ensure these reference levels, it is mandatory to maintain radiation safety during surgical procedures [3].

Air Kerma represents the dose of radiation delivered to the patient, during X-ray imaging and is associated with radiation exposure of imaging professionals [4]. It´s defined as the amount of energy deposited by ionizing radiation in a unit of air and it is measured in units of Gray (Gy) or milliGray (mGy).

Kerma Area Product, which is also known as Dose Area Product (DAP), is the product of the irradiated area and the AK. Unlike AK, KAP remains constant regardless of the distance from the X-ray source. It quantifies the total amount of radiation delivered to the patient during radiological procedures using X-ray beams and is used for risk estimation. The value of KAP* includes a correction factor based on the equipment used and is typically measured in units of $Gy.cm^2$ or $\mu Gy.m^2$. It serves as an important parameter in radiation protection [5].

Exposure Time refers to the total duration of a specific type of exposure, such as fluoroscopy or single shot. Usually measured in seconds, the duration can vary depending on the complexity, nature of the procedure, patient characteristics and the health professional involved.

Minimally invasive procedures reduce the surgical dissection, blood loss and surgical time. They are associated with shorter hospital stays, faster and safer patient recovery, and are considered the best approach for many surgical procedures. Unfortunately, most of them require fluoroscopy-guidance and increase the exposure of the patients to ionizing radiation. [6].

Various options for intraoperative imaging have been developed, each with its own set of unique advantages and disadvantages. Single shot exposure involves producing radiation in short pulses. This results in lower radiation exposure compared to continuous fluoroscopy, and provides superior temporal resolution for capturing moving objects, as it generates stop-motion images. Although the resulting images may not appear smooth on the monitor, they offer sharper detail than those obtained through continuous fluoroscopy since the object is virtually still during each X-ray pulse [7]. Fluoroscopy represents the view of a continuous X-ray image on a monitor, much like an X-ray movie, which results in superior interpretation in real time. Dose reduction techniques like pulsed or intermittent fluoroscopy and last image hold can effectively reduce the exposure of patient and health professionals [8].

1.1. Motivation

Radiation procedures are necessary to produce images for medical diagnostic and interventional procedures, and it is crucial to monitor and record radiation doses, due to the development of reference radiation dose levels. This is especially important when fluoroscopic guidance is employed, since radiation doses can be higher compared to other procedures. This work aims to examine the factors that influence radiation exposure, it's correlation with individual patients' cases variables and the type of procedure, and how machine learning-based regression models can provide a precise (in real-time) estimations of AK and KAP* values. These estimates can assist surgeons and technicians in monitoring radiation quantities and modifying their procedures to minimize radiation risks, whenever appropriate.

1.2. Problem Identification

During surgery, it is common for the surgeon to ask for images, obtained through radiation. However, there are often difficulties during the imaging process, causing the patient to receive more radiation than needed. Patients who are exposed to radiation over long periods of time and several times, are at a higher risk of serious health problems, with the highest exposure levels occurring during spine surgery [9].

The consequences of overexposure to radiation during surgery is relatively low, however, in different surgical specialties, difficulties arise in some procedures, exposing the patient to

higher levels of radiation. For example, in endovascular aortic procedures every Sievert (Sv) of radiation absorbed by the body increases the lifetime risk of cancer by 5.5%. There is a latent period of around 10 years for malignant transformation following radiation exposure [10].

A substantial quantity of research has focused on radiation exposure during surgery, and this dissertation will focus the investigation of predictive regression models for optimizing radiation dosage and exposure time across four distinct surgical specialties: Spine Surgery, Orthopedics, Urology, and General Surgery/Gastro-Enterology. The dissertation will examine the application of these models to various surgical procedures within each specialty.

1.3. Objectives

The aim of this study is to explore machine learning techniques to estimate the amount of radiation exposure that a patient will receive during a surgical procedure / operation. To achieve this objective, our research seeks to accomplish the following specific objectives:

1) Based on mathematical analysis and statistical techniques, identifying the key factors/features that greatly influence radiation exposure.
2) To machine learn a dataset of selected features, producing precise regression models to allow the previous estimation of radiation exposure that a patient will receive during a specific surgical procedure.
3) Select appropriate metrics to measure the accuracy of the machine learning models to be developed.

By achieving these objectives, the aim is to contribute to the advancement of radiation dosage prediction in surgical settings. It is expected that developed machine learning models will have the potential to enhance patient safety by providing accurate and timely information on radiation dosage, enabling healthcare professionals to minimize risks and optimize surgical outcomes.

1.4. Dissertation Contributions

The contributions of the Dissertation are as follow:

- Based on a previous statistical analysis, it was explored, developed and fine-tuned several high performances machine learning regression models to estimate / predict the radiation exposure that a patient will receive during a specific surgical procedure.
- In collaboration with Hospital da Luz Setúbal, it was created a reference dataset that includes approximately 1,000 cases of surgical patients with radiation exposure doses.

1.5. Dissertation Structure

This dissertation is divided into five chapters. Chapter 1 introduces the dissertation topic, objectives, problem identification, motivation, as well as the main dissertation contributions. Chapter 2 includes a review of the relevant literature and a statement of the research in question. In Chapter 3 it is presented the methodology used to conduct the research, including the dataset used and regression models procedures. Chapter 4 presents the results and discussion of the research, including tables and figures that summarize the key findings, it also discusses the results considering the existing literature and draws out the implications of the research. Finally, Chapter 5 concludes the main findings and directions for future research.

Chapter 2

Chapter 2 covers the literature review of the main themes of this dissertation. Topics such as radiation in the surgical environment and the specialties covered, as well as the radiation-related problems. Machine learning models and related evaluation metrics are also looked in depth.

2. Literature Review

Intraoperative imaging is essential for various surgical specialties, but it exposes patients, surgeons, and staff to radiation. This radiation can cause cell damage and lead to various health problems, such as cancer, skin damage, reproductive system effects [8]. Patients in fluoroscopically guided interventions (FGI), may also be exposed to substantial radiation doses. The most commonly reported adverse reactions are skin injuries [11].

There are several ways to reduce radiation exposure, through shielding, optimization of fluoroscopic settings and techniques, and fluoroscopic dose reduction. To help with the later it is important to acquire more information of the procedure in question and the patient subjected to it, since it can help in the development of a patient-specific dose prediction model through ML techniques.

2.1. Intraoperative Imaging

Most surgeries use intraoperative imaging techniques to guide the surgeon and provide real-time feedback on the progress of the procedure. It can be used to identify and localize the surgical target, assess the extent of the disease, guide the surgical procedure, and verify the results [12]. A variety of intraoperative imaging techniques are available, including fluoroscopy, ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI).

In this spectrum of techniques, X-rays have become the most widely used assessment tool during surgical procedures, specifically musculoskeletal cases. However, it is used in many procedures because of the support it gives surgeons during the operation.

The type of exposure in the different surgical procedures encompassed in the specialties analyzed throughout this study, consists of fluoroscopy or steady radiographic images, which are called single shots. The difference between them is their purpose in surgery, since radiography produces a single image of the patient's internal structure, and fluoroscopy uses X-rays to create a real-time video [13].

There are several medical imaging devices capable of assisting the surgeons and operative room staff, but the C-arm is more practical for both minimally invasive and more complex surgeries. A C-arm consists of an X-ray source and a detector at opposite ends of a C-shaped arm. Most C-arms currently use an image intensifier because they can convert X-rays into visible light very quickly [14]. The X-rays pass through the object between the two ends of the C-arm and are weakened depending on the object's radiopacity, as shown in Figure 1. The X-rays that pass through the object hit a fluorescent surface, which produces light photons. These photons are accelerated in a tube, electronically intensified, and detected by a video camera. The camera is either directly connected to a TV screen or, as used in current technology, digitalized, and sent to a computer for post-processing and storage [15].

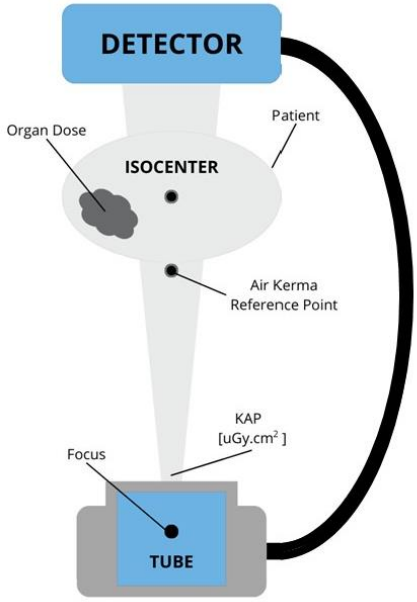

Figure 1. Schematic representation of a C arm. Adapted from [12].

Fluoroscopy is used for a wide variety of clinical purposes, including spine and orthopedic imaging, to better guide the placement of osteosyndissertation (OTS) materials, the identification of several types of lesions in the musculoskeletal system, and to guide minimally invasive procedures such as stent placement, cystoscopy and ureteroscopy [17].

2.1.1. Surgical Specialties using Intraoperative Imaging

Intraoperative radiographic images are very important in spine surgery, to ensure accurate placement of spinal instrumentation. Since minimally invasive procedures need the insertion of instruments percutaneously without anatomic visualization, the use of intraoperative imaging is especially necessary. Although open procedures allow the surgeon to check the exact location

of the injured area, radiographic imaging importance is not diminished, since it helps providing live images to confirm the affected area [8].

There are several procedures performed in both the anterior and posterior regions of the spine, meant to treat degenerative, traumatic, and neoplastic pathologies. These would be more difficult to perform without intraoperative imaging, due to the imprecise positioning of structures such as pedicle screws, potentially leading to complications during surgery and in the postoperative period [8].

In the field of Orthopedics, various procedures are performed, of which the accurate placement and removal of OTS material are the most necessary. The use of intraoperative imaging is essential for this type of minimally invasive surgeries [18]. This kind of procedures is normally performed in peripheral structures of the body or small area segments, since the most affected ones are the hands, feet, knees, and shoulders.

Fluoroscopy is almost mandatory for Urology, since the procedures performed in this specialty include the placement and removal of stents, where it is necessary to see a real-time video throughout the surgical procedure. With this technique, it is common for radiation doses to increase due to the continuous image obtained by fluoroscopy, increasing the risks associated with such exposure [19].

2.1.2. Radiation-Related Problems

Radiation is a form of energy that can be dangerous if not properly controlled. This can cause several health problems for both patients and Operating Room (OR) professionals. In the OR, there are three main sources of radiation: direct radiation, scattered radiation, and leakage radiation [20].

Direct radiation is emitted from the X-ray tube toward the patient. This type of radiation is the most intense and can cause the most damage. Scattered radiation is radiation that has been deflected off of the patient or other objects in the room. This type of radiation is less intense than direct radiation, but it can still be harmful. Leakage radiation is radiation that escapes from the X-ray tube housing, which is not originating from the beam path. This type of radiation is the least intense, but it can still be harmful if it is not properly shielded [20].

The amount of radiation exposure a person receives depends on several factors, including the strength of the X-ray source, the distance from the source, and the amount of time the person is exposed to the radiation, which depends on specialty and surgery. For instance, orthopedic surgeons and OR staff typically receive higher occupational radiation exposure than patients

due to the small area exposed to radiation from the patient. However, the risk of adverse health effects from radiation exposure in this case is low for both patients and medical personnel [20].

Ionizing radiation can damage cells and lead to various health problems. This damage occurs when ionizing radiation interacts with DNA, causing either direct or indirect DNA lesions. The resulting cellular stress response can trigger cell death or, if cell death does not occur, increase the risk of cancer due to the persistence of damaged cells.

The effects of ionizing radiation exposure can be classified as either deterministic or stochastic. Deterministic effects are immediate and worsen with increasing radiation exposure. Examples of deterministic effects include hair loss, skin redness, skin burns, and cataract formation. The threshold for these reactions is usually large and rarely exceeded in doses to which OR staff are subjected. Stochastic effects, on the other hand, are random and increase the risk of cancer or other health problems over time. The risk of stochastic effects increases with radiation exposure, but does not have a clear threshold (i.e., 10 Gy has a higher risk than 1 Gy) [8].

Unlike deterministic effects, which are short-term and worsen with increasing radiation exposure, stochastic effects manifest later in life and are unrelated to the radiation dose. However, the likelihood of these long-term effects increases with higher cumulative radiation exposure. The long-term consequences of chronic low-dose radiation exposure remain unclear, but [21] research suggests a slightly elevated risk of developing cancer, cataracts, and potentially even hereditary diseases.

2.1.3. Radiation Safety

Radiation safety during surgical procedures is important to protect patients and healthcare workers from the harmful effects of radiation. All types of intraoperative imaging involve some degree of radiation exposure. It is therefore important to take steps to minimize radiation exposure during surgical procedures.

- **Shielding**

In attempting to reduce this exposure, a variety of simple methods should be employed by all practitioners. Like the case of shielding, a method of reducing radiation exposure by using physical barriers to absorb scatter radiation.

It is primarily accomplished by wearing lead aprons and thyroid shields, but other methods can also include lead gloves, lead skirts, and mobile shielding screens.

Research has shown that shielding can reduce radiation exposure by up to 96.9%. For example, a study of three surgeons found that lead aprons and collars reduced radiation

exposure to the upper body and thyroid by 94.2% and 96.9%, respectively. The use of lead aprons was also estimated to increase the number of total operations that a surgeon could perform before reaching occupational exposure limits by 5088 procedures [8].

Radiation operators must wear personal protective equipment, including a lead apron, thyroid shield, and leaded glasses. The lead apron should have a 0.25-mm thickness equivalence, which means that it provides 0.5 mm thick protection at the front (when folded) and 0.25 mm thick protection at the back. Leaded glasses can prevent radiation from the front and side, and a thyroid shield can provide additional protection. Wearing this protective equipment can reduce radiation exposure during ureteroscopy by up to 98%. This allows surgeons to perform up to 500 procedures in a year without taking on an excessive amount of radiation [22].

- **Distance**

Increasing the distance between the radiation source and patients reduces radiation exposure. This happens because radiation intensity decreases in proportion to the square of the distance (inverse square law), as presented in equation (2.1), where "*I*" represents Radiation Intensity and "*d*" represents distance. For example, doubling the distance from the source reduces radiation exposure four times.

$$Inverse\ Square\ Law = I\alpha\frac{1}{d^2} \tag{2.1}$$

Portable C-arm fluoroscopy is superior to fixed table fluoroscopy in terms of radiation exposure. This is because the C-arm can be positioned further away from the patient. Which should be positioned as far away from the radiation source as possible, while still maintaining a clear view of the surgical area. Image capturing device should be positioned as close to the patient as possible to reduce radiation scattering, as shown in Figure 2 [22].

With appropriate shielding, scatter radiation can be reduced to 0.1% and 0.025% of the primary radiation at a distance of 0.9 meters and 1.82 meters, respectively. Further recommendations for OR setup to reduce radiation exposure through distance were reviewed by Narain et al., 2017 [8].

These authors suggested that the image intensifier be placed on the same side of the operative table as the surgeon, so as to increase the distance between the radiation source and OR personnel.

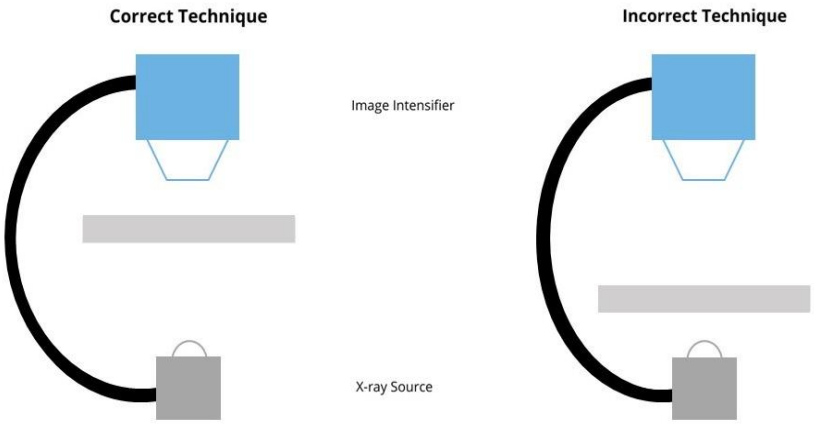

Figure 2. Positioning of C-arm to reduce radiation dosage. Adapted from [23].

It is important to note that radiation exposure is not solely a concern for surgical staff; patients also receive significant doses of radiation during intraoperative imaging procedures. Therefore, it is essential to implement strategies to minimize radiation exposure for patients as well.

- **Fluoroscopic dose reduction techniques**

There are several dose reduction techniques, which are important for reducing radiation exposure during fluoroscopy-guided procedures and should always follow the ALARA principle. For example, pulsed fluoroscopy mode applies power to the radiation source intermittently, producing short pulses of radiation, while low-dose fluoroscopy mode reduces the peak kilovolts and milliamperes needed to create the radiation beam.

A research study was conducted to investigate the effectiveness of pulsed and low-dose fluoroscopy in reducing radiation exposure during spinal interventional procedures. The study found that using pulsed and low-dose fluoroscopy together reduced radiation exposure time by an average of 56.7% [23].

In a separate study, it was investigated the effects of pulsed fluoroscopy together with shielding. They found that the combination of these two methods reduced the effective dose of radiation to all OR staff by 97.3% [24].

An emerging approach, investigated in the present study, is radiation dose prediction based on ML techniques. This approach uses machine learning algorithms to predict the radiation dose that a patient will receive during a medical imaging procedure. Helping radiation technicians to try different approaches, since the optimal doses and exposure time can be predicted previously [25].

2.2. Machine Learning Algorithms

Artificial intelligence (AI) is a field of technology that aims to create intelligent systems capable of performing tasks that typically require (mimics) human intelligence. The evolution of AI has been driven by advancements in electronics and programming, empowering computers to process information more efficiently and handle intricate tasks. The field of biomedical engineering is also rapidly transforming, where AI is playing a crucial role in advancing research, improving healthcare practices, and enhancing patient outcomes. AI's impact is evident across various aspects of biomedical engineering, including drug discovery and development, medical imaging analysis, personalized medicine, medical device development and predictive analytics [26].

A subset of AI, ML, enables systems to learn from data without explicit programming. ML has emerged as a powerful and practical tool, enabling systems to learn from data and refine their performance over time [27]. It can be broadly categorized into two main approaches: supervised and unsupervised learning.

Supervised learning involves algorithms that are trained on a set of labeled data. The labeled data consists of input data and corresponding output data. Supervised algorithms perform analytical tasks first using the training data and subsequently construct contingent functions for mapping new instances of the attribute. Once the algorithm is trained, it can be used to predict the output data for new input data [28].

