ScienceDirect

Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology

Supports open access

3.6
CiteScore

2.4
Impact Factor

Submit your article ↗

Guide for authors

Menu

Search in this journal

Editorial board

Editorial board by country/region

87 editors and editorial board members in 12 countries/regions

1 United States (61)

2 Brazil (9)

3 Canada (4)

See more editors by country/region

Editor in Chief

Mark W. Lingen, DDS, PhD

The University of Chicago, Chicago, IL, United States

Oral and Maxillofacial Surgery

Section Editor

Antonia Kolokythas, DDS, MSc, MSed, FACS

Vanderbilt University Medical Center, Nashville, Tennessee, United States

Associate Editor

Pooja Gangwani, DDS, MPH

Temple University, Philadelphia, Pennsylvania, United States

Editorial board

Kevin Arce, DMD, MD

Mayo Foundation for Medical Education and Research, Rochester, Minnesota, United States

Martin Batstone

Royal Brisbane and Women's Hospital, Herston, Australia

Jeffrey D. Bennett, Anesthesia/Facial Pain

Regenstrief Institute Inc, Indianapolis, Indiana, United States

George Blakey

Allen Cheng, MD, DDS

Stephanie J. Drew, DMD, FACD, FACS

Emory University School of Medicine, Atlanta, Georgia, United States

Sean Edwards, MD, DDS, FRCD(C), FACS

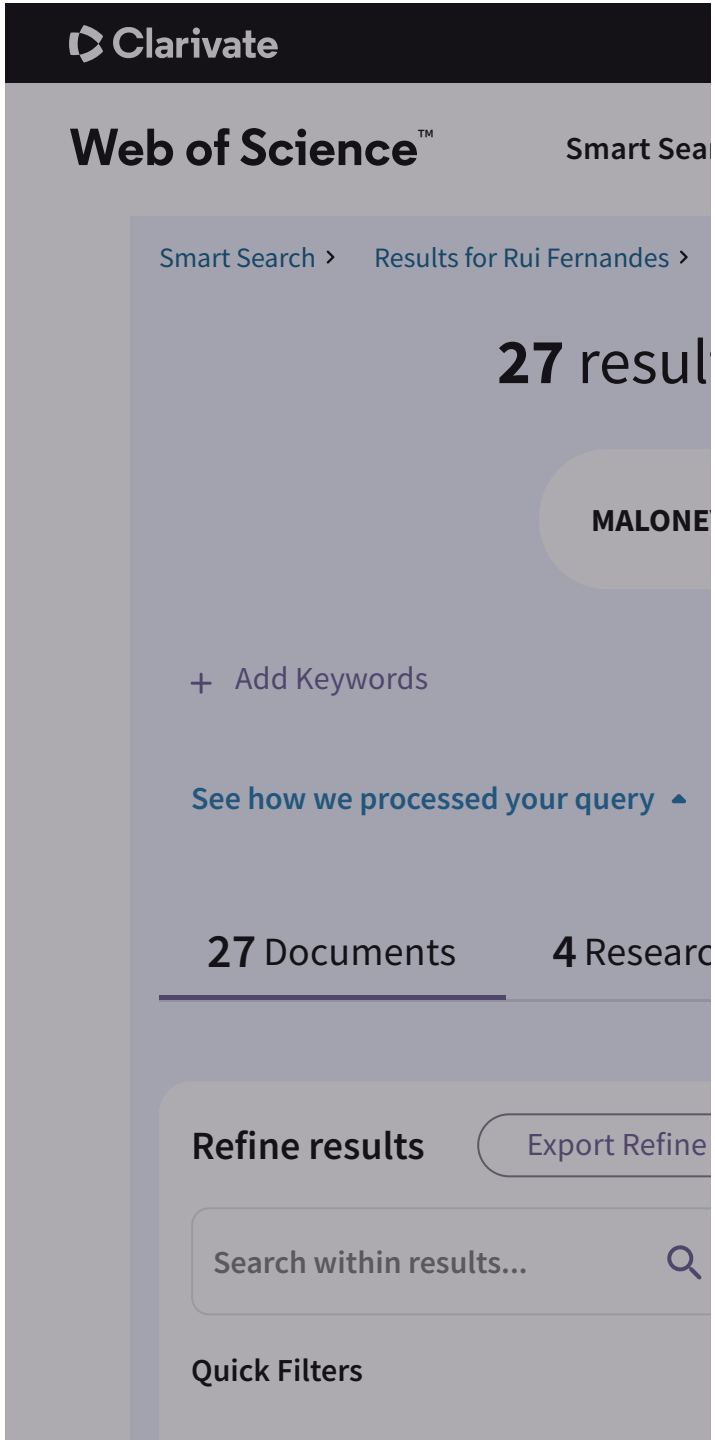
University of Michigan Michigan Medicine, Ann Arbor, Michigan, United States

Rui Fernandes, DDS

University of North Florida, Jacksonville, Florida, United States

Kim E. Goldman

University of California San Francisco, San Francisco, California, United States



Journal information

ORAL SURGERY ORAL MEDICINE ORAL PATHOLOGY ORAL RADIOLOGY

Publisher name: ELSEVIER SCIENCE INC

Journal Impact Factor ™

2.4	2.3
2025	Five Year

JCR Category	Category Rank	Category Quartile
DENTISTRY, ORAL SURGERY & MEDICINE *in SCIE edition*	63/165	Q2

Source: Journal Citation Reports 2025. Go to Journal Citation Reports

Journal Citation Indicator ™

0.82	0.84
2025	2024

JCI Category	Category Rank	Category Quartile
DENTISTRY, ORAL SURGERY &	72/165	Q2

Review Article
Open Access
Enriched Cited References

Publication Years

Show Final Publication Year

2026
2025
2024
2023
2022

See all >

Document Types

Article
Review Article
Meeting Abstract
Book Review
Correction

See all >

Journal information

ORAL SURGERY ORAL MEDICINE ORAL PATHOLOGY ORAL RADIOLOGY

Publisher name: ELSEVIER SCIENCE INC

Journal Impact Factor ™

2.4	2.3
2025	Five Year

JCR Category	Category Rank	Category Quartile
DENTISTRY, ORAL SURGERY & MEDICINE *in SCIE edition*	63/165	Q2

Source: Journal Citation Reports 2025. Go to Journal Citation Reports

Journal Citation Indicator ™

0.82	0.84
2025	2024

JCI Category	Category Rank	Category Quartile
DENTISTRY, ORAL SURGERY &	72/165	Q2

36

Vol. 00 No. 00 ■■ 2026

Differentiation of benign and malignant parotid gland tumors using T2-weighted MRI–based radiomics: the role of feature selection and dimensionality reduction

Özgür Çakır, MD,[a] Coşku Öksüz, PhD,[b] Simge Şahin, MD,[a] Alican Kuran, MSc,[c] Demet Doğan, MD,[d] Büşra Yaprak Bayrak, MD,[e] Fatih Mutlu, MD,[f] Murat Öztürk, MD,[f] and Oğuzhan Urhan, PhD[g]

Objective. To identify the optimal machine learning strategy, including feature selection and dimensionality reduction methods, for differentiating benign and malignant parotid gland tumors using radiomic features from T2-weighted magnetic resonance images (MRI).

Material and Methods. This retrospective study included 95 patients with histopathologically confirmed parotid gland tumors (49 benign, 46 malignant). A total of 107 radiomic features (morphological, first-order, and texture features) were extracted using PyRadiomics. Seven feature selection and dimensionality reduction methods were evaluated in combination with five classifiers. Model performance was assessed using five-fold cross-validation, all data preprocessing and feature selection steps were entirely embedded within the cross-validation framework for data leakage. Performance metrics included accuracy, precision, recall, F1-score, specificity, and area under the receiver operating characteristic curve (AUC).

Results. The full feature set showed limited performance (AUC: 0.65). PCA, PCR, and mRMR did not significantly improve results, whereas recursive feature elimination (RFE) consistently identified more informative features. The RFE–SVM model achieved the best performance with an AUC of 0.85.

