Identifying an optimal machine learning generated image marker to predict survival of gastric cancer patients

Huong Pham[a,*], Meredith A. Jones[b], Tiancheng Gai[a], Warid Islam[a], Gopichandh Danala[a], Javier Jo[a], Bin Zheng[a]

[a] School of Electrical and Computer Engineering, University of Oklahoma, Norman, OK 73019, USA

[b] Stephenson School of Biomedical Engineering, University of Oklahoma, Norman, OK 73019, USA

ABSTRACT

Computer-aided detection and/or diagnosis (CAD) schemes typically include machine learning classifiers trained using handcrafted features. The objective of this study is to investigate the feasibility of identifying and applying a new quantitative imaging marker to predict survival of gastric cancer patients. A retrospective dataset including CT images of 406 patients is assembled. Among them, 162 patients have more than 5-year survival. A CAD scheme is applied to segment gastric tumors depicted in multiple CT image slices. After gray-level normalization of each segmented tumor region to reduce image value fluctuation, we used a special feature selection library of a publicly available Pyradiomics software to compute 103 features. To identify an optimal approach to predict patient survival, we investigate two logistic regression model (LRM) generated imaging markers. The first one fuses image features computed from one CT slice and the second one fuses the weighted average image features computed from multiple CT slices. Two LRMs are trained and tested using a leave-one-case-out cross-validation method. Using the LRM-generated prediction scores, receiving operating characteristics (ROC) curves are computed and the area under ROC curve (AUC) is used as index to evaluate performance in predicting patients' survival. Study results show that the case prediction-based AUC values are 0.70 and 0.72 for two LRM-generated image markers fused with image features computed from a single CT slice and multiple CT slices, respectively. This study demonstrates that (1) radiomics features computed from CT images carry valuable discriminatory information to predict survival of gastric cancer patients and (2) fusion of quasi-3D image features yields higher prediction accuracy than using simple 2D image features.

Keywords: Computer aided diagnosis, machine learning, handcrafted features, principal component analysist, mammography, gastric cancer survival prediction, quasi-3D image features.

1. INTRODUCTION

Gastric cancer (GC) is one of the most commonly malignant tumors and the second leading cause of cancer related deaths worldwide [1]. Conventional treatment methods include surgery and follow-up chemotherapies. In order to develop more effective and personalized treatment plans and reduce patients' mortality rates, it is important to accurately predict aggressiveness of the malignant gastric tumors and likelihood of patients' survival at an early stage of cancer diagnosis and before surgery. Since computed tomography (CT) is the most popular imaging modality used to detect and diagnose gastric cancer in current clinical practice, we hypothesize that CT images contain valuable tumor phenotype information associated with tumor aggressiveness, which can be identified and extracted to help predict the likelihood of patients' overall survival (OS).

The scientific rationale to support this hypothesis comes from the well-tested radiomics concept that indicates some of radiomics features computed from radiographic images (i.e., CT images) highly associate with cancer prognosis [2] and applying computer-aided detection (CAD) schemes can more effectively extract clinically relevant radiomics features to build machine leaning models to predict cancer prognosis. For example, in our previous studies, we have developed CAD-supported machine learning models to predict progression-free survival of ovarian cancer patients [3] and risk of cancer recurrence of lung cancer patients after surgery [4].

CAD schemes are developed to classify suspicious lesions in previous studies. The goal of these techniques is to extract meaningful imaging markers or features to increase the performance. In previous studies, many quantitative and qualitative predictors were used to predict the survival of gastric cancer individuals. Used independent variables in these studies [5, 6] are age, gender, smoking, surgical history, place of living, etc. Not many studies have used radiomic features extracted from CT scan images to predict survival rate. Our study in this paper will assess only radiomic features as the potential markers for survival rate assumed that no other clinical information such as age, gender, surgical history, and place of living are available. In this study [7], radiomic features combined with other clinical variables can be used to predict splenic hilar lymph node metastasis in advanced proximal gastric cancer patients. The common approach of these CAD schemes will extract features of one single slice (2D-CT scan) to classify the lesions while their shapes in nature have three dimensions (3D). Even though 3D objects have more information than one single slice of the object, marking more slices of the lesions could introduce more noise into the data.

