HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2022 June 24.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2021 February ; 11597: . doi:10.1117/12.2582147.

Multiparametric Radiomics for Predicting the Aggressiveness of Papillary Thyroid Carcinoma Using Hyperspectral Images

Ka'Toria Edwards[1],

Martin Halicek[1],

James V. Little[2],

Amy Y. Chen[3],

Baowei Fei[1,4,*]

[1]Department of Bioengineering, University of Texas at Dallas, Richardson, TX

[2]Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA

[3]Department of Otolaryngology, Emory University School of Medicine, Atlanta, GA

[4]Department of Radiology, The University of Texas Southwestern Medical Center, Dallas, TX

Abstract

Papillary thyroid carcinoma (PTC) is primarily treated by surgical resection. During surgery, surgeons often need intraoperative frozen analysis and pathologic consultation in order to detect PTC. In some cases pathologists cannot determine if the tumor is aggressive until the operation has been completed. In this work, we have taken tumor classification a step further by determining the tumor aggressiveness of fresh surgical specimens. We employed hyperspectral imaging (HSI) in combination with multiparametric radiomic features to complete this task. The study cohort includes 72 *ex-vivo* tissue specimens from 44 patients with pathology-confirmed PTC. A total of 67 features were extracted from this data. Using machine learning classification methods, we were able to achieve an AUC of 0.85. Our study shows that hyperspectral imaging and multiparametric radiomic features could aid in the pathological detection of tumor aggressiveness using fresh surgical spemens obtained during surgery.

Keywords

Cancer; Radiomics; Hyperspectral Imaging; Tissues; Head; Neck; Tumor Aggression

* bfei@utdallas.edu, website: https://fei-lab.org.

Disclosures

The authors have no relevant financial interests in this article and no potential conflicts of interest to disclose. Informed consent was obtained from all patients in accordance with Emory Institutional Review Board policies under the Head and Neck Satellite Tissue Bank (HNSB, IRB00003208) protocol.

1. PURPOSE

Over the past 20 years, the rate of new thyroid cancer cases has increased 200% within the United States [1]. Papillary thyroid carcinoma (PTC) accounts for approximately 80% of these cases [2,3]. PTC is not only the most common form of thyroid cancer but also among the most malignant. Of individuals diagnosed with PTC, approximately one in four will not surpass the five-year survival mark [4]. This is due to PTC invading adjacent structures such as the lymphatic system [5]. Individuals who undergo tumor resection have recurrent disease up to 15% of the time [6], and 35% of individuals who suffer from recurrent disease die from the cancer [7]. Metastasis is a key characteristic of aggressive PTC, which makes early detection and classification critical. Tumor aggressiveness is routinely determined by intraoperative pathological evaluation of papillary thyroid tissue specimens via fine-needle biopsy in conjunction with pre-operative notes [8]. Up to 30% of fine-needle biopsies have inconclusive results, which leads to inefficient treatment of aggressive PTC tumors [6]. Inefficient treatment is partly due to the incomplete resection of tumors [6]. Partial resection is caused by the tumor expanding past what the physician determines to be the tumor boundary. Among other factors, this boundary issue must be mitigated during routine tumor resections to prevent the invasion of PTC in nearby structures.

In previous works, machine learning techniques, such as convolutional neural networks, have been implemented to detect head and neck cancer using hyperspectral imaging (HSI) [9–14]. Currently, there is limited research demonstrating PTC tumor aggression classification. A previous study was able to classify PTC on magnetic resonance images (MRI) with an AUC of 0.56, utilizing categorical features (gender, tumor size, presence of multiple lesions, etc.), as outlined in the pathological report. By adding radiomic features, they were able to obtain an AUC 0.92 [15]. One research group utilized similar techniques to determine tumor risk in head and neck cancer. This group was able to identify high-risk, or aggressive, tumors with an AUC of 0.86 [16]. A literature review indicates that this is the first work to present a novel tool for PTC tumor aggressiveness classification by implementing multiparametric radiomics based machine-learning classification method. To the best of our knowledge, this is the first work to investigate tumor aggressiveness classification of PTC utilizing radiomics on hyperspectral images.

