frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 21 August 2025
DOI 10.3389/fonc.2025.1557382

OPEN ACCESS

EDITED BY
Rajesh Kumar Tripathy,
Birla Institute of Technology and Science,
India

REVIEWED BY
Ana Margarida Mota,
University of Lisbon, Portugal
Kamuran Arslan Ibis,
Istanbul University, Türkiye

*CORRESPONDENCE
Sefika Dincer
sefikadincer94@gmail.com

RECEIVED 08 January 2025
ACCEPTED 27 June 2025
PUBLISHED 21 August 2025

CITATION
Dincer S, Akmansu M and Akyol O (2025) Machine learning modeling of cancer treatment-related cardiac events in breast cancer: utilizing dosiomics and radiomics. *Front. Oncol.* 15:1557382.
doi: 10.3389/fonc.2025.1557382

Machine learning modeling of cancer treatment-related cardiac events in breast cancer: utilizing dosiomics and radiomics

Sefika Dincer[1]*, Muge Akmansu[2] and Oya Akyol[2]

[1]Department of Radiation Oncology, Van Yuzuncu Yil University School of Medicine, Van, Türkiye,
[2]Department of Radiation Oncology, Gazi University School of Medicine, Ankara, Türkiye

Background: Personalized medicine has transformed disease management by focusing on individual characteristics, driven by advancements in genome mapping and biomarker discoveries.

Objectives:: This study aims to develop a predictive model for the early detection of treatment-related cardiac side effects in breast cancer patients by integrating clinical data, high-sensitivity Troponin-T (hs-TropT), radiomics, and dosiomics. The ultimate goal is to identify subclinical cardiotoxicity before clinical symptoms manifest, enabling personalized surveillance strategies. It is the first study to utilize heart-segmented dosiomics in breast cancer patients.

Methods and Materials: This retrospective study included clinical, dosimetric, radiomic, and dosiomic data from 42 women with localized breast cancer. Heart-specific Troponin T levels were measured 2–3 weeks post-radiotherapy, with 14 ng/L as the cutoff. Patients were grouped on this threshold to identify potential treatment-related cardiac events. Radiomics and dosiomics were extracted using PyRadiomics. Machine learning models were optimized using the Tree-based Pipeline Optimization Tool (TPOT), identifying the gradient-boosted classification as the best-performing algorithm. Feature selection was conducted using gradient-boosted recursive feature elimination. Model performance is assessed by the area under the curve (AUC).

Results: A total of 111 dosiomic and 119 radiomic features were extracted per patient. The highest predictive accuracy was achieved using clinical, dosiomic, and radiomic parameters (validation cohort-AUC = 0.96), outperforming the clinical + dosimetric model (validation cohort-AUC = 0.67). Permutation tests confirmed the non-randomness of these two models results ($p < 0.05$). Cross-validation indicated that the clinical + dosiomic + radiomic model had a fair-to-good generalizable performance (mean AUC = 80.33 $\pm$ 21%).

Discussion: This study may demonstrate that radiomics and dosiomics provide superior predictive capabilities for cardiac events in breast cancer patients compared to traditional parameters.

KEYWORDS

dosiomics, oncologic treatment-related cardiotoxicity, machine learning, radiomics, oncology cardiology

Introduction

Cardiovascular complications following cancer treatment represent a significant cause of mortality among cancer patients, mainly when therapies such as chemotherapy and radiotherapy induce structural and functional damage to the heart. This concern is particularly substantial for breast cancer patients, where treatments like anthracyclines, trastuzumab, pertuzumab, immune checkpoint inhibitors (ICIs), and radiotherapy can lead to cancer therapy-related cardiac dysfunction (CTRCD) (1–6). Consequently, regularly monitoring the cardiovascular status in breast cancer survivors is imperative, using diagnostic tools such as left ventricular ejection fraction (LVEF) and cardiac biomarkers like troponin and echocardiography (7–11).

Echocardiography is currently the preferred non-invasive imaging modality for diagnosing CTRCD (7, 12). Furthermore, this method often detects cardiac dysfunction only at advanced stages, leading to irreversible damage in 58% of cases (13, 14). A meta-analysis highlights that hs-cTnT measured at the 3-6th month of treatment provides superior early diagnostic value to echocardiography, emphasizing the need for its use in predicting cardiac events (15).

The European Society of Cardiology recommends evaluating serum cardiac biomarkers, such as Trop-T and NT-pro-BNP, before, during, and after treatment to detect subclinical CTRCD and tailor oncologic therapies to individual risk profiles (16). Elevated levels of these biomarkers in cancer patients without coronary artery disease are indicative of subclinical cardiac damage and are linked to reduced left ventricular function (17).

In the last decade, non-invasive imaging biomarkers have been developed to assist in personalized oncology decision support systems. Radiomics and dosiomics are just two of these methods. These techniques have shown promising results in predicting various clinical outcomes, including overall survival and recurrence in cancers such as brain, head and neck, and lung (18–22). However, the scientific field still lacks sufficient dosiomic-radiomic studies, particularly those focused on side effects rather than disease prognosis.

Accordingly, all this literature predicting treatment-related cardiac events before cancer therapy begins would be highly significant. This study aims to bridge this gap by developing a machine learning-based prediction model for early detection of treatment-related cardiac events in breast cancer patients. The model incorporates radiomics, dosiomics, and clinical parameters to identify subclinical cardiotoxicity. To our knowledge, this is the first study to utilize heart-segmented dosiomic data in adult breast cancer patients. To our knowledge, this is the first study to utilize heart-segmented dosiomic data in adult breast cancer patients, and it has the potential to serve as a precursor for future research in this area.

