RESEARCH ARTICLE

Editorial Process: Submission:09/09/2024 Acceptance:01/22/2025

Analyzing Secondary Cancer Risk: A Machine Learning Approach

Erfan Hatamabadi Farahani[1], Hossein Sadeghi[1], Fatemeh Seif[2]*, Mahdi Azad Marzabadi[1], Reza Rezaee[1]

Abstract

Objective: Addressing the rising cancer rates through timely diagnosis and treatment is crucial. Additionally, cancer survivors need to understand the potential risk of developing secondary cancer (SC), which can be influenced by several factors including treatment modalities, lifestyle choices, and habits such as smoking and alcohol consumption. This study aims to establish a novel relationship using linear regression models between dose and the risk of SC, comparing different prediction methods for lung, colon, and breast cancer. **Methods:** Machine learning (ML) models have demonstrated their usefulness in forecasting the likelihood of SC risks based on effective doses in the organ. Linear regression analysis is a widely utilized technique for examining the relationship between predictor variables and continuous responses, particularly in scenarios with limited sample sizes. This study employs linear regression models to analyze the relationship between effective dose and the risk of SC, comparing different prediction methods across lung, colon, and breast cancer. **Result:** The results indicate that the risk of SC increases with the effective dose in the organ. The linear regression model provides coefficients that mirror the radiation sensitivity of the specific organ, demonstrating the model's effectiveness in predicting SC risk based on dose. **Conclusion:** The study highlights the significance of using linear regression models to predict the risk of SC based on effective doses in the organ. The findings underscore the importance of considering the radiation sensitivity of specific organs in SC risk prediction, which can aid in better understanding and managing the long-term health of cancer survivors.

Keywords: Machine learning- Radiation dosage- Precision medicine- Decision trees

Asian Pac J Cancer Prev, **26 (1)**, 239-248

Introduction

Today's society is facing a significant challenge with the rising number of individuals diagnosed with cancer. The uncontrolled proliferation of malignant cells can result in the development of cancer. However, advancements in medical science have enabled the timely diagnosis and treatment of cancer, to minimize mortality rates associated with the disease. While some cancer survivors may remain disease-free following initial treatment, others may experience non-cancer-related health issues and side effects from the treatment [1, 2]. A major concern for individuals who have undergone cancer treatment is the possibility of cancer recurrence. All cancer survivors need to be aware of the potential for developing SC following treatment for the initial cancer. This SC is distinct from the primary cancer in terms of its origin and pathology [3, 4]. It can manifest in the same organ or area of the body as the initial cancer, or in a completely different organ. It is crucial to understand that SC is not indicative of metastasis from the primary cancer [5]. Factors such as treatment methods, smoking, alcohol consumption, and overall lifestyle can contribute to the development of SC. The efficacy of the treatment method plays a crucial role in the development of SC [6, 7].

Radiation therapy, which involves the use of ionizing radiation to induce breaks in DNA, effectively halts the growth of cancer cells and leads to their destruction. By targeting cells that exhibit uncontrolled proliferation, this method results in the eradication of the disease [8, 9]. However, the relationship between radiation dosage and the likelihood of cancer must be considered, as this treatment approach can potentially give rise to SC. The scattering of radiation during therapy may impact healthy organs in the body, leading to damage. Organs such as the thyroid and breast, which are particularly sensitive to radiation, are at a higher risk of developing SC. Therefore, it is imperative to assess the risk of SC considering these factors [10, 11].

There are various approaches to assessing the risk of SC. One of the primary methods involves the use of computational models to calculate the excess relative risk

*[1]Department of Physics, Faculty of Sciences, Arak University, Arak, Iran. [2]Department of Radiotherapy and Medical Physics, Arak University of Medical Sciences and Khansari Hospital, Arak, Iran. *For Correspondence: S.Medphy@gmail.com*

(ERR) and the absolute excess risk (EAR). Additionally, nuclear simulator codes are utilized to determine the effective dose in the organ, which is then used to calculate the risk of SC. Another method involves cohort studies, where databases and patient data are used to estimate the risk of SC among individuals who have undergone primary cancer treatment. This method typically involves studying a significant population within a specific region to ensure the accuracy of the results [12, 13].

Considering the progress of technology and the utilization of artificial intelligence, particularly ML models in various medical fields, the prediction of SC risk is among the valuable applications of ML models in the oncology domain [14]. The ML approach is centered on data and its continual enhancement, relying on statistics and probability. However, the outcomes derived from this approach surpass those of statistical methods [15]. ML encompasses diverse models, with the accuracy of results contingent upon the specific models employed. Linear regression analysis, a straightforward and widely used technique for assessing relationships between predictor variables and a continuous response, assumes linearity in the relationships between predictor and target variables. This implies that a consistent unit change in one variable corresponds to a consistent unit change in the other variable. Linear regression is often the preferred option for analyses involving small sample sizes, as these models are straightforward to interpret. Based on the findings related to SC risk and possessing knowledge about the effective dose, the correlation between the effective dose in the organ and the risk of SC can be computed.

Recently, the integration of ML models, particularly decision trees, into the research methodology has led to the development of a practical framework for predicting the incidence of SC using patient data. This framework facilitates the classification of patients into high-risk and low-risk categories, thus supporting the formulation of personalized treatment strategies and interventions. Furthermore, it highlights several factors influencing the probability of SC, including radiation exposure, patient age, and genetic factors, while also pointing out the shortcomings of existing models in accounting for all pertinent variables [16].

The aim of this study was the utilization of ML models to analyze past research data to calculate the risk of SC. The primary objective of this research is to determine the correlation between dosage and the likelihood of developing SC, with a specific focus on establishing a relationship using a linear regression model.

Materials and Methods

The research involving human participants received approval from the Ethics Committee at Arak University of Medical Sciences. This study was conducted by the regulations established by the local authorities and institutional standards. This research was conducted to obtain the relationship between the dose and the risk of SC, and its working method includes two steps: ML and regression model, which are fully explained below.