Unsupervised learning involves algorithms that are trained on a set of unlabeled data. They are suitable for clustering and creating the labels in the data that are subsequently used to implement supervised learning tasks. The algorithm learns to identify patterns or structure in the data [28].

Machine learning classification models and regression models are two types of supervised learning algorithms that are used to predict outcomes based on input data. However, they differ in the type of outcome they predict, as presented in Table 1.

Classification models are used to predict a categorical outcome, such as whether an email is spam or not spam, or whether a patient has a benign or malignant tumor. Classification models learn to assign input data to a predefined set of classes [29].

Regression models are used to predict a continuous outcome, such as the price of a house, the temperature on a given day, or the blood sugar level of a patient. Regression models learn to map input data to numerical continuous values [29].

Table 1. Key differences between classification and regression models.

Characteristic	Classification	Regression
Outcome type	Categorical	Continuous
Prediction goal	Assign input data to a predefined set of classes	Map input data to a numerical value
Evaluation metric	Accuracy, precision, recall, F1-score	Mean squared error, mean absolute error, R-squared

The various ML regression algorithms also differ as they are either parametric or non-parametric. Parametric regression algorithms make strong assumptions about the data, which simplifies the learning process and makes them computationally faster. However, these assumptions can limit the model's ability to learn. Usual examples of parametric algorithms implemented are Linear Regression (LR), Logistic Regression, Ridge regression and Lasso regression. Non-parametric regression algorithms do not make any assumptions about the relationship between the input and output variables. Instead, they learn the relationship directly from the data. This makes them more flexible than parametric algorithms, but they can also be more computationally expensive to train. Some common examples of implemented non-parametric algorithms are Random Forest Regression (RFR), Support Vector Regression (SVR), Artificial Neural Networks (ANN) and Decision Tree (DT) [30].

LR predictive models are statistical models that use linear regression to predict a continuous outcome variable based on one or more independent variables. They are among the most widely used predictive models in machine learning due to their simplicity and interpretability. LR models assume that the relationship between the independent variables and the outcome variable is linear. This means that the predicted value of the outcome variable can be calculated as a linear combination of the independent variables.

The coefficients of the linear combination represent the strength of the relationship between each independent variable and the outcome variable [31].

Least Absolute Shrinkage and Selection Operator (LASSO or L1) regression is a versatile regression technique, derived from the regular LR, that utilizes L1 regularization to penalize the absolute values of the regression coefficients, as demonstrated in Figure 3. This regularization can effectively force certain coefficients to shrink to zero, essentially eliminating them from the model.

This property makes LASSO regression an important tool for feature selection, enabling the identification of the most influential features in a dataset. It effectively identifies the most influential features, produces sparse and interpretable models, and demonstrates robustness to outliers [32].

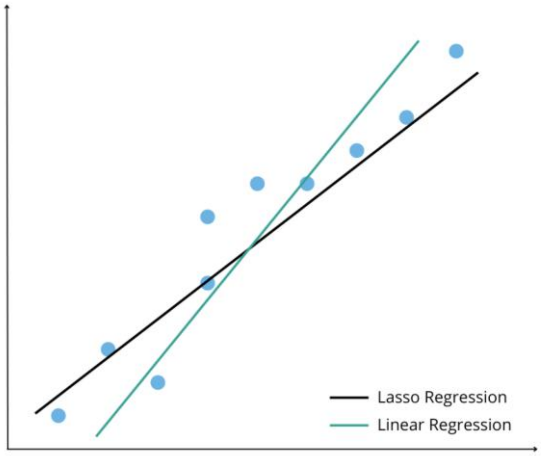

Figure 3. Comparison between Linear Regression and LASSO Regression.

RFR is a supervised learning algorithm that consists of multiple decision trees that are independent of each other and can be used for both classification and regression tasks. To solve a problem using an RFR model, the problem starts at the root of each of the trees and it divides into multiple nodes until a result is obtained on each of the trees [33]. Each decision node is split into two or more child nodes based on the best feature out of a randomly selected subset of features at that node. The data in each child node is then used to predict the values of the dependent variable in that node. Finally, the predictions from all child nodes are aggregated to produce the final prediction, by calculating the average of the values from all the trees [34], as schematized in Figure 4.

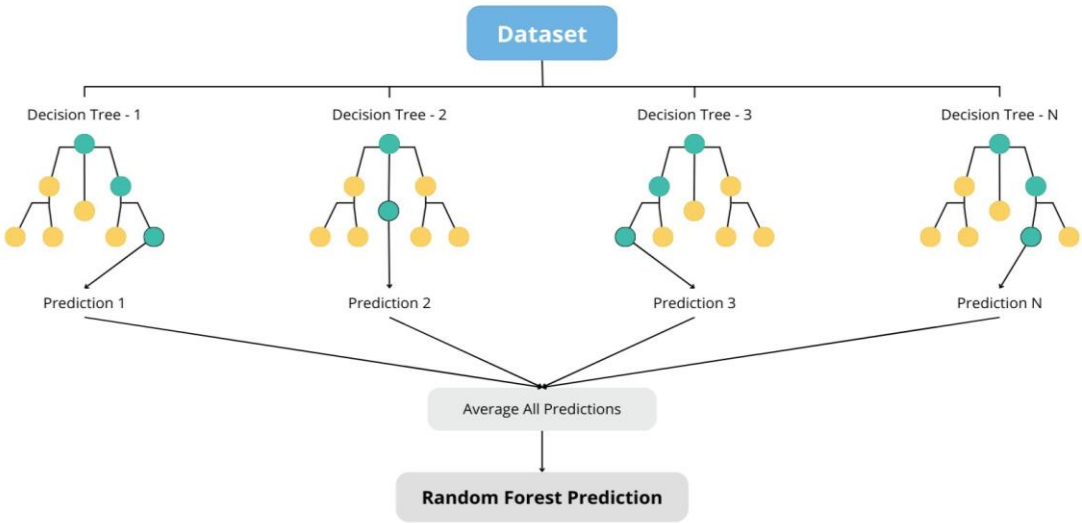

Figure 4. Example of random forest regression tree. Adapted from [35].

SVR is a powerful regressor method for both linear and nonlinear regression tasks that has gained a lot of attention lately. Built with the same principles of Support Vector Machine, where support vectors are data points that are closest to the decision boundary, known as the hyperplane, in a high-dimensional feature space, as shown in Figure 5. SVR is used for regression analysis, mainly in predicting continuous numerical variables [36]. This approach plays a crucial role in understanding how variables work and making predictions based on observed data [36], it leads to robust and generalizable models that can capture complex nonlinear relationships between the input and output variables [37].

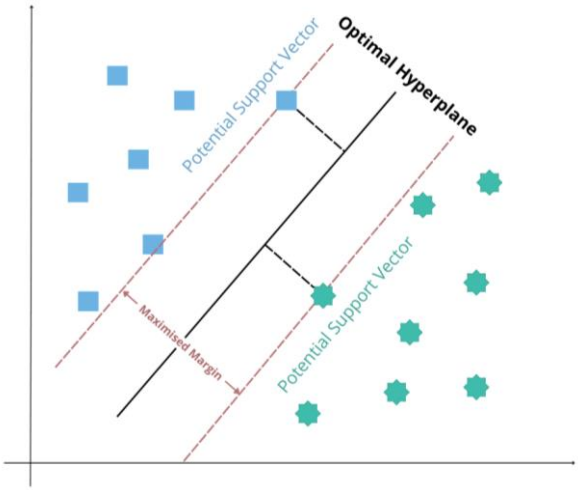

Figure 5. One-dimensional linear SVR. Adapted from [36].

ANNs are a type of AI that simulate the structure and function of the human brain. They are composed of interconnected nodes, called neurons, that process and transmit information. The architecture of ANNs is inspired by the biological neural network, with neurons organized into layers: input, hidden, and output [38].

ANNs are mostly used for processing large datasets [36]. They are built of multiple connected layers and each layer may in turn contain several neurons, forming a complex network. ANNs typically includes an input layer, one or multiple hidden layers, and an output layer [39], as shown in Figure 6. When operating, data is inserted into the input layer, goes through all the hidden layers, and eventually comes out from the output layer, providing the desired predictions [39].

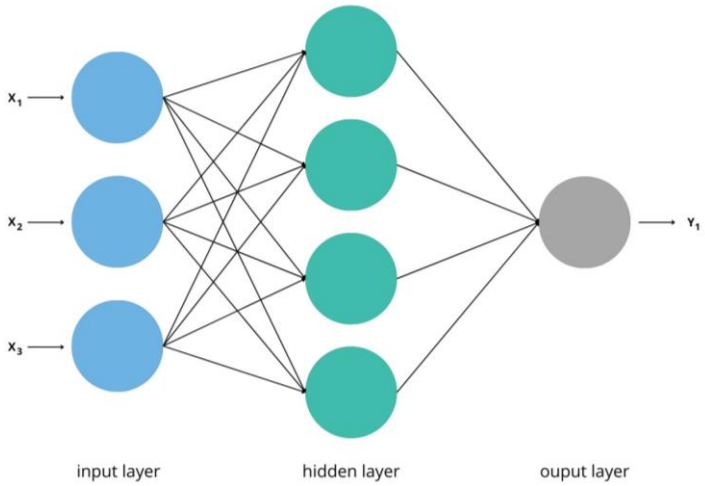

Figure 6. Feedforward neural network. Adapted from [40].

DTs are a type of machine learning algorithm that can be used to classify data by recursively partitioning the data into smaller and smaller subsets. DTs are particularly well-suited for classifying discrete-valued target functions and can be expressed as a set of if-then rules. Each DT consists of nodes, which can be classified as root nodes, decision nodes, or leaf nodes, as shown in Figure 7.

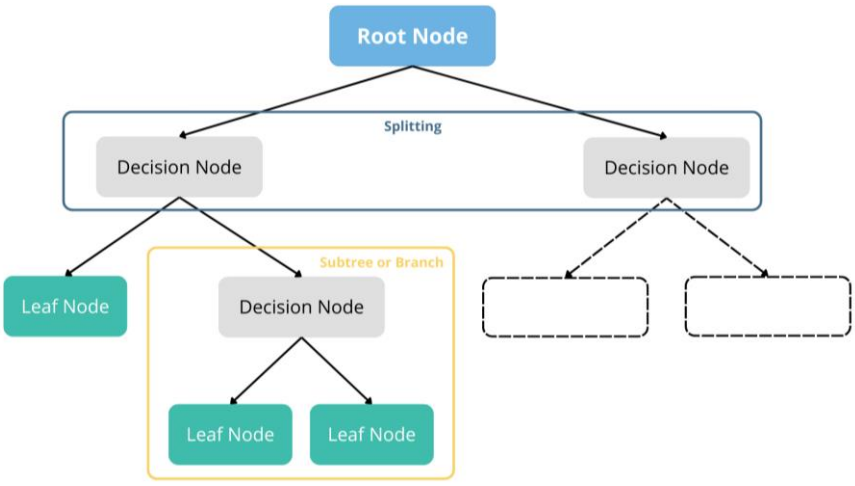

Figure 7. DT diagram and its terminology. Adapted from [41].

The root node is the starting point of the tree, and the leaf nodes represent the final classifications of the data. Decision nodes, on the other hand, contain decision rules that split the data into smaller subsets based on specific attribute values [41]. The classification of an instance using a decision tree involves traversing the tree from the root node to a leaf node. At each decision node, the value of a specific attribute is checked, and the instance is moved down the subtree corresponding to that value. Once the instance reaches a leaf node, the classification associated with that leaf node is assigned to the instance [41].

2.3. Evaluation Metrics and Methods

Evaluation metrics, also known as error measures, play an important role in evaluating the effectiveness of different models and algorithms across diverse fields. These metrics provide a quantitative assessment of how closely the actual results align with the expected or predicted outcomes. Among the most commonly used performance metrics are Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R-Squared (R^2), as presented in Table 2.

Table 2. Evaluation Metrics for Regression.

Metric Description	Equations	Variables	References
MAE Average magnitude of errors between predicted and actual values	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - y\rvert$	n - Total number of data points. y_i - Real measured data point. $\hat{y}_i$ - Estimated or predicted value for that data point.	[42]
MSE Average squared difference between predicted and actual values	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	n - Total number of data points. y_i - Real measured data point. $\hat{y}_i$ - Estimated or predicted value for that data point.	[38]
RMSE Square root of average squared difference between predicted and actual values	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	n - Total number of data points. y_i - Real measured data point. $\hat{y}_i$ - Estimated or predicted value for that data point.	[43]
R^2 Measure of how well the regression model fits the data	$R^2 = 1 - \frac{SSR}{SST}$	SSR - Variance in the dependent variable that is not explained by the regression model. SSL - Total variance in the dependent variable.	[44]

It is possible to closely tie evaluation metrics to the scientific concepts of distance and similarity. In ML regression models, evaluation metrics are employed to compare the predictions generated by the trained model with the actual data from the testing dataset. The results of these comparisons directly influence the decision-making process when selecting the most suitable ML algorithms for implementation [42].

Mean Absolute Error (MAE) serves as a measure of the average absolute difference between the actual and predicted values. As the model's performance is assessed using MAE, the objective is to minimize its value, which signifies a smaller average deviation between the predictions and the actual values. A lower MAE value corresponds to a better-performing

model. When comparing both evaluation metrics, the MSE value is usually lower than the MAE value, but both are equally good to calculate on a regression model [39]. MAE was calculated using the equation mentioned in Table 2.

However, one disadvantage of MAE is that it is not differentiable. A differentiable function is one in which the slope can be defined at every point along its curve. Non-differentiable functions, on the other hand, have points where the slope is undefined or infinite. This can make it difficult to optimize MAE using gradient-based algorithms, which rely on calculating the slope of the function at different points. To overcome this disadvantage, MAE is usually conjugated with other metrics like MSE [43].

MSE is an evaluation metric that measures the total of errors in a statistic model. It works by calculating the average squared error between the real and the predicted values of a regression model. The way that MSE calculates the accuracy of the model, is by having the lowest value possible because it means that the model has very few or no errors, i.e., the closer it is to 0 the better. As the error increases, the more errors the model will likely have [44]. MSE was calculated using the equation mentioned in Table 2.

Since the function produced by MSE is differentiable and can be easily used as a loss function, it conjugates nicely with metrics whose graph is not differentiable. However, this metric greatly penalizes outliers, which makes a database with many outliers backfire and the calculated MSE be bigger. This shows that this metric is not as robust as MAE [43].

The RMSE is the square root of the MSE and measures the average magnitude of the errors between the predicted and actual values. It is calculated by taking the square of the difference between each predicted value and its corresponding actual value, then averaging all of these squared differences, and finally taking the square root of the average. While taking the square root does not alter the relative ranking of different models, it results in a metric that shares the same units as the target variable '*y*', as presented in the equation of Table 2. This makes the RMSE convenient for representing the typical or "standard" error when the errors are normally distributed. It is more sensitive to large errors than the MAE, as the squaring operation amplifies the effect of large errors [45].

R-squared, is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model. It is calculated by squaring the correlation coefficient between the predicted and actual values of the dependent variable, as presented in the equation of Table 2. R-squared values range from 0 to 1, where 0 indicates that the independent variable does not explain any of the variance in the

dependent variable, and 1 indicates that the independent variable explains all of the variance in the dependent variable.

It is a commonly used metric for evaluating the goodness of fit of a regression model. A higher R-squared value generally indicates a better fit, however a disadvantage of the metric is that R-squared can be artificially inflated by adding more independent variables to the model. Therefore, it should be interpreted in conjunction with other measures of model fit, such as adjusted R-squared [46].

Cross-Validation (CV) is a resampling method used to estimate the true prediction error and the performance of ML models, which splits the data in two segments, one responsible (used) for training and the second for testing [47]. There are several CV methods, however the most used in average size databases, is the K-Fold Cross-Validation.

K-Fold CV is normally used to evaluate machine learning models on a limited data. It works by splitting the data into different k folds, training the model on k-1 folds, and evaluating its performance on the remaining fold. This number of folds can be as many as the user wants, but the most common number used is 10, as represented in Figure 8. After selecting the number of folds, the algorithm develops a regression or a prediction, like RFR, SVR or ANN, splits the data into the number of desired folds and then it uses all the folds to both training and testing [47].

By applying K-fold, it is possible to avoid overfitting, which is when a model performs well on the training data but poorly on unseen data. By training and evaluating the model on different subsets of the data, k-fold cross-validation can help to ensure that the model is generalizable to new data [48].

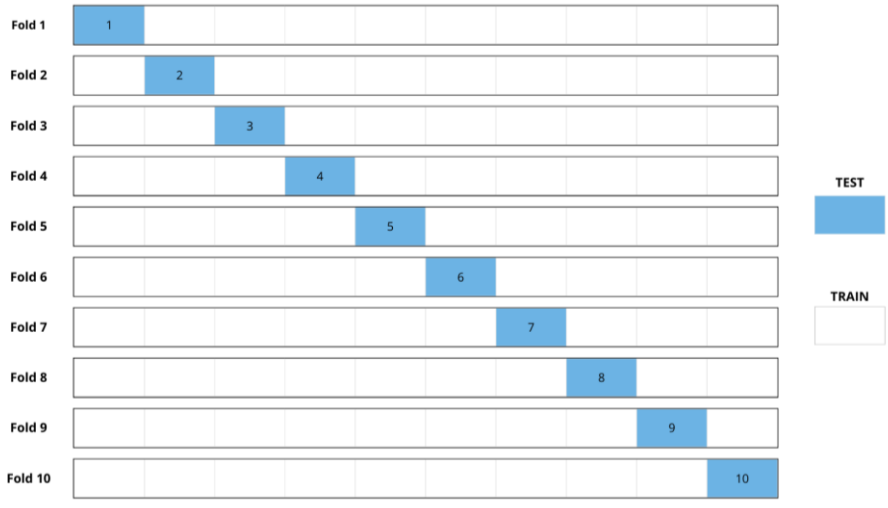

Figure 8. Example of CV with 10 folds.

2.4. Related Work

Radiation dose prediction based on ML techniques is an emerging approach in medical imaging. Using ML algorithms can analyze vast amounts of patient data and imaging parameters to predict the optimal radiation dose for each patient. This information can then be used to personalize radiation treatment plans, reducing the amount of radiation exposure to patients while still ensuring that they receive diagnostically useful images.

