



OPEN Machine learning-driven imaging data for early prediction of lung toxicity in breast cancer radiotherapy

Tamás Ungvári¹✉, Döme Szabó¹, András Gyórfi², Zsófia Dankovics¹, Balázs Kiss¹, Judit Olajos^{3,4} & Károly Tőkési⁵

One possible adverse effect of breast irradiation is the development of pulmonary fibrosis. The aim of this study was to determine whether planning CT scans can predict which patients are more likely to develop lung lesions after treatment. A retrospective analysis of 242 patient records was performed using different machine learning models. These models showed a remarkable correlation between the occurrence of fibrosis and the hounsfield units of lungs in CT data. Three different classification methods (Tree, Kernel-based, k-Nearest Neighbors) showed predictive values above 60%. The human predictive factor (HPF), a mathematical predictive model, further strengthened the association between lung hounsfield unit (HU) metrics and radiation-induced lung injury (RILI). These approaches optimize radiation treatment plans to preserve lung health. Machine learning models and HPF can also provide effective diagnostic and therapeutic support for other diseases.

Keywords Artificial intelligence, Breast cancer, Radiotherapy, Machine learning, Lung injury, CT image

Despite notable advances in therapeutic strategies that have considerably reduced mortality rates, breast cancer remains the most prevalent malignancy among women. According to the National Cancer Institute's SEER program, the five-year survival rate now exceeds 90%, emphasizing the need to mitigate treatment-related side effects to maintain patients' quality of life^{1,2}. Radiotherapy is a fundamental element of breast cancer management, with most patients undergoing this treatment.

Radiotherapy planning relies on non-contrast computed tomography (CT) images, which provide indispensable data for dose calculations, including tissue density. The Hounsfield Unit (HU) is a fundamental parameter in CT interpretation, quantifying tissue density based on X-ray attenuation. HU values are calculated through a linear transformation of attenuation coefficients, with distilled water defined as 0 HU, air as -1000 HU, and denser materials like bone exhibiting higher positive values^{3,4}.

Chest radiotherapy can induce interstitial lung damage. Radiation-induced lung injury (RILI) is a dose-limiting toxicity in thoracic radiotherapy, characterized by a progressive sequence of inflammatory, damaging, and fibrotic responses in lung tissue⁵. Ionizing radiation affects both alveolar epithelial cells and the pulmonary vasculature, leading to the release of pro-inflammatory cytokines and the subsequent activation of fibroblasts and other mesenchymal cells. These changes drive extracellular matrix remodeling and excessive collagen deposition, culminating in radiation-induced fibrosis. The risk and severity of RILI depend on various clinical and dosimetric factors, including total radiation dose, irradiated lung volume, pre-existing pulmonary conditions, smoking history, and overall patient health. The mechanisms contributing to RILI involve inflammation and fibrosis through processes such as alveolar damage, reactive oxygen species (ROS) toxicity, and immune-mediated damage^{6–14}. Pulmonary fibrosis can significantly impair lung function, resulting in chronic breathlessness, reduced physical activity, and long-term deterioration in quality of life.

The application of artificial intelligence (AI) and deep learning to medical imaging, particularly in the diagnosis of lung disease, has seen significant advances in recent years. Zarei et al. (2024) performed quantitative analysis on native chest CT images to accurately measure lung lesions, aiding diagnosis and treatment monitoring¹⁵.

¹Markusovszky University Teaching Hospital, Markusovszky str. 5, Szombathely 9700, Hungary. ²Miskolci SZC Bláthy Ottó Villamosipari Technikum, Soltész Nagy Kálmán str 7, Miskolc 3527, Hungary. ³Jósa András University Teaching Hospital, Szent István str. 68, Nyíregyháza 4400, Hungary. ⁴University of Nyíregyháza, Sóstói str.31/B, Nyíregyháza 4400, Hungary. ⁵HUN-REN Institute for Nuclear Research, Bem square 18/c, Debrecen 4026, Hungary. ✉email: ungvari.tamas@markusovszky.hu

Parameter	RILI status	Mean	Std. Dev	Median	Min	Max	p-value	Effect size (r)
HU minimum	No	-986.39	23.48	-990	-1043	-882	0.123	0.10
	Yes	-978.62	29.71	-988	-1020	-866		
HU maximum	No	242.96	152.25	229	29	840	0.123	0.25
	Yes	266.32	176.29	247	47	1275		
HU mean	No	-746.11	64.13	-749.11	-869.28	-430.38	0.001	0.23
	Yes	-716.03	78.18	-714.68	-846.76	-140.36		
HU STD	No	142.78	18.65	136.45	116.66	192.08	0.001	0.21
	Yes	148.95	17.63	146.47	114.81	195.43		

Table 1. Summary of Hounsfield unit (HU) parameters, including minimum and maximum mean values, standard deviation (SD), and lung volumes, among patients with and without radiation-induced lung injury (RILI). The mean, SD, minimum, and maximum values are presented.

Parameter	RILI status	Mean	Std. Dev	Median	Min	Max	p-value	Effect size (r)
Lung volume (V cm ³)	No	1680.30	450.75	1667.1	693.6	2964.8	0.003	0.19
	Yes	1532.47	466.57	1453.97	668.1	2794.3		
HPF	No	0.14	0.07	0.13	0.03	0.35	<0.001	0.25
	Yes	0.11	0.05	0.10	0.	0.3		

Table 2. Lung volume (cm³) and human predictive factor (HPF) lung volume and human predictive factor (HPF) values in patients with and without radiation-induced lung injury (RILI). Mean, standard deviation (SD), median, minimum, and maximum values are reported. Statistical significance was assessed using the Mann–Whitney U-test. Patients with RILI had significantly lower lung volumes and HPF values, indicating a potential link between these parameters and radiation-induced fibrosis risk.

