Original Article

Radiomics Analysis on Computed Tomography Images for Prediction of Chemoradiation-induced Heart Failure in Breast Cancer by Machine Learning Models

Farzaneh Ansari[1], Ali Neshasteh-Riz[1,2], Reza Paydar[1,2], Fathollah Mohagheghi[3], Sahar Felegari[4], Manijeh Beigi[5], Susan Cheraghi[1,2]

[1]Department of Radiation Sciences, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran, [2]Radiation Biology Research Center, Iran University of Medical Sciences, Tehran, Iran, [3]Department of Oncology, Arak University of Medical Sciences, Arak, Iran, [4]Department of Information Technology, K. N. Toosi University of Technology, Tehran, Iran, [5]Department of Radiation Oncology, Shohadaye Haftome Tir Hospital, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

Abstract

Background: This study aimed to evaluate the effectiveness of clinical, dosimetric, and radiomic features from computed tomography (CT) scans in predicting the probability of heart failure in breast cancer patients undergoing chemoradiation treatment. **Materials and Methods:** We selected 54 breast cancer patients who received left-sided chemoradiation therapy and had a low risk of natural heart failure according to the Framingham score. We compared echocardiographic patterns and ejection fraction (EF) measurements before and 3 years after radiotherapy for each patient. Based on these comparisons, we evaluated the incidence of heart failure 3 years postchemoradiation therapy. For machine learning (ML) modeling, we first segmented the heart as the region of interest in CT images using a deep learning technique. We then extracted radiomic features from this region. We employed three widely used classifiers – decision tree, K-nearest neighbor, and random forest (RF) – using a combination of radiomic, dosimetric, and clinical features to predict chemoradiation-induced heart failure. The evaluation criteria included accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (area under the curve [AUC]). **Results:** In this study, 46% of the patients experienced heart failure, as indicated by EF. A total of 873 radiomic features were extracted from the segmented area. Out of 890 combined radiomic, dosimetric, and clinical features, 15 were selected. The RF model demonstrated the best performance, with an accuracy of 0.85 and an AUC of 0.98. Patient age and V5 irradiated heart volume were identified as key predictors of chemoradiation-induced heart failure. **Conclusion:** Our quantitative findings indicate that employing ML methods and combining radiomic, dosimetric, and clinical features to identify breast cancer patients at risk of cardiotoxicity is feasible.

Keywords: *Chemotherapy echocardiography, heart failure, machine learning, radiomics, radiotherapy*

Submitted: 04-Aug-2024 **Revised:** 01-Oct-2024 **Accepted:** 29-Oct-2024 **Published:** 01-May-2025

Introduction

Breast cancer is the most prevalent cancer among women worldwide.[1,2] The risk of developing heart failure after radiation therapy depends on the radiation dose; as the dose increases, so does the likelihood of complications.[3] The highest dose received by the heart in most patients undergoing breast radiation therapy affects its anterior portion, including the left anterior descending coronary artery.[4] Periodic cardiovascular system monitoring during oncological treatment is crucial to prevent these complications.[5] Echocardiography and electrocardiography are essential tools for assessing left ventricular systolic function. Echocardiography, a noninvasive method of evaluating heart function, can determine heart valve disorders, pericardial disease, left ventricular ejection fraction (LVEF), and diastolic and systolic function. It can detect heart damage in patients undergoing radiation therapy before clinical signs become apparent.[6]

Over the past decade, medical image analysis has significantly advanced due to improvements in diagnostic tools and the expansion of datasets.[7] Machine learning (ML), a subset of artificial intelligence, involves the development of computer algorithms that can replicate human intelligence.[8,9] Radiomics, a new and promising approach to personalized

How to cite this article: Ansari F, Neshasteh-Riz A, Paydar R, Mohagheghi F, Felegari S, Beigi M, *et al*. Radiomics analysis on computed tomography images for prediction of chemoradiation-induced heart failure in breast cancer by machine learning models. J Med Signals Sens 2025;15:14.

Address for correspondence:
Dr. Susan Cheraghi, Iran University of Medical Sciences, Shahid Hemmat Highway, Tehran, Iran.
E-mail: cheraghi.s@iums.ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.jmss_51_24

Quick Response Code:

treatment, involves extracting invisible and quantitative features from medical images. Analyzing these features can provide predictive information about a patient's prognosis and aid in clinical decision-making for many cancers. Despite the large volume of data in medical imaging, ML algorithms are increasingly utilized in oncology. They have applications in cancer diagnosis, staging, treatment simulation, treatment design, quality assurance, response to treatment, and predicting complications.[10-12] Given the significant advances in radiotherapy (RT), the improvement of oncological treatment methods, and the extension of patients' lives, managing the side effects of long-term treatment is essential. Since cardiovascular complications are the most crucial chronic side effects in breast cancer treatment, predicting these complications before starting therapy is vital.[4,13,14]

In 2014, the American Society of Echocardiography (ASE) and the European Association of Cardiovascular Imaging (EACVI) issued an expert consensus statement on multimodality imaging assessment in adult patients during and after cancer treatment, attempting to define cardiotoxicity. According to this statement, cancer treatment-related cardiac dysfunction (CTRCD) is characterized by a reduction in LVEF of at least 5% to <55%, accompanied by signs or symptoms of heart failure or a decrease in LVEF of at least 10% to <55% without evident signs and symptoms.[15-17] Studies have shown that chemoradiotherapy (CRT) for breast cancer can lead to cardiac toxicity in 10%–30% of patients, with a decrease in cardiac EF by approximately 10%–50%.[15,16] In 2011, a study examined 42 breast cancer patients (aged 38–56 years) treated with trastuzumab and RT with an unspecified average anthracycline dose. After treatment, a significant reduction in LVEF of 10% or more to <55% was observed during the 3rd, 6th, 9th, and 12th months, accompanied by signs and symptoms of heart failure and the cessation of chemotherapy. The incidence of cardiac toxicity was 10%.[17]

The goal of our current research was to predict the development of heart failure caused by chemoradiation in breast cancer patients by analyzing their computed tomography (CT) scans using various ML methods. The reduction in EF indicates cardiac impairment, and we applied well-established supervised learning algorithms to assess medical images.

