**Samson DO[1,2]*, Mahayu I[1,3], Mohd Ariff MH[3], Eznal IMM[3], Hanani AM[4], Noorazrul Y[1]***

[1]Department of Diagnostic Imaging and Radiotherapy, Centre for Diagnostic, Therapeutic and Investigative Studies, Faculty of Health Sciences, National University of Malaysia, Kuala Lumpur 50300, Malaysia

[2]Department of Physics, Faculty of Sciences, University of Abuja, Abuja 900211, Nigeria

[3]Department of Radiotherapy and Oncology, Hospital Kuala Lumpur, Kuala Lumpur 50586, Malaysia

[4]Functional Image Processing Laboratory, Department of Radiology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras, Kuala Lumpur 56000, Malaysia

*Corresponding author
Noorazrul Yahya
azrulyahaya@ukm.edu.my
Samson Damilola Oluwafemi
damilola.samson@uniabuja.edu.ng

# Assessing the impact of predictive models on forecasting radiation-induced late xerostomia in head-and-neck cancer patients using CBCT-based delta radiomics features

***Abstract –*** Accurately predicting xerostomia risk remains a challenge for personalized radiotherapy, as conventional dose-volume techniques often underperform in prediction and lack clinical interpretability. This study investigates the impact of different machine learning algorithms (MLAs) on the prediction accuracy of late xerostomia treated with VMAT-CBCT-based delta-radiomics features. A retrospective analysis was performed on 131 HNC patients receiving VMAT-based chemoradiotherapy (60–70 Gy, 30–35 fractions) between 2020 and 2023. Radiomic features were obtained from longitudinal CBCT images following the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) and Image Biomarker Standardization Initiative (IBSI) guidelines, with images resampled to a common resolution and computed delta radiomics. The dataset was split into training (70%) and testing (30%) sets using stratified bootstrapping. Feature selection was performed using recursive feature elimination with logistic regression, followed by Synthetic Minority Oversampling Technique (SMOTE). MLAs predicting 12-month xerostomia (CTCAE v5.0, grade ≥ 2) were developed in Python (v7.0.8) with cross-validation. Models incorporating delta radiomics outperformed baseline clinical models, achieving an AUC of 0.75 (SVM), 0.70 (RF), and 0.68 (LASSO). The SVM demonstrated the best classification metrics, with recall (0.91), accuracy (0.89), sensitivity (0.84), specificity (0.63), and Brier score (0.19). The most influential predictors were temporal changes in specific radiomic features of the parotid glands, along with patient comorbidities. Delta-radiomic features significantly improved the prediction of moderate-to-severe xerostomia at 12 months post-radiotherapy. These findings support the integration of delta radiomics with clinical data for personalized treatment planning in HNC.

***Keywords – delta radiomics, predictive performance, machine learning algorithms, xerostomia, HNCRT, CBCT-VMAT***

## 1 INTRODUCTION

Despite significant advancements in conformal radiotherapy (RT) techniques such as volumetric modulated arc therapy (VMAT), which are designed to minimize radiation exposure to critical structures like the salivary glands, a substantial number of head and neck cancer (HNC) patients still experience late xerostomia (dry mouth). This condition remains one of the most common long-term side effects of RT, negatively impacting patients' quality of life. Accurate prediction of xerostomia risk continues to be a major challenge in personalizing RT planning and delivery. Traditional dose-volume histogram–based models, while widely used, often fall short in terms of predictive performance and clinical interpretability [1–7].

Emerging evidence suggests that delta radiomics, which involves the extraction of quantitative imaging biomarkers reflecting temporal changes during treatment, holds significant potential for improving the prediction of therapeutic outcomes. Studies of baseline CT/MR radiomics of the parotid and submandibular glands have shown prognostic significance for 3-month xerostomia, according to Sheikh et al. [8]. In the meantime, studies using weekly CT images of the contralateral parotid gland that incorporate delta/radiomics features have started to indicate promise for later xerostomia endpoints (12 months), according to van Dijk et al. [1]. Additionally, Abdollahi et al. [9] showed increased performance using pretreatment, mid-treatment, and delta-feature sets in patients with head and

neck cancer by combining radiomics and dosiomics techniques. Although cone-beam computed tomography (CBCT) is routinely acquired during treatment for image guidance, its potential as a source of delta radiomic biomarkers remains underutilized [2–3]. Integrating CBCT-derived delta radiomics with clinical and dosimetric variables in machine learning frameworks offers a novel strategy to enhance the prediction of late xerostomia [6–12]. These models could enable the early identification of patients at elevated risk, paving the way for adaptive RT interventions and personalized supportive care.