In the area of deep learning and medical imaging integration, Huang et al. (2020) developed a comprehensive system that allows the application of deep learning to combine electronic medical records and imaging to improve diagnosis and treatment processes¹⁶. Albers et al. (2023) applied high-resolution propagation-based lung imaging at clinically relevant X-ray dose levels, enabling the study of the fine structure of the lung with low radiation exposure¹⁷. In addition, examining the reliability of CT texture analysis, Adelsmayr et al. (2023) concluded that 3D segmentation and the use of Hounsfield unit thresholds increased the accuracy of texture analysis in the diagnosis of lung lesions¹⁸. These technological advances and artificial intelligence-based applications will contribute to a faster and more accurate diagnosis of lung diseases, as well as personalisation of patient care and treatment, helping to improve treatment efficiency and increase patient survival. Advanced machine learning techniques, including convolutional neural networks (CNNs) and texture-based classification methods, have significantly enhanced the characterization of lung diseases^{19–29}. In this work the pulmonary fibrosis in breast cancer radiotherapy was investigated, focusing on HU of lungs metrics (minimum, maximum, mean, and standard deviation) and lung volume changes. Three machine learning models and a human predictive factor (HPF) were utilized to predict radiation-induced lung injury (RILI). The HPF is a score derived from Hounsfield unit and lung volume measurements, representing an individual's risk of fibrosis through a simple method. A retrospective analysis of 242 patients was conducted to evaluate these parameters and their implications for treatment-related side effects. Among the 242 patients, 113 (46.7%) exhibited detectable radiation-induced lung damage, while 129 (53.3%) showed no radiologically visible lung damage. The study further categorized patients based on the presence or absence of visible lesions, distinguishing between those with lesions and those without^{30,31}.

Results

To investigate the association between CT parameters, lung volume and RILI, three machine learning models (Fine Tree, Kernel-based and k-Nearest Neighbours [kNN]) were used together with a human-generated predictive factor (HPF). The lung volumes (Vcm³) of patients, along with the minimum and maximum, mean, standard- deviation of the Hounsfield unit of the affected side lung with and without RILI, were assessed using planning CT scans. The data employed for the construction of the models and the subsequent statistical evaluation are presented in tabular form in (Table 1). Additional comparative data on lung volume and HPF are included in (Table 2). The Mann-Whitney U-test indicated that the mean HU parameter was significantly higher in patients who developed RILI (median = -714.68) compared to those who did not (median = -749.11), $p=0.001$, $r=0.23$ (medium effect). Additionally, the HU standard deviation (SD) scores were higher in the RILI group (median = 146.47) than in the non-RILI group (median = 136.45), $p=0.001$, $r=0.21$ (medium effect). However, no statistically significant difference was observed in HU min and HU max values ($p=0.123$).

Lung volumes were significantly lower in the RILI group (median = 1453.97 cm³) compared to the control group (median = 1667.1 cm²), $p = 0.003$, $r = 0.19$ (small effect). Additionally, human predictive factor (HPF) values were lower in patients with fibrosis (median = 0.10 vs. 0.13), with a statistically significant difference, $p < 0.001$, $r = 0.25$ (medium effect).

The HPF is a predictive factor that estimates the probability of developing pulmonary fibrosis based on a number of variables (see Eq. (1)).

The algorithms were trained on 159 samples and validated on 83 samples using five-fold cross-validation. The human-generated mathematical predictive factor (HPF) model displays a slight reduction in accuracy relative to the AI models. Conversely, the straightforward formula facilitates greater ease of use.

Table 3 provides an overview of the accuracy of the models, while Fig. 1 illustrates the test and validation of the models and Fig. 2 shows ROC curves of HPF.

Discussion

The application of diverse machine learning algorithms in the medical domain has witnessed a surge in popularity^{32–36}.

The statistical analyses, driven by advanced AI models, provided robust evidence of a correlation between RILI and CT-based lung volume parameters. This relationship was further supported by the Human Predictive Factor (HPF), demonstrating that variations in lung Hounsfield Unit (HU) values are strongly associated with fibrosis development.

The significant correlation between lung volume and HU mean (e.g. HU > -720 for fibrosis risk) suggests that these parameters may reflect not only the presence of fibrosis but also the overall health of the lungs. This finding may encourage wider use of radiological data in the prediction of other lung diseases^{15,37–39}. In their study, O’Callaghan and colleagues showed that chest CT correlates with the function and radiological features of idiopathic pulmonary fibrosis (IPF) and may serve as a potential biomarker for assessing the severity of IPF disease. Our results on radiation-induced lung injury agree with the above study, where a similar correlation was observed. In order to analyse the radiation-induced changes in Hounsfield Units, Wuschner and colleagues demonstrated in their study that radiation three months after RT causes changes in lung anatomy that exhibit a strong linear correlation with dose. The observed changes in Hounsfield units in the vascular lung parenchyma suggest that this measure may be a potential biomarker of changes in perfusion⁴⁰.

The present study did not investigate the effect of dose on lung tissue. Instead, it focused on lung densitometry based on HU value and lung volume to determine whether this is a biomarker. By comparing the results, it can be confirmed that there may be significant changes in lung HU values before and after radiation treatment. The radiogenic lesion of the lung depends on a number of factors, including the dose received, the lung condition, and the ability to regenerate. Further development of the models is possible, as are tests of other models, for example, the model described by Kadeethum et al.⁴¹. It is conceivable that the reconstruction and imaging protocols employed in CT scans may influence the HU values of deep learning models, which could subsequently impact the output values.

The **Fine Tree** model demonstrated a high test accuracy of 83.1%, highlighting its potential in accurately predicting fibrosis risk. However, the validation accuracy was significantly lower at 54.1%, indicating limited generalizability due to overfitting. The model’s strength lies in its interpretability, providing clear decision rules such as: “If HU mean > x value, fibrosis is likely to develop.” This rule-based structure may aid clinicians in making rapid, data-driven decisions, though the model’s over-learning of the training dataset presents challenges for broader clinical applicability.

Optimal Kernel’s deep learning model represents a flexible and robust approach, predominantly utilised for classification tasks.

The Kernel-based model achieved a test accuracy of 81.9% and a validation accuracy of 55.4%. Although the validation performance is modest, the model’s capacity to capture non-linear relationships renders it particularly well-suited to complex datasets. This feature is advantageous when analysing the intricate interplay between HU values, lung volume, and patient-specific factors such as age. While the model’s flexibility is advantageous, its sensitivity to outliers underscores the necessity for robust preprocessing in clinical implementations.