2. METHODS

2.1 Hyperspectral Imaging

A CRI Maestro HS System (Perkin Elmer Inc., Waltham Massachusetts) was used to acquire HSI data of the *ex-vivo* specimens. The HS System obtains images by performing spectral scanning from 450 to 900 nm using a Xenon light source in combination with a liquid crystal tunable filter (LCTF) with a 5 nm spectral resolution [8,17–21]. The image size of the HSI was 1040×1392×91 pixels (height × width × spectral bands) with a 25 μm per pixel spatial resolution. The scanning duration was approximately one minute per HSI.

2.2 Papillary Thyroid Carcinoma Tissue Database

For this study, 72 specimens were acquired from 44 patients undergoing routine resection of papillary thyroid tumors. We acquired the tissue after pathological analysis from the primary tumor as well as tissue located at the tumor margin, which contained both tumor and normal tissue. Tissues were categorized as tumor or tumor and normal, by a histopathological analysis performed by an experienced pathologist. Upon the exclusion of benign tumors, 39 tumor and 33 tumor and normal specimens were selected for this investigation. Specimens were stored on histological slides approximately 10×6×2 mm (height × width × depth) in size. In addition to the histological slides, pathology reports were made available after the redaction of personally identifiable information.

PTC tumor aggressiveness was determined according to the American Thyroid Association (ATA) 2015 Risk Stratification System for differentiating thyroid carcinomas [15, 22]. Aggressive tumors are those which can be identified as intermediate-risk or high-risk. PTC Tumors are assigned to the intermediate-risk or high-risk categories by having one or more of the following histopathological features: aggressive histology subtype (*e.g.*, tall cell, hobnail, columnar cell), vascular invasion, tumor capsular invasion, extra-thyroidal extension (ETE), regional metastases, or distant metastases. Non-aggressive, or low-risk, tumors were classified as not containing any aggressive features. Of the 44 patients, 35 were classified as aggressive, and nine were classified as non-aggressive. This corresponds to 58 aggressive and 14 non-aggressive tissue specimens. In total, there are four groups of tissues being analyzed in this study. Specimen are classified as aggressive tumor-normal interface tissue, non-aggressive tumor-nomal interface tissue, aggressive tumor tissue, and non-aggressive tumor tissue.

2.3 Radiomic Feature Extraction and Selection

Images were loaded into the PyRadiomics package where 120 features were extracted. Several of these features revealed constant values between specimen. As a result, we chose the 67 features that displayed uniqueness for this classification task [23]. These features included first order (15), shape-based (14), gray-devel dependence matrix (11), gray-level co-occurrence matrix (3), gray-level run length matrix (12), and gray-level size zone matrix (12). Features were pre-defined by the Imaging Biomarker Standardization Initiative (IBSI) [24].

After feature extraction, radiomic feature selection was performed to minimize problems associated with high dimensionality, allow the machine learning algorithms to train faster, and to improve the reproducibility of the results [25]. A total of six feature selection methods were implemented: analysis of variance (ANOVA), forward elimination, backward elimination, Pearson correlation, ridge regression, and least absolute shrinkage and selection operator (LASSO). These feature selection methods can be split into three groups: filter, wrapper, and embedded methods. Pearson correlation and ANOVA are filter methods which select features based on their correlation to tumor aggression, determined by statistical tests. The wrapper methods include forward and backward elimination. In wrapper methods, features are chosen by generating a subset that iteratively changes based on thresholding. This generates a dataset that consists of a combination of variables with the highest

predictive power. The embedded methods implemented in this study include LASSO and ridge regression. These methods work by using unique calculations that combine both wrapper and filter qualities.