Data Collection and Study Populations

The analysis encompasses a comprehensive review of 21 experimental and computational studies focused on radiotherapy patients, drawing from a larger dataset of 65 studies conducted between 1980 and 2000. These studies investigated various types of secondary cancers (SCs), identifying 23 distinct SC types, with a notable concentration on secondary breast cancer (SBC) and stomach cancer. The compiled dataset, which includes 113 studies, provides critical insights into SC incidence and mortality rates, and demographic factors such as the percentage of female participants, the duration of radiation exposure, follow-up ages, and average radiation doses. This dataset is particularly valuable for training ML models, consisting of 113 instances categorized into incidence and mortality classes.

The geographical distribution of the studies reveals a predominance of research from the United States, followed by contributions from several other countries, including Sweden, Israel, and the United Kingdom. The methodologies employed in these studies varied significantly, with a majority being population-based. The total participant count across all studies reached 371,992, highlighting the extensive nature of the research.

Key findings emphasize the importance of linking SC risk with factors such as radiation dose, age, and sex at exposure. Evidence suggests that younger individuals exposed to radiation face a heightened risk of developing certain cancers, such as lung cancer among underground miners and primary hypothyroidism in childhood cancer survivors. Additionally, the analysis indicates that cumulative radiation doses have varied over time, with men generally receiving higher doses than women.

The study underscores the necessity of utilizing past research to inform current investigations into radiation exposure and its health implications. The adoption of frameworks like TRIPOD and TRIPOD-AI aims to enhance the transparency and quality of reporting in prediction model studies, particularly as artificial intelligence becomes increasingly integrated into health research. By leveraging data from previous studies, the research seeks to apply AI and ML techniques to better predict SC risk, ultimately contributing to improved health outcomes and reduced research inefficiencies.

Machine learning

Predicting the risk of SC using ML methods is an important and active research field in the field of oncology and medical sciences. This method uses training of its algorithms on patient data to check the risk of SC. Due to the significant importance of data in using the ML method and the information that this data provides us, the data sets were selected with high sensitivity. Table 1 shows details about the data used in this research.

Algorithm Description

The ML method consists of several steps, and we schematically present the steps taken by this method to predict the risk of SC in Figure 1) Start by defining the problem that needs to be solved using ML methods. 2) Gather and preprocess the data needed for the

Table 1(a). Information about Secondary Lung Cancer Data Details

Cancer Site	Lung		
Publication	Van Leeuwen et al. (1995) [17]	Mattsson et al. (1997) [18]	Davis et al. (1989) [19]
all case	1939	1216	13385
cases/death	30	19	69
women in study%	41	100	48.7
Age at exposure (year)	<45 - >55	8-74	<24 -> 38
Follow-up (year)	1->10	5-61	0- 50
Average dose (Sv)	7.2	0.75	0.84
Dose Range (Sv)	0->21	0-8.98	0->8

Table 1(b). Information about Secondary Colon Cancer Data Details

Cancer Site	Colon		
Publication	Inskip et al. (1990) [20]	Darby et al. (1995) [21]	Weiss et al. (1994) [22]
all case	4153	2067	14109
cases/death	73	47	226
women in study%	100	100	17.8
Age at exposure(year)	13 - 88	23 - 65	<25 - >55
Follow up(year)	0 - 60	5 _ 49	5 - >35
Average dose (Sv)	1.2	3.2	4.1
Dose Range (Sv)	<0.6 - 6.65	<2.41 - >3.73	0 - >7.85

Table 1(c). Information about Secondary Breast Cancer Data Details

Cancer Site	Breast	
Publication	Boice et al. (1989) [23]	Hildreth et al. (1989) [24]
all case	12040	1201
cases/death	140	34
women in study%	100	100
Age at exposure(year)	<30 ->75	<1
Follow up(year)	5 ->40	0 ->52
Average dose (Sv)	0.31	0.69
Dose Range (Sv)	0 - 0.98	0.01 -7.1

algorithm. 3) Choose the appropriate ML method based on the type of problem and data. 4) Split the data into training and testing sets. Train the model on the training set using the chosen algorithm. 5) Use cross-validation

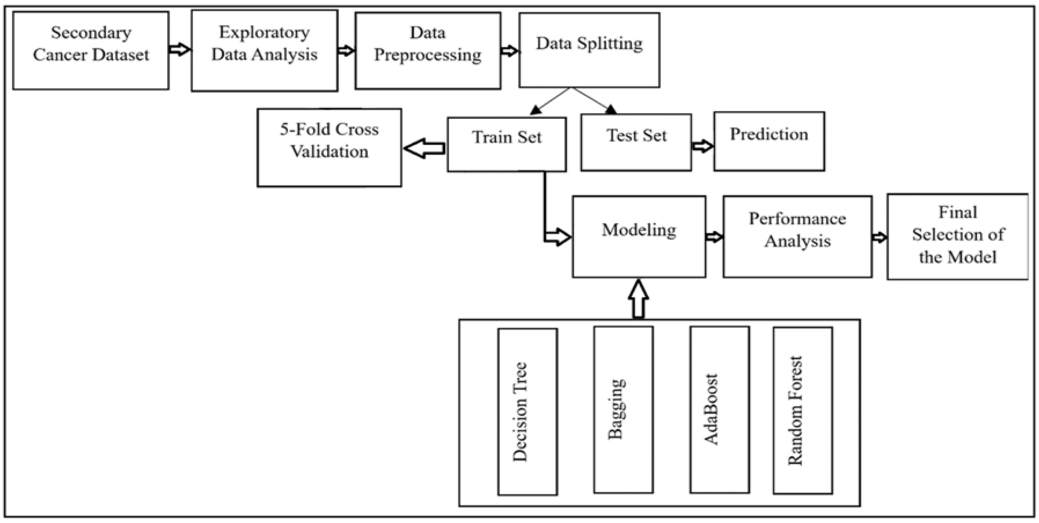

Figure 1. The Figure Depicts the Overall Workflow Diagram

to assess the performance of the model. 6) Fine-tune the model parameters to improve its performance. Evaluate the model on the testing set to determine its generalization ability. 7) If the model's performance is satisfactory, deploy it for real-world use. 8) If not, repeat steps 5-7 with different parameter settings or change the chosen algorithm. 9) Continuously monitor and evaluate the model's performance and make necessary improvements.10) Use the trained model to make predictions on new, unseen data. Finally, document the algorithm and its results for future reference.

