An Explainable Deep Learning Framework for Predicting Postoperative Radiotherapy-Induced Vaginal Stenosis in Surgically Treated Cervical Cancer Patients

Ann. Ital. Chir., 2025 96, 5: 602–616
https://doi.org/10.62713/aic.4011

Hua Han[1], Honger Zhou[1], Jing He[1], Xiang Zhang[2]

[1]Department of Gynaecology, The First People's Hospital of Fuyang District, 311400 Hangzhou, Zhejiang, China
[2]Department of Gynecological Oncology Radiotherapy, Zhejiang Cancer Hospital, 310022 Hangzhou, Zhejiang, China

AIM: Surgery (e.g., radical hysterectomy) combined with radiotherapy is the mainstay of treatment strategy for locally advanced cervical cancer. However, the beneficial effects of adjuvant radiotherapy are frequently offset by late-onset toxicities, such as vaginal stenosis (VS), which significantly impact patients' quality of life. Although imaging techniques like computed tomography (CT) and magnetic resonance imaging (MRI) are key for both surgical planning and radiotherapy targeting, their ability to predict VS risk before treatment remains limited. This challenge underscores the need for accurate and interpretable predictive models specifically adapted to surgical oncology contexts. This study aims to develop and validate an explainable deep learning framework, integrating Squeeze-and-Excitation (SE) networks and Gradient-weighted Class Activation Mapping (Grad-CAM) visualization, for predicting radiotherapy-induced VS to enable early, personalized intervention strategies.

METHODS: Pre-treatment (i.e., post-surgical, pre-radiotherapy) CT images of cervical cancer patients diagnosed between January 2017 and March 2022 were retrospectively collected. These patients underwent radical hysterectomy (or equivalent surgical resection) followed by radiotherapy. Each patient was categorized as either positive or negative for subsequent VS development. Following normalization and augmentation, we employed a Squeeze-and-Excitation enhanced Inception network (SE-Inception) to distinguish between high- and low-risk cases. Model performance was compared to a conventional Random Forest and a deep learning baseline (ResNet50). Additionally, Grad-CAM visualization was integrated to highlight discriminative image regions for enhanced interpretability and clinical validation.

RESULTS: Among the 140 patients included in the study, 51 developed VS after treatment, representing an incidence rate of 36.4%. The SE-Inception model yielded superior performance (accuracy: 0.93; area under the receiver operating characteristic curve [AUC]: 0.95), surpassing both ResNet50 (accuracy: 0.85; AUC: 0.90) and Random Forest (accuracy: 0.59; AUC: 0.65). Recall and F1 scores also improved markedly, indicating robust sensitivity and precision. Calibration curves demonstrated excellent agreement between predicted and observed risks, while decision curve analysis (DCA) consistently indicated superior net clinical benefits of the SE-Inception model across various threshold probabilities compared to ResNet50 and Random Forest. Grad-CAM consistently localized to anatomically relevant regions correlating with surgeon- and radiologist-identified risk sites, strengthening the clinical interpretability and trustworthiness of the predictive framework.

CONCLUSIONS: Taking the surgical context into account, our SE-Inception framework demonstrated enhanced accuracy and interpretability in identifying patients at risk for postoperative radiotherapy-induced VS. Through alignment with expert clinical assessments and enabling early, personalized intervention strategies, this approach has the potential to improve outcomes and long-term quality of life in cervical cancer survivors, supporting more proactive, surgery-informed treatment planning.

Keywords: cervical cancer; vaginal stenosis; radiotherapy; deep learning; Squeeze-and-Excitation networks; Grad-CAM

Introduction

Cervical cancer (CC) remains one of the most frequently diagnosed gynecologic malignancies worldwide, exerting a profound impact on women's health, fertility, and psychosocial well-being [1,2]. While prevention strategies such as human papillomavirus (HPV) vaccination and improved screening methods have made headway, many patients still present with locally advanced disease, particularly in resource-limited settings. In this scenario, radical hysterectomy, or other appropriate surgical approaches, is often employed as the primary treatment modality, potentially followed by radiotherapy with or without chemotherapy to mitigate recurrence risks [3,4]. Despite improving local control and survival outcomes, adjuvant radiotherapy can lead to late-onset, radiation-induced toxicities that may arise months to years after treatment completion and profoundly affect a patient's long-term quality of life.

Among these late sequelae, radiotherapy-induced vaginal stenosis (VS) is especially significant. Pathologically, VS

Submitted: 19 February 2025 Revised: 24 March 2025 Accepted: 2 April 2025 Published: 10 May 2025
Correspondence to: Hua Han, Department of Gynaecology, The First People's Hospital of Fuyang District, 311400 Hangzhou, Zhejiang, China (e-mail: hanhua0228@163.com).

is characterized by fibrotic and proliferative tissue changes in the vaginal canal—often accelerated by prior surgical manipulation, subsequent tissue remodeling, and exposure to ionizing radiation [5,6]. The resulting cascade involves endothelial damage (injury to the inner lining of blood vessels), impaired microcirculation, chronic inflammation, and progressive fibrosis, which ultimately diminishes elasticity and narrows the vaginal lumen. Clinically, the implications for patients can include dyspareunia (painful sexual intercourse), difficulties during gynecologic examinations (medical procedures to assess the female reproductive system), sexual dysfunction, and compromised quality of life.

The prevalence of VS is substantial, with up to one-third of cervical cancer survivors treated with pelvic surgery and radiotherapy potentially developing some degree of clinically relevant stenosis [7,8]. Despite its impact, the ability to identify individuals at heightened risk for VS before beginning radiotherapy remains limited. Conventional imaging modalities—ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and nuclear imaging—lack the sensitivity and specificity necessary for accurate risk stratification [9]. Consequently, high-risk patients are not always flagged early enough for preventive measures such as surgical technique refinements (e.g., nerve-sparing hysterectomy), organ-preserving maneuvers, or the proactive use of vaginal dilators.

The convergence of artificial intelligence (AI), machine learning (ML), and deep learning has catalyzed advancements in medical image analysis and outcome prediction [10]. By leveraging large, high-dimensional data, these methods can discern subtle imaging biomarkers beyond the capacity of standard radiologic evaluations [11]. Deep learning methods, particularly convolutional neural networks (CNNs), have shown great potential in analyzing medical images. For instance, Squeeze-and-Excitation (SE) networks can enhance the analysis by focusing on the most important image features [12], while Inception modules capture multi-scale feature representations [13]. Merging these strengths in an "Squeeze-and-Excitation enhanced Inception network (SE-Inception)" architecture can reveal nuanced features—potentially those linked to post-surgical tissue changes—in the CT scans of cervical cancer patients at risk for VS.

Equally important, interpretability bolsters the clinical utility of deep learning models. Gradient-weighted Class Activation Mapping (Grad-CAM) offers heatmaps that visualize the model's focus, highlighting image regions most influential to its predictions [14]. By mapping these high-risk areas to post-surgical anatomy and potential radiation damage zones, clinicians gain transparent insights into the model's decision-making. This fosters confidence in AI-driven recommendations and facilitates early, individualized interventions for preventing VS.

In this study, we present a deep learning-based system aimed at predicting radiotherapy-induced VS in patients who have undergone primary surgery (e.g., radical hysterectomy) for cervical cancer. Our approach combines the SE-Inception architecture with Grad-CAM visualization for both high performance and enhanced interpretability. The ultimate goal is to identify patients at higher risk pre-radiotherapy, thereby enabling proactive measures to preserve vaginal function and improve long-term survivorship outcomes.

Materials and Methods

The Proposed System

Our objective was to develop and validate a fully automated, interpretable deep learning system that predicts the risk of VS prior to adjuvant radiotherapy, in patients who had already undergone surgical resection for locally advanced cervical cancer. The pipeline encompasses data acquisition, preprocessing, SE-Inception model training, and Grad-CAM visualization.

Study Population

In this study, we retrospectively collected clinical baseline data of cervical cancer patients diagnosed at The First People's Hospital of Fuyang District between January 2017 and March 2022. Data sources included both electronic medical record systems and paper-based medical records to ensure comprehensiveness and accuracy. The variables collected primarily included patient age, tumor size (in centimeters), tumor location, pathological type, clinical stage (based on the International Federation of Gynecology and Obstetrics [FIGO] staging system), tumor local metastasis (metastasis to lymph nodes), radiotherapy dose, vaginal tumor invasion, and follow-up information. Vaginal stenosis was diagnosed based on clinical symptoms, such as dyspareunia and vaginal narrowing observed during gynecological examinations, or radiological findings during follow-up imaging. The diagnosis of VS was confirmed by a multidisciplinary team comprising gynecologists and radiologists within 6–12 months post-radiotherapy. All data were entered independently by two researchers, with cross-checks performed to minimize the possibility of human error.

Inclusion criteria adopted in this study are as follows: (i) Patients diagnosed with locally advanced cervical cancer who underwent a surgical procedure such as radical hysterectomy; (ii) Patients aged 18–75 years; and (iii) Patients with available pre-adjuvant-radiotherapy CT scans (performed within three months after surgery).