In a study by Ahn et al. [16], a deep learning method for dose prediction was developed and successfully applied to accurately predict patient-specific doses for left-sided breast cancer. The patient-specific dose prediction was achieved using a U-net-based modified dose prediction neural network, leveraging a contour image of the planning target volume and organs at risk. A database of 50 volumetric modulated arc therapy plans for left-sided breast cancer patients was employed to quantitatively assess the dose prediction performance using various parameters. The efficiency and accuracy of the deep learning-based dose prediction framework were compared against those of a conventional knowledge-based planning system (RapidPlan™). The results demonstrated superior dose predictions using deep learning, as evaluated through the MAE and standard deviation (STD).

In their study, Huettenbrink et al. [49] developed a Machine Learning (ML) algorithm to accurately predict a patient's personal radiation exposure during ureterorenoscopy. Their research included 995 patients with ureterorenoscopy, and the radiation features (DAP) were categorized into two groups: 'low doses' $\leq$ 2.8 Gy·cm2 and 'high doses' > 2.8 Gy·cm2. To predict the level of radiation exposure during treatment, six different machine learning models were trained and evaluated using 10-fold cross-validation. The results showed that the neuronal network model achieved the highest accuracy, predicting whether a patient would receive a low dose of radiation with 94% accuracy. The study also identified the factors that most influenced radiation exposure, including age, gender, weight, stone size, surgeon experience, number of stones, stone density, use of a flexible endoscope, and preoperative stone position.

The application of diverse ML models and techniques holds immense potential to enhance the reliability and practicality of predictive models in real-time surgical procedures [16, 49].

The use of ML for radiation dose prediction has several potential benefits, such as:

- Reduced Patient Radiation Exposure - By optimizing radiation dose delivery, ML can help to minimize the amount of radiation exposure that patients receive during imaging procedures. Which is particularly important for children and pregnant women, who are at higher risk of radiation-induced health risks.

- Increased efficiency - ML can automate many of the tasks involved in radiation dose optimization, which can save radiation technicians time and effort.

Various methods have been explored to achieve these benefits, such as the work by Garcia-Sanchez et al., which employed a variety of well-known machine learning techniques to reduce unnecessary radiation without compromising diagnosis through Computed Tomography images. Their study utilized a large-scale dataset of over 50,000 patients. They classified all patients into standardized protocols (skull, abdomen, thorax, pelvis, etc.), extracted the parameters of interest, eliminated atypical information (outliers), to later generate regression curves through ML methods. Finally, a decision-making process to select the better ML techniques for each protocol, which included Gaussian processes, neural networks, and regression trees [25].

Chapter 3

Chapter 3 covers the materials and methods used to develop the present study. The paper discusses topics such as the dataset used, the proposed framework for achieving the research objectives, the preprocessing steps taken, and the models employed.

3. Materials and Methods

A radiation dataset was collected and curated, consisting of examples of radiation data along with the corresponding ground truth output, i.e., the target variables. Feature selection was performed to identify the most important features and remove the most irrelevant in the radiation dataset. Machine learning algorithms were then applied to the radiation dataset to learn from the data, along with HT, which was performed to find the optimal values for the hyperparameters of the machine learning algorithms. After that, the trained machine learning models were used to predict the output for new exposure time and radiation exposure data. Finally, the predicted output was analyzed to obtain insights into the radiation data. The proposed system architecture is represented in Figure 9.

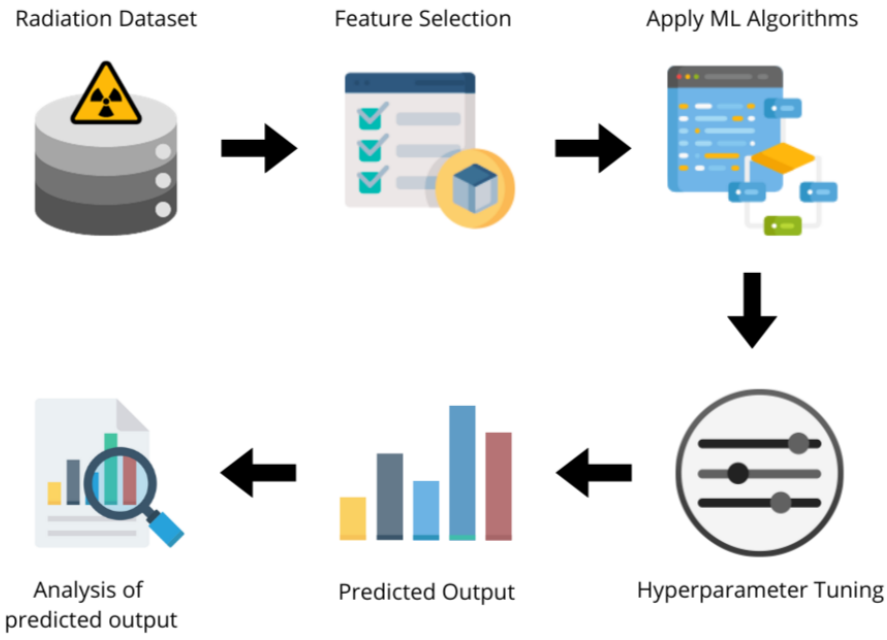

Figure 9. Proposed workflow.

3.1. Dataset

Data regarding radiation exposure was prospectively collected from 1125 consecutive patients' cases submitted to surgical procedures in the "Hospital da Luz Setúbal", Portugal between 2021 and June 2023. However, 101 patients' cases (9.9% of the total patients) were not considered due to missing data problems. The final dataset included data from 1024 patients.

Apart from the AK and KAP* variables, it was also acquired and documented a comprehensive set of features, encompassing exposure parameters (Exposure Time, Number of Shots, Type of Exposure), imaging Equipment used, Surgical Procedure, anatomical Segments involved, Laterality, affected vertebral levels, and patient demographic data (age, gender, height, weight, and body mass index).

To ensure the trustworthiness and dependability of the dataset, it underwent a thorough data quality evaluation process. This process encompassed Data Cleaning, where missing values, outliers, and inconsistencies were identified and eliminated; Error Detection, which involved checking for data entry errors and logical inconsistencies; and Data Validation, which verified the accuracy and completeness of the data against the original source.

The dataset comprises the following features:

- Continuous features such as:
  - Number of Shots: Records the number of radiation shots used during the imaging process.
  - Exposure Time: Measures the duration of radiation exposure.
  - AK: Quantifies the radiation dose delivered to the patient.
  - KAP:* Represents the product of the irradiated area, after adjustments by an equipment-specific correction factor.
  - Height: Records the patient's height.
  - Weight: Records the patient's weight.
  - BMI: Represents the patient's body mass index.
  - Age: Records the patient's age.
- Categorical features such as:
  - Surgeon: Represents the surgeon's encoded identifier.
  - Technician: Represents the technician's encoded identifier.
  - NHC: Records the patient's hospital number.
  - Specialty: Indicates the medical specialty associated with the surgical procedure.

- Surgery: Specifies the type of medical procedure performed.
- Levels: Indicates the depth at which the surgical procedure was performed.
- Segment: Identifies the anatomical segment targeted for surgery.
- Equipment: Names the equipment utilized for intraoperative imaging.
- Type of Exposure: Indicates the type of radiation exposure employed for imaging.
- 3D: Indicates whether a three-dimensional image of the segment was acquired.
- Sex: Indicates the patient's gender.

3.2. Strategies Employed

The present study was developed and refined using Python (Anaconda Inc.), a widely used programming language, and a dataset on radiation. The algorithms employed in the regression models were developed through the help of *Python's Scikit-Learn* library, a machine learning toolkit.

The algorithms were also fine-tuned through hyperparameter Tuning (HT) which involves exploring and selecting the optimal values for the parameters within each model. HT methods include manual tuning, which involves manually experimenting with different values for each hyperparameter and evaluating the model's performance; Grid Search, which systematically evaluates every possible combination of values for a selected set of hyperparameters; and Random Search, which randomly selects and evaluates different combinations of values for a set of hyperparameters [50]. While this process can be time-consuming, it can be crucial in transforming a poorly performing model into a well-performing one.

To prepare the data for analysis, a preprocessing step was performed, followed by a feature selection process. This feature selection process involved a statistical analysis of the data provided by "Hospital da Luz Setúbal", Portugal, to identify the most relevant features and understand their correlations. The analysis was conducted by examining the behavior of each feature and its relationship with the independent variables (AK and KAP*).

To further investigate the correlation of certain features with KAP*, which is considered a more reliable measure of radiation exposure, a study was conducted using all surgeries in the dataset. However, entries with a "Number of Shots" less than 5 were excluded due to their potential to influence the analysis. Additionally, surgeries with less than 25 patients were also removed due to insufficient sample size.

3.3. Preprocessing

To prepare the dataset for training, several preprocessing operations were performed to ensure that all the data was ready for training Artificial Intelligence regression models. During data preprocessing, one column, Exposure Time, was transformed from its original 24-hour clock format to integral numbers to enhance the applicability of statistical analysis.

In the models trained to predict AK, the KAP* feature was removed from the dataset and vice versa. Since KAP* is directly proportional to AK, as they are dependent on each other, it was necessary to carry out the predictions without the opposite, as this would influence the prediction results. In order to obtain the most realistic Exposure Time prediction possible, features that could not be obtained before the surgical procedure were also removed, Type of Exposure and Number of Shots, as those features data can't be obtained pre-operatively.

In the case of Regression Models the numerical columns are scaled to a range between 0 and 1 using the *MinMaxScaler* library, and non-numerical columns are transformed into numerical labels with the *LabelEncoder* library. These preprocessing steps are essential for preparing data for ML by ensuring it's in a suitable format for model training and evaluation. After preprocessing, the dataset consisted of 1024 data rows.

3.4. Regression Models

This research employed supervised regression algorithms implemented in Python's Scikit-Learn library, a widely used machine learning toolkit, to predict exposure time. Subsequently, radiation values were also predicted from recorded data, aiming to enhance safety measures and mitigate radiation-related issues.

Five ML regression models were trained, fine-tuning and tested to identify the optimal predictive model with high performance on the collected dataset, as it is explained in Figure 10. The models evaluated were: (i) Random Forest; (ii) Support Vector Regression; (iii) Artificial Neural Networks; (iv) Decision Tree; and (v) Lasso Regression. For each model, two evaluation metrics were calculated: MSE and MAE, as these metrics were deemed to complement each other effectively given the nature of the data obtained from the dataset. Additionally, a ten-fold cross-validation procedure was employed to ensure that all patient data were utilized for both training and testing, and to assess the accuracy of the generated models.

Initially, the dataset is loaded and subjected to preprocessing, as outlined in the preceding section. Subsequently, the data is analyzed, and the correlation between each feature is examined to facilitate the selection of relevant and irrelevant data points.

Following this, a crucial step, Data Splitting, divides the dataset into two distinct subsets: a Training Set comprising a bigger part of the original dataset and a Test Set consisting of a smaller part of the original dataset. During model fine-tuning, the dataset split was adjusted from the conventional 70/30 ratio to a more restrictive 90/10 division. This alteration was implemented to focus the model's training on a more refined subset of data, enabling enhanced performance during subsequent testing. The 70/30 split was chosen as it is a common choice for datasets with sufficient size, however it was perceptible that the present dataset didn't perform at its best. The 90/10 split is normally employed when data is limited or when a more rigorous evaluation is desired.

The Training Set, employing the CV method, identifies the optimal trained model. This best trained model is then applied on the (unused for training) Testing Set to predict the values of the chosen label, effectively simulating a new case using unseen data. Finally, the chosen evaluation metrics enable the assessment of the employed ML models.

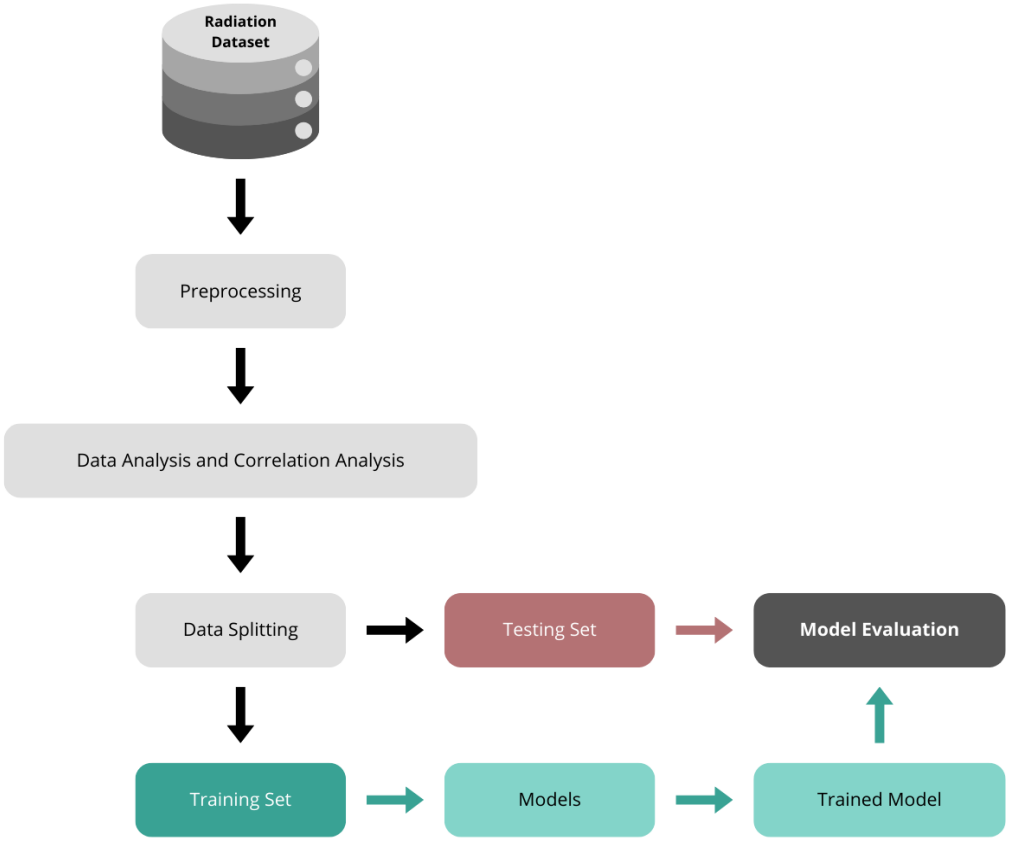

Figure 10. Algorithm workflow.

In order to improve as much as possible, the performance of the models HT was also employed within each model. To achieve a more dynamic implementation, two techniques were employed (i) Randomized Search (*RandomizedSearchCV*) and (ii) Grid Search (*GridSearchCV*) [30] were used to optimize the model's hyperparameters, as they are implemented in the *Scikit-Learn* library.

In the RF model, imported from sklearn.ensemble, it was used six parameters for HT, as presented in Table 3.

Table 3. Parameters used for HT in RF model.

Parameter	Description	Values Used
n_estimators	Number of trees in the forest	100, 200, 300, 400, 500
max_features	Number of features to consider for the best split	'auto', 'sqrt', 'log2'
max_depth	Maximum depth of the trees	None, 10, 20, 30, 40, 50
min_samples_split	Minimum number of samples required to split an internal node	2, 5, 10
min_samples_leaf	Minimum number of samples required for a leaf node	1, 2, 4
bootstrap	Whether to use bootstrap samples	True, False

In the SVR model, imported from *sklearn.svm*, four parameters for hyperparameter tuning were used, shown in Table 4.

Table 4. Parameters used for HT in SVR model.

Parameter	Description	Values Used
kernel	Type of kernel function	'linear', 'poly', 'rbf', 'sigmoid'
C	Regularization parameter (cost) that controls the trade-off between model complexity and loss reduction	0.1, 1, 10
epsilon	Tolerance for miscalculations in the epsilon-SVR model	0.01, 0.1, 0.2
degree	Degree of the polynomial kernel (applicable only when using the 'poly' kernel)	2, 3, 4

In this work, the ANN model was developed using the *keras.models* module. The model consists of a single dense layer with 32 neurons, utilizing the Rectified Linear Unit (ReLU) activation function, and trained for 10 epochs. The model was subjected to hyperparameter tuning using two parameters, seen in Table 5.

Table 5. Parameters used for HT in ANN model.

Parameter	Description	Values Used
hidden_units	Number of neurons in the hidden layer	100, 200, 300
dropout_rate	Percentage of neurons randomly dropped out during training	0.1, 0.2, 0.3

The DT method employed a decision tree model imported from the *sklearn.tree* library. The algorithm was subjected to hyperparameter tuning using five parameters, described in Table 6.

Table 6. Parameters used for HT in DT model.

Parameter	Description	Values Used
max_depth	Maximum depth of the tree, which controls the complexity of the model	None, 10, 20, 30, 40, 50
min_samples_split	Minimum number of samples required to split an internal node	2, 5, 10
min_samples_leaf	Minimum number of samples required for a leaf node	1, 2, 4
max_features	Number of features to consider when selecting the best split at each internal node	'auto', 'sqrt', 'log2'
max_leaf_nodes	Maximum number of leaf nodes in the tree	None, 10, 20, 30, 40, 50

The LASSO method employed a Linear Regression model imported from the *sklearn.linear_model* library. The model was subjected to hyperparameter tuning using only the 'alpha' parameter which represents the regularization strength (alpha), with values of [0.001, 0.01, 0.1, 1, 10].

As previously mentioned, a CV method was employed in this study. Specifically, a 10-fold CV system was implemented, ensuring that all data points were utilized for training, resulting in 10 distinct model evaluations. The model subsequently identified the fold with the optimal performance, designating that fold as the final model for predicting the target variable alongside the test set. Additionally, the mean of all fold-specific performance values was calculated to provide an overall assessment of the model's performance.

To provide a clearer understanding of the data and following the calculation of MSE and MAE values, the evaluation metric values for each regressor were subtracted from 1.0 (1-MSE or 1-MAE). This transformation ensured that the regressor with the closest value to 1.0, rather than the one with the smallest value, was considered the best performer when visualizing the summary of all data. This adjustment provided a clearer representation of the relative performance of each regressor.

Chapter 4

Chapter 4 presents a comprehensive statistical analysis of the dataset to elucidate the characteristics and interrelationships of the included features. Additionally, the performance of various prediction models is evaluated using their respective metrics. Furthermore, the regression models employed for predicting intraoperative imaging exposure time are subsequently utilized to predict AK and KAP.*

4. Results and Discussion

It was trained ML algorithms using a dataset of patient cases undergoing surgical procedures that includes patient demographics, imaging parameters, and radiation dose. The trained algorithms/models were then used to predict the radiation dose of new patient cases. With this, it is possible to identify patients who are at risk of receiving high radiation doses, these patients can then be targeted with dose reduction strategies, such as pulsed fluoroscopy or using a lower X-ray energy, as mentioned in the previous techniques. Optimizing imaging parameters to reduce radiation exposure, is also possible, through the developed machine learning models predicting the minimum exposure time required to produce an image of acceptable quality.