The novelty of this paper lies in using Artificial Intelligence to analyze CT imaging, dosimetrists, and clinical and demographic parameters. This approach may help the early prediction of cardiovascular toxicity, enabling the prevention of its progression and the precise customization of therapeutic pathways for patients.

Materials and Methods

Data collection

This study included 54 adult breast cancer patients (21–80 years) treated for Stage I-IIIA breast cancer who received CRT between January 2018 and January 2020 [Table 1]. The sample size was determined based on statistical calculations, this value of patients has also been used in similar previous studies.[11,18] The study excluded patients with distant lymph node metastases, solid metastases, or missing treatment-related information.

Patient demographics such as age, body mass index, and comorbid conditions were gathered from the Sina and Khansari Medical Centers. Detailed information on breast cancer therapy, including chemotherapy and total radiation doses, was obtained from hospital medical records. Survival and cardiovascular outcomes were determined at the Sina Medical Center.

In this study, by selecting patients based on the Framingham risk score, we purposefully included patients from the low cardiovascular risk group to reduce the influence of other influential variables in heart failure and focus our analysis on the effects of radiomics and chemoradiation variables. Patients with Framingham risk score higher than 5% were excluded from the study.[19] The Framingham risk score is a well-established tool that evaluates the risk of developing coronary issues by considering factors such as gender, age, total cholesterol, high-density lipoprotein cholesterol, systolic blood pressure, and smoking status. It calculates a patient's risk of myocardial infarction and cardiac death, as well as the probability of developing clinical cardiovascular diseases, stroke, peripheral vascular disease, chronic heart failure, and cardiac death over 10 years.

Treatment planning

The patients were subjected to three-dimensional (3D) conformal radiation therapy, receiving a total dose of

Table 1: Demographic and clinical data of patients

Parameters	Value
Age (years), median (range)	54 (21–80)
Chemoradiation therapy	34 patients
Only radiation therapy	20 patients
V_5 (%)	6 (average)
V_{10} (%)	0.54 (average)
V_{25} (%)	1 (average)
Total dose (Gy)	50
Dose/fraction (Gy)	2
Number of fractions	25
Hypertension	9 patients
HER2 positive	12

V5 – The volume of heart (%) receiving 5 Gy; V10 – The volume of heart (%) receiving 10 Gy; V25 – The volume of heart (%) receiving 25 Gy; HER2 – Human epidermal growth factor receptor 2

Table 2: Segmentation accuracy of heart with 2 different deep learning models on dataset

Deep learning model	Dice coefficient, mean±SD
nnUNet	0.82±0.02
SwinUNETR	0.84±0.01

SD – Standard deviation

50 Gy in five fractions per week. Among these patients, 34 received chemotherapy before RT with the following drugs: cyclophosphamide (23 patients), doxorubicin (17 patients), herceptin (15 patients), paclitaxel (15 patients), epirubicin (seven patients), carboplatin (three patients), and cisplatin (one patient). Some patients were treated with multiple chemotherapy drugs. The remaining 20 patients received only RT.

Before starting CRT, all patients underwent echocardiography and CT imaging. CT scans were performed with 100 kV, 220 mAs, and a slice thickness of <5 mm.

Follow-up examinations

Follow-up examinations were conducted 36 months after CRT. These results were compared to each patient's initial EF, which served as the baseline. All echocardiographic images were obtained using a Philips EPIQ 7c echocardiography apparatus. Calculation of LVEF is done with two-dimensional echocardiography using manual biplane measurement according to the Simpson method by an expert cardiologist. This method requires the measurement of LVEF by tracing the endocardial border in both the apical four-chamber and two-chamber views in end-systole and end-diastole.[20] In defining cardiotoxicity, we relied on the expertise of the ASE and the EACVI. According to their statement, CTRCD is characterized by a reduction in LVEF of at least 5% to <55%, accompanied by signs or symptoms of heart failure, or a decrease in LVEF of at least 10% to <55% without evident signs and symptoms.[15-17] Following this definition, we categorized patients into two groups: 1 (with heart toxicity) and 0 (without heart toxicity).

Image preprocessing

We prepared the images for feature extraction during image preprocessing by resampling the voxel size to create an isotropic dataset. Interpolation was applied to all patients' images to facilitate comparisons between data acquired from various samples and scanners.[21] We resampled the images to a resolution of 1 mm × 1 mm × 1 mm, followed by the application of several filters, including Wavelet Decomposition (WAV)[22] and Laplacian of Gaussian (LOG).[23] Different sigma values, ranging from 0.5 to 5 in increments of 0.5, were used with the LOG filter to extract features at fine, medium, and coarse scales. The use of Wavelet Decomposition and LOG filters was motivated by their complementary abilities to capture fine and coarse image features, crucial for identifying subtle structural changes in the heart tissue. Furthermore, they can capture multiscale textures and edges, offering spatial and frequency information crucial for conducting detailed radiomic analyses. Wavelet decomposition allows multiresolution analysis of the images, whereas LOG filter effectively detects important anatomical boundaries at multiple scales. These filters greatly improved the quality of the preprocessed images and enabled the extraction of valuable radiomic features for our ML models.[21-23] The Wavelet filter provided eight decompositions at all levels, encompassing all possible combinations of high-pass and low-pass filters in each of the three dimensions, resulting in combinations such as HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL. Traditional threshold functions have limitations in wavelet denoising; a hard threshold function is characterized by discontinuity, while a soft threshold function results in constant deviation. To address these issues, we employed an enhanced wavelet threshold for denoising.[22] The preprocessing algorithms, including interpolation, WAV, and LOG, significantly improved the quality of the CT images.