Materials and methods

Patients

In this retrospective study, data from patients treated for localized breast cancer at our institution between January 2020 and February 2024 were analyzed. Of the 235 patients initially screened, those who had hs-TropT levels measured in the blood 2–3 weeks after the completion of radiotherapy were considered eligible. Only patients who received radiotherapy as part of their treatment were included; those treated exclusively with systemic therapies such as chemotherapy (e.g., anthracyclines, trastuzumab) were not considered. Patients with a prior history of cardiac events resulting in elevated hs-TropT levels were not eligible. Additionally, cases involving breath-controlled radiotherapy or significant metal artifacts on planning CT scans (n = 3) were omitted due to potential data quality concerns. Ultimately, 42 patients met all inclusion criteria and were retained for final analysis. As frequently recommended in the literature, the ratio between the training and validation cohort groups was set at %80/20 (33 patients in the training group and 9 patients in the validation group). The patient selection process is summarized in Figure 1. Clinical data included age, comorbidities, smoking status, use of cardioprotective medications, chemotherapy regimens, surgical procedures, and laterality of breast cancer. Dosimetric data included the dose of radiotherapy delivered to the heart, the presence of a boost dose, the fractionation scheme, the radiotherapy technique used, and the mean and maximum radiation doses received by the heart, as well as the heart volumes receiving 5 Gy and 25 Gy (V5-V25). Radiomic features were extracted from pre-contoured heart regions in the planning CT scans using advanced statistical methods, and dosiomic features were derived from the radiotherapy plan dose distributions through similar statistical processing. The dosiomic data were derived from the originally planned dose distribution, with the fractionation value (hypofractionation vs. normofractionation) included as a separate variable in the model. The left-right breast distinction was not

FIGURE 1
Patient selection process. N, number of patients; Trop-T, Troponin-T.

defined as an independent variable but incorporated as a separate parameter in the clinical and clinical+radiomic models (Table 1). The Gazi University Ethics Committee approved the study on April 16, 2024 (Research Code No: 2024-619).

Study design

Figure 2 illustrates the overall workflow of this study. Our study set the threshold value for high-sensitivity Troponin T (hs-TropT) at greater than 14 ng/L (23, 24). The dependent variable (target), hs-TropT, was classified into two groups: those above the threshold of 14 ng/L and those equal to or below 14 ng/L. Troponin-T levels exceeding 14 ng/dL were associated with treatment-related cardiac events. This proportion is consistent with the literature when considering our study data; hs-TropT values exceeding 14 ng/L comprised 16.66% of the patients (reported ranges in the literature vary between 10% and 47%) (15, 23). Sixteen dual (categorical) and five numerical variables were incorporated into the model, derived from dosimetric and clinical parameters. Clinical data were obtained from the electronic hospital record system, while dosimetric data were extracted from the dose-volume histograms (DVHs) of the 3D planned dose distributions. 119 radiomic and 111 dosiomic numerical features were individually extracted for each patient. Gradient-boosted recursive feature elimination, a hybrid method, was used to select important features for prediction.

Clinical and dosimetric variables

The clinical data included patients' age, comorbidities, smoking status, use of cardioprotective medications, administered chemotherapeutic agents, performed surgical procedures, and laterality (right or left breast). The dosimetric data encompassed the dose of radiotherapy delivered to the heart, the presence of a boost dose, fractionation scheme, radiotherapy technique employed, mean and maximum radiation doses received by the heart, and the heart volumes receiving 5 Gy and 25 Gy of radiation (V5 and V25) (2). Chemotherapy-related variables, including the administration of potentially cardiotoxic agents such as anthracyclines, trastuzumab, and taxanes, were incorporated into the model as binary categorical inputs (administered vs. not administered). Due to limitations in the available clinical data, cumulative dose information could not be retrieved and was therefore not used in the analysis.

Medical image preprocessing and segmentation

3D dose distributions, radiotherapy planning CT scans, and organ contours were obtained from the Treatment Planning System (TPS). Before feature extraction, DICOM files from the TPS were converted to NRRD format using the plastimatch library (25). To address variations in slice thickness ranging from 2 mm to 5 mm, both dose

TABLE 1 Machine learning model evaluations.

Models	Model AUC (Gradient Boosted Classification)	5-Fold Cross-Validation Mean AUC	AUC Standard Deviation (Cross-Validation)	Nonparametric Permutation Test p-value (Cross-Validation)
Clinical + Radiomic + Dosiomic	0.96	0.8	0.21	0.021
Clinical + Dosimetric	0.67	0.76	0.28	0.029
Clinical + Dosiomic	0.75	0.66	0.26	0.12
Clinical + Radiomic + Dosimetric	0.67	0.35	0.26	0.84
Clinical	0.89	0.38	0.33	0.77
Clinical + Radiomic	0.5	0.76	0.25	0.06

AUC, Area Under the Curv; p-value, Probability value from the t-test; indicating statistical significance ($p \leq 0{,}05$).

distributions and CT images were resampled to an isotropic 1 mm resolution using B-spline interpolation. No filtering was applied; the original images were included in the model.

Voxel addition was performed for patients who underwent sequential radiotherapy to combine the boost dose with the whole-breast radiotherapy dose distributions using the SimpleITK library (26). During the extraction of these features, heart contours delineated by the patient's physician in clinical practice were used (27). Dose and contouring files were processed in the plastimatch library for masking, and the ITK-SNAP interface was employed for accuracy control and visualization (28). Masking of CT images was conducted using the 3D Slicer interface (29). Patients with significant metal artifacts in their CT images were excluded from the study.

