Diagn Interv Radiol 2026; DOI: 10.4274/dir.2026.263954

ARTIFICIAL INTELLIGENCE AND INFORMATICS

ORIGINAL ARTICLE

©Copyright 2026 Author(s) - Available online at dirjournal.org.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Predicting drainage success of peritonsillar abscesses: a radiomics-based machine learning approach using contrast-enhanced computed tomography

İzzet Selçuk Parlak[1]
Bökebatur Ahmet Raşit Mendi[2]
Alperen Özkan[1]

[1]Bilkent City Hospital, Clinic of Radiology, Ankara, Türkiye

[2]Yıldırım Beyazıt University Medical School, Department of Radiology, Ankara, Türkiye

PURPOSE

To develop and validate radiomics-based machine learning models combined with clinical parameters derived from venous phase contrast-enhanced computed tomography (CECT) for predicting drainage success in patients with peritonsillar abscess (PTA), aiming to reduce unnecessary invasive procedures.

METHODS

This retrospective study included 94 adult patients with PTA who underwent venous phase CECT followed by incision and drainage within 24 hours. Patients were categorized into drainage (n = 52) and non-drainage (n = 42) groups based on procedural outcomes. Clinical parameters (age, sex, trismus, and uvula deviation) were integrated with 107 extracted radiomics features. Three-dimensional manual segmentation was performed using 3D Slicer. Feature selection was conducted using the least absolute shrinkage and selection operator (LASSO) regression with 10-fold cross-validation. Four machine learning algorithms—support vector machine (SVM), random forest classifier (RFC), decision tree, and extreme gradient boosting (XGBoost)—were developed, and a combined clinical–radiomics model was constructed. Model performance was evaluated using sensitivity, specificity, and area under the receiver operating characteristic curve (AUC).

RESULTS

The LASSO selected 10 discriminative features (optimal λ: 0.043). Inter- and intraobserver reproducibility demonstrated good agreement (intraclass correlation coefficient ≥ 0.75; Spearman's $\rho > 0.8$). Among the radiomics-only models, XGBoost achieved the highest diagnostic performance (AUC: 0.919). The combined clinical–radiomics model further improved diagnostic performance, reaching an AUC of 0.934, with a sensitivity of 90.2% and a specificity of 94.1%. The SVM yielded an AUC of 0.899, whereas the decision tree and RFC demonstrated AUCs of 0.887 and 0.850, respectively. Notably, the high specificity of the combined model suggests strong potential for identifying non-drainable collections.

CONCLUSION

Integrating clinical parameters with radiomics-based machine learning, particularly the combined clinical–radiomics model, can accurately predict drainage success in PTA using pre-procedural CECT images. This quantitative approach may improve patient selection for invasive interventions.

CLINICAL SIGNIFICANCE

Pre-procedural identification of non-drainable collections may help avoid unnecessary drainage attempts, reduce procedure-related morbidity, and support more informed clinical decision-making in the management of PTA.

KEYWORDS

Peritonsillar abscess, radiomics, machine learning, contrast-enhanced computed tomography, drainage

Handling editor: Barbaros Çil

Corresponding author: İzzet Selçuk Parlak

E-mail: selcukparlakdr@gmail.com

Received 25 February 2026; revision requested 30 March 2026; accepted 04 May 2026.

Epub: 11.05.2026

Publication date: 02.11.2026

DOI: 10.4274/dir.2026.263954

You may cite this article as: Parlak İS, Mendi BAR, Özkan A. Predicting drainage success of peritonsillar abscesses: a radiomics-based machine learning approach using contrast-enhanced computed tomography. *Diagn Interv Radiol*. 2026;32(6):757-764.

Peritonsillar abscess (PTA) is defined as a collection of pus within the peritonsillar space, representing the most common deep neck infection in the adult population. Its clinical presentation includes severe unilateral sore throat, fever, dysphagia, and odynophagia, often accompanied by trismus.[1] The established management for PTA is the evacuation of purulent material through an invasive procedure—either needle aspiration or incision and drainage (I&D)—supported by systemic antibiotic therapy.[2,3]

Although clinical assessment is the first step, contrast-enhanced computed tomography (CECT) is frequently employed to confirm the diagnosis, especially in ambiguous cases, and to differentiate a suspected abscess from non-suppurative peritonsillar cellulitis (phlegmon).[4] The typical radiological finding suggestive of an abscess is a peripherally rim-enhancing fluid collection. This finding is crucial because it directly informs the decision to proceed with an invasive drainage attempt.[5]

However, a critical clinical challenge arises from the fact that a decision to intervene based on current diagnostic standards does not guarantee a successful outcome. A non-negligible proportion of drainage procedures, performed on lesions that appear as well-defined abscesses on CECT, result in a "dry tap" or failure to evacuate purulent material despite surgical incision.[6,7] This outcome means that the patient underwent an invasive procedure without the intended therapeutic benefit; furthermore, these attempts are not without their own potential for morbidity. These procedural risks include considerable pain, iatrogenic bleeding from local vasculature, and potential injury to critical nearby structures.[1,8]

Main points

- Conventional contrast-enhanced computed tomography findings alone may not reliably predict whether a peritonsillar abscess (PTA) is amenable to successful drainage.
- Computed tomography-based radiomics can quantitatively analyze internal lesion heterogeneity and provide additional information beyond visual assessment.
- Among the tested models, the radiomics-based extreme gradient boosting model demonstrated high diagnostic performance [area under the receiver operating characteristic curve (AUC): 0.919], which was further improved by integrating clinical parameters in the combined model (AUC: 0.934).
- Pre-procedural prediction of drainage success may help reduce unnecessary invasive interventions and improve clinical decision-making in patients with PTA.

The mechanisms underlying failed drainage remain incompletely understood. It is conceivable that some collections, although radiologically consistent with an abscess, may represent partially organized or inspissated material with increased viscosity rather than freely drainable fluid.[9] However, conventional imaging assessment primarily focuses on the presence of a rim-enhancing collection and does not directly evaluate internal fluid characteristics, such as viscosity or organization, which may influence drainability.[10] This limitation highlights an unmet clinical need for quantitative, non-invasive imaging biomarkers capable of predicting drainability before intervention.