As mentioned, the ML method is data-oriented, and you can see in Figure 1 that the most basic steps of this method are the selection of data to train the algorithm on them. The dataset used in this research is based on work done in the past decades and collected by databases and includes information such as gender, radiation dose, and age of radiation exposure.

One of the most essential steps in the data mining process, which has a significant impact on the selection of models for prediction and conclusions, is data processing and the relationship between them, which is examined in Figure 2. After this step, the data is ready for analysis. The development of intelligent predictive models aimed at health outcomes necessitates meticulous attention to data collection, preprocessing, and the selection of relevant features. Techniques for feature selection, such as optimization-based methods and swarm intelligence algorithms, are crucial for minimizing dimensionality and clarifying intricate causal relationships, especially within healthcare contexts. Machine learning algorithms are commonly utilized for tasks involving classification and prediction. A well-structured study design, along with appropriate data manipulation and evaluation techniques, constitutes fundamental aspects of creating and validating predictive models.

This method examines a part of the data set for training and after this step examines the remaining part of the data set for testing and obtaining results. In this research, the training dataset consists of 70% of the data, while the testing dataset comprises the remaining 30%.

In this research, to choose the best model for predicting the risk of SC, we have examined four models: decision tree, random forest, bagging, and AdaBoost. One of the tests that helped us choose the best model is the calculation of the AUC and ROC curve, the minimum value of AUC is zero and the maximum value is one, and the closer this number is to one, it means that the model has strong power. Predictability is therefore very satisfactory if it is in the range of (0.8-0.9) and excellent if it is (0.9-1).

Machine learning methods

Decision Tree (DT)

One of the best classification algorithms is DT, which has features such as interpretability, analysis, and simplicity. In this method, a tree structure should be used, which has special rules for the collective implementation of classification processes, in the tree structure, there are three important parts internal nodes, branches, and leaf nodes, which respectively indicate the characteristics, values of the characteristics and the classes that exist in the data set. The internal node that produces the output is called a branch and can be the input of another internal node [25]. Figure 3 shows the results of the AUC and ROC curve for the Decision Tree model.

Bagging

This technique generates final predictions using a random selection of subsets of the data. Breiman introduced the concept of bagging, also referred to as bootstrap aggregation [26]. Figure 4 shows the results of the AUC and ROC curve for the Bagging model.

AdaBoost

Adaboost was first introduced as a classification algorithm in 1997 by Freund and Schapire. For training, this method first creates a decision tree in which the data has equal weight at each point, then uses the appropriate model to classify the training set. If this model correctly predicts the weight of the data, it remains unchanged, and if this diagnosis is wrong, the weight of the samples is changed, and after creating a balance between the