2.4 Machine Learning

Eighteen machine learning algorithms were tested for each feature extraction method. The machine learning algorithms were implemented using the python scikit-learn package [24]. The 108 combinations of feature extraction and machine learning methods are shown in the heatmap in Figure 3. Classifiers were trained using an independent training set (N = 54 specimens). The predictive performance was evaluated based on an independent testing set (N = 18 specimens) using accuracy analysis. The independent testing set was randomly selected on a patient basis. To aid in reproducibility of these results, there was no patient overlap between the training and testing groups. Each method was optimized by adjusting the parameters as shown in Table 1 to obtain maximum accuracy.

3. RESULTS

PTC aggression was classified with an accuracy of 0.83 for patients in the testing group (Figure 3) along with an AUC of 0.85. The number of features outlined in Column three of Table 1 illustrates that the feature selection methods presented here produced similar results despite using different numbers of features. Fom this, we cannot determine the optimal feature selection method. Results were obtained through optimization of various parameters, as shown in the fourth column of Table 1. The number of true-positive, true-negative, false-positive, and false-negative predictions are also shown, where positive is aggressive, and negative is non-aggressive. Quadratic Discriminant Analysis (QDA) proved to be the most accurate machine learning classification tool in two of the top three classification algorithms.

Of the features selected, gray-level dependence matrix (GLDM) variance was unanimously chosen as an important feature to classify tumor aggression. This feature measures the variance, or amount of variability, found in an image. GLDM variance is calculated for the HSI by finding the GLDM as defined as a the number of connected voxels within a distance of one voxel, that are dependent on the center voxel for each HSI. Next, the variance for every voxel in the HSI is calculated. The summation of these variances result in one value that represents the variability within the HSI [23]. The GLDM variance average ± standard deviation for non-aggressive tumor-nomal interface tissue was 33.5 ± 9.53, aggressive tumor-normal interface tissue was 26.7 ± 9.97, non-aggressive tumor tissue was 30.8 ± 14.8, and aggressive tumor tissue was 22.5 ± 10.3. After performing a one-way ANOVA test, we found that there was a significant difference between each of these groups with a p-value < 0.03. The results of the follow up ad-hoc Student’s unpaired, two-tailed t-test assuming equal variances are show in Figure 3. There is a significant difference with a p-value < 0.01 between the GLDM variance of non-aggressive tumor-normal interface tissue and aggressive tumor tissue, and a significant difference between GLDM variance of non-aggressive tumor tissue and aggressive tumor tissue was found with p-value < 0.05.

4. DISCUSSION & CONCLUSION

Tumor aggression was independently classified with an accuracy of 0.83 and AUC of 0.85. Although various classification methods produced similar results, they utilized different numbers of features. This implies that the significance of the feature is more important than the number of features used. Again, the most significant feature across the three different methods was GLDM variance, a measure of image variability. As expected, the variability of an image is greater when the tumor-normal interface is present, rather than a tumor alone. This is demonstrated by the difference of the means displayed in Figure 4.

A limitation of this study, as with similar classification problems, is the limited data set. In the event that a larger data set was available, the models presented here would have likely performed better. In addition, the number of aggressive tissue specimen out-numbered the non-aggressive tissue specimen by a large margin. This data imbalance limited the success of the results presented here. Most of the models predicted that all specimen were aggressive, which results in the accuracy of 72% observed frequently in Figure 3. The rate of false-negative cases, or patients that were identified as non-aggressive when it was aggressive, was zero for forward elimination and LASSO, and one for backward elimination (FN column of Table 1). This demonstrates the conservative nature of the models presented here. It is better to check a tumor that is falsely classified as aggressive, rather than fail to treat a tumor that is falsely classified as non-aggressive. The sooner aggressive tumors are treated, the more likely the physician will be able to fully resect the tumor.