To investigate this issue and discover a better CAD scheme for predicting not only the survival rate but also other factors in advanced gastric cancer, we hypothesize that quasi-3D image features extracted and computed from the multiple tumor related CT image slices can yield better results than conventional method using single slice feature extraction. Hence, the objective of this study is to validate our hypothesis and develop a new CAD scheme to segment tumors, compute and select effective radiomics features, and identify an optimal approach to train or build multi-feature fusion-based machine learning models to predict likelihood of survival of gastric cancer patients using a relatively large and diverse clinical dataset.

2. MATERIALS AND METHODS

2.1 Image Dataset

In this study, a retrospective dataset of gastric cancer patients is assembled and used. The details of this image dataset including demographic and clinical information of patients, as well as CT image scanning protocols and characteristics have been reported in previous study [7]. In brief, this image dataset includes abdominal CT perfusion images acquired from 406 patients with the pathologically confirmed gastric cancer. Three clinicopathological findings or markers yielded from the tumor resection specimen after surgery were also retrospectively collected for all these patients. For each case, based on RECIST guidelines used in current clinical practice, the radiologists have identified and draw a single region of interest (ROI) on the transverse image section of a tumor that depicts the maximum tumor size (or diameter). We use the radiologists' annotated CT image slices as ground truth to identify the locations of the gastric tumors. Other clinical diagnostic results and treatment outcomes are also recorded and collected in this dataset. Based on the 5-year survival data, we divide these 406 patients into two groups of Yes (survival) and No (non-survival). Patients' follow-up data show that among these patients, 168 survived and 238 did not survive. Following steps are applied to develop and test the new CAD and machine learning supported quantitative image markers to predict likelihood of patients' survival in this study.

2.2 Image Preprocessing

We first design and implement a graphic user interface (GUI) to upload and display CT images of each patient. Based on radiologists' annotated tumor region in one CT slice, we place a tumor segmentation seed into the center of the tumor. From this seed, an adaptive multi-layer topographic region growing algorithm is applied to segment tumor region on this CT slice across the central region of the tumor. This algorithm has been used to segment different lesions depicting on different medical images in our previous studies (i.e., [8, 9]). In order to reduce the potential errors in computing radiomics image features from the tumors, the segmented tumor is visually reviewed and manually edited if necessary. Then, the algorithm is continued being applied to adjacent CT slices in two directions (up and down) to segment tumor regions depicting on each CT slice until no tumor region is detected. Figure 1 shows examples of one gastric tumor segmentation results in 3 adjacent CT image slices in which the segmented tumor boundary contours are marked in red color.

2.3 Feature Extraction:

The gray-level normalization of each segmented tumor region is first performed using the limitation of dynamics to $\mu \pm 3s$ (μ, gray-level mean; and s, gray-level standard deviation) to minimize the impact of contrast and brightness variation, which may otherwise over curtain the true image texture. Next, to compute radiomics features from each segmented tumor region, we use a publicly available libraries Pyradiomics [10]. It extracts radiomics features from medical images and provides research community a reference standard for radiomics analysis tasks. Pyradiomics can initially compute up to more than 1,500 features. In this study, we compute 103 radiomics features, which are computed from the first order statistics, shape factors, gray level concurrence matrix, gray level run length matrix, neighboring gray tone difference matrix, and gray level dependence matrix.

Figure 1: Example of illustrating the segmented tumor regions on three adjacent CT image slices of one tumor.

Specifically, we computed large number of the first order statistics, which include energy, entropy (defined Image Biomarker Standardization Initiative – ISBS), minimum, maximum, mean, standard deviation, skewness, Kurtosis, and uniformity. Shape 2D features include mesh surface, pixel surface, perimeter, perimeter to surface ratio, sphericity, spherical disproportion, maximum 2D diameter, major axis length, elongation. Details of gray level concurrence matrix, gray level run length matrix, neighboring gray tone difference matrix, and gray level dependence matrix can be found in the documentation published in [10].