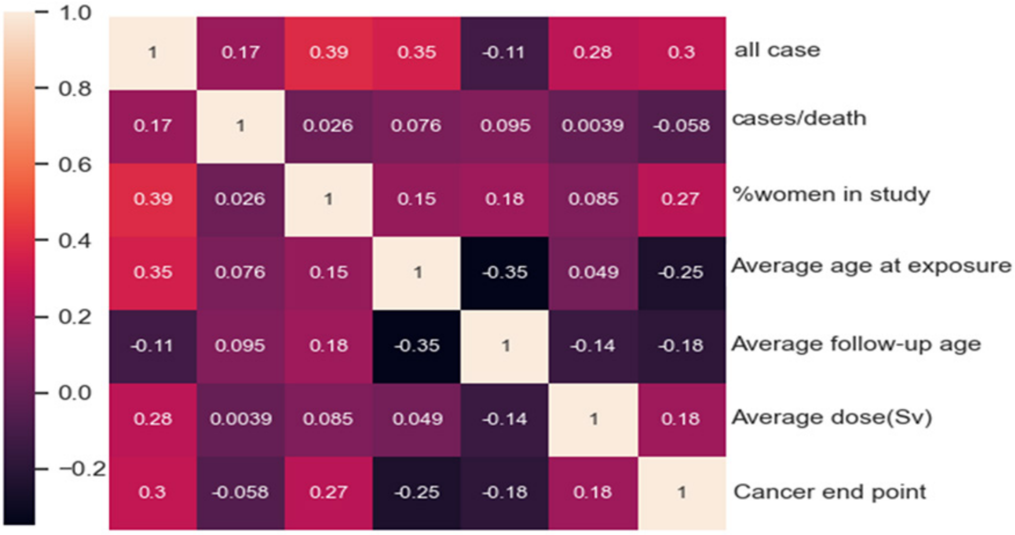

Figure 2. Overall Workflow Diagram

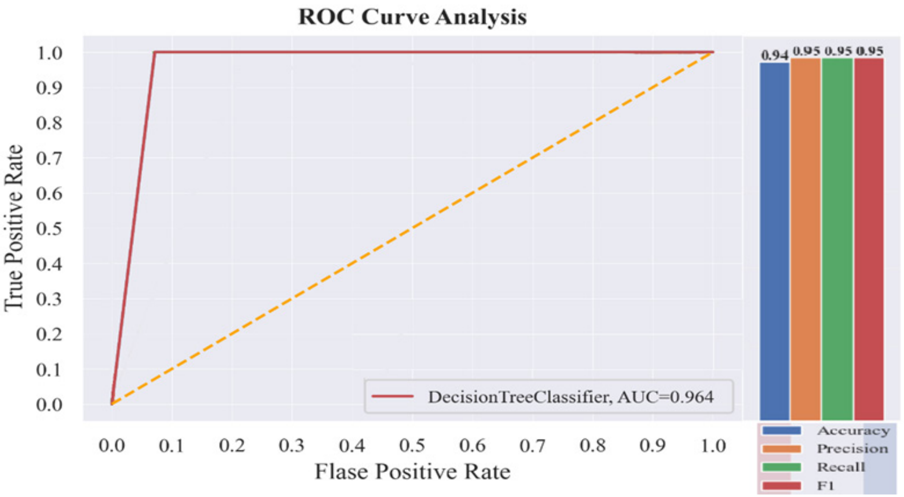

Figure 3. AUC and ROC Curve for the Decision Tree Model

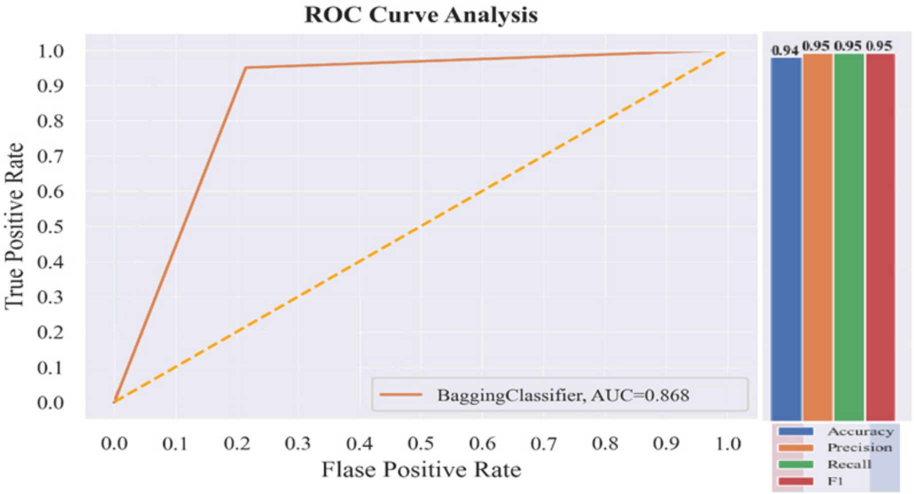

Figure 4. AUC and ROC Curve for the Bagging Model

weights (normalization), a new decision tree is created. This process is repeated until the ideal conditions are reached [27]. Figure 5 shows the results of the AUC and ROC curves for the AdaBoost model.

Random Forest (Rf)

In 1995, Hu introduced the RF model with the idea of taking stochastic methods that include sub-decision trees that function as an ensemble learning classification

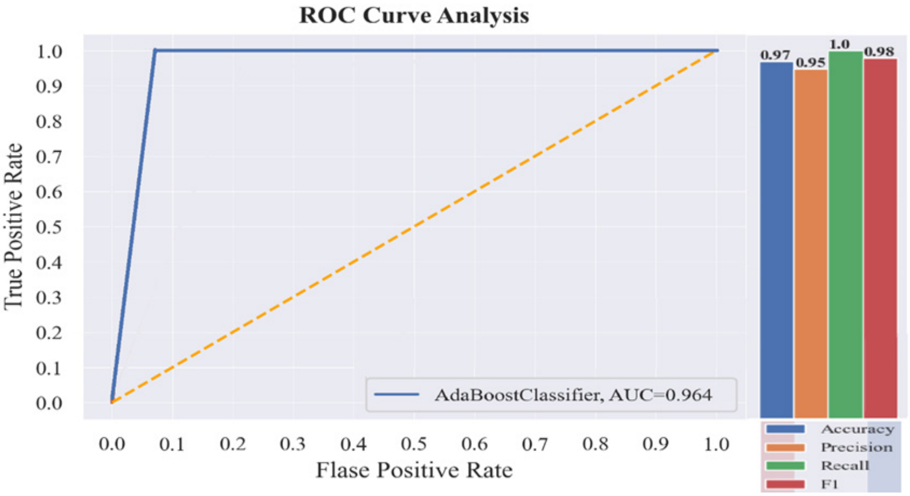

Figure 5. AUC and ROC Curve for the AdaBoost Model

algorithm. Finally, this method has a higher prediction accuracy than the methods that use a single decision tree [28]. Figure 6 shows the results of the AUC and ROC curve for the Random Forest model.

Regression Model

Linear regression is a statistical technique used to estimate the linear association between a single response variable and one or more explanatory variables, which are also referred to as dependent and independent variables. When the model involves only one explanatory variable, it is known as simple linear regression [29].

This research, for obtaining the relationship between dose and SC risk has been done by linear regression method, so that, first, the risks obtained by ML are placed in one list and the available doses from previous research are placed in another list. After creating these two lists, we begin by creating a linear regression model in our Python program. We classify the values in these lists as independent variables (in this case, SC risk) and dependent variables (dose) and input them into the model to determine the relationship between these variables.

The linear regression model works by initially creating a default first-order linear equation.

$$Y= Ax + B \quad (1)$$

After this step, the program considers the slope of the line (A) to be 1 and the distance from the origin (B) to be 0. It then predicts the relationship between risk and dose as a line and calculates the standard deviation for the values. After calculating the standard deviation, the program again predicts a new line equation by changing the slope and width from the origin and calculates the standard deviation for the new values obtained, if the new standard deviation is smaller than before, it means that this new equation It is more optimal and suitable than the previous equation and until the lowest standard deviation is obtained, the program automatically changes the values of A and B and finally predicts the most optimal and most suitable equation for this relationship and finally the line equation can be drawn and compared with other experimental values.