This study aims to evaluate predictive models for late xerostomia using CBCT-derived delta radiomic features with machine learning techniques, with the ultimate goal of supporting personalized radiotherapy and improving patient outcomes. The predictive performance of the models (support vector machine, random forests, and LASSO – least absolute shrinkage and selection operator) was compared with the baseline model using the mean area under the receiver operating characteristic curve (ROC-AUC) with 95% confidence intervals, as well as recall, sensitivity, specificity, accuracy, and Brier score.

## 2 MATERIALS & METHODS

### 2.1 Patient and treatment characteristics

A total of 173 patients were initially assessed from archived data of HNC patients treated with VMAT-RT between 2020 and 2023 at the Department of Radiotherapy and Oncology, Kuala Lumpur General Hospital, Malaysia. Eligible patients (n = 131) with histologically confirmed HNC, subsequent radiographically-proven CTCAE v5.0 grades ≥ 2 xerostomia (12 months endpoint), availability of planning CT, and treatment imaging were retrospectively analyzed. Patients with unidentified primary tumors, single-fractionation, inadequate imaging data, or previous salivary gland surgery were excluded (n = 42). The research was carried out in compliance with the Image Biomarker Standardization Initiative (IBSI) and TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) criteria [13–14].

All patients were treated on the RapidArc RT system (Eclipse, Varian Medical Systems, USA), with on-board kilovoltage imaging. Prescribed dose ranged from 60–70 Gy, in 30–35 fractions. For every three weeks, induction chemotherapy of cisplatin (100 mg/m$^2$) and carboplatin (400 mg/m$^2$) was administered, this was followed by weekly concurrent treatment.

Predictor parameters included: demographic (age: 23–87 years; gender: male (101), female (30)), clinical (tumor site: nasopharyngeal (93), tongue (9), tonsillar (7), glottis (6), maxillary sinus (5), nasal cavity (4), oropharyngeal (3), alveolar ridge (1), buccal mucosa (1), hard palate (1), laryngeal (1)); smoking (yes = 41, no = 90), alcohol (yes = 29, no = 102), comorbidity (yes = 77, no = 54)), and treatment-related factors (CHRT type, dose, BMI-body mass index). The National University of Malaysia Research Ethics Committee granted ethical approval (UKM PPI/111/8/JEP-2023-817).

### 2.2 Imaging, CT simulation, and Segmentation

Planning CTs and CBCT images acquired during RT were registered to a common reference frame (Digital Imaging and Communications in Medicine (DICOM)). All CBCT scans were attained with a Canon LB Acquilion 32-slice CT simulator (Canon Medical Systems Corporation, Japan). Scan parameters were as follows: slice thickness reconstruction of 3mm, X-ray tube current of 300mA at 120 kVp with 512 x 512 pixels, and followed by a 90s delay after intravenous contrast administration. A standard quality assurance was performed according to the guidelines of the American Association of Physicists in Medicine (AAPM) (TG-142 and TG-179) [15] using a quality assurance phantom (CatPhan). All CT scans were spatially resampled (1 x 1 x 1 mm$^3$), intensity normalized (–1000 to 3000 HU), and discretized in order to standardize the image voxel sizes for use in texture feature computations. Oncologists used the treatment planning system (Eclipse, Varian Medical Systems, USA) to contour target volumes and organs at risk, such as the left and right parotid gland, on the pCT scans. During treatment planning, the dosimetric data were used to derive the mean dose ($D_{mean}$) of the left and right parotid glands, which were then independently limited.

### 2.3 Feature extraction, selection, and calculation of delta radiomics

For each patient, radiomic features were extracted using LIFEx v7.8.0 software (https://www.lifexsoft.org/) from the segmented left and right parotid glands that developed xerostomia of baseline CBCT (day 1) and CBCT (week 4). This was done with the 3D processing selection. Delta radiomics was calculated as the absolute

differences by subtracting the radiomic features of CBCT (week 4) from those of CBCT (day 1).