Extraction of radiomic and dosiomic features

Radiomic features were extracted from 3D CT images using heart regions previously contoured by patients' physicians during routine clinical practice, employing the 3D Slicer interface for segmentation, resampling, and feature extraction. Dose distributions were isotropically resampled to 1 mm within the Treatment Planning System, utilizing a cumulative dose intensity range of 1 Gy for dosiomic texture features and a voxel intensity bin width of 64 Hounsfield Units for radiomics. Features were extracted using the PyRadiomics library under eight categories, including First-Order Statistics, Shape Features, and various Gray Level Matrices (GLCM, GLDM, GLSZM, GLRLM, NGTDM), and are consistent with the Image Biomarker Standardisation Initiative (IBSI) guidelines (30).

Model building and evaluation

Machine learning model selection and hyperparameter optimization were performed using the Tree-based Pipeline Optimization Tool (TPOT), an automated machine learning library in Python that utilizes genetic programming and evolutionary algorithms to identify optimal models and data processing pipelines (31, 32). The TPOTClassifier automatically explores a wide range of supervised classifiers including linear models (e.g., logistic regression), naïve Bayes variants, decision trees, ensemble methods (e.g., random forest, gradient boosting, XGBoost), and support vector machines. The model search is performed via genetic programming to select optimal preprocessing steps, feature selection, classifier type, and hyperparameters. Gradient-boosted classification yielded the highest validation AUC among the tested pipelines and was thus selected as the final model. Model performance was assessed using the area under the curve (AUC) method on data classified according to the hs-TropT> 14 ng/L threshold. To evaluate the model's generalizability, we employed 5-fold cross-validation (33), and a non-parametric permutation test was conducted to assess the model's robustness against randomness (34).

Results

Patient characteristics

42 patients (training: 33, validation: 9) were enrolled in this study. The characteristics of the patients are presented in Table 2. This was presented concerning hs-Troponin T levels, divided into four categories: those with hs-Troponin T levels above 14 ng/L and those below 14 ng/L, both with and without elevated clinical parameters.

Dosiomic features and dose-volume factors

The features extracted from the original dose distributions in the treatment planning system through advanced statistical methods included first-order statistics (19), 2D shape features (16), 3D shape features (10), Gray Level Co-occurrence Matrix (GLCM) features (24), Gray Level Dependence Matrix (GLDM) features (16), Gray Level Size Zone Matrix (GLSZM) features (16), Gray Level Run Length Matrix (GLRLM) features (5), and

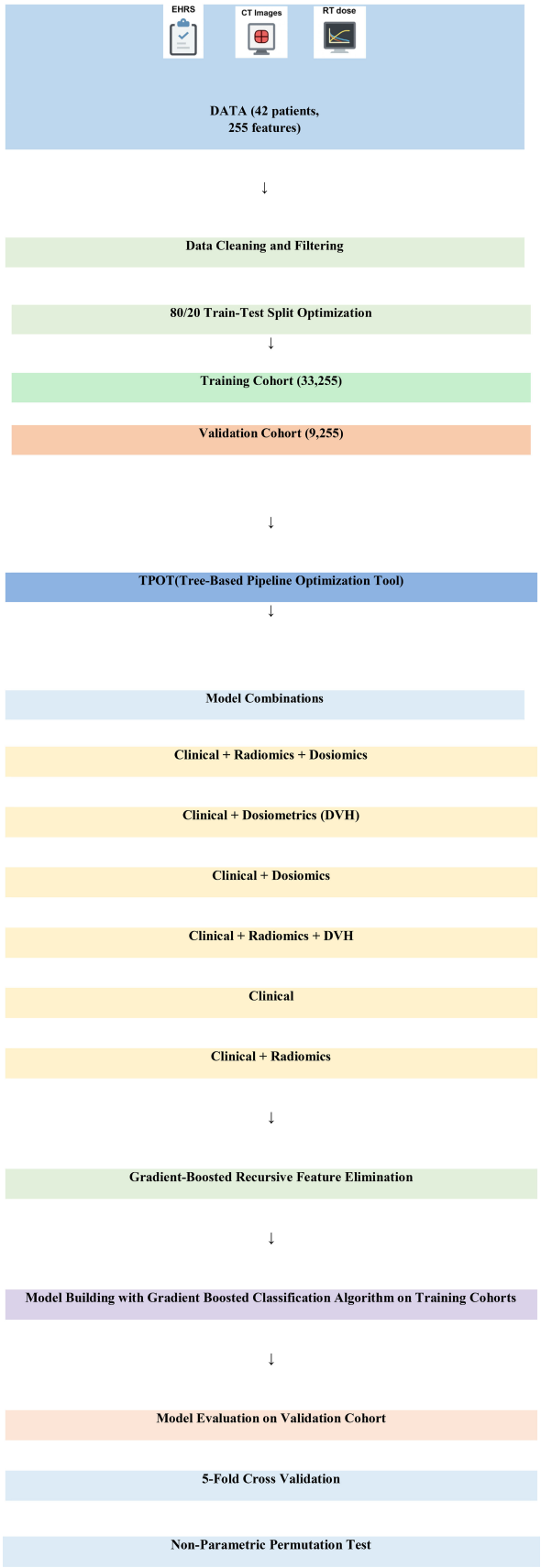

FIGURE 2
Overall workflow of the study. DVH, dose-volume histogram; EHRS, electronic hospital record system.

TABLE 2 Clinical characteristics.