Individuals who fit the following criteria were excluded from this study: (i) Having a history of prior pelvic surgery or radiation (unrelated to their primary cervical cancer diagnosis); (ii) Having undergone concurrent chemotherapy for other malignancies; (iii) Having incomplete medical records or CT scans of poor quality; and (iv) Having non-VS severe toxicities that might confound assessment.

Non-VS severe toxicities refer to adverse events resulting from radiotherapy that significantly impact patients' quality

of life but do not involve VS. These include, but are not limited to, severe rectal bleeding, fistula formation, bowel obstruction, or other Grade 3 or higher toxicities as defined by the Common Terminology Criteria for Adverse Events (CTCAE). The exclusion of these cases ensures that the analysis focuses exclusively on risk factors and outcomes associated with VS, without interference from other severe toxicities that may involve overlapping anatomical or treatment-related factors.

Data Collection and Preprocessing

We collected a dataset of 1200 pre-treatment CT images from cervical cancer patients. Each CT image had an original resolution of 1024×1024 pixels. Prior to modeling, we resampled and resized these images according to the input size required by our deep learning framework. The dataset was split into training, validation, and test sets at a ratio of 60:20:20, ensuring balanced class representation.

Preprocessing involved standard intensity normalization and contrast enhancement steps aimed at highlighting relevant anatomical structures while reducing noise and irrelevant variations [10]. Additional preprocessing procedures included the conversion to grayscale and the application of Gaussian noise augmentation where appropriate. Data augmentation techniques, including random rotations (± 15 degrees), horizontal and vertical flips, and intensity shifts (brightness adjustment range: 0.8 to 1.2), were employed to simulate real-world variability in imaging conditions. We carefully adjusted these parameters to avoid altering important structures, such as the vaginal canal and nearby tissues, ensuring that the anatomical details remained accurate. The calibration process involved consultation with clinical radiologists to retain inter-image spatial relationships necessary for recognizing anatomic features relevant to VS. Augmented images were further reviewed by a panel of radiologists to confirm their anatomical accuracy and clinical relevance. We utilized the Albumentations library (version 1.3.1, Albumentations Team, San Francisco, CA, USA)—a fast and flexible image augmentation toolkit—to perform random rotations ($\pm 15°$), horizontal and vertical flips, random shifts (up to 10% of image size), and contrast manipulations (brightness adjustment range: 0.8 to 1.2). These augmentations were carefully calibrated to maintain anatomical fidelity while increasing the effective size and diversity of the training dataset [15]. Fig. 1 illustrates the differences between the original and preprocessed CT images. The preprocessing steps, including intensity normalization, contrast enhancement, and Gaussian noise augmentation, aim at improving the visibility of key anatomical structures while maintaining spatial fidelity. These adjustments can be seen in the enhanced contrast and reduced noise in the preprocessed image.

Fig. 2 illustrates the overall workflow of the study, spanning from patient selection performed in consideration of inclusion and exclusion criteria, as well as data preprocessing, to the development of three predictive models Random Forest and a deep learning baseline (ResNet50), and SE-Inception). The Fig. 2 highlights how each stage of the pipeline—data collection, augmentation, model training, and Grad-CAM interpretability analysis—was designed to ensure robust predictions of VS risk.

Model Comparison Framework

To rigorously evaluate the performance of the proposed SE-Inception model, we compared it against two baselines: a Random Forest classifier and a standard deep CNN model, ResNet50 [16]. For the Random Forest classifier, we extracted handcrafted radiomic features from the preprocessed CT images using the Pyradiomics library. A total of 1000 features encompassing first-order statistics, texture features, and shape descriptors were initially considered. Feature selection was performed using Recursive Feature Elimination with Cross-Validation (RFECV) to identify the most predictive features, reducing the feature set to 50 optimal variables. Hyperparameter tuning for the Random Forest model was conducted using grid search with 5-fold cross-validation to optimize parameters such as the number of estimators, maximum depth, and minimum samples split. The Random Forest model served as a conventional machine learning baseline, while the ResNet50 architecture represented a widely adopted deep learning benchmark in medical image analysis. By contrasting SE-Inception's performance with these two models, we aimed to highlight the added value of combining SE blocks with Inception modules.

We employed several performance metrics to comprehensively assess and compare the models' effectiveness:

(i) Accuracy: Measures the proportion of correctly classified instances out of the total instances, providing a general sense of the model's correctness.
(ii) Area under the receiver operating characteristic curve (AUC): Evaluates the model's ability to distinguish between classes across various threshold settings, offering insight into its discriminative power.
(iii) Recall (sensitivity): Assesses the model's ability to correctly identify positive instances; it is crucial for minimizing false negatives in clinical settings.
(iv) F1 score: Balances precision and recall, providing a single metric that accounts for both false positives and false negatives, which is especially useful in cases of class imbalance.
(v) Receiver operating characteristic (ROC) curves: Graphically represent the trade-off between the true positive rate and the false positive rate at different threshold levels, aiding in the visual assessment of model performance.

By utilizing these metrics, we ensured a comprehensive evaluation of each model's strengths and weaknesses, fa-

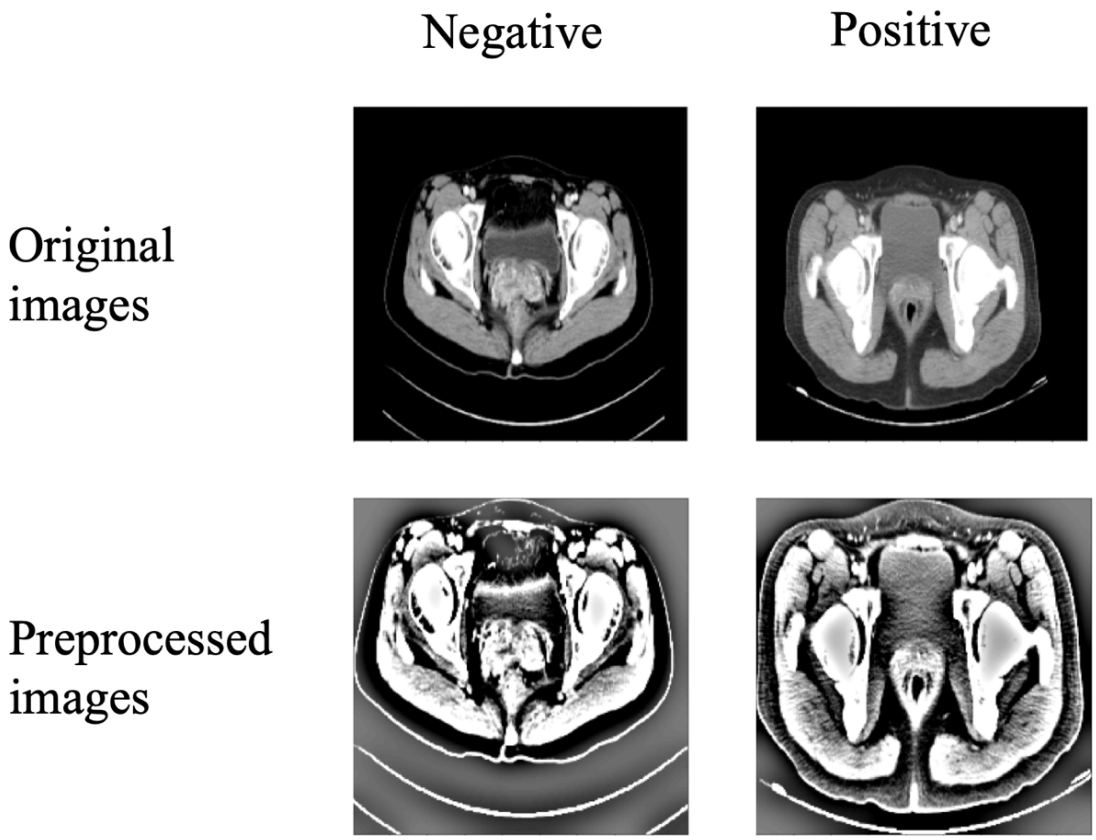

Fig. 1. Comparison of original and preprocessed CT images for negative and positive patients. CT, computed tomography.

cilitating a thorough comparison between SE-Inception, ResNet50, and Random Forest.

The Integration of Squeeze-and-Excitation Networks With Inception

Rationale

Convolutional neural networks have achieved remarkable success in medical image classification tasks, but further improvements often come from innovative architectural enhancements. Two widely recognized techniques are the SE blocks [12] and the Inception modules [13]. SE blocks focus on channel-wise feature recalibration, allowing the network to adaptively emphasize the most informative channels. This mechanism enhances representational power and has been shown to improve performance across various CNN architectures. Inception modules, on the other hand, capture multi-scale spatial information by applying parallel filters of different sizes within the same layer, allowing the model to learn both coarse- and fine-grained features simultaneously [13].