To address this challenge and minimize unnecessary invasive interventions, we propose a novel approach using radiomics and machine learning. Radiomics enables quantitative characterization of lesion heterogeneity, texture, and spatial complexity, which may reflect underlying differences in fluid composition, viscosity, and organization.[11,12] We hypothesize that radiomics-derived features may enable quantitative differentiation between liquefied, drainable collections and organized, non-drainable collections.

This study utilizes I&D as the reference standard for drainage success, as it provides a more definitive evaluation of the abscess cavity than needle aspiration, thereby minimizing potential technical failures. Therefore, the primary aim of this study is to develop and validate machine learning models—specifically random forest classifier (RFC), support vector machine (SVM), extreme gradient boosting (XGBoost), and decision tree—based on radiomics features extracted from venous phase CECT scans integrated with key clinical parameters (age, sex, trismus, and uvula deviation) to predict drainage success in patients with suspected PTA. By identifying non-drainable collections pre-procedurally, this approach may help avoid unnecessary invasive interventions, reduce morbidity, and optimize clinical decision-making.

Methods

The study was conducted at Ankara Bilkent City Hospital, a tertiary care center. Ethical approval was granted by the Bilkent City Hospital Medical Research Scientific and Ethical Review Board (TABED) (decision number: 2/1912/2026, date: 04.02.2026). Given the retrospective nature of the study, the use of anonymized data, and the lack of potential risk to participants, the requirement for informed consent was waived by the Ethical Review Board. All procedures were conducted in accordance with the principles outlined in the Declaration of Helsinki.

Study population

The institutional database was screened for patients who underwent CECT with a preliminary diagnosis of PTA. The exclusion criteria comprised age younger than 18 years, evaluation with imaging modalities other than CECT, prior drainage attempts or surgical interventions before imaging, prolonged antibiotic therapy (> 72 hours) before CT acquisition, presence of coexisting head and neck malignancy, and significant motion or beam-hardening artifacts impairing image quality.

The inclusion criteria consisted of adult patients diagnosed with PTA who underwent neck CECT acquired in the venous phase and subsequently received I&D with a clearly documented clinical outcome. Only patients whose CT examinations were performed within 24 hours before the intervention were included to ensure temporal consistency between imaging and treatment.

A retrospective analysis was performed on 94 eligible patients between January 2022 and December 2025. Of these, 42 patients were classified as the non-drainage group and 52 as the drainage group. Given the balanced distribution of the two groups (52 drainage vs. 42 non-drainage), specific oversampling or undersampling techniques were not required.

In the non-drainage group, there were 23 men and 19 women aged 21–74 years (mean age: 45.8 years). The drainage group consisted of 28 men and 24 women aged 19–76 years (mean age: 47.2 years).

Patient demographics (age and sex) and physical examination findings at admission (trismus and uvula deviation) were retrospectively retrieved from the institutional database. Due to the retrospective nature of the study, clinical findings were analyzed based on available documentation; data for trismus and uvula deviation were available for 79/94 (84%) and 82/94 (87%) patients, respectively.

The imaging data of the included patients were independently reviewed by two radiologists (İSP, with 26 years of experience, and BARM, with 8 years of experience).

The flowchart of the study is depicted in Figure 1.

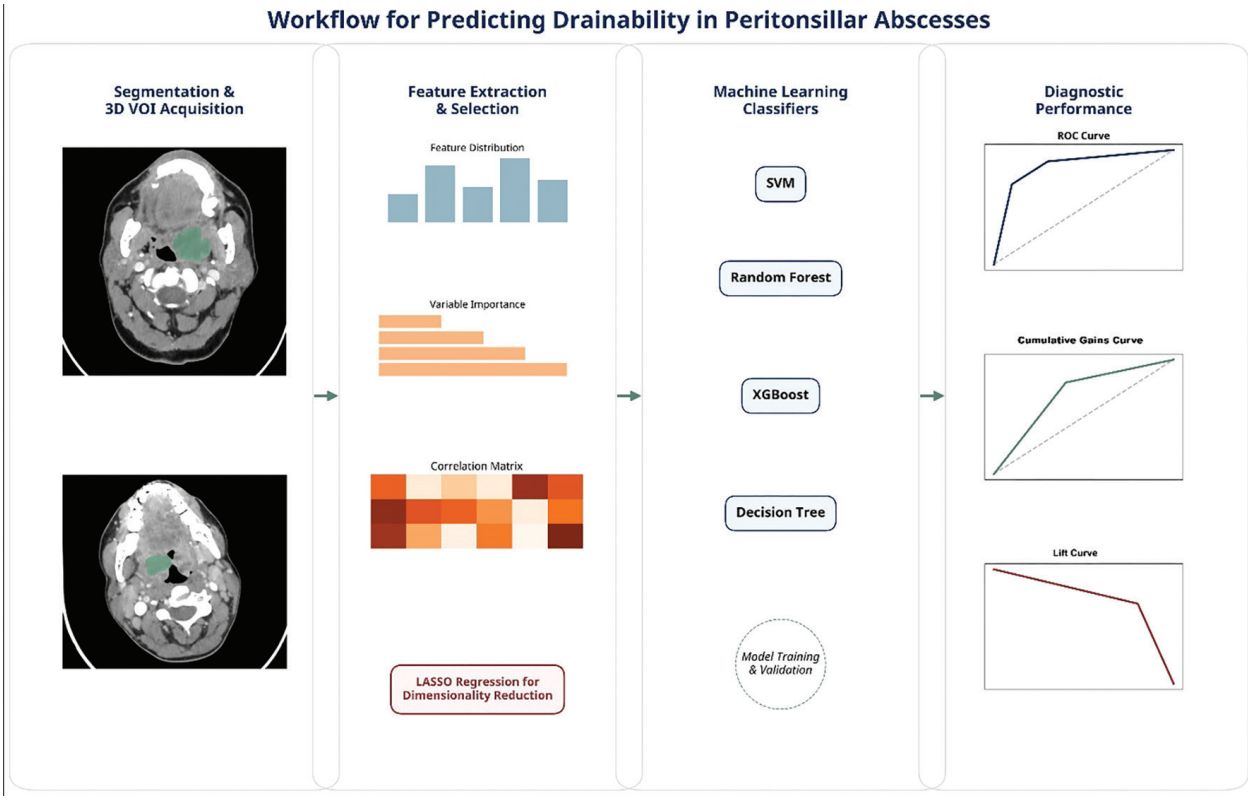


Figure 1. Flowchart of the study. 3D, three-dimensional; VOI, volume of interest; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; XGBoost, extreme gradient boosting; ROC, receiver operating characteristic.