The dataset was randomly split into training (70%) and testing (30%) sets using stratified bootstrapping. The features were selected, and dimensionality was decreased using recursive feature elimination with logistic regression. Class imbalance was resolved in all models using the Synthetic Minority Oversampling Technique (SMOTE). Five-fold cross-validation with GridSearchCV and randomized search (≤ 1000 iterations) was used to optimize the hyperparameters. Python v7.0.8 was used to implement the models using scikit-learn.

### 2.4 Assessment of model predictive power

Model predictive power was assessed using standard metrics for classification accuracy and discriminative ability, including mean area under the receiver operating characteristic curve (ROC-AUC) with 95% confidence intervals (CI), recall, sensitivity, specificity, accuracy, and Brier score.

## 3 RESULTS

Accurate prediction of xerostomia following radiotherapy remains critical for individualizing treatment planning and minimizing long-term toxicity. In this study, we compared the predictive performance of a baseline model with three data-driven approaches (support vector machine, random forests, and LASSO), at clinically relevant time points (12 months post-treatment), using the AUC, recall, accuracy, sensitivity, specificity, and Brier score as an indicator of overall probabilistic accuracy. Table 1 displays the most relevant predictive features found during modeling, along with their significance coefficients.

Models incorporating delta radiomics outperformed baseline clinical models (Fig. 1), with the support vector machine model having the optimal AUC of 0.75. The support vector machine also demonstrated the strongest classification metrics for predicting moderate-to-severe xerostomia at 12 months, with recall (0.91), accuracy (0.89), sensitivity (0.84), and specificity (0.63). It also consistently displayed the lowest Brier scores (0.19), indicating superior calibration and discrimination compared to both random forests and LASSO. Remarkably, random forests performed similarly in terms of AUC (0.70), recall (0.85), accuracy (0.70), sensitivity (0.79), specificity (0.55), and Brier score (0.23). Despite having a high Brier score (0.28) and a lower specificity (0.41), LASSO produced comparable predictive power. The three models outperformed the baseline model (AUC: 0.34–0.49, recall: 0.43–0.62, accuracy: 0.41–0.55, sensitivity: 0.44–0.53, specificity: 0.23–0.40, Brier score: 0.30–0.43).

The most influential predictors were temporal changes in specific radiomic features of the parotid glands, along with patient comorbidities. While RT-related parameters contributed to model performance, their predictive value was comparatively lower than imaging and clinical variables.

## 4 DISCUSSION

As detailed in Table 1, the integration of demographic data with first-order radiomic features significantly improved predictive accuracy. The $D_{mean}$ LT_PG and $D_{mean}$ RT_PG, in particular, showed the highest feature coefficients (1.578, 1.741), underscoring the established dose-response relationship between radiation exposure and salivary gland toxicity HNC and confirming their significance as important indicators of late-onset xerostomia. Patient-specific variables such as alcohol consumption (coefficient: 1.299) and comorbidities (coefficient: 0.961) also emerged as significant predictors, highlighting the impact of individual health and lifestyle factors on the risk of late toxicity. Higher-order radiomic features, including intensity histogram kurtosis (IBSI: C3I7; coefficient: 1.515), intensity kurtosis (IBSI: IPH6; coefficient: 1.346), and the 90$^{th}$ percentile of the intensity histogram (IBSI: 8DWT; coefficient: 1.071), were also strongly predictive. Elevated kurtosis values, in particular, suggest uniformity in tissue composition, which may be indicative of radiation-induced glandular injury [6–7,16]. Sum entropy (IBSI: P6QZ; coefficient: 1.403), inverse variance (IBSI: E8JP; coefficient: 1.029), maximum histogram gradient (IBSI: 12CE; coefficient: 0.817), and minimum histogram gradient grey level (IBSI: RHQZ; coefficient: 0.615) were additional important features. These imaging biomarkers have the potential to be useful in identifying and forecasting late-stage xerostomia because they provide insightful information on tissue heterogeneity and structural changes [7–12].