Architecture

As shown in Fig. 3, the SE-Inception model refines the conventional Inception architecture by seamlessly integrating SE blocks to enhance feature selection. Initially, the input image is processed through a standard CNN backbone, which reduces spatial dimensions and extracts lower-level features. Each Inception module then applies multiple parallel convolutional filters of varying kernel sizes to these feature maps, capturing information at different spatial scales. The outputs of these parallel branches are concatenated, yielding a rich, multi-scale representation. At this stage, the SE block is introduced to recalibrate channel-wise feature responses. First, a global average pooling operation condenses the spatial dimensions of the concatenated feature maps into a single vector, effectively summarizing each feature channel's global context. This global descriptor is then passed through two fully connected (FC) layers. The first FC layer reduces the dimensionality to a bottleneck representation, which enables the model to efficiently learn feature interdependencies. A rectified linear unit (ReLU) activation follows to introduce nonlinearity. The second FC layer restores the original channel dimension and employs a sigmoid activation to produce a set of channel-wise weights between 0 and 1. By element-wise multiplying the Inception output by these learned weights, each channel's contribution is adaptively scaled. Channels that effectively capture discriminative patterns are prioritized, whereas those containing less relevant or redundant information are downweighted. This channel recalibration mechanism ensures that the downstream classifier focuses on the most informative cues, thereby boosting the model's

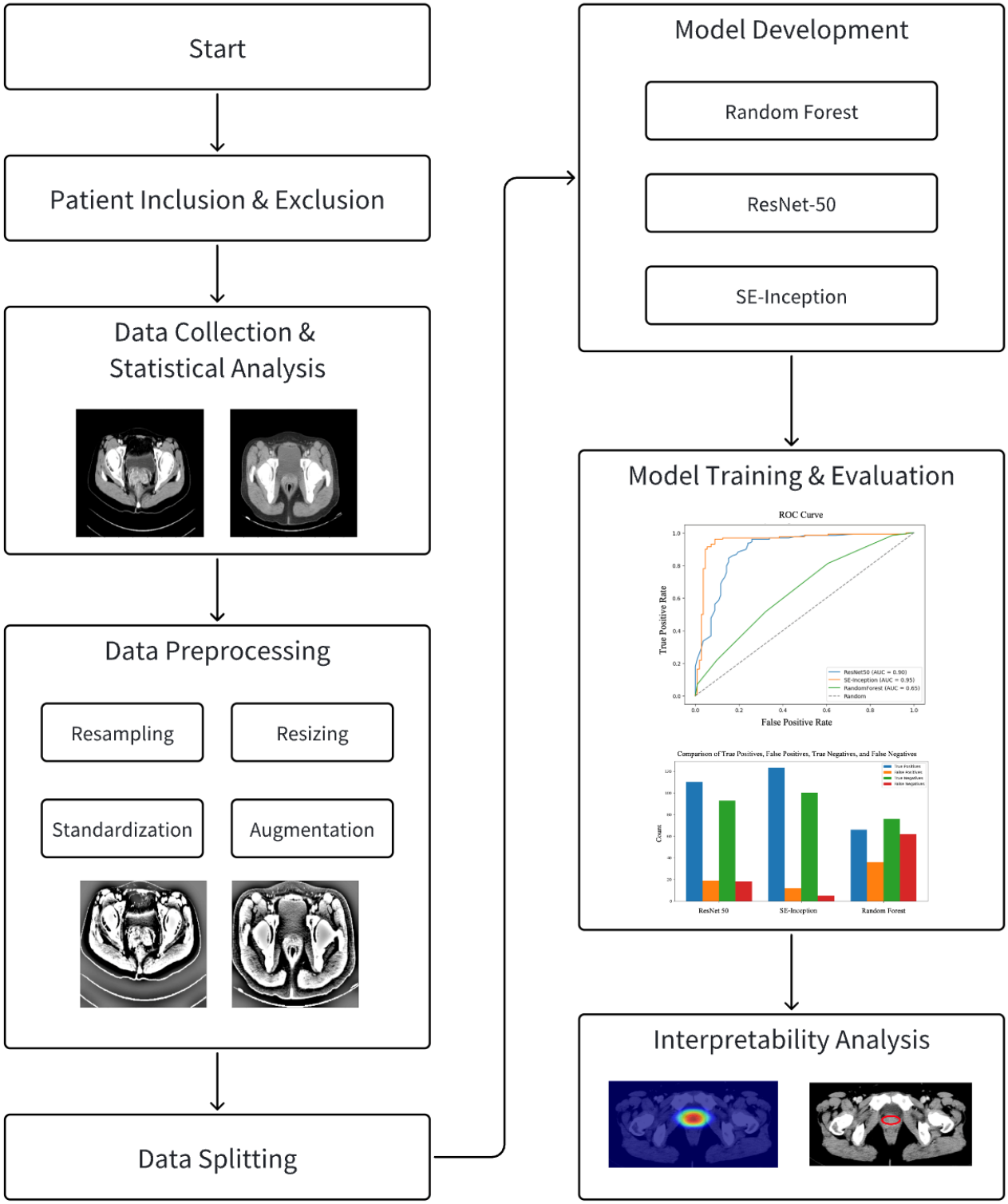

Fig. 2. Overview of the study workflow: evaluation and interpretability analysis for predicting radiotherapy-induced vaginal stenosis in cervical cancer patients. Overview of the study workflow, including patient inclusion, data preprocessing, model development, training, evaluation, and interpretability analysis. This comprehensive pipeline highlights the data-driven approach used to predict radiotherapy-induced VS in cervical cancer patients. VS, vaginal stenosis; SE-Inception, Squeeze-and-Excitation enhanced Inception network; ROC, Receiver operating characteristic; ResNet50, Random Forest and a deep learning baseline; AUC, area under the receiver operating characteristic curve.

capacity to detect subtle biomarkers associated with VS risk. The final output of the SE-Inception block is then passed through subsequent layers, culminating in a binary classification of VS risk.

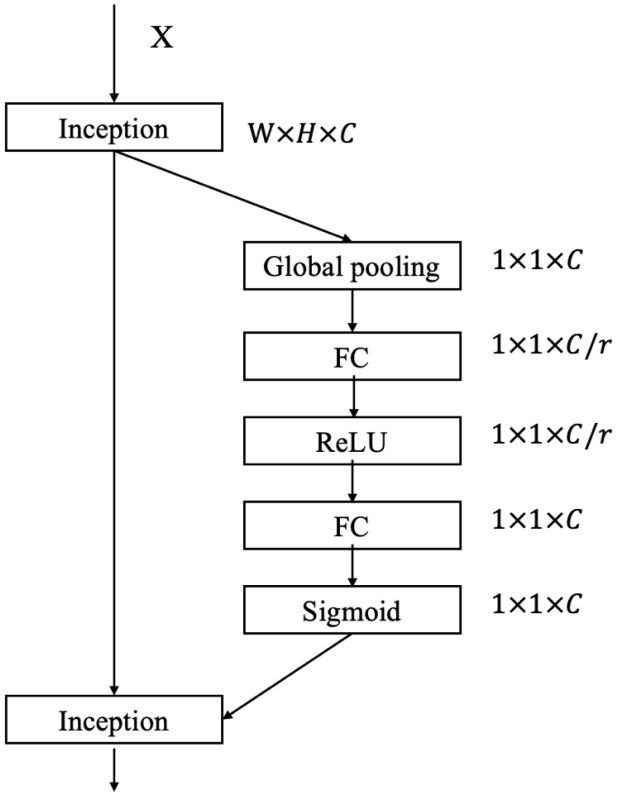

Fig. 3. Schematic representation of the SE-Inception network architecture. FC, fully connected; ReLU, rectified linear unit; X, represents the input feature map or image; W, the width of the input feature map; H, the height of the input feature map; C, the number of channels in the input feature map, representing the depth of features.

Implementation Details

The network was implemented in PyTorch [17]. The architecture began with a series of convolutional and max-pooling layers to reduce the spatial dimensions, followed by several SE-Inception blocks. The final representation was fed into a global average pooling layer and a fully connected classifier to produce a binary prediction: the risk of developing VS (positive or negative). To prevent overfitting, dropout layers were included after the FC layers, with a dropout rate set to 0.5. While this value is higher than the conventional range of 0.2–0.3 for CNNs, prior studies have shown that higher dropout rates can be particularly effective for small, domain-specific datasets in medical imaging tasks. For instance, dropout rates as high as 0.5 have been demonstrated to significantly reduce overfitting in complex architectures, particularly for tasks involving limited data [18]. Similarly, regularization strategies, including higher dropout rates, have been highlighted as critical for improving performance in medical image analysis tasks with small datasets.

The specific hyperparameters used are as follows:

(i) Learning rate: 0.0001
(ii) Batch size: 32
(iii) Weight decay: $1e^{-5}$
(iv) Dropout rate: 0.5

We employed 5-fold cross-validation to ensure the robustness and generalizability of the model. The dataset was divided into five equal folds, where each fold was used once as the validation set while the remaining four folds constituted the training set. This process was repeated five times, and the hyperparameter combinations that yielded the best average performance across the folds were selected for the final model training. Early stopping was implemented based on validation loss with a patience of 10 epochs to prevent overfitting and ensure the selection of the best-performing model checkpoint.

Integration of Grad-CAM Technology

Rationale for Interpretability

While deep learning models can achieve impressive classification performance, their inherent complexity often impedes clinical adoption. Interpretability is crucial, as clinicians must understand the rationale behind a model's prediction to trust and effectively utilize its outputs. Grad-CAM offers a solution by producing visual heatmaps that indicate which regions of the input image most strongly influenced the model's decision [14]. Such interpretability not only instills clinical confidence but also provides insights into the model's internal feature representations, potentially leading to further improvements in diagnostic pathways.

