


Article

Deep Texture Analysis Enhanced MRI Radiomics for Predicting Head and Neck Cancer Treatment Outcomes with Machine Learning Classifiers

Aryan Safakish ^{1,2}, Amir Moslemi ², Daniel Moore-Palhares ^{2,3,4} , Lakshmanan Sannachi ^{2,3,4}, Ian Poon ^{3,4}, Irene Karam ^{3,4}, Andrew Bayley ^{3,4}, Ana Pejovic-Milic ¹ and Gregory J. Czarnota ^{3,4,5,*}

¹ Department of Physics, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; asafakis@torontomu.ca (A.S.)

² Physical Sciences, Sunnybrook Research Institute, Toronto, ON M4N 3M5, Canada

³ Department of Radiation Oncology, Sunnybrook Health Sciences Center, Toronto, ON M4N 3M5, Canada

⁴ Department of Radiation Oncology, University of Toronto, Toronto, ON M5T 1P5, Canada

⁵ Department of Medical Biophysics, University of Toronto, Toronto, ON M5T 1P5, Canada

* Correspondence: gregory.czarnota@sunnybrook.ca

Simple Summary: Treating head and neck cancers is challenging due to the complex anatomy and often late-stage diagnosis. Early symptoms can be misattributed to those associated with common ailments like the flu, and unsurprisingly, many patients already have lymph node (LN) involvement by diagnosis. Treatment options include radiation and chemotherapy; nevertheless, there remains a subset of patients for whom these treatments fail to adequately reduce tumor size. In this investigation, phenotypic insights regarding treatment outcomes from pathologically involved LNs were explored and “mined” as radiomics features. Said features were used to train machine learning (ML) classifiers, enabling the identification of informative feature combinations and the prediction of treatment outcomes. Moreover, iterative *deep texture analysis* (DTA) was explored to evaluate deeper-layer radiomics features and proved useful for further enhancement of predictive models. Robust generalized models could enhance personalized cancer care, and the preliminary results in this work support further research into radiomics-trained predictive models and DTA to boost performance.



Citation: Safakish, A.; Moslemi, A.; Moore-Palhares, D.; Sannachi, L.; Poon, I.; Karam, I.; Bayley, A.; Pejovic-Milic, A.; Czarnota, G.J. Deep Texture Analysis Enhanced MRI Radiomics for Predicting Head and Neck Cancer Treatment Outcomes with Machine Learning Classifiers. *Radiation* **2024**, *4*, 192–212. <https://doi.org/10.3390/radiation4020015>

Academic Editor: Sam Payabvash

Received: 6 May 2024

Revised: 7 June 2024

Accepted: 12 June 2024

Published: 14 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Head and neck cancer treatment does not yield desired outcomes for all patients. This investigation aimed to explore the feasibility of predicting treatment outcomes from routine pre-treatment magnetic resonance images (MRIs). Radiomics features were “mined” and used to train machine learning (ML) classifiers to predict treatment outcomes. Moreover, iterative *deep texture analysis* (DTA) was explored to boost model performances. **Methods:** Radiomics features were determined from T1-weighted post-contrast MRIs of pathologically involved lymph node (LN) segmentations for $n = 63$ patients. SVM, k -NN, and FLD classifier models were trained, selecting for 1–10 features. The model with the top balanced accuracy was chosen for an iteration of DTA. New feature sets were used to retrain and test the ML. Radiomics features were explored for a total of three layers through two iterations of DTA. **Results:** Models proved useful in predicting treatment outcomes. The best model was a nine-feature multivariable k -NN model with a sensitivity ($\%S_n$) of 93%, specificity ($\%S_p$) of 74%, 86% accuracy ($\%Acc$), and 86% precision ($\%Per$). The best model for two of the three classifiers (k -NN and FLD) was trained using features from three layers. The performance of the average k -NN and FLD models trained with features was boosted significantly with the inclusion of deeper-layer features. **Conclusions:** Pre-treatment LN MRIs contain quantifiable texture information that can be used to train ML models to predict cancer treatment outcomes. Furthermore, DTA proved useful to boosting predictive models.

Keywords: radiomics; deep texture analysis; cancer treatment; response prediction; machine learning prediction; texture features; head and neck cancer; texture of texture

1. Introduction

Pattern recognition is among of a plethora of intricate cognitive capabilities that has evolved over thousands of years, by encoding and integrating information acquired from environmental sensory inputs, to guide behavioral responses [1]. Some pattern processing functions—like the ability to create cognitive maps of physical environments and to distinguish individuals of the same species and the use of gestures to communicate with other individuals—are common among many animal species, whereas others—like creativity and invention, spoken and written language, imagination and comprehension of the passing of time—are uniquely human, in large part due to relatively large size of the cerebral cortex [1]. Specifically for visual information, photoreceptors in the retina activate in response to incoming light, and details are transiently encoded in nerve cell circuits of the visual regions of the cerebral cortex, which are then transferred and stored in the hippocampus [1]. As a result of this evolutionary development, humans have an innate ability to recognize and empirically describe textural properties (coarseness, rippling, contrast, etc.) associated with visual inputs. In 1973, with advancements in computing and digital photography, Haralick et al. foreshadowed the need and potential utility in quantifying textural characteristics of an image and pioneered the field of radiomics by publishing a seminal work defining a process to calculate a set of features from an image [2]. Based on the assumption that the textural information of images is contained in the overall or “average” relationship of pixel intensities within the image, Haralick et al. computed a spatially dependent probability–distribution matrix based on grayscale intensities of neighboring pixels, in what is called a gray-level co-occurrence matrix (GLCM), and defined features based on said matrix [2]. From that work, other matrices have been developed to quantify relationships between image pixels, including the gray-level run length matrix (GLRLM) [3], the gray-level size zone matrix (GLSZM) [4], and the gray-level dependence matrix (GLDM) [5], all with distinctly defined features. Features can be calculated for entire images or for regions of interest (ROI) within an image.

Modern personalized medicine is emerging in parallel with advancements in biomedical imaging, unlocking visual insights beyond what is accessible to the naked eye. Observation of internal anatomy with medical imaging allows physicians to better understand the nature of diseases, plan treatment, and monitor treatment efficacy. Current cancer care standard protocols involve the acquisition of a biopsy sample from the suspected diseased tissue for up-front genetic profiling and confirmation of cancer. However, three major hurdles limit potential, namely: (i) acquiring a biopsy sample is an invasive procedure for patients, and (ii) samples can ultimately fall short in adequately characterizing the entirety of a suspicious mass, which is crucial, considering genetic, anatomic, and physiological heterogeneity drastically influences cancer growth and treatment outcomes [6,7]; and (iii) clinically validated biological biomarkers have yet to be established for most cancer types. Globally, considerable resources are directed towards researching the role tumor composition and microenvironments play with respect to treatment outcomes, disease progression, metastasis, recurrence, and more [6,8,9].

In contrast to biopsies, radiomic analysis of suspect masses presents a few noteworthy potential advantages; mainly, (i) it presents a potential method of characterizing the entirety of a region in question, (ii) it is non-invasive, (iii) it allows clinicians to maximize the utility of often previously acquired images (computed tomography (CT) and magnetic resonance imaging (MRI), in particular) required for qualitative diagnostic and treatment planning purposes, (iv) it can be applied to multiple masses within a single patient, and (v) it can be used at different time points, providing longitudinal data such as treatment response over time [10]. Theoretically, radiomics features could reveal promising imaging biomarkers associated with numerous biological and clinical conditions of interest. The identification and combining of meaningful features to distinguish between populations of interest (for example, recurrence, metastasis, treatment efficacy, etc.) is a task that can be carried out by common machine learning (ML) classifiers after labeling the “mined” textural features with known outcomes and training predictive models [11].

Epithelial malignancies originating in the oral cavity, pharynx, larynx, paranasal sinuses, nasal cavity, and salivary glands are broadly categorized as head and neck (H&N) cancers [12], approximately 90% of which are squamous cell carcinomas (SCC) [13]. A 25-year-long analysis of cancers in Canada, published in 2022, revealed that the vast majority (~35,000 or 70%) of roughly 48,000 incidences of H&N cancers occurred in male subjects [14]. Tobacco and alcohol consumption [15–17], human papilloma virus (HPV) infection [18], and p53 and p16 gene mutations [19,20] are risk factors that make H&N cancers the seventh most common type of cancer [18]. The World Health Organization's International Agency for Research on Cancer estimated globally—in the year 2020—there would be in excess of 900,000 newly diagnosed H&N cancer patients, and just under 500,000 individuals would succumb to complications of H&N cancer [14]. Though distant metastasis is rare at the time of diagnosis (~10%), the majority of HPV-related cancer patients exhibit regional spread of cancerous cells to lymph nodes (LN) by the time of diagnosis [12]. Five-year mortality rates are approximately 50% but vary depending on clinical factors such as the primary site of disease, tumor stage, and HPV status, as well as non-clinical factors like geographical and socioeconomic status influencing access to healthcare [21,22]. It is worth noting that these statistics and estimations do not take into account the impact of the COVID-19 pandemic, which should not be underestimated. One study on the impact of the COVID-19 pandemic on H&N cancer diagnosis and treatment reported a decrease in the proportion of H&N patients with early-stage diagnosis, with authors positing these trends to be associated with patients' unwillingness to visit a doctor during uncertain times or lack of access [23].

Treatments for H&N cancer are multimodal and may involve a combination of surgery, radiotherapy (RT), and or systemic therapy, depending on factors related to the tumor and the patient, including tumor site, stage, operability, and patients' overall health. A typical up-front RT prescription consists of 70 Gy to the areas containing gross disease, 63 Gy to intermediate-risk areas (i.e., areas adjacent to the gross tumor volume [GTV]), and 54–56 Gy to low-risk areas (i.e., elective treated nodal regions), delivered in 33–35 fractions [24]. It is worth noting that the complexity of H&N cancers, which may originate in a number of sub-sites, warrant precise considerations and oncological approaches both for treatment planning and recovery [25]. Despite the shift towards personalized medicine and advancements in precise RT delivery techniques, such as newer treatment-planning software, improved imaging methods, and innovations like intensity-modulated radiation therapy (IMRT) [26] and volumetric-modulated arc therapy (VMAT) [27], there remains a subset of patients who do not respond to treatment and experience tumor resistance to treatment or recurrence. Therefore, having reliable and robust predictions regarding the probability of tumor response to radiotherapy is potentially valuable for tailoring treatment and improving results. For instance, treatment de-escalation, including reductions in dose or treatment volume, could be considered for patients predicted to exhibit complete responses to radiation, and those predicted to have poor responses could be offered treatment intensification.

Based on a working hypothesis that index LNs (detailed in methods) of H&N cancer patients exhibit phenotypic signals associated with treatment efficacy, the aim in this study was (i) to determine the feasibility of predicting tumor response to up-front RT by building predictive ML models trained using pre-treatment MRI radiomics features from index LN segmentations, (ii) to identify top- k features, and (iii) to test previously developed *deep texture analysis* (DTA) methodology to improve feature sets and enhance predictive capabilities. In short, DTA is an iterative process to investigate the spatial distributions of top- k features by creating feature map images and subsequently “mining” deeper-layer texture features from said feature maps. Previously, DTA methodology has shown promise in predicting H&N cancer treatment outcomes with index LN quantitative ultrasound spectroscopic (QUS) parametric maps [28] and treatment-planning CT scans [29], but this was yet explored on MRI images (to our knowledge). Figure 1 presents a graphical abstract for the DTA methodology.