Of all the models, the support vector machine model demonstrated the best predictive power. Random forests showed modest improvement over the support vector machine model. The LASSO model, on the other hand, consistently underperformed, suggesting that linear regularized regression has limited use in this

particular clinical setting. These results imply that the support vector machine may offer incremental gains in predictive accuracy, particularly for long-term outcomes. The relative underperformance of LASSO may reflect its limited capacity to capture complex, nonlinear interactions between dose-volume and clinical variables that are better handled by tree-based ensemble methods.

These findings align with recent studies highlighting the relevance of delta radiomics in xerostomia prediction. van Dijk et al. [1] showed that mid-treatment changes in parotid gland surface area significantly improved models predicting moderate-to-severe xerostomia at 12 months compared to pre-treatment models (AUC: 0.69–0.78), while Khajetash et al. [2] validated the effectiveness of using multiple classifiers with CT and MRI radiomics to enhance predictions of both early and late symptoms of xerostomia and sticky saliva, compared to single-model approaches. Additionally, Berger et al. [12] evaluated day-to-day kinetics of radiomics features using megavoltage CT (MVCT) and found that radiomics features extracted during the first half of RT improved prediction of moderate-to-severe xerostomia at 6, 12, and 24 months, outperforming dose-only models (AUC: 0.69–0.86) when early treatment imaging features were included. Our results suggest that integrating delta radiomics with clinical data enhances the ability to predict late xerostomia, potentially guiding more personalized treatment adjustments.

This study is limited by its retrospective design, single-institution dataset, and relatively small sample size. External validation with larger, multi-center cohorts is necessary to confirm the generalizability of these findings.

## 5 CONCLUSIONS

Delta radiomics-based models markedly improved the prediction of moderate-to-severe xerostomia at 12 months compared with baseline clinical models. The support vector machine achieved the best overall performance, with superior discrimination (AUC = 0.75) and calibration (Brier score = 0.19), followed by random forests and LASSO regression. Temporal alterations in parotid gland radiomic features and patient comorbidities were the most influential predictors, while radiotherapy dose parameters contributed less substantially. These findings highlight the prognostic significance of incorporating longitudinal imaging biomarkers into predictive models, assisting in the development of personalized and adaptive radiation approaches to reduce treatment-related xerostomia.

## FUNDING

This research was supported by the Premier Impact Fund 2.0 (DIP 2.0) [Grant number: DIP-2024-032].

## ACKNOWLEDGEMENT

We acknowledge the assistance of the Department of Radiotherapy and Oncology, Kuala Lumpur General Hospital, Malaysia, for the extraction of treatment and patient data.

## REFERENCES

[1] Van Dijk LV, Brouwer CL, van der Laan HP, Burgerhof JGM, Langendijk JA, Steenbakkers RJHM, Sijtsema NM. Geometric image biomarker changes of the parotid gland are associated with late xerostomia. Int. J. Radiat. Oncol. Biol. Phys. 2017;99(5):1101–1110, https://doi.org/10.1016/j.ijrobp.2017.08.003.

[2] Khajetash B, Hajianfar G, Talebi A, Ghavidel B, Mahdavi SR, Lei Y, Tavakoli M. A comparison of different machine learning classifiers in predicting xerostomia and sticky saliva due to head and neck radiotherapy using a multi-objective, multimodal radiomics model. Biomed. Phys. Eng. Express, 2025;11(025027), https://doi.org/10.1088/2057-1976/adafac.

[3] Ismail M, Hanifa MAM, Mahidin EIM, Manan HA, Yahya N. Cone beam computed tomography (CBCT) and megavoltage computed tomography (MVCT)-based radiomics in head and neck cancers: a systematic review and radiomics quality score assessment. Quant. Imaging Med. Surg. 2024;14(9):6963–6977, https://dx.doi.org/10.21037/qims-24-334.

[4] Deasy JO, Moiseenko V, Marks L, Chao KSC, Nam J, Eisbruch A. Radiotherapy dose-volume effects on salivary gland function. Int J Radiat Oncol Biol Phys 2010;76(3 Suppl):S58–63. https://doi.org/10.1016/j.ijrobp.2009.06.090.