Grad-CAM Integration

We integrated Grad-CAM into the SE-Inception pipeline by computing gradients of the target class score with respect to the final convolutional layer's feature maps. Specifically, once the model generated a VS risk prediction, Grad-CAM used the gradients of this output with respect to the feature maps to produce a spatial localization map. This map was then overlaid on the original CT image to highlight the regions that were pivotal in the model's prediction. To ensure the accuracy and clinical relevance of the Grad-CAM visualizations, the highlighted regions were independently reviewed and validated by a panel of three experienced radiologists. This joint review process served as a reference standard to assess whether the model correctly identified anatomically significant areas associated with VS risk. Any discrepancies between the model's focus and the radiologists' assessments were discussed and addressed to refine the interpretability of the model's predictions. In practice,

Table 1. Patient characteristics between the two groups.

	Non-VS (n = 89)	VS (n = 51)	Statistical methods	Z/t/χ^2	p-value
Age (years), median (P25, P75)	53.00 (44.00, 58.00)	51.50 (41.25, 59.25)	Mann–Whitney U test	–0.108	0.916
BMI (kg/m^2), mean ± SD	23.7 ± 4.5	24.1 ± 4.2	t-test	0.518	0.605
Tumor type			Chi-square	0.460	0.498
Squamous cell carcinoma	69 (77.5)	42 (82.4)			
Adenocarcinoma	20 (22.5)	9 (17.6)			
Tumor stage, n (%)			Chi-square	7.752	0.021
I	29 (32.6)	6 (11.8)			
II	49 (55.1)	35 (68.6)			
III	11 (12.4)	10 (19.6)			
Tumor size			Chi-square	1.235	0.267
≥4	30 (33.7)	22 (43.1)			
<4	59 (66.3)	29 (56.9)			
Local metastasis			Chi-square	1.437	0.231
Yes	8 (9.0)	8 (15.7)			
No	81 (91.0)	43 (84.3)			
Vaginal tumor invasion			Chi-square	5.213	0.022
Yes	66 (74.2)	46 (90.2)			
No	23 (25.8)	5 (9.8)			
Intracavitary brachytherapy			Chi-square	10.300	0.001
Yes	36 (40.4)	35 (68.6)			
No	53 (59.6)	16 (31.4)			
Treatment duration			Chi-square	10.300	0.001
≥45 days	36 (40.4)	35 (68.6)			
<45 days	53 (59.6)	16 (31.4)			

BMI, body mass index; VS, vaginal stenosis; SD, standard deviation.

the Grad-CAM procedure was applied post hoc to images in the validation and test sets, offering clinicians and researchers interpretable model outputs.

Clinical Utility

By revealing the spatial focus of the SE-Inception model, Grad-CAM can guide clinicians to previously overlooked anatomical areas that may predispose to VS. The resulting heatmaps serve as a valuable tool for hypothesis generation and can potentially inform more targeted interventions. For instance, if consistent patterns of radiotherapy-induced damage are localized in specific vaginal sub-regions, prophylactic measures or modified radiation planning could be considered to protect these vulnerable areas. The synergy of high predictive performance and interpretability thus has the potential to enhance clinical decision-making, improve patient outcomes, and foster trust in AI-driven medical solutions.

Calibration Curve and Decision Curve Analysis

Calibration curves were generated by plotting the predicted probabilities of VS risk against the observed outcomes to evaluate the agreement between predicted and actual risks. The calibration performance was assessed using the Hosmer-Lemeshow goodness-of-fit test. Decision curve analysis (DCA) was conducted to assess the net clinical benefit of the SE-Inception model across a range of threshold probabilities. This analysis quantifies the trade-offs between the true positive and false positive rates, helping clinicians evaluate the practical utility of the model in different clinical scenarios.

Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics software (version 22.0, IBM Corp., Armonk, NY, USA). Python packages (version 3.6, Python Software Foundation, Wilmington, DE, USA) were used to generate figures for data visualization. Categorical variables were presented as frequencies and percentages, and analyzed via the chi-square tests. Normality of all continuous variables was initially assessed using the Kolmogorov-Smirnov test. Normally distributed continuous variables were expressed as mean ± standard deviation (SD) and compared between groups using the t-test. Non-normally distributed continuous variables were expressed as median (IQR, interquartile range) and analyzed using the Mann-Whitney U test. Statistical significance was defined as a two-tailed p-value < 0.05.

Results

Patient Characteristics

A total of 140 patients who underwent surgery (e.g., radical hysterectomy) followed by definitive radiotherapy were retrospectively analyzed. Table 1 summarizes the demographic and clinical characteristics of the study population, comparing 89 patients who did not develop VS and 51 patients who did. In our study, the overall incidence of VS was 36.4%.

Notably, no statistically significant differences were observed between the two groups in terms of age, body mass index (BMI), tumor type, tumor size, and local metastasis. However, tumor stage, vaginal tumor invasion, intracavitary brachytherapy and treatment duration showed significant statistical differences. Patients who developed VS were primarily categorized as tumor stages II and III, while patients in stage I exhibited a relatively lower incidence of VS. In this study, the total radiotherapy dose administered across all included patients was similar, ranging from 45 to 50 Gy.

Comparison of Model Performance

Table 2 provides a quantitative performance comparison among the three evaluated models—Random Forest, ResNet50, and the proposed SE-Inception architecture—using accuracy, AUC, recall, and F1 score. These results were derived from the test set, which was exclusively reserved for evaluating the final performance of the models to ensure an unbiased assessment of their generalizability and effectiveness. As shown, the SE-Inception model achieved a notably higher accuracy (0.93) and AUC (0.95, 95% CI: 0.92–0.97) relative to both the Random Forest (0.59 Accuracy; 0.65 AUC, 95% CI: 0.60–0.70) and the ResNet50 baseline (0.85 Accuracy; 0.90 AUC, 95% CI: 0.87–0.93). The 95% confidence intervals (CIs) for the AUC values were calculated using a bootstrapping method with 1000 resamples from the test set. This non-parametric approach ensures robust estimation of variability in the AUC metrics, particularly given the limited size of the test set. Clinically, this improvement translates to fewer false negatives and false positives, enabling more accurate identification of high-risk patients who may benefit from preemptive interventions, such as vaginal dilator therapy or modified radiotherapy planning.

The imaging data were divided into training, validation, and test sets with the following parameters:

(i) Training set: 720 images (60% of the total dataset) used to train the models.
(ii) Validation set: 240 images (20% of the total dataset) utilized for hyperparameter tuning and model selection.
(iii) Test set: 240 images (20% of the total dataset) reserved for evaluating the final model performance.

This partitioning ensured that each subset maintained a balanced representation of both positive and negative classes, thereby providing a reliable assessment of the models' generalizability and performance across different data segments.

To further demonstrate the robustness of our SE-Inception model, we conducted an accuracy convergence analysis during the training process. The SE-Inception model consistently demonstrated faster convergence and greater stability, reaching peak accuracy by the 30th epoch and maintaining it thereafter. In contrast, the Random Forest model exhibited slower convergence with significant fluctuations, while the ResNet50 baseline achieved moderate convergence but did not surpass the SE-Inception model in final accuracy. This convergence behavior underscores the SE-Inception model's superior ability to learn and generalize from the training data efficiently.

The observed performance gap between the SE-Inception and ResNet50 models, particularly in AUC (0.95 vs. 0.90), can be attributed to the architectural enhancements introduced by the SE-Inception model. The integration of SE blocks within the Inception modules allows for more effective channel-wise feature recalibration, which enables the network to emphasize more informative features while suppressing irrelevant ones.

Table 2. Comparative performance metrics for Random Forest, ResNet50, and SE-Inception models.

Model	Accuracy	AUC (95% CI)	Recall	F1 score
Random Forest	0.59	0.65 (0.60–0.70)	0.52	0.57
ResNet50	0.85	0.90 (0.87–0.93)	0.86	0.86
SE-Inception	0.93	0.95 (0.92–0.97)	0.96	0.94

AUC, area under the receiver operating characteristic curve; CI, confidence interval; SE-Inception, Squeeze-and-Excitation (SE) enhanced Inception network; ResNet50, Random Forest and a deep learning baseline.

Fig. 4 illustrates the ROC curves for the three models. The SE-Inception model's curve closely approaches the top-left corner, signifying a superior true positive rate over a wide range of false positive rates. In contrast, the ResNet50 model exhibits a lower, yet still robust, ROC curve. The Random Forest model displays a more modest performance with fewer turning points in its ROC curve, which is indicative of its limited discriminative ability in this specific application. Despite this, the Random Forest serves as a valuable baseline, providing a reference point to demonstrate the enhanced performance of the more complex SE-Inception and ResNet50 models. The inclusion of the Random Forest model underscores the effectiveness of deep learning approaches in capturing intricate patterns within the CT imaging data that simpler models may overlook.