The findings of the statistical analysis and predictive regression models for assessing radiation exposure in medical imaging procedures are presented and discussed in this chapter. Statistical analysis involved a comprehensive exploration of the relationships between radiation exposure and a range of variables, including patient demographic factors, imaging equipment used and procedural characteristics. This analysis employed descriptive statistics, correlation analysis, and regression modeling techniques to elucidate the patterns and associations within the data.

To develop predictive models for radiation exposure, we utilized a variety of ML algorithms, as mentioned before. These algorithms were trained on a subset of the data and subsequently evaluated for their predictive accuracy on a separate holdout sample. The performance of the models was assessed using evaluation metrics.

The results of the statistical analysis and predictive regression models will be presented in detail in this chapter. It will be discussed the key findings of the analysis, the performance of the predictive models, and the implications of these findings for clinical practice and radiation safety.

4.1. Statistical Analysis

The dataset regarding radiation exposure produced / developed through a narrow scientific research collaboration with the "Hospital da Luz Setúbal", Portugal was preprocessed and curated to convert it in a "benchmarking radiation dataset". A total of 1024 patients meets the inclusion criteria and included in the present study. Patients' cases data is shown in Table 7, through the features Age, Sex, BMI, Specialties, Equipment, Type of Exposure, KAP*, Exposure Time and Air Kerma. An analysis of the patient sample revealed that women constituted the majority, accounting for 50.9% of the participants. The mean age of the entire patient population was 56.5 years, with a standard deviation of 16.26 years.

Table 7. Characteristics of patients, surgeries, and radiation features obtained from the dataset.

Variables	Total Dataset (*n* = 1024)
Age	56.5 ± 16.26
Sex	
Male	503 (49.1%)
Female	521 (50.9%)
BMI	26.8 (24.0 – 29.7)
< 18.5 kg/m^2	11 (1.1%)
[18.5, 24.9] kg/m^2	39 (31.9%)
[25, 29.9] kg/m^2	447 (43.7%)
[30, 34.9] kg/m^2	187 (18.3%)
[35, 40] kg/m^2	43 (4.2%)
> 40 kg/m^2	9 (0.9%)
Specialties	
Spine Surgery	411 (40.1%)
Orthopedics	444 (43.4%)
Urology	134 (13.1%)
General Surgery / Gastro-Enterology	35 (3.4%)
Equipment	
Cios Connect Bloco	688 (67.2%)
Siemens Cio Spin	336 (32.8%)
Type of Exposure	
Single Shot	25 (2.4%)
Fluoroscopy	999 (97.6%)
KAP* (uGy.m^2)	31.81 (7.44 – 174.74)
Exposure Time (s)	23.5 (7.0 – 56.0)
Air Kerma (mGy)	1.90 (0.40 – 10.93)

In terms of BMI, 1.1% of patients were classified as underweight, 31.9% were normal weight, 43.7% were overweight, 18.3% were obese, and 5.2% fell into the class 2 and 3 obesity categories. The median BMI was 26.8 kg/m^2, with an interquartile range (IQR) of 24.0 to 29.7 kg/m^2. The distribution of patients across specialties revealed that Orthopedics and Spine Surgery represented the majority of the study population, accounting for 43.3% and 40.1% of the patients, respectively. Urology constituted a smaller proportion, with 13.1% of the patients, while General Surgery / Gastro-Enterology had the smallest representation, with 3.4% of the patients. The primary fluoroscopic equipment used in the study was the Cios Connect Bloco, employed in 67.2% of the surgeries. The Siemens Cio Spin was utilized in the remaining 32.8% of surgeries. Notably, fluoroscopy was the predominant exposure modality, employed in 97.6% of the surgeries. Single-shot fluoroscopy was used less frequently, accounting for only 2.4% of the surgeries.

The median exposure time for all procedures was 23.5 seconds, with an IQR of 7.0 to 56.0 seconds. The median AK was 1.9 mGy (IQR: 0.4 to 10.9 mGy), and the median KAP* was 31.8 mGy.m^2 (IQR: 7.4 to 174.7 mGy.m^2). Consistent with the approach in Miller et al., 2020 [51], continuous variables were summarized as mean ± Standard Deviation (SD) or median and IQR, while categorical variables were presented as relative number and percentage, in the previous table.

Initially a correlation matrix was employed to identify features in the database that exhibited stronger correlations with Air Kerma and KAP*. As shown in Table 8, the correlation coefficients of numeric features, calculated through a Python function, allowed for the selection of parameters that had the most significant influence on these target variables. Given the direct proportionality between AK and KAP, where one variable is dependent on the other, as researched in Know et al. Analyzing a single variable is sufficient as it inherently reflects the behavior of the other [52].

Table 8. Correlation Coefficients with Air Kerma and KAP*.

Features	Air Kerma	KAP*
Exposure Time	0.7233	0.7774
Number of Shots	0.5467	0.5727
Age	0.1894	0.1884
BMI	0.0866	0.0990
Weight	0.0520	0.0710
Height	-0.0301	-0.0149

Certain surgical procedures in the dataset only required minimal radiation support for localization purposes, such as level verification or patient positioning. These procedures represent a subset where radiation has no substantial impact on the execution of the medical intervention and should be analyzed separately.

Exposure Time was the surgery-related feature with the strongest correlation to KAP*, exhibiting a correlation coefficient of 0.7774 (Table 8). This strong / moderate positive correlation was also evident in Figure 11, where some correlation values approached 0.8 across various surgical procedures.

All surgeries demonstrated a significant correlation, with the highest value of 0.79 observed in stent replacement for urology, followed by a correlation of 0.76 in posterior arthrodesis and the weakest correlation being of 0.31 in anterior arthrodesis, both for spine surgery.

The confidence intervals applied to each line indicate a 95% probability that the true best-fit line for the population falls within the confidence interval.

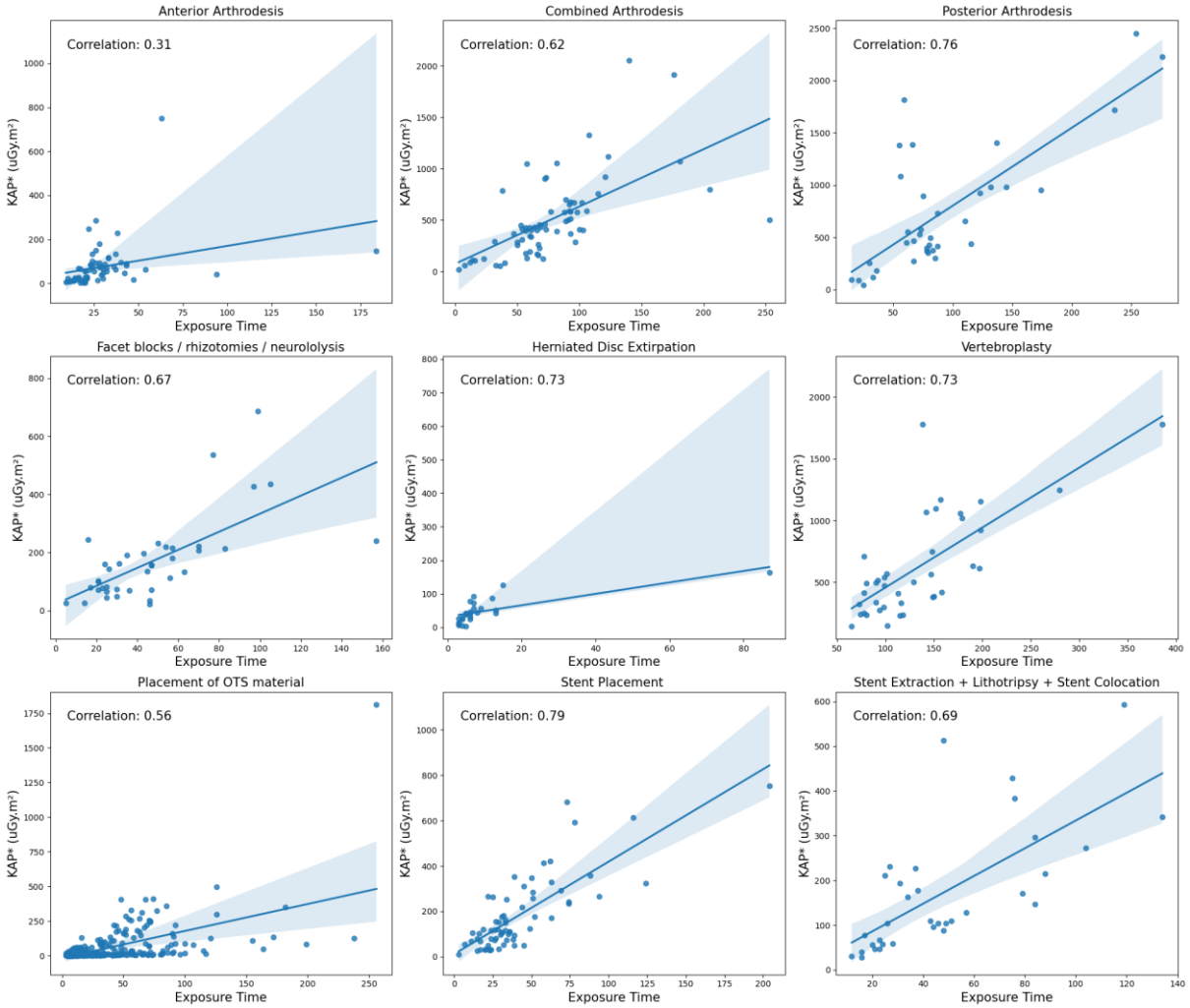

Figure 11. Exposure Time vs KAP* for each Surgery.

The second variable with the highest correlation was the Number of Shots, exhibiting a coefficient of 0.5727, as presented in Table 8. As evident in Figure 12, this variable also significantly impacts the amount of radiation exposure. These surgery-related features were also expectedly correlated, as an increased number of shots during the procedure, usually exposes the patient to more radiation.

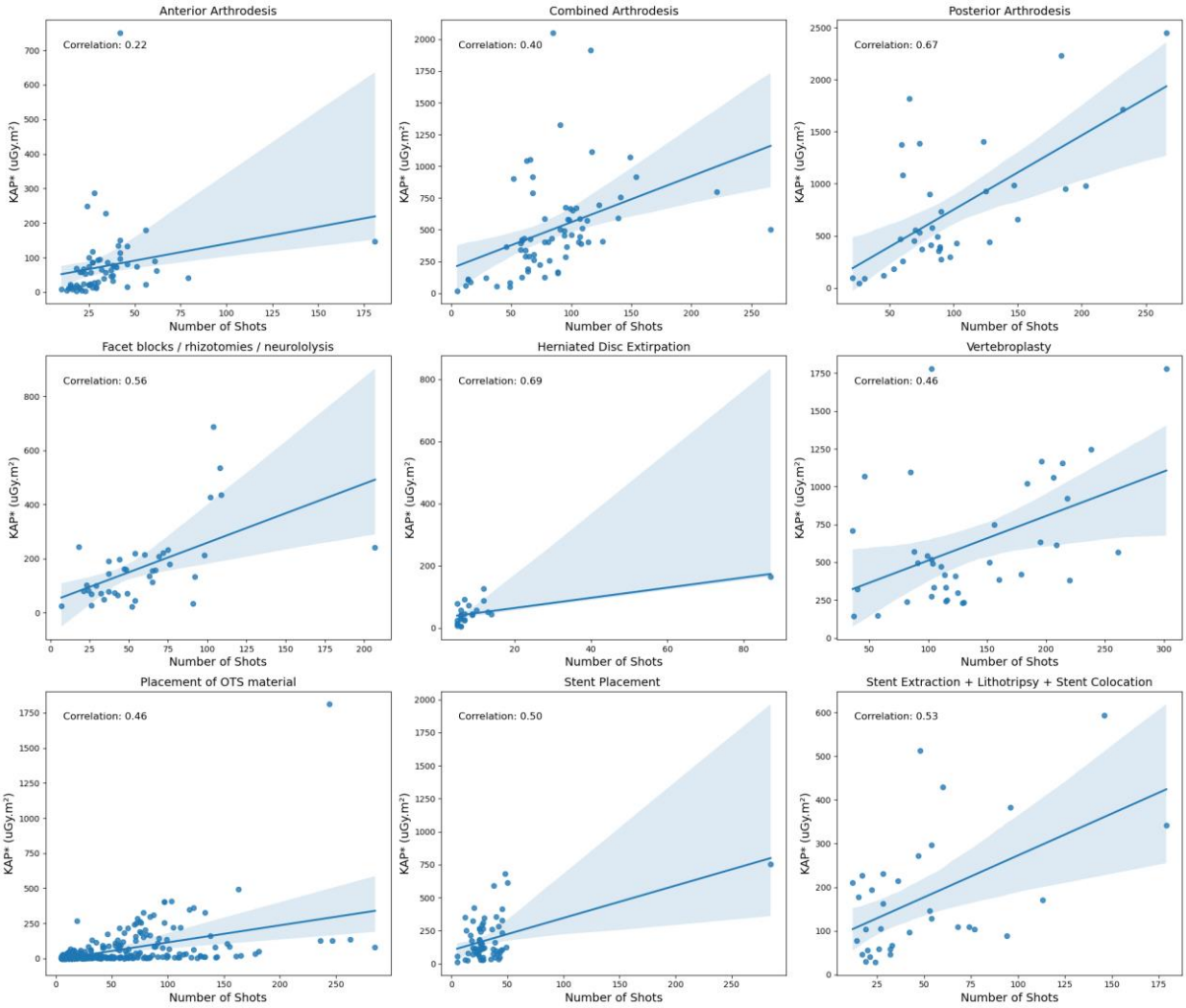

Figure 12. Number of Shots vs KAP* for each Surgery.

Despite exhibiting a lower correlation compared to the surgery-related features discussed earlier, BMI still influences the amount of radiation utilized during surgical procedures, as explained later in the study. Body mass index impacts the amount of radiation required to obtain a clear image of the patient. However, its relevance is limited to specific specialties and procedures, particularly spine surgery. As depicted in Figure 13, anterior and posterior arthrodesis of the lumbar spine demonstrate a moderate association between BMI and KAP*, which corroborates with the study carried out by Kukreja et al in 2015 on minimally invasive spine surgeries [53].

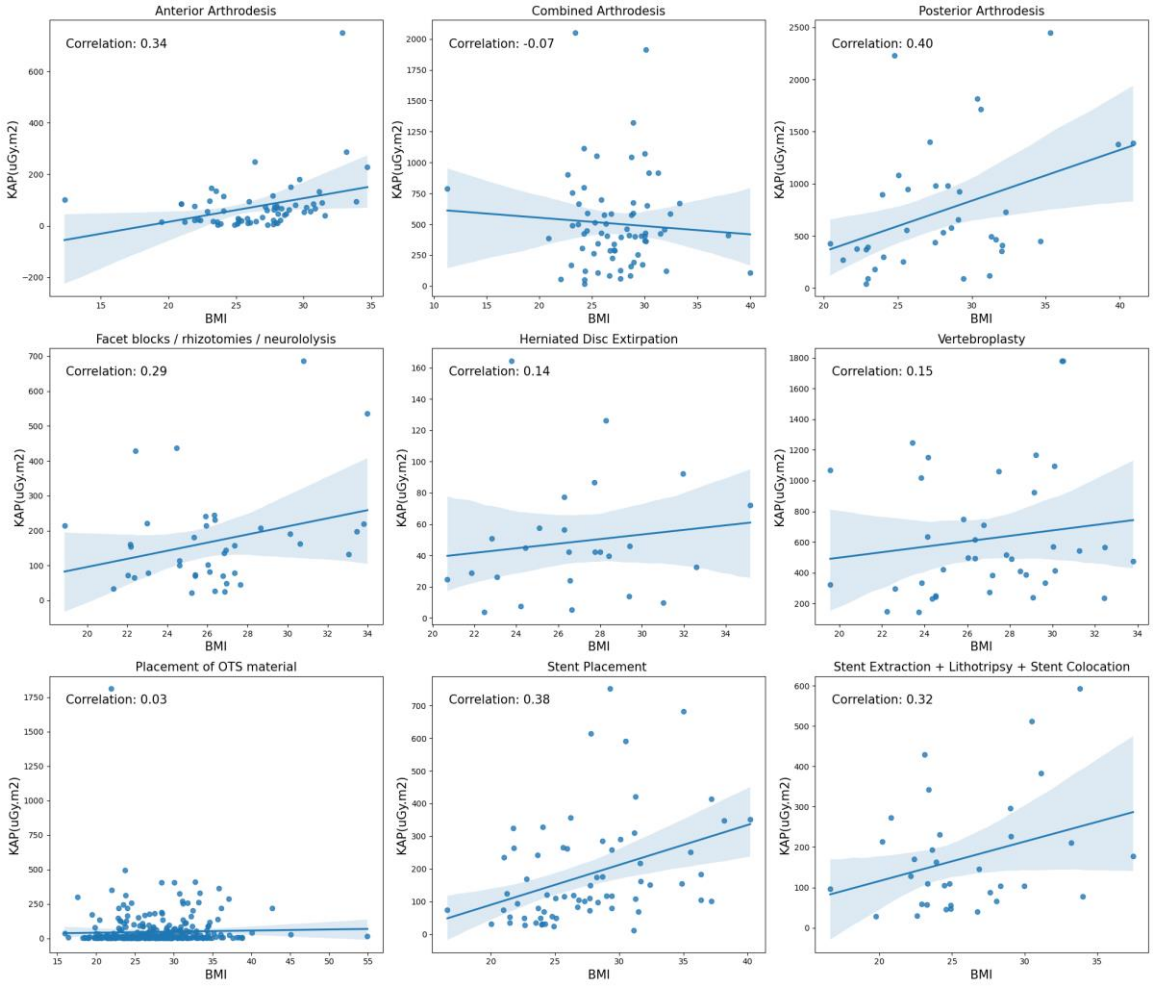

Figure 13. BMI vs KAP* for each Surgery.

According to the literature presented in Table 9, there is a relationship between BMI and KAP, however, the correlation values (R2) are generally less than 0.36 [53].

Table 9. BMI relevance dependent of each procedure, reported in the literature.