The **kNN** model demonstrated optimal test accuracy (100%), reflecting its strength in adapting to the training data. However, this exceptional performance underscores a significant limitation: overfitting. The model struggles to generalize to unseen data due to its reliance on proximity-based classification. Nonetheless, its simplicity and efficiency make it a valuable tool for small, well-curated datasets.

Model	Accuracy valid %	Accuracy test %
Tree	54.1	83.1
Kernel	55.4	81.9
kNN	62	100
HPF	62.81	72

Table 3. Validation and test accuracy of machine learning models and the human predictive factor (HPF). The accuracy values reflect the models’ ability to predict fibrosis risk based on CT-derived parameters and lung volumes.

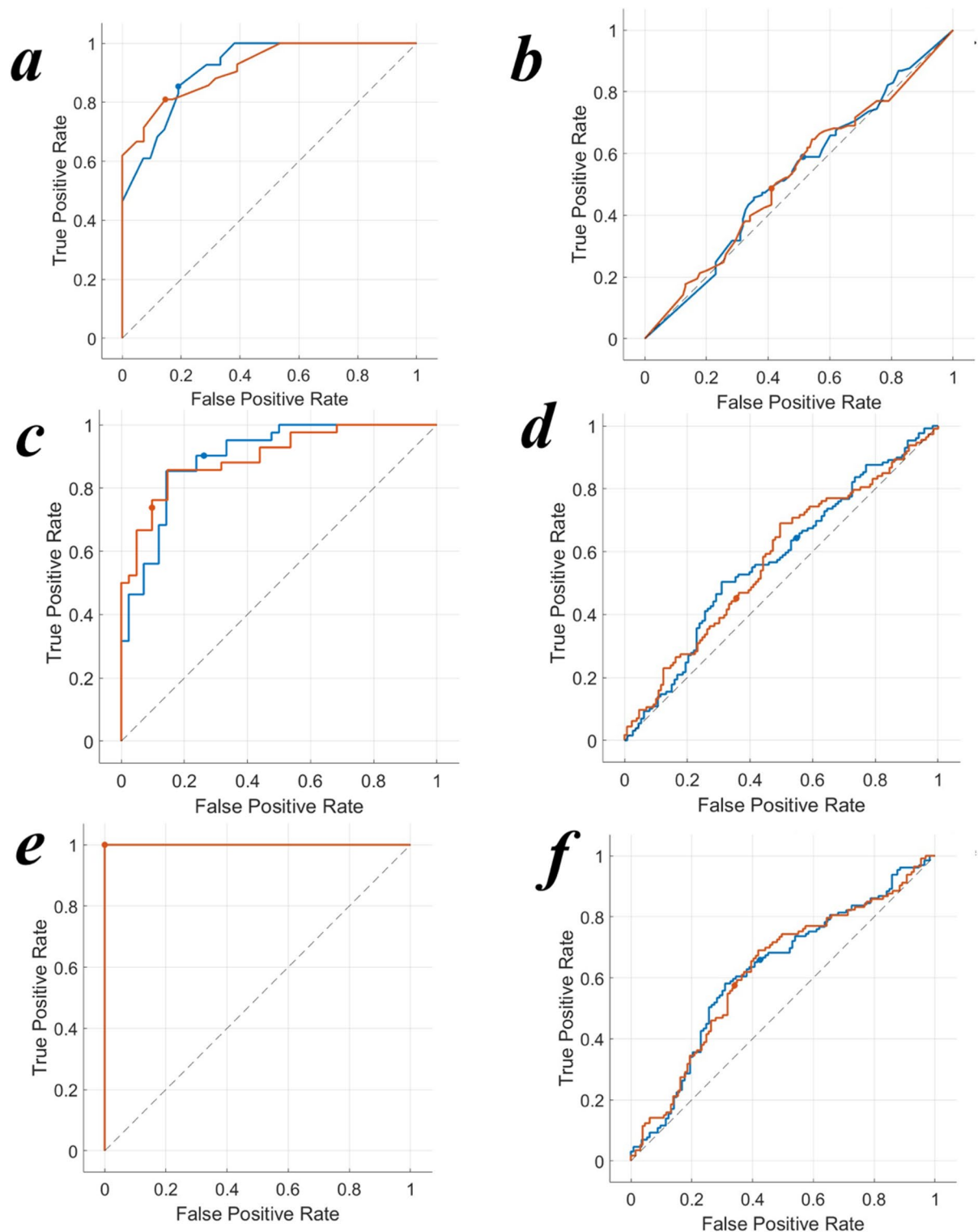


Fig. 1. Receiver operating characteristic (ROC) curves for fine tree, test (a), and validation (b), Kernel-based, test (c), validation (d), k-Nearest neighbors (kNN) models test (e), validation (f). Curves represent test and validation performance in fibrosis prediction.

A comparative analysis of the **HPF model** reveals that it exhibits superior clinical performance due to its straightforward interpretation and expeditious application. Although its accuracy is inferior to that of machine learning models, with testing accuracy at 72% and validation accuracy at 62.81%, the straightforward formulaic approach reduces the potential for overly complex decision-making processes. A high HPF value mean 0.14