In Fig. 5, the comparative distribution of true positives (TP), false positives (FP), true negatives (TN), and false nega-

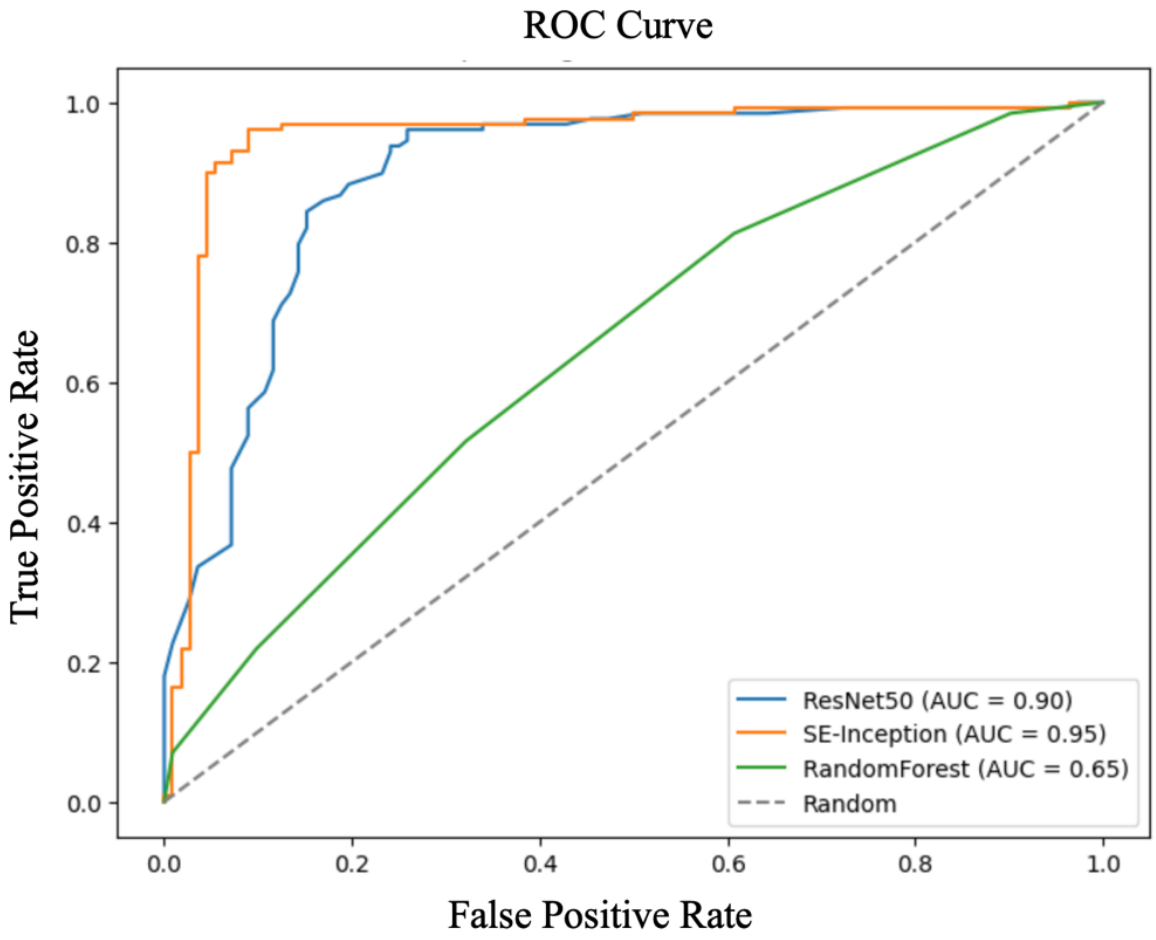

Fig. 4. ROC curves demonstrating the performance of SE-Inception, ResNet50, and Random Forest models. AUC, area under the receiver operating characteristic curve; ROC, receiver operating characteristic.

tives (FN) provides additional insight into the models' predictive characteristics. The SE-Inception model not only demonstrated the highest count of correctly identified true positives (TP) but also maintained a low FP and FN count, reflecting its robust discriminative ability. ResNet50, while still performing relatively well, showed a higher FN and FP count compared to SE-Inception. Meanwhile, the Random Forest model performed significantly worse, misclassifying a substantially larger proportion of cases and showing a notably less favorable balance between TP and FP.

To further assess the clinical applicability of the proposed SE-Inception model, calibration curves and DCA were conducted.

The calibration curves (Fig. 6) evaluate the agreement between predicted probabilities and observed outcomes across the three models—SE-Inception, ResNet50, and Random Forest. As shown in Fig. 7, the SE-Inception model demonstrates excellent calibration, closely aligning with the ideal calibration line (perfect calibration). This indicates that the predicted probabilities generated by the SE-Inception model reflect the true risk of VS with high accuracy, further confirming its reliability for clinical application.

The DCA results, depicted in Fig. 7, provide a comprehensive evaluation of the net clinical benefits of the three models across a range of threshold probabilities. As illustrated in Fig. 7, the SE-Inception model consistently outperforms both ResNet50 and Random Forest across all clinically relevant thresholds. The SE-Inception model achieves the highest net benefit, indicating its superior ability to guide clinical decision-making and stratify patients at high risk of VS for early intervention.

These results highlight not only the predictive accuracy of the SE-Inception model but also its practical utility in clinical settings, aligning with the goal of improving patient outcomes and enabling proactive management of VS.

In addition to quantitative metrics and model comparisons, the interpretability of the SE-Inception model was further assessed using Grad-CAM visualizations, allowing us to qualitatively compare the model's decision-making process against expert clinical judgments. Fig. 8 provides a representative Grad-CAM heatmap depicting the specific subregions within the vagina corresponding closely to areas identified by experienced radiologists as indicative of early tissue changes predictive of VS.

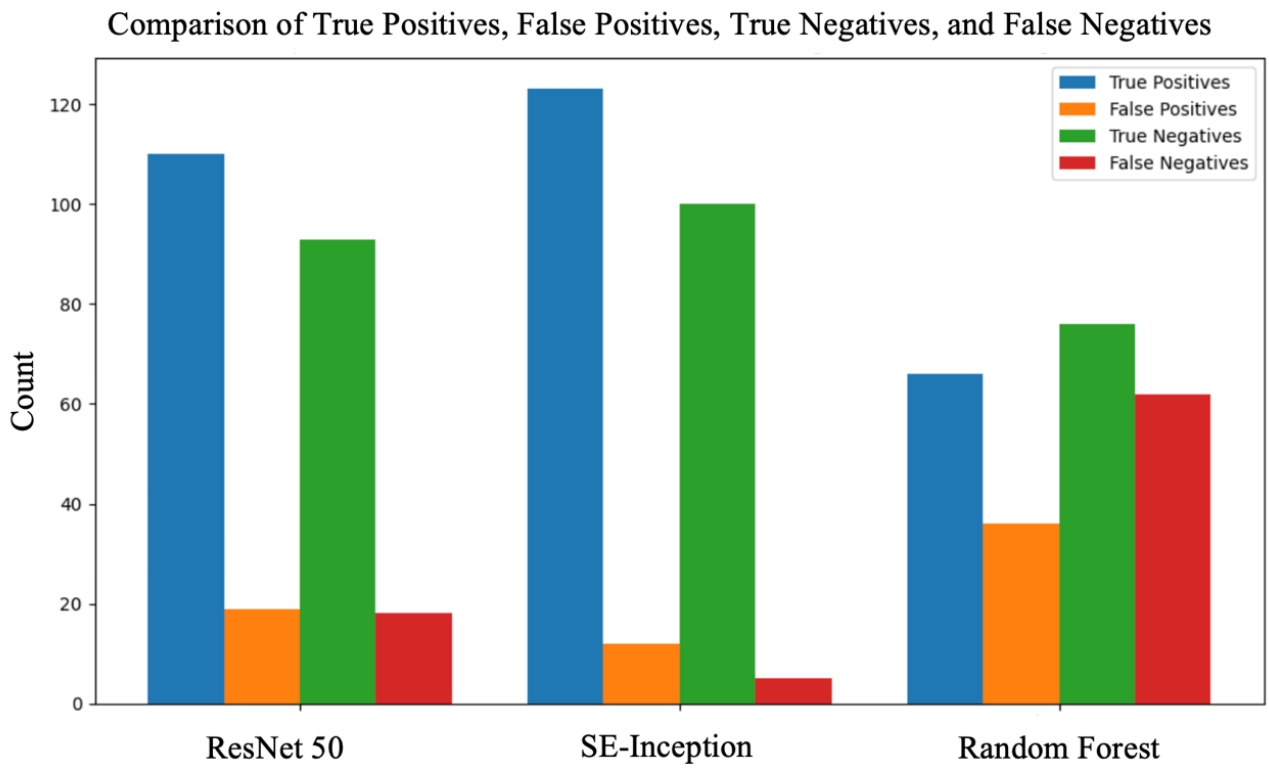

Fig. 5. Comparison of true positives, false positives, true negatives, and false negatives for each model.

Discussion

The present study demonstrates that the integration of SE networks with an Inception-based architecture, coupled with Grad-CAM visualization, can significantly improve the prediction of post-surgical and radiotherapy-induced VS in cervical cancer patients. In these patients, the development of VS typically follows a combination of surgical resection and subsequent radiotherapy, both of which can induce anatomical and functional changes. The ability to predict VS before initiating radiotherapy after surgery enables more proactive and personalized treatment planning [19]. Compared to both a conventional machine learning model (Random Forest) and a commonly employed deep learning baseline (ResNet50), our proposed SE-Inception model achieved superior classification metrics, including higher accuracy, AUC, recall, and F1 score. This performance advantage underscores the capacity of advanced deep learning architectures to identify subtle pre-treatment imaging features predictive of VS, enabling better-informed clinical decision-making.

