JULY 12–16
2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

An empirical comparison of Weka classifiers for outcome prediction using an Imaging Habitats Definition and Feature Extraction method on MRI

Q. Han[1], **R. F. Palm**[2], **K. Latifi**[2], **E. G. Moros**[2], **A. O. Naghavi**[2] **and G. G. Zhang**[2]
1University of South Florida, Tampa, FL
2H. Lee Moffitt Cancer Center, Tampa, FL

MOFFITT CANCER CENTER
UNIVERSITY of SOUTH FLORIDA

INTRODUCTION

We propose an Imaging Habitats Definition and a Feature Extraction method from MRI images.

We hypothesize that utilizing Machine Learning (ML) and MRI images is possible to predict the outcome of cancer patients treated with radiation therapy (RT).[1,2]

We also present a practical method to build classification models from small and imbalanced patient datasets. The proposed method uses the Waikato Environment for Knowledge Analysis (Weka) employing simple graphical users interfaces (GUI) instead of coding with a programming language.[3] The method provides good prediction power with simple operations for medical researchers without a strong background in ML. We tested the method using radiomics features extracted using an Imaging Habitats Definition and a Feature Extraction method on pretreatment MRI images of Soft Tissue Sarcomas (STS) patients treated with neoadjuvant RT.

METHOD

Patients: A total of 97 STS patients were included, with 68 and 29 in the training and testing cohorts, respectively. Gross-tumor volumes (GTV) were manually segmented on pre-RT T1-post contrast and T2 STIR sequences.

Features: The habitats method dichotomized the MRI images by signal intensity to form 4 imaging habitats: T1 high-T2 high, T1 high-T2 low, T1 low-T2 high and T1 low-T2 low; 154 habitat features were extracted, and 11 clinical features recorded.

Theory: We hypothesize that features computed from medical images may represent biological information.[2] A group of carefully chosen features may therefore contain enough information to represent the status of a tumor and thereby allow to predict the tumor response to RT.[4,5,6] We also hypothesize that some classifiers are more efficient in solving a given problem than others.[7] Therefore, we develop a method to find representative features and preferred classifiers.

Method Design:

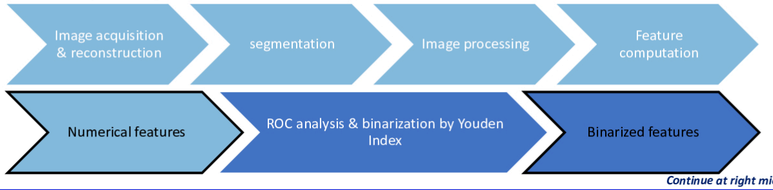
- Multi-step dimensionality reduction and Weka machine learning classifiers were utilized to identify habitat features predictive of pathological necrosis rates (PNR) >90% (35 in 97) or >95% (20 in 97) at the time of surgery.
- ROC analysis and binarization by Youden Index, Fisher's exact test, leave-one-out cross-validation were used to overcome the limitation of the small sample size.
- Weka Cost Sensitive Classifier and Wrapper Subset Evaluator were used to overcome the limitation of imbalanced datasets. More than 30 Weka Classifiers (Table 1) were tested in Wrapper Subset Evaluator (limited to 7 features) to select a group of representative features forming a radiomic signature.
- The group contained 14 features, including both habitat features and clinical features; 5 to 8 features were selected for each classifier.
- Five performance metrics (Accuracy, Area under receiver operative characteristic (AUROC), Area under precision-recall curve (AUPRC), F-Measure, Matthews correlation coefficient (MCC)) were used to evaluate the different properties of each model.

A group of representative features was selected. They are thought to contain more information relevant to our objectives.

A group of preferred Weka standard classifiers was selected for objective(s) of interest. They are thought to be more efficacious in making predictions.

Cost matrices were recorded for each prediction model that reduce the influence of the imbalanced datasets.

Feature Pre-processing Scheme:



Continue at right middle

RESULTS

We found that a combination of radiomics features and clinical features using a combination of Weka classifiers produced models with predictive powers. Furthermore, the selection of representative features is more beneficial than using AdaBoost or Bagging to improve a model. AdaBoost tended to be overfitting compare to Stochastic Gradient Descent (SGD). Support Vector Machine with SGD (SGD(SVM)) gave the best performance, Local Weighted Learning (LWL) was second and Multilayer Perceptron was third. The latter did not show superiority. (Table 2) The 8-feature models built with the SGD(SVM) and the representative features performed better overall. We also evaluated patients treated without prior neoadjuvant chemotherapy. (Table 3) We found models incorporating pretreatment MRI-habitats and clinical features can predict response to RT in STS.

Table 1: Some Weka classifiers

J48	AdaBoost(J48)	LogitBoost
JRip	AdaBoost(JRip)	Bagging(REPTree)
LMT	AdaBoost(LMT)	Bagging(LWL(DecisionStump))
RandomForest	AdaBoost(RandomForest)	Bagging(MultilayerPerceptron)
MultilayerPerceptron	AdaBoost(MultilayerPerceptron)	Bagging(SGD(SVM))
VotedPerceptron	AdaBoost(VotedPerceptron)	
SGD(SVM)	AdaBoost(SGD(SVM))	
SGD(logistic)	AdaBoost(SGD(logistic))	
LWL(DecisionStump)	AdaBoost(LWL(DecisionStump))	
LWL(logistic)	AdaBoost(LWL(logistic))	
LWL(J48)	AdaBoost(LWL(J48))	
LWL(JRip)	AdaBoost(LWL(JRip))	