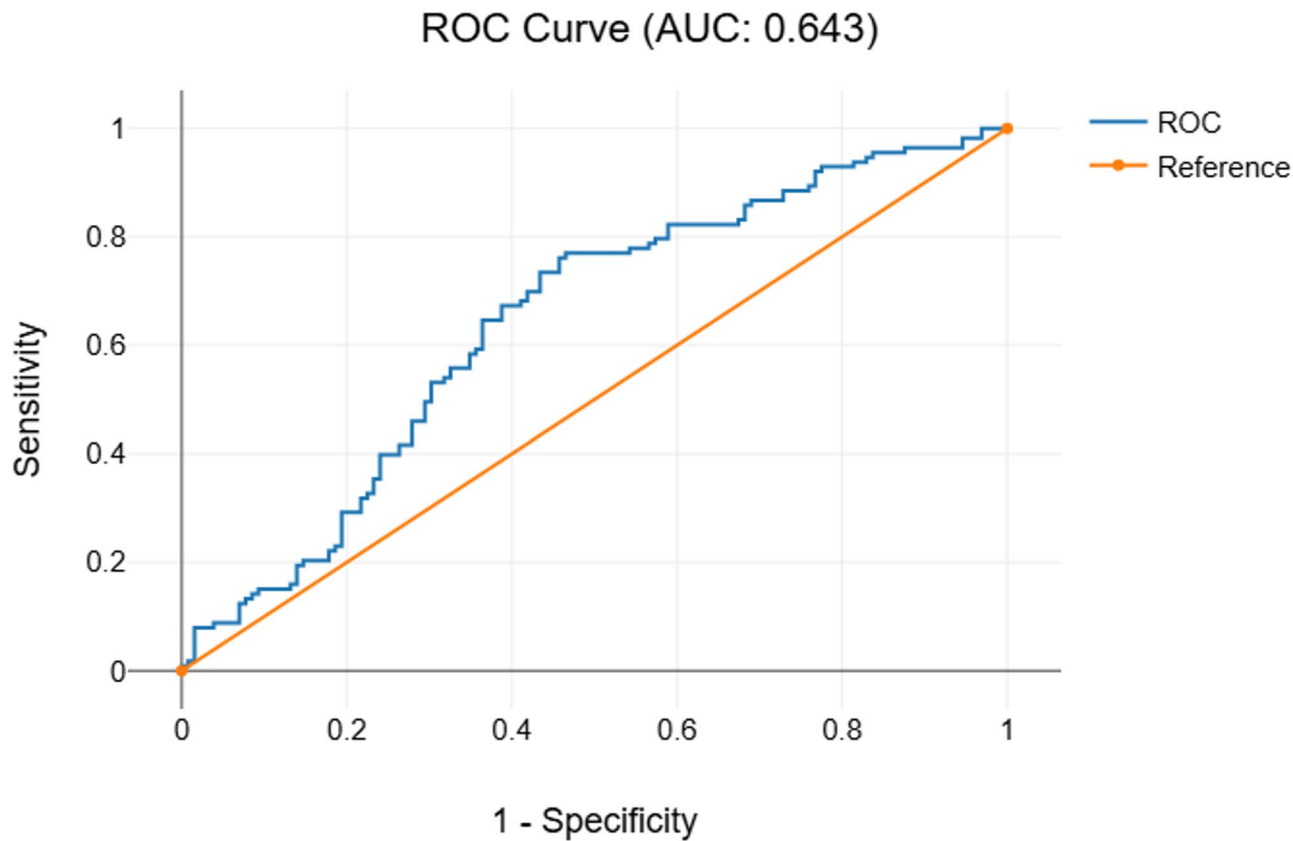


Fig. 2. ROC curve for the human predictive factor (HPF) model. The curve demonstrates the trade-off between sensitivity and specificity in fibrosis prediction.

Model	Validation accuracy (%)	Test accuracy (%)	Advantage	Disadvantage
<i>Fine Tree</i>	54.1	83.1	Easy to interpret decision rules	Over-learning, limited generalisability
<i>Kernel-based</i>	55.4	81,9	Handling non-linear relationships	Sensitive to outliers
<i>kNN</i>	62	100	Simplicity, efficiency on small data sets	Significant over-learning
<i>HPF</i>	62.81	72	Simple, fast to implement	Lower accuracy than machine learning models

Table 4. A comparative analysis of the performance of machine learning models and the human predictive factor (HPF).

(± 0.13) indicates a healthier lung structure with a reduced risk of fibrosis. A low HPF mean $0.11(\pm 0.05)$ indicates the presence of damaged tissue and an elevated risk of fibrosis.

The advantage of the HPF model is that it can be incorporated as an input variable in machine learning algorithms, which can increase their prediction accuracy. This combined approach can increase the efficiency of decision support systems over time.

Unfortunately, its simplicity makes it less accurate than machine kernel or kNN model. On the other hand, the results of the analysis showed a statistically significant difference in HPF scores between groups with and without fibrosis.

While automated prediction cannot be considered a replacement for medical decision-making, the results can assist doctors in making more rapid and objective diagnostic and therapeutic decisions. The Table 4 presents the test and validation accuracies, as well as an overview of the key advantages and limitations of each model.

Clinical utility

The findings of this study substantiate the correlation between fibrosis and HU parameters, as well as the clinical significance of lung volume. CT parameters may serve as potential biomarkers for predicting lung injury.

The results of the study are as follows: The predictive models (both automated and manual) facilitate the expeditious identification of fibrosis risk. This is of particular importance when developing treatment strategies.

In regard to therapy decision-making, the all model can be employed to optimise radiation treatment plans with the objective of preserving lung health. This may entail the prioritisation of the deep inspiration-breath-holding (DIBH) technique in cases where there is a high risk of adverse effects.

Our models based on human patients yielded comparable results to those observed in Drayson's study on mice. While the accuracy of our models was inferior, both studies demonstrated the potential of radiomics in identifying radiation-induced lung injury and predicting therapeutic efficacy at early time points⁴². While the accuracy of the models on the current dataset is promising, the small sample size and lack of population diversity limit the generalisability of the results to other patient populations. Although the integration of artificial intelligence in medical imaging has expanded rapidly, the specific application of deep learning to predict radiation-induced lung injury (RILI) remains underexplored. Our study focuses on planning CT data and utilizes both machine learning and a human-derived predictive factor (HPF) to contribute to the early identification of patients at risk of fibrosis, providing a foundation for future, larger-scale investigations.

Further research is required, including a larger and more diverse patient population, to increase the validity of the models. Additionally, attention should be directed towards the radiation dose to the lungs and other factors that can further refine and enhance the models.