In the TP, TN, FP, FN columns of Table 1, where positive is aggressive and negative is non-aggressive, a closer look at the data reveals that the models correctly identify non-aggressive data on tumor-normal interface tiussue, but struggle when a non-aggressive image of tumor tissue was presented. This is not likely due to the data imbalance, since the non-aggressive images of tumor-normal interface tissue were able to achieve a high accuracy. From this, we can conclude that the GLDM variance operates better at the tumor-normal margin. This finding is applicable to the boundary issue of tissue resection cases, where physicians may not completely remove the tumor. The method presented here offers a tool that can identify the textural features that are unique to the tumor-normal boundary of aggressive or non-aggressive tumors. A p-value < 0.01 supports the significant difference of boundary texture observed between aggressive and non-aggressive of tumor-normal interface tissue. Using this information, we can prove that we have identified a numerical characteristic to describe ETE. ETE occurs when the tumor extends outside of the thyroid capsule and invades surrounding structures [22]. This phenomenon can be observed at the tumor-normal interface tissue and can therefore be characterized by the variance presented in this region. This is beneficial for patient treatment since most PTC related deaths occur from post-resection disease recurrence. As a result, the physician can be notified that the resection may not have encompassed the tumor boundary and tailor the PTC tumor treatment plan accordingly.

In conclusion, radiomics has proven capable of differentiating aggressive and non-aggressive PTC tumors with an AUC of 0.85 for patients in the testing group independent of pathological data. Together, the hyperspectral imaging methods presented along with

preliminary results of this work demonstrate the potential for such methods to be implemented as a tool to increase the efficiency and accuracy of pathologists performing PTC tumor classification on histological slides for treatment planning.

ACKNOWLEDGEMENTS

This research was supported in part by the U.S. National Institutes of Health (NIH) grants (R01CA156775, R01CA204254, R01HL140325, and R21CA231911) and by the Cancer Prevention and Research Institute of Texas (CPRIT) grant RP190588.

ABBREVIATIONS

SVC	Support-Vector Classifier
Log	Logarithmic
RF	Random Forest
KNN	K- Nearest Neighbor
GaussNB	Gaussian Naïve Bayes
DT	Decision Trees
XT	Extra Trees
ABC	AdaBoost Classifier
GB	Gradient Boosting
SGDC	Stochastic Gradient Descent
GPC	Gaussian Process Classification
BNB	Bernoulli Naïve Bayes
L_SVC	Linear Support-Vector Classifier
LDA	Linear Discriminant Analysis
QDA	Quadratic Discriminant Analysis
XBG	eXtreme Gradient Boosting
SVM	Support-Vector Machine
ANOVA	Analysis of Variance
Forward	Forward Elimination
Backward	Backward Elimination
LASSO	Least Absolute Shrinkage and Selection Operator
Pearson	Pearson Correlation

Ridge Reg Ridge Regression

REFERENCES

[1]. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975–2017, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April (2020).

[2]. Bikas A, Burman KD Epidemiology of Thyroid Cancer. In: Luster M, Duntas L, Wartofsky L (eds) The Thyroid and Its Diseases. Springer, Cham (2019).

[3]. Guo Zhenying MD*,†; Ge Minghua MD*,†; Chu Ying-Hsia MD‡; Asioli Sofia MD§; Lloyd Ricardo V. MD, PhD‡ Recent Advances in the Classification of Low-grade Papillary-like Thyroid Neoplasms and Aggressive Papillary Thyroid Carcinomas: Evolution of Diagnostic Criteria, Advances In Anatomic Pathology: July - Volume 25 - Issue 4 - p 263–272 doi: 10.1097/PAP.0000000000000198 (2018). [PubMed: 29762157]

[4]. American Cancer Society. Cancer Facts & Figures 2020. Atlanta: American Cancer Society; (2020).

[5]. Limaiem F, Rehman A, Mazzoni T. Cancer, Papillary Thyroid Carcinoma (PTC) [Updated 2020 Apr 21]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536943/(2020).