We used the values obtained for the risk of secondary breast, colon, and lung cancer by ML model and considering that these values were obtained using the available data from previous research and the average dose was also available in these data. We obtained the relationship between the risk of SC and the dose using the linear regression method, and after drawing the line, we compared this relationship with other studies.

Results

The findings obtained from the ML method are displayed in Table 2 and are further enhanced by integrating results from other computational and simulation methods to enable a comprehensive comparison. Donovan et al. conducted an experimental study in 2012, utilizing thermoluminescence dosimeters (TLD) to measure the effective dose in the organ using a phantom. This study encompassed various radiotherapy techniques, including whole breast radiotherapy (WBRT), partial breast irradiation (APBI), and simultaneous integrated enhancement (SIB) with two and three-volume models. Subsequently, the risk of SC was determined using computational models (BEIR VII) [30]. Mendes et al. utilized the MCNP code to calculate the risk of SC, employing a virtual phantom known as the VW phantom, which represented a woman with 63 organs, a height of 165 cm, and a weight of 98 kg [13]. The absorbed dose in each organ after radiotherapy was calculated using the MCNP code, considering a parallel field of 6 Mv as the radiation source.

The risk of SC was then determined using the BEIR VII computational model. The linear regression model outcomes concerning the correlation between dosage and the risk of SC have been visually represented through graphs displayed in Figures 7 to 9, focusing on SC occurrences in the breast, colon, and lung. The importance of genetic and hormonal factors is immense, as they have not been previously explored through machine learning techniques. The likelihood of health effects caused by radiation exposure varies based on the age

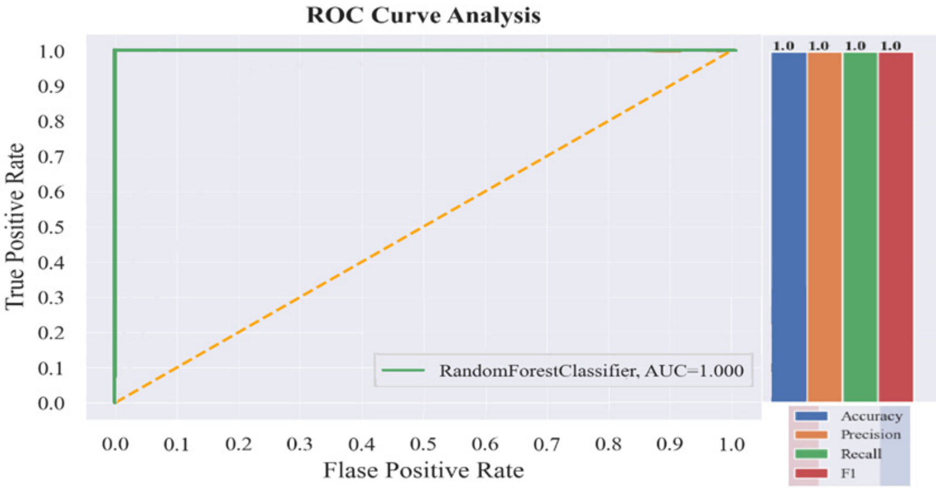

Figure 6. AUC and ROC Curve for the Random Forest Model

Table 2(a). Comparison of Different Methods for Predicting the Risk of Second Cancer in the Lung

Second Cancer	Publishers	Method	Average Dose(sv)	Second Cancer Risk (%)
Lung	Donovan et al. (2012) [30]	WBRT	0.68	0.12
	Donovan et al. (2012) [30]	APBI	0.7	0.07
	Donovan et al. (2012) [30]	SIB 2 volume	1.9	1.11
	Donovan et al. (2012) [30]	SIB 3 volume FP IMRT	0.27	0.3
	Donovan et al. (2012) [30]	SIB 3 volume IP IMRT	1.8	0.68
	Mendes et al. (2017) [13]	MCNP	0.22	0.38
	This work (2024)	ML	7.2	0.77
	This work (2024)	ML	0.75	0.59
	This work (2024)	ML	0.84	0.44

Table 2(b). Comparison of Different Methods for Predicting the Risk of Second Cancer in the Colon

Second Cancer	Publishers	Method	Average Dose(sv)	Second Cancer Risk (%)
Colon	Donovan et al. (2012) [30]	WBRT	0.08	0.06
	Donovan et al. (2012) [30]	APBI	0.07	0.05
	Donovan et al. (2012) [30]	SIB 2 volume	0.15	0.09
	Donovan et al. (2012) [30]	SIB 3 volume FP IMRT	0.21	0.14
	Donovan et al. (2012) [30]	SIB 3 volume IP IMRT	0.11	0.06
	Mendes et al. (2017) [13]	MCNP	0.06	0.03
	This work (2024)	ML	1.2	0.38
	This work (2024)	ML	3.2	0.41
	This work (2024)	ML	4.1	0.41

Table 2(c). Comparison of Different Methods for Predicting the Risk of Second Cancer in the Breast

Second Cancer	Publishers	Method	Average Dose(sv)	Second Cancer Risk (%)
Breast	Donovan et al. (2012) [30]	WBRT	0.6	0.6
	Donovan et al. (2012) [30]	APBI	0.19	0.18
	Donovan et al. (2012) [30]	SIB 2 volume	0.74	0.69
	Donovan et al. (2012) [30]	SIB 3 volume FP IMRT	0.43	0.61
	Donovan et al. (2012) [30]	SIB 3 volume IP IMRT	1.17	1.1
	Mendes et al. (2017) [13]	MCNP	0.27	0.55
	This work (2024)	ML	0.31	0.59
	This work (2024)	ML	0.69	0.7