Methods

Population

The study population comprised breast cancer patients with stage I–III invasive adenocarcinoma or carcinoma in situ who underwent radiotherapy for breast cancer between April 2021 and December 2023⁴³. The study included 242 patients for whom a chest CT scan or chest X-ray was available. We did not exclude anyone from the study. The primary outcome was the presence of a shadow (opacity) or bundle (reticulation) in the lung, as observed on chest CT or X-ray imaging, resulting from irradiation.

Data collection

The data pertaining to the patients, their treatment plans, the follow-up information and the details of the endpoints for pulmonary fibrosis were obtained from the patient registry of the Department of Radiation Oncology at Markusovszky University Teaching Hospital in Szombathely, Hungary. The methodology was approved by the Regional and Institutional Research Ethics Committee of the Markusovszky University Teaching Hospital of Szombathely on 19 September 2022, under the protocol number 26/2022. All experimental procedures were conducted in accordance with the recommendations set forth by the ICH-GCP guidelines. We hereby confirm that the subjects provided informed consent for the appropriate analysis to be conducted. The mean follow-up time was 13.5 months (range: 6–24 months).

Data definitions and statistics

The radiological images from chest CT scans and chest X-rays were assessed for the presence of lung fibrosis after breast irradiation. The Mann-Whitney U test was used to assess statistical differences between groups with and without RILI for various CT parameters, including HU mean, standard deviation, and lung volume. MATLAB's built-in cross-validation methods were applied to mitigate overfitting. Statistical analyses were performed using JASP (v. 0.19.0) (JASP 2024 Amsterdam, The Netherlands) and DATAtab: Online Statistics Calculator. (DATAtab e.U. Graz, Austria.)

Machine learning models

In this study, three machine learning models—Fine Tree, Kernel-based, and k-Nearest Neighbors (kNN)—were employed to analyze the correlation between CT-derived parameters and the risk of radiation-induced lung injury (RILI). The models were implemented using MATLAB v. R2024b, Classification Learner tool (The Mathworks Inc Natick, Ma US) was used for modelling with a five-fold cross-validation approach to validate their performance. Matlab's built-in Cross Validation method is recommended for small sample sizes.

Tree model

A Fine Tree model employs a recursive process of data division into smaller groups based on potential features, with the objective of achieving maximum homogeneity within each final group (leaf). The Fine Tree model, which is based on the decision tree algorithm and is a form of machine learning, is used to address a range of classification and regression issues. This model represents a sophisticated variant of the decision tree paradigm, facilitating the construction of more intricate and comprehensive decision-making structures through the incorporation of multiple branches.

In regression problems, the methodology is analogous, albeit with the final leaves providing a numerical prediction. As the tree becomes more complex and contains a greater number of nodes, it is able to discern subtle differences in the data, and the decision rules are logical and well-interpreted. It is important to note, however, that deep trees are susceptible to overfitting the learning dataset, which can result in a reduction in the model's generalisability to new data. This is an intrinsic challenge associated with overlearning. Furthermore, the Fine Tree model necessitates greater computational resources due to its enhanced complexity⁴⁴.

Principle: Fine Tree is a decision tree-based model that progressively classifies data according to target variables (e.g., fibrosis present/absent). For this study, the Fine Tree model was configured with a maximum of 100 splits to partition the data. The split criterion used was Gini's diversity index, and surrogate decision splits were turned off.

Kernel-based model

This model is an optimizable kernel-based learning technique integrated into a machine learning framework, capable of being combined with features from deep learning architectures⁴⁵. The optimizable kernel is designed to create a unique, problem-specific kernel function that enhances the model's performance on the given dataset. By using a mathematical function to transform the data into a higher-dimensional space, this approach enables

the handling of complex non-linear relationships. In MATLAB, kernel parameters are automatically optimized during the learning process, reducing the need for manual tuning.

While not a traditional deep neural network, this model effectively combines the flexibility of kernel methods with the strengths of deep learning techniques. It is often employed for feature learning, where features extracted from a deep learning network are processed by a kernel-based algorithm. Optimization procedures further improve the model by identifying the best hyperparameters, eliminating the need for extensive manual experimentation. The kernel-based approach is particularly effective for small to medium-sized datasets and has demonstrated superior performance in these cases. This model has been selected for its suitability in analyzing biological and medical data.

For this study, the model was configured with the following hyperparameters: multiclass coding was set to “One vs One,” and the iteration limit was fixed at 1000. During optimization, logistic regression was chosen as the learner, with 115 expansion dimensions, a regularization strength (Lambda) of 0.55035, and a kernel scale ranging from 0.001 to 1000. Data standardization was enabled. The hyperparameter search explored a range of learners, including SVM and logistic regression. The number of expansion dimensions was varied from 100 to 10,000, while the regularization strength (Lambda) ranged from 4.1322×10^{-6} to 4.1322. The standardized data was true and false.

k-nearest neighbors (kNN) model

The k-Nearest neighbours (kNN) algorithm is a fundamental non-parametric machine learning method widely used for classification and regression tasks. This model was chosen due to its straightforward implementation, as it does not require explicit model training. The algorithm is highly adaptable to various types of data and dimensions, making it a versatile choice in machine learning applications⁴⁶. However, filtering out outliers during preprocessing could negatively influence the results. Despite this potential limitation, the algorithm retains its adaptability and effectiveness.

In this study, the kNN model was configured with 84 neighbors, using the Euclidean distance metric and squared inverse distance weighting. During hyperparameter optimization, the number of neighbors was varied between 1 and 121, and several distance metrics were explored, including City Block, Chebyshev, Correlation, Cosine, Euclidean, Hamming, Jaccard, Mahalanobis, Minkowski (cubic), and Spearman. Additionally, the impact of standardizing the data was evaluated by testing both standardized and non-standardized versions.

Human predictive factor

HPF is a predictive factor that estimates the likelihood of developing pulmonary fibrosis based on several variables:

$$HPF = \frac{V \frac{HU_{mean} + HU_{sd}}{HU_{mean} - HU_{sd}}}{HU_{max} - HU_{min}} \quad (1)$$

Lung volume (V in cm³) indicates the degree of exertion, with a larger volume associated with a reduced relative risk of tissue damage.