[6]. Bates MF, Lamas MR, Randle RW, Long KL, Pitt SC, Schneider DF, & Sippel RS Back so soon? Is early recurrence of papillary thyroid cancer really just persistent disease?. Surgery, 163(1), 118–123. 10.1016/j.surg.2017.05.028 (2018). [PubMed: 29128176]

[7]. Song Eyun, Jeon Min Ji, Oh Hye-Seon, Han Minkyu, Lee Yu-Mi, Kim Tae Yong, Chung Ki-Wook, Kim Won Bae, Shong Young Kee, Song Dong Eun, and Kim Won Gu. “Do aggressive variants of papillary thyroid carcinoma have worse clinical outcome than classic papillary thyroid carcinoma?”. European Journal of Endocrinology 179.3: 135–142. <10.1530/EJE-17-0991>. Web. 29 Jun. (2020).

[8]. Halicek Martin, Dormer James D., Little James V., Chen Amy Y., and Fei Baowei, “Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning,” Biomed. Opt. Express 11, 1383–1400 (2020) [PubMed: 32206417]

[9]. Ortega S, Halicek M, Fabelo H, Camacho R, Plaza MD, Godtliebsen F, M Callicó G, Fei BW (Corresponding author). Hyperspectral imaging for the detection of glioblastoma tumor cells in H&E slides using convolutional neural networks. Sensors; 20(7):1911.(2020).

[10]. Ma L, Halicek M, Fei BW. In vivo cancer detection in animal model using hyperspectral image classification with wavelet feature extraction. Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging; 11317(113171C). International Society for Optics and Photonics,(2020).

[11]. Ortega S, Halicek M, Fabelo H, Guerra R, Lopez C, Lejeune M, Godtliebsen F, Callico GM, Fei BW (Corresponding author). Hyperspectral imaging and deep learning for the detection of breast cancer cells in digitized histological images. Medical Imaging 2020: Digital Pathology; 11320(113200V). International Society for Optics and Photonics, (2020).

[12]. Ma L, Halicek M, Zhou X, Dormer J, Fei BW (Corresponding author). Hyperspectral microscopic imaging for automatic detection of head and neck squamous cell carcinoma using histologic image and machine learning. Medical Imaging 2020: Digital Pathology; 11320(113200W). International Society for Optics and Photonics, (2020).

[13]. Halicek M, Little JV, Wang X, Chen AY, Fei B. Optical biopsy of head and neck cancer using hyperspectral imaging and convolutional neural networks. J Biomed Opt. 2019 Mar;24(3):1–9. doi: 10.1117/1.JBO.24.3.036007. (2019).

[14]. Halicek M, Lu G, Little JV, Wang X, Patel M, Griffith CC, El-Deiry MW, Chen AY, Fei B. Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. J Biomed Opt. 2017 Jun 1;22(6):60503. doi: 10.1117/1.JBO.22.6.060503. (2017). [PubMed: 28655055]

[15]. Wang Hao, Song Bin, Ye Ningrong, Ren Jiliang, Sun Xilin, Dai Zedong, Zhang Yuan, Chen Bihong T., Machine learning-based multiparametric MRI radiomics for predicting the aggressiveness of papillary thyroid carcinoma, European Journal of Radiology, Volume 122,108755,ISSN 0720–048X, 10.1016/j.ejrad.2019.108755 (2020).

[16]. Vallières M, Kay-Rivest E, Perrin LJ et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 7, 10117. 10.1038/s41598-017-10371-5 (2017). [PubMed: 28860628]

[17]. Zhang Y, Wu X, He L, Meng C, Du S, Bao J, & Zheng Y (2020). Applications of hyperspectral imaging in the detection and diagnosis of solid tumors. Translational Cancer Research, 9(2), 1265–1277. doi:10.21037/tcr.2019.12.53 (2020) [PubMed: 35117471]

[18]. Halicek M, Fabelo H, Ortega S, Callico GM, Fei B. In-Vivo and Ex-Vivo Tissue Analysis through Hyperspectral Imaging Techniques: Revealing the Invisible Features of Cancer. Cancers (Basel).;11(6):756. Published 2019 May 30. doi:10.3390/cancers11060756 (2019).