Computed tomography acquisition and preprocessing

The examinations were performed using a 128-detector CT scanner (GE Revolution, General Electric, Milwaukee, WI, USA). Acquisition parameters consisted of a tube voltage of 100–120 kVp, a tube current of 250–270 mAs, a 512 × 512 reconstruction matrix, a 1.25-mm slice thickness, a body filter, and a pitch of 1.375. Intravenous iodinated contrast material (300 mgI/mL; 80–100 mL per patient) was administered at a rate of 3 mL/s. Contrast-enhanced images were acquired during the venous phase, approximately 60–90 seconds after contrast injection.

The selected CT datasets were imported into 3D Slicer software (version 5.6.2)[13] for three-dimensional manual segmentation of the PTA. Segmentations were carefully reviewed by the same radiologists to identify and correct potential inaccuracies, including the inadvertent inclusion of adjacent fat, air, or bone structures, as well as the unintended partial exclusion of the abscesses. Example images illustrating PTA segmentation are presented in Figure 2.

To assess intra- and interobserver reproducibility, 20 randomly chosen cases were re-segmented following the same standardized protocol.

Texture analysis

Radiomics features were extracted using the SlicerRadiomics extension based on the PyRadiomics library after completion of segmentations.[14] Feature extraction was performed on original, non-filtered images. All images were resampled to an isotropic voxel size of 1 × 1 × 1 mm^3 using linear interpolation. Intensity discretization was applied using a fixed bin width of 64 Hounsfield units, which was selected to balance noise reduction and the preservation of meaningful intensity heterogeneity, thereby improving the stability and reproducibility of radiomics features. No additional image filters (e.g., wavelet or Laplacian of Gaussian) were applied.

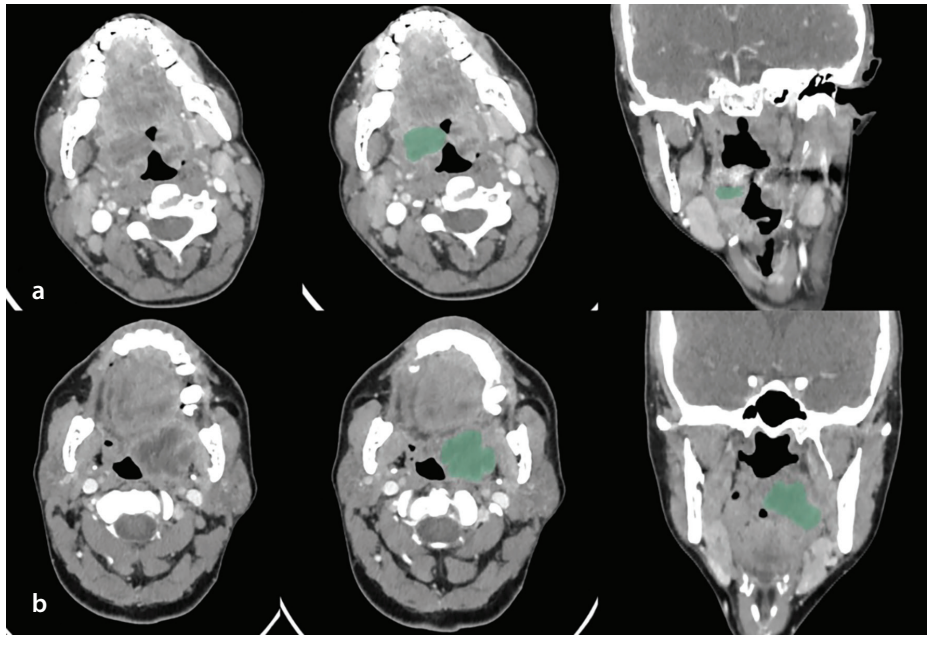


Figure 2. (**a, b**) Examples illustrating peritonsillar abscess segmentation.

Following the preset configuration, a total of 107 radiomics features were extracted. These features were categorized as shape-based features, first-order statistics, gray level co-occurrence matrix (GLCM), gray level run length matrix, neighboring gray tone difference matrix, gray level size zone matrix (GLSZM), and gray level dependence matrix.

Statistical analysis

The normality of data distribution was evaluated using the Kolmogorov–Smirnov test. Variables with a normal distribution were expressed as mean ± standard deviation, whereas non-normally distributed variables were reported as median and interquartile range.

Intraobserver agreement was assessed using the intraclass correlation coefficient (ICC), with values ≥ 0.75 considered indicative of good reproducibility.[15] Interobserver agreement of the selected features was assessed using Spearman's rank correlation coefficient (ρ), with values greater than 0.8 considered indicative of good reproducibility for each parameter.

Before feature selection and model training, feature standardization (Z-score normalization) was applied. Normalization parameters (mean and standard deviation) were calculated using the training set and subsequently applied to the validation set to prevent data leakage.

The least absolute shrinkage and selection operator (LASSO) regression was applied for dimensionality reduction and feature selection. The optimal regularization parameter (λ) was determined using 10-fold cross-validation on the training set. The relationship between the mean squared error (MSE) and λ values was examined to identify the most appropriate model. Feature importance was quantified based on the magnitude of the regression coefficients.

Missing clinical data were managed using complete-case analysis for the respective sub-cohorts.