[5] Sim CPC, Soong YL, Pang EPP, Lim C, Walker GD, Manton DJ, Reynolds EC, Wee JTS. Xerostomia, salivary characteristics and gland volumes following intensity-modulated radiotherapy for nasopharyngeal carcinoma: a two-year follow up. Aust Dent J. 2018;63(2):217–223. https://doi.org/10.1111/adj.12608.

[6] Yahya N, Manan HA. Quality of life and patient-reported outcomes following proton therapy for oropharyngeal carcinoma: A systematic review. Cancers 2023;15(8):2252. https://doi.org/10.3390/cancers15082252.

[7] Tan D, Salleh M, Manan HA, Yahya N. Delta-radiomics-based models for toxicity prediction in radiotherapy: A systematic review and meta-analysis. J. Med. Imaging Radiat Oncol 2023;67(5):564–579. https://doi.org/10.1111/1754-9485.13546.

[8] Sheikh K, Lee SH, Cheng Z, Lakshminarayanan P, Peng L, Han P, McNutt TR, Quon H, Lee J. Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands. Radiat Oncol 2019;14(1):131. https://doi.org/10.1186/s13014-019-1339-4.

[9] Abdollahi H, Dehesh T, Abdalvand N, Rahmim A. Radiomics and dosiomics-based prediction of radiotherapy-induced xerostomia in head and neck cancer patients. Int J Radiat Biol 2023;99(11):1669–1683. https://doi.org/10.1080/09553002.2023.2214206.

[10] Ghazali SNA, Manan HA, Chan CMH, Eezamuddeen M, Yahya N. Predictors of survival among head and neck cancer patients after radiotherapy based on health-related quality of life indices and symptoms-specific outcomes: a systematic review. Qual Life Res 2025. https://doi.org/10.1007/s11136-025-03933-y.

[11] Wang H, Fan M, Yan L, Du X, Li L, Lai X, Yin J, Orlandini LC, Ren J, Yin Q, Pei J, Ren Y, Lang J, Zhou Q, Feng M. Early prediction of parotid glands secretory function based on ADC variations during radiotherapy for nasopharyngeal carcinoma: a phase II prospective study. Radiat Oncol. 2025;20(117):1–10. https://doi.org/10.1186/s13014-025-02696-7.

[12] Berger T, Noble DJ, Shelley LEA, McMullan T, Bates A, Thomas S, Carruthers LJ, Beckett G, Duffton A, Paterson C, Jena R, McLaren DB, Burnet NG, Nailon WH. Predicting radiotherapy-induced xerostomia in head and neck cancer patients using day-to-day kinetics of radiomics features. Phys Imaging Radiat Oncol 2022;24:95–101. https://doi.org/10.1016/j.phro.2022.10.004.

[13] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circulat 2015;131(2):211–219. https://doi.org/10.1161/CIRCULATIONAHA.114.014508.

[14] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, Bogowicz M, Boldrini L, Buvat I, Cook GJR, Davatzikos C, Depeursinge A, Desseroit M-C, Dinapoli N, Dinh CV, Echegaray S, Naqa IE, Fedorov AY, Gatta R, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiol 2020;295(2):328–338. https://doi.org/10.1148/radiol.2020191145.

[15] Klein EE, Hanley J, Bayouth J, Yin F-F, Simon W, Dresser S, Serago C, Aguirre F, Ma L, Arjomandy B, Liu C, Sandin C, Holmes T. Task Group 142 report: Quality assurance of medical accelerators. Med Phys 2009;36(9):4197–4212. https://doi.org/10.1118/1.3190392.

[16] Nardone V, Tini P, Nioche C, Mazzei MA, Carfagno T, Battaglia G, Pastina P, Grassi R, Sebaste L, Pirtoli L. Texture analysis as a predictor of radiation-induced xerostomia in head and neck patients undergoing IMRT, Radiol. Med., 123(6) (2018) 415– 423, https://doi.org/10.1007/s11547-017-0850-7.