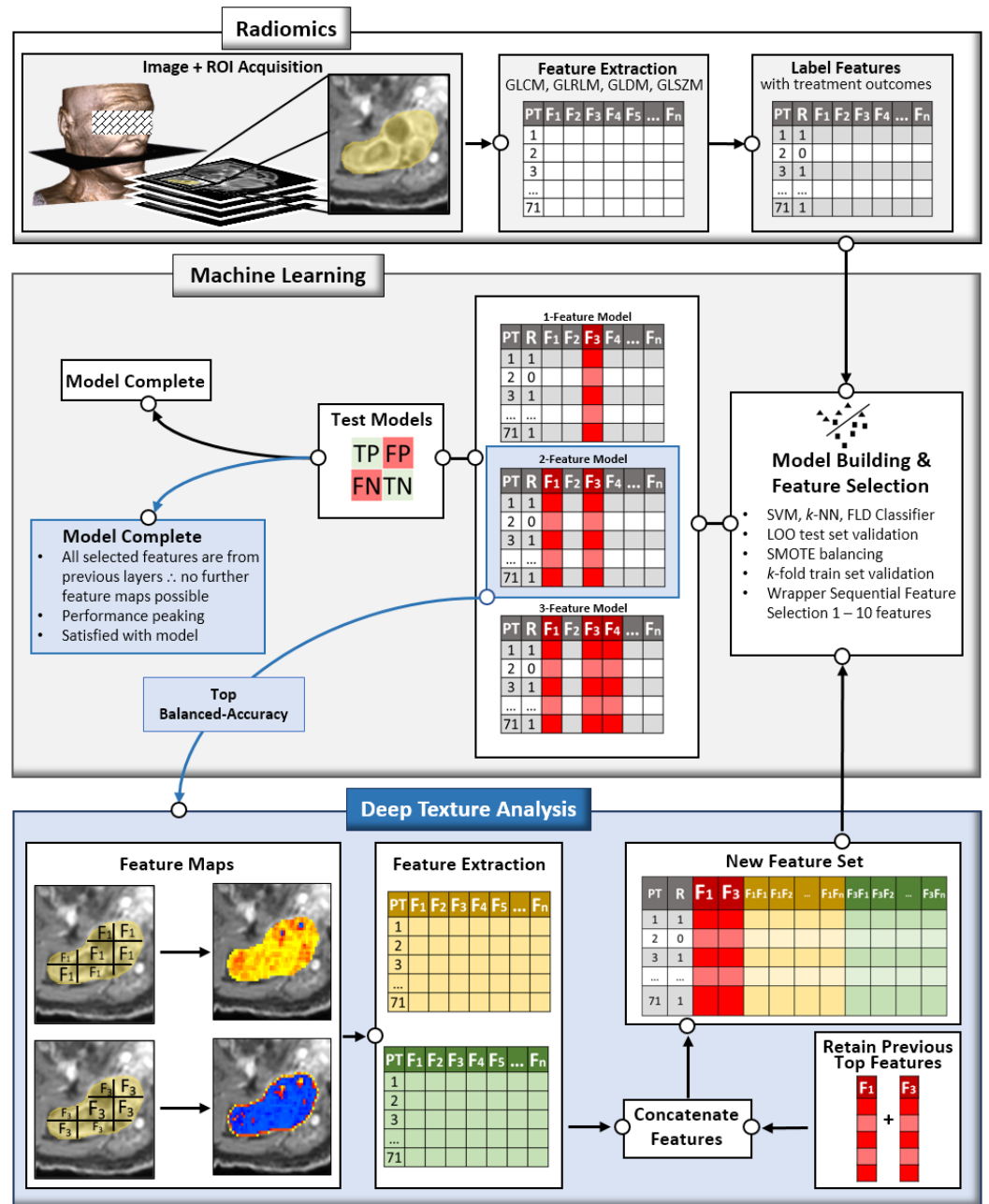


Figure 1. Graphical summary of the workflow presented in this study. The three main stages included radiomic feature extraction, the creation of ML models, and the iterative DTA method, which was carried out two times in this work, to extract features from three total layers.

2. Materials and Methods

This constitutes an analysis of a prospective cohort of patients with biopsy-proven de novo cT0-4N1-3M0 primary H&N cancer who underwent a standard course of up-front radiotherapy with 70 Gy in 33–35 fractions, with or without concurrent chemotherapy, between 2015 and 2020 ([ClinicalTrials.gov](https://clinicaltrials.gov) (accessed on 11 June 2024), identifier: NCT03908684). For the purpose of this current study, we included a subgroup of patients from that cohort who had enlarged regional lymph nodes measuring ≥ 15 mm in short axis, as assessed by CT scan, and underwent baseline magnetic resonance imaging at the time of diagnosis or radiation planning [30], while excluding patients whose MRI contained motion and dental artifacts.

As mentioned earlier, radiomics models for this patient cohort were trained using US and CT scans and published previously [28,29]. Recruiting patients for the acquisition of US scans—which are not part of the standard treatment pipeline—solely for scientific research purposes proved challenging and limited the number of participants to $n = 63$. The main objective of this study was to create machine learning algorithms using baseline MRI texture features determined from the index regional lymph node for predicting complete response three months after radical intent radiation therapy. This is a single-institution study conducted at Sunnybrook Health Sciences Center (Toronto, ON, Canada) and approved by the Research Institute Ethics Board (SUN-3047).

2.1. Treatment Approach and Follow-Up Imaging

Patients were treated with radical intent definitive up-front radiotherapy. Contouring and treatment plans were made in line with institutional guidelines and in accordance with standard practice. As part of patient clinical care for radiation treatment planning, the primary and lymph node gross tumor volumes (GTV) were contoured by a clinical team on planning CT scans, co-registered with diagnostic or planning magnetic resonance imaging data that were acquired from a 1.5 T Philips Ingenia Elition X (Philips Medical Systems, Amsterdam, Netherlands) device [30]. The typical contouring approach consisted of (i) expanding the GTV by 3–5 mm to generate the high-dose clinical target volume (CTV) and further expanding it by 5 mm to generate the high-dose planning target volume (PTV); (ii) expanding the GTV by 10 mm and including the elective nodal regions to generate the low-dose CTV. The contouring of an intermediate-risk CTV was at the physician's discretion. The PTV margins consisted of a 5 mm expansion of the respective CTVs. A typical prescription included 70 Gy to the PTV high-risk, 63 Gy to the intermediate-risk, and 56 Gy to the low-risk PTV, delivered in 33–35 fractions. RT was administered using IMRT or VMAT techniques, and cone-beam CT scans were used to verify and adjust the patient's position before each radiation fraction [30].

After completing treatment, patients were regularly monitored according to institutional clinical guidelines. The standard follow-up regimen typically included physical examinations and restaging imaging every 3 months during the first year post-treatment, every 4 months in the second year, every 6 months in the third year, and annually during years 4 and 5. The initial restaging imaging modality was MRI, and in the case of a complete response, CT scans were performed subsequently.

2.2. Tumor Response Definition and Segmentation

Tumor response was assessed for the primary tumor and regional lymph nodes on the 3-month follow-up contrast-enhanced magnetic resonance imaging following radiotherapy completion and was based on the Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 criteria [31]. In brief, complete response (CR) was defined as disappearance of the primary tumor and reduction of the involved lymph nodes in short axis to less than 10mm, whereas partial response (PR) was defined as at least a 30% decrease in the sum of diameters of tumors compared to the baseline MR imaging. The protocol outlined evaluation of patients that exhibited stable and progressive disease as well, but the cohort included no such instances and was thus not worth detailing in this methodology.

Radiation oncologists with over 5 years' experience (DMP, IP, IR or AB) used pre-treatment T1-weighted post-contrast MRI for manual segmentation of the region of interest (ROI) in a radiation planning system (Pinnacle, Philips Medical System) that was copied for analysis using open-source software ITK-SNAP (version 3.6.0, www.itksnap.org (accessed on 11 June 2024) [32]). The ROI consisted of the largest affected regional lymph node, and when multiple lymph nodes were closely adjacent or formed a conglomerate, all were included into the ROI.

2.3. Texture Extraction and Machine Learning Algorithms

LN segmentation ROIs were used as a mask with associated MRI scans. In order to account for varying and arbitrary pixel intensities between scans, images were normalized by normalizing values to the mean with one standard deviation of pixel values in the whole image [33]. Subsequently, 24 GLCM, 16 GLRLM, 16 GLDM, and 14 GLSZM 2D 1st-layer texture features (1LTFs) were calculated (as an S_1 feature-set) from axial slices using Pyradiomics version 3.0.1, an open-source Python (Python Software Foundation, Wilmington, DE, USA, version 3.7.10) package [33]. Lastly, before proceeding to ML model building, radiomics features were concatenated with retrospective binary treatment outcomes.

Models were trained using S_1 1LTFs and three common classifiers: Fischer's Linear Discriminant (FLD), Support Vector Machines (SVM), and k -Nearest Neighbors (k -NN). To account for varying magnitudes within the features, z-score scaling was carried out at the feature level. To split test and training data, a leave-one-out approach was used, whereby iteratively, each sample was left out, and models were trained, validated, and subsequently tested on the held-out sample. After the test sample was left out, but before training, the synthetic minority oversampling method (SMOTE) was utilized to account for the imbalance of data (23 CR/40 PR) [34]. During each iteration, a 5-fold split was used on the remaining data to train and validate the models. To avoid overfitting and increase model robustness, a sequential forward selection (SFS) method in a wrapper framework was used to reduce dimensionality and identify valuable S_1 features based on an increasing balanced accuracy metric. Throughout all of the leave-one-out iterations, the most frequently selected combinations of single-feature and multi-feature models were identified and tested on the left-out samples for models with 1–10 features.

The attention mechanism is one of the most influential developments within deep learning for text- and image-based studies. This mechanism is based on amplifying important areas in image or word inputs to receive more attention from the learning networks to enhance the performance [35]. Inspired by attention mechanism, we proposed deep texture analysis based on the extraction of important features within the image. To draw parallels, in this technique, radiomics features are extracted from images coded by important selected radiomics features from previous layers. Whereas in the attention mechanism, the scaled dot-product is used to obtain the importance of each token, here, the output of the classifiers determines the important features.

After identifying the most valuable 1LTFs from the S_1 feature set, texture features were calculated once again, this time, however, using a sliding window technique (window size of 3×3 pixels) for sub-ROI windows rather than for the entire LN ROI. For each of the three classifiers, S_1 feature maps were made for the models that resulted in the best-balanced accuracy test score among the 10 models. If multiple models demonstrated the same highest balanced accuracy, preference was given to the model using fewer features.

Next, using Pyradiomics, 2nd-layer texture features (2LTFs) were "mined" from the S_1 feature maps (rather than the MRI scans directly). Previously identified top- k 1LTFs from each classifier's top model were concatenated with newly determined 2LTFs to create new feature sets, $S_{2,FLD}$, $S_{2,SVM}$, and $S_{2,kNN}$, respectively. Subsequently, identical model building and testing procedures were repeated to evaluate the influence of implementing deeper-layer features on predictive models. In addition to 2LTFs, the iterative DTA methodology was repeated with the aforementioned steps to evaluate the inclusion of 3rd-layer texture features (3LTFs) as well.