[19]. Fei B, Lu G, Wang X, Zhang H, Little JV, Patel MR, Griffith CC, El-Diery MW, Chen AY. Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients. J Biomed Opt. 2017 Aug;22(8):1–7. doi: 10.1117/1.JBO.22.8.086009. (2017).

[20]. Lu G, Little JV, Wang X, Zhang H, Patel MR, Griffith CC, El-Deiry MW, Chen AY, Fei B. Detection of Head and Neck Cancer in Surgical Specimens Using Quantitative Hyperspectral Imaging. Clin Cancer Res. 2017 Sep 15;23(18):5426–5436. doi: 10.1158/10780432.CCR-17-0906. Epub 2017 Jun 13. (2017). [PubMed: 28611203]

[21]. Halicek M, Fabelo H, Ortega S, Little JV, Wang X, Chen AY, Callico GM, Myers L, Sumer BD, Fei B. Hyperspectral imaging for head and neck cancer detection: specular glare and variance of the tumor margin in surgical specimens. J Med Imaging (Bellingham). 2019 Jul;6(3):035004. doi: 10.1117/1.JMI.6.3.035004. Epub 2019 Sep 14. (2019). [PubMed: 31528662]

[22]. Haugen Bryan R., Alexander Erik K., Bible Keith C., Doherty Gerard M., Mandel Susan J., Nikiforov Yuri E., Pacini Furio, Randolph Gregory W., Sawka Anna M., Schlumberger Martin, Schuff Kathryn G., Sherman Steven I., Sosa Julie Ann, Steward David L., Tuttle R. Michael, and Wartofsky Leonard.Thyroid.Jan.1–133.10.1089/thy.2015.0020 (2016).

[23]. Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillon-Robin JC, Pieper S, Aerts HJWL (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research, 77(21), e104–e107. 10.1158/0008-5472.CAN-17-0339 <10.1158/0008-5472.CAN-17-0339>(2017). [PubMed: 29092951]

[24]. Zwanenburg A, Leger S, Vallières M, and Löck S (2016). Image biomarker standardisation initiative - feature definitions. In eprint arXiv:1612.07003 [cs.CV] (2016).

[25]. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825–2830, (2011).

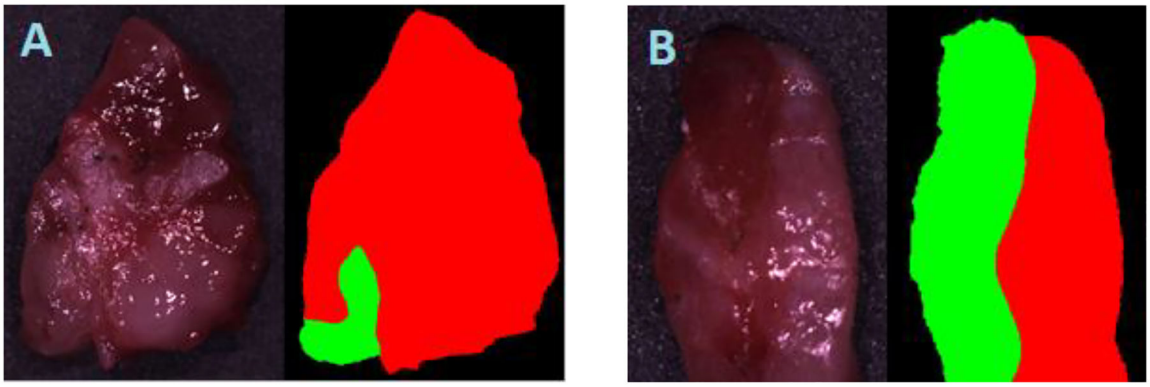

Figure 1.

Representative images of tumor-normal interface of a resected tissue specimen. (A) Aggressive tissue specimen exhibiting ETE. (B) Non-aggressive tissue specimen. The tumor-normal classification is indicated where red is tumor tissue and green is normal tissue.