The predictive performance of the combined clinical–radiomics model was compared against the radiomics-only and clinical-only models using the area under the receiver operating characteristic curve (AUC), sensitivity, and specificity to determine if the integration of clinical variables provided a significant improvement in predicting drainage success. DeLong's test was applied to compare the AUCs between different models.

Decision curve analysis was performed according to the method described by Vickers and Elkin[16] using Python (version 3.11), with custom-written code implementing standard net benefit calculations. The decision curve was visualized using the Matplotlib library to assess the clinical utility of the models across a range of threshold probabilities.

All statistical analyses were performed using XLSTAT statistical and data analysis software (version 2024.2.2; Addinsoft, NY, USA). A two-tailed *P* value of < 0.05 was considered statistically significant.

Machine learning

SVM,[17] RFC,[18] decision tree,[19] and XGBoost[20] algorithms were implemented for model development. The dataset was randomly partitioned into training and validation cohorts using a 70:30 split. Model performance was evaluated by calculating sensitivity, specificity, and the AUC.

The SVM model was configured using a linear kernel with a regularization parameter (C) set to 1.0 and a tolerance value of 0.001. Given the substantial variability in radiomics feature values, feature standardization was applied as a preprocessing step before model training. To reduce the potential influence of variability and selection bias on performance estimates, 10-fold bootstrapping was performed in accordance with the approach described by Vrigazova and Ivanov.[21]

The RFC model was constructed using a bagging approach with bootstrap sampling (random sampling with replacement). A total of 200 decision trees were generated, with the maximum tree depth limited to 20. Model evaluation additionally included calculation of the misclassification rate, analysis of the out-of-bag (OOB) error progression, and assessment of feature importance based on the mean decrease in accuracy.

Decision tree analysis was performed using a classification tree based on the chi-square Automatic Interaction Detection algorithm. The significance level was set at 0.05, with a maximum tree depth of 3 and a merge threshold of 0.05. Model validation was conducted using a two-fold cross-validation. In addition, model performance was evaluated using gain and lift curves.

The XGBoost model was configured with a maximum of 100 boosting iterations, a learning rate of 0.3, and a maximum tree depth of 6. The objective function was specified as quadratic, with the minimum loss reduction (gamma) set to 0. In addition to the primary performance metrics, gain and lift curves were generated for further evaluation of model performance.

To evaluate the added value of clinical integration, three distinct modeling approaches were developed: (1) a clinical-only model, (2) radiomics-only models using four machine learning algorithms (SVM, RFC, decision tree, and XGBoost), and (3) a combined clinical–radiomics model. The clinical parameters were integrated with the selected radiomics features using logistic regression to develop a combined clinical–radiomics nomogram.

Results

Study population

A total of 94 patients with PTA were included in this retrospective study, comprising 42 patients in the non-drainage group and 52 in the drainage group. The mean age was 45.8 ± 12.4 years (range: 21–74 years) in the non-drainage group and 47.2 ± 13.1 years (range: 19–76 years) in the drainage group, with no significant difference between the groups (independent-samples t-test, $P = 0.621$).

The non-drainage group included 23 men (54.8%) and 19 women (45.2%), whereas the drainage group consisted of 28 men (53.8%) and 24 women (46.2%). Sex distribution did not differ significantly between the groups (χ^2 test, $P = 0.929$).

Regarding physical examination findings, trismus was documented in 79 patients, with a prevalence of 73.1% (32/44) in the drainage group compared with 61.9% (22/35) in the non-drainage group ($P = 0.349$). Uvula deviation was observed in 82 patients, with a prevalence of 80.8% (36/45) in the drainage group and 71.4% (26/37) in the non-drainage group ($P = 0.307$).

Inter- and intraobserver reproducibility

Twenty cases were randomly selected during the segmentation process. Interobserver reproducibility was assessed using the first-order radiomics features "mean" and "range." Both features showed strong agreement between the two observers, as reflected by Spearman's rank correlation coefficients (mean: $\rho = 0.837$, $P = 0.016$; range: $\rho = 0.822$, $P = 0.023$).

The same randomly selected cases were used to evaluate intraobserver reproducibility. Each observer repeated the

segmentation procedure twice, and the first-order "mean" and "range" features were analyzed. Intraobserver agreement was good for both observers: the mean ICC was 0.852 (range: 0.773–0.946) for Observer 1 (İSP) and 0.844 (range: 0.762–0.927) for Observer 2 (BARM).

Dimensionality reduction of the parameters

The LASSO regression selected 10 features that exhibited low intercorrelation and strong discriminatory power. The optimal λ value was determined to be 0.043. The correlation matrix is provided in Supplementary Figure 1, whereas the most relevant features and their standardized coefficients are summarized in Table 1. Supplementary Figure 2 shows the association between the λ parameter and the MSE.

Diagnostic efficacy of machine learning algorithms

The diagnostic performance of the SVM model was as follows: sensitivity of 87.1% [95% confidence interval (CI): 84.1–90.3], specificity of 87.25% (95% CI: 83.9–90.25), and an AUC of 0.899 (95% CI: 0.869–0.928).

The diagnostic accuracy metrics of the RFC model consisted of a misclassification rate of 0.15, sensitivity of 88.3% (95% CI: 85.25–91.35), specificity of 82.2% (95% CI: 78.63–85.74), and an AUC of 0.85 (95% CI: 0.811–0.892). Figure 3 displays the progression of the OOB error over iterations.

For the XGBoost model, sensitivity was 88.65% (95% CI: 85.05–92.19), specificity was 92.55% (95% CI: 89.4–95.6), and the AUC reached 0.919 (95% CI: 0.887–0.951). Figure 4 illustrates the gain and lift curves associated with the model.

Table 1. The most significant radiomics parameters based on LASSO regression

Parameter	Standardized coefficient
Small dependence emphasis	0.207
Coarseness	0.180
Run percentage	0.166
Mean absolute deviation	–0.140
Short run high gray level emphasis	–0.120
90th percentile	–0.088
Interquartile range	–0.083
Least axis length	0.074
Cluster shade	–0.062
Run length non-uniformity	0.059

LASSO, least absolute shrinkage and selection operator.