Conclusion. RFE combined with SVM improves the differentiation of benign and malignant parotid gland tumors on T2-weighted MRI. These findings highlight the importance of model-based feature selection for developing clinically applicable radiomic models. (Oral Surg Oral Med Oral Pathol Oral Radiol 2026;000:1–14)

Salivary gland tumors represent a heterogeneous group of neoplasms with complex histopathological features and account for approximately 3% to 6% of all tumors in the head and neck region.[1] Parotid gland tumors comprise about 80–85% of salivary gland tumors and therefore constitute the most frequently encountered subgroup in clinical practice.[2] Approximately 80% of parotid gland tumors are benign, with pleomorphic adenoma and Warthin tumor being the most common entities, whereas the remaining ~20% are malignant.[3,4] Preoperative differentiation between benign and malignant parotid tumors is of major importance because it helps determine the extent of surgery, ranging from superficial parotidectomy to radical excision that may require facial nerve resection, and it may also influence the need for neck dissection.[5] Although fine-needle aspiration cytology is widely used in the diagnostic workup, it may be insufficient for accurate classification of certain tumor subtypes due to nonrepresentative sampling and cytomorphologic overlap, particularly between low-grade malignancies and benign lesions.[6,7]

Magnetic resonance imaging (MRI) is the preferred modality for evaluating parotid gland masses because of its superior soft tissue contrast and multiplanar capability.[5-8] T2-weighted (T2w) sequences provide valuable information about tumor cellularity and stromal composition in parotid gland imaging.[9] However, qualitative interpretation of signal intensity alone may be misleading: benign lesions such as Warthin tumors can mimic imaging features associated with malignancy, whereas low-grade carcinomas may demonstrate deceptively "benign appearing" T2 hyperintensity.[10] This overlap in conventional MRI findings highlights

Statement of Clinical Relevance

Radiomic features extracted from T2-weighted MRI images of parotid gland tumors can accurately differentiate benign from malignant lesions when appropriate feature elimination strategies and well-selected machine learning algorithms are applied.

[a]Department of Radiology, Faculty of Medicine, Kocaeli University, Kocaeli, Türkiye.
[b]Department of Electrical and Electronics Engineering, Izmir Bakircay University, Izmir, Türkiye.
[c]Department of Oral and Maxillofacial Radiology, Faculty of Dentistry, Kocaeli University, Kocaeli, Türkiye.
[d]Department of Radiology, Faculty of Medicine, Okan University, İstanbul, Türkiye.
[e]Department of Medical Pathology, Faculty of Medicine, Kocaeli University, Kocaeli, Türkiye.
[f]Department of Otorhinolaryngology, Faculty of Medicine, Kocaeli University, Kocaeli, Türkiye.
[g]Department of Electronics and Telecom Engineering, Kocaeli University, Kocaeli, Türkiye.
Corresponding author: Alican Kuran. E-mail address: alican.kuran@outlook.com
Received for publication Feb 1, 2026; returned for revision Apr 19, 2026; accepted for publication Apr 22, 2026.
© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
2212-4403/$-see front matter
https://doi.org/10.1016/j.oooo.2026.04.021

the need for more objective, quantitative diagnostic biomarkers.

Radiomics has emerged as a promising solution to these diagnostic challenges by extracting high-dimensional quantitative data from medical images that are imperceptible to the human eye.[11] By transforming radiological images into mineable quantitative data, radiomics allows for the assessment of subtle intratumoral heterogeneity that may aid in the differentiation of salivary gland tumors.[12-14] Nevertheless, the clinical translation of radiomics is hindered by the "curse of dimensionality," where the vast number of extracted features exceeds the number of available samples, leading to overfitting and poor model generalization.[15]

To develop a robust and reliable model, redundant and irrelevant features must be eliminated. Accordingly, the vast majority of contemporary radiomics studies employ either feature selection methods or dimensionality reduction/feature projection techniques.[16-18] Feature selection methods aim to identify features that are strongly associated with the target problem and outcome variable and that exhibit high predictive power.[16] In contrast, dimensionality reduction/feature projection methods seek to recombine the original features to generate new, lower-dimensional representations.[17] A comprehensive comparison across diverse radiomic datasets reported that, overall, feature selection approaches outperformed feature projection methods, although the optimal strategy was dataset-dependent; therefore, both approaches should be evaluated to achieve maximal predictive performance.[17] Although feature selection is widely recognized as one of the most critical steps in the radiomics analysis pipeline previous studies focusing on parotid gland tumors have predominantly compared the performance of different machine learning algorithm and no study to date has systematically compared feature selection and dimensionality reduction methods to determine which approach yields more effective radiomic outputs for this topic.[12,19-27] Moreover, there is currently no consensus in the literature regarding the most effective radiomic feature selection or dimensionality reduction/feature projection approach for the differential diagnosis of benign and malignant tumors originating from the parotid gland on the MRI.

In this retrospective study, we aim to address this methodological gap by systematically evaluating feature selection and dimensionality reduction strategies for differentiating histopathologically confirmed benign and malignant parotid gland tumors using T2-weighted MRI. We also compare the diagnostic performance of machine learning models trained on feature sets derived from the evaluated feature reduction strategies.

MATERIALS AND METHODS

Study design

The workflow of this retrospective study consists of the following stages: (1) segmentation of benign and malignant parotid gland tumors on T2-weighted MR images, (2) extraction of radiomic features from the regions of interest (ROIs) obtained from segmentation using the PyRadiomics software, (3) application of various feature selection and dimensionality reduction methods to identify the most discriminative radiomic features for benign−malignant tumor differentiation, (4) development and training of different machine learning algorithms using the selected feature sets, (5) comparison of model performances to determine the most appropriate feature reduction strategy and machine learning algorithm. The study protocol was approved by the Non-Interventional Ethics Committee of Kocaeli University (Decision No: KÜ GOKAEK-2025/22/22, Date: 6.11.2025), and the study was conducted in accordance with the principles of the Declaration of Helsinki.

Dataset and segmentation of the tumors

Patients with available preoperative T2-weighted magnetic resonance (MR) images were retrospectively included from the magnetic resonance imaging archive of the Department of Radiology, Faculty of Medicine, Kocaeli University. In all included cases, the diagnosis of parotid gland tumor was histopathologically confirmed, and the tumors were verified as benign or malignant. Inclusion criteria required that the entire tumor be clearly evaluable on MR images and that image quality be sufficient to allow reliable segmentation. The exclusion criteria included cases with pronounced motion or metal artifacts on MR images that impaired radiomic analysis, cases in which tumor boundaries could not be clearly delineated and reliable segmentation could not be performed, recurrent parotid gland tumors, and metastatic lesions or tumors with a primary origin outside the parotid glands.

Based on the inclusion and exclusion criteria, a total of 95 patients with parotid gland tumors, including 49 benign and 46 malignant cases, were enrolled in the study. MR images were exported in DICOM format, and manual tumor segmentation was performed using the 3D Slicer software package (version 5.6.2).[28] To capture the full extent of tumor heterogeneity, segmentation was performed slice by slice on axial images rather than on a single slice, thereby generating 3-dimensional volumes of interest (VOIs). The segmentation of benign and malignant parotid gland tumours and the 3D reconstruction of the segmented area are shown in Figure 1.

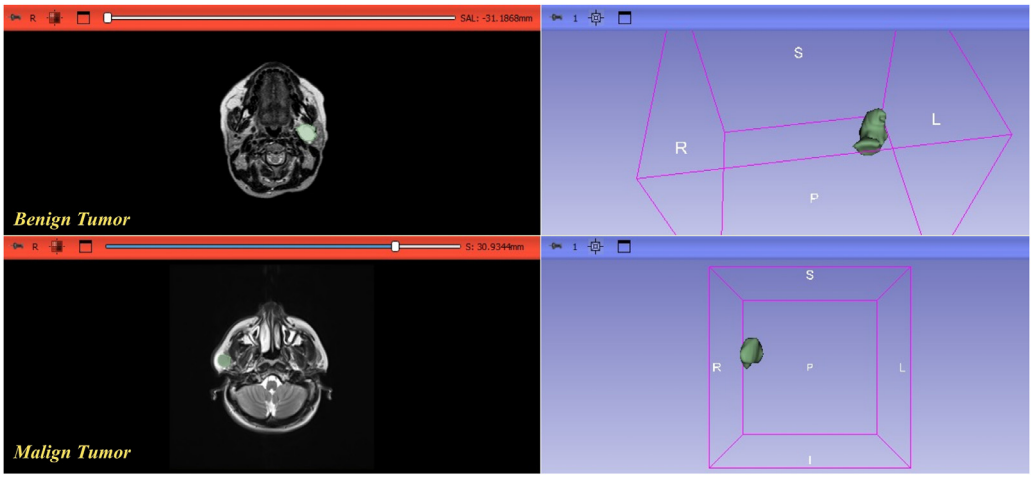


Fig. 1. Segmentation of benign and malignant parotid gland tumors on axial MR images using 3D Slicer, followed by 3-dimensional (3D) reconstructions of the segmented tumor volumes.