Clinical/Dosimetric Variable	Trop-T > 14 ng/L (+)	Trop-T ≤ 14 ng/L (+)	Trop-T > 14 ng/L (-)	Trop-T ≤ 14 ng/L (-)
Hypertension	3 (7%)	8 (19%)	4 (10%)	27 (64%)
Diabetes Mellitus	4 (10%)	6 (14%)	3 (7%)	29 (69%)
Smoking	1 (3%)	9 (21%)	3 (7%)	26 (62%)
Hyperlipidemia	1 (3%)	3 (7%)	6 (14%)	32 (76%)
ACE-i	0	1 (3%)	7 (17%)	34 (80%)
ARB	3 (7%)	6 (14%)	4 (10%)	29 (69%)
B-blocker	3 (7%)	3 (7%)	6 (14%)	30 (72%)
Trastuzumab	1 (3%)	6 (14%)	1 (3%)	34 (80%)
Cyclophosphamide	5 (12%)	24 (57%)	7 (17%)	11 (26%)
5FU/Capecitabine	0	0	3 (7%)	44 (93%)
Anthracycline	3 (7%)	27 (64%)	4 (10%)	24 (57%)
Taxane	5 (12%)	27 (64%)	4 (10%)	22 (52%)
Surgery (MRM)	5 (12%)	12 (28%)	4 (10%)	23 (55%)
RT Dose (2 Gy × 15)	4 (10%)	18 (43%)	8 (19%)	12 (28%)
Boost Dose (2 Gy × 5)	3 (7%)	12 (28%)	4 (10%)	23 (55%)
RT Technique (IMRT)	2 (5%)	19 (45%)	5 (12%)	16 (38%)

(+) and (-) indicate the presence and absence of the clinical condition or treatment, respectively; Trop-T, Troponin-T; ACE-i, Angiotensin Converting Enzyme Inhibitors; ARB, Angiotensin Receptor Blocker; MRM, Modified Radical Mastectomy; IMRT, Intensity Modulated Radiation Therapy.

Neighboring Gray Tone Difference Matrix (NGTDM) features (5), resulting in a total of 111 dosiomic based features calculated for each patient. The dosimetric features extracted from the heart contour are shown in Table 3.

Radiomic features

The features extracted from the CT scans taken for RT planning through advanced statistical methods included first-order statistics (18), 2D shape features (16), 3D shape features (10), Gray Level Co-occurrence Matrix (GLCM) features (24), Gray Level Dependence Matrix (GLDM) features (16), Gray Level Size Zone Matrix (GLSZM) features (16), Gray Level Run Length Matrix (GLRLM) features (5), and Neighboring Gray Tone Difference Matrix (NGTDM) features (14), resulting in a total of 119 radiomic features calculated for each patient.

Univariate analysis

The chi-square test was used for univariate analysis of clinical-dosimetric categorical variables (Table 2). All variables were statistically non-significant for an association with elevated cardiac troponin ($p>0.05$). The independent samples t-test results indicated that the mean age was significantly higher in individuals with elevated troponin T levels ($p = 0.044$). Specifically, the mean age was 65.43 ± 14.57 years for those with Troponin T levels >14 ng/dL, compared to 55.31 ± 11.21 years for those with Troponin T levels ≤ 14 ng/dL. Additionally, an elevated heart V5 value was found to be borderline statistically significant in association with elevated troponin T levels among the dosimetric variables ($p = 0,05$) (Table 3). Recursive feature elimination was performed on six different models. The top five essential features extracted from the best-performing model (clinical + dosiomics + radiomics) were analyzed for their univariate association with troponin T elevation, as shown in Table 4.

Prediction performance

The prediction models were evaluated using Gradient-Boosted Classification, with 5-fold cross-validation and non-parametric permutation tests to assess generalizability and model randomness. A total of six different models incorporating clinical, radiomic, dosiomic, and dosimetric parameters were constructed. The combined model of clinical, radiomic, and dosiomic parameters yielded the highest predictive performance, achieving an AUC of 0.96 for the validation cohort. This value was much lower for the clinical + dosimetric model (validation cohort-AUC = 0.67) (Figure 3). The internal validation using 5-fold cross-validation showed that the generalizability performance of the clinical + dosiomic + radiomic model was fair-to-good (mean AUC = 80.33 ± 21%). However, the relatively high standard deviation suggests considerable variability across folds, which may reflect model instability and potential

10.3389/fonc.2025.1557382

TABLE 3 Dosimetric variables.

Dosimetric Features	Trop-T ≤ 14 (mean±std)	Trop-T > 14 (mean±std)	T-test p-value
Heart/Dmean	3.06±1.85	3.91±1.99	0.27
Heart/Dmax	31.26±17.76	30.76±13.96	0.94
Heart/V5	12.05±11.48	22.33±16.42	0.05
Heart/V25	2.53±3.21	1.28±1.85	0.32

Dmean, Mean dose; Dmax, Maximum dose; V5, Volume of the heart receiving 5 Gy or more; V25, Volume of the heart receiving 25 Gy or more; Trop-T, Troponin-T; mean±std, Mean value ± standard deviation; p-value, Probability value from the t-test, indicating statistical significance (p ≤ 0,05).

TABLE 4 Most important features extracted from the most effective model (clinical + dosiomics + radiomics).

Biomarker	Classes	Features	T-test (Target = Troponin-T)
Dosiomic	First-order Statistics	Maximum	P<0.001
Radiomic	First-order Statistics	Minimum	P= 0.002
Dosiomic	First-order Statistics	10th Percentile	P= 0.406
Radiomic	Gray-Level Run Length Matrix	Run Entropy	P=0.159
Radiomic	First-order Statistics	Maximum	P=0.88

overfitting due to the limited sample size. These two models were non-coincidences based on confirmation from nonparametric permutation tests (p<0,05). Other models were considered random based on permutation testing (p>0,05) (Table 1).