To evaluate whether the inclusion of deeper-layer features was worthwhile for model building (independently of model complexity with regards to the number of features), average performances of models trained using the S_1 dataset were compared to the average performance of models trained using S_2 and S_3 datasets, using a one-tailed t -test with the significance level (α) set at $p = 0.05$, assessing the null hypothesis that there was no difference in average performance metrics.

3. Results

3.1. Patient Characteristics

A total of 63 patients were involved in this study. The median age of patients at the time of diagnosis was 61 years (range, 36–80), and 94% were male. Primary tumors were located in the oropharynx (68%, $n = 43$), larynx (6%, $n = 4$), hypopharynx (5%, $n = 3$), and nasopharynx (8%, $n = 5$). A total of 13% ($n = 8$) of tumors were of unknown or unspecified primary locations. A total of 94% of tumors ($n = 59$) were squamous cell carcinomas. A total of 86% received platinum-based concurrent chemotherapy ($n = 54$). Tumor staging, p16 status, and alcohol and tobacco consumption habits were summarized in Table 1. Supplementary Table S1 includes anonymized medical and treatment characteristics for individual patients. A total of 23 patients achieved CR and 40 PR at 3 months after radiotherapy. Figure 2 provides representative MRI scans and index LN ROIs before and after treatment for a CR and PR patient.

Table 1. Summary of demographic and clinical characteristics analyzed in patients in study cohort.

Demographics and Clinical Characteristics	<i>n</i> (%)
Age (years)	
Median (range)	61 (36–80)
Mean	60.7 ± 9.8
Sex	
Male	59 (94)
Female	4 (6)
Primary Tumor Type	
Squamous cell carcinoma	59 (94)
Other	4 (6)
Primary Tumor location	
Nasopharynx	5 (8)
Oropharynx	43 (68)
Hypopharynx	3 (5)
Larynx	4 (6)
Unspecified	8 (13)
HPV/p16 Status	
p16+	35 (55)
p16–	1 (1.5)
p16+ and p63+	1 (1.5)
HPV+	1 (1.5)
Epstein–Barr virus	2 (3)
Unknown/not tested	23 (36)
Tumor (T) and Node (N) Staging	
T ₁	4 (6)
T ₂	22 (35)
T ₃	5 (8)
T ₄	13 (21)
Unspecified	19 (30)
N ₁	8 (13)
N ₂	23 (44)
N ₃	4 (6)
Unspecified	23 (37)
Chemotherapy Regimen	
Cisplatin	43 (68)
Carboplatin	6 (10)
No chemotherapy	9 (14)
Combination Cis/Carboplatin	5 (8)

Table 1. Cont.

Demographics and Clinical Characteristics	n (%)
Smoking Habits	
Smoker	36 (57)
Non-smoker	21 (33)
Unspecified	6 (10)
Drinking Habits	
Occasional drinker	23 (36)
Heavy drinker	17 (27)
Non-drinker	13 (21)
Unspecified	10 (16)
Post-Treatment Assessment	
Complete responder (CR)	23 (36.5)
Partial responder (PR)	40 (63.5)

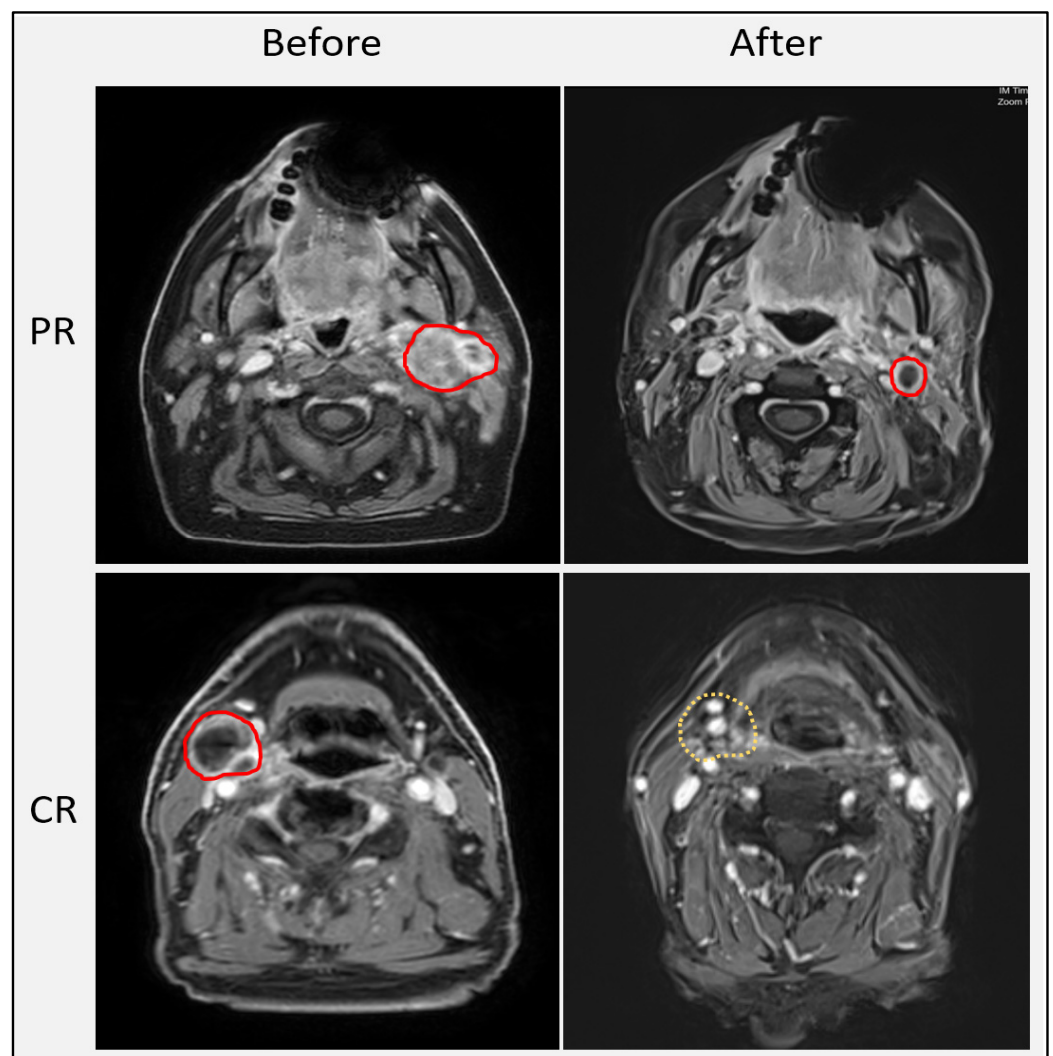


Figure 2. Pre- and post-treatment axial T₁-weighted MRIs for one CR and one PR patient, with diseased LN ROIs in red and treated area in yellow.

3.2. Models Trained with S_1 Feature Set

Seventy 1LTFs were determined from LN segmentations from axial slices of MRI scans to create the S_1 feature set used to train models. As presented in Table 2, FLD, k -NN, and SVM classifier predictive models demonstrated a capability to differentiate between CR- and PR-labeled patients, with varying effectiveness. For each classifier, the best-performing model (highest balanced accuracy with the lowest number of features) was identified and is highlighted in Table 2.

Table 2. Results from SVM, k -NN, and FLD classifier models trained using S_1 feature set selecting for 1–10 features. For each classifier, the highest balanced accuracy is bolded, and highlighted in gray are the corresponding metrics for said models. For each classifier, features from the top balanced accuracy model were used to create texture feature maps for DTA.

No. of Features	Sensitivity (%)			Specificity (%)			Accuracy (%)			Precision (%)			Balanced Accuracy (%)			AUC (%)		
	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD
1	80	75	82.5	56.5	65.2	65.2	71.4	71.4	76.2	76.2	78.9	80.5	68.3	70.1	73.9	72.1	68.2	79.5
2	85	75	82.5	60.9	69.6	65.2	76.2	73	76.2	79.1	81.1	80.5	72.9	72.3	73.9	77.9	70.7	77.3
3	85	72.5	80	60.9	69.6	69.6	76.2	71.4	76.2	79.1	80.6	82.1	72.9	71	74.8	78.7	71.9	78.4
4	85	75	80	69.6	69.6	69.6	79.4	73	76.2	82.9	81.1	82.1	77.3	72.3	74.8	80.1	75.3	77.7
5	82.5	75	80	65.2	69.6	65.2	76.2	73	74.6	80.5	81.1	80	73.9	72.3	72.6	79.9	70.9	78.4
6	77.5	75	82.5	73.9	69.6	69.6	76.2	73	77.8	83.8	81.1	82.5	75.7	72.3	76	78.6	70.3	77.7
7	72.5	75	82.5	73.9	65.2	69.6	73	71.4	77.8	82.9	78.9	82.5	73.2	70.1	76	79.1	71.4	79.5
8	70	80	80	78.3	69.6	65.2	73	76.2	74.6	84.8	82.1	80	74.1	74.8	72.6	77.5	72.5	79.3
9	70	80	77.5	82.6	73.9	65.2	74.6	77.8	73	87.5	84.2	79.5	76.3	77	71.4	77.3	72	79
10	90	80	82.5	56.5	69.6	69.6	77.8	76.2	77.8	78.3	82.1	82.5	73.3	74.8	76	78.5	72.5	79.3

The best performance among the ten SVM classifier models was with the four-feature multivariable model, with a sensitivity ($\%S_n$) of 85%, specificity ($\%S_p$) of 70%, accuracy ($\%Acc$) of 79%, precision ($\%Pre$) of 83%, balanced accuracy ($\%BA$) of 77%, and an area under ($\%AUC$) the receiver operator characteristic (ROC) curve of 80%. Of the four selected features, three were GLCM features, including “Difference Entropy”, “Joint Entropy”, and “Joint Energy”, and one was a GLRLM feature, “Short Run Emphasis”. Feature maps were made for the aforementioned four features, using the sliding window technique, from which 2LTFs were determined in preparation of the $S_{2,SVM}$ feature set.

The best performance among the ten k -NN ($k = 5$) classifier models was seen with the nine-feature multivariable model, with $\%S_n = 80\%$, $\%S_p = 74\%$, $\%Acc = 78\%$, $\%Pre = 84\%$, $\%BA = 77\%$, and $\%AUC = 72\%$. The nine selected features were three GLCM features, i.e., “Maximum Probability”, “IDM”, and “Joint Energy”; two GLRLM features, i.e., “Run Percentage” and “Low Gray Level Run Emphasis”; three GLSZM features, i.e., “Small Area Emphasis”, “Small Area Low Gray Level Emphasis”, and “Size Zone Non-Uniformity Normalized”; and one GLDM feature, i.e., “Low Gray Level Emphasis”. Feature maps were made for the aforementioned nine features, using the sliding window technique, from which 2LTFs were determined in preparation for the $S_{2,KNN}$ feature set.

The best performance among the ten FLD classifier models was seen with the six-feature multivariable model, with $\%S_n = 83\%$, $\%S_p = 70\%$, $\%Acc = 78\%$, $\%Pre = 83\%$, $\%BA = 76\%$, and $\%AUC = 78\%$. The six selected features were three GLCM features, i.e., “Difference Average”, “MCC”, and “Inverse Variance”; two GLDM features, i.e., “Small Dependence Low Gray Level Emphasis” and “Large Dependence Low Gray Level Emphasis”; and one GLRLM feature, i.e., “Run Variance”. Feature maps were made for the aforementioned six features, using the sliding window technique, from which 2LTFs were determined in preparation for the $S_{2,FLD}$ feature set.