Radiomic feature extraction

A total of 107 radiomic features were extracted from each segmented volume using the PyRadiomics extension of the 3D Slicer software, in accordance with the Image Biomarker Standardization Initiative (IBSI) guidelines.[29] The extracted feature set consisted of 14 morphological features, 18 first-order features, 24 Gray Level Co-occurrence Matrix (GLCM) features, 14 Gray Level Dependence Matrix (GLDM) features, 16 Gray Level Run-Length Matrix (GLRLM) features, 16 Gray Level Size Zone Matrix (GLSZM) features, and 5 Neighboring Gray Tone Difference Matrix (NGTDM) features. Following feature extraction, all features were normalized using z-score standardization to ensure comparability across patients.

Feature selection/ dimensionality reduction strategies

To enable the use of radiomics features for distinguishing between benign and malignant parotid gland−related tumors, it is first necessary to meaningfully reduce the high-dimensional and complex nature of these data. This can be achieved either by selecting features that are reliable, likely to be associated with diagnostic or prognostic endpoints, and not redundantly correlated with each other, or by generating a smaller number of new components through linear or nonlinear combinations of the original features while preserving as much information as possible from the high-dimensional radiomics data. Accordingly, to evaluate which strategy yields more successful results in differentiating benign and malignant parotid tumors, four different feature selection methods: correlation-based analysis,[30] minimum redundancy maximum relevance (mRMR),[31] Boruta,[32] and recursive feature elimination (RFE)[33]; and three different dimensionality reduction methods: principal component analysis (PCA),[34] partial least squares (PLS),[35] and principal component regression (PCR)[36] were applied. In addition, machine learning algorithms were also evaluated using the original dataset consisting of 107 radiomics features, without applying any feature selection or dimensionality reduction strategy, in order to provide a baseline comparison.

- Correlation-based analysis: In the initial step, Pearson correlation coefficients were calculated among all features, and highly correlated feature pairs satisfying the threshold $|\rho| \geq t$ were identified. For each highly correlated pair, the feature with the lower variance was removed from the dataset. This procedure was repeated using correlation thresholds of 0.6, 0.7, 0.8, and 0.9, respectively.
- Minimum redundancy−maximum relevance (mRMR): mRMR is a filter-based feature selection method that aims to select radiomics features with high mutual information with the target variable (maximum relevance) while minimizing redundancy among the selected features (minimum redundancy) in distinguishing between benign and malignant parotid gland tumors.
- Boruta: Boruta is an embedded feature selection method based on the Random Forest algorithm, in which feature importance is calculated during model training. For each original feature, “shadow” features are generated through random permutations, and the importance of the real features is statistically compared with that of the shadow features. Based on this comparison, features are classified as confirmed, rejected, or tentative. Overall, Boruta follows a conservative selection strategy to identify radiomics features relevant for benign−malignant discrimination,

evaluating feature importance through the internal dynamics of the model.

- Recursive feature elimination (RFE): RFE is a wrapper-based method in which the model is initially trained using all features. Subsequently, the features contributing least to the model are iteratively removed, and this process continues until a predefined stopping criterion, or the target number of features is reached, with the aim of identifying the most informative radiomics features for distinguishing between benign and malignant tumors.
- Dimensionality reduction (PCA, PLS, PCR): These methods aim to reduce data redundancy by projecting the high-dimensional feature space into a lower-dimensional representation. Principal Component Analysis (PCA) constructs components by maximizing the total variance in the data without directly considering the target variable. Principal Component Regression (PCR) follows a similar approach by incorporating PCA-derived components into the classification process. In contrast, Partial Least Squares (PLS) seeks to maximize the covariance between the independent variables and the target variable when constructing components, thereby indirectly incorporating label information into the dimensionality reduction process.

Development of machine learning algorithms

Five different machine learning classifiers were evaluated on the selected feature subsets: Decision Tree, Support Vector Machine (SVM), Ensemble Learning, k-Nearest Neighbors (k-NN), and Neural Network (NN). To improve the performance of the algorithms, the core hyperparameters specific to each model were systematically optimized over wide ranges. During the optimization process, cross-validation−based evaluations were employed, and a balance between model performance and model complexity was taken into consideration.

The SVM model was addressed within the framework of a binary classification problem aimed at distinguishing between benign and malignant tumors. The kernel function (linear, Gaussian, quadratic, and cubic), box constraint parameter, kernel scale, and data standardization settings were optimized.

For the Decision Tree model, the maximum number of splits determining the model complexity was searched within the range of 1−94, and the Gini diversity index and maximum deviance reduction were evaluated as splitting criteria. The most balanced performance was achieved using a configuration with a maximum of 21 splits and the Gini index.

In ensemble learning models, Bagging, GentleBoost, LogitBoost, AdaBoost, XGBoost, and RUSBoost methods were evaluated. The number of learners, learning rate, maximum number of splits, and the number of sampled features were optimized. The final ensemble models were trained using the optimal hyperparameter combinations determined based on cross-validation results.

For the k-Nearest Neighbors (k-NN) model, the number of neighbors, distance metrics (City block, Euclidean, Correlation, Cosine, Mahalanobis, etc.), distance weighting strategies, and data standardization methods were evaluated.

The Neural Network model was implemented based on a fully connected multilayer perceptron architecture, and the number of layers, layer sizes, activation functions (ReLU, Tanh, Sigmoid), L2 regularization coefficient, and data standardization hyperparameters were optimized.

The hyperparameters of each classifier were determined using Bayesian optimization over 30 iterations.[37] This approach constructs a probabilistic model of the parameter space, enabling the identification of optimal hyperparameter combinations with fewer evaluations. Subsequently, all models were trained and tested using five-fold cross-validation (CV5). This strategy allowed for a statistically reliable assessment of model generalizability despite the limited sample size.

Performance metrics

To evaluate the classification performance of the developed machine learning algorithms, Accuracy, Specificity, Recall, Precision, and F1-score metrics were calculated based on the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values obtained from the classification results. In addition, Receiver Operating Characteristic (ROC) curves were generated, and the Area Under the Curve (AUC) values were calculated separately for each model as an important metric for assessing classification performance in distinguishing between benign and malignant parotid gland tumors.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - Score = \frac{2.Recall.Precision}{Recall + Precision}$$

RESULTS

When the overall performance of the machine learning models developed using all 107 radiomic features without applying dimensionality reduction was evaluated for discriminating between benign and malignant parotid gland tumors, it was observed that the NN model achieved the highest accuracy (64.2%), specificity (75.5%), precision (66.7%), and F1-score (58.5). In terms of recall, the most successful algorithm was k-NN with a value of 54.3%, whereas the highest value for the ROC−AUC metric, reflecting discriminative performance, was obtained by the SVM model at 65.9%. It was observed that the SVM model ranked second after the NN model in all performance metrics except recall, while the NN model ranked second in terms of ROC−AUC with a value of 63.3%. The correlation structure of the radiomic features obtained without dimensionality reduction is presented in Figure 2**A.** In addition, the detailed performance results of the machine learning models developed using these features are provided in Table I.