Heart autosegmentation through deep learning

Experienced radiotherapists, under the supervision of a specialist, used 3D-Slicer software to annotate the heart and define the region of interest (ROI). The entire cardiac muscle was delineated as the ROI. This process involved iteratively going through each image slice and drawing contours along the boundaries to delineate areas of interest accurately.

Since manual heart segmentation can be time-consuming while dealing with big datasets, two deep-learning models were implemented to make this process automatically applicable [Figure 1]. For this mean, we use both SwinUNETR and nnUNet models for heart autosegmentation.[24,25] This decision was motivated by the need to evaluate multiple state-of-the-art deep learning architectures and determine which one could provide the most accurate segmentation for our specific dataset. Both models have been widely recognized for their performance in medical image segmentation tasks, but they use different architectures and approaches that offer complementary strengths. SwinUNETR was chosen because it leverages the Swin Transformer architecture, which is particularly efficient in handling large-scale medical imaging datasets. The hierarchical vision transformer used in SwinUNETR enhances the computational efficiency of self-attention mechanisms, making it well suited for complex tasks like heart segmentation. nnUNet was included in our study due to its flexibility and adaptability to a wide range of segmentation tasks. nnUNet automatically configures its architecture, preprocessing, and training strategies to the dataset at hand, ensuring an optimal model for the specific segmentation task.[24,26] Both models were trained and validated on the same dataset to ensure a fair comparison. The performance of these models was evaluated using the Dice coefficient, a widely accepted metric for segmentation tasks [Table 2].

To optimize the model hyperparameters, we systematically explored various combinations, including different learning

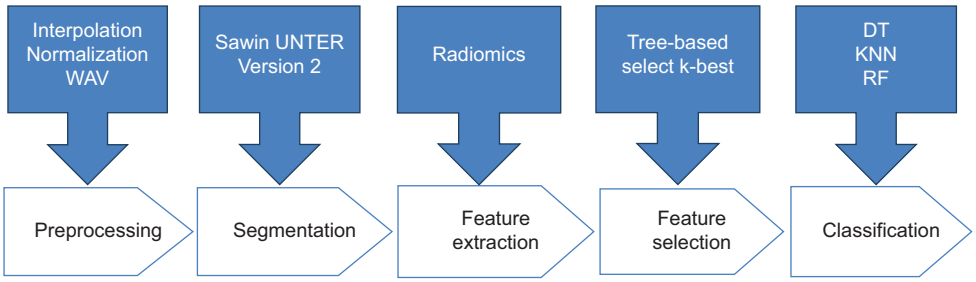

Figure 1: Proposed framework for radiomics analysis on computed tomography images for predicting chemoradiation-induced heart failure by machine learning models

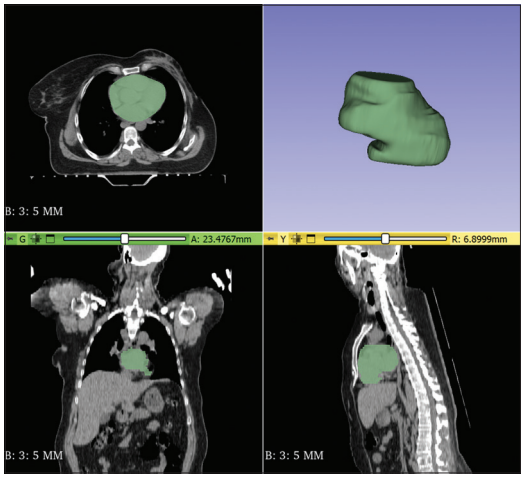

Figure 2: Segmented area of heart. This step is done by experienced radiotherapists under the supervision of a specialist to define the region of interest

rates, batch sizes, and data augmentation strategies. Preprocessing steps included intensity normalization (scaling intensities from − 175 to 250), cropping the foreground to focus on the heart, reorienting images to RAS orientation, and resampling them to uniform spacing of 1.5 mm × 1.5 mm × 2.0 mm. Random augmentations, such as flips and intensity shifts, were also applied during training to enhance model robustness and prevent overfitting. During SwinUNETR training, various data augmentation techniques were applied, including random flips across spatial axes 1 and 2, random rotations with a 10% flipping probability, and random intensity shifts with a 50% flipping probability within a 10% offset range. Patch-based SwinUNETR utilized patch sizes of 96 × 96 × 32 for 30,000 training iterations, implemented using the MONAI library.[27] A batch size of 2 for training and 1 for testing was selected to balance memory usage and training time while maintaining stable model convergence. The Adam optimizer was employed with a weight decay of 1e-5 and a learning rate of 1e-4, ensuring stable convergence without the instabilities seen with higher rates or the slower convergence of lower rates.[28] In addition, SwinUNETR was initialized using pretrained self-supervised weights to improve model performance from the start of training. These hyperparameter choices were based on multiple trials, and the selected values yielded optimal segmentation accuracy. The dataset was partitioned into 12% for evaluation, 10% for testing, and 78% for training.

The nnUNet version 1 was trained for 1000 epochs using the stochastic gradient descent method with an initial learning rate of 0.01 and a batch size of 2. Data augmentation techniques included random rotations, random scaling, elastic deformations, and random cropping, along with intensity augmentations such as brightness and contrast adjustments to improve model robustness. Based on the evaluation metrics in Table 2, SwinUNETR was selected for segmentation due to its superior performance in terms of segmentation accuracy compared to nnUNet.