3.3. Models Trained with S₂ Feature Sets

Feature maps were made from the selected features from each classifiers’ best-performing model. From each of the newly created feature maps, once again, 70 radiomics features were determined to create second-layer texture features (2LTFs). S₂ feature set preparation included retaining top-*k* selected 1LTFs from the first iteration of model building and concatenating with newly mined 2LTFs. The S_{2,SVM} feature set was comprised of 284 total features ((4 × 1LTFs) + (70 2LTFs × 4 sets of 1LTF maps)). The S_{2,KNN} feature set was comprised of 639 total features ((9 × 1LTFs) + (70 2LTFs × 9 sets of 1LTF maps)). The S_{2,FLD} feature set was comprised of 426 total features ((6 × 1LTFs) + (70 2LTFs × 6 sets of 1LTF maps)).

After S₂ feature set preparation was completed, predictive models were trained, validated, and tested again, with model building parameters unchanged. Ten models were built for each classifier with 1–10 features, and the results are presented in Table 3. The best performances were seen with the nine-, four-, and six-feature multivariable models for the SVM, *k*-NN, and FLD classifiers, respectively.

Table 3. Results from SVM, *k*-NN, and FLD classifier models trained using S_{2,SVM}, S_{2,kNN}, S_{2,FLD} feature sets and selecting for 1–10 features. For each classifier, the highest balanced accuracy is bolded, and highlighted in gray are the corresponding metrics for said models. 2LTFs selected for in each classifiers’ top model were used to create texture feature maps for DTA.

No. of Features	Sensitivity (%)			Specificity (%)			Accuracy (%)			Precision (%)			Balanced Accuracy (%)			AUC (%)		
	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD
1	80	87.5	72.5	65.2	60.9	78.3	74.6	77.8	74.6	80	79.5	85.3	72.6	74.2	75.4	72	65	75.3
2	80	85	72.5	65.2	69.6	73.9	74.6	79.4	73	80	82.9	82.9	72.6	77.3	73.2	74.3	78.6	76.2
3	82.5	82.5	67.5	60.9	56.5	87	74.6	73	74.6	78.6	76.7	90	71.7	69.5	77.2	74.9	69.4	80.8
4	85	90	70	65.2	69.6	87	77.8	82.5	76.2	81	83.7	90.3	75.1	79.8	78.5	78.5	75.5	81.2
5	80	90	70	65.2	69.6	87	74.6	82.5	76.2	80	83.7	90.3	72.6	79.8	78.5	79.1	75.5	82.4
6	80	87.5	72.5	69.6	69.6	87	76.2	81	77.8	82.1	83.3	90.6	74.8	78.5	79.7	77.6	74.1	82.4
7	80	75	72.5	69.6	60.9	87	76.2	69.8	77.8	82.1	76.9	90.6	74.8	67.9	79.7	79.9	67.7	82.9
8	82.5	72.5	72.5	69.6	52.2	87	77.8	65.1	77.8	82.5	72.5	90.6	76	62.3	79.7	79.2	59.7	82.9
9	85	70	70	69.6	52.2	87	79.4	63.5	76.2	82.9	71.8	90.3	77.3	61.1	78.5	80	57.7	82.1
10	82.5	87.5	72.5	69.6	65.2	87	77.8	79.4	77.8	82.5	81.4	90.6	76	76.4	79.7	80.2	75.9	84

The best performance among the ten SVM classifier models trained with the S_{2,SVM} feature set was seen with the nine-feature multivariable model, with %S_n = 85%, %S_p = 70%, %Acc = 79%, %Pre = 83%, %BA = 77%, and %AUC = 80%. Of the nine selected features, four were the originally selected 1LTFs, and five were newly introduced 2LTFs. Of the selected five 2LTFs, three were determined from the GLCM “Joint Energy” feature map. Those features were GLCM “Sum Entropy” and “Joint energy”, as well as GLSZM “Gray Level Non-Uniformity Normalized”. Of the remaining two 2LTFs, one was GLCM “Joint Energy” determined from the GLCM “Difference Entropy” feature map, and the other was GLRLM “Long Run Emphasis” determined from the GLRLM “Short Run Emphasis” feature map.

The best performance among the ten *k*-NN classifier models trained with the S_{2,kNN} feature set was seen with the four-feature multivariable model, with %S_n = 90%, %S_p = 70%, %Acc = 83%, %Pre = 84%, %BA = 80%, and %AUC = 76%. Of the four selected features, two were the originally selected 1LTFs, and two were newly introduced 2LTFs. The selected 1LTFs were GLCM “IDM” and “Maximum Probability”, and the two 2LTFs were GLCM “Maximum Probability” determined from the GLRLM “Run Percentage” feature map and GLSZM “Size Zone Non-Uniformity Normalized” determined from the GLSZM “Small Area Low Gray Level Emphasis” feature map.

The best performance among the ten FLD classifier models trained with the S_{2,FLD} feature set was seen with the six-feature multivariable model, with %S_n = 73%, %S_p = 87%, %Acc = 78%, %Pre = 91%, %BA = 80%, and %AUC = 82%. All six of the selected features were newly introduced 2LTFs. These features were GLCM “MCC” and GLRLM “Run

Entropy” determined from the GLCM “MCC” feature map, GLCM “IDN” and GLSZM “Zone Entropy” from the GLCM “Inverse Variance” feature map, and GLCM “Joint Entropy” from both the GLRLM “Run Variance” feature map and the GLDM “Small Dependence Low Gray Level Emphasis” feature map, respectively.

3.4. Models Trained with S₃ Feature-Sets

Selected features for each classifiers’ top-performing model were retained for S₃ feature sets. Using Pyradiomics, feature maps were made from the selected 2LTFs and their associated 1LTF map. Figure 3 shows an example of a CR and PR patient with their associated MRI, LN ROI, an example of an 1LTF map, and an 2LTF map. From each of the newly created 2LTF maps associated with the three classifiers’ best-performing model, seventy third-layer texture features (3LTFs) were determined and concatenated to the previously retained first-layer and or second-layer features selected for by models trained with S₂ feature-sets. The S_{3,SVM} feature set was comprised of 359 total features ((4 × 1LTFs) + (5 × 2LTFs) + (70 3LTFs × 5 sets of 2LTF maps)). The S_{2,KNN} feature set was comprised of 144 total features ((2 × 1LTFs) + (2 × 2LTFs) + (70 3LTFs × 2 sets of 2LTF maps)). S_{3,FLD} consisted of 426 total features ((6 × 2LTFs) + (70 3LTFs × 6 sets of 2LTFs)).

After S₃ feature set preparation was completed, predictive models were trained, validated, and tested again, with all model-building parameters unchanged. For each classifier, ten models were built with 1–10 features, and the results can be seen in Table 4. The best performances were seen with the six-, nine-, and four-feature multivariable models for the SVM, k-NN, and FLD classifiers, respectively.

Table 4. Results from SVM, k-NN, and FLD classifier models trained using S_{3,SVM}, S_{3,kNN}, S_{3,FLD} feature sets and selecting for 1–10 features. For each classifier, the highest balanced accuracy is bolded, and highlighted in gray are the corresponding metrics for said models.

No. of Features	Sensitivity (%)			Specificity (%)			Accuracy (%)			Precision (%)			Balanced Accuracy (%)			AUC (%)		
	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD	SVM	KNN	FLD
1	80	85	77.5	69.6	60.9	73.9	76.2	76.2	76.2	82.1	79.1	83.8	74.8	72.9	75.7	73.3	64.7	76.2
2	80	87.5	72.5	65.2	69.6	78.3	74.6	81	74.6	80	83.3	85.3	72.6	78.5	75.4	74.7	79	79.8
3	82.5	90	75	65.2	69.6	82.6	76.2	82.5	77.8	80.5	83.7	88.2	73.9	79.8	78.8	74.8	75.3	79
4	82.5	87.5	75	65.2	69.6	87	76.2	81	79.4	80.5	83.3	90.9	73.9	78.5	81	74.9	74.4	80.7
5	77.5	90	75	65.2	69.6	87	73	82.5	79.4	79.5	83.7	90.9	71.4	79.8	81	73.6	73.8	81.4
6	85	90	75	69.6	73.9	87	79.4	84.1	79.4	82.9	85.7	90.9	77.3	82	81	77.2	74.3	80.4
7	82.5	90	72.5	69.6	69.6	87	77.8	82.5	77.8	82.5	83.7	90.6	76	79.8	79.7	78.5	76.4	81.1
8	85	90	72.5	69.6	73.9	87	79.4	84.1	77.8	82.9	85.7	90.6	77.3	82	79.7	79.2	74.7	79.3
9	77.5	92.5	72.5	69.6	73.9	87	74.6	85.7	77.8	81.6	86	90.6	73.5	83.2	79.7	79.6	74.9	80.7
10	85	90	72.5	69.6	69.6	82.6	79.4	82.5	76.2	82.9	83.7	87.9	77.3	79.8	77.6	79.8	74.7	79.6

The best performance among the ten SVM classifier models trained with the S_{3,SVM} feature set was seen with the six-feature multivariable model, with %S_n = 85%, %S_p = 70%, %Acc = 79%, %Pre = 83%, %BA = 77%, and %AUC = 77%. All six of the selected features were first- and second-layer texture features. None of the 350 available 3LTFs were selected, and the top performance of the S_{3,SVM} feature set-trained SVM classifier model matched but did not exceed the performance of the best SVM model trained with the S_{2,SVM} feature set. The selected features were two 1LTFs (GLCM “Joint Energy” and “Joint Entropy”) and four 2LTFs (GLCM “Joint Energy” and “Sum Entropy” from the GLCM “Joint Energy” feature map, GLCM “Joint Energy” from the GLCM “Difference Entropy” feature map, and GLRLM “Long Run Emphasis” from the GLRLM “Short Run Emphasis” feature map).