Furthermore, the suboptimal performance of the Random Forest model highlights the inherent limitations of traditional machine learning approaches in medical imaging tasks. Random Forest relies on a fixed set of handcrafted radiomic features, which, while useful for dimensionality reduction, restrict the model's ability to dynamically adapt to the high-dimensional and complex nature of imaging data. Additionally, Random Forest lacks hierarchical feature extraction capabilities, preventing it from capturing the multi-scale spatial information critical for distinguishing subtle imaging biomarkers associated with VS. In contrast, deep learning models such as SE-Inception and ResNet50 can automatically learn and refine features from raw image data, thereby achieving higher accuracy and generalizability. This limitation of Random Forest demonstrates the challenges of applying traditional machine learning techniques to high-dimensional medical imaging datasets without extensive feature engineering.

One of the most critical aspects of our findings is the robust discriminative power of the SE-Inception model. Specifically, the model's recall of 0.96 ensures that nearly all high-risk patients are correctly identified, minimizing the risk of untreated VS. In addition to its technical performance, the clinical implications of VS prediction are particularly critical for improving patient outcomes. Beyond its physical effects, VS can significantly impact a patient's sexual function and psychological well-being, leading to reduced quality of life [20]. Early identification of at-risk individuals through predictive models like SE-Inception allows clinicians to initiate interventions such as vaginal dilator therapy or pelvic floor physiotherapy, which have been shown to mitigate these consequences. Additionally, incorporating patient education on the potential risks and management strategies for VS could further empower cervical cancer survivors in their recovery journey.

Furthermore, the integration of Grad-CAM visualizations allows clinicians to verify the anatomical regions contributing to the prediction, fostering confidence in the model's

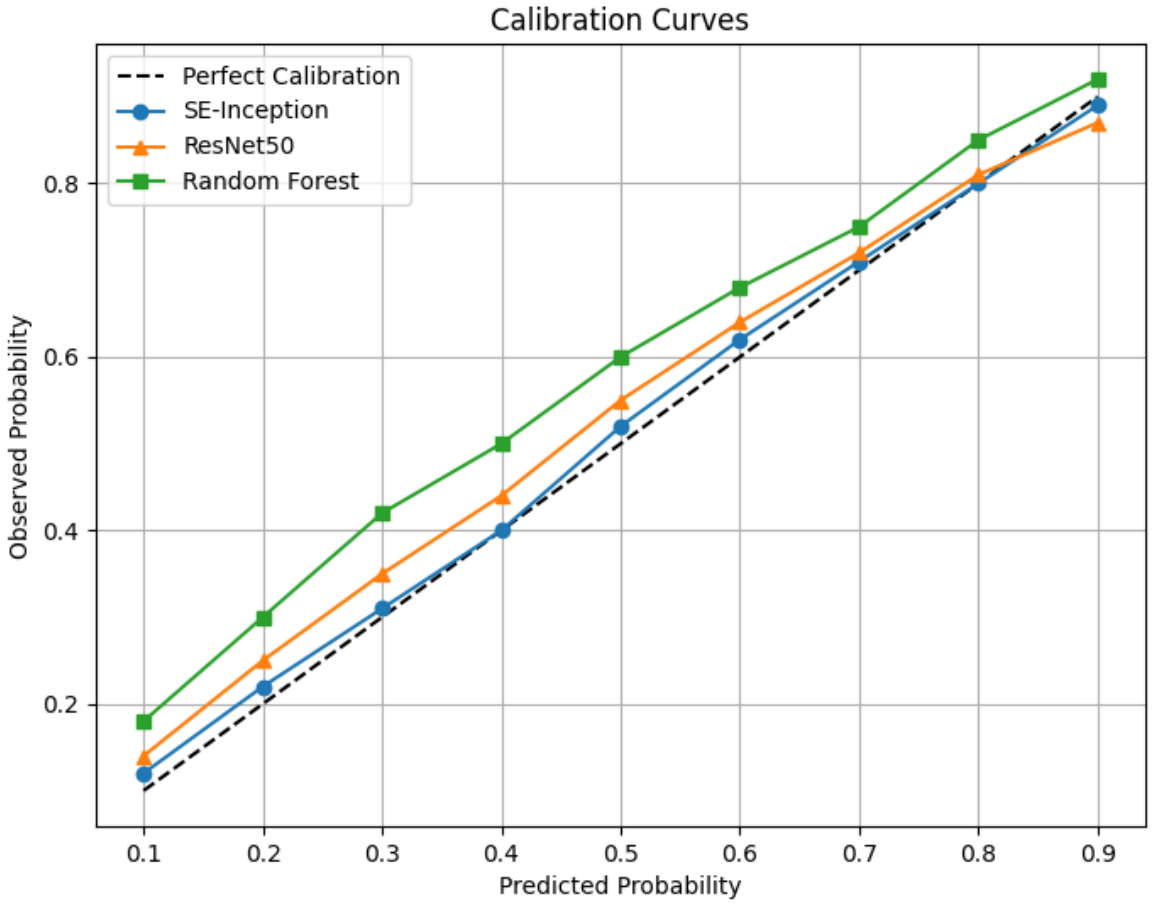

Fig. 6. Calibration curves comparing the predictive performance of the SE-Inception, ResNet50, and Random Forest models. The SE-Inception model demonstrates the highest degree of calibration, closely aligning with the perfect calibration line.

outputs and supporting personalized care strategies. In surgical contexts, multi-scale feature extraction is vital as it enables the model to capture both coarse- and fine-grained details from pre-surgical imaging, reflecting structural changes due to surgery [21]. The Inception modules employed in our architecture facilitate this multi-scale analysis, while the SE blocks re-weight channel-wise features, emphasizing clinically relevant information and attenuating redundant features. Such a combination proved instrumental in uncovering latent patterns associated with the subsequent onset of VS. Moreover, advanced computational approaches, including those leveraging radiomics pipelines, have demonstrated the ability to decode complex imaging phenotypes, offering insights into treatment outcomes following surgical intervention and radiotherapy [22]. Similarly, emerging self-configuring deep learning techniques have shown state-of-the-art performance in various segmentation and classification tasks, demonstrating potential in predicting surgical outcomes in oncology [23]. In the realm of radiation oncology, machine learning-driven models have been applied successfully for early prediction of therapy outcomes in other cancer types, particularly in relation to surgical resection and postoperative treatment [24]. These models help improve the efficiency and precision of clinical decision-making, especially for high-risk patients undergoing multi-modal treatment.

In particular, the high recall achieved by the SE-Inception model indicates its effectiveness in identifying true positives—patients likely to develop VS post-surgery and radiotherapy. This is clinically significant, as failing to identify individuals at high risk before initiating radiotherapy can lead to long-term functional impairment and reduced quality of life. Previous research emphasizes the importance of mitigating late treatment-induced toxicities in cervical cancer survivors, particularly in the context of surgery and radiotherapy [6,7]. By accurately stratifying risk, clinicians can consider preemptive measures, such as modified radiation planning, tailored dose distribution, or early introduction of vaginal dilator therapy, to preserve vaginal function post-surgery and enhance patient outcomes. High-risk patients identified by the SE-Inception model could benefit from early interventions, such as proactive vaginal dilator therapy, tailored radiotherapy protocols, and pelvic floor physiotherapy. These strategies have been shown to mitigate the progression of VS, improve long-term quality of life, and support personalized treatment ap-