The term “HU mean” refers to the average Hounsfield units of a given lung, which provides an indication of the overall density of lung tissue. A reduction in the mean HU value suggests a healthier lung structure. The HU standard deviation (sd) represents the variability of Hounsfield units in the lung, with higher values indicating greater tissue heterogeneity and potentially the presence of more fibrotic lesions.

The HU maximum and minimum values represent the upper and lower limits of the Hounsfield unit range for the lung.

Formula interpretation

The following formula should be interpreted as follows:

The numerator represents the lung volume, calculated by dividing the ratio of the Hounsfield unit values. This indicates the degree of risk associated with the condition and volume of the tissue in question.

The denominator is defined as the difference between the maximum and minimum HU values, serving as a normalization factor that accounts for the upper and lower limits of the standard deviation of the HU values.

The most extreme values within the HU range indicate the densest and least dense areas of lung tissue. A broader range of HUs suggests greater tissue diversity.

A predictor is constructed as a function of five variables. Based on the findings from the statistical analysis and prior experience, the objective is to quantify the conditions under which lesions may develop.

Radiation technique and dosimetry

The breast irradiation was performed in all patients with 3D conformal RT using a CT-based design with mixed energy of 6 and 10 MV or 6 and 18 MV, respectively. The irradiation plan composed of tangential fields and additional beams to optimize the coverage of the target volume of the design and to minimize the dose to the organs at risk: the heart, lungs and contralateral breast.

We note that hypo fractionated total breast irradiation can be used as an equivalent tool to standard radiotherapy for women who have undergone breast conservation surgery for invasive breast cancer with a clear surgical margin and negative axillary nodes^{47,48}.

Determining the design target volume and the volume of the “organ at risk” is a critical part of radiotherapy. Identifying volumes on the planning CT is often not an easy task. In our work we followed the European Society for Radiotherapy and Oncology (ESTRO) recommendations⁴⁷.

For the whole breast, a dose of 40.05 Gy^{48–51} was prescribed in 15 fractions (223 cases 92.4%), 50 Gy^{46,52,53} in 25 fractions (17 cases 7.02%), and in one case 43.2 Gy in 24 fractions, depending on pathological risk factors.

The supraclavicular lymphatic area and boost dosimetry for the tumor bed were omitted because tangential field placement was used and did not contribute significantly to the lung burden. Lung contours were generated using Siemens SOMATOM Go.Sim (Siemens Erlangen Germany) CT simulator software (SynGO CT VA40). The image slice width was 3 mm, in accordance with the thorax protocol.

Toxicity assessment

Subsequent imaging was conducted following the conclusion of the radiotherapy course. The patient presented the radiological results for the consultation. In light of the radiographic findings, the toxicity was assessed as fibrosis, as described by the radiologist, or the presence of reticulation or opacity in the lung at the chest wall, where the irradiation field may have been. All cases were graded as 1 according to the Common Terminology Criteria for Adverse Events v.5. No medical intervention was required, and only radiological lesions were visible.

Data availability

The datasets generated and/or analysed during the current study are available in the [https://github.com] repository, [https://github.com/gyorfiandras/Ungvary.git]. The underlying Matlab file [and training/validation datasets] for this study is available in https://github.com and can be accessed via this link [https://github.com/gyorfiandras/Ungvary.git].