References	Procedure	BMI Relevance
Kukreja et al., 2015 [53]	Minimally invasive spine surgery	Severe obesity (BMI ≥ 35) is associated with a significantly risk of radiation exposure compared with other weight categories (R^2 = 0.36).
Dolenc et al., 2022 [54]	General radiography	BMI and body type (overweight and obese) have a strong influence on the radiation dose received by patients.
Madder et al., 2019 [55]	Coronary angiography	Increasing patient BMI was associated with a significant increase in physician radiation dose.
Shah et al., 2015 [56]	Coronary Angiogram	Obese patients require more than double the radiation dose in comparison to those with normal BMI (R^2 =0.20).
Liu et al., 2020 [57]	Atrial Fibrillation	Patients with overweight/obesity undergoing RFA experienced a significantly increased procedure duration and radiation dose.

Consistent with the literature, the data presented in this study shows that as BMI increases, the median KAP values also increase, demonstrating a correlation between the two variables. These results are in agreement with the values obtained by Shah et al. in 2015 [56].

A significant positive correlation was observed between BMI and median KAP*, indicating that as BMI increased, median KAP* also increased, as shown in Figure 14. However, this correlation was not evident in the underweight category, possibly due to the small sample size and the influence of outliers. Notably, obese patients demonstrated a significantly higher median KAP* (103.42 mGy.m2), compared to all other BMI categories. This observation supports the direct proportionality between KAP and BMI features, as also concluded by the study from Kukreja et al., 2015 [53].

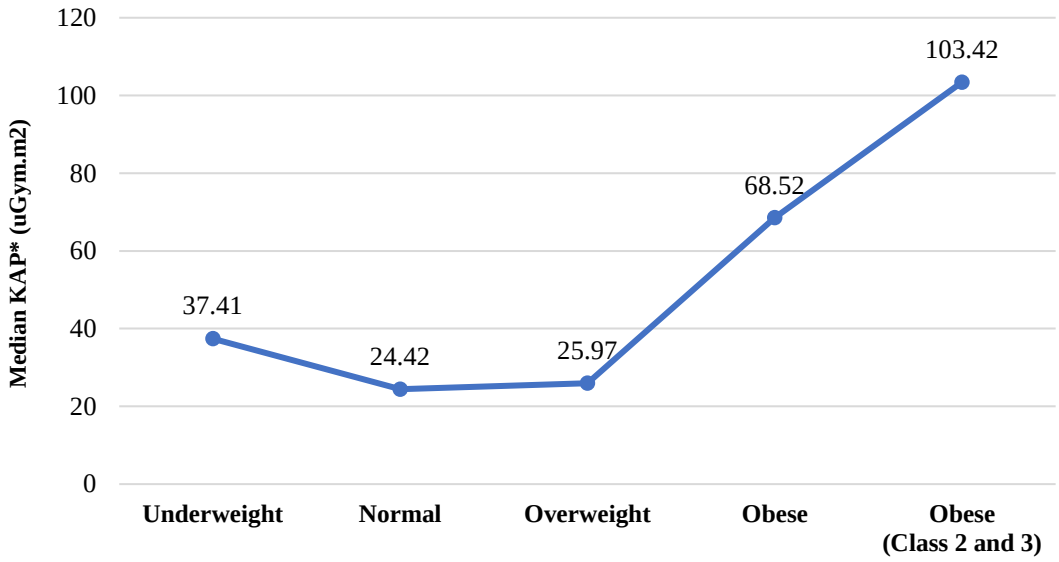

Figure 14. Median KAP* according to each weight category.

While Age exhibits an average correlation overall, its association with any of the target variables is highly dependent on the specific surgical procedure. As shown in Figure 15, only one surgery, Combined Arthrodesis, demonstrates a notable correlation and a slightly inclined slope.

Initial analysis revealed that Exposure Time and Number of Shots were among the most influential features affecting radiation exposure. However, these features cannot be preoperatively determined as they are recorded postoperatively. The same applies to Type of Exposure, as the operating room staff decides whether to use fluoroscopy or single shots during surgery.

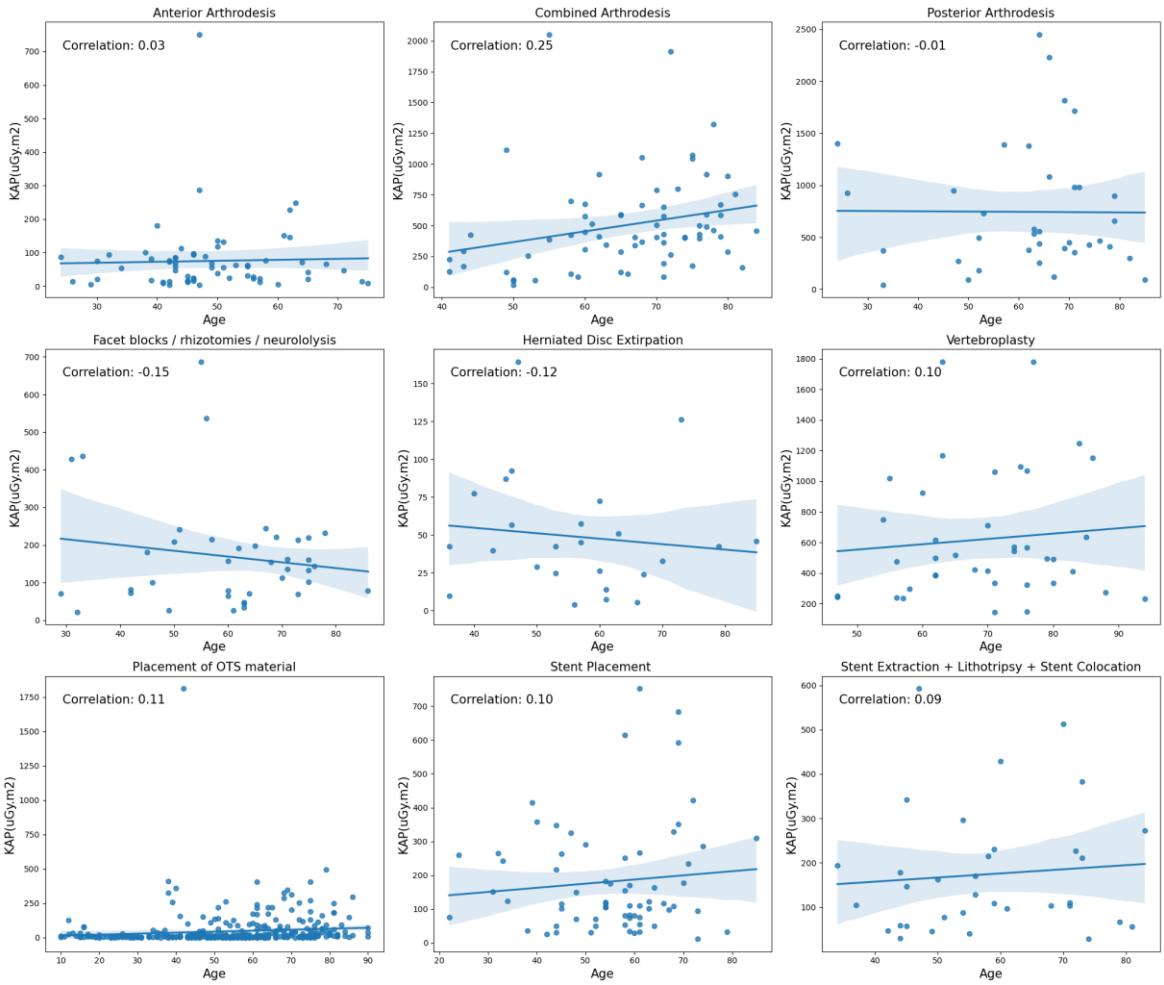

Figure 15. Age vs KAP* for each Surgery.

Given that Exposure Time is the most critical feature, a focused analysis was conducted to identify the most influential factors contributing to it. A variety of techniques were employed, including the 'Feature Importance' function from RFR, as shown in Figure 16. This function is a metric used in several embedded filter-based feature selection methods, which through training determines features that reduce impurity. It is important to note that the results obtained from this function represent the most important features only for the model in question.

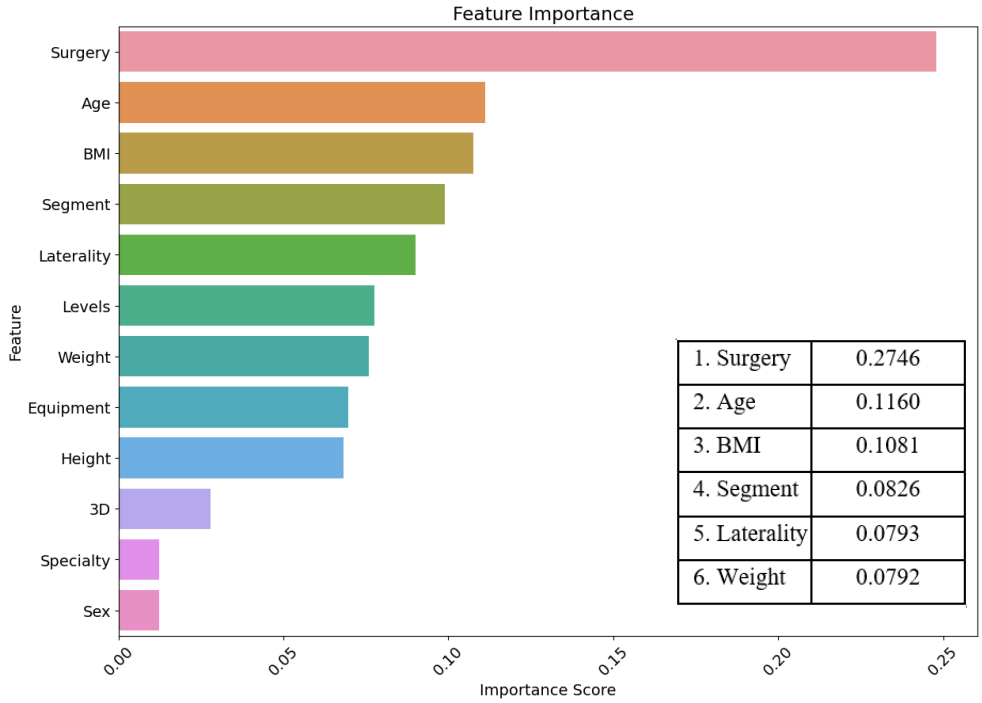

Figure 16. Feature Importance for Exposure Time.

Upon evaluating various features, it becomes apparent that Exposure Time is contingent upon the Surgical Procedure performed. Given that different specialties encompass diverse types of procedures, it is understandable that each specialty exhibits an average radiation exposure time, as illustrated in Figure 17. Although General Surgery / Gastro-Enterology represents a smaller number of cases, its Exposure Time of 89.66 seconds stands out in comparison to other specialties with substantially more cases but simpler procedures or lower radiation requirements. This observation underscores the influence of surgical complexity on radiation exposure.

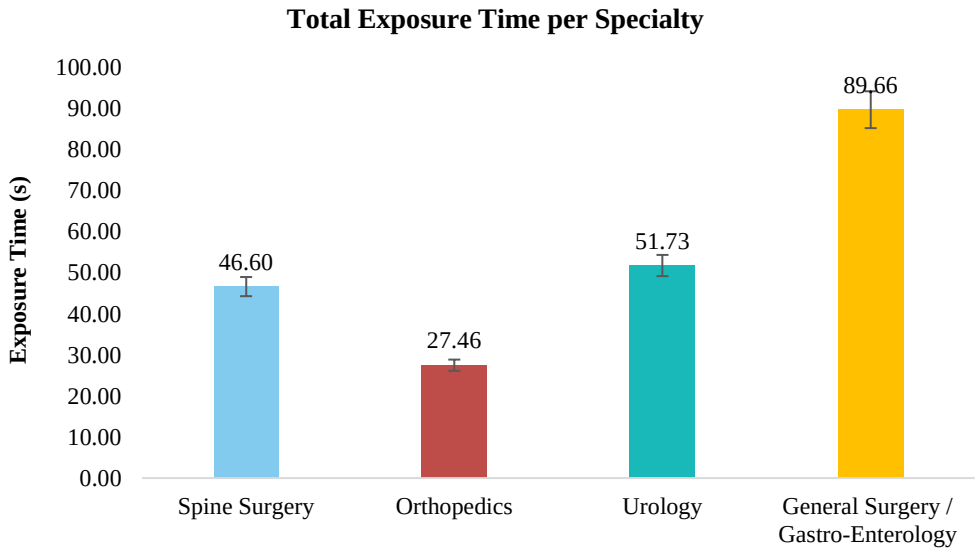

Figure 17. Total Exposure Time per Specialty in ratio.

Another technique employed was the calculation of *p-value* metrics, which utilizes a LR model to identify the most significant features for prediction, as presented in Table 10. The *p-value* is a statistical measure that evaluates the strength of evidence against a null hypodissertation. It represents the probability of obtaining results at least as extreme as the observed result, assuming that the null hypodissertation is true. A lower *p-value* indicates stronger evidence against the null hypodissertation, as explained in Table 11 [58].

Table 10. *p-value* feature significance for Exposure Time.

Feature	***p-value***
Laterality	<0.001
3D	<0.001
Levels	<0.001
Surgery	<0.001
Segment	<0.001
Age	<0.05
Equipment	<0.05
Specialty	>0.05
Height	>0.05
Sex	>0.05
BMI	>0.05
Weight	>0.05

Table 11. Significance level in *p-value* metric. Adapted from [59].

Significance Level	**Specification**
$p > 0.05$	Not significant
$p \leq 0.05$ (5%)	Significant
$p \leq 0.01$ (1%)	Very significant
$p < 0.001$ (0.1%)	Highly significant

The preceding results (Table 10) reveal the significant impact of Laterality on Exposure Time. This is unsurprising as bilateral procedures typically take approximately twice as long and require double dose of radiation as unilateral procedures. Additionally, the specialty in which the procedure is performed influences Laterality, as demonstrated in Table 12.

Table 12. Frequency of Laterality in each Specialty.

Specialty	Laterality			Total Cases
	Not Applicable	Unilateral	Bilateral	
Spine Surgery	66	122	223	411
Orthopedics	0	439	5	444
Urology	0	132	2	134
General Surgery	1	34	0	35
Total	67	727	230	1024

The 3D feature demonstrates high significance, as evidenced by an increase in Exposure Time when a surgeon requests a 3D image of the affected area to generate a three-dimensional representation. However, the relatively low utilization of this technique, as shown in Table 13, may influence this result.

Table 13. Frequency of 3D in each Specialty.

Specialty	3D		Total Cases
	Yes	No	
Spine Surgery	12	399	411
Orthopedics	0	444	444
Urology	0	134	134
General Surgery	0	35	35
Total	12	1012	1024

As previously discussed, Surgery Performed and Levels significantly impact Exposure Time. Procedures involving deeper levels typically exhibit longer exposure times, as evident in Figure 18.

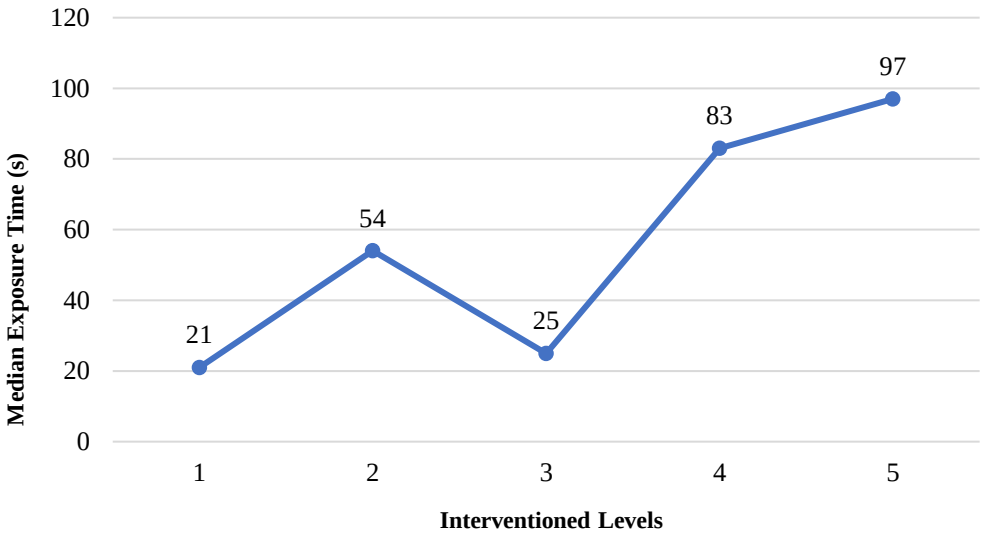

Figure 17. Frequency of intervened levels and median exposure time per level.

Furthermore, the body segment undergoing exposure significantly influences the duration of radiation exposure. Larger body segments generally require longer exposure times compared to smaller extremities like hands or feet. This observation is illustrated in Figure 18, where the Torso segment necessitates more than triple the exposure time compared to the Hand. This difference stems from the varying radiation attenuation properties of different body tissues and the anatomical complexity of larger segments.

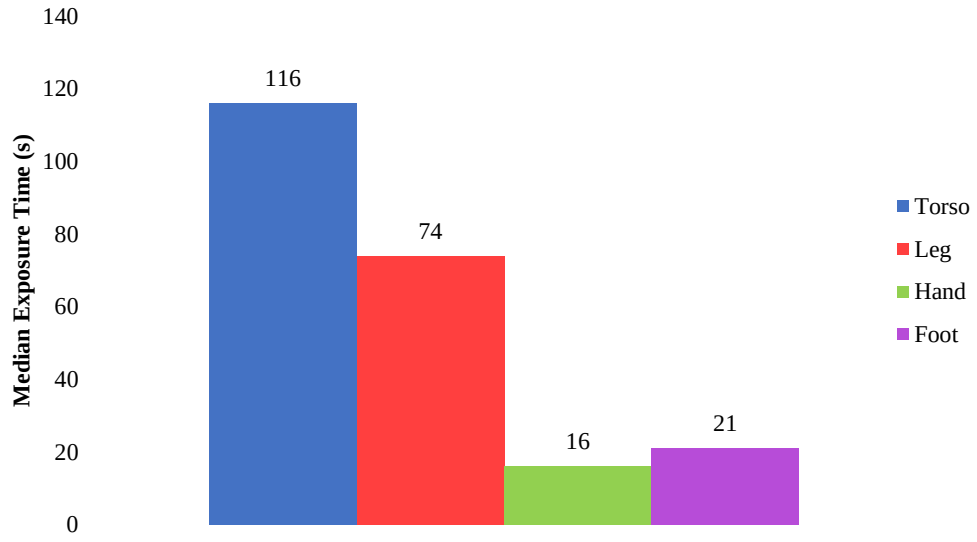

Figure 18. Median Exposure Time for the body segments under procedure.