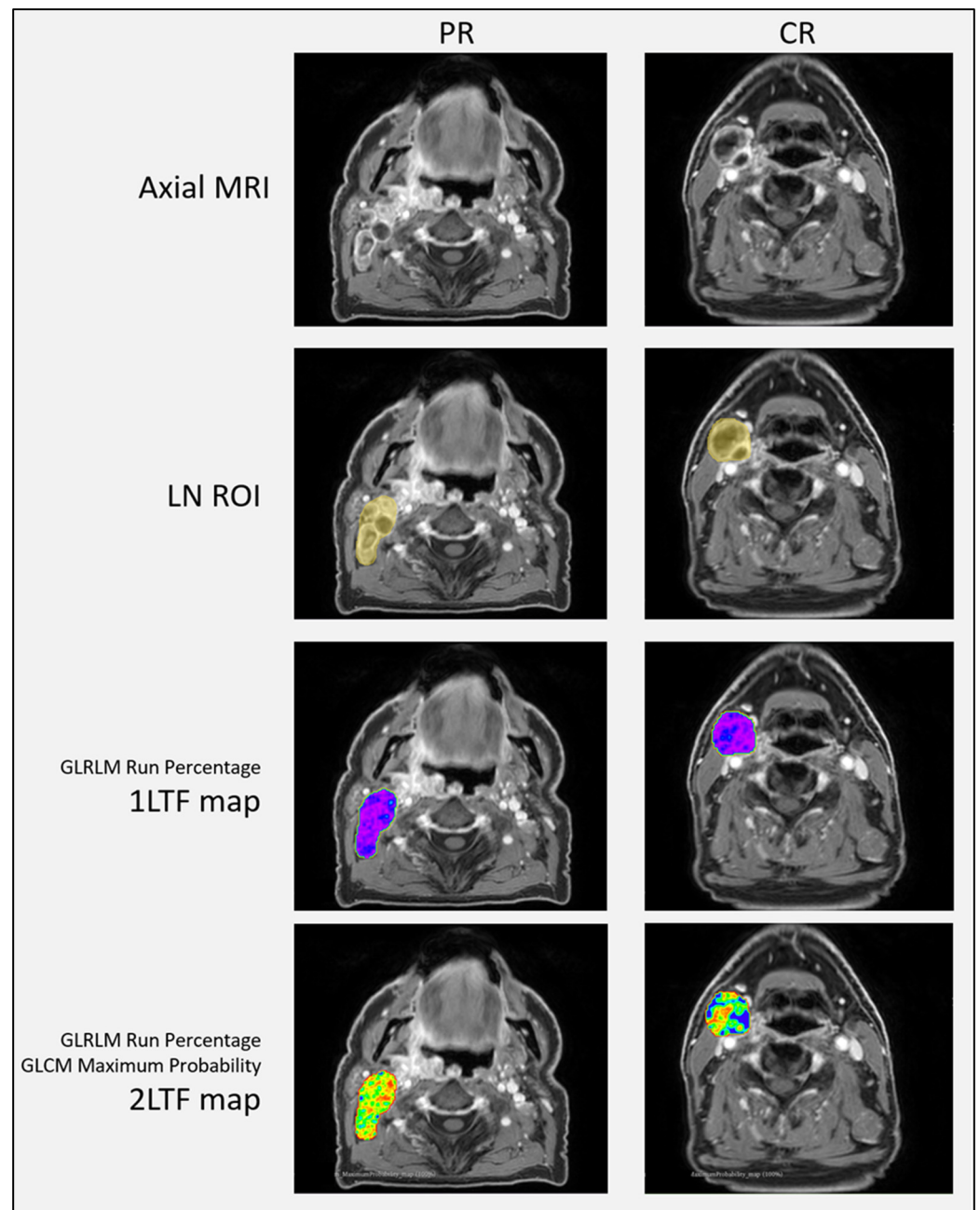


Figure 3. Axial MRI and associated LN ROI, 1LTF map, and 2LTF map for one PR and one CR patient showing the potential importance of spatial distribution of features as a source of phenotypic signaling.

The best performance among the ten k -NN classifier models trained with the $S_{3,kNN}$ feature set was seen with the nine-feature multivariable model, with $\%S_n = 93\%$, $\%S_p = 74\%$, $\%Acc = 86\%$, $\%Pre = 86\%$, $\%BA = 83\%$, and $\%AUC = 75\%$. This was also the best-performing model in the study. Of the nine selected features, two were 1LTFs, two were 2LTFs, and the remaining five were 3LTFs. The two 1LTFs were GLCM “Maximum Probability” and “IDM”. The two 2LTFs were GLCM “Maximum Probability” from the GLRLM “Run Percentage” feature map and GLSZM “Size Zone Non-Uniformity Normalized” from the GLSZM “Small Area Low Gray Level Emphasis” 1LTF map. Of the five 3LTFs, four were determined from the GLRLM “Run Percentage” GLCM “Maximum Probability” 2LTF map, and those were GLCM “Joint Energy”, “Inverse Variance”, “IDMN”, and “IDN”. The final selected 3LTF was GLRLM “Low Gray Level Run Emphasis” determined from the

GLSZM “Small Area Low Gray Level Emphasis” and GLSZM “Size Zone Non-Uniformity Normalized” 2LTF maps.

The best performance among the ten FLD classifier models trained with the $S_{3,FLD}$ feature set was seen with the four-feature multivariable model, with $\%S_n = 75\%$, $\%S_p = 87\%$, $\%Acc = 79\%$, $\%Pre = 91\%$, $\%BA = 81\%$, and $\%AUC = 81\%$. Of the four selected features, two were 2LTFs, and two were 3LTFs. The two 2LTFs were GLCM “MCC” from the GLCM “MCC” 1LTF map and GLDM “Small Dependence Low Gray Level Emphasis” from the GLCM “Joint Entropy” 1LTF map. The two 3LTFs were GLCM “MCC” from the GLCM “MCC” GLRLM “Run Entropy” 2LTF map and GLCM “Difference Average” from the GLRLM “Run Variance” GLCM “Joint Entropy” 2LTF map.

3.5. Evaluating Deeper-Layer Features

Next, an assessment was conducted to determine whether incorporating and processing deeper-layer texture features was beneficial. For the ten models created at each layer, the average performance of models trained using S_n feature sets was compared to $S_n + 1$ and $S_n + 2$ (when possible) using a one-tailed *t*-test with the significance level (α) set at $p = 0.05$, assessing the null hypothesis that there was no difference in average performance metrics. Tables 5–7 summarize the findings for the inclusion of deeper-layer texture features for the SVM, *k*-NN, and FLD classifiers, respectively.

Table 5. Average of 10 SVM classifier models trained using S_1 , $S_{2,SVM}$, and $S_{3,SVM}$ feature sets as well as associated *p*-values from one-tailed *t*-test to evaluate enhancement.

Average Performance (%)	S_1	$S_{2,SVM}$	<i>p</i> Value	$S_{2,SVM}$	$S_{3,SVM}$	<i>p</i> Value	S_1	$S_{3,SVM}$	<i>p</i> Value
Sensitivity	79.8 ± 6.7	81.8 ± 1.9	0.208	81.8 ± 1.9	81.8 ± 2.8	0.5	79.8 ± 6.7	81.8 ± 2.8	0.205
Specificity	67.8 ± 2.6	67.0 ± 2.9	0.367	67.0 ± 2.9	67.8 ± 2.2	0.083	67.8 ± 2.6	67.8 ± 2.2	0.5
Accuracy	75.4 ± 2.3	76.3 ± 1.7	0.148	76.3 ± 1.7	76.7 ± 2.2	0.339	75.4 ± 2.3	76.7 ± 2.2	0.134
Precision	81.5 ± 3.3	81.2 ± 1.4	0.349	81.2 ± 1.4	81.5 ± 1.2	0.141	81.5 ± 3.3	81.5 ± 1.2	0.486
Balanced accuracy	73.8 ± 2.3	74.4 ± 1.8	0.207	74.4 ± 1.8	74.8 ± 2.0	0.253	73.8 ± 2.3	74.8 ± 2.0	0.178
AUC	78 ± 2.2	77.6 ± 2.7	0.294	77.6 ± 2.7	76.5 ± 2.4	0.071	78 ± 2.2	76.5 ± 2.4	0.090

Table 6. Average of 10 *k*-NN classifier models trained using S_1 , $S_{2,KNN}$, and $S_{3,KNN}$ feature sets and associated *p*-values from one-tailed *t*-test to evaluate enhancement. Bolded are *p*-values for metrics that demonstrated statistically significant enhancement with inclusion of features from deeper layers.

Average Performance (%)	S_1	$S_{2,KNN}$	<i>p</i> Value	$S_{2,KNN}$	$S_{3,KNN}$	<i>p</i> Value	S_1	$S_{3,KNN}$	<i>p</i> Value
Sensitivity	76.3 ± 2.6	82.8 ± 7.1	0.025	82.8 ± 7.1	89.3 ± 1.9	0.022	76.3 ± 2.6	89.3 ± 1.9	<0.001
Specificity	69.1 ± 2.3	62.6 ± 6.8	0.014	62.6 ± 6.8	70.0 ± 3.6	0.012	69.1 ± 2.3	70.0 ± 3.6	0.171
Accuracy	73.7 ± 2.2	75.4 ± 6.7	0.264	75.4 ± 6.7	82.2 ± 2.4	0.016	73.7 ± 2.2	82.2 ± 2.4	<0.05
Precision	81.1 ± 1.5	79.3 ± 4.3	0.149	79.3 ± 4.3	83.8 ± 1.9	0.014	81.1 ± 1.5	83.8 ± 1.9	<0.05
Balanced accuracy	72.7 ± 2.1	72.7 ± 6.7	0.498	72.7 ± 6.7	79.6 ± 2.7	0.015	72.7 ± 2.1	79.6 ± 2.7	<0.001
AUC	71.6 ± 1.7	69.9 ± 6.9	0.245	69.9 ± 6.9	74.2 ± 3.5	0.043	71.6 ± 1.7	74.2 ± 3.5	0.013

Table 7. Average of 10 FLD classifier models trained using S_1 , $S_{2,FLD}$, and $S_{3,FLD}$ feature sets and associated results for one-tailed *t*-test *p*-value. Bolded are *p*-values for metrics that demonstrated statistically significant enhancement with inclusion of features from deeper layers.

Average Performance (%)	S_1	$S_{2,FLD}$	<i>p</i> Value	$S_{2,FLD}$	$S_{3,FLD}$	<i>p</i> Value	S_1	$S_{3,FLD}$	<i>p</i> Value
Sensitivity	81.0 ± 1.7	71.3 ± 1.7	<0.05	71.3 ± 1.7	74.0 ± 1.7	<0.01	81.0 ± 1.7	74.0 ± 1.7	<0.05
Specificity	67.4 ± 2.2	84.8 ± 4.5	<0.01	84.8 ± 4.5	83.9 ± 4.4	0.171	67.4 ± 2.2	83.9 ± 4.4	<0.05
Accuracy	76.0 ± 1.5	76.2 ± 1.6	0.405	76.2 ± 1.6	77.6 ± 1.5	<0.01	76.0 ± 1.5	77.6 ± 1.5	<0.05
Precision	81.2 ± 1.2	89.2 ± 2.6	<0.05	89.2 ± 2.6	89 ± 2.5	0.351	81.2 ± 1.2	89.0 ± 2.5	<0.05
Balanced accuracy	74.2 ± 1.6	78.0 ± 2.0	<0.05	78.0 ± 2.0	79.0 ± 2.0	<0.05	74.2 ± 1.6	79.0 ± 2.0	<0.05
AUC	78.6 ± 0.8	81.0 ± 2.8	0.012	81 ± 2.8	79.8 ± 1.4	0.061	78.6 ± 0.8	79.8 ± 1.4	<0.05

The one-tailed *t*-test to evaluate meaningful improvement suggests that the inclusion of deeper-layer texture features and the DTA method did not statistically improve or degrade the performances of the SVM models. Figure 4 is a visual representation of the results from Table 5.

As seen in Table 6, comparing the *k*-NN classifier models trained using the S_1 and $S_{2,KNN}$ feature sets showed that average sensitivity increased significantly ($p = 0.025$) from $\overline{\%S_n} = 76\%$ to 83% , whereas average specificity decreased significantly ($p = 0.014$) from $\overline{\%S_p} = 69\%$ to 63% . Accuracy, precision, balanced accuracy, and AUC did not change significantly. After the next iteration of DTA, comparing the average performance of the *k*-NN classifier models trained using the $S_{2,KNN}$ and $S_{3,KNN}$ feature sets, every metric demonstrated a statistically significant improvement. Average sensitivity increased from $\overline{\%S_n} = 83\%$ to 89% , average specificity from $\overline{\%S_p} = 63\%$ to 70% , average accuracy from $\overline{\%Acc} = 75\%$ to 82% , average precision from $\overline{\%Pre} = 79\%$ to 84% , average balanced accuracy from $\overline{\%BA} = 73\%$ to 80% , and average AUC from $\overline{\%AUC} = 70\%$ to 74% .