Features extraction

Radiomic features were extracted from the segmented heart region using a custom extension of 3D Slicer software. 3D Slicer is an open-source medical image computing platform for biomedical research and freely downloadable (www.slicer.org). A total of 873 radiomic features were acquired from ROI using the 3D Slicer Figure 2.

Feature selection

In our study, we employed multiple feature selection methods to identify the most important features for predicting chemoradiation-induced heart failure from a combined set of 890 radiomic, dosimetric, and clinical features. Specifically, a combination of filter and embedded methods was used in the selection process. First, using Kendall's rank correlation coefficient, features that had the highest correlation with the target variable (heart failure) were selected. Features with a correlation coefficient exceeding a specified threshold were chosen for the next stage. Next, to further reduce dimensionality and select the best features, we applied cross-validation along with the random forest (RF) algorithm. RF allowed us to evaluate the importance of each feature based on its predictive power in the model. The final features were selected based on their importance in different models, including RF and decision tree (DT). These methods ensured that only the most impactful features for predicting heart failure were chosen. This combination of feature selection methods enabled us to reduce the original 890 features to 15 key features, dimensionality reduction helped to mitigate overfitting and

reduce computational complexity by prioritizing the most informative features. Ultimately, it improves the accuracy of the model.[29]

Classification

In our study, we utilized Python 3.7 and PyCharm 2019.3.1 to implement DT, K-Nearest Neighbor (KNN), and RF ML algorithms to analyze the correlations between radiomic, dosimetric, and clinical features with heart damage.[27]

Choosing multiple different algorithms such as DT, KNN, and RF allows us to have greater diversity in modeling and benefit from the advantages of each. We can use their results to compare prediction accuracy and select the best model. DT model provides a clear, interpretable model structure, making it easier for clinicians to understand decision-making processes. The KNN algorithm uses a learning method based on the nearest neighbors of each query point, with the number of neighbors set to five in this investigation. RF, recognized as one of the most commonly used ML techniques, serves as a meta-estimator that trains multiple DT classifiers on different subsets of the dataset and uses averaging to enhance predictive accuracy and manage overfitting. The test size for this study was set to 0.2. In our experiments, we trained the algorithms using 5-fold cross-validation. During cross-validation, the data are repeatedly divided, and several models are trained rather than splitting the dataset into separate training and test sets. This approach provides insight into our model's best and worst-case performance when applied to new data. We used four evaluation criteria for comparison: accuracy, sensitivity, specificity, and the area under the curve (AUC). Accuracy represents the percentage of total correct predictions relative to the total samples, whereas sensitivity measures the model's ability to identify true positive samples, and specificity indicates the model's ability to identify true negative samples. AUC is a useful measure to evaluate the performance of a binary classifier at different threshold values, which reflects the ability of the model to discriminate between classes, regardless of the class distribution. By considering these parameters, we can gain a better understanding of the model's performance under different conditions and identify its strengths and weaknesses, which is crucial for clinical decision-making or other applications.

Datasets

We categorize the features into three categories (radiomic, clinical, and dosimetric). The clinical features included age, type of chemotherapy drugs, blood pressure, and body mass index, whereas dosimetric features consisted of the total dose, dose per fraction, number of fractions, V5, V10, and V25. We tested different data combinations in the models and implemented them accordingly. The integration of these features was based on empirical results and theoretical considerations. Radiomic features, extracted from medical images, provided detailed insights into tissue characteristics. Clinical features, such as patient age, were considered for that potential impact on heart failure outcomes, whereas dosimetric features related to radiation exposure were included to capture the effects of radiation therapy. Initially, we analyzed each feature category separately and then combined them to assess their contributions to model performance. Combining all datasets as input to the model yields better results and, in this case, the patient's age and numerous radiomic features emerged as important. Therefore, we ran the 3 ML algorithms with four datasets to confirm this result. The first dataset included radiomic and clinical features; the second dataset included only radiomic features. The third dataset combined radiomic features and patient age as the only important nonradiomic feature. The last dataset included the top 10 selected features, combining radiomic, clinical, and dosimetric features. The flow chart of this study is provided in Figure 1.

Results

Based on the EF from echocardiography, 25 of the patients (46%) experienced heart failure.

The segmented area yielded 873 radiomic features, of which 17 were selected. The RF algorithm using the dataset combining radiomic features and patient age, and the DT algorithm using the dataset combining radiomic, clinical, and dosimetric features, provided the best predictions for heart failure (AUC = 0.98 and AUC = 0.92, respectively) [Table 3].

Important features for predicting chemoradiation-induced heart failure using the RF classifier, derived from the

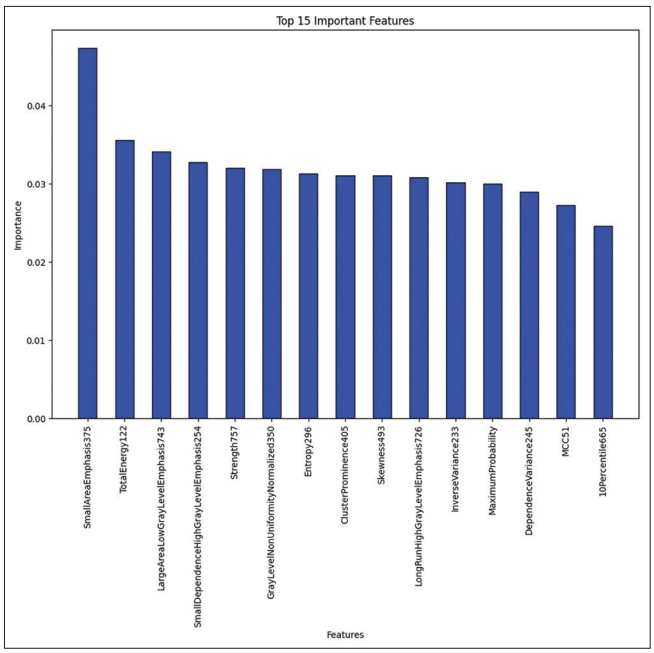

Figure 3: Important features for prediction of chemoradiation-induced heart failure in random Forest classifier resulting from the first dataset (radiomic and clinical features combination)