4.2. Machine Learning Analysis

In this subsection, it was explored the application of the five ML algorithms previously mentioned, to a radiation exposure dataset, examining their performance in predicting various radiation-related metrics. By employing regression analysis, the aim was to gain insights into the factors influencing radiation exposure levels and establish models that can accurately predict these values.

4.2.1. First Approach

In the initial phase of the study, five ML algorithms were trained and evaluated to analyze the radiation exposure dataset under investigation using regression analysis. The values obtained by each of the algorithms are summarized in Table 14. Upon comparison of the results, both MSE and MAE values exhibited the expected pattern, with MSE values being higher than MAE values. The results obtained with CV were marginally lower than those obtained without, as CV provides a more accurate representation of the regression models' true accuracy by employing the entire dataset for both training and testing.

All five machine learning models employed in this study were feeding with a training set representing a 70% of the total dataset instances (patients' cases) and a 30% of the total

instances were keeping on the test set. All models were run with their default settings and evaluated using 10-fold cross-validation to compare their performance to models without cross-validation. No features were dropped, except when the model was predicting AK, where KAP* was excluded and vice versa, as these features are mutually exclusive and including both could introduce bias into the prediction.

Comparing the performance of the five models (Table 14), the RF model achieved the closest values to 1 for both MSE and MAE, indicating its superior performance compared to the other four models. Since the evaluation metric values for each regressor were subtracted from 1.0 (1-MSE or 1-MAE). The ML models developed in this study were further compared with similar algorithms used by Ahn et al., 2021 [16]. The results obtained in the present work exhibited higher accuracy, despite the database used by Ahn et al., 2021 [16] containing fewer instances than the one employed in this study.

Table 14. Median Exposure Time for the body segments under procedure.

Metric	MSE	MAE
Air Kerma	RF: 0.9975 ± 0.0010	RF: 0.9809 ± 0.0029
	SVR: 0.9958 ± 0.0013	SVR: 0.9472 ± 0.0008
	ANN: 0.9966 ± 0.0012	ANN: 0.9670 ± 0.0029
	DT: 0.9959 ± 0.0020	DT: 0.9772 ± 0.0057
	LR: 0.9979 ± 0.0010	LR: 0.9733 ± 0.0035
Air Kerma (CV)	RF: 0.9955 ± 0.0078	RF: 0.9751 ± 0.0269
	SVR: 0.9933 ± 0.0065	SVR: 0.9456 ± 0.0181
	ANN: 0.9943 ± 0.0084	ANN: 0.9612 ± 0.0267
	DT: 0.9920 ± 0.0086	DT: 0.9658 ± 0.0294
	LR: 0.9958 ± 0.0061	LR: 0.9662 ± 0.0204
KAP*	**RF: 0.9981 ± 0.0003**	**RF: 0.9825 ± 0.0016**
	SVR: 0.9961 ± 0.0008	SVR: 0.9518 ± 0.0032
	ANN: 0.9969 ± 0.0012	ANN: 0.9706 ± 0.0093
	DT: 0.9964 ± 0.0001	DT: 0.9772 ± 0.0014
	LR: 0.9980 ± 0.0001	LR: 0.9759 ± 0.0020
KAP* (CV)	RF: 0.9975 ± 0.0031	RF: 0.9792 ± 0.0169
	SVR: 0.9945 ± 0.0035	SVR: 0.9453± 0.0150
	ANN: 0.9944 ± 0.0057	ANN: 0.9520 ± 0.0213
	DT: 0.9963 ± 0.0040	DT: 0.9742 ± 0.0197
	LR: 0.9977 ± 0.0021	LR: 0.9720 ± 0.0130

The algorithm additionally generated two scatter plots for each model, one for KAP* and one for AK, with actual values on the x-axis and predicted values by the models on the y-axis. Several patterns can be readily observed upon examining the scatter plots that compare the actual vs. predicted values on KAP* (Figure 19).

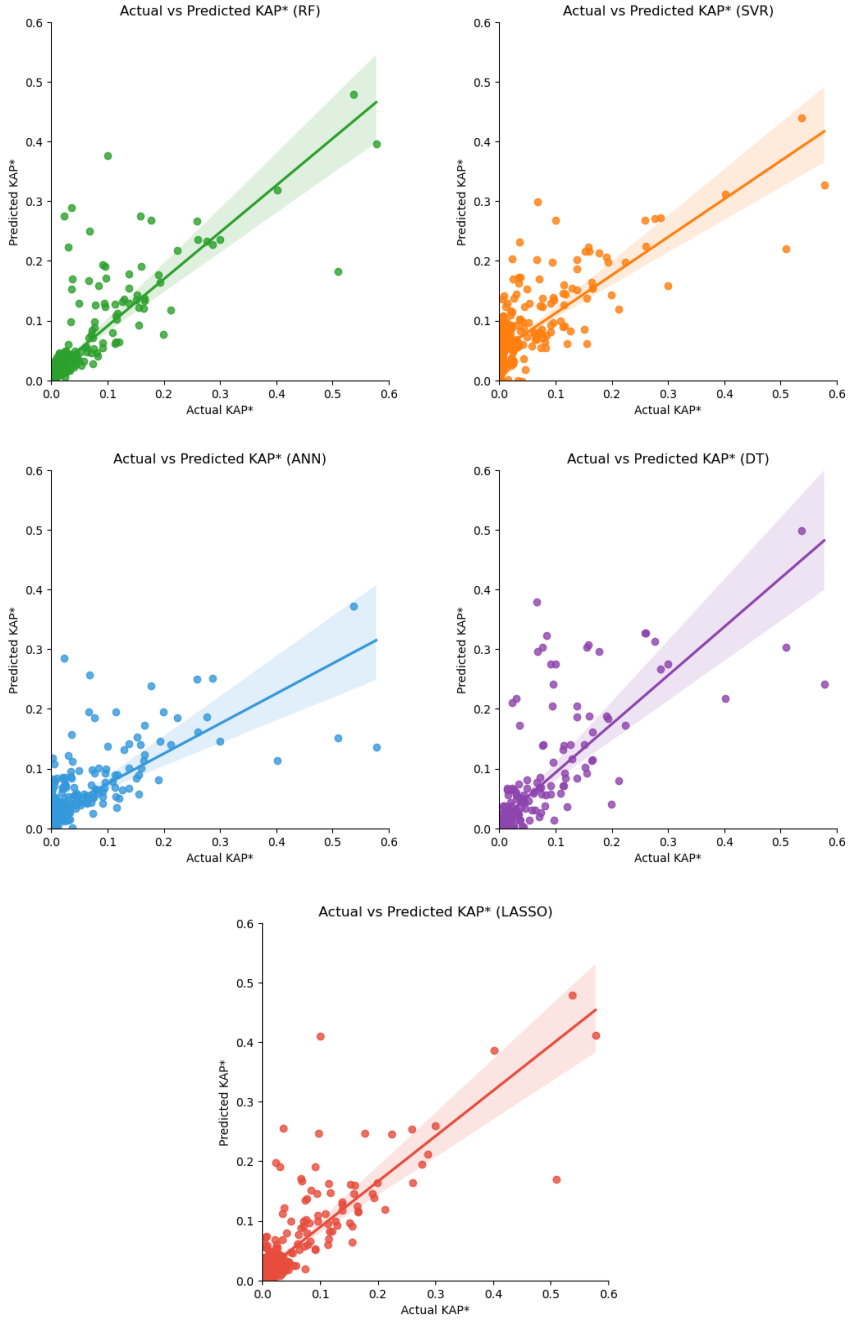

Figure 19. Comparison of Actual vs Predicted Values of KAP* for initial Regression Models.

All plots exhibit a positive slope, which suggests that the predicted values for KAP* are closer to the actual values, indicating a good prediction. Furthermore, the narrow confidence intervals for KAP* prediction support these results, indicating that the predicted KAP* values are likely to be closer to the actual KAP* values. This suggests that the model has good predictive accuracy and can be used for reliable estimation of KAP* values.

Analyzing the scatter plots obtained from all models for the prediction of AK (Figure 20) reveals positive slopes as well, indicating that a satisfactory prediction was made. However, the discrepancy between expected and actual results is greater than that observed for KAP*. Additionally, the confidence intervals for AK prediction are slightly wider than those for KAP* prediction, suggesting that while the predicted values are likely to be close to the actual values, the predictive accuracy is not as high as for KAP*. However, while wider confidence intervals indicate greater uncertainty in the predicted values, they do not necessarily imply lower prediction accuracy.

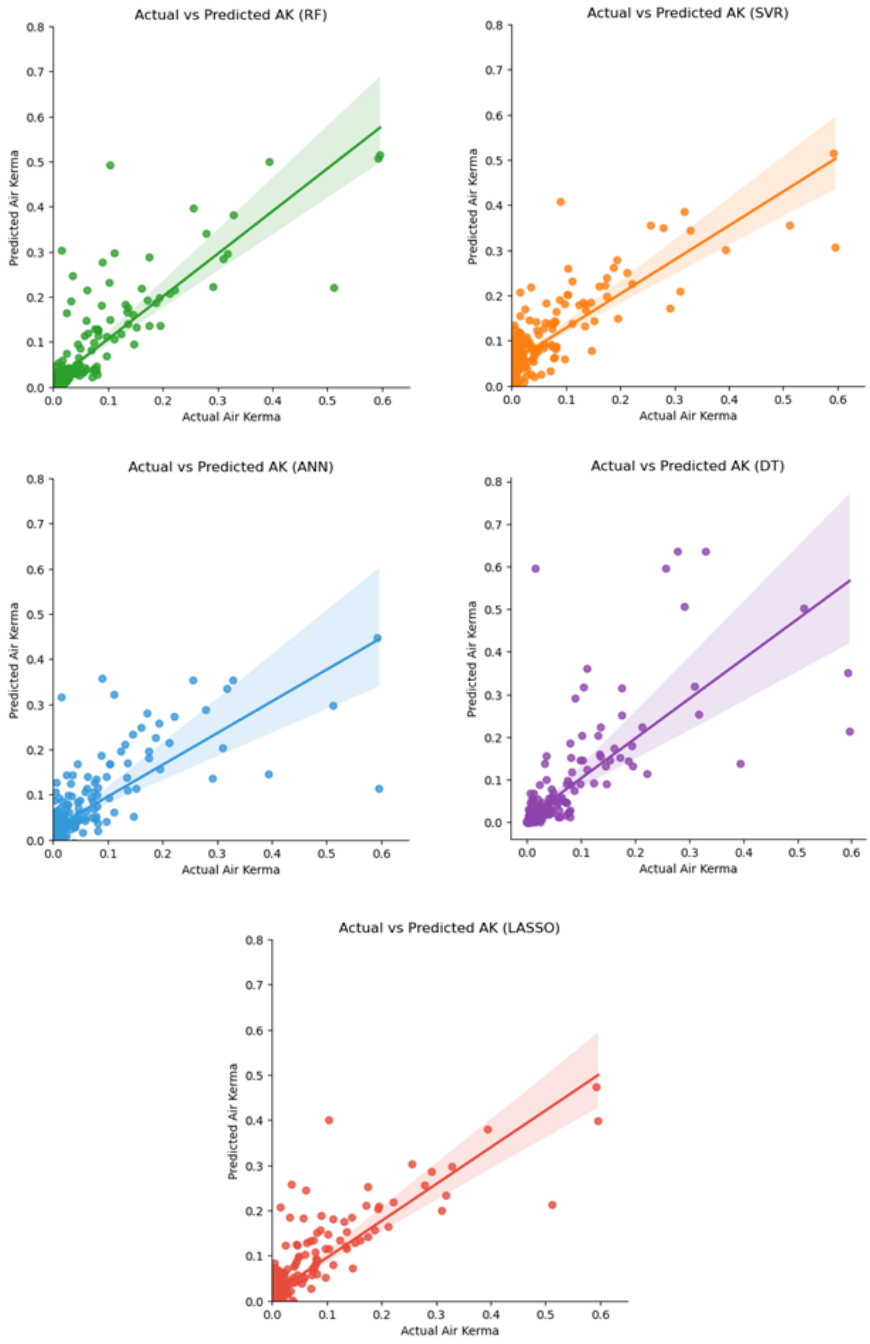

Figure 20. Comparison of Actual vs Predicted Values of AK for initial Regression Models.

AK outcomes tend to be more variable than KAP* outcomes, resulting in a wider spread in the data and potentially contributing to broader confidence intervals for AK prediction. Additionally, the relationship between AK and the predictors may be more intricate than the connection between KAP* and the predictors, making it more challenging to accurately model AK and leading to wider confidence intervals.

The scatter plots generated using the RF model exhibit a higher concentration of points clustered close to the regression line compared to other models. This model presents a more focused distribution of data clusters, enhancing its reliability. In contrast, the scatter plots produced by the SVR and ANN models display a more dispersed pattern of points.

A comprehensive analysis of all ten scatter plots reveals the superiority of the RF model for predicting KAP*. The slope of the regression line in the RF scatter plot approaches a 45-degree angle, indicating a strong positive correlation between the predicted and actual KAP* values. Additionally, the narrow confidence intervals associated with the RF model suggest a high level of precision in its predictions. Furthermore, a significant portion of the data points in the RF scatter plot align closely with the regression line, demonstrating its ability to capture the underlying relationship between the variables. Although a few outliers are present, the overall pattern indicates the RF model's superior performance for predicting KAP*.

4.2.2. Second Approach

Following an in-depth analysis of the initial results and a thorough exploration of strategies to enhance the performance of the models, a different approach was adopted. We train and tested all algorithms using a 10-fold split CV method to ensure a well-distributed allocation of data between training and testing. Feature selection was also performed, revealing that four features exhibited low importance for both KAP* and AK, as illustrated in Figure 21. Consequently, these features were removed from the dataset with the aim of enhancing model interpretability and potentially improving model performance. Additionally, the data split was modified to a 90/10 ratio, with 90% of the dataset assigned to training and 10% to testing. This modification allows the model to learn from a larger volume of data, potentially improving its ability to generalize to new, unseen examples. A larger training set enables the model to capture more intricate patterns and reduces the risk of overfitting.

A validation set was also introduced, which is incorporated within the K-fold CV process. This set, typically comprising around 10% of the training set, is utilized to evaluate the model's performance during training. The validation set provides valuable information into the model's ability to generalize to unseen data and helps prevent overfitting.

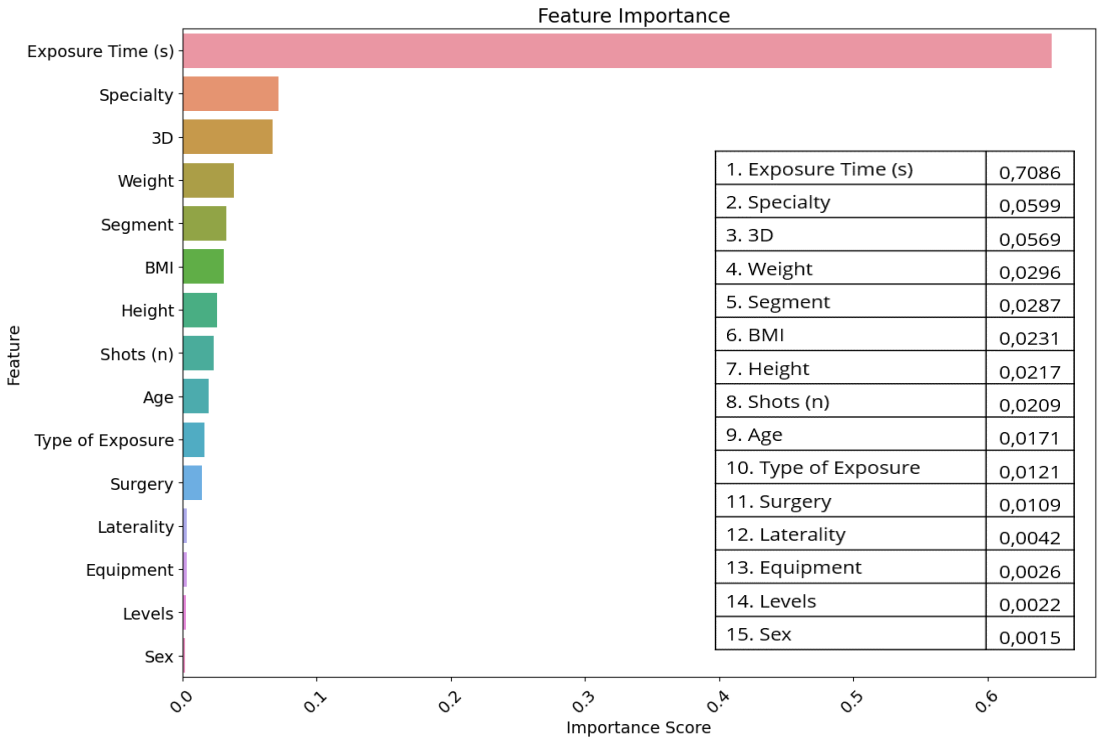

Figure 21. Feature importance for KAP* and AK.

The results obtained by the algorithms are summarized in Table 15. While some models exhibited a slight decline in performance compared to the previous results without CV, it is important to note that the application of CV provided a more accurate representation of the regression models' true accuracy.

Table 15. Results for AK and KAP* for regression models.

Metric	MSE	MAE
Air Kerma	RF: 0.9975 ± 0.0017	RF: 0.9804 ± 0.0040
	SVR: 0.9962 ± 0.0026	SVR: 0.9494 ± 0.0061
	ANN: 0.9968 ± 0.0021	ANN: 0.9676 ± 0.0043
	DT: 0.9959 ± 0.0020	DT: 0.9772 ± 0.0057
	LR: 0.9976 ± 0.0026	LR: 0.9712 ± 0.0059
KAP*	**RF: 0.9984 ± 0.0016**	**RF: 0.9844 ± 0.0038**
	SVR: 0.9963 ± 0.0014	SVR: 0.9496 ± 0.0029
	ANN: 0.9976 ± 0.0014	ANN: 0.9726 ± 0.0052
	DT: 0.9960 ± 0.0042	DT: 0.97899 ± 0.0090
	LR: 0.9981 ± 0.0006	LR: 0.9749 ± 0.0026

Comparing the performance of the five models, the RF model once again achieved the closest values to 1 for both MSE and MAE in KAP*, demonstrating its superior performance compared to the other four models.

The performance of these models was evaluated against similar algorithms developed by Huettenbrink et al. The results obtained in the current study demonstrate comparable or superior predictive performance to those reported by Huettenbrink et al. [49].

By utilizing the entire dataset for both training and testing, the CV method provides a tool to have a more realistic assessment of the models' ability to generalize to unseen data.