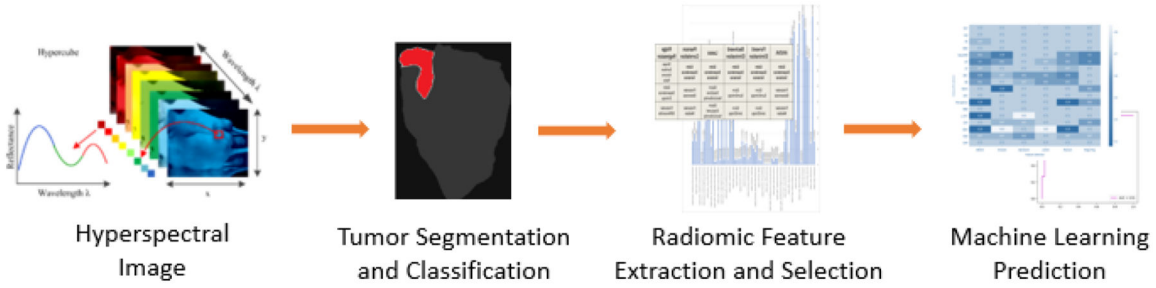

Figure 2.
Pipeline of radiomics analysis of papillary thyroid carcinoma on HSI.

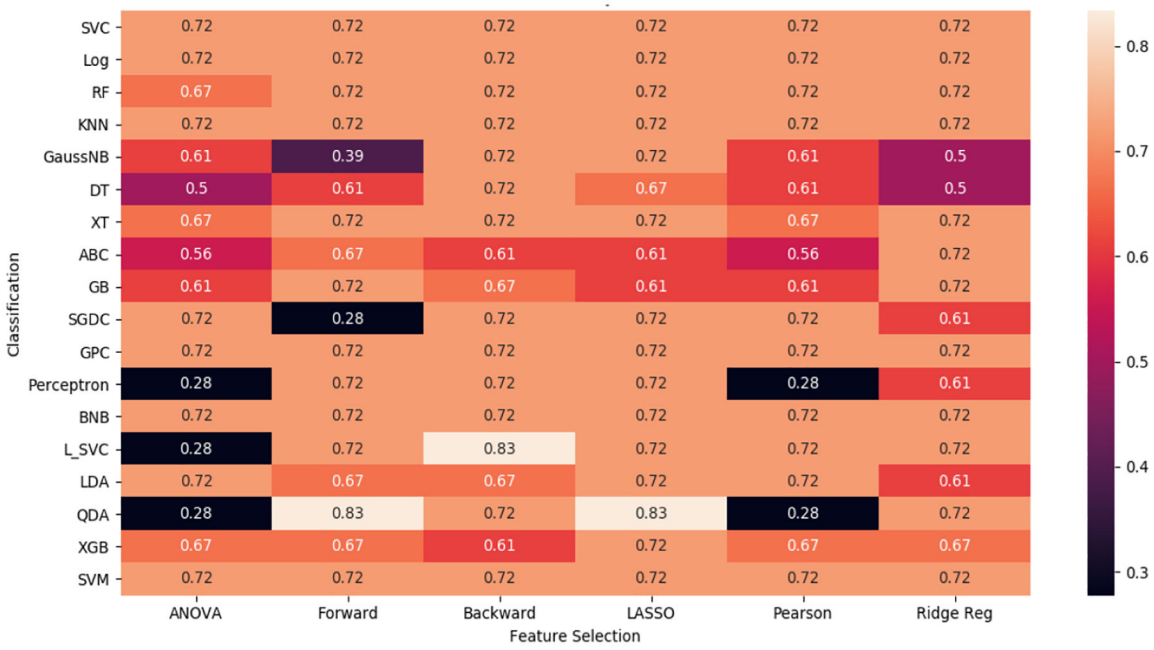

Figure 3.