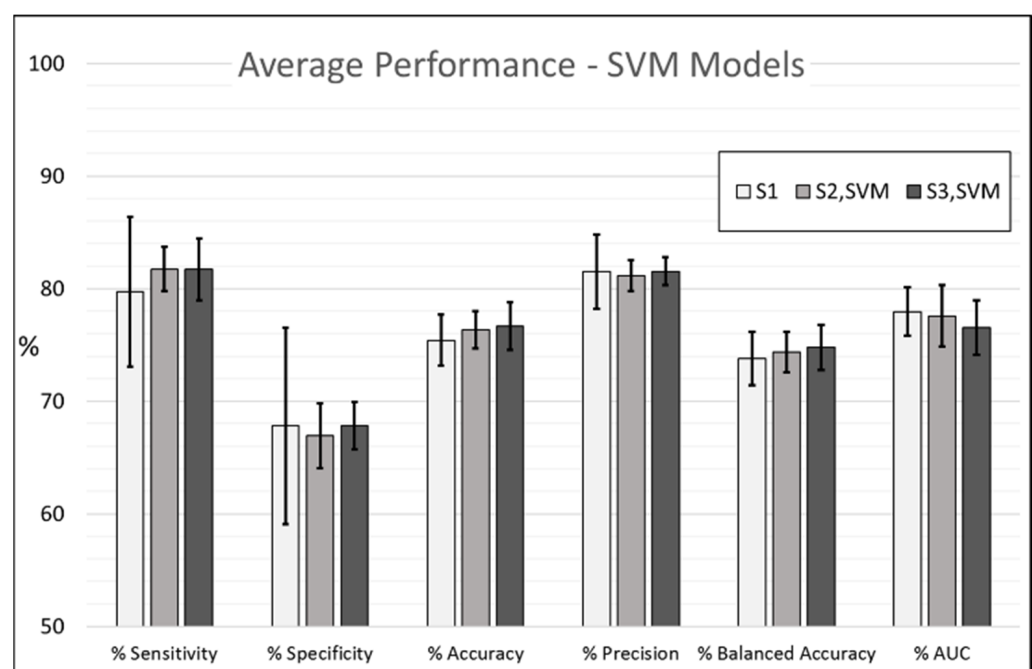


Figure 4. Average performance metrics to predict treatment outcomes with SVM classifier models trained using S_1 , $S_{2,SVM}$, and $S_{3,SVM}$ feature sets.

Figure 5 visually presents the results from Table 6. Comparing the performances of the *k*-NN classifier models trained using S_1 to models trained using $S_{3,KNN}$ shows that with the exception of average specificity, all other metrics improved significantly with the inclusion of 2LTFs and 3LTFs. Average sensitivity increased from $\overline{\%S_n} = 76\%$ to 89% , average accuracy from $\overline{\%Acc} = 74\%$ to 82% , average precision from $\overline{\%Pre} = 81\%$ to $\bar{e} = 84\%$, average balanced accuracy from $\overline{\%BA} = 73\%$ to 80% , and average AUC from $\overline{\%AUC} = 72\%$ to 74% .

Finally, the results for the FLD classifier models trained using the S_1 and $S_{2,FLD}$ feature sets are shown in Table 7, where it can be seen that average sensitivity decreased significantly from $\overline{\%S_n} = 81\%$ to 71% , whereas average specificity, average precision, average balanced accuracy, and average AUC increased significantly from $\overline{\%S_p} = 67\%$ to 85% , $\overline{\%Pre} = 81\%$ to 89% , $\overline{\%BA} = 74\%$ to 78% , and $\overline{\%AUC} = 79\%$ to 81% , respectively.

Evaluating the DTA methodology one layer deeper and comparing the average performance of FLD classifier models trained using the $S_{2,FLD}$ feature set to models trained using the $S_{3,FLD}$ feature set, the one-tailed *t*-test revealed that average sensitivity, average accuracy, and average balanced accuracy improved significantly, from $\overline{\%S_n} = 71\%$ to

$\overline{\%S_n} = 74\%$, $\overline{\%Acc} = 76\%$ to $\overline{\%Acc} = 78\%$, and $\overline{\%BA} = 78\%$ to $\overline{\%BA} = 79\%$, respectively. Specificity, precision, and AUC did not change significantly. Figure 6 visually presents the results of Table 7. Comparing average FLD classifier models trained with two iterations of DTA ($S_{3,FLD}$) to average models trained with the original radiomics features (S_1) results in significant enhancement of the average specificity, average accuracy, average precision, and average, while showing a significant decrease in sensitivity.

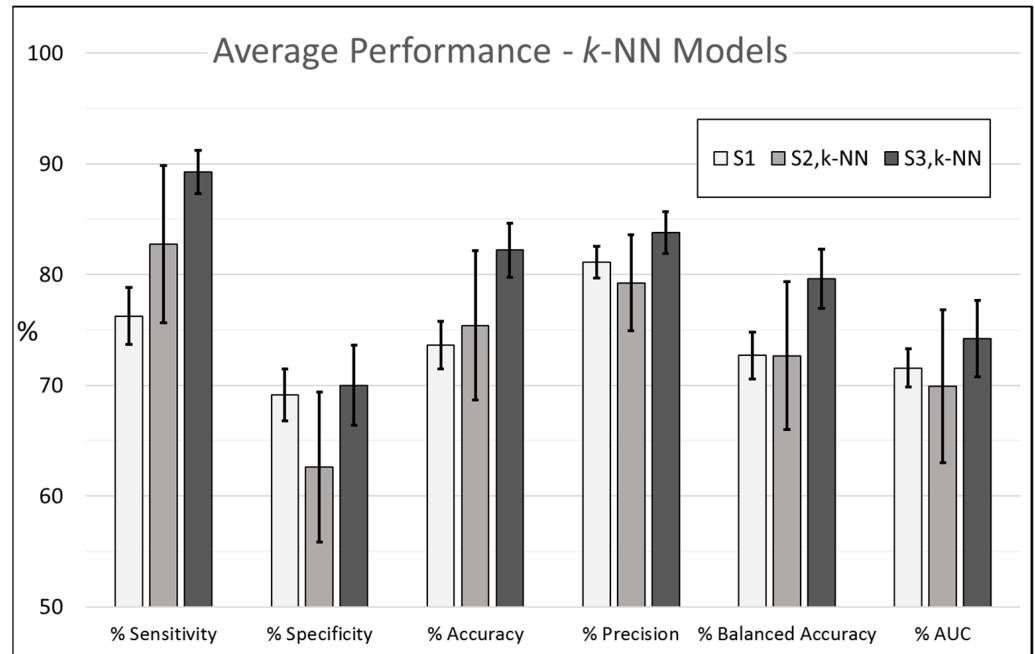


Figure 5. Average performance metrics to predict treatment outcomes with *k*-NN classifier models trained using S_1 , $S_{2,KNN}$, and $S_{3,KNN}$ feature sets.

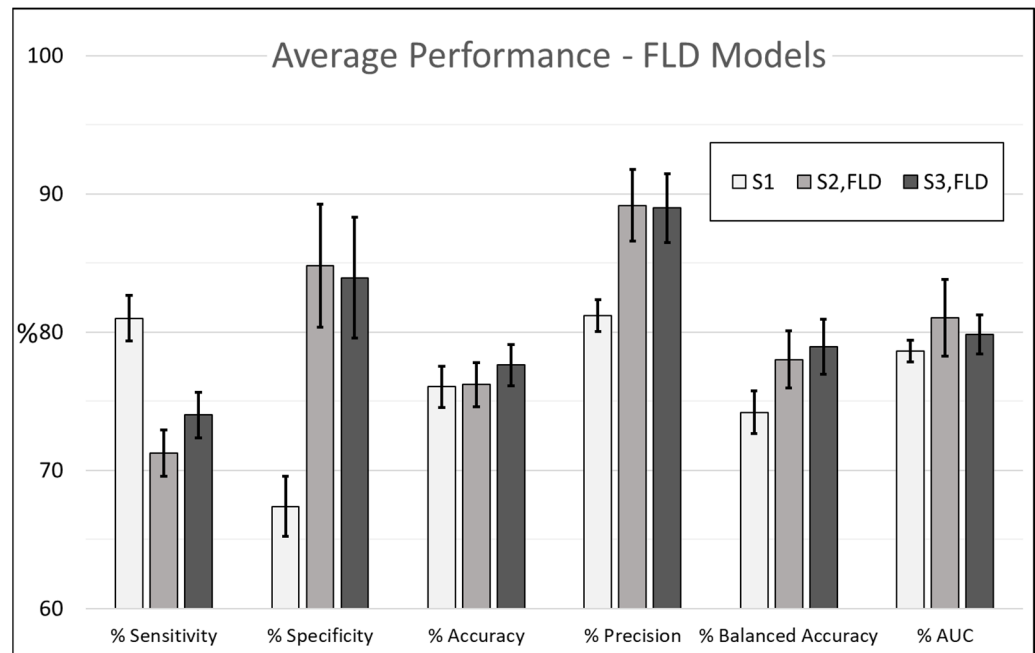


Figure 6. Average performance metrics to predict treatment outcomes with FLD classifier models trained using S_1 , $S_{2,FLD}$, and $S_{3,FLD}$ feature sets.

4. Discussion

With growing appreciation of and a shift toward personalized medicine, clinicians and researchers have come to embrace the idea that a “one-size-fits-all” approach is often suboptimal, and that identifying clinically relevant biomarkers can lead to valuable breakthroughs, on both population levels and, by extension, on individual patients’ outcomes. These biomarkers can include clinical biomarkers (such as HPV status), genomic biomarkers, and/or imaging biomarkers, referred to as radiomics features [36]. Though RT and systemic therapy (such as chemotherapy) are two of the most prominent and widely implemented approaches for treating cancer, they also present the possibility for patients to experience several undesirable side effects, such as cardio-cytotoxicity, nephrotoxicity, myelosuppression, neurotoxicity, hepatotoxicity, gastrointestinal toxicity, and mucositis, to name a few, and they may exaggerate the already challenging burden of cancer on patients [37]. An in-depth review of the physiological mechanisms associated with negative treatment side effects can be found in the work of Liu et al. [37]. Similarly, Rocha et al. provided sub-site specific insights regarding treatment side effects of H&N cancers and qualitative imaging features that may aid clinicians in identifying said symptoms from medical images [38]. Although vast improvements with precision delivery and patient-specific customization have markedly improved survival and recurrence rates, the unfortunate reality of cancer treatment is that it does not necessarily yield desired outcomes, as shown in Figure 2. Accurate and reliable a priori insights about an individual’s response to upcoming treatment could provide clinicians with a valuable tool in aiding with both treatment planning and decision making. For example, if a patient is predicted to respond well to treatment, they could be given reassurances in regards to the likelihood for success and overcome potential hesitations or fears of ineffective treatment. Consequently, if a patient is predicted to exhibit an insufficient response to treatment, physicians may adjust treatment doses or fractionation, or perhaps advise against undergoing treatment and sparing the patient the potential aforementioned undesirable side effects.

This study explored phenotypic insights from regional involved LNs of H&N cancer patients ($n = 63$) by “mining” 2D radiomics features from axial slices of pre-treatment T1-weighted post-contrast MRI scans of cancer patients (S_1 feature set) and using ex post binary treatment outcomes to train and test predictive models with three common ML classifiers (SVM, k -NN, FLD). Crucially, the DTA methodology was explored to evaluate whether the inclusion of deeper-layer features enhanced model performances, suggesting that the proposed method may be a worthwhile consideration for future radiomics studies.