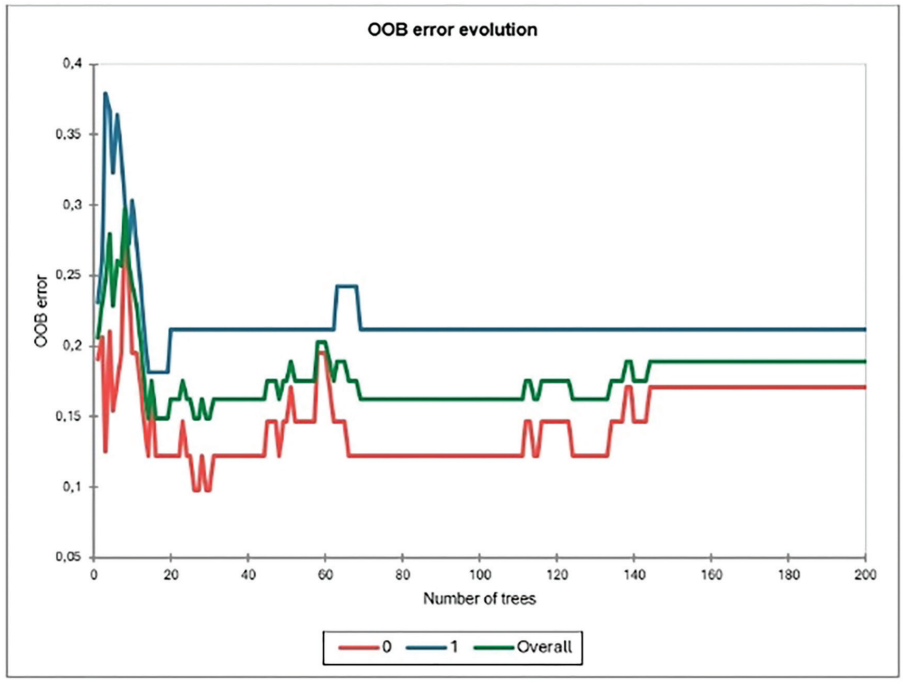


Figure 3. Evolution of the out-of-bag (OOB) error in the random forest classifier. The x-axis represents the number of constructed trees, and the y-axis indicates the error rate. The red curve corresponds to the error rate for the drainage group, the blue curve represents the error rate for the non-drainage group, and the green curve shows the overall OOB observations.

Decision tree analysis demonstrated a sensitivity of 90.24% (95% CI: 86.41–93.85), a specificity of 88.75% (95% CI: 85.05–92.55), and an AUC of 0.887 (95% CI: 0.861–0.914). Figure 5 presents the corresponding gain and lift curves.

Performance of the combined clinical–radiomics model

Although the clinical-only model demonstrated limited predictive performance, with an AUC of 0.642 (95% CI: 0.581–0.703), integrating clinical parameters into the radiomics-based XGBoost model resulted in a significant improvement. The combined clinical–radiomics model achieved the highest diagnostic performance, yielding an AUC of 0.934 (95% CI: 0.902–0.966), a sensitivity of 90.2%, and a specificity of 94.1%. This combined approach outperformed the radiomics-only XGBoost model (AUC: 0.919), highlighting the synergistic value of clinical variables in predicting drainage success. DeLong's test demonstrated that the combined clinical–radiomics model achieved a statistically significantly higher AUC than all other models (Table 2).

Decision curve analysis

Decision curve analysis showed that the combined model provided the highest net benefit across most threshold probabilities (Figure 6). The radiomics-only model also performed well but remained below the combined model. In contrast, the clinical-only model showed a lower net benefit and more limited clinical usefulness than the other models. Both the combined and radiomics-only models outperformed the "treat all" and "treat none" strategies over a wide range of threshold probabilities.

Discussion

This study explores the potential of radiomics-based machine learning models, derived from venous phase CECT, to predict the drainage success of PTA. Among the evaluated algorithms, the XGBoost model exhibited the highest diagnostic performance with an AUC of 0.919, followed by SVM (0.899) and the decision tree (0.887). These results suggest that quantitative radiomics analysis may capture subtle internal characteristics of peritonsillar collections that are not readily apparent to the human eye, potentially offering a non-invasive tool to inform clinical decision-making and reduce unnecessary invasive procedures.