Heatmap displaying embedded accuracy values for the testing group. The feature selection methods (column) and classification tools (row) with a maximum accuracy of 0.83 are shown. Acronyms are outline in the abbreviations section.

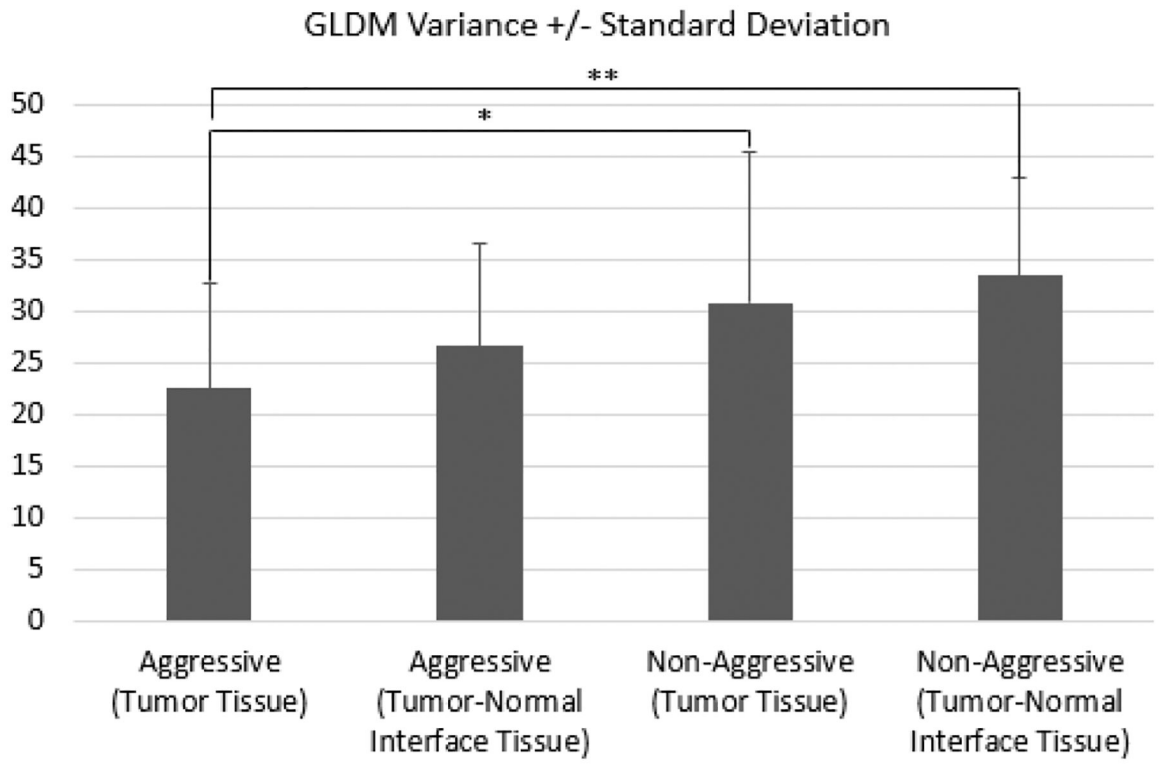

Figure 4.

Gray-level dependence matrix (GLDM) variance of tissue specimen classification based on tissue type. The significance between groups are denoted by asterisks, where ** $p < 0.01$, * $p < 0.05$.

Table 1.

The corresponding true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) values for feature selection and machine learning method combinations that resulted in 83.3% accuracy for the testing group.

Feature Selection	Machine Learning Classification	# of Features	Optimized Feature Selection Parameter	TP	TN	FP	FN
Forward Elimination	QDA	9	Threshold in = 0.3 Threshold out = 0.35	13	2	3	0
Backward Elimination	L_SVC	39	Threshold out = 0.35	12	3	2	1
LASSO	QDA	3	alpha = 0.1 threshold = |value| > 1E-3	13	2	3	0