Table 3: Results of machine learning models performances in predicting chemoradiation-induced heart failure

Models	Datasets	Accuracy	Sensitivity	Specificity	AUC
RF	17 radiomic features + clinical features	0.85	0.83	0.85	0.90
	Only 17 selected radiomic features	0.85	0.83	0.85	0.92
	15 selected radiomic features + patient age	0.85	0.83	0.85	0.98
	10 selected radiomic features + selected clinical and dosimetry features	0.85	0.83	0.85	0.96
DT	17 radiomic features + clinical features	0.85	0.83	0.85	0.92
	Only 17 selected radiomic features	0.85	0.83	0.85	0.88
	15 selected radiomic features + patient age	0.85	0.83	0.85	0.89
	10 selected radiomic features + selected clinical and dosimetry features	0.85	0.83	0.85	0.96
KNN	17 radiomic features + clinical features	0.67	0.57	0.75	0.64
	Only 17 selected radiomic features	0.67	0.57	0.75	0.64
	15 selected radiomic features + patient age	0.67	0.57	0.75	0.64
	10 selected radiomic features + selected clinical and dosimetry features	0.70	0.80	0.60	0.76

Plus (+) means combination. AUC – Area under the curve; DT – Decision tree; KNNs – K nearest neighbors; RF – Random forest

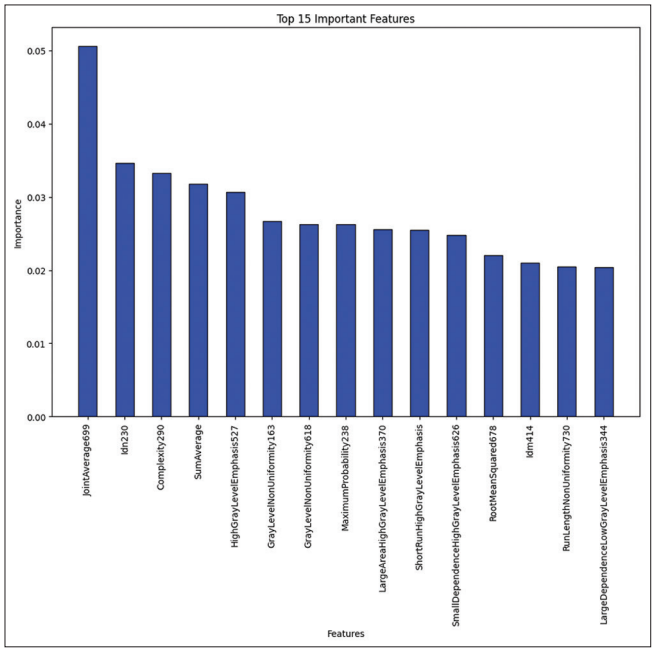

Figure 4: Important features for prediction of chemoradiation-induced heart failure in random forest classifier resulting from the third dataset (radiomic features and patient age combination)

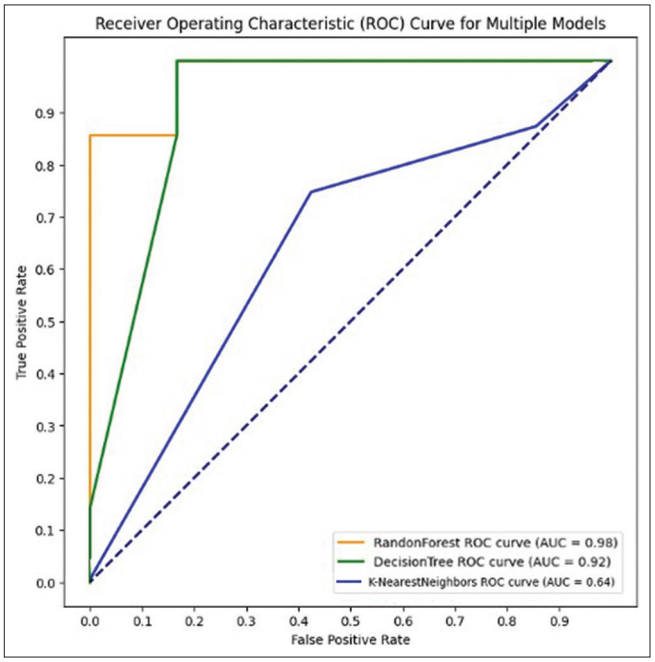

Figure 5: Majoring performance in three classifiers based on datasets identified by the area under the receiver operating characteristic curve. The random forest is shown in orange, K-nearest neighbors in blue, and the decision tree in green. ROC – Receiver operating characteristic

first and third datasets (radiomic and clinical feature combinations), are shown in Figures 3 and 4.

Essential features for predicting chemoradiation-induced heart failure using the RF classifier from the fourth dataset (10 selected radiomic features combined with selected clinical and dosimetric features) are shown in Table 4.

Overall, the RF classifier outperformed the DT and KNN classifiers based on high AUC values, accuracy, sensitivity, and specificity, with an average AUC of 0.94. RF classifier identified 15 essential features from the radiomic features and age dataset. A comparison of the three classifiers was conducted based on the AUC of the receiver operating characteristic curves, which plot the model's features against sensitivity and are created by adjusting classification thresholds [Figure 5].