Discussions

Traditional clinical studies in oncology typically rely on long-term follow-up and clinical, radiological, and dosimetric data. However, over the past decade, advancements in genomics and biomarker discovery have positioned personalized medicine as a central focus of scientific research. In 2017, the U.S. National Cancer Institute emphasized the potential of using biomarkers to evaluate the effects of radiotherapy on tissues, highlighting their capacity to revolutionize the assessment of treatment efficacy and average tissue damage (35). Despite the advances in precision medicine, predicting adverse events and disease prognosis using biomarkers remains an unmet need in medical research. This study focuses on predicting cardiac events that may arise as a result of breast cancer treatment before they become clinically evident. We utilized cardiac biomarkers, such as high-sensitivity Troponin T (hs-TropT), and imaging biomarkers derived from advanced statistical analyses of radiological images and dose distribution data (radiomics and dosiomics).

According to the literature, cardiovascular causes account for a significant proportion of deaths in breast cancer survivors (36, 37). Therefore, early prediction of cardiac events is essential for tailoring treatment and follow-up strategies in this patient population. While echocardiography remains the standard non-invasive method for detecting cancer therapy-related cardiac dysfunction (CTRCD) (7, 12), several studies suggest that it may not detect subclinical myocardial injury in its early stages, which can lead to missed opportunities for intervention (13, 14). The European Society of Cardiology has recommended that evaluating serum cardiac biomarkers, such as Troponin-T (Trop-T) and NT-proBNP, both at baseline and during or after treatment, may serve as a valuable diagnostic tool in the detection of subclinical cancer therapy-related cardiac dysfunction (CTRCD) and in tailoring oncological therapies based on individual risk profiles (16). Our findings corroborate previous reports, emphasizing the significance of imaging biomarkers like radiomics and dosiomics in improving early detection rates of such adverse events.

Our findings may emphasize the significant potential of models based on imaging biomarkers for early predicting cardiac events related to breast cancer treatment, compared to traditional clinical and dosimetric features. Furthermore, this may highlight the importance of a multidisciplinary approach, demonstrating that integrating clinical, dosimetric, imaging, and cardiological data can

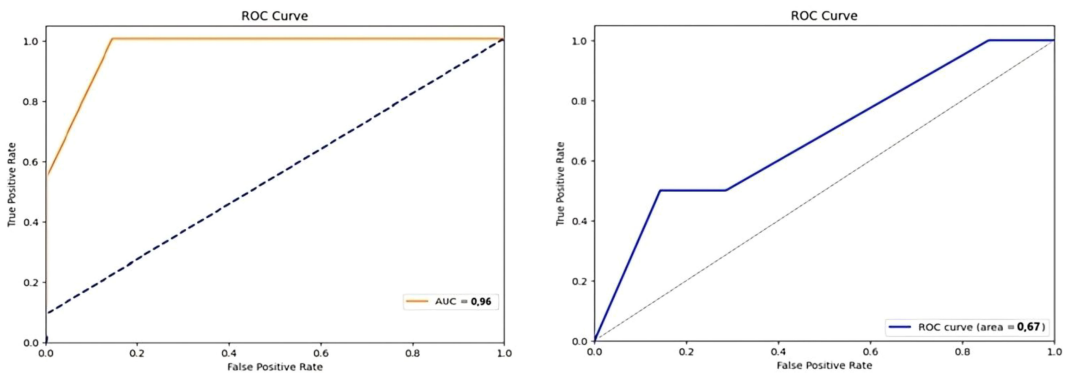

FIGURE 3
Receiver operating characteristic (ROC) curves for prediction models. Left: ROC curve for the clinical + radiomic + dosiomic model with an Area Under the Curve (AUC) of 0.96. Right: ROC curve for the clinical + dosimetric model with an AUC of 0.67. ROC, Receiver Operating Characteristic; AUC, Area Under the Carve.

contribute to better management and personalization of patient treatment processes. The results of this study could provide a valuable foundation for future research focused on breast cancer treatment and side effect management.

Extracting three-dimensional (3D) radiomic and dosiomic features from imaging data provides a more comprehensive and objective assessment than traditional two-dimensional approaches. This study used PyRadiomics, an open-source Python library, to extract 3D features from medical images. While various methods can address calibration and noise issues in multi-center imaging studies, our images were acquired from a single center, minimizing these challenges. Consequently, we extracted radiomic features from the original planning CT images without additional filtering.

The only previous study utilizing dosiomics to investigate cardiac toxicity was conducted in a pediatric cohort to predict cardiac valvulopathy (38). This study is the first to develop a predictive model for adult breast cancer patients by integrating clinical, dosimetric, radiomic, and dosiomic data. The strength of this study is that no previous studies have specifically employed segmented cardiac dosiomic data in adult patients, underscoring the novelty of our research and its potential to serve as a foundation for future investigations.

Despite these promising results, our study has several limitations. The cumulative dose of cardiotoxic agents, particularly anthracyclines, is known to be strongly associated with cancer therapy-related cardiac dysfunction (CTRCD), with a recommended cumulative dose threshold of 550 mg/m². In our study, chemotherapy-related variables—including the use of anthracyclines, trastuzumab, and taxanes—were encoded as binary categorical inputs (administered vs. not administered). However, detailed cumulative dosing data were not available. This limitation may have hindered the predictive capacity of the clinical model regarding treatment-related cardiac toxicity (39, 40).

Moreover, the relatively small sample size, particularly the 9-patient validation cohort, presents a limitation regarding model generalizability and the potential risk of overfitting. As this study was conducted within the scope of a completed and ethically approved academic thesis, adding new patients retrospectively or implementing synthetic data augmentation techniques (e.g., SMOTE) was not feasible. Therefore, while the findings are promising, they should be interpreted as exploratory and hypothesis-generating, warranting validation in larger prospective cohorts.

This limitation was also reflected in the internal validation results. The relatively high standard deviation in AUC values across the cross-validation folds (80.33 ± 21%) indicates potential model instability and overfitting risk. Such variability is commonly observed in small datasets, and highlights the importance of future validation using larger, independent cohorts or more robust validation schemes such as repeated or nested cross-validation. Future studies should employ repeated or nested cross-validation frameworks to enhance the reliability of performance estimates, especially in datasets with limited sample sizes.