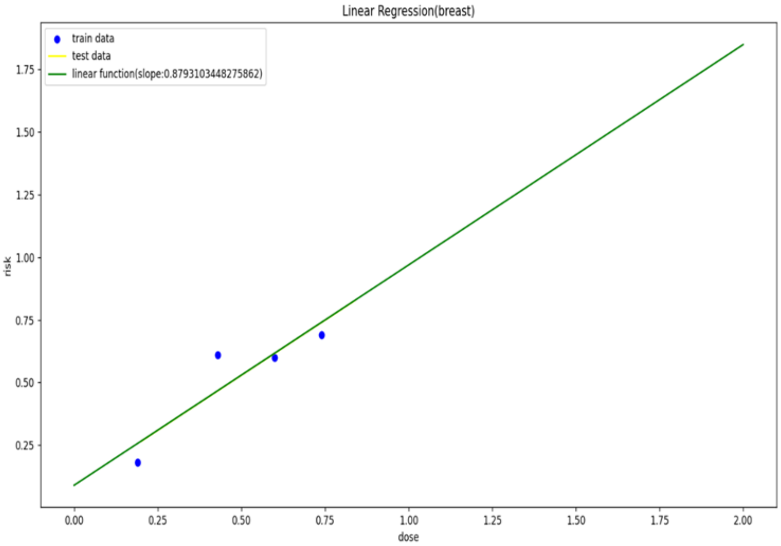

Figure 7. Results of Linear Regression Model for Secondary Breast Cancer

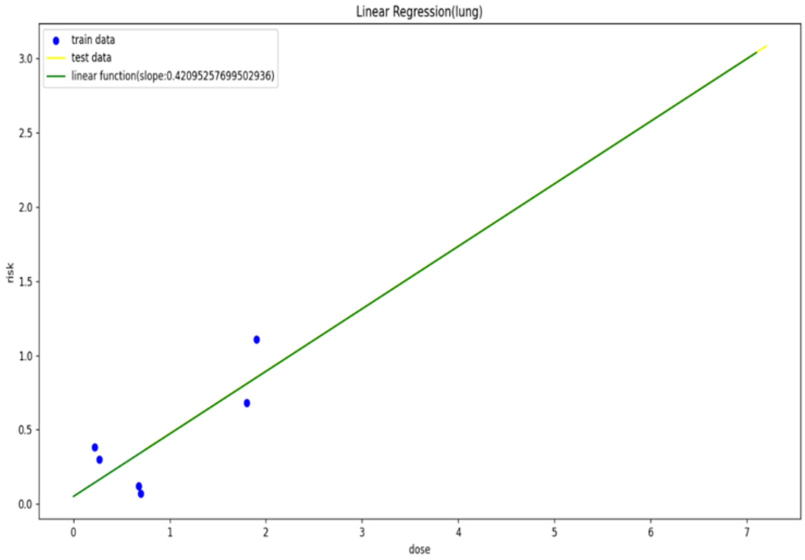

Figure 8. Results of Linear Regression Model for Secondary Lung Cancer

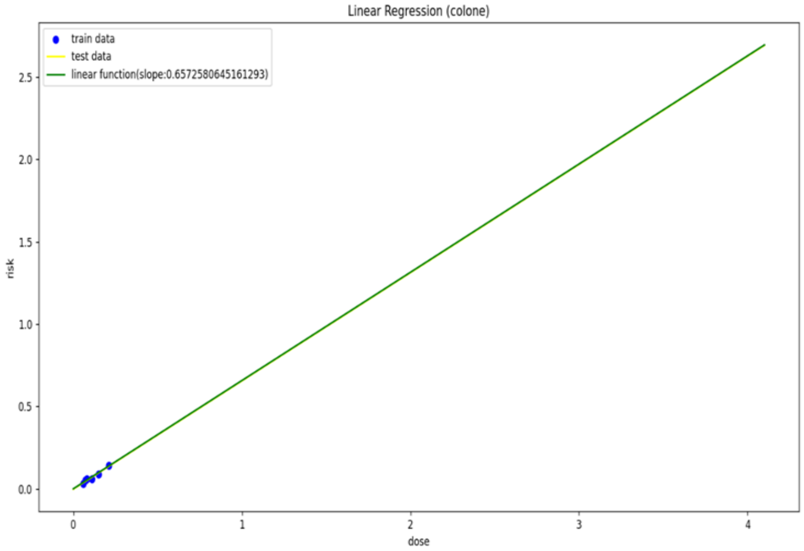

Figure 9. Results of the Linear Regression Model by Python Program for Secondary Colon Cancer

at which exposure occurs, gender, and the specific year. Research on underground miners has shown that younger individuals at the time of exposure have a higher risk of developing lung cancer (Tomášek, 2014). Similarly, among childhood cancer survivors, the risk of developing primary hypothyroidism after radiation therapy is greater for females and those exposed at ages over 15 years. The cumulative red bone marrow dose from diagnostic radiation exposure is influenced by the calendar year, with peaks around 1950 and post-1980, and is generally higher in men compared to women.

Based on the graphs provided, the outcomes of the regression model demonstrate a satisfactory level of concordance with the outcomes of the computational models. It is evident that, in line with expectations, the likelihood of developing SC escalates with the augmentation of the effective dose in the organ. The model recommended by the BEIR committee has outlined the definitions of excess relative risk (ERR) and excess absolute risk (EAR) is articulated as follows:

$$\textbf{ERR and EAR} = \beta\textbf{SD exp}\,(\gamma\ \textbf{e}*)\ \left(\frac{a}{60}\right)^{\eta}, \tag{2}$$

In the equation provided, D represents the dose administered, while βS, γ, and η are parameters specific to excess relative risk (ERR) and excess absolute risk (EAR) for different organs based on sex. The variable e* denotes the age at exposure, and a represents the attained age. Focusing solely on the linear component of Equation (2), which pertains to the correlation between dose and risk, the equation can be reformulated as [30].

$$\text{ERR and EAR} = \beta\text{SD}. \tag{3}$$

The linear regression model has yielded a coefficient (A) from Eq. (1) that represents the slope of the graphs depicting the relationship between dose and the risk of SC. Utilizing this model allows for the calculation of the βS coefficient for SC. Furthermore, the slope of the regression line for SBC is 0.879, while for secondary lung cancer, it is 0.420. The slope for secondary colon cancer is 0.657, indicating a potential correlation with the radiation sensitivity of the respective organs.