Received: 30 January 2025; Accepted: 14 May 2025

Published online: 27 May 2025

References

1. National Cancer Institute. (n.d.). *Breast cancer statistics*. Surveillance, epidemiology, and end results (SEER) program. <https://seer.cancer.gov/statfacts/html/breast.html>
2. Mohamed, R. F., Abdelhameed, D. H. & Mohamed, M. A. Combination of anatomical and biological factors to predict disease-free survival in breast cancer. *JCO Global Oncol.* **9**, e2200269. <https://doi.org/10.1200/GO.22.00269> (2023).
3. Ambrose, J. & Hounsfield, G. Computerized transverse axial tomography. *Br. J. Radiol.* **46** (542), 148–149 (1973).
4. DenOtter, T. D. & Schubert, J. Hounsfield unit. In StatPearls. (StatPearls Publishing, 2023).
5. Wilson, M. S. & Wynn, T. A. Pulmonary fibrosis: pathogenesis, etiology, and regulation. *Mucosal Immunol.* **2** (2), 103–121. <https://doi.org/10.1038/mi.2008.85> (2009).
6. Burkhardt, A. Alveolitis and collapse in the pathogenesis of pulmonary fibrosis. *Am. Rev. Respir. Dis.* **140** (2), 513–524. <https://doi.org/10.1164/ajrccm/140.2.513> (1989).
7. Carver, J. R. et al. American society of clinical oncology clinical evidence review on the ongoing care of adult cancer survivors: cardiac and pulmonary late effects. *J. Clin. Oncol.* **25** (25), 3991–4008. <https://doi.org/10.1200/JCO.2007.10.9777> (2007).
8. Vågane, R. et al. Radiological and functional assessment of radiation-induced pulmonary damage following breast irradiation. *Acta Oncol.* **47** (2), 248–254. <https://doi.org/10.1080/02841860701630267> (2008).
9. Ghafoori, P., Marks, L. B., Vujaskovic, Z. & Kelsey, C. R. Radiation-induced lung injury: assessment, management, and prevention. *Oncol. (Williston Park)*. **22** (1), 37–53 (2008).
10. Rødningsen, O. K. et al. Radiation-induced gene expression in human subcutaneous fibroblasts is predictive of radiation-induced fibrosis. *Radiother. Oncol.* **86** (3), 314–320. <https://doi.org/10.1016/j.radonc.2007.09.013> (2008).
11. Beinert, T. et al. Oxidant-induced lung injury in anticancer therapy. *Eur. J. Med. Res.* **4** (2), 43–53 (1999).
12. Lemay, A. M. & Haston, C. K. Radiation-induced lung response of AcB/BcA Recombinant congenic mice. *Radiat. Res.* **170** (3), 299–306. <https://doi.org/10.1667/RR1319.1> (2008).
13. Johnston, C. J. et al. Inflammatory cell recruitment following thoracic irradiation. *Exp. Lung Res.* **30** (5), 369–382. <https://doi.org/10.1080/01902140490438915> (2004).
14. Westermann, W. et al. Th2 cells as effectors in postirradiation pulmonary damage preceding fibrosis in the rat. *Int. J. Radiat. Biol.* **75** (5), 629–638. <https://doi.org/10.1080/095530099140276> (1999).
15. Zarei, F. et al. Quantitative analysis of lung lesions using unenhanced chest computed tomography images. *Clin. Respiratory J.* **18** (5), e13759. <https://doi.org/10.1111/crj.13759> (2024).
16. Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *NPJ Digit. Med.* **3**, 136. <https://doi.org/10.1038/s41746-020-00341-z> (2020).
17. Albers, J. et al. High-resolution propagation-based lung imaging at clinically relevant X-ray dose levels. *Sci. Rep.* **13** (1), 4788. <https://doi.org/10.1038/s41598-023-30870-y> (2023).
18. Adelsmayr, G. et al. CT texture analysis reliability in pulmonary lesions: the influence of 3D vs. 2D lesion segmentation and volume definition by a Hounsfield-unit threshold. *Eur. Radiol.* **33** (5), 3064–3071. <https://doi.org/10.1007/s00330-023-09500-8> (2023).
19. Costa, G. et al. Mapping tumor heterogeneity via local entropy assessment: making biomarkers visible. *J. Digit. Imaging.* **36** (3), 1038–1048. <https://doi.org/10.1007/s10278-023-00799-9> (2023).
20. Roisman, L. C. et al. Radiological artificial intelligence - predicting personalized immunotherapy outcomes in lung cancer. *NPJ Precision Oncol.* **7** (1), 125. <https://doi.org/10.1038/s41698-023-00473-x> (2023).
21. Chen, M. L. et al. Is there any correlation between spectral CT imaging parameters and PD-L1 expression of lung adenocarcinoma? *Thorac. Cancer.* **11** (2), 362–368. <https://doi.org/10.1111/1759-7714.13273> (2020).
22. Alksas, A. et al. A novel higher order appearance texture analysis to diagnose lung cancer based on a modified local ternary pattern. *Comput. Methods Programs Biomed.* **240**, 107692. <https://doi.org/10.1016/j.cmpb.2023.107692> (2023).
23. Brown, K. H. et al. Characterisation of quantitative imaging biomarkers for inflammatory and fibrotic radiation-induced lung injuries using preclinical radiomics. *Radiotherapy Oncology: J. Eur. Soc. Therapeutic Radiol. Oncol.* **192**, 110106. <https://doi.org/10.1016/j.radonc.2024.110106> (2024).
24. Tan, H. et al. A study on the differential of solid lung adenocarcinoma and tuberculous granuloma nodules in CT images by radiomics machine learning. *Sci. Rep.* **13** (1), 5853. <https://doi.org/10.1038/s41598-023-32979-6> (2023).
25. Avanzo, M., Stancanella, J., Pirrone, G. & Sartor, G. Radiomics and deep learning in lung cancer. *Strahlentherapie Und Onkologie: Organ. Der Deutschen Röntgengesellschaft ... Et Alj.* **196** (10), 879–887. <https://doi.org/10.1007/s00066-020-01625-9> (2020).
26. Ghonge, N. P. & Chowdhury, V. Minimum-intensity projection images in high-resolution computed tomography lung: technology update. *Lung India.* **35** (5), 439–440. https://doi.org/10.4103/lungindia.lungindia_489_17 (2018).