There are two computational approaches in this paper to create the extracted features data. In the first approach, the features are extracted from the segmentation of a single slice of tumor marked by radiologist while in the second approach, all adjacent slices of the marked tumor also collected and extracted as a second data set. In the second approach, all 2D features of multiple slices of the same case were weighted based on the "Pixel Surface" or area feature and combined as:

$$F_i^k = \sum_{i=1}^{N} w_i \times F_i$$

where w_i is the ratio of the segmented tumor area on a *i th* slice to the total tumor area segmented on all N involved CT image slices. Finally, all computed radiomics image feature values are normalized between 0 to 1 to reduce case-based reliance.

2.4 Feature dimensionality reduction using Principal Component Analysist algorithm:

One of the challenging tasks in analyzing high-dimensional data in cancer research is to reduce the number of extracted features and create relevant important variables. Principal Component Analysist (PCA) is a common tool used to find new patent in high-dimensional data and collect the most important features that has the largest variance and diffusion [11]. PCA was applied on the first approach data to reduce high-dimensional features from 103 to 11 dependent variables while retains 95% variance of the original data set. The same process was applied on the data set of the second approach to reduce the feature from 103 to 7 dependent variables.

2.5 Classification:

We combine above computed radiomics features to build machine learning model or classifier to predict the likelihood of gastric cancer patients' survival. Although many different types of machine learning models have been investigated and developed in many previous CAD of medical images studies, as a prove-of-concept study, we select a few most common machine learning models namely, logistic regression model (LRM), artificial neurol network (ANN), and Support Vector Machine (SVM). Next, performance of these models will be compared to choose the one with highest performance (Logistic Regression) based on data of the first approach.

Next, in order to identify an optimal approach to predict patient survival, we also investigate two logistic regression model (LRM) generated imaging markers. The first one fuses image features computed from one CT slice and the second one fuses the weighted average image features computed from multiple CT slices. The first LRM mimics the method used in current clinical practice based on RECIST guidelines, which only select and analyze one 2D image slice that across the center or maximum region of the tumor volume. In second approach, a quasi-SD image features are generated and used. One image feature in each CT slices will multiple the size of the tumor region. Then, each quasi-3D feature is computed by averaging all weighted feature values in all involved CT image slices in which tumor regions are detected and segmented.

Two LRMs are trained and tested using a standard leave-one-case-out (LOCO) cross-validation method. Since there 406 cases in our dataset, in each training and testing iteration cycle of each LRM, 405 cases are used to train the classifier and one remained case is used to test the classifier. By repeating this process 406 times, each case serves as an independent testing case once and the classifier generates 406 testing or prediction scores. These 406 LRM-generated prediction scores are then analyzed using a receiving operating characteristics (ROC) type data analysis method. As a result, a ROC curve is computed for each LRM and the area under ROC curve (AUC) is used as index to evaluate performance of the LRM in predicting patients' survival. Additionally, since the optimized LRM image marker generates a classification score ranging from 0 to 1, we applied an operation threshold ($T = 0.5$) to LRM-generated classification scores to divide all testing cases into two class of survival and not survival. From the numbers of testing cases in two classes, we compute and create confusion matrix. From the confusion matrices, the classification accuracy, sensitivity and specificity are computed and compared.

3. RESULTS

In this study, the tumor segmentation results of all testing cases are visually examined. Although our graphic user interface (GUI) of the CAD tool has installed manual correction functions that allow the users to correct the segmentation errors if detected, in this study we do not detect any major or significant tumor segmentation errors we thus accepted all automated tumor segmentation results. As a result, all radiomics features used to train and test LRM-based classifiers are computed based on the automated segmentation results. Figures 2 and 3 illustrate ROC curves and confusion matrices of different LRM-based image markers generated using different types (i.e., 2D and 3D) of radiomics features. Based on the confusion matrices, the summarized classification performance results of different LRM image markers and/or approaches are shown in Table 1. From the summarized classification performance data as shown in Table 1, we observed that radiomic features computed from CT images can contain valuable discriminatory information to predict survival of gastric cancer patients.