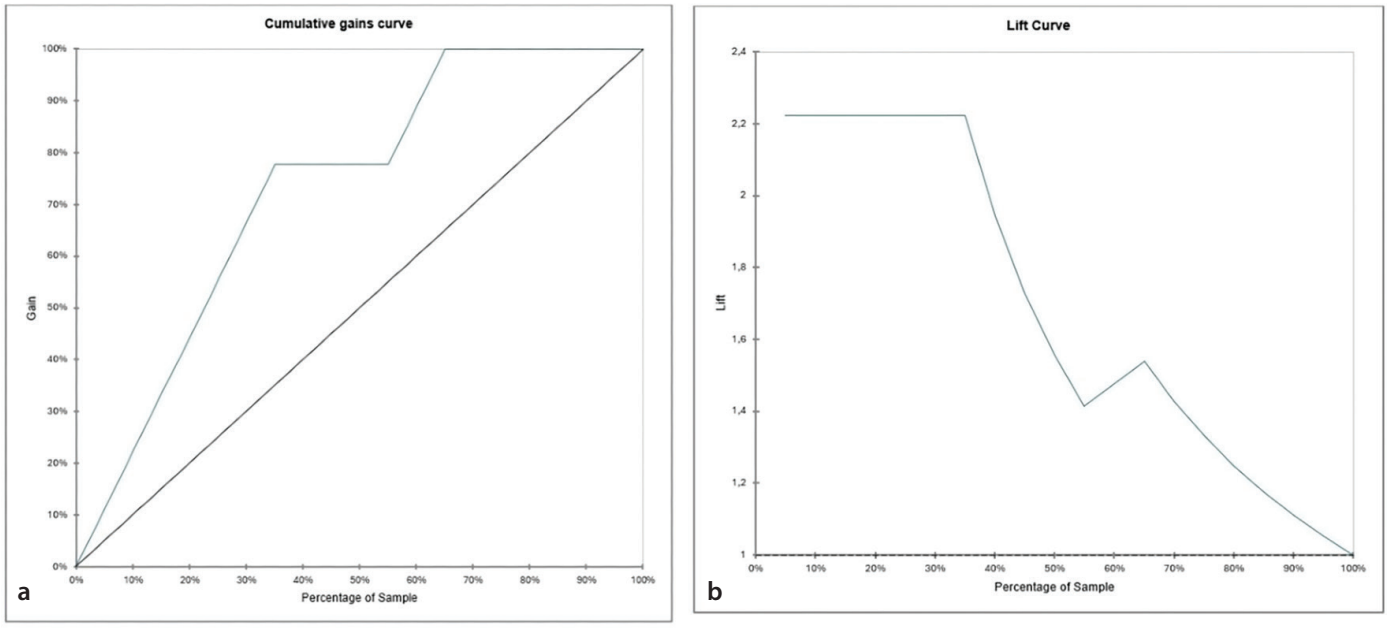


Figure 4. The diagnostic efficacy of XGBoost depicted with **(a)** gain and **(b)** lift curves. The gain curve illustrates the true positive rate in percent versus the percent of total observations. The lift curve reflects the ratio of the model's gain to the percentage of the dataset utilized. XGBoost, extreme gradient boosting.

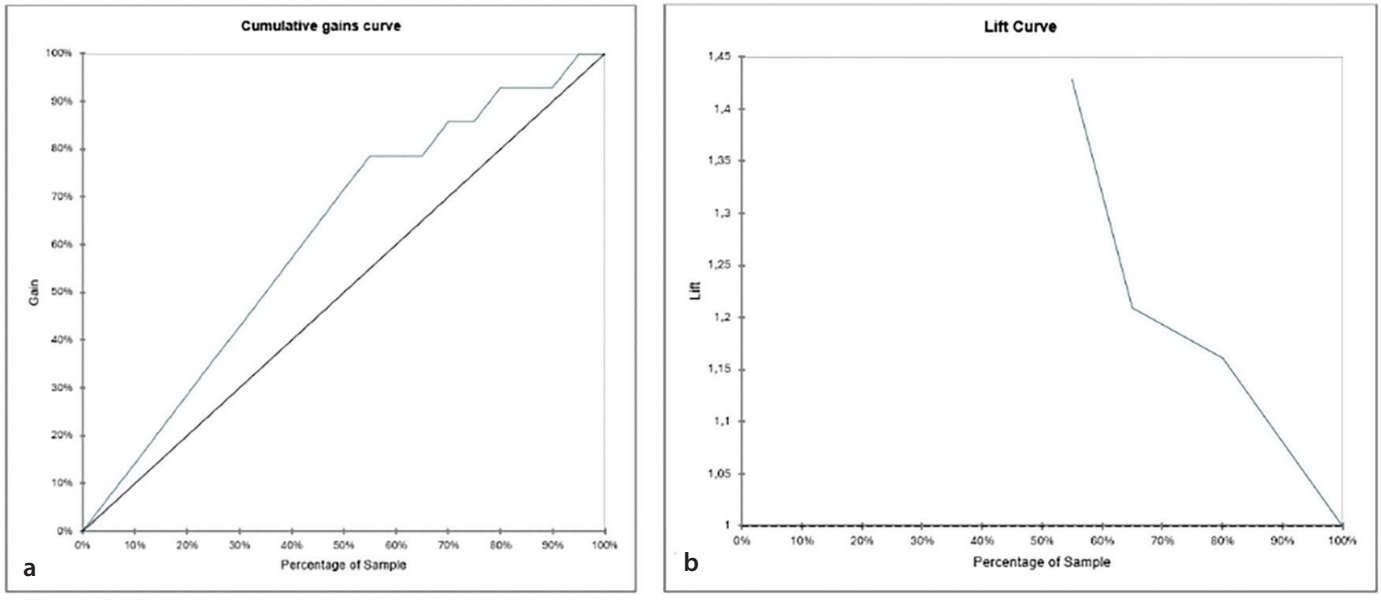


Figure 5. The diagnostic performance of the decision tree demonstrated with **(a)** gain and **(b)** lift curves. The gain curve depicts the true-positive rate in percent versus the percentage of total observations, and the lift curve represents the ratio of the model's gain to the percentage of the dataset used.

Table 2. Diagnostic performance comparison of the models and DeLong's test results

Model type	AUC (95% CI)	Sensitivity (recall) (%)	Specificity (%)	Precision (%)	F1 score	DeLong's *P* value*
Clinical only	0.642 (0.581–0.703)	62.4%	58.1%	64.8	0.636	<0.001
SVM	0.899 (0.869–0.928)	87.1%	87.2%	89.4	0.882	0.038
Decision tree	0.887 (0.861–0.914)	90.2%	88.7%	90.8	0.905	0.029
RFC	0.850 (0.811–0.892)	88.3%	82.2%	86.0	0.871	0.012
XGBoost	0.919 (0.887–0.951)	88.6%	92.5%	93.6	0.910	0.046
XGBoost + clinical	0.934 (0.902–0.966)	90.2%	94.1%	95.0	0.925	Reference

P values represent the statistical comparison of each model's AUC against the combined model (reference) using DeLong's test. A *P* value <0.05 indicates a statistically significant difference in diagnostic performance. AUC, area under the curve; CI, confidence interval; SVM, support vector machine; RFC, random forest classifier; XGBoost, extreme gradient boosting.