Small-sample machine learning studies are increasingly reported in emerging fields such as cardio-oncology, where access to large annotated datasets is limited. Despite their limitations, such studies often play a critical role in hypothesis generation and feasibility assessment, especially when novel biomarkers or data types are explored. In our case, the integration of heart-segmented dosiomics represents a unique and innovative dimension that justifies the use of a pilot-scale dataset. We believe that our findings contribute valuable preliminary insights that warrant future validation in larger, prospective studies.

However, dosiomic features were extracted from the original treatment plan's dose distribution, and the model still achieved high performance (AUC = 0.96). Extracting features from Biological Effective Dose (BED) distributions could be considered as an alternative approach. Evaluating the impact of BED-based dosiomic features on model performance warrants further investigation in future studies.

Class imbalance is a significant challenge in medical side-effect prediction studies involving machine learning. In our dataset, the target variable hs-TropT> 14 ng/L represented only 16.66% of the data, resulting in an imbalanced dataset. Class imbalance can cause models to favor the majority class, thereby reducing accuracy in predicting minority class instances. Gradient-boosted classification algorithms help mitigate class imbalance by automatically adjusting weights to define minority classes better, and the AUC metric remains unaffected by class imbalance. However, despite using these methods, performance in other models remained lower, suggesting that alternative techniques (e.g., synthetic data generation or data reduction) may be necessary to address the imbalance, representing a key area for improvement in our study.

Another challenge in radiomic studies is respiratory and cardiac motion, which can reduce the quality of extracted features. Our study did not include patients with controlled breathing; future research could explore this area. Moreover, considering the potential of respiratory control to reduce cardiac events, evaluating its impact on predictive models would be valuable. We also excluded patients with significant metal artifacts in their planning CT scans, as such artifacts could affect both dose distribution and radiomic feature extraction. Investigating prediction models for this patient group remains an area for future research.

Future perspectives

This study highlights the potential of integrating imaging biomarkers such as radiomic and dosiomic features into predictive models for early detection of treatment-related cardiac events in breast cancer patients. However, several areas require further investigation to realize these models' clinical applicability fully. One key area for future research is the inclusion of more precise cumulative anthracycline doses in predictive models. As anthracycline dosage is closely associated with cancer therapy-related cardiac dysfunction (CTRCD), moving beyond categorical variables to incorporate exact dosage values could significantly enhance the predictive accuracy of such models. This will enable

more personalized treatment adjustments, potentially mitigating the risk of cardiac events for high-risk patients.

Another avenue for future research is the exploration of Biological Effective Dose (BED) distributions for dosiomic feature extraction. While our study demonstrated high model performance using traditional dose distributions, BED-based dosiomics may provide an even more accurate representation of the actual biological impact of radiotherapy on cardiac tissues. Evaluating the contribution of BED-based dosiomic features could improve model precision and further tailor treatment plans to individual patient risk profiles.

Moreover, future studies should aim to conduct more extensive, multi-center investigations to validate the generalizability of the proposed models across diverse patient populations and clinical settings. This broader scope would provide more robust evidence and ensure that the models are applicable in varied healthcare environments. Additionally, incorporating the effects of cardiac and respiratory motion into predictive models is essential. Since these motions can impact dose distribution and radiomic feature extraction, addressing these factors in future research could significantly enhance model accuracy and the early detection of subclinical cardiac events.

Lastly, addressing the challenge of class imbalance remains critical in enhancing the robustness of machine learning models in medical side-effect studies. Future studies should explore alternative techniques, such as synthetic data generation or advanced data reduction methods, to mitigate this issue. By overcoming these challenges and continuing to refine predictive models by integrating diverse clinical, imaging, and dosimetric data, future research could significantly improve the prediction and prevention of treatment-related cardiac events, ultimately leading to better patient outcomes and personalized medicine applications.

Conclusions

This study can demonstrate that imaging biomarkers (radiomic and dosiomic parameters) significantly outperform traditional clinical and dosimetric parameters in predicting treatment-related cardiac events in breast cancer patients (96% vs. 67% AUC). When integrated into clinical practice, these radiomic and dosiomic parameters could become critical patient factors, potentially altering treatment and follow-up strategies by predicting cardiac side effects with nearly perfect (%96) accuracy. However, the moderate-to-good generalizability observed in cross-validation suggests that these models need to be supported by real-world data. This underscores the necessity for further research to integrate imaging biomarkers, which currently require complex statistical processes and have yet to be widely adopted in personalized medicine applications.

Data availability statement

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving humans were approved by The Gazi University Ethics Committee approved the study on April 16, 2024 (Research Code No: 2024-619). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SD: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MA: Project administration, Resources, Supervision, Writing – review & editing. OA: Data curation, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The authors thank the Gazi University Department of Radiation Oncology team.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. We utilized ChatGPT to assist in language editing and improving the manuscript's clarity. All scientific content and conclusions were authored and reviewed by the study team.