Each models' algorithm also generated two scatter plots of the best fold captured by the CV. Upon examining the scatter plots that compare the actual vs. predicted values on KAP* (Figure 22), several patterns can be readily observed.

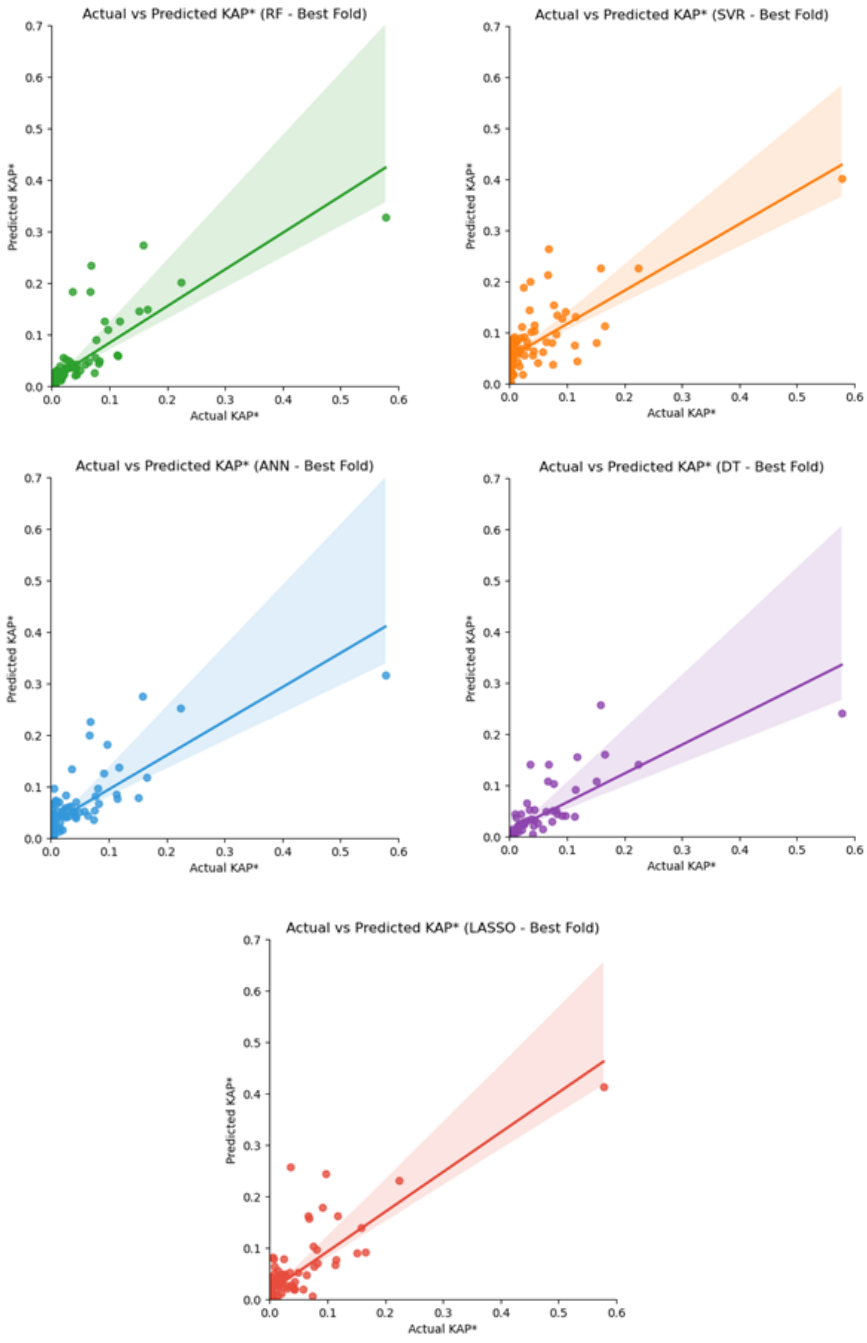

Figure 22. Comparison of Actual vs Predicted Values of KAP* for Regression Models.

All plots exhibit a positive slope, indicating a satisfactory prediction. Furthermore, the confidence intervals for KAP* prediction corroborate these results, implying that the predicted values are likely to be closer to the actual values. However, compared to the previous results, the confidence intervals are wider, suggesting that the models are less certain about their predictions, potentially resulting in a lower level of accuracy compared to the previous results. This difference in accuracy is likely attributed to the presence of outliers and multicollinearity.

Outliers can disrupt the overall pattern of the data, making it more difficult for the models to capture the underlying relationship between the variables. Multicollinearity, on the other hand, occurs when two or more variables are highly correlated, making it difficult to distinguish the independent effects of each variable on the target variable.

Analyzing the scatter plots obtained from all models for the prediction of AK (Figure 23) also reveals positive slopes, indicating that a satisfactory prediction was made. However, the discrepancy between expected and actual results is greater than that observed for KAP* and for the previous plots of AK prediction. This suggests that the models are less accurate at predicting AK values compared to KAP* values.

Additionally, the confidence intervals for AK prediction are wider than those for KAP* prediction and for the previous plots of AK prediction. This implies that while the predicted values are likely to be close to the actual values, the predictive accuracy is not as high as KAP*.

The wider confidence intervals for AK prediction suggest that the models are less certain about their predictions, potentially resulting in a lower level of accuracy compared to KAP* prediction. This difference in accuracy is likely attributed to the inherent complexity of predicting AK, as it is influenced by multiple factors, including individual characteristics and environmental exposures.

A comprehensive review of the latest results reinforces the superiority of the RF model in predicting both KAP* and AK. The MSE and MAE metrics presented in Table 15 demonstrate the superior performance of the RF model compared to the other models, further supported by its low standard deviation.

Due to the excessive correlation of Exposure Time on both KAP* and AK, the prediction of radiation exposure without considering Exposure Time proved to be infeasible, also because of the bias inserted in the prediction. As Exposure Time is an intrinsic parameter that cannot be determined prior to the procedure, an alternative target variable was sought. Analysis of correlations and feature importance, as presented in Table 10 and Figure 16, revealed that Exposure Time prediction was possible due to the strong correlation between certain features and the new target variable.

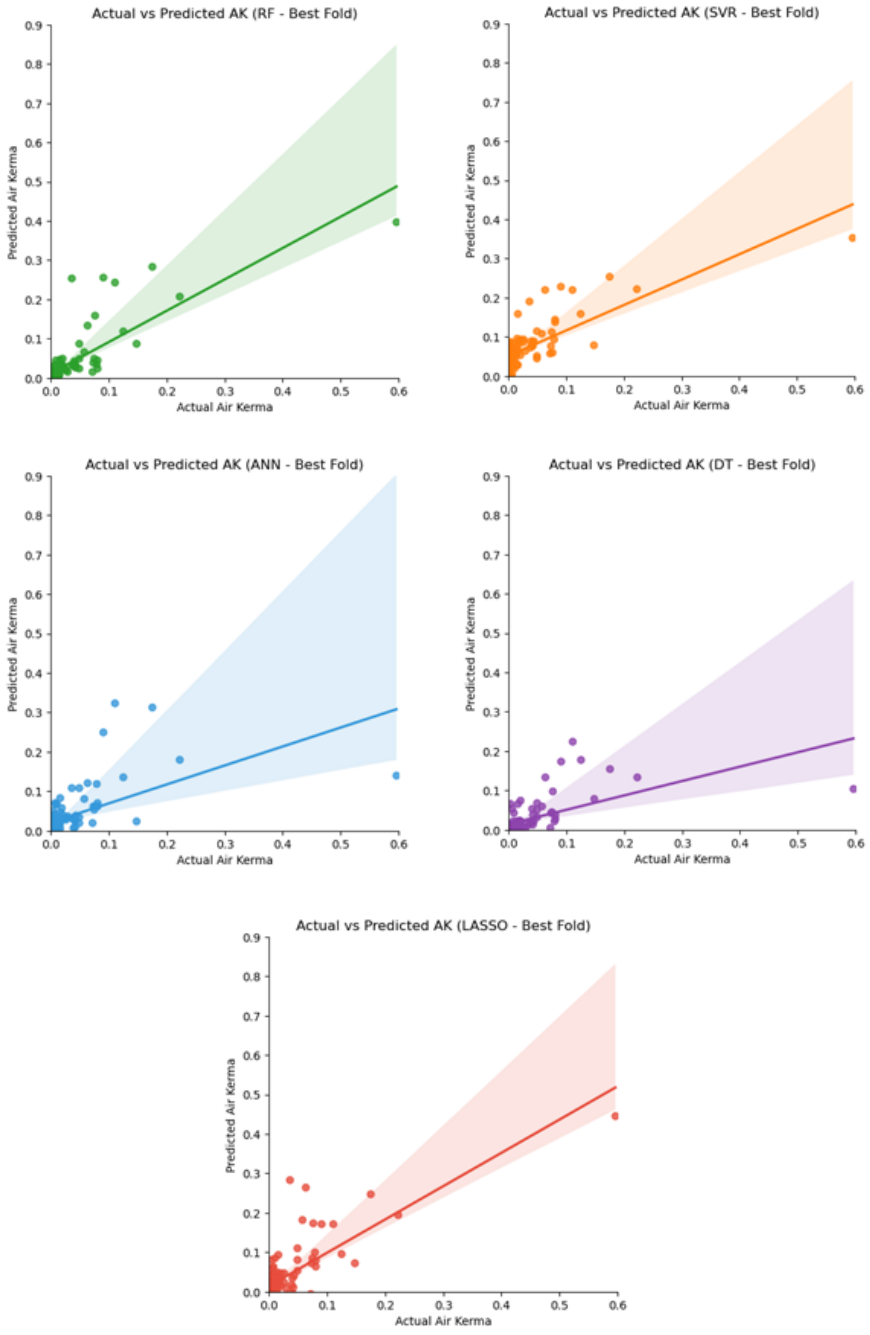

Figure 23. Comparison of Actual vs Predicted Values of AK for Regression Models.

4.2.3. Third Approach

The prediction of Exposure Time was initially attempted using five regression models trained on a dataset of 900 cases and evaluated on a separate test set of approximately 150 entries. Despite the model's promising performance on the training data, its generalization ability proved to be suboptimal. The model exhibited unsatisfactory performance on unseen data, as evidenced by the MSE and MAE values falling below 0.92. This suggests that the model may be overfitting to the training data, hindering its ability to effectively predict on new, unfamiliar data.

To improve predictive capacity for real-world data, the entire dataset of 1024 cases was utilized, with 90% allocated for training and 10% for testing. As expected, the results improved significantly due to the larger and more comprehensive dataset, shown in Table 16.

Table 16. Results for Exposure Time for regression models.

Metric	Exposure Time		Mean	STD
MSE	RF	**0.9932**	**0.9933**	**0.0034**
	SVR	0.9908	0.9924	0.0042
	ANN	0.9915	0.9926	0.0038
	DT	0.9930	0.9919	0.0033
	LASSO	0.9908	0.9917	0.0045
MAE	RF	**0.9515**	**0.9506**	**0.0065**
	SVR	0.9462	0.9505	0.0061
	ANN	0.9449	0.9443	0.0075
	DT	0.9476	0.9453	0.0080
	LASSO	0.9407	0.9423	0.0071

Feature selection was conducted for these algorithms, resulting in the removal of four features from the dataset: KAP*, AK, Number of Shots, and Type of Exposure. This choice was made due to the unavailability of these features prior to surgery, requiring Exposure Time prediction to rely solely on patient information and preoperative parameters. HT was also implemented using the *RandomizedSearchCV* and *GridSearchCV* functions, as previously mentioned. Despite the reduction of highly correlated features, the prediction performance remained satisfactory, with MSE values ranging around 0.99 and relatively low STDs. Additionally, the RF model consistently outperformed the other models in terms of both metrics. To enhance the robustness of the results, a comparison was conducted with a RF algorithm similar to the one developed in this study, incorporating both HT and feature selection methodology as proposed by Kaliappan et al. [60]. The findings of the current investigation demonstrate superior predictive accuracy, despite employing a smaller database.

An analysis of the Exposure Time prediction plots (Figure 24) reveals a small scattering of data in all models, with a cluster near the low values reflecting low radiation exposure times. The DT model exhibits a pattern of horizontal lines/dots, which is attributed to the regions where the tree splits or where the tree's predictions transition from one leaf to another.

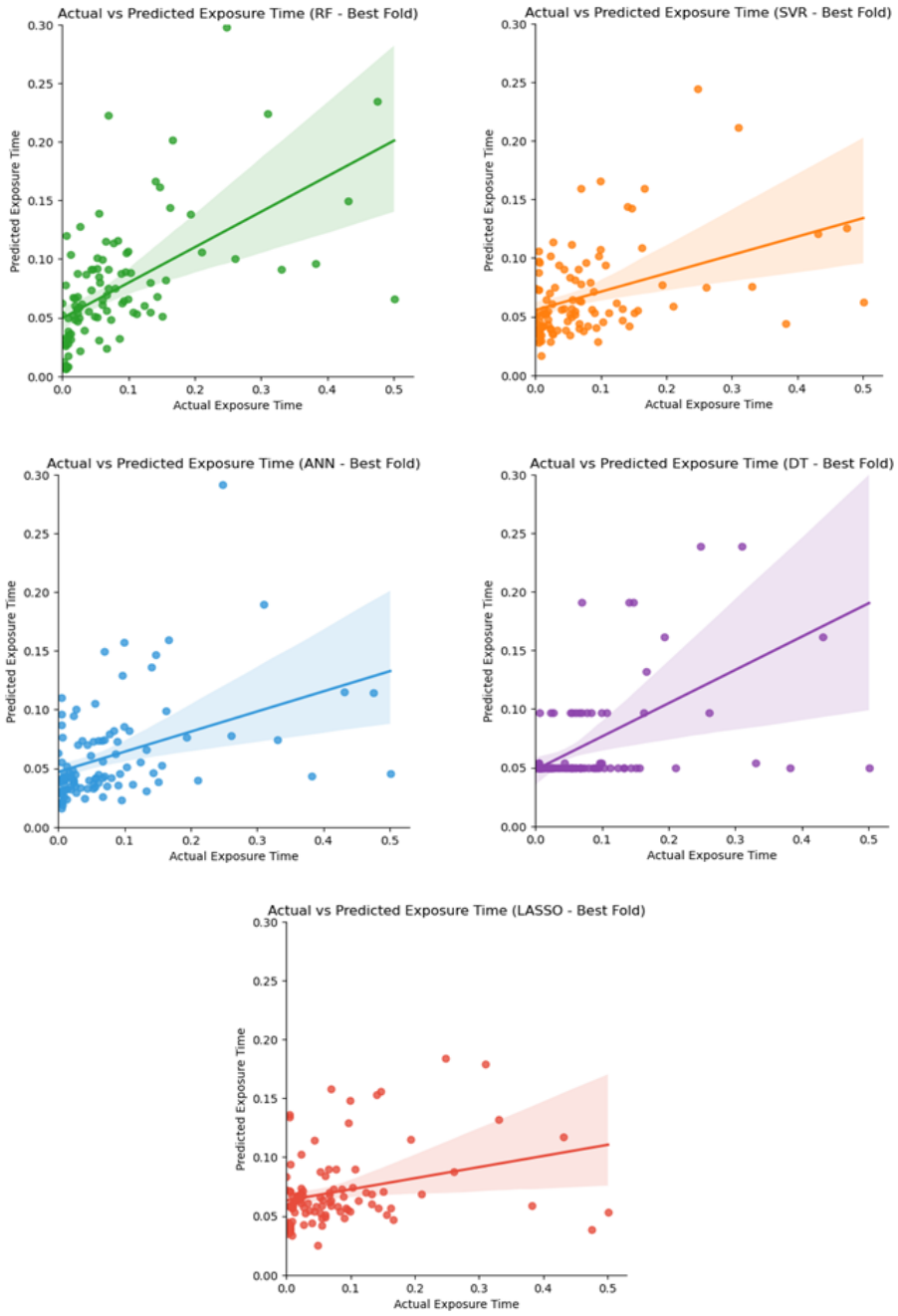

Figure 24. Comparison of Actual vs Predicted Exposure Time in Regression Models.

Based on these findings, a single model can be selected for predicting Exposure Time using an alternative approach. The ultimate objective is to employ the predicted Exposure Time to estimate both KAP* and AK, as it exhibits the strongest correlation among the features.

4.2.4. Fourth Approach

For this purpose, a two-layer train-test split was implemented using only the RF model since it had the best predictions for Exposure Time. The entire dataset was initially divided into a 10% test split and a 90% train split. Subsequently, the train split was further subdivided into

two equal subsets: one for predicting KAP* and AK using the Original Exposure Time (Te), and the other one for predicting KAP* and AK using the Predicted Exposure Time (Tep).

The previous selection of features remained unchanged, except that KAP* and AK were removed when they are being predicted. HT was also performed as previously mentioned.

An examination of the Exposure Time prediction results obtained using the final approach (Figure 25) reveals a larger cluster of data and a narrower confidence interval. The corresponding metric values, MSE = 0.9925 and MAE = 0.9478, indicate satisfactory performance compared to the previous model evaluations.

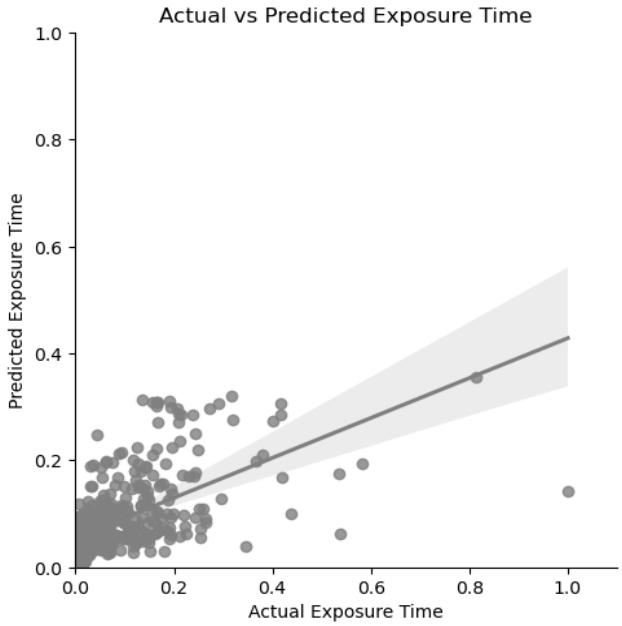

Figure 25. Prediction of Exposure Time.

