Abstract

Purpose:

There is increasing interest in the use of radiomic models for non-invasive evaluation of head and neck squamous cell carcinoma (HNSCC) for prediction of various clinical, pathological, or molecular endpoints. In order to increase patient numbers, radiomic studies frequently combine HNSCC from different primary sites, with the implicit assumption that the tumor texture features are not site dependent. In this study, we investigated whether there are statistical differences in texture features from different sites and evaluated whether site-specific variations in texture features can affect the performance of machine learning models based on those texture features.

Materials and Methods:

603 contrast enhanced neck CT scans were evaluated from patients diagnosed with head and neck squamous cell carcinoma (HNSCC). Inclusion criteria were patient primary untreated tumors arising in the larynx or hypopharynx (LHP), lip & oral cavity (OC), and oropharynx (OP), further stratified based on HPV status to avoid confounding effects of HPV status. Tumor texture features were extracted from each image and used in conjunction with patient age, smoking status, drinking status, tumor T-stage to construct prediction models for predicting nodal status and the presence of lymphovascular invasion (LVI) and perineural invasion (PNI). Statistical analysis was performed using Wilks test and Roy’s largest root test to evaluate for variations in texture features based on tumor primary site. Two machine learning approaches (Random Forests (RF) and support vector machine (SVM)) were used for constructing prediction models. For constructing prediction models, the patients were randomly divided into a training set (70%) and an independent testing set (30%).

Results:

There were statistically significant differences (*P* < 0.05) between texture features of tumors arising in the OC, LHP, and OP. In order to evaluate whether the differences in texture features could affect prediction model performance, prediction models were constructed using texture data from the entire population or texture data stratified based on primary tumor site. Sub-stratification of texture data based on primary tumor site resulted in up to 14% improvement in prediction model performance compared to models combined using the combined datasets.

Conclusion:

There are differences in hand crafted texture features of HNSCC arising from different primary sites below the hard palate and these differences can impact the performance of radiomic models constructed based on those features. These results suggest that HNSCC texture features from different primary sites may not be equivalent. Therefore, for optimal performance and reliability, radiomic studies may have to stratify patients based on primary tumor site.

Clinical Relevance: Radiomic analysis can be used to predict various clinical endpoints of interest but these features can vary based on HNSCC primary site, and this should be taken into account in clinical investigations using radiomic analysis of HNSCC.