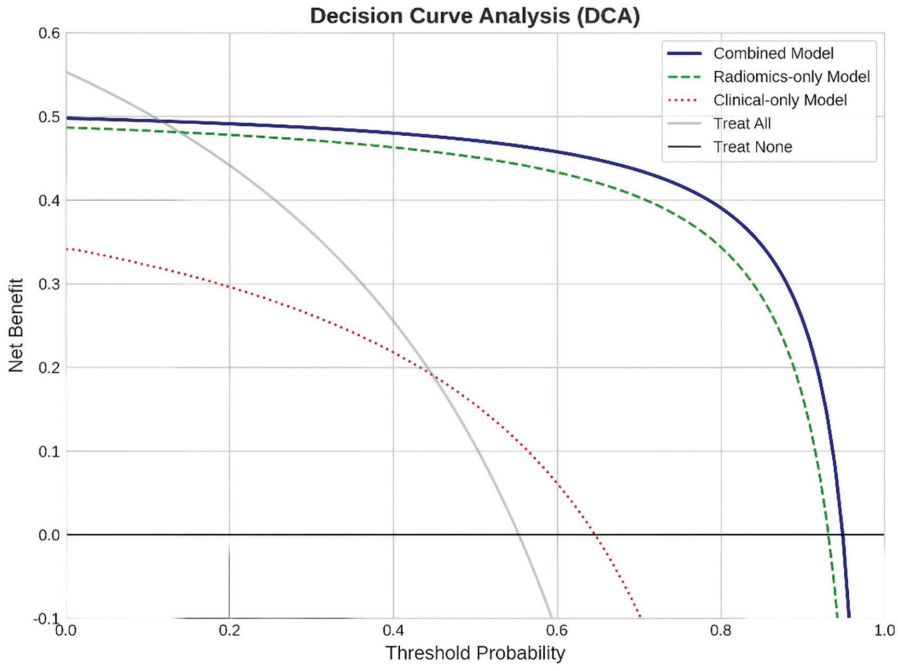


Figure 6. Decision curve analysis showing the net benefit of the clinical model, radiomics model, and combined clinical–radiomics model across a range of threshold probabilities, along with the "treat all" and "treat none" strategies. The combined model demonstrates superior clinical utility across most threshold probabilities.

In this study, CT-based radiomics features were analyzed in conjunction with clinical parameters—including age, sex, trismus, and uvula deviation—to evaluate their collective contribution to predicting drainage success in PTA. The combined clinical–radiomics model demonstrated an AUC of 0.934, compared with 0.919 for the radiomics-only model. This observation suggests that while radiomics captures internal lesion heterogeneity, clinical variables provide additional context that contributes to the model's overall predictive performance and clinical applicability. The inclusion of clinical signs such as trismus and uvula deviation is relevant, as these findings are traditional indicators of PTA severity. Although these clinical parameters alone showed limited predictive value in our cohort (AUC: 0.642), their integration with radiomics data suggests a multi-dimensional approach to evaluating peritonsillar collections.

We utilized I&D as the reference standard to potentially reduce the technical uncertainties associated with needle aspiration, such as needle malposition, the limitations of small-gauge needles in aspirating viscous material, or the presence of multiloculated septations. This approach was intended to provide a more consistent ground truth for our models by focusing on the intrinsic characteristics of the collection rather than procedural variables.

The management of PTA traditionally relies on the identification of a rim-enhancing collection on CECT to justify needle aspiration or I&D.[4,5] However, the diagnostic accuracy of CECT in identifying truly drainable collections remains a subject of debate. Previous studies have reported that while CECT is highly sensitive (approaching 100%), its specificity and positive predictive value (PPV) in identifying successful drainage outcomes can be as low as 75%.[4,10] Hagelberg et al.,[10] in a recent meta-analysis, highlighted the substantial variability in the PPV of CECT for neck abscesses, suggesting that the classic rim enhancement sign is not always synonymous with the presence of freely drainable pus.

This diagnostic uncertainty often leads to a "dry tap," where no purulent material is obtained despite radiological evidence of an abscess. This challenge is echoed in previous literature, where reported false-positive rates for identifying drainable collections range from approximately 20% to 25%, reflecting the inherent limitations of conventional diagnostic standards.[6,7] In our cohort, 44.7% (42/94) of the cases resulted in a non-drainable outcome—a rate higher than these reported averages. This discrepancy likely reflects the complex nature of cases managed at our tertiary center, where collections may represent more organized or viscous material that is resistant to successful drainage.[9]

Our findings align with and extend recent research into predictive markers for PTA drainage. Although previous studies have explored machine learning for PTA diagnosis or surgical intervention prediction using clinical and conventional imaging features,[9,22] to our knowledge, this is the first study to utilize CT radiomics-based machine learning to predict the success of PTA drainage. For instance, Wilson et al.[22] developed a machine learning model incorporating clinical symptoms and findings—such as trismus, neck pain, and otalgia—to predict PTA, demonstrating moderate predictive performance with an accuracy of up to 0.72. However, clinical signs are often subjective and may not always clearly differentiate collections requiring drainage from those that do not. In contrast, our approach leverages high-dimensional quantitative data directly from the lesion. Li et al.[9] recently identified that specific CT findings, such as soft palate effacement, continuous ring enhancement, and larger abscess size, are significantly associated with successful drainage. Although conventional imaging features are clinically valuable, they are inherently subject to interobserver variability and may not adequately capture the internal structural complexity of an abscess.