Discussion and Conclusion

This research investigated three ML algorithms to predict heart failure in breast cancer patients using their medical imaging, clinical, and dosimetric data. Our main objective was to present predictive models that can identify breast cancer patients at high risk of developing heart problems due to cancer treatments, with the goal of preventing or minimizing the occurrence of such issues. Prior research has utilized ML techniques to predict heart failure in general patient populations without explicitly targeting any particular treatment or

Table 4: Important features for prediction of chemoradiation-induced heart failure in random forest classifier resulting from the fourth dataset (10 selected radiomic features and selected clinical and dosimetric features combination)

	Important features	Importance value
1	V_5 (%)	0.628021
2	GrayLevelNonUniformityNormalized443	0.186887
3	LargeDependenceHighGrayLevelEmphasis343	0.179115
4	Mean	0.1719
5	ShortRunHighGrayLevelEmphasis549	0.15135
6	V_{10} (%)	0.141213
7	ZonePercentage658	0.118181
8	RunLengthNonUniformityNormalized452	0.097925
9	Contrast663	0.083257
10	HighGrayLevelEmphasis341	0.0673

disease.[30-32] However, none of these methods have been able to predict radiation-induced heart problems based on medical imaging. Our study demonstrated the potential of using ML approaches to predict heart failure in breast cancer patients by analyzing CT scan images and LVEFs. The advantage of cohort studies like ours is that they reduce the effect of inherent parameters and latent variable effects, thereby enabling relatively high accuracy in predicting complications.

From the first dataset in our study, Short Run High Gray Level Emphasis 735 was identified as the most important feature. Other essential radiomic features related to cardiac damage included Small Dependence Low Gray Level Emphasis, Run Length Non Uniformity, and Joint Energy. The best performance for this dataset was achieved by the DT algorithm, with an AUC of 0.90. To comparison of the results, we used a second dataset that included only radiomic features. We had expected that combining clinical features with radiomic data would improve the performance of diagnostic models. However, the results showed that the models produced similar outputs in both cases. Combining all the data in the modeling showed us that, except for age, none of the clinical characteristics were among the most important characteristics identified by the models. Other studies also confirm that age is a critical predictor, with studies indicating that older patients are at a higher risk for complications postradiation therapy.[33,34] To confirm this result, we re-ran the models with a third dataset, which included radiomic features and the age column, excluding other dosimetric and clinical features. The model outputs were the same as when the combination of radiomic and clinical features was applied. These results indicate that clinical features do not play a significant role in improving the accuracy of our prediction models. This could be due to the high influence of features derived from medical images that are directly related to pathological features. Therefore, necessary information for heart failure classification and prediction is likely to be found in radiomic features. How clinical and dosimetric characteristics influence predictive models led to the creation of the fourth dataset. For this purpose, first, clinical and dosimetric characteristics were presented to the models and six critical characteristics were identified. Then, these features were combined with ten radiomic features that were recognized as essential by the models, and this combined dataset, known as the fourth dataset, was used to compare the results. Based on the obtained results, the V5 (%) feature was identified as the most crucial feature. This finding aligns with other studies that have recognized V5 as a marker of heart muscle damage, which can serve as a predictor for patients receiving radiation therapy for left-sided breast cancer and lung cancer.[35] In radiation biology, the heart is considered an organ with both serial and parallel array substructures.[36] This classification is based on the volume effect and organization of various tissues and cells within the heart that can be affected by radiation. In parallel organs, the percentage volume receiving radiation such as V5 is a critical parameter. Knowing this can help in assessing risks and protective measures in radiation exposure scenarios. In our study, in addition to emphasizing the parallel aspect of the heart, it was shown that the V5 metric, which represents the volume of the heart receiving at least 5 Gy of radiation, can effectively predict cardiac disease. In another study, a V5 threshold ≥49% was associated with a significantly higher risk of Grade 3 cardiac complications.[37]

To optimize the performance of the three ranking algorithms (KNN, RF, and DT), we employed four distinct evaluation criteria: accuracy, sensitivity, specificity, and AUC. Accuracy indicates how effectively a test can differentiate between cardiac complications and normal cases. To assess a test's accuracy, it is essential to calculate the true positive rate and true negative rate for all cases under evaluation. Based on accuracy, both DT and RF models outperformed the KNNs algorithm, achieving respective accuracy rates of 0.85 and 0.67. The AUC values for DT and RF are significant indicators of their predictive performance. Given that all AUC values are >0.5, it can be inferred that these models are effective predictors. Notably, RF exhibits an AUC value that is closer to 1, indicating that it is the superior predictor among the evaluated models. The DT model achieved the highest accuracy when using both clinical and radiomic datasets, whereas the RF model performed best when using only radiomic datasets. Patients' ages were among the best predictors for heart failure, with AUC values of 0.92 and 0.98, respectively. The analysis of the extracted features revealed that in the third dataset, which includes radiomic features and patients' ages, the Gray Level Co-occurrence Matrix class contains important features. This class examines the statistical relationships between pairs of pixels with different gray levels at a specific distance, providing information about the diversity and homogeneity of tissues in the image. The

next important class is the Gray Level Dependence Matrix, which quantifies gray-level dependencies in the image. The number of voxels connected at a specific distance (δ) and dependent on the central voxel is called gray-level dependence. The Gray Level Run Length Matrix (GLRLM) quantifies the lengths of consecutive runs of gray levels, indicating the number of successive pixels with the same gray levels. The Gray Level Size Zone Matrix identifies regions with different gray levels and provides an advanced statistical matrix for describing tissues. This order also holds for the second dataset, which includes only radiomic features. For the first dataset, which consists of all radiomic and clinical features, GLRLM is the essential feature.

The analysis of the extracted features from different algorithms shows that some common features consistently contribute to the predictive power of the models. In the combination of clinical and radiomic features, both DT and RF models identify Short Run High Gray Level Emphasis as an essential feature with an AUC of 0.97. This feature captures short runs with high intensity in the image, potentially indicating small and bright areas associated with pathological changes. When we focused solely on radiomic features, the RF models performed better and highlighted Gray Level Non-uniformity Normalized 443 as the key feature. This feature measures the variability in gray level intensity values normalized by the total number of gray level values, indicating the uniformity of intensity distribution within an image; for the third dataset, which combined radiomic features and age, both the DT and RF models identified Short Run High Gray Level Emphasis as an important feature. This feature helps classify high-intensity short runs in the image, potentially indicating small and bright areas associated with pathological changes.