To test for a statistically significant difference between the performance of each approach, we also used ROCKIT Software (http://metz-roc.uchicago.edu/MetzROC/software) to compare the difference of AUC values of ROC curve in means between two approaches. ROC data analysis results show that the first LRM-based classifier trained using 2D CT image features yields an AUC = 0.70 with a 95% confidence interval (CI) of [0.65, 0.75] to predict overall survival of these 406 gastric cancer patients. Meanwhile, the second LRM-based classifier an AUC = 0.72 with a 95% CI of [0.66, 0.76]. The approximate 95% confidence interval for the difference between single slice and multiple slices approach is within [-.0373, .0156] with p-value = 0.2103 which is greater than 0.05. Thus, the result indicates that there is no significant improvement between those two approaches.

Table 1: Average ROC AUC, Accuracy, Sensitivity, and Specificity using Leave-One-Out Cross-Validation.

	Model	Accuracy	Sensitivity	Specificit y	AUC
Single Slice	SVM	0.641	0.381	0.828	0.653
	ANN	0.65	0.47	0.777	0.679
	LGM	0.667	0.5	0.786	0.704
Multiple Slices	LGM	0.655	0.423	0.819	0.717

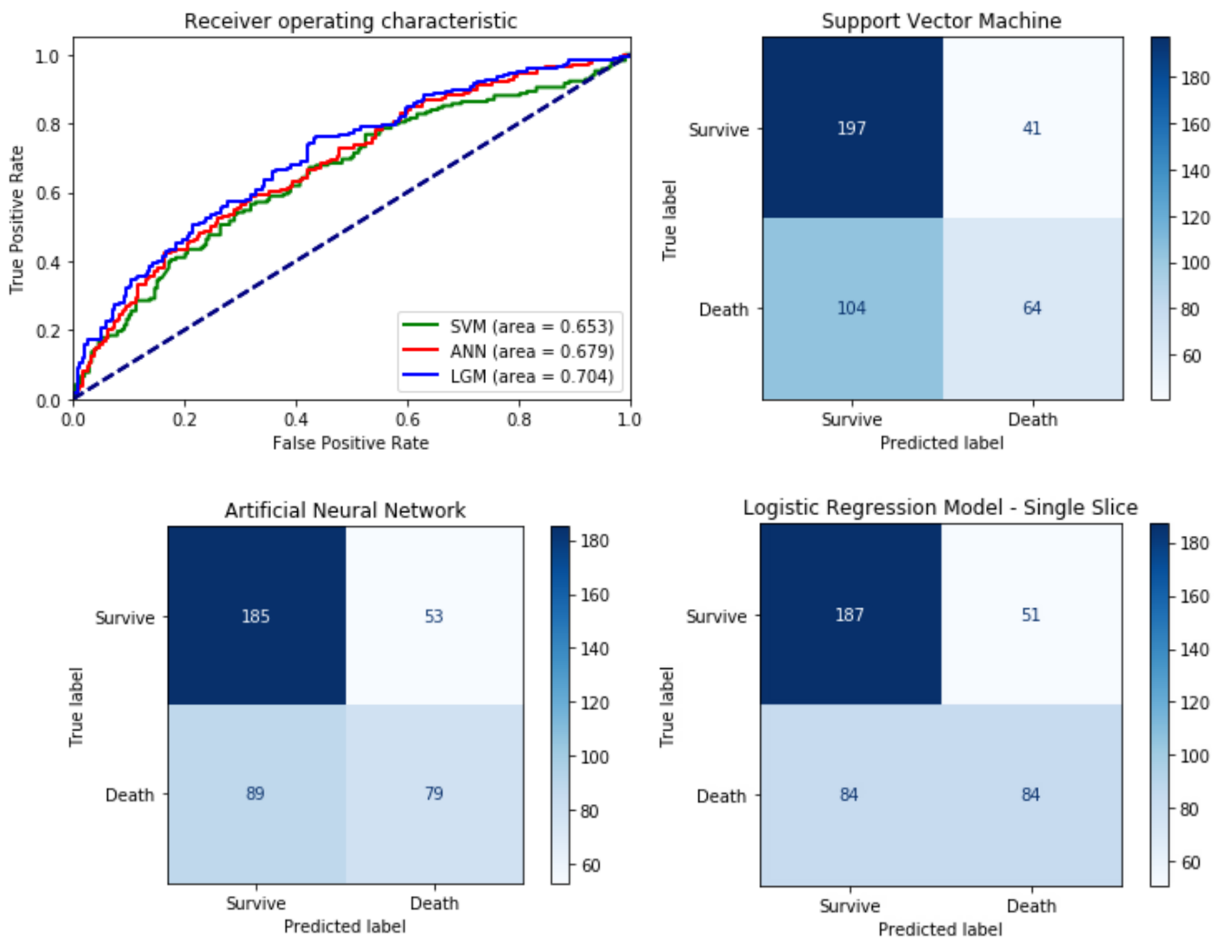

Figure 2: Illustration of ROC Curves and confusion matrices for all 3 classification models. Confusion matrices are calculated based on Leave-One-Out Cross-Validation.

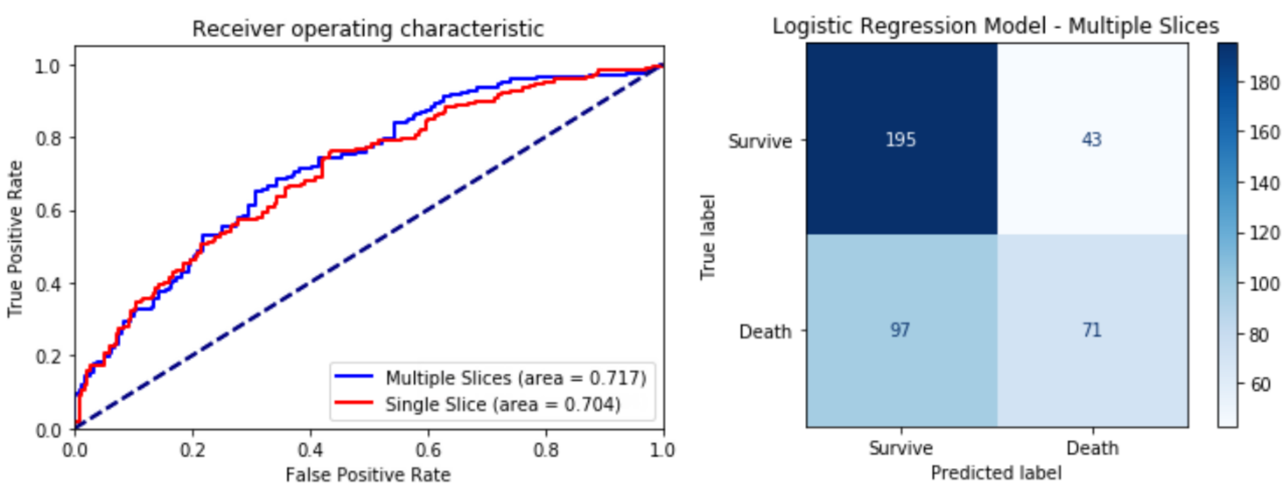

Figure 3: Illustration of ROC Curves and confusion matrices comparation between two approaches.

4. DISCUSSION

Although number of studies have been published in the recent literature [5-7, 12] to develop CAD schemes of radiomics features aiming to help diagnosis of gastric cancer and prediction of patients' prognosis or survival, we investigated a different approach and conducted a unique proof-of-concept preliminary study. The contribution of this new study to CAD or medical imaging informatics research field include:

1. We investigated a new application topic of applying a new quantitative image marker generated by a radiomics-based machine learning model or classifier to predict overall survival of gastric cancer patients using diagnostic CT images. To the best of our knowledge, no study in the similar topic has been reported in the literature.
2. Although previous radiomics studies have indicated that it is possible to identify a group of radiomics features that highly associate with tumor prognosis or response to the specific treatment, this study demonstrates that radiomics features computed from CT images also carry valuable discriminatory information to predict overall survival of gastric cancer patients.
3. Logistic Regression Model can achieve higher performance in predicting overall survival of gastric cancer patients in compare with commonly used models such as support vector machine and artificial neural networks.
4. Based on RECIST guidelines used in current clinical practice, large number of previous radiomics studies used image features computed from one image slice that crosses the center or maximum size region of the tumor volume. This study compares the performance of using conventional 2D and new quasi-3D image features to develop the same machine learning models to predict overall survival of gastric cancer patients. The comparison is made using the same dataset and the same LOCO cross-validation method. We observe that involving multiple image slices to build quasi-3D features can extract higher discriminatory power but not statistically significant.