Predicting Exposure Time beforehand enhances the reliability of radiation exposure estimation through KAP* and AK, as time significantly impacts the radiation received by the patient. The plots generated by this algorithm (Figure 26) present the predicted KAP* and AK values obtained using the predicted Exposure Time, demonstrating a favorable slope with minimal data dispersion and an acceptable confidence interval. The metric results, MSE = 0.9965 and MAE = 0.9668 for AK, and MSE = 0.9968 and MAE = 0.9704 for KAP*, are relatively satisfactory considering the fact that this is a prediction based on a previous prediction.

In the case of predicted AK and KAP*, the regression is nearly perfect, with results surpassing those obtained in previous approaches. This improvement is likely attributed to the smaller set of data used and the strong correlation between all features, as the Exposure Time utilized is directly derived from procedure data. The metric results for the target variables are

remarkably close to perfect, with MSE = 0.9999 and MAE = 0.9988 for both, differing only in the fifth decimal place.

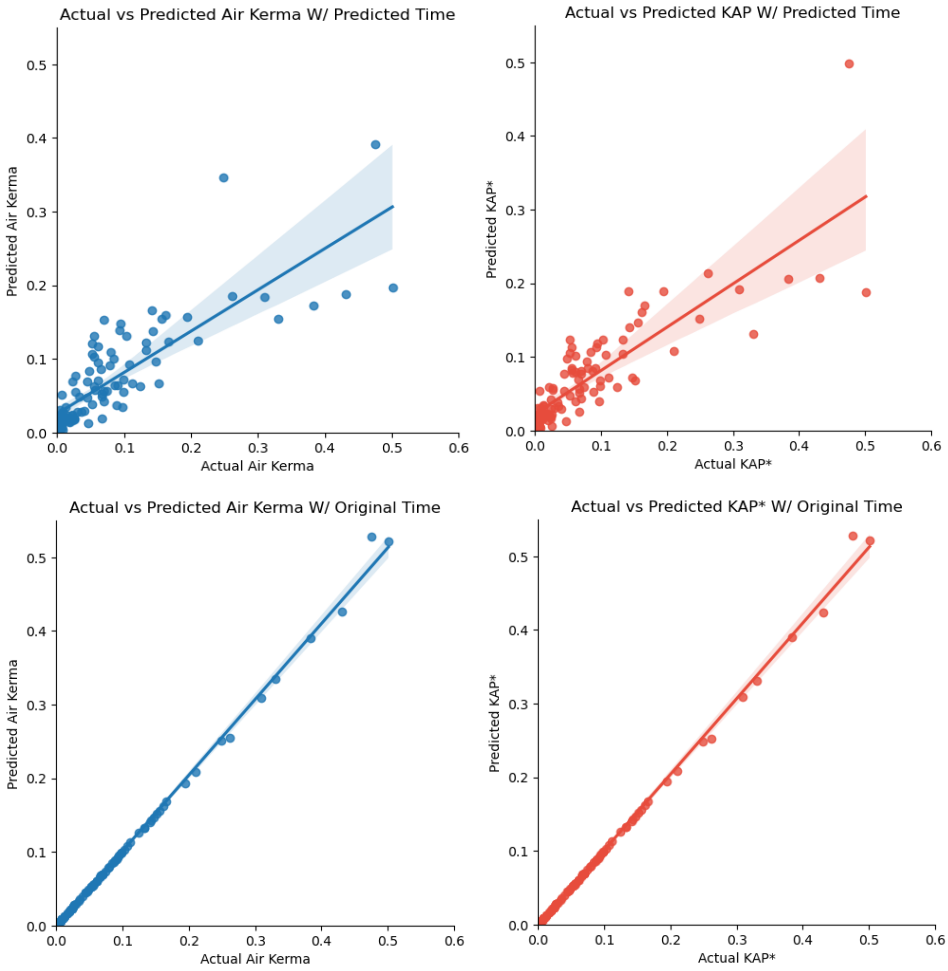

Figure 26. Comparison between AK and KAP* Prediction using Predicted and Original Exposure Time.

The findings presented here demonstrate that predicting radiation exposure received by the patient through the variables KAP* and AK is achievable, even when solely employing features obtained prior to the medical procedure. Furthermore, the prediction of Exposure Time has also been established as successful, with a slightly lower error margin compared to the other target variables, likely due to the absence of highly correlated features.

Chapter 5

Chapter 5 provides a comprehensive conclusion of the current research, outlining its limitations and exploring future work.

5. Conclusions

This study demonstrates the feasibility of predicting radiation exposure received by patients, even when relying solely on preoperatively obtained features. Moreover, the final developed / proposed regression model to predict the “Radiation Exposure Time”, at least experimentally have proved to be highly promising, as it enables radiation technicians to employ different techniques and optimize radiation exposure timing based on “individual patient characteristics and the specific procedure”. Notably, this approach can be applied across four distinct medical specialties, encompassing a wide range of surgical procedures.

Through the analysis of feature relevance, it is revealed that Exposure Time and Number of Shots exhibit strong correlation coefficients, indicating their significance in predicting radiation exposure variables AK and KAP*. Additionally, as shown in previous reported works, it is demonstrated that BMI emerges as a relevant feature depending on the medical specialty and surgical procedure, due to, it can contribute to increased radiation exposure depending on patient BMI.

A comparative analysis of the five regression models revealed the superior performance of the RF model, consistently generating values closer to 1 for both evaluation metrics (MSE and MAE). This finding is further corroborated by the analysis of scatter plots, which revealed the RF model's superior predictive performance, particularly for the KAP* feature.

As an important contribution of this work, in narrow collaboration with the Hospital da Luz, Setúbal, we have developed a benchmarking dataset comprising a total of 1024 patients’ cases for exploring radiation exposure dose, which will be released to public domain.

This study was potentially limited by the availability of a larger dataset with a broader range of features, as some features exhibited negligible correlation with the target variables. Additionally, the inability to incorporate features related to surgeon and technician experience due to data protection restrictions imposed by the hospital board presented a challenge.

Future research is needed to validate these models in larger and more diverse patient populations, as well as to develop models for other imaging modalities. Subsequently, the implementation of these predictive models into real-time surgical procedures, providing immediate feedback to surgeons and technicians, represents the next final step of this research.

References

1. Jenkins NW, Parrish JM, Sheha ED, Singh K (2021) Intraoperative risks of radiation exposure for the surgeon and patient. Ann Transl Med 9:84–84. https://doi.org/10.21037/atm-20-1052
2. Ojodu I, Ogunsemoyin A, Hopp S, et al (2018) C-arm fluoroscopy in orthopaedic surgical practice. European Journal of Orthopaedic Surgery & Traumatology 28:1563–1568. https://doi.org/10.1007/s00590-018-2234-7
3. Canon Medical Systems USA (2017) Dose Tracking System: Patient Skin Dose Estimates in Real Time
4. Paulo G, Bartal G, Vano E (2021) Radiation Dose of Patients in Fluoroscopically Guided Interventions: an Update. Cardiovasc Intervent Radiol 44:842–848. https://doi.org/10.1007/s00270-020-02667-3
5. Toshiba America Medical Systems Inc (2014) Dose Tracking System. A Paradigm Shift in Patient Dose Monitoring
6. Falco MD, Masala S, Stefanini M, et al (2015) Patient skin dose measurements using a cable free system MOSFETs based in fluoroscopically guided percutaneous vertebroplasty, percutaneous disc decompression, radiofrequency medial branch neurolysis, and endovascular critical limb ischemia. J Appl Clin Med Phys 16:298–310. https://doi.org/10.1120/jacmp.v16i1.5020
7. Brown PH, Thomas RD, Silberberg PJ, Johnson LM (2000) Optimization of a fluoroscope to reduce radiation exposure in pediatric imaging. Pediatr Radiol 30:229–235. https://doi.org/10.1007/s002470050728
8. Narain AS, Hijji FY, Yom KH, et al (2017) Radiation exposure and reduction in the operating room: Perspectives and future directions in spine surgery. World J Orthop 8:524. https://doi.org/10.5312/wjo.v8.i7.524
9. Bratschitsch G, Leitner L, Stücklschweiger G, et al (2019) Radiation Exposure of Patient and Operating Room Personnel by Fluoroscopy and Navigation during Spinal Surgery. Sci Rep 9:17652. https://doi.org/10.1038/s41598-019-53472-z
10. Howells P, Eaton R, Patel AS, et al (2012) Risk of Radiation Exposure during Endovascular Aortic Repair. European Journal of Vascular and Endovascular Surgery 43:393–397. https://doi.org/10.1016/j.ejvs.2011.12.031

11. Loose R, Wucherer M (2021) How to Measure/Calculate Radiation Dose in Patients? Cardiovasc Intervent Radiol 44:835–841. https://doi.org/10.1007/s00270-021-02772-x
12. Noh T, Mustroph M, Golby AJ (2021) Intraoperative Imaging for High-Grade Glioma Surgery. Neurosurg Clin N Am 32:47–54. https://doi.org/10.1016/j.nec.2020.09.003
13. Rednam M, Tiwari V. (2023) Fluoroscopy Orthopedic Assessment, Protocols, and Interpretation. In: StatPearls. https://www.ncbi.nlm.nih.gov/books/NBK573065/. Accessed 14 Oct 2023
14. Orth RC, Wallace MJ, Kuo MD (2009) C-arm Cone-beam CT: General Principles and Technical Considerations for Use in Interventional Radiology. Journal of Vascular and Interventional Radiology 20:S538–S544. https://doi.org/10.1016/j.jvir.2009.04.026
15. Keil H, Beisemann N, Swartman B, et al (2018) Intra-operative imaging in trauma surgery. EFORT Open Rev 3:541–549. https://doi.org/10.1302/2058-5241.3.170074
16. Ahn SH, Kim E, Kim C, et al (2021) Deep learning method for prediction of patient-specific dose distribution in breast cancer. Radiation Oncology 16:154. https://doi.org/10.1186/s13014-021-01864-9
17. Christopher J. Williams, Walter Sussman, John Pitts (2022) Atlas of Interventional Orthopedics Procedures, 1st ed
18. Asari T, Rokunohe D, Sasaki E, et al (2022) Occupational ionizing radiation-induced skin injury among orthopedic surgeons: A clinical survey. Journal of Orthopaedic Science 27:266–271. https://doi.org/10.1016/j.jos.2020.11.008
19. Bhanot R, Hameed ZBM, Shah M, et al (2022) ALARA in Urology: Steps to Minimise Radiation Exposure During All Parts of the Endourological Journey. Curr Urol Rep 23:255–259. https://doi.org/10.1007/s11934-022-01102-z
20. Jenkins NW, Parrish JM, Sheha ED, Singh K (2021) Intraoperative risks of radiation exposure for the surgeon and patient. Ann Transl Med 9:84–84. https://doi.org/10.21037/atm-20-1052
21. Frane N, Megas A, Stapleton E, et al (2020) Radiation Exposure in Orthopaedics. JBJS Rev 8:e0060–e0060. https://doi.org/10.2106/JBJS.RVW.19.00060
22. Soison J, Usawachintachit M (2022) Radiation safety and protection in urology. Insight Urology 43:140–149. https://doi.org/10.52786/isu.a.60
23. Goodman BS, Carnel CT, Mallempati S, Agarwal P (2011) Reduction in Average Fluoroscopic Exposure Times for Interventional Spinal Procedures Through the Use of

Pulsed and Low-Dose Image Settings. Am J Phys Med Rehabil 90:908–912. https://doi.org/10.1097/PHM.0b013e318228c9dd

24. Plastaras C, Appasamy M, Sayeed Y, et al (2013) Fluoroscopy procedure and equipment changes to reduce staff radiation exposure in the interventional spine suite. Pain Physician 16:E731-8
25. Garcia-Sanchez A-J, Garcia Angosto E, Llor JL, et al (2019) Machine Learning Techniques Applied to Dose Prediction in Computed Tomography Tests. Sensors 19:5116. https://doi.org/10.3390/s19235116
26. Bhardwaj KK, Banyal S, Sharma DK (2019) Artificial Intelligence Based Diagnostics, Therapeutics and Applications in Biomedical Engineering and Bioinformatics. In: Internet of Things in Biomedical Engineering. Elsevier, pp 161–187
27. Battineni G, Sagaro GG, Chinatalapudi N, Amenta F (2020) Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. J Pers Med 10:21. https://doi.org/10.3390/jpm10020021
28. Michael W. Berry, Azlinah Mohamed, Bee Wah Yap (2020) Supervised and Unsupervised Learning for Data Science. Springer International Publishing, Cham
29. Dönmez P (2013) Introduction to Machine Learning. Nat Lang Eng 19:. https://doi.org/10.1017/S1351324912000290
30. Moustafa SSR, Abdalzaher MS, Yassien MH, et al (2021) Development of an Optimized Regression Model to Predict Blast-Driven Ground Vibrations. IEEE Access 9:31826–31841. https://doi.org/10.1109/ACCESS.2021.3059018
31. Kohli S, Godwin GT, Urolagin S (2021) Sales Prediction Using Linear and KNN Regression. pp 321–329
32. Meng Z, Wang M, Guo S, et al (2021) Development and Validation of a LASSO Prediction Model for Better Identification of Ischemic Stroke: A Case-Control Study in China. Front Aging Neurosci 13:. https://doi.org/10.3389/fnagi.2021.630437
33. El Mrabet Z, Sugunaraj N, Ranganathan P, Abhyankar S (2022) Random Forest Regressor-Based Approach for Detecting Fault Location and Duration in Power Systems. Sensors 22:458. https://doi.org/10.3390/s22020458
34. Yuchi W, Gombojav E, Boldbaatar B, et al (2019) Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. Environmental Pollution 245:746–753. https://doi.org/10.1016/j.envpol.2018.11.034

35. Pangarkar DJ, Sharma R, Sharma A, Sharma M (2020) Assessment of the Different Machine Learning Models for Prediction of Cluster Bean (Cyamopsis tetragonoloba L. Taub.) Yield. Adv Res 98–105. https://doi.org/10.9734/air/2020/v21i930238

36. Awad M, Khanna R (2015) Support Vector Regression. In: Efficient Learning Machines. Apress, Berkeley, CA, pp 67–80

37. Parbat D, Chakraborty M (2020) A python based support vector regression model for prediction of COVID19 cases in India. Chaos Solitons Fractals 138:109942. https://doi.org/10.1016/j.chaos.2020.109942

38. Kadam AK, Wagh VM, Muley AA, et al (2019) Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Model Earth Syst Environ 5:951–962. https://doi.org/10.1007/s40808-019-00581-3

39. Robeson SM, Willmott CJ (2023) Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. PLoS One 18:e0279774. https://doi.org/10.1371/journal.pone.0279774

40. Agatonovic-Kustrin S, Beresford R (2000) Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal 22:717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

41. Alakbari FS, Mohyaldinn ME, Ayoub MA, et al (2023) A decision tree model for accurate prediction of sand erosion in elbow geometry. Heliyon 9:e17639. https://doi.org/10.1016/j.heliyon.2023.e17639

42. Botchkarev A (2019) A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. Interdisciplinary Journal of Information, Knowledge, and Management 14:045–076. https://doi.org/10.28945/4184

43. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7:e623. https://doi.org/10.7717/peerj-cs.623

44. Hodson TO, Over TM, Foks SS (2021) Mean Squared Error, Deconstructed. J Adv Model Earth Syst 13:. https://doi.org/10.1029/2021MS002681

45. Hodson TO (2022) Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geosci Model Dev 15:5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

46. Karch J (2020) Improving on Adjusted R-Squared. Collabra Psychol 6:. https://doi.org/10.1525/collabra.343
47. Refaeilzadeh P, Tang L, Liu H (2009) Cross-Validation. In: Encyclopedia of Database Systems. Springer US, Boston, MA, pp 532–538
48. Berrar D (2019) Cross-Validation. In: Encyclopedia of Bioinformatics and Computational Biology. Elsevier, pp 542–545
49. Huettenbrink C, Hitzl W, Distler F, et al (2023) Personalized Prediction of Patient Radiation Exposure for Therapy of Urolithiasis: An Application and Comparison of Six Machine Learning Algorithms. J Pers Med 13:643. https://doi.org/10.3390/jpm13040643
50. Elgeldawi E, Sayed A, Galal AR, Zaki AM (2021) Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. Informatics 8:79. https://doi.org/10.3390/informatics8040079
51. Miller DL (2020) Review of air kerma-area product, effective dose and dose conversion coefficients for non-cardiac interventional fluoroscopy procedures. Med Phys 47:975–982. https://doi.org/10.1002/mp.13990
52. Kwon D, Little MP, Miller DL (2011) Reference air kerma and kerma-area product as estimators of peak skin dose for fluoroscopically guided interventions. Med Phys 38:4196–4204. https://doi.org/10.1118/1.3590358
53. Kukreja S, Haydel J, Nanda A, Sin AH (2015) Impact of body habitus on fluoroscopic radiation emission during minimally invasive spine surgery: Presented at the 2014 AANS/CNS Joint Section on Disorders of the Spine and Peripheral Nerves. J Neurosurg Spine 22:211–218. https://doi.org/10.3171/2014.10.SPINE14163
54. Dolenc L, Petrinjak B, Mekiš N, Škrk D (2022) The impact of body mass index on patient radiation dose in general radiography. Journal of Radiological Protection 42:. https://doi.org/10.1088/1361-6498/ac9f1f
55. Madder RD, VanOosterhout S, Mulder A, et al (2019) Patient body mass index and physician radiation dose during coronary angiography: Is the obesity epidemic impacting the occupational risk of physicians in the catheterization laboratory? Circ Cardiovasc Interv 12
56. Shah A, Das P, Subkovas E, et al (2015) Radiation Dose During Coronary Angiogram: Relation to Body Mass Index. Heart Lung Circ 24:21–25. https://doi.org/10.1016/j.hlc.2014.05.018

57. Liu M, Mei K, Liu X, Zhao Y (2020) Impact of body mass index on procedural complications, procedure duration, and radiation dose in patients with atrial fibrillation undergoing radiofrequency ablation: A systematic review and meta-analysis. Clin Cardiol 43:1067–1075

58. Andrade C (2019) The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. Indian J Psychol Med 41:210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19

59. Schmidt J-S, Osebold R (2017) Environmental Management Systems as A Driver for Sustainability: State Of Implementation, Benefits And Barriers In German Construction Companies. Journal Of Civil Engineering and Management 23:150–162. Https://Doi.Org/10.3846/13923730.2014.946441

60. Kaliappan J, Srinivasan K, Mian Qaisar S, et al (2021) Performance Evaluation of Regression Models for the Prediction of the COVID-19 Reproduction Rate. Front Public Health 9:. https://doi.org/10.3389/fpubh.2021.729795