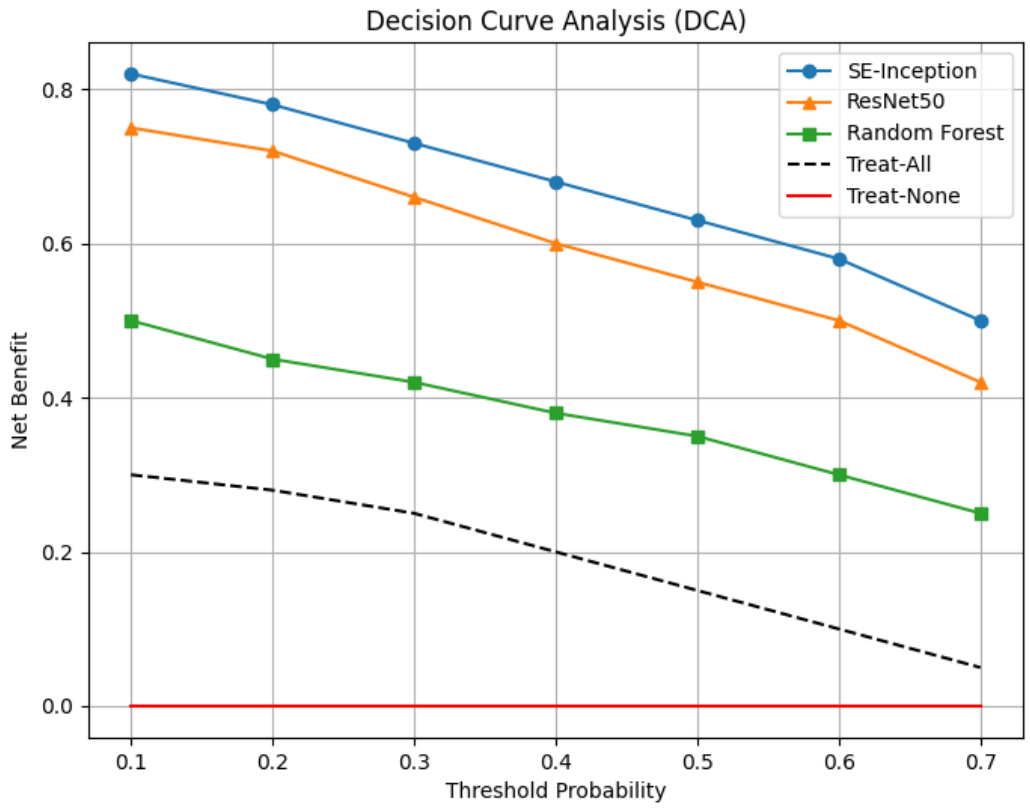

Fig. 7. DCA comparing the net clinical benefit of the SE-Inception, ResNet50, and Random Forest models across a range of threshold probabilities. The SE-Inception model consistently provides the highest net benefit, indicating its superior clinical applicability. DCA, decision curve analysis.

proaches. The high recall of the model empowers clinicians to implement these preemptive strategies at an earlier stage, thereby enhancing patient outcomes and advancing individualized care. Integrating the model into clinical workflows could further provide actionable insights, enabling clinicians to identify high-risk patients and design targeted care plans [25,26]. Additionally, radiotherapy protocols could be adjusted based on model predictions to minimize exposure to high-risk anatomical areas. Future studies should aim to prospectively validate the model's impact on patient outcomes, such as improved quality of life, reduced incidence of VS, and long-term survivorship benefits. Such research would further establish the clinical utility of the model and its potential to transform patient-centered care in cervical cancer treatment. This omission may affect the model's comprehensive predictive capability and its applicability across different treatment protocols. Future studies should consider integrating these radiotherapy-related variables to enhance the model's accuracy and generalizability.

Moreover, the interpretability offered by Grad-CAM visualizations stands out as a key advantage in our work. The lack of transparency in deep learning models has historically been a barrier to their clinical adoption, especially in surgical contexts where anatomical specificity is critical [27]. With Grad-CAM, we provided spatially resolved heatmaps indicating the regions of interest that informed the model's predictions. The observed alignment between these highlighted regions and the areas identified by surgical experts as anatomically relevant for VS risk further supports the model's clinical utility. This congruence helps build trust among clinicians, bridging the "black box" gap that often hinders the integration of AI-driven solutions into routine clinical practice. Interpretability not only validates our findings but also may guide future hypothesis-driven research into the pathophysiology of VS and potential interventions following surgery and radiotherapy.

The improvement in predictive performance observed here aligns with broader trends in medical imaging research, where deep learning models have outperformed traditional algorithms in detecting subtle markers predictive of treatment outcomes. These models are becoming increasingly capable of detecting pre-symptomatic changes in tissue caused by both surgical resection and radiation therapy, providing insights into the post-treatment risks that are critical for personalized oncology care [28]. By leveraging large datasets and advanced architectures, deep learning models can surpass human-level sensitivity and provide clinicians with more accurate predictions to inform their treatment choices.

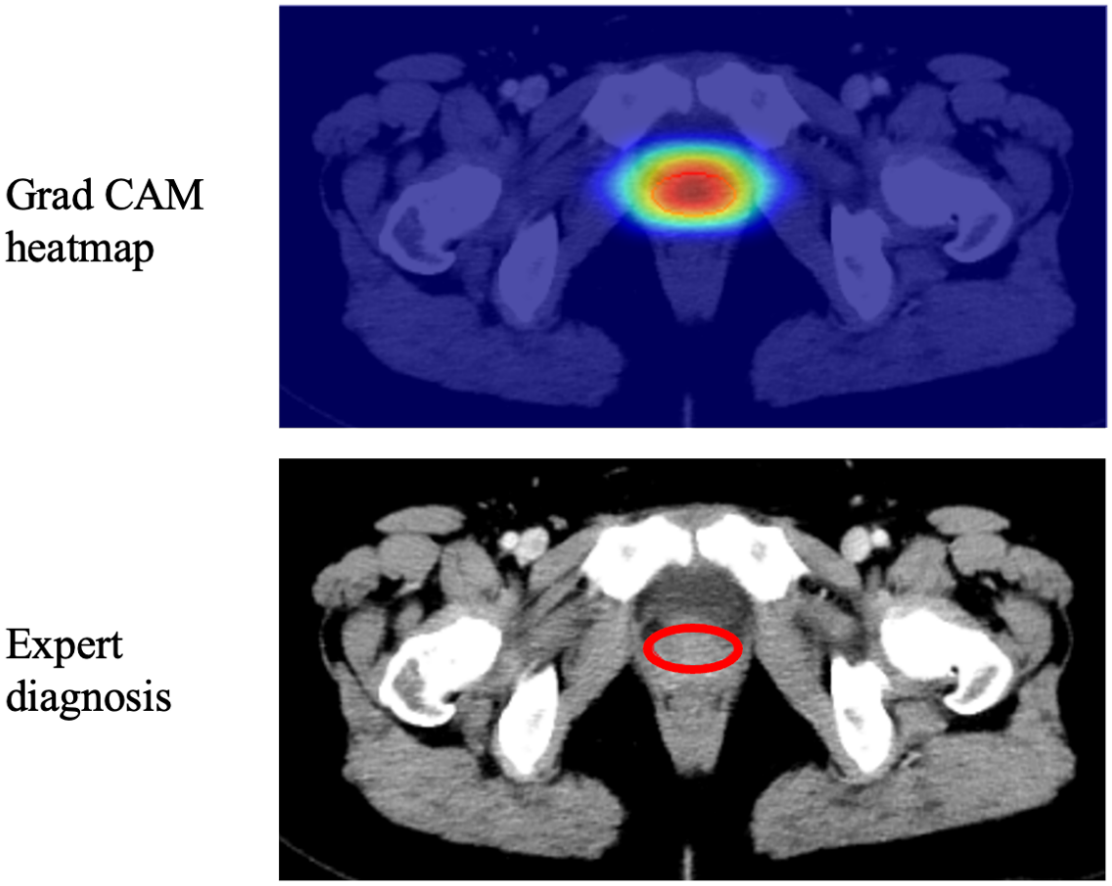

Fig. 8. Grad-CAM heatmap highlighting model-focused regions compared to expert-identified VS risk areas. The red box indicates the region identified by clinical experts as the area of interest for VS risk assessment. Grad-CAM, Gradient-weighted Class Activation Mapping.

The calibration of data augmentation strategies underscores the importance of preserving the anatomical integrity of CT images in medical imaging applications [29]. By carefully limiting transformation parameters such as rotation angles and intensity adjustments, the methodology ensures that augmented images retain the spatial and anatomical features critical for recognizing VS risk. This approach not only enhances the model's performance on the current dataset but also enhances its generalizability to external datasets by maintaining clinical relevance. Future studies incorporating multi-institutional data may further validate the effectiveness of these augmentation strategies across diverse patient populations and imaging protocols.

Compared to prior research in predicting radiotherapy-induced toxicities, our study offers significant improvements in both predictive performance and result interpretability. Previous models primarily focused on achieving high accuracy and AUC without providing mechanisms for understanding the decision-making process [30]. In contrast, our integration of Grad-CAM visualization not only maintains superior predictive capabilities but also enhances the interpretability of the model by highlighting the specific anatomical regions influencing its predictions. This dual focus on accuracy and interpretability facilitates greater trust and usability in clinical settings, enabling healthcare professionals to make informed decisions based on both the model's outcomes and the underlying reasoning. By aligning model focus with expert-identified risk areas, our approach bridges the gap between complex deep learning models and practical clinical application, setting a new standard for explainable AI in medical image analysis.

In clinical applications, the SE-Inception model demonstrates significant potential. Predictive results can be presented through intuitive interfaces, such as heatmaps or risk scores, enabling clinicians to quickly interpret the findings and seamlessly integrate them into their workflows. The computational efficiency of the model ensures timely generation of results, while integration with electronic medical records (EMR) could further facilitate real-time predictions and support clinical practice.

However, certain challenges must be addressed for successful clinical implementation, including data standardization, applicability to diverse populations, and implementation costs. Variations in imaging protocols and patient demographics across institutions may affect the model's performance. Future research should prioritize external valida-

tion, the use of multi-center datasets, and clinician training to ensure the robustness of the model and its widespread adoption in routine medical practice.

Several limitations of the current study should be acknowledged. First, the model has not undergone external validation using multi-center datasets, which may limit its generalizability to diverse clinical settings with varying imaging protocols and patient demographics. Future studies should prioritize the inclusion of data from multiple institutions to rigorously assess the model's robustness and applicability across different populations and clinical environments.