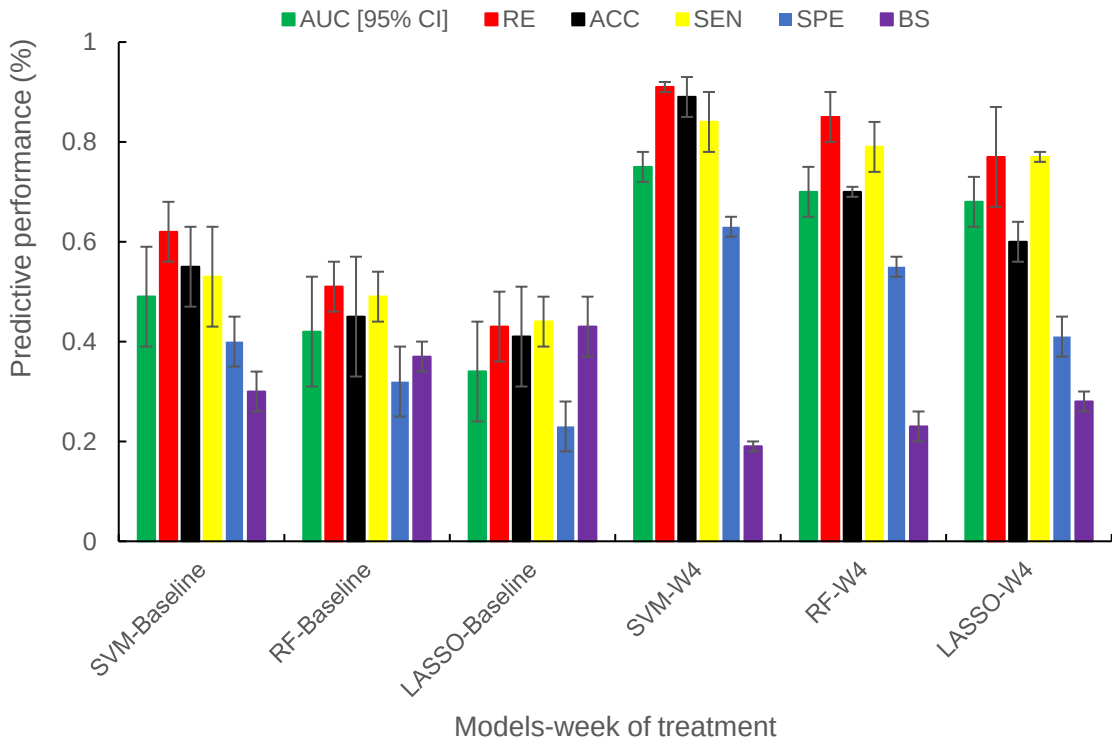


**Figure 1.** Performance parameters for each model.
Note: AUC: mean area under the receiver operating characteristic curve, CI: confidence intervals, RE: recall, ACC: accuracy, SEN: sensitivity, SPE: specificity, BS: Brier score, and W4: week 4.

**Table 1.** Significant predictor features and importance coefficients at 12-month endpoint.

| Features name | Category | Feature coefficients |
|---|---|---|
| Comorbid | Demographic | 0.961 |
| Alcohol habits | Demographic | 1.299 |
| $D_{mean}$ LT_PG | Sum dose | 1.338 |
| $D_{mean}$ RT_PG | Sum dose | 1.741 |
| Intensity histogram kurtosis (IBSI: C317) | First order | 1.575 |
| Intensity histogram 90$^{th}$ percentile (IBSI: 8DWT) | First order | 1.071 |
| Intensity kurtosis (IBSI: IPH6) | First order | 1.346 |
| Intensity quartile coefficient of dispersion (IBSI: 9S40) | First order | 0.791 |
| Root mean square intensity (IBSI: 5ZWQ) | First order | 0.557 |
| Maximum histogram gradient (IBSI: 12CE) | First order | 0.817 |
| Minimum histogram gradient grey level (IBSI: RHQZ) | First order | 0.615 |
| Compactness (IBSI: SKGS) | NGTDM | 0.957 |
| Complexity (IBSI: HDEZ) | NGTDM | 0.837 |
| Large zone high grey level emphasis (IBSI: J17V) | GLSZM | 0.785 |
| Zone size variance (IBSI:3NSA) | GLSZM | 0.799 |
| Zone size entropy (IBSI: GU8N) | GLSZM | 0.644 |
| Sum entropy (IBSI: P6QZ) | GLCM | 1.403 |
| Inverse variance (IBSI: E8JP) | GLCM | 1.029 |