Publisher's note

References

1. Broder H, Gottlieb RA, Lepor NE. Chemotherapy and cardiotoxicity. *Rev Cardiovasc Med*. (2008) 9:75. doi: 10.1016/s1470-2045(08)70003-2

2. 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. *New Engl J Med*. (2013) 368:987–98. doi: 10.1056/NEJMoa1209825

3. Baillet F, Housset M, Maylin C, Boisserie G, Bettahar R, Delanian S, et al. The use of a specific hypofractionated radiation therapy reg imen versus classical fractionation in the treatment of breast cancer: A randomized study of 230 patients. *Int J Radiat Oncol Biol Physics*. (1990) 19:1131–3. doi: 10.1016/0360-3016(90)90216-7

4. Wu Q, Bai B, Tian C, Li D, Yu H, Song B, et al. The molecular mechanisms of cardiotoxicity induced by HER2, VEGF, and tyrosine kinase inhibitors: an updated review. *Cardiovasc Drugs Ther*. (2022) 36(3):511–24. doi: 10.1007/s10557-021-07181-3

5. Sara JD, Kaur J, Khodadadi R, Rehman M, Lobo R, Chakrabarti S, et al. 5-fluorouracil and cardiotoxicity: a review. *Ther Adv Med Oncol*. (2018) 10:1758835918780140. doi: 10.1177/1758835918780140

6. Romond EH, Jeong J-H, Rastogi P, Swain SM, Geyer CE Jr., Ewer MS, et al. Seven-year follow-up assessment of cardiac function in NSABP B-31, a randomized trial comparing doxorubicin and cyclophosphamide followed by paclitaxel (ACP) with ACP plus trastuzumab as adjuvant therapy for patients with node-positive, human epidermal growth factor receptor 2–positive breast cancer. *J Clin Oncol*. (2012) 30:3792. doi: 10.1200/JCO.2011.40.0010

7. Leong DP, Lenihan DJ. Clinical practice guidelines in cardio-oncology. *Heart Failure Clinics*. (2022) 18:489–501. doi: 10.1016/j.hfc.2022.02.002

8. Zamorano JL, Lancellotti P, Rodriguez Munoz D, Aboyans V, Asteggiano R, Galderisi M, et al. 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. doi: 10.1093/eurheartj/ehw211

9. Zhang P, Hu X, Yue J, Meng X, Han D, Sun X, et al. Early detection of radiation-induced heart disease using 99mTc-MIBI SPECT gated myocardial perfusion imaging in patients with oesophageal cancer during radiotherapy. *Radiotherapy Oncol*. (2015) 115:171–8. doi: 10.1016/j.radonc.2015.04.009

10. Ferreira de Souza T, Quinaglia AC Silva T, Osorio Costa F, Shah R, Neilan TG, Velloso L, et al. Anthracycline therapy is associated with cardiomyocyte atrophy and preclinical manifestations of heart disease. *JACC: Cardiovasc Imaging*. (2018) 11:1045–55. doi: 10.1016/j.jcmg.2018.05.012

11. Galán-Arriola C, Lobo M, Vílchez-Tschischke JP, López GJ, de Molina-Iracheta A, Pérez-Martínez C, et al. Serial magnetic resonance imaging to identify early stages of anthracycline-induced cardiotoxicity. *J Am Coll Cardiol*. (2019) 73:779–91. doi: 10.1016/j.jacc.2018.11.046

12. Curigliano G, Lenihan D, Fradley M, Ganatra S, Barac A, Blaes A, et al. Management of cardiac disease in cancer patients throughout oncological treatment: ESMO consensus recommendations. *Ann Oncol*. (2020) 31:171–90. doi: 10.1016/j.annonc.2019.10.023

13. Curigliano G, Cardinale D, Dent S, Criscitiello C, Aseyev O, Lenihan D, et al. Cardiotoxicity of anticancer treatments: epidemiology, detection, and management. *CA: Cancer J Clin*. (2016) 66:309–25. doi: 10.3322/caac.21341

14. Moudgil R, Hassan S, Palaskas N, Lopez-Mattei J, Banchs J, Yusuf SW. Evolution of echocardiography in subclinical detection of cancer therapy–related cardiac dysfunction. *Echocardiography*. (2018) 35:860–8. doi: 10.1111/echo.14012

15. Lv X, Pan C, Guo H, Chang J, Gao X, Wu X, et al. Early diagnostic value of high-sensitivity cardiac troponin T for cancer treatment-related cardiac dysfunction: a meta-analysis. *ESC Heart Failure*. (2023) 10:2170–82. doi: 10.1002/ehf2.14373

16. Pudil R, Mueller C, Čelutkienė J, Henriksen PA, Lenihan D, Dent S, et al. Role of serum biomarkers in cancer patients receiving cardiotoxic cancer therapies: a position statement from the Cardio-Oncology Study Group of the Heart Failure Association and the Cardio-Oncology Council of the European Society of Cardiology. *Eur J Heart failure*. (2020) 22:1966–83. doi: 10.1002/ejhf.2017

17. Cardinale D, Sandri MT, Colombo A, Colombo N, Boeri M, Lamantia G, et al. Prognostic value of troponin I in cardiac risk stratification of cancer patients undergoing high-dose chemotherapy. *Circulation*. (2004) 109:2749–54. doi: 10.1161/01.CIR.0000130926.51766.CC

18. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, et al. Dosiomics: extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis. *Front Oncol*. (2019) 9:269. doi: 10.3389/fonc.2019.00269

19. Ren W, Liang B, Sun C, Wu R, Men K, Xu Y, et al. Dosiomics-based prediction of radiation-induced hypothyroidism in nasopharyngeal carcinoma patients. *Physica Medica*. (2021) 89:219–25. doi: 10.1016/j.ejmp.2021.08.009

20. Wu A, Li Y, Qi M, Lu X, Jia Q, Guo F, et al. Dosiomics improves prediction of locoregional recurrence for intensity modulated radiotherapy treated head and neck cancer cases. *Oral Oncol*. (2020) 104:104625. doi: 10.1016/j.oraloncology.2020.104625

21. Lee SH, Han P, Hales RK, Voong KR, Noro K, Sugiyama S, et al. Multi-view radiomics and dosiomics analysis with machine learning for predicting acute-phase weight loss in lung cancer patients treated with radiotherapy. *Phys Med Biol*. (2020) 65:195015. doi: 10.1088/1361-6560/ab8531