Machine learning models performance using correlation-based feature selection

The correlation structures of feature sets derived using different correlation thresholds are shown in Figure 2**B-E.** Overall evaluation of correlation-based feature selection indicated that the most favourable and balanced performance was achieved with the 17-feature representation obtained at a moderate threshold ($|\rho| \geq 0.7$). At this level, the SVM model demonstrated the most stable performance, yielding an accuracy of 66.3%, recall of 60.9%, specificity of 71.4%, F1-score of 63.8%, precision of 66.7%, and a ROC−AUC of 0.6975 (Table I**).** Applying a lower threshold ($|\rho| \geq 0.6$), which resulted in a more aggressive reduction to nine features, led to decreased performance and reduced model stability. In contrast, higher thresholds ($|\rho| \geq 0.8$ and $|\rho| \geq 0.9$), retaining 20 and 34 features respectively, did not provide meaningful performance gains despite the increased feature numbers. These findings indicate that the moderate correlation threshold ($|\rho| \geq 0.7$) provides the most suitable trade-off between feature redundancy reduction and preservation of discriminative information for benign−malignant classification of the parotid gland tumors. Across these scenarios, ROC−AUC values showed modest variations, with SVM and k-NN generally exhibiting stronger discriminative ability; however, none of the correlation-based feature sets outperformed the classification performance achieved using the full feature set (Table I**).**

Machine learning models performance using mRMR-based feature selection

Using the mRMR feature selection algorithm, 10 radiomic features were identified for distinguishing benign and malignant parotid gland tumors. As illustrated in Figure 3, the feature importance distribution demonstrates that the discriminative information is largely concentrated in the first few features, whereas the remaining features contribute only marginally to the overall classification performance. This pronounced imbalance represents a structural explanation for the limited generalization performance observed in some models when using the mRMR-based feature subset. The ROC−AUC values of the developed machine learning models generally ranged between 0.52 and 0.63, indicating that mRMR-based feature selection provided only limited discriminatory power for differentiating benign and malignant parotid tumours. In terms of performance metrics, the k-NN model achieved the most balanced results (Table I), with an accuracy of 63.2%, a recall of 52.2%, and an F1 score of 58.0, suggesting that the dominant features offered a more meaningful representation for similarity-based classification approaches. Although the SVM model achieved very high specificity (98.0%) and precision (90.9%), its recall remained low (21.7%), resulting in a reduced F1 score of 35.1; this outcome indicates a shift of the decision boundary toward the negative class, leading to a substantial proportion of positive cases being misclassified. The Tree, NN, and Ensemble models exhibited moderate performance overall.

Machine learning models performance using Boruta-based feature selection

Using the Boruta feature selection method, 12 radiomic features (five confirmed and seven tentative) were selected and subsequently used to evaluate the classification performance of models for differentiating benign and malignant parotid gland tumors. Analysis of the ROC−AUC values indicated that the models developed using the 12 Boruta-selected features achieved discriminatory performance in the range of 0.54−0.72, with the highest ROC−AUC value observed for the SVM model (0.7197). The Ensemble model demonstrated the most favourable overall performance in terms of positive class detection and classification balance, achieving an accuracy of 66.3%, a recall of 65.2%, and an F1 score of 65.2% (Table I**).** The SVM model excelled in reducing false positives, as reflected by a specificity of 75.5% and a precision of

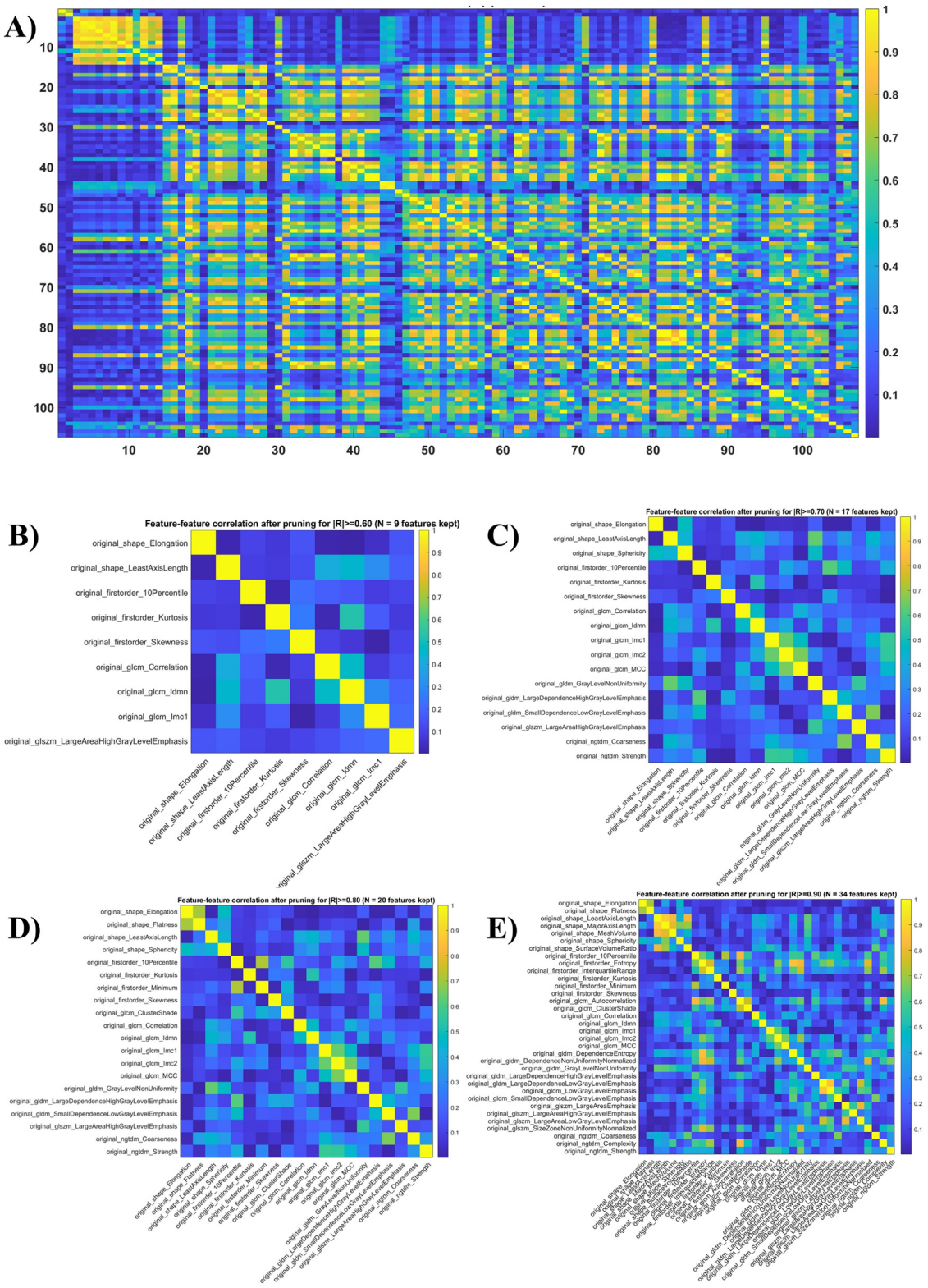


Fig. 2. Heatmaps illustrating the pairwise correlation matrices among radiomic features. (A) Pairwise correlation matrix of radiomic features in the original dataset before dimensionality reduction, showing a high degree of correlation among many features, indicating the presence of redundant information. (B) Correlation heatmap of the 9 radiomic features retained after removing highly correlated features using a correlation threshold of $|\rho| \geq 0.6$. (C) Correlation heatmap of the 17 radiomic features retained after applying a correlation threshold of $|\rho| \geq 0.7$. (D) Correlation heatmap of the 20 radiomic features retained after applying a correlation threshold of $|\rho| \geq 0.8$. (E) Correlation heatmap of the 34 radiomic features retained after applying a correlation threshold of $|\rho| \geq 0.9$.

Table I. Performance of 5 machine learning algorithms developed using all radiomic features and after applying different feature selection methods