The superior performance of the RF and DT models can be attributed to their tree-like structure, including decision and leaf nodes. This structure involves a series of consecutive choices to achieve a specific outcome. The element of randomness introduces diversity among individual trees, reducing the risk of overfitting and enhancing overall predictive performance. Consequently, these tree-learning methods have become influential in ML. Over the years, numerous studies have focused on radiomic models in RT due to their ability to offer personalized patient treatment by providing quantitative descriptions of medical images.[7] In addition, several studies have applied radiomics in diagnosing and prognosing heart failure.[8-10] For example, Parmar *et al*.[38] conducted a similar study and found that identifying the best ML algorithm for analyzing predictive radiomic features could expand the field of radiomics in oncology and cancer treatment.[12] Another study found that CT features could distinguish sensory-neural hearing loss in head-and-neck tumors, achieving an accuracy of more than 0.70 using ten ML methods.[39] Moreover, Zhou *et al*. outlined risk assessments based on ML, which included patients' ages, hypertension, glucose levels, LVEF, creatinine, and aspartate aminotransferase levels for detecting cancer treatment-related CTRCD.[40]

The study's limitations included a relatively small number of patients and the need to more follow-up for myocardial function evaluation. Small patient sample sizes may affect the generalizability of our findings to other populations or situations, and the findings may not accurately represent the behaviors and characteristics of the entire population. Short-term follow-up was also required for better prediction in times before 3 years. In addition, although we used the Framingham risk score test to exclude background individual cardiovascular risk factors, it is recommended that future studies employ multivariate prediction models to assess all risks of heart failure.

In conclusion, our understanding of the mechanisms underlying cancer therapy-induced cardiotoxicity is limited. This research serves as an initial step in evaluating ML models for predicting heart failure in this patient population.

Code, data, and materials availability

The data that support the findings of this article are not publicly available due to privacy concerns. They can be requested from the corresponding author at the E-mail listed above.

Acknowledgments

The authors acknowledge the funding provided by the Research Chancellor of Iran University of Medical Sciences (Grant number 1400-3-5-22475).

Financial support and sponsorship

This study is funded by the Iran University of Medical Sciences (IUMS).

Conflicts of interest

There are no conflicts of interest.

References

1. Fouladi N, Amani F, Harghi AS, Nayebyazdi N. Five year survival of women with breast cancer in Ardabil, North-West of Iran. Asian Pac J Cancer Prev 2011;12:1799-801.
2. Riihimäki M, Thomsen H, Brandt A, Sundquist J, Hemminki K. Death causes in breast cancer patients. Ann Oncol 2012;23:604-10.
3. Saiki H, Petersen IA, Scott CG, Bailey KR, Dunlay SM, Finley RR, *et al.* Risk of heart failure with preserved ejection fraction in older women after contemporary radiotherapy for breast cancer. Circulation 2017;135:1388-96.
4. Berkman A, Cole BF, Ades PA, Dickey S, Higgins ST, Trentham-Dietz A, *et al.* Racial differences in breast cancer, cardiovascular disease, and all-cause mortality among women with ductal carcinoma *in situ* of the breast. Breast Cancer Res Treat 2014;148:407-13.
5. Kim H, Chung WB, Cho KI, Kim BJ, Seo JS, Park SM, *et al.* Diagnosis, treatment, and prevention of cardiovascular toxicity related to anti-cancer treatment in clinical practice: An opinion