The role of imaging in predicting drainability has also been explored through other modalities, most notably ultrasound (US). Kim et al.[23] and Costantino et al.[24] have shown that intraoral US has high sensitivity but moderate specificity in identifying PTA, with the advantages of being radiation-free, cost-effective, and capable of real-time visualization of fluid dynamics to assess the presence of drainable pus. However, the application of US can be technically challenging in patients with severe trismus and is inherently operator-dependent. Although US remains a preferred initial tool for many clinicians due to its accessibility, CT is frequently performed in clinical practice to evaluate the extent of the infection into deeper neck spaces or when the diagnosis remains ambiguous. Our radiomics-based approach aims to enhance the utility of these already-acquired CT scans by providing a level of content characterization that traditional visual inspection might lack. Feature groups such as the GLCM and GLSZM, which contributed to our selected model features, quantify the spatial relationships and distribution patterns of voxel attenuation values. These parameters likely reflect the internal heterogeneity of the abscess cavity, enabling differentiation between freely drainable purulent collections and more organized, highly viscous, or septated components.[11,12]

The XGBoost model achieved the best performance among the evaluated algorithms, consistent with its success in other radiomics applications. Its ability to handle non-linear relationships and its robust regularization parameters make it particularly well-suited for high-dimensional radiomics datasets.[20] The high specificity achieved by our models (up to 92.5% for XGBoost) is relevant from a clinical perspective. In the context of PTA, high specificity results in a low rate of false positives for drainage, effectively identifying patients who are unlikely to benefit from an invasive procedure. By opting for conservative medical management in these cases, clinicians can avoid the risks of iatrogenic bleeding, pain, and injury to the internal carotid artery.[8]

Despite the promising results, this study has several limitations. First, its retrospective, single-center design and relatively small sample size may limit the generalizability of the findings. Although we employed bootstrapping and cross-validation to ensure model robustness, external validation in larger, multi-center prospective cohorts is essential to confirm broader applicability. Second, the manual segmentation process, though reviewed by experienced radiologists, is time-intensive and inherently prone to interobserver variability. The integration of automated or semi-automated segmentation tools in future iterations could improve efficiency, minimize observer dependency, and facilitate clinical adoption. Third, the retrospective nature of the study led to incomplete clinical data. Although trismus and uvula deviation were included in the analysis, they were documented for only 84% and 87% of the cohort, respectively. Furthermore, other relevant parameters—such as symptom duration and specific laboratory markers—could not be analyzed due to missing records. Future prospective studies utilizing standardized clinical reporting protocols may further refine and improve model performance.

In conclusion, radiomics-based machine learning models offer a highly accurate, quantitative approach to predicting drainage success in PTA. By identifying non-drainable collections before intervention, this method has the potential to optimize treatment strategies, reduce patient morbidity, and enhance the efficiency of emergency care delivery. Future research should prioritize prospective validation and the development of intuitive software interfaces to support real-time clinical decision-making.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

Supplementary Figures 1-2: https://d2v96fxpocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/85c3ced9-ef05-4c99-93d2-9ae-fa69b526c.pdf

References

1. Galioto NJ. Peritonsillar abscess. *Am Fam Physician*. 2017;95(8):501-506. [Crossref]
2. Johnson RF, Stewart MG. The contemporary approach to diagnosis and management of peritonsillar abscess. *Curr Opin Otolaryngol Head Neck Surg*. 2005;13(3):157-160. [Crossref]
3. Herzon FS, Martin AD. Medical and surgical treatment of peritonsillar, retropharyngeal, and parapharyngeal abscesses. *Curr Infect Dis Rep*. 2006;8(3):196-202. [Crossref]
4. Scott PM, Loftus WK, Kew J, Ahuja A, Yue V, van Hasselt CA. Diagnosis of peritonsillar infections: a prospective study of ultrasound, computerized tomography and clinical diagnosis. *J Laryngol Otol*. 1999;113(3):229-232. [Crossref]
5. Steyer TE. Peritonsillar abscess: diagnosis and treatment. *Am Fam Physician*. 2002;65(1):93-96. Erratum in: *Am Fam Physician* 2002;66(1):30. [Crossref]
6. Powell J, Wilson JA. An evidence-based review of peritonsillar abscess. *Clin Otolaryngol*. 2012;37(2):136-145. [Crossref]
7. Ban MJ, Jung JY, Kim JW, et al. A clinical prediction score to determine surgical drainage of deep neck infection: a retrospective case-control study. *Int J Surg*. 2018;52:131-135. [Crossref]
8. Thomas JA, Ware TM, Counselman FL. Internal carotid artery pseudoaneurysm masquerading as a peritonsillar abscess. *J Emerg Med*. 2002;22(3):257-261. [Crossref]
9. Li AC, Wadley A, Aciu S, Zhao H, Jain N, Soliman AMS. Computed tomographic and clinical findings predictive of successful peritonsillar abscess drainage. *Laryngoscope*. 2026;136(1):151-157. [Crossref]
10. Hagelberg J, Pape B, Heikkinen J, Nurminen J, Mattila K, Hirvonen J. Diagnostic accuracy of contrast-enhanced CT for neck abscesses: a systematic review and meta-analysis of positive predictive value. *PLoS One*. 2022;17(10):e0276544. [Crossref]
11. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446. [Crossref]
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [Crossref]
13. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-1341. [Crossref]
14. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107. [Crossref]
15. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. Erratum in: *J Chiropr Med*. 2017;16(4):346. [Crossref]
16. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. [Crossref]
17. Borstelmann SM. Machine learning principles for radiology investigators. *Acad Radiol*. 2020;27(1):13-25. [Crossref]
18. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. [Crossref]
19. Navada A, Ansari AN, Patil S, Sonkamble BA. Overview of use of decision tree algorithms in machine learning. In: 2011 IEEE Control and System Graduate Research Colloquium; 2011; Shah Alam, Malaysia. p. 37-42. [Crossref]
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785-794. [Crossref]
21. Vrigazova B, Ivanov I. Tenfold bootstrap procedure for support vector machines. *Comput Sci*. 2020;21(2):241-257. [Crossref]
22. Wilson MB, Ali SA, Kovatch KJ, Smith JD, Hoff PT. Machine learning diagnosis of peritonsillar abscess. *Otolaryngol Head Neck Surg*. 2019;161(5):796-799. [Crossref]
23. Kim DJ, Burton JE, Hammad A, et al. Test characteristics of ultrasound for the diagnosis of peritonsillar abscess: a systematic review and meta-analysis. *Acad Emerg Med*. 2023;30(8):859-869. [Crossref]
24. Costantino TG, Satz WA, Dehnkamp W, Goett H. Randomized trial comparing intraoral ultrasound to landmark-based needle aspiration in patients with suspected peritonsillar abscess. *Acad Emerg Med*. 2012;19(6):626-631. [Crossref]