	Model	*Accuracy (%)*	*Recall (%)*	*Specificity (%)*	*F1-Score (%)*	*Precision (%)*	*AUC*
Using all 107 features	SVM	63.2	52.2	73.5	57.7	64.9	***0.6599***
	Tree	56.8	47.8	65.3	51.8	56.4	0.5770
	Ensemble	58.9	52.2	65.3	55.1	58.5	0.5896
	k-NN	60.0	***54.3***	65.3	56.8	59.5	0.6043
	NN*	***64.2***	52.2	***75.5***	***58.5***	***66.7***	0.6331
Correlation threshold ($\|\rho\| \geq 0.6$)	SVM	53.7	4.3	***100***	8.25	***100***	0.4766
	Tree	57.9	***71.7***	44.9	***62.3***	55.0	0.5612
	Ensemble	49.5	45.7	53.1	46.7	47.7	0.4201
	k-NN*	***60.0***	28.3	89.8	40.6	72.2	***0.5984***
	NN	57.9	32.6	81.6	42.9	62.5	0.5424
Correlation threshold ($\|\rho\| \geq 0.7$)	**SVM***	***66.3***	60.9	71.4	63.8	66.7	***0.6975***
	Tree	52.6	50.0	55.1	50.5	51.1	0.5612
	Ensemble	61.1	52.2	69.4	56.6	61.5	0.5632
	k-NN	62.1	23.9	***98.0***	38.0	***91.7***	0.6599
	NN	65.3	***65.2***	65.3	***64.5***	63.8	0.6522
Correlation threshold ($\|\rho\| \geq 0.8$)	SVM	60.0	***54.3***	65.3	***56.8***	59.5	0.5964
	Tree	52.6	45.7	59.2	48.3	51.2	0.5306
	Ensemble	55.8	37.0	73.5	44.7	56.7	0.5949
	k-NN*	***61.1***	21.7	***98.0***	35.3	***90.9***	***0.6594***
	NN	57.9	43.5	71.4	50.1	58.8	0.5853
Correlation threshold ($\|\rho\| \geq 0.9$)	SVM	58.9	45.7	71.4	51.7	60.0	0.6050
	Tree	55.8	***58.7***	53.1	56.3	54.0	0.5504
	Ensemble	56.8	54.3	59.2	54.9	55.6	0.5284
	k-NN*	***64.2***	50.0	***77.6***	***57.6***	***67.6***	***0.6667***
	NN	56.8	45.7	67.3	50.6	56.8	0.5544
mRMR	SVM	61.1	21.7	***98.0***	35.1	***90.9***	0.5714
	Tree	54.7	34.8	73.5	43.0	55.2	0.5606
	Ensemble	55.8	39.1	71.4	46.1	56.2	0.6273
	k-NN*	***63.2***	***52.2***	73.5	***58.0***	64.9	***0.6340***
	NN	55.8	41.3	69.4	47.5	55.9	0.5193
Boruta	SVM	66.3	56.5	***75.5***	61.8	***68.4***	***0.7197***
	Tree	54.7	45.7	63.3	49.5	53.8	0.5370
	Ensemble*	***66.3***	***65.2***	67.3	***65.2***	65.2	0.6919
	k-NN	65.3	60.9	69.4	63.0	65.1	0.6807
	NN	62.1	47.8	***75.5***	55.0	64.7	0.6694
RFE	**SVM***	***78.9***	***78.3***	79.6	***78.3***	78.3	***0.8535***
	Tree	53.7	50.0	57.1	51.1	52.3	0.5211
	Ensemble	54.7	45.7	63.3	49.4	53.8	0.5461
	k-NN	67.4	50.0	***83.7***	59.7	74.2	0.6961
	NN	76.8	71.7	81.6	75.0	***78.6***	***0.8549***

68.4%; however, its recall remained limited at 56.5%, resulting in an F1 score of 61.8%.

Machine learning models performance using RFE-based feature selection

RFE was applied iteratively to the original 107-feature space, progressively removing the least contributive features at each step. The results indicate a clear improvement in classification performance as the number of features decreased, with performance metrics remaining relatively low during the early stages of RFE ($k \geq 72$) and increasing markedly as the feature count was reduced. The most pronounced improvement was observed in the range of 22 to 17 features, where the highest F1 score (71.0%) and ROC−AUC value (0.815) was achieved using 17 features, suggesting an optimal discriminative representation at this level. Although the highest accuracy (74.4%) was obtained with 12 features, a concurrent decline in F1 score and ROC−AUC indicated a loss of class balance. Further reduction of the feature set to seven or fewer features resulted in a substantial deterioration across all performance metrics, reflecting the loss of discriminative information due to over-elimination. Table II summarizes the performance changes associated with feature subsets obtained at each iteration of the RFE process.

Based on the combined evaluation of accuracy, F1 score, and ROC−AUC, the 17-feature subset was selected as the final RFE-based representation. Using this subset, the SVM model achieved the most balanced and highest performance, with an accuracy of 78.9%, a recall of 78.3%, an F1 score of 78.3%, and a

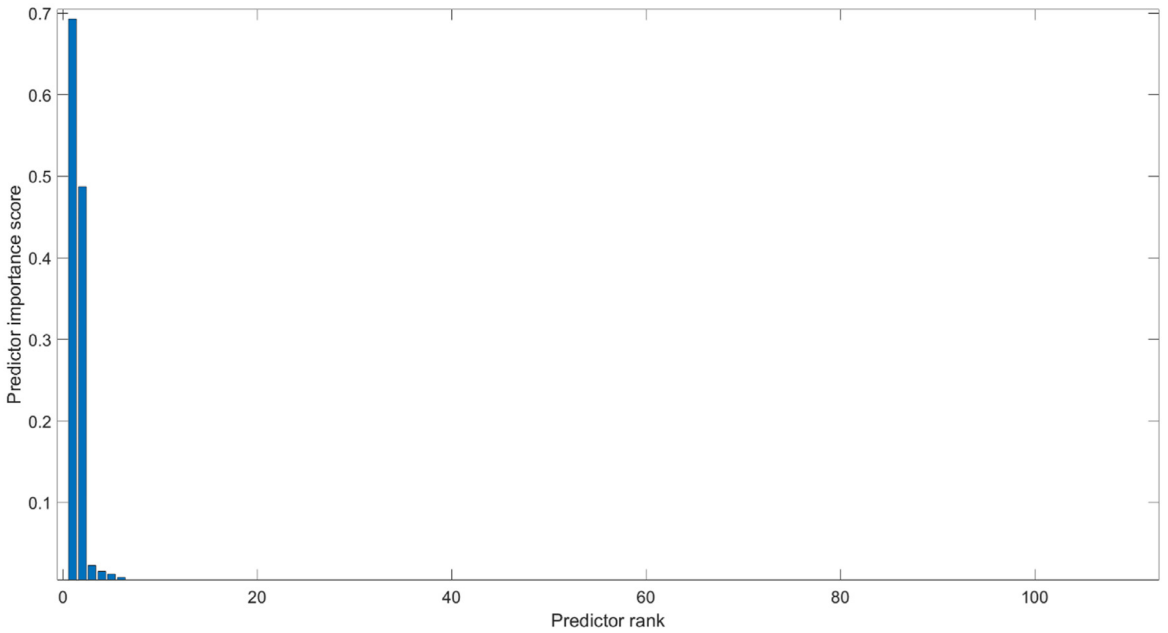


Fig. 3. Importance Ranking of Radiomic Features Based on mRMR Analysis.

Table II. Model performance achieved by iterative feature subsets selected using the Recursive Feature Elimination (RFE) method

RFE iteration	*Number of features (k)*	*Accuracy*	*F1 score*	*ROC-AUC*
1	107	0.551	0.530	0.527
2	102	0.551	0.532	0.547
3	97	0.614	0.611	0.598
4	92	0.604	0.607	0.608
5	87	0.606	0.591	0.608
6	82	0.626	0.621	0.633
7	77	0.638	0.620	0.647
8	72	0.659	0.644	0.656
9	67	0.659	0.641	0.656
10	62	0.660	0.637	0.676
11	57	0.680	0.645	0.709
12	52	0.691	0.666	0.704
13	47	0.679	0.662	0.727
14	42	0.701	0.666	0.730
15	37	0.724	0.676	0.733
16	32	0.724	0.684	0.774
17	27	0.714	0.670	0.793
18	22	0.733	0.706	0.803
19	17	0.732	**0.710***	**0.815***
20	12	**0.744***	0.708	0.811
21	7	0.660	0.568	0.766
22	2	0.586	0.422	0.697

*The RFE iterations at which the highest values for the corresponding performance metrics were achieved

ROC−AUC of 0.8535 (Table I**).** Similarly, the NN model demonstrated performance comparable to that of the SVM, achieving an accuracy of 76.8% and an F1 score of 75.0%, while standing out in terms of discriminative ability with a ROC−AUC value of 0.8549. In contrast, although the k-NN model was effective in reducing false-positive predictions, as reflected by a specificity of 83.7% and a precision of 74.2%, its recall remained limited at 50.0%, resulting in a lower F1 score of 59.7%. The Tree and Ensemble models exhibited relatively limited performance, with lower accuracy and F1 scores. Overall, these findings indicate that

Table III. Performance of 5 machine learning algorithms developed using all radiomic features and after applying different dimensionality reduction methods for radiomic features