Discussion

It is crucial to tackle the rise in cancer incidences within the community by promptly diagnosing and treating the disease. Furthermore, it underscores the significance of cancer survivors being mindful of the potential risk of developing SC, which may be impacted by a range of factors including treatment modalities, lifestyle decisions, and behaviors such as smoking and alcohol intake. The use of ML models in predicting the risk of SC based on effective doses in the organ is an effective application in the field of oncology. Linear regression analysis is a popular method for measuring the relationship between predictor variables and continuous response and is often the best choice for analyses with small sample sizes. These results highlight the critical need to account for age, gender, and temporal variables when evaluating radiation-related health risks in epidemiological research. Our goal is to assemble a more comprehensive dataset that includes these factors and to perform modeling analyses based on this data. We will integrate this information into the Introduction, Results, and Conclusions sections of the manuscript to enable comparisons and emphasize its importance. The research aims to establish a new relationship using the linear regression model between the dose and the risk of SC. We compare different methods for predicting the risk of SC in the lung, colon, and breast. The results indicate that the risk of SC increases with the effective dose in the organ, and the linear regression model provides coefficients that are related to the radiation sensitivity of the specific organ.

Author Contribution Statement

Erfan Hatamabadi Farahani: Methodology, Investigation, Data curation, Writing – review & editing, Conceptualization, Validation, Software, Methodology. Hossein Sadeghi: Methodology, Investigation, Data curation, Writing – review & editing, Supervision. Fatemeh Seif: Investigation, Data curation, Conceptualization, Writing– review & editing, Supervision. Mahdi Azad: Conceptualization, Validation, Software, Methodology. Reza Rezaee: Investigation, Data curation, Conceptualization..

Acknowledgements

General

The ML approach to forecasting the risk of SC was executed using data and tools obtained from various research groups. We express our gratitude to the international modeling and experimental groups for generously sharing their data for analysis. The authors thank the Arak University of Medical Sciences (project no. 4529). The successful completion of this research is attributed to the patients who provided their information and the continuous efforts of organizations and scientists in advancing this valuable community resource.

Scientific Approval

The studies involving humans were approved by the Ethics Committee of Arak University of Medical Sciences (Approval number IR.ARAKMU.REC.1403.158).

Ethical Declaration

The studies involving humans were approved by the Ethics Committee of Arak University of Medical Sciences (Approval number IR.ARAKMU.REC.1403.158). The research was conducted in compliance with the regulations of the local district and the standards set by the institution. Due to the retrospective nature of the study, the ethics committee or institutional review board decided to exempt the need for written informed consent from the participants or their legal guardians/next of kin.

Data Availability

All data supporting the findings of this study are available within the paper. Source code and Datasets are available from: https://github.com/Sorayakhan/khanmohammadi/blob/main/Hatam-FINAL-Paper-GitHub.ipynb.

Study Registration

This study is not registered in any dataset.

References

1. Devita vt. Devita, hellman, and rosenberg's cancer: Principles & practice of oncology. Lippincott williams & wilkins; 2008.18: 581-92.
2. Feller A, Matthes KL, Bordoni A, Bouchardy C, Bulliard JL, Herrmann C, et al. The relative risk of second primary cancers in switzerland: A population-based retrospective cohort study. BMC Cancer. 2020;20(1):51. https://doi.org/10.1186/s12885-019-6452-0.
3. Dasu A, Toma-Dasu I. Models for the risk of secondary cancers from radiation therapy. Phys Med. 2017;42:232-8. https://doi.org/10.1016/j.ejmp.2017.02.015.
4. Hall EJ. Intensity-modulated radiation therapy, protons, and the risk of second cancers. Int J Radiat Oncol Biol Phys. 2006;65(1):1-7. https://doi.org/10.1016/j.ijrobp.2006.01.027.
5. Davis RH. Production and killing of second cancer precursor cells in radiation therapy: In regard to hall and wuu (int j radiat oncol biol phys 2003;56:83-88). Int J Radiat Oncol Biol Phys. 2004;59(3):916. https://doi.org/10.1016/j.ijrobp.2003.09.076.
6. Mertens AC, Liu Q, Neglia JP, Wasilewski K, Leisenring W, Armstrong GT, et al. Cause-specific late mortality among 5-year survivors of childhood cancer: The childhood cancer survivor study. J Natl Cancer Inst. 2008;100(19):1368-79. https://doi.org/10.1093/jnci/djn310.
7. Yerramilli D, Xu AJ, Gillespie EF, Shepherd AF, Beal K, Gomez D, et al. Palliative radiation therapy for

oncologic emergencies in the setting of covid-19: Approaches to balancing risks and benefits. Adv Radiat Oncol. 2020;5(4):589-94. https://doi.org/10.1016/j.adro.2020.04.001.