Table 2: Selected 7-feature models' performance on Training and Testing subsets

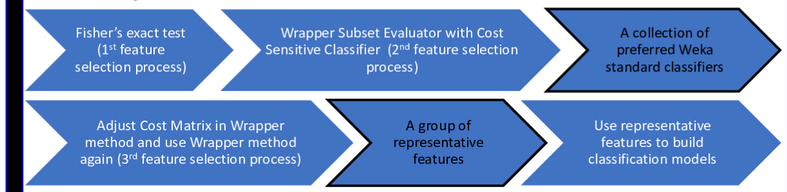
7-feature model	Weka Classifier	Training					Testing				
		Accuracy	ROC	PRC(1)	F-Measure	MCC	Accuracy	ROC	PRC(1)	F-Measure	MCC
PN90	MultilayerPerceptron	88%	0.949	0.925	0.882	0.747	52%	0.474	0.334	0.528	0.070
	SGD(SVM)	69%	0.689	0.503	0.696	0.367	65%	0.689	0.469	0.663	0.362
	LWL(DecisionStump)	69%	0.752	0.625	0.694	0.452	59%	0.516	0.384	0.595	0.216
	AdaBoost(SGD(SVM))	84%	0.874	0.812	0.841	0.674	48%	0.461	0.352	0.488	-0.119
	AdaBoost(LWL(DecisionStump))	76%	0.800	0.703	0.769	0.557	66%	0.582	0.411	0.663	0.362
	LogitBoost	75%	0.879	0.846	0.753	0.478	55%	0.511	0.405	0.562	0.075
PN95	MultilayerPerceptron	84%	0.893	0.762	0.836	0.492	62%	0.286	0.192	0.631	-0.089
	SGD(SVM)	84%	0.898	0.560	0.852	0.668	69%	0.620	0.270	0.710	0.209
	LWL(DecisionStump)	93%	0.829	0.714	0.923	0.764	62%	0.399	0.195	0.607	-0.233
	AdaBoost(SGD(SVM))	84%	0.907	0.690	0.840	0.519	69%	0.504	0.244	0.679	-0.008
	AdaBoost(LWL(DecisionStump))	85%	0.923	0.780	0.853	0.550	69%	0.330	0.167	0.647	-0.173
	LogitBoost	93%	0.959	0.863	0.927	0.782	52%	0.333	0.175	0.555	-0.191

Table 3: Performance of SGD(SVM) model of all patients and No prior neoadjucant chemotherapy patients.

	Class	Feature number	Weka Classifier	Training					Testing				
				Accuracy	ROC	PRC(1)	F-Measure	MCC	Accuracy	ROC	PRC(1)	F-Measure	MCC
All patients (n=97)	PN90	8	SGD(SVM)	72%	0.746	0.549	0.725	0.474	66%	0.713	0.484	0.658	0.418
	PN95	7	SGD(SVM)	84%	0.898	0.560	0.852	0.668	69%	0.620	0.270	0.710	0.209
No prior neoadjuvant chemotherapy (n=84)	PN90	7	SGD(SVM)	71%	0.732	0.558	0.715	0.455	60%	0.663	0.458	0.595	0.336
	PN95	8	SGD(SVM)	78%	0.721	0.392	0.787	0.412	72%	0.600	0.253	0.729	0.187

Continued

Model Building Scheme:



CONCLUSIONS

The approach we employed proved useful to build classification models with clinically useful prediction power when using a relatively small and imbalanced patient dataset. Radiomic features suggestive of habitat diversity within a tumor appear to be associated with RT outcome.

DISCUSSIONS

Many trade-offs were made in the development of this method. Only common standard Weka classifiers and default parameters were used for simplicity. The general approach, however, has the potential for building more predictive models for medical and non-medical applications with more advanced ML classifiers and optimized parameters.

The Wrapper method is flexible because users can select one or multiple configurations to adjust the priority of the performance metric or cost matrix. The user can select one or many classifiers in the 2nd feature selection process.

It is important to understand the definition/meaning of each feature after the 3rd feature selection process. Some features correlated. So it is possible to arrive at 20 representative features but only 5 of them can make a good model.

SGD(SVM), LWL and Multilayer Perceptron had good performance in this study. We will study these more for other objectives in the future.

REFERENCES

1. Naghavi, Arash O. et al. "The Future of Radiation Oncology in Soft Tissue Sarcoma." *Cancer Control : Journal of the Moffitt Cancer Center* vol. 25,1 1073274818815504. 11 Dec. 2018, doi:10.1177/1073274818815504
2. Gillies, Robert J et al. "Radiomics: Images Are More than Pictures, They Are Data." *Radiology* vol. 278,2 (2016): 563-77. doi:10.1148/radiol.2015151169
3. Witten, Ian H., et al. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Fourth Edition, 2016
4. Peeken, Jan C et al. "Tumor grading of soft tissue sarcomas using MRI-based radiomics." *EBioMedicine* vol. 48 (2019): 332-340. doi:10.1016/j.ebiom.2019.08.059
5. Peeken, Jan C et al. "CT-based radiomic features predict tumor grading and have prognostic value in patients with soft tissue sarcomas treated with neoadjuvant radiation therapy." *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* vol. 135 (2019): 187-196. doi:10.1016/j.radonc.2019.01.004
6. Spraker, Matthew B et al. "MRI Radiomic Features Are Independently Associated With Overall Survival in Soft Tissue Sarcoma." *Advances in radiation oncology* vol. 4,2 413-421. 23 Feb. 2019, doi:10.1016/j.adro.2019.02.003
7. Deist, Timo M et al. "Machine learning algorithms for outcome prediction in (chemo)radiotherapy: An empirical comparison of classifiers." *Medical physics* vol. 45,7 (2018): 3449-3459. doi:10.1002/mp.12967

CONTACT INFORMATION

Qixing Han
University of South Florida, TAMPA, FL 33620
EMail: qhan@mail.usf.edu