	Model	*Accuracy (%)*	*Recall (%)*	*Specificity (%)*	*F1-Score (%)*	*Precision (%)*	*AUC*
Using all 107 features	SVM	63.2	52.2	73.5	57.7	64.9	***0.6599***
	Tree	56.8	47.8	65.3	51.8	56.4	0.5770
	Ensemble	58.9	52.2	65.3	55.1	58.5	0.5896
	k-NN	60.0	***54.3***	65.3	56.8	59.5	0.6043
	NN*	***64.2***	52.2	***75.5***	***58.5***	***66.7***	0.6331
PCA	SVM	57.9	43.5	71.4	50.0	58.8	0.5737
	Tree*	***64.2***	***71.7***	57.1	***66.0***	61.1	***0.6200***
	Ensemble	61.1	65.2	57.1	61.8	58.8	0.5885
	k-NN	58.9	37.0	79.6	46.6	63.0	0.5873
	NN	55.8	13.0	***95.9***	22.2	***75.0***	0.5760
PLS	SVM	70.2	60.0	***79.6***	66.2	***73.0***	***0.7320***
	Tree	69.1	68.9	69.4	68.2	67.4	0.6297
	Ensemble*	***71.3***	***71.1***	71.4	***70.3***	69.6	0.7313
	k-NN	***71.3***	68.9	73.5	69.7	70.5	0.7016
	NN	68.1	57.8	77.6	63.5	70.3	0.7315
PCR	SVM	61.7	37.8	***83.7***	48.6	***68.0***	***0.6349***
	Tree	57.4	57.8	57.1	56.5	55.3	0.5794
	Ensemble	61.7	60.0	63.3	60.0	60.0	0.5692
	k-NN*	***63.8***	***64.4***	63.3	***63.0***	61.7	0.6272
	NN	54.3	51.1	57.1	51.7	52.3	0.5415

the RFE method provides an effective model simplification strategy by eliminating redundant and low-contribution features from radiomic characteristics extracted from T2-weighted MR images of the parotid gland, while preserving generalization performance in the differentiation of benign and malignant tumors.

Machine learning model performance under PCA, PLS, and PCR dimensionality reduction methods

When the results of the models trained using a single PCA component were examined, it was observed that the one-component representation derived from the original 107-dimensional feature space, explaining 95% of the variance, generally limited classification performance, particularly in terms of the recall−precision balance. Among the developed models, the Decision Tree classifier achieved relatively better performance in capturing the positive class, with a recall of 71.7% and an F1 score of 66.0% (Table III), whereas the Ensemble approach produced a more balanced outcome, yielding a recall of 65.2% and an F1 score of 61.8.

For the PLS model, the estimated mean squared prediction error (MSEP) reached its minimum when 2 components were used, after which the error gradually increased with the addition of further components (Figure 4**A**). This indicates that components beyond the second did not contribute meaningful information to the model and instead increased the risk of overfitting, thereby adversely affecting generalization performance. When the results of different classifiers trained using the optimal number of PLS components were evaluated, the models exhibited generally balanced and comparable performance. The Ensemble model achieved the most favourable overall results, with an accuracy of 71.3%, a recall of 71.1%, and an F1 score of 70.3%, indicating balanced performance in both positive class detection and overall classification accuracy (Table III).

For the PCR model, the MSEP reached its minimum when 7 components were used (Figure 4**B**). Evaluation of classifiers trained with this optimal number of PCR components revealed persistent performance differences across models, suggesting that the PCR-based representation provided only limited gains for certain classifiers. Among these, the k-NN model demonstrated the best overall performance, achieving an accuracy of 63.8%, a recall of 64.4%, and an F1 score of 63.0%, reflecting a more balanced classification outcome (Table III).

DISCUSSION

In this study, we investigated the performance of radiomic features for differentiating benign and malignant parotid gland tumors using T2-weighted MRI. The primary endpoint was to evaluate the impact of different feature selection methods and dimensionality reduction strategies on classification performance. The secondary endpoint was to determine which machine learning algorithm achieved the best performance for benign/malignant discrimination.

Regarding the primary endpoint, we found that feature selection and dimensionality reduction strategies markedly influenced classification performance. The

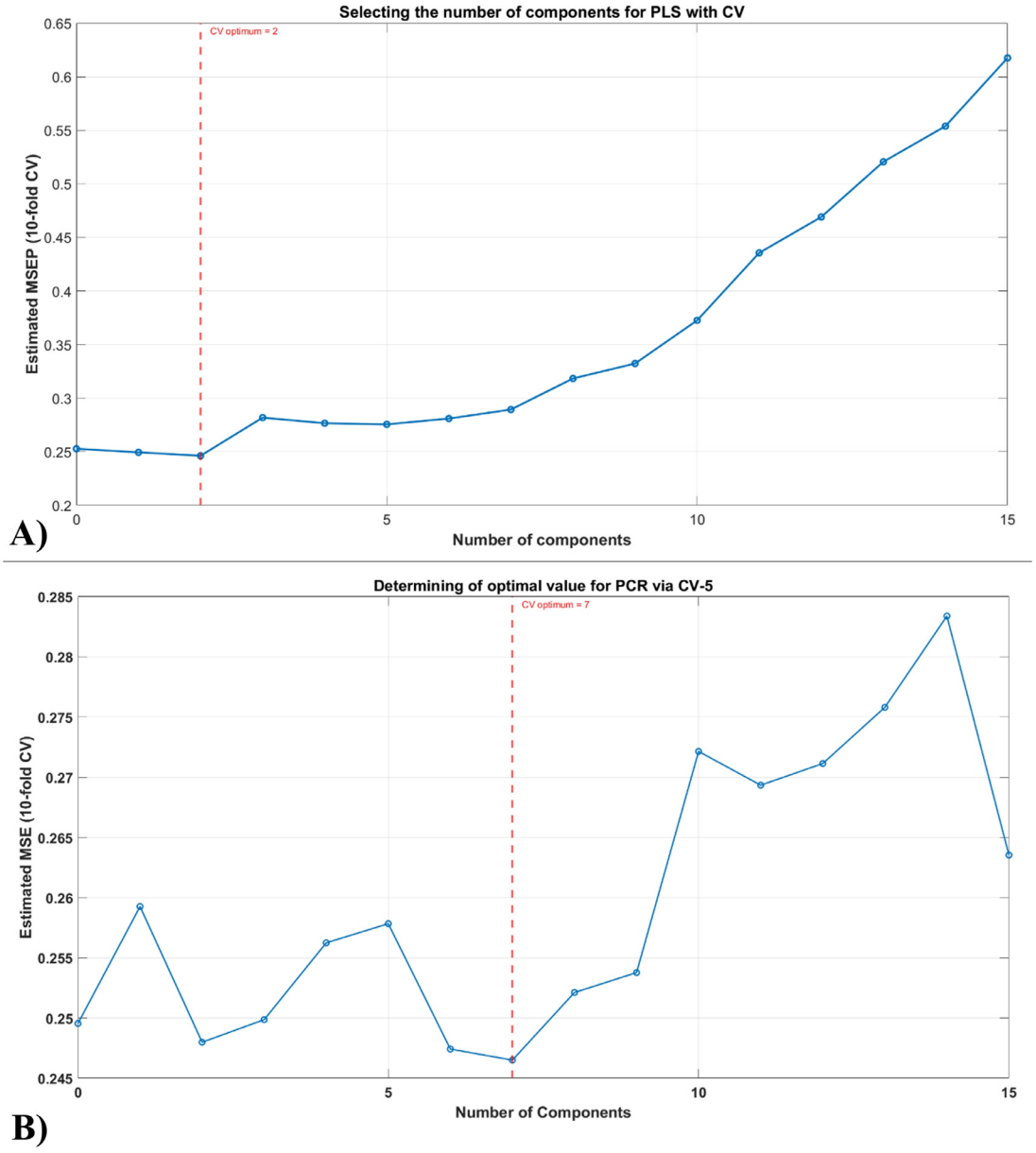


Fig. 4. Variation of the estimated mean squared error (MSEP) as a function of the number of components. (A) For the PLS model, the optimal number of components was determined as two, as indicated by the red dashed line. (B) For the PCR model, the red dashed line denotes the optimal number of components, identified as 7.