22. Jin X, Zheng X, Chen D, Jin J, Zhu G, Deng X, et al. Prediction of response after chemoradiation for esophageal cancer using a combination of dosimetry and CT radiomics. *Eur radiology*. (2019) 29:6080–8. doi: 10.1007/s00330-019-06193-w

23. Demissei BG, Hubbard RA, Zhang L, Smith AM, Sheline K, McDonald C, et al. Changes in cardiovascular biomarkers with breast cancer therapy and associations with cardiac dysfunction. *J Am Heart Assoc*. (2020) 9:e014708. doi: 10.1161/JAHA.119.014708

24. Petricciuolo S, Delle Donne MG, Aimo A, Chella A, De Caterina R. Pre-treatment high-sensitivity troponin T for the short-term prediction of cardiac outcomes in patients on immune checkpoint inhibitors. *Eur J Clin Invest*. (2021) 51:e13400. doi: 10.1111/eci.13400

25. Sharp GC, Li R, Wolfgang J, Chen G, Peroni M, Spadea MF, et al. Plastimatch: an open source software suite for radiotherapy image processing. In: *Proceedings of the XVI'th international conference on the use of computers in radiotherapy (ICCR)*. Amsterdam, Netherlands (2010). Vol. 3.

26. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of simpleITK. *Front neuroinformatics*. (2013) 7:45. doi: 10.3389/fninf.2013.00045

27. Kong FM, Quint L, Machtay M, Bradley J. *Atlases for organs at risk (OARs) in thoracic radiation therapy*. Radiation Therapy Oncology Group (RTOG) (2010). Available online at: https://www.eviq.org.au/getmedia/a4c012a8-d6a7-4d87-93f7-d3f465b49889/RTOG-heart-contouring-atlas. pdf. aspx.

28. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015

29. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*. (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001

30. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *arXiv preprint arXiv:161207003*. (2016). doi: 10.48550/arXiv.1612.07003

31. Bishop CM. Pattern recognition and machine learning. *Springer google schola*. (2006) 2:645–78. doi: 10.1117/1.2819119

32. Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*. (2020) 36:250–6. doi: 10.1093/bioinformatics/btz470

33. Garcia S, Herrera F. An Extension on" Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. *J Mach Learn Res*. (2008) 9:2677–94. Available online at: http://www.jmlr.org/papers/v9/garcia08a.html.

34. Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, et al. Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. *J Thorac Oncol*. (2017) 12:467–76. doi: 10.1016/j.jtho.2016.11.2226

35. Ahmed MM, Coleman CN, Mendonca M, Bentzen S, Vikram B, Seltzer SM, et al. Workshop report for cancer research: Defining the shades of Gy: Utilizing the biological consequences of radiotherapy in the development of new treatment approaches—meeting viewpoint. *Cancer Res*. (2018) 78:2166–70. doi: 10.1158/0008-5472.CAN-17-3760

36. Bikiewicz A, Banach M, von Haehling S, Maciejewski M, Bielecka-Dabrowa A. Adjuvant breast cancer treatments cardiotoxicity and modern methods of detection and prevention of cardiac complications. *ESC Heart Failure*. (2021) 8:2397–418. doi: 10.1002/ehf2.13365

37. Abdel-Qadir H, Austin PC, Lee DS, Amir E, Tu JV, Thavendiranathan P, et al. A population-based study of cardiovascular mortality following early-stage breast cancer. *JAMA Cardiol*. (2017) 2:88–93. doi: 10.1001/jamacardio.2016.3841

38. Chounta S, Allodji R, Vakalopoulou M, Bentriou M, Do DT, De Vathaire F, et al. Dosiomics-based prediction of radiation-induced valvulopathy after childhood cancer. *Cancers*. (2023) 15:3107. doi: 10.3390/cancers15123107

39. Manrique CR, Park M, Tiwari N, Plana JC, Garcia MJ. Diagnostic strategies for early recognition of cancer therapeutics–related cardiac dysfunction. *Clin Med Insights: Cardiol*. (2017) 11:1179546817697983. doi: 10.1177/1179546817697983

40. Von Hoff DD, Layard MW, Basa P, Davis HL Jr, Von Hoff AL, Rozencweig M, et al. Risk factors for doxorubicin-lnduced congestive heart failure. *Ann Internal Med*. (1979) 91:710–7. doi: 10.7326/0003-4819-91-5-710

Glossary

ACE-i	Angiotensin Converting Enzyme Inhibitors
ARB	Angiotensin Receptor Blocker
AUC	Area Under the Curve
BED	Biological Effective Dose
CTRCD	Cancer Therapy-Related Cardiac Dysfunction
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
DVH	Dose-Volume Histogram
Dmax	Maximum Dose;
Dmean	Mean Dose
EHRS	Electronic Hospital Record System
GLCM	Gray Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
hs-TropT	High-Sensitivity Troponin-T
IBSI	Image Biomarker Standardization Initiative
IMRT	Intensity Modulated Radiation Therapy
ITK	Insight Segmentation and Registration Toolkit
LVEF	Left Ventricular Ejection Fraction
MRM	Modified Radical Mastectomy
NGTDM	Neighboring Gray Tone Difference Matrix
NT-proBNP	N-Terminal Pro B-Type Natriuretic Peptide
PET	Positron Emission Tomography
PyRadiomics	Python-based Radiomics Extraction Library
ROC	Receiver Operating Characteristic
RT	Radiotherapy
TPS	Treatment Planning System
TPOT	Tree-Based Pipeline Optimization Tool
V5	Volume of the Heart Receiving 5 Gy or More
V25	Volume of the Heart Receiving 25 Gy or More