For each of the three classifiers, ten models were created using an SFS method for 1–10 features. Throughout the study, the best model was considered as the one with the highest %BA and the lowest number of features. For the SVM, k -NN, and FLD classifier models trained using the S_1 feature set, the outstanding models were the four-, nine-, and six-feature multivariable models, with %BA = 77%, 77%, and 76%, respectively. Interestingly “Joint Energy”, a GLCM feature, was selected for both the SVM and k -NN classifier models. “Joint Energy” is a measure of homogenous patterns, with a greater value implying more instances of pixel intensity pairs in the image that neighbor each other at high frequencies [33]. Taking into account the role of tumor heterogeneity with regards to outcomes, it is interesting that in this study, the average GLCM “Joint Energy” value for the CR and PR group was 3.3×10^{-4} and 1.6×10^{-4} , respectively (arbitrary units), suggesting that treatment was more effective for patients exhibiting more homogeneous patterns.

In recent years, with advancements in computing power and affordability, researching ML modeling has increased in popularity. Coupled with improvements in imaging techniques, this has resulted in growing interest for radiomics investigations aimed at predicting a variety of biological endpoints. For example, Niu et al. found that by using contrast-enhanced T1- and T2-weighted MRI images, preoperative prediction of cavernous sinus invasion by pituitary adenomas was possible, with %AUC = 82.6% and 87.1% for exclusively radiomics features and radiomics + clinical features, respectively [39]. Tang et al. found it possible to predict recurrence within two years of locally advanced esophageal SCC

with radiomics features with a sensitivity of $\%S_n = 87\%$ with a sample size of $n = 220$ [40]. Yet another study reported promising results using H&N patients' pre-operative CT radiomics features to predict metastasis, with $\%Acc = 78\%$ and $\%Acc = 74\%$, and extranodal extension with $\%Acc = 80\%$ and $\%Acc = 70\%$, for the model compared with experienced radiologists, respectively [41]. Previously, ML models trained using radiomics features mined from pre-treatment LN QUS parametric maps for the same patient cohort studied in this investigation were found to predict binary treatment outcomes with $\%S_n = 81\%$ and $\%S_p = 76\%$ using a seven-feature multivariable model trained using the SVM classifier [28]. Whereas CT and MRI radiomics studies mine features from the images directly, it is worth bringing attention to the distinction that QUS radiomics studies utilize raw RF data rather than processing of B-mode images and pixel intensities [42]. Using the sliding window technique, RF data can be converted into a QUS spectrum from which various parameters can be extracted [43]. In the QUS study, parametric maps were created for various quantitative US spectral parameters, before radiomics features were determined for model building. QUS spectral parameters have proven useful for characterizing cellular conditions such as apoptosis [44,45]. Similarly, for the same patient cohort, radiomics features were mined from treatment-planning CT LN segmentations and used to train ML classifiers to create predictive models, resulting in $\%BA = 71\%$ for a six-feature SVM classifier model trained using only 1LTFs [29].

This study is unique in that the potential for MRI deeper-layer features was explored in order to study the heterogeneity of features with a resolution of 0.5 mm. After building the models trained using 1LTFs, the DTA method was initiated, and 1LTF maps were made of the features selected for in each classifier's best model. Subsequently, 2LTF radiomics features were mined from the 1LTF maps. S_2 feature sets were comprised of newly determined 2LTFs concatenated with the retained 1LTFs selected for in the first step. The best k -NN model trained using the $S_{2,KNN}$ feature set (four-feature multivariable model) outperformed any of the models trained using the S_1 feature set, with $\%BA = 80\%$. Similarly, the best FLD model trained using the $S_{2,FLD}$ feature set (six-feature multivariable model) outperformed any of the models trained using the S_1 feature set, with $\%BA = 80\%$. The best $S_{2,SVM}$ -trained SVM model matched but did not improve on the best- S_1 trained SVM model, with $\%BA = 77\%$.

Another iteration of the DTA method was carried out to mine 3LTFs, and again, the best-performing k -NN and FLD models trained using the $S_{3,KNN}$ and $S_{3,FLD}$ feature sets bested the top performances of models trained using the $S_{2,KNN}$ and $S_{2,FLD}$ feature sets, with $\%BA = 83\%$ and $\%BA = 81\%$, respectively. The nine-feature multivariable k -NN model trained using the $S_{3,KNN}$ feature set was the best model in the study. The nine selected features were a combination of two 1LTFs, two 2LTFs, and five 3LTFs.

Although the best performance of SVM models did not improve with the inclusion of 2LTFs and 3LTFs, it is worth emphasizing that it did not degrade either. This is theoretically consistent, mainly because at each iteration of DTA, the best features from the previous iteration are retained, while all other features are discarded. The principle of DTA works such that after the introduction of deeper-layer features, two outcomes are that (i) the models improve due to valuable information gain from new features, or (ii) the new features are not providing additive benefit, at which point performance should match the previous iteration due to retention of the previously selected features, as was the case in this study.

Next, for each classifier, to inspect the effect of including deeper-layer features on all the models, in each iteration, average performance metrics were evaluated for the 10 models created. As seen in Table 5, average SVM classifier models did not benefit from the inclusion of 2LTFs or 3LTFs, as none of the average performance metrics changed in a statistically significant manner ($p > 0.05$). However, once again, it is worth emphasizing that average performances did not degrade in a statistically significant manner either, which is to be expected due to retention of the best features from the previous iteration. On the other hand, k -NN and FLD classifier models exhibited statistically significant improvement

on several metrics, as shown in Tables 6 and 7 and visually represented in Figures 5 and 6. As previously mentioned, the DTA feature enhancement methodology was evaluated for this patient cohort with QUS [28] and CT [29] radiomics as well. In the QUS study, models trained with the inclusion of 2LTFs improved the predictive capacity of the seven-feature multivariable model up to $\%S_n = 85\%$ and $\%S_p = 80\%$, a marked improvement from the seven-feature multivariable model trained solely on 1LTFs, with $\%S_n = 81\%$ and $\%S_p = 76\%$ [28]. Similarly, DTA methodology enhanced the performance of predictive models trained using treatment-planning CT LN segmentations, with the best model performance resulting from inclusion of 2LTFs and 3LTFs, with a seven-feature model SVM classifier model demonstrating $\%S_n = 76\%$ and $\%S_p = 64\%$ [29]. It is worth mentioning that in both of these studies, DTA methodology was carried out for the five-feature model in each iteration, which was a decision made mainly to keep computation time reasonable and not due to any loss function [28,29]. In this study, 1–10 feature multivariable models were created for 1–10 features, and the model with the top BA was selected to proceed for DTA. In the case of multiple models exhibiting the same top BA, the model with fewer features was selected. To our knowledge, this is the first time that DTA methodology was implemented on MRI images.

Computing radiomics features in 2D slices rather than 3D voxels and the use of a small sample size are two factors that reduce the generalizability of the models. Another shortfall involves one necessary step for mining radiomics features, which is an initial discretization of image pixel intensities through binning, which was not optimized in this study and is worth consideration [33]. Similarly, in this study, the window size for calculating the feature maps was 3×3 pixels, which was the smallest window size possible, but future studies could investigate optimizing window size as well [33]. Finally, a clinically feasible model would likely maximize all available information and include clinical and genomics features in addition to radiomics features. We acknowledge that HPV-related carcinomas may have characteristic inhomogeneities as assessed by MRI, which are absent in non-HPC cancers, and that in this work, the unknown HPV status may silently be reflected in radiomic features. The hope is that radiomics features may capture and characterize heterogeneous patterns from involved LNs that impact response to treatment, even absent of knowledge about HPV status. In this study, clinical features were not consistently available in institutional database patient notes and, as such, were not included.

5. Conclusions

A few key findings in this study include the following: (i) demonstration of the index LNs of H&N cancer patients may exhibit a priori phenotypic MRI characteristics that can be quantified with radiomics and provide insights regarding treatment efficacy; (ii) when radiomics features were used to train machine learning classifiers, SVM, k -NN, and FLD models performed with balanced accuracy of 77.3%, 77.0%, and 76%, respectively (see Table 2); (iii) statistically significant improvement was demonstrated for average k -NN and FLD models, with the best models showing 83% and 81% balanced accuracy (see Table 4); (iv) even though average SVM classifier models did not improve significantly, notably, average performance did not decrease either, which is consistent with expectations; (v) it was found that predictive models trained with MRI radiomics demonstrated superior performance to models trained with QUS and CT radiomics for the same cohort of patients in previous works [28,29]. To our knowledge, this is the first time DTA was explored for MRI radiomics, and results suggest that future radiomics investigations should consider the proposed DTA method to enhance discriminating between populations of patients. Accurate and robust predictive models can provide improved cancer care for patients and provide clinicians with additional tools to guide decision making.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/radiation4020015/s1>, Table S1: Anonymized clinical characteristics of interest as reported in patient notes from institutional database.