results of the best-performing models developed using the evaluated strategies are summarized in Table IV**.** Although some models achieved reasonable accuracy in the original 107-feature space without feature selection or dimensionality reduction, the comparatively lower values of metrics that better reflect class separability (e.g., F1 score and ROC-AUC) suggest that redundancy and multicollinearity in the high-dimensional feature space hindered the formation of a well-defined decision boundary. Among models trained directly on all 107 radiomic features, the best-performing model was the NN, with an AUC of 0.63. In contrast, applying dataset-appropriate feature selection or dimensionality reduction substantially improved classification performance. While AUC values for models developed using other strategies ranged from 0.62 to 0.73, the highest overall performance across all evaluated approaches was achieved by an SVM model trained on 17 radiomic features selected via RFE (AUC: 0.85). These findings underscore the importance of applying feature selection/dimensionality reduction methods tailored to the dataset and suggest that multiple strategies should be systematically tested to achieve optimal performance in radiomics studies. Moreover,

Table IV. Summary of the best-performing machine learning models under different feature selection and dimensionality reduction strategies

Method	*Number of features/ components*	*Best ML model*	*Accuracy (%)*	*Specificity (%)*	*Recall (%)*	*Precision (%)*	*F1 score (%)*	*AUC*
No dimensionality reduction	107	NN	64.2	75.5	52.2	66.7	58.5	0.6331
Correlation-based	17 ($\rho = 0.7$)	SVM	66.3	71.4	60.9	66.7	63.8	0.6975
mRMR	10	k-NN	63.2	73.5	52.2	64.9	58.0	0.6340
Boruta	12	Ensemble	66.3	67.3	65.2	65.2	65.2	0.6919
RFE*	**17**	**SVM**	**78.9**	**79.6**	**78.3**	**78.3**	**78.3**	**0.8535**
PCA	1 (%95 variance)	Tree	64.2	57.1	71.7	61.1	66.0	0.6200
PLS	2	Ensemble	71.3	71.4	71.1	69.6	70.3	0.7313
PCR	7	k-NN	63.8	63.3	64.4	61.7	63.0	0.6272

in this dataset, RFE-based feature reduction followed by SVM modeling yielded the most effective benign/malignant discrimination. The results obtained using RFE clearly demonstrate the impact of wrapper-based feature selection on this classification performance. RFE systematically eliminates features that contribute least to the decision boundary by evaluating feature contributions through the parameters of the classification model at each iteration (Table II**).** This approach enables the direct removal of both redundant features and those weakly associated with the target variable based on model performance. Particularly when combined with SVM, the feature subset generated by RFE provided a representation that supports margin-based discrimination.

The closest competing model to the SVM developed using radiomic features selected by RFE was the Ensemble model constructed with dimensionality reduction using the PLS method. The accuracy, F1 score, and AUC of this model were 71.3%, 70.3%, and 0.73, respectively. Although the best performance in the present study was achieved using RFE, a feature selection−based method, the fact that the second-best approach relied on a dimensionality reduction−based method (PLS) indicates that the application of both feature selection and dimensionality reduction strategies plays a critical role in radiomic studies, and that different methods under both strategies should be evaluated to achieve optimal performance. The PLS method was observed to perform markedly better than the other two dimensionality reduction approaches, PCA and PCR. Because PCA constructs components solely by maximizing total variance, it may prioritize directions with high variance that are weakly related to the target variable. PCR, which incorporates these PCA-derived components into the classification process, therefore retains a similar limitation. This structure reduces the representational power of directions that define interclass differences in benign−malignant discrimination and leads to only limited improvements in ROC−AUC values compared with models developed using all radiomic features. The high variance explained by PCA and PCR components does not appear to directly correspond to improved classification performance for differentiating benign and malignant parotid gland tumors. In contrast, the superior performance of the PLS method is consistent with its objective of directly maximizing the covariance between the independent variables and the target variable during component construction. The performance achieved using only two components suggests a low-dimensional representation that is more directly associated with the target variable. These findings indicate that, for the task of classifying parotid gland tumors as benign or malignant on T2-weighted MRI, PLS is capable of generating a more appropriate component structure than purely variance-based approaches.

The results of correlation-based feature selection indicate that a representation with 17 features obtained using a moderate threshold (ρ=0.7) yielded more balanced performance metrics. However, because the selection criterion of correlation-based approaches relies solely on linear relationships between features, features that are strongly associated with the target variable but also correlated with each other may be eliminated. This can result in the exclusion of part of the discriminative information from the representation, thereby limiting potential performance gains. The Boruta algorithm was observed to eliminate noisy features based on statistical importance measures and to provide a relatively balanced performance profile. Nevertheless, because Boruta's selection process operates independently of the classifier, it does not incorporate a mechanism that directly optimizes the decision boundary of the classification model. The mRMR method, while effective in reducing redundancy among features, relies on mutual information as its selection criterion. This can lead to discriminative information being concentrated in a small number of features, which in turn may result in imbalanced performance metrics in

certain models. Overall, methods that directly or indirectly incorporate target variable information into the optimization process tend to produce more consistent classification performance. Pure variance-based dimensionality reduction approaches and filter methods based solely on inter-feature relationships are effective in simplifying the data structure; however, they remain limited in preserving information that defines inter-class separation. Within this framework, model-based and iterative feature selection strategies can be considered to provide a more suitable representation in terms of the balance between performance and complexity.

MRI, despite providing excellent soft-tissue contrast, still poses a challenge in achieving reliable visual differentiation between benign and malignant lesions of the parotid gland.[10] To address this limitation, numerous studies have investigated the extraction and analysis of quantitative radiomic features from MRI-derived tumor regions. In a systematic review evaluating the diagnostic performance of radiomic approaches for distinguishing benign from malignant salivary gland tumors, reported AUC values ranged from 0.73 to 0.95.[38] Similarly, a meta-analysis demonstrated that radiomic analyses achieved high diagnostic accuracy in detecting malignant parotid gland tumors, with a sensitivity of 0.86, a specificity of 0.90, and an AUC of 0.94. In the same study, when benign lesions were analyzed separately, MRI-based models identified Warthin tumors with a sensitivity of 0.87, a specificity of 0.86, and a ROC-AUC of 0.93, while pleomorphic adenomas were detected with a sensitivity of 0.83, a specificity of 0.81, and a ROC-AUC of 0.89.[39] However, it is noteworthy that a substantial proportion of the high-performing models reported in the literature were developed using radiomic features derived from multiple MRI sequences.[21,23,40–42] In this context, the present study employed only T2-weighted images, applying feature selection–based radiomic dimensionality reduction strategies in combination with multiple machine learning algorithms. As a result, the RFE-SVM model achieved a ROC-AUC of 0.85, demonstrating a competitive diagnostic performance among single-sequence approaches. These findings suggest that meaningful clinical support for differentiating benign and malignant parotid gland tumors may be achieved without the need for advanced functional MRI techniques, such as DCE-MRI or DWI, or additional MRI sequences.

This study has several limitations that should be acknowledged. First, the retrospective study design and data collection from a single institution may limit the generalizability of the findings to centers using different MRI scanners and acquisition protocols. Second, although the RFE approach helped mitigate the risk of overfitting, the overall sample size remained relatively limited, which precluded the inclusion of an independent external validation cohort. Therefore, external validation of both the developed machine learning models and the selected radiomic features on larger, multi-center datasets is warranted. Third, the manual segmentation process, despite being performed by experienced observers, is inherently time-consuming and susceptible to inter-observer variability. Future studies may benefit from the integration of fully automated, deep learning–based segmentation approaches to improve reproducibility and clinical applicability.

CONCLUSION

RFE combined with SVM classifiers constitutes a robust framework for differentiating benign and malignant parotid gland tumors on T2-weighted MRI. In addition, the present study highlights the importance of employing multiple feature selection and dimensionality reduction strategies in radiomic-based classification tasks to enhance model stability and performance.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical Committee of Kocaeli University Non-Interventional Clinical Research (Decision date: 06/11/2025, Decision Number: KÜ GOKAEK-2025/22/22. The study was conducted in accordance with the principles of the Declaration of Helsinki.

AVAILABILITY OF DATA AND MATERIALS

The datasets generated and/or analyzed during the current study are not publicly available due to ethical restrictions imposed by the approval conditions from the ethics committee. However, they can be obtained from the corresponding author upon reasonable request.

FUNDING

The authors declare that there is no funding.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

ChatGPT (OpenAI) was employed exclusively for language editing and grammatical refinement of the manuscript. The AI tool was not used for content generation, data analysis, or any other purpose.