8. Rades D, Stalpers LJ, Veninga T, Schulte R, Hoskin PJ, Obralic N, et al. Evaluation of five radiation schedules and prognostic factors for metastatic spinal cord compression. J Clin Oncol. 2005;23(15):3366-75. https://doi.org/10.1200/jco.2005.04.754.
9. Mullenders L, Atkinson M, Paretzke H, Sabatier L, Bouffler S. Assessing cancer risks of low-dose radiation. Nat Rev Cancer. 2009;9(8):596-604. https://doi.org/10.1038/nrc2677.
10. Rades D, Panzner A, Rudat V, Karstens JH, Schild SE. Dose escalation of radiotherapy for metastatic spinal cord compression (mscc) in patients with relatively favorable survival prognosis. Strahlenther Onkol. 2011;187(11):729-35. https://doi.org/10.1007/s00066-011-2266-y.
11. Berrington de Gonzalez A, Gilbert E, Curtis R, Inskip P, Kleinerman R, Morton L, et al. Second solid cancers after radiation therapy: A systematic review of the epidemiologic studies of the radiation dose-response relationship. Int J Radiat Oncol Biol Phys. 2013;86(2):224-33. https://doi.org/10.1016/j.ijrobp.2012.09.001.
12. Doudoo CO, Gyekye PK, Emi-Reynolds G, Adu S, Kpeglo DO, Nii Adu Tagoe S, et al. Dose and secondary cancer-risk estimation of patients undergoing high dose rate intracavitary gynaecological brachytherapy. J Med Imaging Radiat Sci. 2023;54(2):335-42. https://doi.org/10.1016/j.jmir.2023.03.031.
13. Mendes BM, Trindade BM, Fonseca TCF, de Campos TPR. Assessment of radiation-induced secondary cancer risk in the brazilian population from left-sided breast-3d-crt using mcnpx. Br J Radiol. 2017;90(1080):20170187. https://doi.org/10.1259/bjr.20170187.
14. Debnath S, Barnaby DP, Coppa K, Makhnevich A, Kim EJ, Chatterjee S, et al. Machine learning to assist clinical decision-making during the covid-19 pandemic. Bioelectron Med. 2020;6:14. https://doi.org/10.1186/s42234-020-00050-8.
15. Syleouni ME, Karavasiloglou N, Manduchi L, Wanner M, Korol D, Ortelli L, et al. Predicting second breast cancer among women with primary breast cancer using machine learning algorithms, a population-based observational study. Int J Cancer. 2023;153(5):932-41. https://doi.org/10.1002/ijc.34568.
16. Sadeghi H, Seif F, Farahani EH, Khanmohammadi S, Nahidinezhad S. Utilizing patient data: A tutorial on predicting second cancer with machine learning models. Cancer Med. 2024;13(18):e70231. https://doi.org/10.1002/cam4.70231.
17. van Leeuwen FE, Klokman WJ, Stovall M, Hagenbeek A, van den Belt-Dusebout AW, Noyon R, et al. Roles of radiotherapy and smoking in lung cancer following hodgkin's disease. J Natl Cancer Inst. 1995;87(20):1530-7. https://doi.org/10.1093/jnci/87.20.1530.
18. Mattsson A, Hall P, Rudén BI, Rutqvist LE. Incidence of primary malignancies other than breast cancer among women treated with radiation therapy for benign breast disease. Radiat Res. 1997;148(2):152-60.
19. Davis FG, Boice JD, Jr., Hrubec Z, Monson RR. Cancer mortality in a radiation-exposed cohort of massachusetts tuberculosis patients. Cancer Res. 1989;49(21):6130-6.
20. Inskip PD, Monson RR, Wagoner JK, Stovall M, Davis FG, Kleinerman RA, et al. Cancer mortality following radium treatment for uterine bleeding. Radiat Res. 1990;123(3):331-44. https://doi.org/10.2307/3577741
21. Darby SC, Reeves G, Key T, Doll R, Stovall M. Mortality in a cohort of women given x-ray therapy for metropathia haemorrhagica. Int J Cancer. 1994;56(6):793-801. https://doi.org/10.1002/ijc.2910560606.
22. Weiss HA, Darby SC, Doll R. Cancer mortality following x-ray treatment for ankylosing spondylitis. Int J Cancer. 1994;59(3):327-38. https://doi.org/10.1002/ijc.2910590307.
23. Boice JD Jr, Engholm G, Kleinerman RA, Blettner M, Stovall M, Lisco H, et al. Radiation dose and second cancer risk in patients treated for cancer of the cervix. Radiat Res. 1988;116(1):3-55.
24. Hildreth NG, Shore RE, Dvoretsky PM. The risk of breast cancer after irradiation of the thymus in infancy. N Engl J Med. 1989;321(19):1281-4. https://doi.org/10.1056/nejm198911093211901.
25. Wu Y, Ke Y, Chen Z, Liang S, Zhao H, Hong H. Application of alternating decision tree with adaboost and bagging ensembles for landslide susceptibility mapping. CATENA. 2020;187:104396. https://doi.org/https://doi.org/10.1016/j.catena.2019.104396.
26. Alshahrani SM, Fahem Albaghdadi M, Yasmin S, Alosaimi ME, Alsalhi A, Algarni M, et al. Green processing based on supercritical carbon dioxide for preparation of nanomedicine: Model development using machine learning and experimental validation. Case Stud Therm. 2023;41:102620. https://doi.org/https://doi.org/10.1016/j.csite.2022.102620.
27. Mosavi A, Sajedi Hosseini F, Goodarzi M, Dineva A, Rafiei Sardooi E. Ensemble boosting and bagging based machine learning models for groundwater potential prediction. Water Resour Manga. 2021;35:1-15. https://doi.org/10.1007/s11269-020-02704-3.
28. Alsagri H, Ykhlef M. Quantifying feature importance for detecting depression using random forest. Int J Adv Comput Sci Appl. 2020;11. https://doi.org/10.14569/IJACSA.2020.0110577.
29. Freedman da. A simple regression equation has on the right-hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression e right hand side, each with its own slope coefficient. Statistical models: Theory and practice. 2009:26; 16: 581-92.
30. Donovan EM, James H, Bonora M, Yarnold JR, Evans PM. Second cancer incidence risk estimates using beir vii models for standard and complex external beam radiotherapy for early breast cancer. Med Phys. 2012;39(10):5814-24. https://doi.org/10.1118/1.4748332.