Second, the dataset used in this study, while relatively large for a single institution, may not fully capture the heterogeneity of real-world clinical scenarios. Although data augmentation and cross-validation strategies were employed to mitigate overfitting risks associated with the limited dataset size, further improvements could be achieved by increasing the dataset scale and incorporating additional clinical variables. For instance, radiotherapy-related factors such as radiation dose, target volume, and fractionation schedules were not included in the current model, which may have impacted its ability to comprehensively predict VS risk. Integrating these variables in future studies would likely enhance the model's predictive performance and clinical relevance.

Finally, the reliance on pre-treatment imaging data alone may not fully account for post-treatment anatomical or functional changes induced by radiotherapy. The inclusion of longitudinal imaging data or post-treatment variables could provide a more comprehensive understanding of VS risk and improve the model's utility for clinical decision-making. Future research should also explore the integration of advanced imaging modalities and molecular biomarkers to further enhance the predictive accuracy and clinical applicability of the model.

Conclusions

In conclusion, integrating the SE-Inception architecture with Grad-CAM visualization effectively predicts post-surgical and radiotherapy-induced VS in cervical cancer patients. The SE-Inception model outperformed both the Random Forest and ResNet50 baselines, demonstrating superior sensitivity, specificity, and robustness. Grad-CAM heatmaps not only highlighted key anatomical regions linked to VS development but also provided valuable interpretability, aligning with expert clinical assessments. This combination of predictive accuracy and visual clarity suggests that such AI systems can support proactive patient management, reduce long-term morbidity, and improve the quality of life for patients undergoing cervical cancer treatment, particularly those who have undergone surgery and subsequent radiotherapy. These tools have the potential to enhance personalized oncology care and long-term survivorship management.

Availability of Data and Materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Author Contributions

HH, HZ: Methodology and analysis, investigation, visualization, data collection, writing—original draft; JH: Methodology and analysis; XZ, Conceptualization, writing—reviewing. All authors contributed to important editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

We hereby declare that informed consent was obtained from the human subjects involved in this study for the publication of their images. The purpose and nature of the study, as well as the potential risks and benefits, were explained to the participants prior to obtaining their consent. The ethical considerations pertaining to this research have been rigorously examined and approved by the Ethics Committee of The First People's Hospital of Fuyang District, Hangzhou (No. 2022-ky-004). The study was reviewed and approved in accordance with the Declaration of Helsinki. The study involving the utilization of medical images and data adheres to the highest standards of ethical conduct and patient confidentiality. The approval from the Ethics Committee underscores our commitment to upholding the welfare and rights of all individuals involved in this study.

Acknowledgment

Not applicable.

Funding

This work was funded by the Hangzhou Health Science and Technology Plan (No. B20220713).

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, *et al*. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet. Global Health. 2020; 8: e191–e203. https://doi.org/10.1016/S2214-109X(19)30482-6.

[2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a Cancer Journal for Clinicians. 2018; 68: 394–424. https://doi.org/10.3322/caac.21492.

[3] Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet (London, England). 2019; 393: 169–182. https://doi.org/10.1016/S0140-6736(18)32470-X.

[4] Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain J, *et al*. American Cancer Society, American Society for Col-

poscopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA: a Cancer Journal for Clinicians. 2012; 62: 147–172. https://doi.org/10.3322/caac.21139.

[5] Martins J, Vaz AF, Grion RC, Esteves SCB, Costa-Paiva L, Baccaro LF. Factors associated with changes in vaginal length and diameter during pelvic radiotherapy for cervical cancer. Archives of Gynecology and Obstetrics. 2017; 296: 1125–1133. https://doi.org/10.1007/s00404-017-4553-z.

[6] Miles T, Johnson N. Vaginal dilator therapy for women receiving pelvic radiotherapy. The Cochrane Database of Systematic Reviews. 2014; 2014: CD007291. https://doi.org/10.1002/14651858.CD007291.pub3.

[7] Kirwan JM, Symonds P, Green JA, Tierney J, Collingwood M, Williams CJ. A systematic review of acute and late toxicity of concomitant chemoradiation for cervical cancer. Radiotherapy and Oncology: Journal of the European Society for Therapeutic Radiology and Oncology. 2003; 68: 217–226. https://doi.org/10.1016/s0167-8140(03)00197-x.

[8] Rodrigues AC, Teixeira R, Teixeira T, Conde S, Soares P, Torgal I. Impact of pelvic radiotherapy on female sexuality. Archives of Gynecology and Obstetrics. 2012; 285: 505–514. https://doi.org/10.1007/s00404-011-1988-5.

[9] Laganà AS, Garzon S, Raffaelli R, Ban Frangež H, Lukanovič D, Franchi M. Vaginal Stenosis After Cervical Cancer Treatments: Challenges for Reconstructive Surgery. Journal of Investigative Surgery: the Official Journal of the Academy of Surgical Research. 2021; 34: 754–755. https://doi.org/10.1080/08941939.2019.1695987.

[10] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Medical Image Analysis. 2017; 42: 60–88. https://doi.org/10.1016/j.media.2017.07.005.

[11] Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. Radiology. 2019; 292: 60–66. https://doi.org/10.1148/radiol.2019182716.

[12] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020; 42: 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372.

[13] Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, *et al*. Going deeper with convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp.1–9). 2015.

[14] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. IEEE International Conference on Computer Vision (ICCV) (pp.618–626). 2017.

[15] Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Gavrilyuk K. Albumentations: Fast and Flexible Image Augmentations. Information. 2020; 11: 125. https://doi.org/10.3390/info11020125.

[16] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp.770–778). 2016.

[17] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, *et al*. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R (eds.) Advances in Neural Information Processing Systems 32 (NeurIPS 2019) (pp.8024–8035). 2019.

[18] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014; 15: 1929–1958.

[19] Barki C, Alsufyani SJ, Softah A, Labidi S, Rahmouni HB. Advancing radiation therapy safety in cancer-care: Leveraging AI for adverse reaction prediction. Journal of Radiation Research and Applied Sciences. 2024; 17: 101141. https://doi.org/10.1016/j.jrras.2024.101141.

[20] Morris L, Do V, Chard J, Brand AH. Radiation-induced vaginal stenosis: current perspectives. International Journal of Women's Health. 2017; 9: 273–279. https://doi.org/10.2147/IJWH.S106796.

[21] Yang H, Wang Z, Liu X, Li CG, Xin JC, Wang ZQ. Deep learning in medical image super resolution: a review. Applied Intelligence. 2023; 53: 20891–20916. https://doi.org/10.1007/s10489-023-04566-9.

[22] Al Mudawi N, Alazeb A. A Model for Predicting Cervical Cancer Using Machine Learning Algorithms. Sensors (Basel, Switzerland). 2022; 22: 4132. https://doi.org/10.3390/s22114132.

[23] Ding AS, Lu A, Li Z, Sahu M, Galaiya D, Siewerdsen JH, *et al*. A Self-Configuring Deep Learning Network for Segmentation of Temporal Bone Anatomy in Cone-Beam CT Imaging. Otolaryngology–head and Neck Surgery: Official Journal of American Academy of Otolaryngology-Head and Neck Surgery. 2023; 169: 988–998. https://doi.org/10.1002/ohn.317.

[24] Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine Learning-Guided Adjuvant Treatment of Head and Neck Cancer. JAMA Network Open. 2020; 3: e2025881. https://doi.org/10.1001/jamanetworkopen.2020.25881.

[25] Haddad NC, Soares Brollo LC, Pinho Oliveira MA, Bernardo-Filho M. Diagnostic Methods for Vaginal Stenosis and Compliance to Vaginal Dilator Use: A Systematic Review. The Journal of Sexual Medicine. 2021; 18: 493–514. https://doi.org/10.1016/j.jsxm.2020.12.013.

[26] Damast S, Jeffery DD, Son CH, Hasan Y, Carter J, Lindau ST, *et al*. Literature Review of Vaginal Stenosis and Dilator Use in Radiation Oncology. Practical Radiation Oncology. 2019; 9: 479–491. https://doi.org/10.1016/j.prro.2019.07.001.

[27] Morris MX, Fiocco D, Caneva T, Yiapanis P, Orgill DP. Current and future applications of artificial intelligence in surgery: implications for clinical practice and research. Frontiers in Surgery. 2024; 11: 1393898. https://doi.org/10.3389/fsurg.2024.1393898.

[28] Jha AK, Mithun S, Sherkhane UB, Jaiswar V, Osong B, Purandare N, *et al*. Systematic review and meta-analysis of prediction models used in cervical cancer. Artificial Intelligence in Medicine. 2023; 139: 102549. https://doi.org/10.1016/j.artmed.2023.102549.

[29] Islam T, Hafiz S, Jim JR, Kabir M, Mridha MF. A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. Healthcare Analytics. 2024; 5: 100340. https://doi.org/10.1016/j.health.2024.100340.

[30] Ubeira-Gabellini MG, Mori M, Palazzo G, Cicchetti A, Mangili P, Pavarini M, *et al*. Comparing Performances of Predictive Models of Toxicity after Radiotherapy for Breast Cancer Using Different Machine Learning Approaches. Cancers. 2024; 16: 934. https://doi.org/10.3390/cancers16050934.