27. Gao, M. et al. Holistic classification of CT Attenuation patterns for interstitial lung diseases via deep convolutional neural networks. *Comput. Methods Biomech. Biomedical Engineering: Imaging Visualization*. **6** (1), 1–6. <https://doi.org/10.1080/21681163.2015.1124249> (2018).
28. Al-Sheikh, M. H., Dandan, A., Al-Shamayleh, O., Jalab, A. S., Ibrahim, R. W. & H. A., & Multi-class deep learning architecture for classifying lung diseases from chest X-ray and CT images. *Sci. Rep.* **13** (1), 19373. <https://doi.org/10.1038/s41598-023-46147-3> (2023).
29. Liu, X., Shao, C. & Fu, J. Promising biomarkers of radiation-induced lung injury: A review. *Biomedicine* **9** (9), 1181. <https://doi.org/10.3390/biomedicine9091181> (2021).
30. Nishioka, A. et al. Analysis of radiation pneumonitis and radiation-induced lung fibrosis in breast cancer patients after breast conservation treatment. *Oncol. Rep.* **6** (3), 513–517. <https://doi.org/10.3892/or.6.3.513> (1999).
31. Karlsen, J. et al. Pneumonitis and fibrosis after breast cancer radiotherapy: occurrence and treatment-related predictors. *Acta Oncol.* **60** (12), 1651–1658. <https://doi.org/10.1080/0284186X.2021.1976828> (2021).
32. Li, Y. et al. CT image-based texture analysis to predict microvascular invasion in primary hepatocellular carcinoma. *J. Digit. Imaging*. **33** (6), 1365–1375. <https://doi.org/10.1007/s10278-020-00386-2> (2020).
33. Iwasawa, T., Matsushita, S., Hirayama, M., Baba, T. & Ogura, T. Quantitative analysis for lung disease on thin-section CT. *Diagnostics (Basel)*. **13** (18), 2988. <https://doi.org/10.3390/diagnostics13182988> (2023).
34. Jiang, B. et al. Deep learning reconstruction shows better lung nodule detection for ultra-low-dose chest CT. *Radiology* **303** (1), 202–212. <https://doi.org/10.1148/radiol.210551> (2022).
35. Park, D. et al. Importance of CT image normalization in radiomics analysis: prediction of 3-year recurrence-free survival in non-small cell lung cancer. *Eur. Radiol.* **32** (12), 8716–8725. <https://doi.org/10.1007/s00330-022-08869-2> (2022).
36. Krass, S., Lassen-Schmidt, B. & Schenk, A. Computer-assisted image-based risk analysis and planning in lung surgery—A review. *Front. Surg.* **9**, 920457. <https://doi.org/10.3389/fsurg.2022.920457> (2022).
37. Simpson, S., Hershman, M., Nachiappan, A. C., Raptis, C. & Hammer, M. M. The short and long of COVID-19: A review of acute and chronic radiologic pulmonary manifestations of SARS-CoV-2 and their clinical significance. *Rheumatic Disease Clin. North. Am.* **51** (1), 157–187. <https://doi.org/10.1016/j.rdc.2024.09.004> (2025).
38. Gegin, S., Pazarlı, A. C., Özdemir, B., Özdemir, L. & Aksu, E. A. The effect of Hounsfield unit value on the differentiation of malignant/benign mediastinal lymphadenopathy and masses diagnosed by endobronchial ultrasonography. *Cancer Manage. Res.* **16**, 1013–1020. <https://doi.org/10.2147/CMAR.S473653> (2024).
39. O’Callaghan, M. et al. Analysis of tissue lipidomics and computed tomography pulmonary fat Attenuation volume (CT_{FAV}) in idiopathic pulmonary fibrosis. *Respirology* **28** (11), 1043–1052. <https://doi.org/10.1111/resp.14582> (2023).
40. Wuschner, A. E. et al. Radiation-induced Hounsfield unit change correlates with dynamic CT perfusion better than 4DCT-based ventilation measures in a novel-swine model. *Sci. Rep.* **11** (1), 13156. <https://doi.org/10.1038/s41598-021-92609-x> (2021).
41. Kadeethum, T., O’Malley, D., Choi, Y., Viswanathan, H. S. & Yoon, H. Progressive transfer learning for advancing machine learning-based reduced-order modeling. *Sci. Rep.* **14** (1), 15731. <https://doi.org/10.1038/s41598-024-64778-y> (2024).
42. Drayson, O. G. G., Montay-Gruel, P. & Limoli, C. L. Radiomics approach for identifying radiation-induced normal tissue toxicity in the lung. *Sci. Rep.* **14** (1), 24256. <https://doi.org/10.1038/s41598-024-75993-y> (2024).
43. Recht, A. et al. Postmastectomy radiotherapy: an American society of clinical oncology, American society for radiation oncology, and society of surgical oncology focused guideline update. *J. Clin. Oncol.* **34** (36), 4431–4442. <https://doi.org/10.1200/JCO.2016.6.9.1188> (2016).
44. Cho, Y., Molinaro, A. M., Hu, C. & Strawderman, R. L. Regression trees and ensembles for cumulative incidence functions. *Int. J. Biostatistics*. **18** (2), 397–419. <https://doi.org/10.1515/ijb-2021-0014> (2022).
45. Radhakrishnan, A., Ruiz Luyten, M., Prasad, N. & Uhler, C. Transfer learning with kernel methods. *Nat. Commun.* **14** (1), 5570. <https://doi.org/10.1038/s41467-023-41215-8> (2023).
46. Wang, J. & Geng, X. Large margin weighted k-nearest neighbors label distribution learning for classification. *IEEE Trans. Neural Networks Learn. Syst.* **35** (11), 16720–16732. <https://doi.org/10.1109/TNNLS.2023.3297261> (2024).
47. Kaidar-Person, O. et al. Tricks and tips for target volume definition and delineation in breast cancer: lessons learned from ESTRO breast courses. *Radiother. Oncol.* **162**, 185–194. <https://doi.org/10.1016/j.radonc.2021.07.015> (2021).
48. Offersen, B. V. et al. Hypofractionated versus standard fractionated radiotherapy in patients with early breast cancer or ductal carcinoma in situ in a randomized phase III trial: the DBCG HYPO trial. *J. Clin. Oncol.* **38** (31), 3615–3625. <https://doi.org/10.1200/JCO.20.01363> (2020).
49. Freedman, R. A. et al. Association of breast cancer knowledge with receipt of guideline-recommended breast cancer treatment. *J. Oncol. Pract.* **12** (6), e613–e625. <https://doi.org/10.1200/JOP.2015.008508> (2016).
50. Andrade, T. R. M. et al. Meta-analysis of long-term efficacy and safety of hypofractionated radiotherapy in the treatment of early breast cancer. *Breast* **48**, 24–31. <https://doi.org/10.1016/j.breast.2019.08.001> (2019).
51. Rock, K. et al. Local control in young women with early-stage breast cancer treated with hypofractionated whole breast irradiation. *Breast* **41**, 89–92. <https://doi.org/10.1016/j.breast.2018.07.002> (2018).
52. Whelan, T. J. et al. Long-term results of hypofractionated radiation therapy for breast cancer. *N. Engl. J. Med.* **362** (6), 513–520. <https://doi.org/10.1056/NEJMoa0906260> (2010).
53. Bukovszky, B. et al. Radiotherapy instead of axillary lymph node dissection: evaluation of axillary lymph node dose coverage with whole breast radiotherapy. *Rep. Practical Oncol. Radiotherapy*. **27** (3), 458–466. <https://doi.org/10.5603/RPOR.a2022.0043> (2022).

Acknowledgements

We would like to express our gratitude to Ms Alexandra Nagy, Mrs Anita Sági, Ms Alexandra Pesz and Mr József Horváth for their meticulous data collection.

Author contributions

T. U. conceptualized the study, analysed the data, and wrote the first version of the manuscript. D. Sz. collected and analysed the data. (A) Gy. wrote scripts and prepared (Fig. 1). ZS. D. and (B) K. reviewed manuscript. J. O. and K. T. reviewed and finalized manuscript. All authors reviewed the manuscript.

Funding

Open access funding provided by Markusovszky Teaching Hospital of County Vas.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study was approved by the Regional and Institutional Research Ethics Committee of the Markusovszky University Teaching Hospital of Szombathely on 19 September 2022, under the protocol No. 26/2022. Reporting of all experimental procedures complied with recommendations in committee based on the ICH-GCP guidelines.

Additional information

Correspondence and requests for materials should be addressed to T.U.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025