Author Contributions: Conceptualization, A.S., A.M. and D.M.-P.; methodology, A.S.; software, L.S.; formal analysis, A.S.; investigation, G.J.C.; data curation, A.S., D.M.-P., I.P., I.K. and A.B.; writing—original draft preparation, A.S.; writing—review and editing, A.M., D.M.-P. and G.J.C.; visualization, A.S.; supervision, A.P.-M. and G.J.C.; project administration, G.J.C.; funding acquisition, G.J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC), the Terry Fox Research Institute (TFRI)/Lotte and John Hecht Foundation (project #1115), and the Canadian Institute of Health Research (CIHR) (project #162136) Program Project. Terry Fox Foundation New Frontiers Program Project Grant in Ultrasound and MRI for Cancer Therapy. The funding agencies had no role in the study design, study methodology, study results, or in the preparation of the manuscript.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Research Institute Ethics Board (SUN-3047) at Sunnybrook Health Sciences Center (Toronto, ON, Canada).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data can be made available upon request (Contact Czarnota Lab at Sunnybrook Health Sciences Center).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Mattson, M.P. Superior pattern processing is the essence of the evolved human brain. *Front. Neurosci.* **2014**, *8*, 265. [[CrossRef](#)] [[PubMed](#)]
2. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. [[CrossRef](#)]
3. Chu, A.; Sehgal, C.; Greenleaf, J. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognit. Lett.* **1990**, *11*, 415–419. [[CrossRef](#)]
4. Thibault, G.; Fertil, B.; Navarro, C.; Pereira, S.; Cau, P.; Levy, N.; Sequeira, J.; Mari, J.-L. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. *Int. J. Pattern Recognit. Artif. Intell.* **2013**, *27*, 1357002.
5. Sun, C.; Wee, W.G. Neighboring gray level dependence matrix for texture classification. *Comput. Vis. Graph. Image Process.* **1983**, *23*, 341–352. [[CrossRef](#)]
6. Dagogo-Jack, I.; Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 81–94. [[CrossRef](#)] [[PubMed](#)]
7. Lin, G.; Keshari, K.R.; Park, J.M. Cancer metabolism and tumor heterogeneity: Imaging perspectives using MR imaging and spectroscopy. *Contrast Media Mol. Imaging* **2017**, *2017*, 1–18. [[CrossRef](#)] [[PubMed](#)]
8. Denison, T.A.; Bae, Y.H. Tumor heterogeneity and its implication for drug delivery. *J. Control. Release* **2012**, *164*, 187–191. [[CrossRef](#)]
9. Ganeshan, B.; Miles, K.A. Quantifying tumour heterogeneity with CT. *Cancer Imaging* **2013**, *13*, 140–149. [[CrossRef](#)] [[PubMed](#)]
10. Brock, K.K.; Mutic, S.; McNutt, T.R.; Li, H.; Kessler, M.L. Use of Image Registration and Fusion Algorithms and Techniques in Radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med. Phys.* **2017**, *44*, e43–e76. [[CrossRef](#)]
11. Shur, J.D.; Doran, S.J.; Kumar, S.; ap Dafydd, D.; Downey, K.; O’connor, J.P.B.; Papanikolaou, N.; Messiou, C.; Koh, D.-M.; Orton, M.R. Radiomics in oncology: A practical guide. *Radiographics* **2021**, *41*, 1717–1732. [[CrossRef](#)] [[PubMed](#)]
12. Johnson, D.E.; Burtneis, B.; Leemans, C.R.; Lui, V.W.Y.; Bauman, J.E.; Grandis, J.R. Head and neck squamous cell carcinoma. *Nat. Rev. Dis. Primers* **2020**, *6*, 92. [[CrossRef](#)] [[PubMed](#)]
13. Skarsgard, D.P.; Groome, P.A.; Mackillop, W.J.; Zhou, S.; Rothwell, D.; Dixon, P.F.; O’Sullivan, B.; Hall, S.F.; Holowaty, E.J. Cancers of the upper aerodigestive tract in Ontario, Canada, and the United States. *Cancer* **2000**, *88*, 1728–1738. [[CrossRef](#)]
14. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
15. Vineis, P.; Alavanja, M.; Buffler, P.; Fontham, E.; Franceschi, S.; Gao, Y.T.; Gupta, P.C.; Hackshaw, A.; Matos, E.; Samet, J.; et al. Tobacco and cancer: Recent epidemiological evidence. *J. Natl. Cancer Inst.* **2004**, *96*, 99–106. [[CrossRef](#)] [[PubMed](#)]
16. Blot, W.J.; McLaughlin, J.K.; Winn, D.M.; Austin, D.F.; Greenberg, R.S.; Preston-Martin, S.; Bernstein, L.; Schoenberg, J.B.; Stemhagen, A.; Fraumeni, J.F. Smoking and drinking in relation to oral and pharyngeal cancer. *Cancer Res.* **1988**, *48*, 3282–3287. [[PubMed](#)]
17. Hashibe, M.; Brennan, P.; Benhamou, S.; Castellsague, X.; Chen, C.; Curado, M.P.; Dal Maso, L.; Daudt, A.W.; Fabianova, E.; Wunsch-Filho, V.; et al. Alcohol Drinking in Never Users of Tobacco, Cigarette Smoking in Never Drinkers, and the Risk of Head and Neck Cancer: Pooled Analysis in the International Head and Neck Cancer Epidemiology Consortium. *J. Natl. Cancer Inst.* **2007**, *99*, 777–789. [[CrossRef](#)] [[PubMed](#)]

18. Barsouk, A.; Aluru, J.S.; Rawla, P.; Saginala, K.; Barsouk, A. Epidemiology, Risk Factors, and Prevention of Head and Neck Squamous Cell Carcinoma. *Med. Sci.* **2023**, *11*, 42. [CrossRef] [PubMed]
19. Gasco, M.; Crook, T. The p53 network in head and neck cancer. *Oral Oncol.* **2003**, *39*, 222–231. [CrossRef]
20. Bryant, A.K.; Sojourner, E.J.; Vitzthum, L.K.; Zakeri, K.; Shen, H.; Nguyen, C.; Murphy, J.D.; Califano, J.A.; Cohen, E.E.W.; Mell, L.K. Prognostic role of p16 in nonoropharyngeal head and neck cancer. *J. Natl. Cancer Inst.* **2018**, *110*, 1393–1399. [CrossRef]
21. Gormley, M.; Creaney, G.; Schache, A.; Ingarfield, K.; Conway, D.I. Reviewing the epidemiology of head and neck cancer: Definitions, trends and risk factors. *Br. Dent. J.* **2022**, *233*, 780–786. [CrossRef] [PubMed]
22. Cooper, J.S.; Porter, K.; Mallin, K.; Hoffman, H.T.; Weber, R.S.; Ang, K.K.; Gay, E.G.; Langer, C.J. National Cancer Database Report on Cancer of The Head and Neck: 10-Year Update. *Head Neck* **2009**, *31*, 748–758. [CrossRef] [PubMed]
23. Nishimura, N.Y.; Aoki, K.; Koyama, S.; Nishio, M.; Otsuka, T.; Miyazaki, M.; Yoshii, T.; Otozai, S.; Miyabe, J.; Korematsu, M.; et al. The impact of COVID-19 pandemic on head and neck cancer diagnosis and treatment. *J. Dent. Sci.* **2023**, *23*, 1–5. [CrossRef]
24. Chau, L.; McNiven, A.; Arjune, B.; Bracken, G.; Drever, L.; Fleck, A.; Grimard, L.; Poon, I.; Provost, D. Dose Objectives for Head and Neck IMRT Treatment Planning Recommendation Report. 2014. Available online: https://www.cancercareontario.ca/sites/ccocancercare/files/guidelines/full/DoseObj_HN_IMRT_TrtmtPIngRec_0.pdf (accessed on 15 March 2023).
25. Anderson, G.; Ebadi, M.; Vo, K.; Novak, J.; Govindarajan, A.; Amini, A. An updated review on head and neck cancer treatment with radiation therapy. *Cancers* **2021**, *13*, 4912. [CrossRef] [PubMed]
26. Taylor, A.; Powell, M.E.B. Intensity-modulated radiotherapy—What is it? *Cancer Imaging* **2004**, *4*, 68–73. [CrossRef] [PubMed]
27. Teoh, M.; Clark, C.H.; Wood, K.; Whitaker, S.; Nisbet, A. Volumetric modulated arc therapy: A review of current literature and clinical use in practice. *Br. J. Radiol.* **2011**, *84*, 967–996. [CrossRef] [PubMed]
28. Safakish, A.; Sannachi, L.; DiCenzo, D.; Kolios, C.; Pejovic-Milic, A.; Czarnota, G.J. Predicting Head & Neck Cancer Treatment Outcomes with Pre-Treatment Quantitative Ultrasound Texture Features & Optimizing Machine Learning Classifiers with Texture-of-Texture Features. *Front. Oncol.* **2023**, *13*, 1258970. [PubMed]
29. Safakish, A.; Sannachi, L.; Moslemi, A.; Pejovic-Milic, A.; Czarnota, G.J. Deep Texture Analysis—Enhancing CT Radiomics Features for Prediction of Head and Neck Cancer Treatment Outcomes: A Machine Learning Approach. *Radiation* **2024**, *4*, 50–68. [CrossRef]
30. Moore-Palhares, D.; Ho, L.; Lu, L.; Chugh, B.; Vesprini, D.; Karam, I.; Soliman, H.; Symons, S.; Leung, E.; Loblaw, A.; et al. Clinical implementation of magnetic resonance imaging simulation for radiation oncology planning: 5 year experience. *Radiat Oncol.* **2023**, *18*, 27. [CrossRef]
31. Eisenhauer, E.A.; Therasse, P.; Bogaerts, J.; Schwartz, L.H.; Sargent, D.; Ford, R.; Dancey, J.; Arbuck, S.; Gwyther, S.; Mooney, M.; et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur. J. Cancer* **2009**, *45*, 228–247. [CrossRef]
32. Yushkevich, P.A.; Piven, J.; Hazlett, H.C.; Smith, R.G.; Ho, S.; Gee, J.C.; Gerig, G. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* **2006**, *31*, 1116–1128.
33. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef] [PubMed]
34. Elreedy, D.; Atiya, A.F. A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Inf. Sci.* **2019**, *505*, 32–64. [CrossRef]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 5999–6009.
36. Avanzo, M.; Wei, L.; Stancanello, J.; Vallières, M.; Rao, A.; Morin, O.; Mattonen, S.A.; El Naqa, I. Machine and deep learning methods for radiomics. *Med. Phys.* **2020**, *47*, e185–e202. [CrossRef]
37. Liu, Y.-Q.; Wang, X.-L.; He, D.-H.; Cheng, Y.-X. Protection against chemotherapy- and radiotherapy-induced side effects: A review based on the mechanisms and therapeutic opportunities of phytochemicals. *Phytomedicine* **2021**, *80*, 153402. [CrossRef]
38. Rocha, P.H.P.; Reali, R.M.; Decnop, M.; Souza, S.A.; Teixeira, L.A.B.; Lucas Júnior, A.; Sarpi, M.O.; Cintra, M.B.; Pinho, M.C.; Garcia, M.R.T. Adverse Radiation Therapy Effects in the Treatment of Head and Neck Tumors. *RadioGraphics* **2022**, *42*, 806–821. [CrossRef]
39. Niu, J.; Zhang, S.; Ma, S.; Diao, J.; Zhou, W.; Tian, J.; Zang, Y.; Jia, W. Preoperative prediction of cavernous sinus invasion by pituitary adenomas using a radiomics method based on magnetic resonance images. *Eur. Radiol.* **2019**, *29*, 1625–1634. [CrossRef]
40. Tang, S.; Ou, J.; Wu, Y.P.; Li, R.; Chen, T.W.; Zhang, X.M. Contrast-enhanced CT radiomics features to predict recurrence of locally advanced oesophageal squamous cell cancer within 2 years after trimodal therapy A case-control study. *Medicine* **2021**, *100*, e26557. [CrossRef]
41. Huang, T.-T.; Lin, Y.-C.; Yen, C.-H.; Lan, J.; Yu, C.-C.; Lin, W.-C.; Chen, Y.-S.; Wang, C.-K.; Huang, E.-Y.; Ho, S.-Y. Prediction of extranodal extension in head and neck squamous cell carcinoma by CT images using an evolutionary learning model. *Cancer Imaging* **2023**, *23*, 84. [CrossRef]
42. Lizzi, F.L.; Coleman, J.; Greenebaum, M.; Feleppa, E.J.; Elbaum, M. Theoretical framework for spectrum analysis in ultrasonic tissue characterization. *J. Acoust. Soc. Am.* **1983**, *73*, 1366–1373. [CrossRef]
43. Lizzi, F.L.; Astor, M.; Feleppa, E.J.; Shao, M.; Kalisz, A. Statistical framework for ultrasonic spectral parameter imaging. *Ultrasound Med. Biol.* **1997**, *23*, 1371–1382. [CrossRef] [PubMed]

44. Sadeghi-Naini, A.; Papanicolau, N.; Falou, O.; Tadayyon, H.; Lee, J.; Zubovits, J.; Sadeghian, A.; Karshafian, R.; Al-Mahrouki, A.; Giles, A.; et al. Low-frequency quantitative ultrasound imaging of cell death in vivo. *Med. Phys.* **2013**, *40*, 082901. [[CrossRef](#)] [[PubMed](#)]
45. Kolios, M.C.; Czarnota, G.J.; Lee, M.; Hunt, J.W.; Sherar, M.D. Ultrasonic spectral parameter characterization of apoptosis. *Ultrasound Med. Biol.* **2002**, *28*, 589–597. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.