paper from the working group on cardio-oncology of the Korean Society of Echocardiography. J Cardiovasc Ultrasound 2018;26:1-25.
6. Zamorano JL, Lancellotti P, Rodriguez Muñoz D, Aboyans V, Asteggiano R, Galderisi M, *et al.* 2016 ESC position paper on cancer treatments and cardiovascular toxicity developed under the auspices of the ESC Committee for Practice Guidelines: The task force for cancer treatments and cardiovascular toxicity of the European Society of Cardiology (ESC). Eur Heart J 2016;37:2768-801.
7. Brunese L, Mercaldo F, Reginelli A, Santone A. Formal methods for prostate cancer Gleason score and treatment prediction using radiomic biomarkers. Magn Reson Imaging 2020;66:165-75.
8. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol 2019;188:2222-39.
9. Shi Y, Suk H-I, Liu M. Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings: Springer; 2018.
10. Gagliardi G, Constine LS, Moiseenko V, Correa C, Pierce LJ, Allen AM, *et al.* Radiation dose-volume effects in the heart. Int J Radiat Oncol Biol Phys 2010;76:S77-85.
11. Amiri S, Abdolali F, Neshastehriz A, Nikoofar A, Farahani S, Firoozabadi LA, *et al*. A machine learning approach for prediction of auditory brain stem response in patients after head-and-neck radiation therapy. J Cancer Res Ther 2023;19:1219-25.
12. Agheli R, Siavashpour Z, Reiazi R, Azghandi S, Cheraghi S, Paydar R. Predicting severe radiation-induced oral mucositis in head and neck cancer patients using integrated baseline CT radiomic, dosimetry, and clinical features: A machine learning approach. Heliyon 2024;10:e24866.
13. Darby SC, Ewertz M, McGale P, Bennet AM, Blom-Goldman U, Brønnum D, *et al.* Risk of ischemic heart disease in women after radiotherapy for breast cancer. N Engl J Med 2013;368:987-98.
14. Taylor C, Correa C, Duane FK, Aznar MC, Anderson SJ, Bergh J, *et al.* Estimating the risks of breast cancer radiotherapy: Evidence from modern radiation doses to the lungs and heart and from previous randomized trials. J Clin Oncol 2017;35:1641-9.
15. Seidman A, Hudis C, Pierri MK, Shak S, Paton V, Ashby M, *et al.* Cardiac dysfunction in the trastuzumab clinical trials experience. J Clin Oncol 2002;20:1215-21.
16. Plana JC, Galderisi M, Barac A, Ewer MS, Ky B, Scherrer-Crosbie M, *et al.* Expert consensus for multimodality imaging evaluation of adult patients during and after cancer therapy: A report from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2014;15:1063-93.
17. Jiji RS, Kramer CM, Salerno M. Non-invasive imaging and monitoring cardiotoxicity of cancer therapeutic drugs. J Nucl Cardiol 2012;19:377-88.
18. Jang BS, Chang JH, Park AJ, Wu HG. Generation of virtual lung single-photon emission computed tomography/CT fusion images for functional avoidance radiotherapy planning using machine learning algorithms. J Med Imaging Radiat Oncol 2019;63:229-35.
19. D'Agostino RB Sr., Grundy S, Sullivan LM, Wilson P, CHD Risk Prediction Group. Validation of the Framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation. JAMA 2001;286:180-7.
20. Benyounes N, Van Der Vynckt C, Tibi T, Iglesias A, Gout O, Lang S, *et al.* Left ventricular end diastolic volume and ejection fraction calculation: Correlation between three echocardiographic methods. Cardiol Res Pract 2020;2020:1-7.
21. Shiri I, Maleki H, Hajianfar G, Abdollahi H, Ashrafinia S, Hatt M, *et al.* Next-generation radiogenomics sequencing for prediction of EGFR and KRAS mutation status in NSCLC patients using multimodal imaging and machine learning algorithms. Mol Imaging Biol 2020;22:1132-48.
22. Zhang Y, Ding W, Pan Z, Qin J. Improved wavelet threshold for image de-noising. Front Neurosci 2019;13:39.
23. Chen JS, Huertas A, Medioni G. Fast convolution with Laplacian-of-Gaussian masks. IEEE Trans Pattern Anal Mach Intell 1987;9:584-90.
24. Yazdani E, Karamzadeh-Ziarati N, Cheshmi SS, Sadeghi M, Geramifar P, Vosoughi H, *et al.* Automated segmentation of lesions and organs at risk on [(68)Ga]Ga-PSMA-11 PET/CT images using self-supervised learning with Swin UNETR. Cancer Imaging 2024;24:30.
25. Isensee F, Wald T, Ulrich C, Baumgartner M, Saikat Roy, Maier-Hein K *et al*. A Call for Rigorous Validation in 3D Medical Image Segmentation nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention; Springer: 2024.
26. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-11.
27. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, *et al*. Scikit-learn: Machine learning in Python. J Mach Learn Res 2022;12:2825-30.
28. Zhang Z. Improved Adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE; 2018.
29. Whybra P, Zwanenburg A, Andrearczyk V, Schaer R, Apte AP, Ayotte A, *et al.* The image biomarker standardization initiative: Standardized convolutional filters for reproducible radiomics and enhanced clinical insights. Radiology 2024;310:e231319.
30. Kavitha M, Gnaneswar G, Dinesh R, Rohith Sai Y, Sai Suraj R. Heart disease prediction using hybrid machine learning model. In: 2021 6th International Conference on Inventive Computation Technologies (ICICT). IEEE; 2021.
31. Mehta LS, Watson KE, Barac A, Beckie TM, Bittner V, Cruz-Flores S, *et al.* Cardiovascular disease and breast cancer: Where these entities intersect: A scientific statement from the American Heart Association. Circulation 2018;137:e30-66.
32. Ezaz G, Long JB, Gross CP, Chen J. Risk prediction model for heart failure and cardiomyopathy after adjuvant trastuzumab therapy for breast cancer. J Am Heart Assoc 2014;3:e000472.
33. Cella L, D'Avino V, Palma G, Conson M, Liuzzi R, Picardi M, *et al.* Modeling the risk of radiation-induced lung fibrosis: Irradiated heart tissue is as important as irradiated lung. Radiother Oncol 2015;117:36-43.
34. Gaya AM, Ashford RF. Cardiac complications of radiation therapy. Clin Oncol (R Coll Radiol) 2005;17:153-9.
35. Cirnigliaro V, Pietrosanti S, Demofonti C, DE Angeli M, Giovenco D, Cedrone L, *et al.* V5 and high sensitivity cardiac troponin T for early detection of cardiac toxicity during left breast cancer irradiation. Cancer Diagn Progn 2023;3:365-9.
36. Saynak M. Türkkan G, Nurlu D, Özgüven Y. Radiotherapyinduced cardiotoxicity after the treatment of pulmonary and mediastinal solid tumors. Turk J Oncol 2022;37:20-7.
37. Ni L, Koshy M, Connell P, Pitroda S, Golden DW, Al-Hallaq H, *et al.* Heart V5 predicts cardiac events in unresectable lung cancer patients undergoing chemoradiation. J Thorac Dis 2019;11:2229-39.

38. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine learning methods for quantitative radiomic biomarkers. Sci Rep 2015;5;1:1-11.
39. Abdollahi H, Mostafaei S, Cheraghi S, Shiri I, Rabi Mahdavi S, Kazemnejad A. Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: A machine learning and multi-variable modelling study. Phys Med 2018;45:192-7.
40. Zhou Y, Hou Y, Hussain M, Brown SA, Budd T, Tang WH, *et al.* Machine learning-based risk assessment for cancer therapy-related cardiac dysfunction in 4300 longitudinal oncology patients. J Am Heart Assoc 2020;9:e019628.