We also recognize that this preliminary study has several limitations. For example, (1) we only use and test three popular models namely, logistic regression, artificial neural networks, and support vector machine in this study, (2) we only use a simple tumor region size weighted method to generate quasi-3D features, and (3) we only use one feature selection library in Pyradiomics software to compute 103 features and use one method of PCA to reduce the feature dimensionality. Despite of these limitations, the study results are encouraging, which fully supports our study hypothesis and the feasibility of identifying a new radiomics image marker to help predict overall survival of gastric cancer patients. Once combined with other historical clinical data of patients, the performance could be improved. Based on foundation established in this preliminary study, we will continue our studies including but not limited to (1) test more effective machine learning models including applying advanced deep learning models to improve tumor segmentation [13, 14], (2) identify optimal method to integrate features computed from multiple CT image slices to produce better 3D features, (3) extracting more radiomics features and apply more effective machine learning method to identify more clinically relevant and non-redundant image features. Overall, our long-term goal is to further improve performance or accuracy of new quantitative image markers to help more accurately predict overall survival of gastric patients in future studies.

5. ACKNOWLEDGMENT

The authors thank the support from the pilot project funding from the Stephenson Cancer Center, the University of Oklahoma.

6. REFERENCES

*Author to whom correspondence should be addressed: Huong N. Pham (huongnpham@ou.edu).

[1] Bray, F., et al., "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancer in 185 countries," CA Cancer J Clin, 68(6) 394-424 (2018).

[2] Aerts, H., et al., "Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach," Nature Communication, 5, 4006 (2014).

[3] Khuzani, A.Z., et al., "Prediction of chemotherapy response in ovarian cancer patients using a new clustered quantitative image marker," Physics in Medicine and Biology, 63, 155020 (2018).

[4] Emaminejad, N., “Fusion of quantitative image features and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients,” IEEE Transactions on Biomedical Engineering, 63, 1034-1043 (2016).

[5] Kangi, A.K., Bahrampour, A., “Predicting the survival of gastric cancer patients using artificial and Bayesian neural networks,” Asian Pac J Cancer Pre, 19(2), 487-490 (2018).

[6] Charati, J.Y., et al., “Survival prediction of gastric cancer patients by artificial neural network model,” Gastroenterol Hepatol Bed Bench, 11(2), 110-117 (2018).

[7] Wang, L., et al., “CT-based radiomics nomogram for preoperative prediction of No.10 lymph nodes metastasis in advanced proximal gastric cancer,” European Journal of Surgical Oncology, 47(6), 1458-1465 (2021).

[8] G. Danala, et al., “Classification of breast masses using a computer-aided diagnosis scheme of contrast enhanced digital mammograms,” Annals of Biomedical Engineering, 46, 1419-1431 (2018).

[9] S. Mirniaharikandehei, et al., “Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using CT images,” Computer Methods and Programs in Biomedicine, 200, 105937 (2021).

[10] van Griethuysen, J.J.M., et al., “Computational radiomics system to decode the radiographic phenotype,” Cancer Research, 77(21), e104–e107 (2017).

[11] Günaydin, Ö., et al., “Comparison of lung cancer detection algorithms,” 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), 1-4 (2019).

[12] Sun, Z., et al., “Preoperative prediction for Lauren type of gastric cancer: A radiomics nomogram analysis based on CT images and clinical features,” Journal of X-ray Science and Technology, 29, 675-686 (2021).

[13] Yang, T., et al., “DCU-Net: Multi-scale U-Net from brain tumor segmentation,” Journal of X-ray Science and Technology, 28, 709-726 (2020).

[14] Shi, T., et al., “A stacked generalized U-shape network based on zoom strategy and its application in biomedical image segmentation,” Computer Methods and Programs in Biomedicine, 197, 105678 (2020).