DECLARATIONS OF INTEREST

None.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Özgür Çakır: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology,

Investigation, Data curation, Conceptualization. **Coşku Öksüz:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis. **Simge Şahin:** Methodology, Investigation, Data curation. **Alican Kuran:** Writing – review & editing, Writing – original draft, Validation, Data curation. **Demet Doğan:** Methodology, Investigation, Formal analysis. **Büşra Yaprak Bayrak:** Visualization, Methodology, Data curation. **Fatih Mutlu:** Supervision, Methodology, Data curation. **Murat Öztürk:** Data curation, Investigation, Resources, Methodology, Visualization, Writing – review & editing. **Oğuzhan Urhan:** Writing – review & editing, Supervision, Software, Methodology, Investigation, Formal analysis.

REFERENCES

1. El-Naggar AK, Chan JKC, Grandis JR, Takata T, Slootweg PJ. WHO Classification of Head and Neck Tumours. *Int Agency Res Cancer;*. 2017:352.
2. Abdel Razek AAK, Mukherji SK. State-of-the-art imaging of salivary gland tumors. *Neuroimaging Clin N Am*. 2018;28(2):303-317.
3. Kennedy RA. WHO is in and WHO is out of the mouth, salivary glands, and jaws sections of the 4th edition of the WHO classification of head and neck tumours. *Br J Oral Maxillofac Surg*. 2018;56(2):90-95.
4. Lobo R, Hawk J, Srinivasan A. A review of salivary gland malignancies. *Neuroimaging Clin N Am*. 2018;28(2):171-182.
5. Thoeny HC. Imaging of salivary gland tumours. *Cancer Imaging*. 2007;7(1):52-62.
6. Lee YYP, Wong KT, King AD, Ahuja AT. Imaging of salivary gland tumours. *Eur J Radiol*. 2008;66(3):419-436.
7. Schmidt RL, Hall BJ, Wilson AR, Layfield LJ. A systematic review and meta-analysis of the diagnostic accuracy of fine-needle aspiration cytology for parotid gland lesions. *Am J Clin Pathol*. 2011;136(1):45-59.
8. Elsholtz FHJ, Reiter R, Marticorena Garcia SR, et al. Multifrequency magnetic resonance elastography-based tomoelastography of the parotid glands—feasibility and reference values. *Dentomaxillofacial Radiol*. 2022;51(1):20210337.
9. Karaman CZ, Tanyeri A, Özgür R, Öztürk VS. Parotid gland tumors: comparison of conventional and diffusion-weighted MRI findings with histopathological results. *Dentomaxillofacial Radiol*. 2021;50(4):20200391.
10. Christe A, Waldherr C, Hallett R, Zbaeren P, Thoeny H. MR imaging of parotid tumors: typical lesion characteristics in MR imaging improve discrimination between benign and malignant disease. *AJNR Am J Neuroradiol*. 2011;32(7):1202-1207.
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016 Feb;278(2):563-577.
12. Mei ZY, Chen J, Xu Q, et al. Development and validation of an MRI-based radiomics nomogram for distinguishing Warthin's tumour from pleomorphic adenomas of the parotid gland. *Dentomaxillofacial Radiol*. 2021;50(7):20210023.
13. Chen F, Ge Y, Li S, Liu M, Wu J, Liu Y. Enhanced CT-based texture analysis and radiomics score for differentiation of pleomorphic adenoma, basal cell adenoma, and Warthin tumor of the parotid gland. *Dentomaxillofacial Radiol*. 2023;52(2):20220009.
14. Muraoka H, Kaneda T, Kondo T, Okada S, Tokunaga S. Differential diagnosis of parotid gland tumors using apparent diffusion coefficient, texture features, and their combination. *Dentomaxillofacial Radiol*. 2023;52(4):20220404.
15. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012 Nov;30(9):1234-1248.
16. Demircioğlu A. Benchmarking feature selection methods in radiomics. *Invest Radiol*. 2022;57(7):433-443.
17. Demircioğlu A. Benchmarking feature projection methods in radiomics. *Sci Rep*. 2025;15(1):32368.
18. Bommert A, Sun X, Bischl B, Rahnenführer J, Lang M. Benchmark for filter methods for feature selection in high-dimensional classification data. Comput Stat Data Anal;143:106839.
19. Gabelloni M, Faggioni L, Attanasio S, et al. Can magnetic resonance radiomics analysis discriminate parotid gland tumors? A pilot study. *Diagnostics*. 2020;10(11):900.
20. Qi J, Gao A, Ma X, et al. Differentiation of benign from malignant parotid gland tumors using conventional MRI based on radiomics nomogram. *Front Oncol*. 2022;12:937050.
21. He Z, Mao Y, Lu S, et al. Machine learning–based radiomics for histological classification of parotid tumors using morphological MRI: a comparative study. *Eur Radiol*. 2022;32(12):8099-8110.
22. Piludu F, Marzi S, Ravanelli M, et al. MRI-based radiomics to differentiate between benign and malignant parotid tumors with external validation. *Front Oncol*. 2021;11:656918.
23. Fathi Kazerooni A, Nabil M, Alviri M, et al. Radiomic analysis of multi-parametric MR images (MRI) for classification of parotid tumors. *J Biomed Phys Eng*. 2022;12(6):599-610.
24. Faggioni L, Gabelloni M, De Vietro F, et al. Usefulness of MRI-based radiomic features for distinguishing Warthin tumor from pleomorphic adenoma: performance assessment using T2-weighted and post-contrast T1-weighted MR images. *Eur J Radiol Open*. 2022;9:100429.
25. Hu Z, Guo J, Feng J, Huang Y, Xu H, Zhou Q. Value of T2-weighted-based radiomics model in distinguishing Warthin tumor from pleomorphic adenoma of the parotid. *Eur Radiol*. 2023;33(6):4453-4463.
26. Muntean DD, Dudea SM, Băciuț M, et al. The role of an MRI-based radiomic signature in predicting malignancy of parotid gland tumors. *Cancers*. 2023;15(13):3319.
27. Wen B, Zhang Z, Zhu J, et al. Apparent diffusion coefficient map–based radiomics features for differential diagnosis of pleomorphic adenomas and warthin tumors from malignant tumors. *Front Oncol*. 2022;12:830496.
28. 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.
29. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.
30. Kowshalya AM, Madhumathi R, Gopika N. Correlation based feature selection algorithms for varying datasets of different dimensionality. *Wirel Pers Commun*. 2019;108(3):1977-1993.
31. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*. 2005;03(02):185-205.
32. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *J Stat Softw*. 2010;36:1-13.
33. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn*. 2002;46(1):389-422.
34. Principal Component Analysis and Factor Analysis. In: Jolliffe IT, ed. *Principal Component Analysis [Internet]*, New York, NY: Springer; 2002:150-166. https://doi.org/10.1007/0-387-22440-8_7.
35. Rosipal R, Krämer N. Overview and recent advances in partial least squares. In: Saunders C, Grobelnik M,

Gunn S, Shawe-Taylor J, eds. *Subspace, Latent Structure and Feature Selection*, Berlin, Heidelberg: Springer; 2006:34-51.

36. Principal components in regression analysis. In: Jolliffe IT, ed. *Principal Component Analysis*, New York, NY: Springer; 2002:167-198. https://doi.org/10.1007/0-387-22440-8_8.
37. Wu J, Chen XY, Zhang H, Xiong LD, Lei H, Deng SH. Hyperparameter optimization for machine learning models based on bayesian optimizationb. *J Electron Sci Technol*. 2019;17(1):26-40.
38. Mao K, Wong LM, Zhang R, et al. Radiomics analysis in characterization of salivary gland tumors on MRI: a systematic review. *Cancers*. 2023;15(20):4918.
39. Rao Y, Ma Y, Wang J, et al. Performance of radiomics in the differential diagnosis of parotid tumors: a systematic review. *Front Oncol*. 2024;14:1383323.
40. Shao S, Mao N, Liu W, et al. Epithelial salivary gland tumors: utility of radiomics analysis based on diffusion-weighted imaging for differentiation of benign from malignant tumors. *J X-Ray Sci Technol*. 2020;28(4):799-808.
41. Zheng YM, Li J, Liu S, et al. MRI-based radiomics nomogram for differentiation of benign and malignant lesions of the parotid gland. *Eur Radiol*. 2021;31(6):4042-4052.
42. Zhang R, Ai QYH, Wong LM, et al. Radiomics for discriminating benign and malignant salivary gland tumors; which radiomic feature categories and MRI sequences should be used? *Cancers*. 2022;14(23):5804.